Time to Diagnosis and Persistence: The Two Major Determinants of Effective Tuberculosis Control

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Biological Sciences

at

The University of Waikato

by

ALAINE TESS RUTHE

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2015

Abstract

The greatest challenge confronting effective tuberculosis (TB) eradication is the time to diagnosis, and duration of treatment of chronically infected individuals which represent a pool of infection. In an attempt to help limit the spread of TB in New Zealand, a fast SNP based diagnostic test was developed, to quickly identify the highly transmissible and virulent endemic Rangipo strain. The role of VapBC toxin-antitoxin systems in *M. tuberculosis* has been the subject of great interest recently, due to their expanded number in the genome and links with virulence and the regulation of cell growth in response to environmental stress. Their ability to regulate growth under adverse conditions for presumed survival advantages possibly leading to dormancy or persistence, make them ideal candidates for the development of new *M. tuberculosis* treatments. To establish differential expression of *vapC*, and therefore identify possible pathways and functions of the VapBC proteins, RT-qPCR was used to assess the expression levels of *vapB* and *vapC* in *M. smegmatis* under conditions of stress. No consistant changes in *vapC* mRNA levels were observed, resulting in the hypothesis that it is not the transcriptional differences which are important in the regulation of VapC, but post-transcriptional factors. In order to investigate the function(s) of *M. tuberculosis* VapBCs, these VapBC proteins were expressed and purified in *M. smegmatis*, and the VapC toxin tested for RNase activity. The purification, expression, RNase testing and bioinformatic analysis of *M. tuberculosis* VapCs suggested that $VapC_{Rv2530c}$, $VapC_{Rv0065}$ and $VapC_{Rv0617}$ may all target the same recognition sequence, UA*GG. Bioinformatic analysis revealed an abundance of this target sequence in horizontal gene transfer and TA genes, raising the possibility that VapC toxins could be functioning as selfish elements, or initiating transcriptional regulation cascades when a rapid change in the proteomic response and metabolic state of the cell is required. It is intriguing that the three *M. tuberculosis* VapC proteins tested thus far appear to target the same recognition sequence, possibly suggesting that all 47 VapCs are RNases and are targeting the same sequence. Alternatively; VapCs may belong to sub-groups targeting different sequences, allowing *M. tuberculosis* to exude both gross and fine metabolic control; or, they may share the same target, but are regulated by different activators triggered in response to different environmental stimuli.

Acknowledgements

The kind efforts of numerous people over the years have enabled me to get here, here are but a few of those I could not have done it without.

I would like to thank my supervisors Professor Vic Arcus and Dr Ray Cursons for their ongoing guidance, support and patience. I feel incrediably lucky to have beared witness to their pragmatism, encouragement and distain for mediocrity.

To everyone in the labs who have been there, thank you. I know I am lucky to have undergone this enduring process with such great people. No where else is good food, laughter and escapism (even if it is by dumpster diving) an every day occurance during your PhD. Judith - your no-love (or sugar)-spared baking has fuelled many a busy day at the bench, Greg and Olivia, well done on keeping me sane, The Jo's - your critiquing and support have been invaluable, and to everyone for putting up with watching my children grow (yes, this has been a while in the making). Thank you also to the Health Research Council, Waikato Medical Research Foundation and Trust Waikato for funding.

And last but by no means least, my family. Thank you being there and being proud of me, and for inspiring in me the magic of life as well as science. To my boys for their continual schooling and making me rich in ways I never knew the importance of, and Dan, thank you for making this possible and for not being acrimonious (to the point at which I noticed).

That's how you devour a whale, one bite at a time. F. Underwood 2013

Table of Contents

List of Figures

List of Tables

List of Abbreviations

SI (Systeme Internationale d‟Unités) abbreviations for units and standard notations for chemical elements and formulae are used throughout this thesis. Other abbreviations are listed below.

A	adenosine
ACN	acetonitrile
ADC	albumin dextrose catalase
ATP	adenosine triphosphate
bp	base pair(s)
BLAST	basic local alignment search tool
C	cytosine
Cq	quantification cycle
CTAB	hexadecyltrimethylammonium bromide
C-terminal	carboxy terminus
DEPC	diethyl pyruvate carbonate
DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
doc	death on curing
ds	double-stranded
DTT	dithiothreitol
EDTA	ethylene diamine tetraacetic acid (disodium salt)
EMSA	electrophoretic mobility shift assay
EtBr	ethidium bromide
fit	fast intracellular trafficking
g	times the force of gravity
G	guanine
GITC	guanidium isothiocyanate
GOI	genes of interest

HEPES	N-2-hydroxyethylpiperazine-N‟-2-ethanesulphonic acid
HGT	horizontal gene transfer
His-tag	poly-histidine tag
HIV	human immunodeficiency virus
HPA	hydroxypicolinic acid
HPLC	High-performance liquid chromatography
Hyg^r	hygromycin B resistance
IEF	isoelectric focussing
IMAC	immobilised metal affinity chromatography
IPTG	isopropylthio-β-D-galactosidase
IR	inverted repeat
kb	kilobase
kDa	kilodalton
Km^r	kanamycin resistance
LB	luria bertani
LSP	large sequence polymorphism
MAD	multiple wavelength anomalous diffraction
MALDI	matrix assisted laser desorption ionisation
mAU	milli-absorbance units
Mb	megabase
MDR	multidrug resistant
MES	2-(N-morpholino)-ethanesulfonic acid
MESH	medical subject headings
MIRU	mycobacterial interspersed repetitive units
MIQE	Minimum Information for Publication of Quantitative Real-Time PCR Experiments
ML	mother liquor
MPD	2-methyl 2, 4-pentanediol
mRNA	messenger RNA
MTBC	Mycobacterium tuberculosis complex
MS	mass spectrometry

MQ	milli Q water – ion exchanged purified water
m/z	mass to charge ratio
Native-PAGE	non-denaturing PAGE
N-terminal	amino terminus
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
par	partitioning
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pI	isoelectric point
PIN	PilT N-terminal domain
ppGpp	guanosine tetraphosphate
qPCR	real-time polymerase chain reaction
RD	region of difference
RFLP	restriction fragment length polymorphism
RHH	ribbon helix helix
rpm	revolutions per minute
RNA-Seq	RNA sequencing
RNase	ribonuclease
rNTP	ribonucleotide
rRNA	ribosomal RNA
rrs	16S ribosomal ribonucleic acid (rRNA) gene
RT	reverse transcriptase/transcription
RT-qPCR	reverse transcription quantitative real-time polymerase chain reaction
qPCR	quantitative real-time polymerase chain reaction
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulphate
ss	single stranded
T	thymine

TA	toxin antitoxin
TAE	tris-acetate-EDTA
TB	tuberculosis
TBDB	TB database
TE	tris EDTA buffer
TEMED	N, N, N, N,-tetramethylethylenediamine
TOF	time of flight
U	uracil
UTR	untranslated region
UV	ultra violet
Vap	virulence associated protein
VNTR	variable number tandem repeats
v/v	volume per volume
WGS	whole genome sequencing
WHO	world health organisation
wt	wild type
w/v	weight per volume
XDR	extensively drug resistant

Chapter One:

An Introduction to Tuberculosis and Toxin-Antitoxin systems

1.1 The Pathology of Disease

Tuberculosis (TB) is an ancient disease responsible for the death of 1.3 million people in 2012 (WHO 2013). Approximately one-third of the world's population are thought to be infected with the tubercle bacilli, though only 5 – 10 % of infected individuals will develop active TB disease during their lives, in 2012 there were an estimated 8.6 million cases of active TB (WHO 2013). Most infected individuals will be asymptomatic, experiencing latent tuberculosis infection (LTBI), from which reactivation may occur at a later date, except in certain geographical areas where the risk of developing active disease is greatly increased due to HIV co-infection (Harries & Dye 2006).

The rate of TB incidence at country level varies enormously, as illustrated in Figure 1.1. In 2012, 58 % of the world's TB cases occurred in the South-East Asia and Western Pacific regions, and approximately one quarter were from the African region. While India and China had the largest number of cases (26 % and 12 % of the global total, respectively), the African region contained the highest numbers of cases and deaths relative to population size (averaging 255 incident cases per 100,000, approximately double the global average of 122), with South Africa and Swaziland having the highest incidence rate per capita (approximately

1,000 per 100,000) (WHO 2013). The 2012 TB incidence rate in New Zealand (7.6 per 100,000) was similar to other developed regions in parts of the Americas, several countries in western Europe, Japan and Australia, which all had fewer than 10 per 100,000 population (WHO 2013).

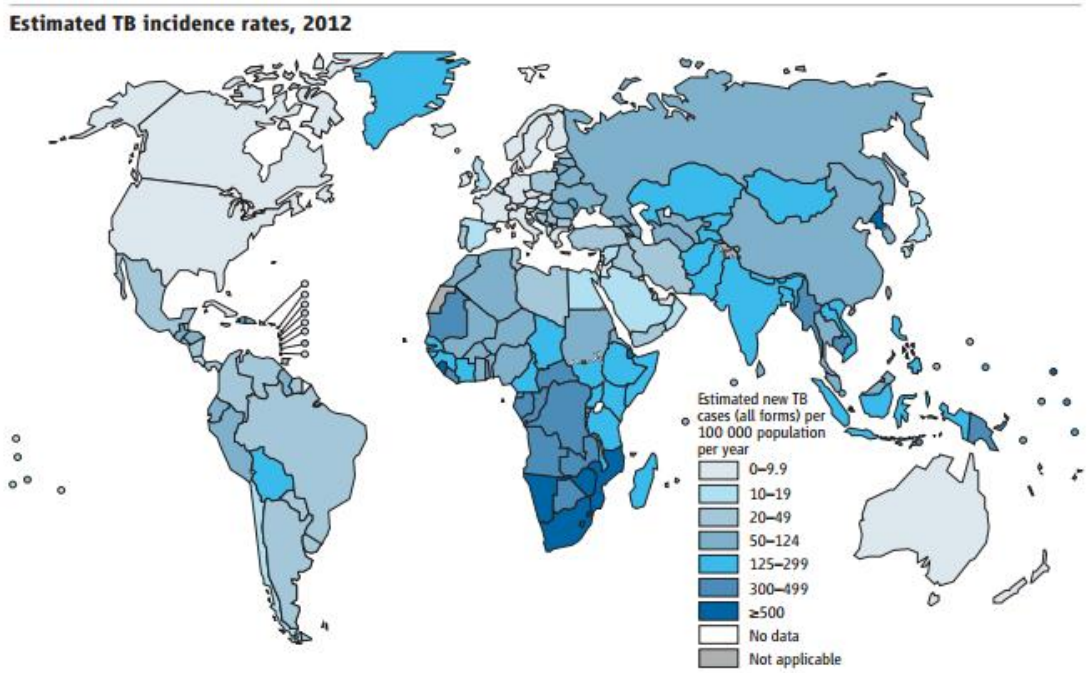


Figure 1.1: Global estimated TB incidence rates for 2012. Figure obtained from the WHO Global Tuberculosis Report 2013.

There are many factors which influence a person's susceptibility to TB disease including poor living conditions, drug and alcohol abuse and immunosuppression, but non more so than the HIV pandemic. HIV infected individuals accounted for an estimated 13 % of the 8.6 million cases of TB in 2012 (with approximately 75 % of those occuring in the African Region) (WHO 2013), and are more than 20 times more likely to reactivate TB than HIV negative individuals (Girardi *et al.* 2000).

TB is an airborne disease which typically affects the lungs (pulmonary), but can infect most sites in the body (extrapulmonary) such as the central nervous system, lymph nodes, urinary tract, bones and joints. TB is caused primarily by the bacterium, *Mycobacterium tuberculosis*. *M. tuberculosis* is a member of the *M. tuberculosis* complex (MTBC), along with the very

closely related mycobacterial species *M. bovis*, *M. africanum*, *M. microti*, *M. caprae*, *M. pinnipedii*, *M. canetti* and *M. mungi* which all cause tuberculosis in humans or animals.

1.2 Organism

M. tuberculosis is a Gram-positive, rod-shaped, non-motile aerobe of the order Actinomycetales. This bacterium has a characteristic hydrophobic cell wall which renders it acid-fast, and hinders the transfer of nutrients and waste into and out of the cell resulting in an extremely slow generation time of 15 - 20 hours. The impressively thick and complex cell wall (Figure 1.2) is composed of an inner plasma membrane, a unique large cell wall core, and an outer layer or capsule. The cell wall core structure contains layers of covalently linked peptidoglycan, arabinogalactan and mycolic acids which result in a hydrophobic permeability barrier making the mycobacteria resistant to detergents and a wide variety of antimicrobial agents. The outer part of the cell-wall core contains multiple free lipids including cord factor, and is covered by the capsule which consists mainly of polysaccharides (Brennan 2003; Zuber *et al.* 2008).

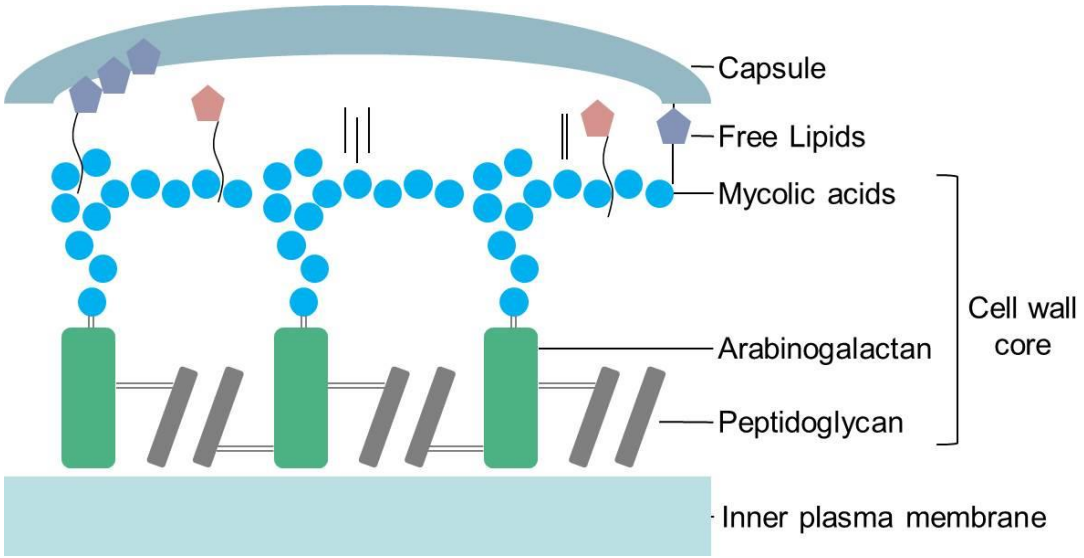


Figure 1.2: Schematic of the *M. tuberculosis* cell wall structure. The cell wall is composed mainly of a large cell-wall core flanked on either side by an inner membrane, and a polysaccharide based outer capsule. The cell wall core is comprised of three covalently linked structures: peptidoglycan (grey), arabinogalactan (green) and mycolic acids (blue), covered by a layer of multiple free lipids including cord factor.

The bacilli vary in their size and shape from coccobacilli to long rods of 1 - 10 µm in length (generally 3 - 5 µm), and 0.2 - 0.6 µm in width (Velayati *et al.* 2009). Morphological variations such as: round, oval, V and Y-shape; stunted or extremely long bacilli; or bacilli with extra-ordinary thick cell walls occur under conditions of stress such as starvation, hypoxia and old age, as well as in un-stressed but drug resistant strains (Ntolosi *et al.* 2001; Velayati *et al.* 2009; Farnia *et al.* 2010; Shleeva *et al.* 2011).

Initial infection is caused by inhalation of bacilli which have been aerosolized by an infected person. The bacilli are engulfed by alveolar macrophages from where the majority of the bacilli are destroyed or inhibited. Some bacilli remain however, replicating intracellularly and are released when the macrophages die, spreading by way of the lymphatic system or bloodstream to other organs. Activation of the immune response however, slows replication of the bacilli within macrophages and in turn the bacilli become walled off within developing granulomas where they can persist for long periods of time. This is referred to as latent tuberculosis infection (LTBI). If a host's immune system is compromised (by factors such as old age, malnutrition or co-infection with HIV), containment fails and the granulomas progress towards disease, becoming necrotic and eventually collapsing, in turn releasing the infectious bacilli into the airways (Russell 2007).

Globally, there has been an increase in the number of cases of multidrug-resistant tuberculosis (MDR-TB), which have a much higher morbidity outcome than non-MDR-TB. In 2012 the morbidity outcome for MDR-TB cases was 37 % with an estimated 450,000 cases of MDR-TB and 170,000 deaths (WHO 2013) compared to 14.6 % for non-MDR-TB.

1.3 Evolution of the *M. tuberculosis* genome

The first Mycobacterium isolated was *Mycobacterium leprae*, and was done so almost 150 years ago by Hansen, eight years before Robert Koch discovered the tubercle bacillus, *M. tuberculosis* in 1882. It was originally thought all mammalian tubercle bacilli originated from the same source, until in 1898 when Theobald Smith showed that bacilli isolated from humans were different morphologically to those isolated from cattle, and those isolated from swine,

horse and cat were “either identical with the bovine bacilli or very closely resembled them” (Smith 1898). *M. tuberculosis* was later designated as the typical human pathogen and *M. bovis* the animal pathogen, which is now believed to consist of a series of host-adapted ecotypes (Smith *et al.* 2006a; Smith *et al.* 2006b) affecting animals such as seals (*M. pinnipedii*), mongooses (*M. mungi*), voles (*M. microti*) and goats (*M. caprae*).

The MTBC is comprised of the typical human-associated pathogens *M. tuberculosis* and *M. africanum*, the so-called ‘smooth TB bacilli’ (including *M. canettii*), and the host-adapted ecotype lineages including *M. bovis*, *M. microti*, *M. caprae*, *M. orygis*, and *M. pinnipedii*. While *M. tuberculosis* has been isolated from other mammals, it has not been shown to be able to sustain its population (by animal-to-animal transmission) in any of these species, and similarly, while *M. bovis* can cause infection in humans, it does not readily transmit between them (Gagneux 2012). It was originally assumed that human TB strains had evolved from an animal origin (*M. bovis*) due to the expanded host range of the animal-associated MTBC. However, comparative genomic analyses has shown that the animal MTBC strains fit within the more diverse human MTBC strains, advocating the alternative scenario of animal MTBC strains evolving from a human origin (Gagneux *et al.* 2006; Smith *et al.* 2006b; Coscolla *et al.* 2013).

One unique member of the MTBC is *M. canettii*. Along with the other ‘smooth TB bacilli’, *M. canettii* produces smooth and shiny colonies (not rough, characteristic of the rest of the MTBC) and is more genetically diverse. Unusually, *M. canettii* shows evidence of recombination via horizontal gene transfer - a trait considered not to occur for the rest of the *M. tuberculosis* complex (Supply *et al.* 2003; Hirsh *et al.* 2004). Recombination was proposed as one possible explanation for an unusually high synonymous divergence rate observed between *M. tuberculosis* strains in one study, and homoplasies present within the intergenic regions in another study analysing single-nucleotide mutations (Gutacker *et al.* 2006), however these occurances on their own were not considered definitive evidence of recombination within strains belonging to the MTBC (Smith *et al.* 2006a), so as such *M. tuberculosis* is considered clonal and not to undergo recombination. A consequence of clonality is reduced diversity. Despite the MTBC adapting to different host species, these mycobacteria are extremely closely related displaying 99.9 % similarity at the nucleotide level, and possessing almost identical

16S rRNA sequences (Brosch *et al.* 2002). In fact, the MTBC bacilli exhibit such low sequence diversity they have been referred to as 'genetically monomorphic' along with other pathogenic bacteria such as *Bacillus anthracis* (anthrax)*, Yersinia pestis* (the plague), and *Salmonella enterica* serovar Typhi (typhoid fever) (Achtman 2008).

H37Rv and H37Ra are 'virulent' (Rv) and 'avirulent' (Ra) *M. tuberculosis* strains derived (in 1934) from the original human-lung H37 strain which was isolated in 1905, and are used extensively worldwide as reference strains (Bifani *et al.* 2000). The first *M. tuberculosis* genome sequenced was H37Rv in 1998 (Cole *et al.* 1998). The H37Rv genome is 4.41 Mb in size, contains approximately 4000 genes, and as with other mycobacteria and closely related Actinomycetales it has an unusually high genomic DNA GC content of 65.6 %.
Another mycobacterium widely used in mycobacteriological studies is *Mycobacterium smegmatis*, a fast-growing opportunistic mycobacteria with a much larger genome of 6.99 Mb. *M. smegmatis* was initially isolated from syphilitic chancres and normal genital secretions in 1885 (Talaat *et al.* 1999) and is now considered an environmental saprophyte (common in soil and water) which rarely causes skin and soft-tissue lesions (Brown-Elliott & Wallace 2002). *M. smegmatis* mc^2155 is a non cell clumping, high transformation efficiency mutant strain originally isolated from its 'wild type' progenitor mc^26, which is a single colony isolate of the *M. smegmatis* reference strain, ATCC 607 (Snapper *et al.* 1990). Because it is non-pathogenic, fast growing and easily transformed *M. smegmatis* mc^2155 is used extensively as a model organism in mycobacteriological studies. The frequent use of *M. smegmatis* mc^2155 as a surrogate host is not always fully considered however, as it also contains an expansion in the copy number of IS1096 (containing 13 extra) as well as a large duplicated fragment (56 kb) and deletions (Wang *et al.* 2008) which may have an effect on the outcome of the study. As long as these differences are taken into consideration, *M. smegmatis* mc^2155 can be immensely valuable in the analysis of mycobacterial gene function, expression and replication.

1.3.1 Virulence

Virulence, by Medical Subject Headings (MeSH) definition, is "the degree of pathogenicity within a group or species of microorganisms or viruses as indicated by case fatality rates

and/or the ability of the organism to invade the tissues of the host. The pathogenic capacity of an organism is determined by its virulence factors". Virulent MTBC species have acquired virulence factors and developed strategies to evade altogether, or alter the host immune response in their favour. Much research is focussing on determining what these virulence factors are (for review, see Forreland *et al*. (2013)), so as to better understand the mechanisms involved, and to aid in the development of improved vaccines and antimicrobial therapies. There is no universally accepted definition of what constitutes a virulence gene, but it is generally considered that its loss reduces the virulence of the pathogen, without affecting the growth rate.

Genes or proteins associated with virulence have been identified in the following categories: lipid and fatty acid metabolism, cell envelope proteins, proteins inhibiting antimicrobial effectors of the macrophage, protein kinases, proteases, metal-transporter proteins, gene expression regulators, proteins of unknown function, and other virulence proteins (Forrellad *et al.* 2013). An area of particular interest is virulence factors which are involved in the interaction of *M. tuberculosis* with the host macrophage. Many mycobacterial species have adapted to survive and replicate inside amoebae and macrophages (Zhang *et al.* 1999; Greub & Raoult 2004), which is one of the reasons *M. tuberculosis* is able to persist inside its host for such extended periods. Human alveolar macrophages (which are the primary host cell for inhaled *M. tuberculosis*) infected with *M. tuberculosis* undergo apoptosis *in vitro* (Keane *et al.* 1997; Danelishvili *et al.* 2003). This cell death removes an environment conducive for bacterial growth, therefore limiting the growth of the intracellular pathogen and ultimately benefiting the host. It has been demonstrated however, that there are mycobacterial virulence determinants which reduce the apoptotic response of alveolar macrophages to intracellular infection, in order to evade this host defense mechanism (Keane *et al.* 2000; Danelishvili *et al.* 2003). This negative correlation between mycobacterial virulence and the ability to induce apoptosis explains why the majority of *M. tuberculosis* infections cause necrosis, while attenuated mutant infections mainly induce apoptosis (Forrellad *et al.* 2013). Virulence factors have also been implicated in the switching and regulation of carbon metabolism which must occur when *M. tuberculosis* resides inside the nutrient poor host macrophage (Forrellad *et al.* 2013). The toxin–antitoxin (TA) genes were identified during the course of this research as playing an

important role in the regulation of carbon metabolism in mycobacteria (McKenzie *et al.* 2012b). These genes have also been linked with virulence (Arcus *et al.* 2005).

1.4 Toxin-Antitoxin Systems

1.4.1 History and Function of Toxin-Antitoxin systems

TA systems were originally discovered in the 1980's on plasmids, where they acted to maintain the plasmids by post-segregational killing of plasmid-free cells (Gerdes *et al.* 1986). They have since been found to be highly abundant on the chromosomes and plasmids of Eubacteria and Archaea (Pandey & Gerdes 2005; Fozo *et al.* 2010). A TA system is comprised of two genes in an operon: one encodes a stable toxic protein, and the other a labile but neutralising antitoxin. Due to the relative stability of the toxin compared with the antitoxin, the antitoxin must be continually expressed to maintain its neutralising effect on the toxin and prevent cell stasis or death.

Early research on TA systems was predominantly performed with *Escherichia coli* where their function has been linked to the response to environmental stress, programmed cell death and bacterial persistence as well as plasmid and gene stabilisation (Korch *et al.* 2003; Hazan *et al.* 2004; Gerdes *et al.* 2005; Szekeres *et al.* 2007; Maisonneuve *et al.* 2011). It has also been suggested that TA loci may simply be selfish genetic elements with no biological function, functioning just to maintain their presence in a genome (Mine *et al.* 2009). The role of TA systems in other organisms, however, has not been studied as intensively and many questions remain over their physiological function in these microbes, especially *M. tuberculosis*.

TA systems, while not generally essential for normal cell growth, are widely prevalent in bacterial pathogens, often in expanded numbers and closely linked to mobile genetic elements (Pandey & Gerdes 2005). The number of TA systems in *M. tuberculosis* are massively expanded, suggesting they perform roles which are advantageous for cell survival in its many and varied environments, and may be involved in *M. tuberculosis* pathogenesis (Ramage *et al.* 2009).

There are five types of TA systems, classified by the nature and mode of action of the antitoxin: type I and III where the antitoxins are small RNAs, and types II, IV and V which utilise protein antitoxins (the toxins are proteins for all TA system types). Type I antitoxin RNAs function by binding to toxin-encoding mRNA, thereby suppressing toxin translation, while type III antitoxins inhibit their cognate protein by binding to it directly (Fineran *et al.* 2009). Type II antitoxin proteins bind directly to the toxin causing its inactivation, type IV antitoxin proteins compete with the toxin for the same binding target, and the type V antitoxin proteins cleave the toxin-encoding mRNA (Short *et al.* 2013). Type II TA systems were the first TA systems discovered and are the most prevalent (Yamaguchi & Inouye 2011). They are believed to move from one genome to another by horizontal gene transfer (Pandey & Gerdes 2005) and are the focus of research in this thesis.

Type II TA systems were initially classified into TA system families but bioinfomatic approaches have since predicted many more novel putative families of toxins and antitoxins, and examples of 'hybrid' TAs (where the toxin belongs to one family and the antitoxin to another) have since been characterised (Schmidt *et al.* 2007). Therefore, the practice is to now classify the antitoxin and toxin families independently (Leplae *et al.* 2011). Toxins belonging to type II TA systems have been sub-divided into at least 12 super-families based on similarities of the toxin sequences and three-dimensional structure prediction, and antitoxins into 20 super-families (Leplae *et al.* 2011). Most of the characterised type II toxins are translation inhibitors (Doc, HicA, HipA, MazF, RelE, VapC, YafO and GinA,B,C and D), although there are a number of exceptions: ζ toxins inhibit peptidoglycan synthesis, and CcdB and ParE toxins inhibit chromosome replication through inhibition of DNA gyrase (Korch & Hill 2006; Hayes & Van Melderen 2011). Table 1.1 summarises the type II toxin super-families and their targets. Phylogenetic distribution of the toxin super-families varies from one super-family to another. ParE/RelE, Zeta, VapC, Doc, CcdB/MazF, HipA and GinB are all distributed across a wide range of bacterial phyla (Proteobacteria, Firmicutes (except GinB), Cyanobacteria and Actinobacteria), while other toxin super-families are only present in one or two phyla (Leplae *et al.* 2011).

Table 1.1: The super-families of type II toxins, their targets, activities and cellular process they affect.

Super-family	Target	Activity	Cellular process
CcdB/MazF	DNA gyrase/mRNAs	Generates DS breaks/ Endoribonuclease	Replication/Translation
Doc	Translating ribosome	Induces mRNA cleavage	Translation
Gin families	ND	ND	Translation
HipA	EF-Tu	Protein kinase	Translation
HicA*	ND	mRNA cleavage	Translation
ParE/RelE	DNA gyrase/Translating ribosome	Generates DS breaks/ Induces mRNA cleavage	Replication/Translation
VapC	mRNAs	Endoribonuclease	Translation
VapD	ND	ND	ND
YafO	ND	mRNA cleavage**	Translation
ζ	ND	Phosphotransferase	Peptidoglycan synthesis

Table adapted from (Van Melderen & Saavedra De Bast 2009) and (Hayes & Van Melderen 2011), * = (Jørgensen *et al.* 2009), ** = (Zhang *et al.* 2009). Abbreviations: ND, not determined; Ccd, coupled cell division; Doc, death on curing; Gin, growth inhibition; Hic, *Haemophilus influenzae* (*hif*) contiguous; Hip, high persistence; Par, partitioning; Vap, virulence associated protein. ParE/RelE includes the HigB, PasB, StbE, Txe, YoeB and YafQ subfamilies, and CcdB/MazF includes the YdcE, PemK and ChpBK subfamilies (Leplae *et al.* 2011).

Of the type II TA systems, RelBE and MazEF have been studied in the most detail. Both the RelE and MazF toxins inhibit translation by cleaving mRNA, as does VapC, however only RelE requires the ribosome for RNA cleavage (Pedersen *et al.* 2003; Zhang *et al.* 2003; Arcus *et al.* 2004; Daines *et al.* 2007). VapBC is the largest of the type II toxin and antitoxin superfamilies (Gerdes *et al.* 2005; Van Melderen & Saavedra De Bast 2009), but the least well characterised. *vapBC* loci account for approximately one third of TA loci identified in a study involving 900 prokaryotic genomes (Shao *et al.* 2011), and are over-represented in the *M. tuberculosis* genome where they account for 47 of the (at least) 88 type II TA loci (Ramage *et al.* 2009; Ahidjo *et al.* 2011). So why are there so many TA systems, in particular *vapBCs*, in *M. tuberculosis*? The majority of TA systems found in *M. tuberculosis* are only found in other pathogenic mycobacteria (Ramage *et al.* 2009), implying they play a fundamental role in the diversification and evolution of the pathogenic *M. tuberculosis* species from its non-pathogenic relatives. One theory proposed is that obligate host-associated organisms do not keep their TA systems, while they are beneficial to free-living organisms. This trend is supported by

Mycobacterium leprae (an obligate intracellular pathogen) having no functional TA loci and *M. tuberculosis* (which has extra- and intracellular growth phases) having many. Similarly the free-living spirochetes *Leptospira interrogans* and *Treponema denticola* have five and 33 TA systems, respectively, while the obligate parasitic spirochetes *Treponema pallidum* and *Borrelia burgdorferi* have none (Pandey & Gerdes 2005; Mitchell *et al.* 2010). Another trend observed by Pandey and Gerdes (2005) was that almost all free-living bacteria living in varied and changing environments contained TA systems as well as having a propensity for dormancy or slow growth, so they postulated that the number of TA loci may be correlated with the cell growth rate. This is supported by *M. tuberculosis* having large numbers of TA systems but a slow growth rate, while the related but non-pathogenic species, *M. smegmatis* has only three TA systems (Robson *et al.* 2009) and is relatively fast growing. The function of TA systems in microbial physiology is still unknown and remains the subject of great debate, although the ability of pathogenic bacteria to slow or regulate growth in order to survive in stressful environments is an appealing hypothesis for the function of TA systems.

1.4.2 Mechanism of Action and the Regulation of Type II TA Systems

The toxin and antitoxin genes constituting the type II TA systems are small in all prokaryotic species. The antitoxins are usually shorter than their partnered toxin (approximately 210 to 450 and 270 to 900 bp long respectively (Pandey & Gerdes 2005)) and are located upstream of the toxin (Yamaguchi & Inouye 2011). Exceptions to this general arrangement are the *higBA, hicAB* and *mqsRA* loci where they are in the reverse order, and the HigBA toxin which is unusually large (1320 bp) (Tian *et al.* 1996; Brown *et al.* 2009; Gerdes 2012). The TA genes are arranged in an operon with the start and stop codons usually overlapping by a few bases (reflecting translational coupling which is common in prokaryotes), although sometimes instead with a spacer region of up to 30 nucleotides (Sevin & Barloy-Hubler 2007; Goulard *et al.* 2010). The toxin and its cognate antitoxin form a stable TA complex, which inhibits toxicity. Generally antitoxin proteins contain unstructured regions and, as such, are unstable and more prone to proteolysis by cellular proteases than the toxins. Therefore, an antitoxin must be constantly expressed to neutralise the effect of the toxin which is most likely why antitoxins are

arranged (and therefore produced) first in the TA operon. In cases where the antitoxin is located downstream of the toxin, the antitoxin and toxin levels are balanced by rare start codons or an additional promoter. Antitoxins are subject to degradation by one of the cytosolic ATP-dependent proteases (Lon, ClpPX or ClpPA), assumed to be induced by various environmental stresses (Christensen *et al.* 2003; Buts *et al.* 2005; Gerdes *et al.* 2005; Wang & Wood 2011). Dissociation of the TA complex results in the release of free toxin which is then able to bind a divalent metal ion, leading to metal-ion dependent nuclease activity which slows the metabolism of the cell ultimately leading to cell growth arrest (Bunker *et al.* 2008; Yamaguchi & Inouye 2009).

The TA complex also functions to autoregulate the TA operon via an N-terminal DNA binding domain in the antitoxin which binds to a repeat sequence in the promoter region, repressing transcription of the TA operon (Yamaguchi & Inouye 2011). Therefore, expression of the TA operon is activated as the antitoxin is degraded and this drives the production of more toxin and antitoxin causing autoregulation of TA expression once again. During this process, free toxin is released into the cell where it is able to act on its molecular target (Figure 1.3).

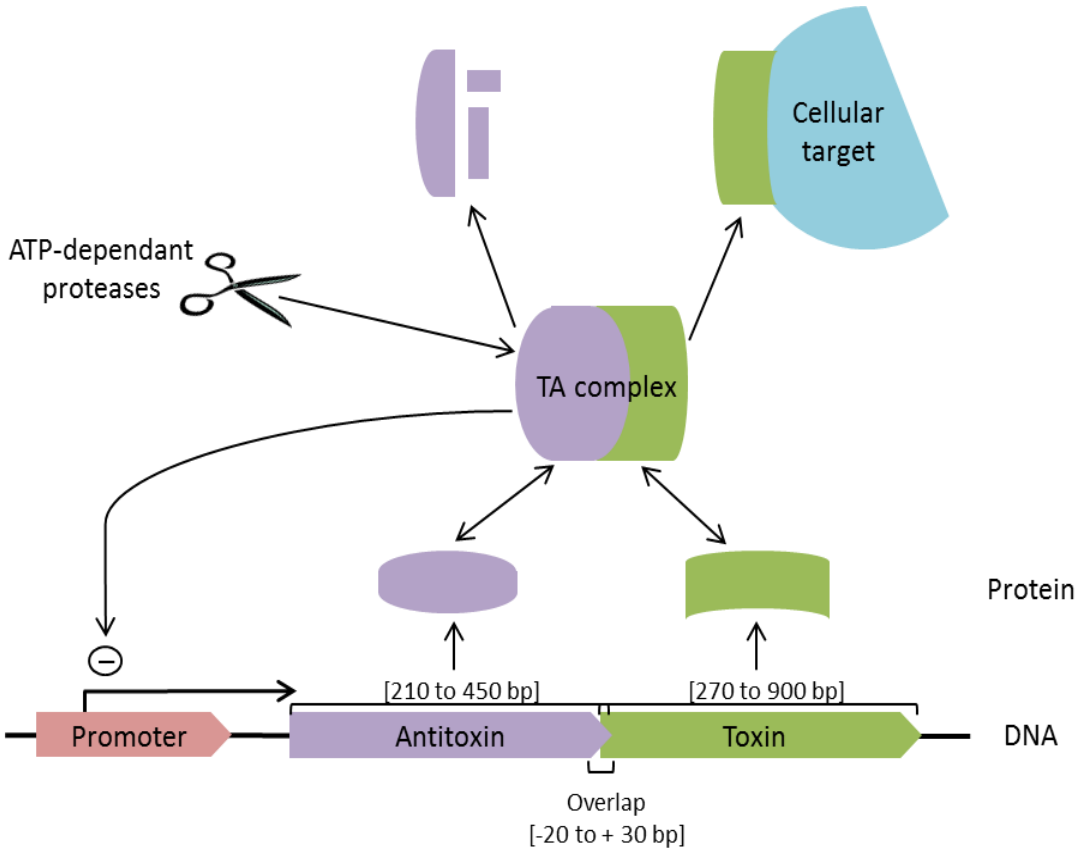


Figure 1.3: General set up of a TA loci. TA loci are organised in an operon, usually overlapping by a few bases but sometimes with a spacer region instead. The stable TA complex binds to the promoter region to autoregulate transcription, while the labile antitoxin is subject to degradation by ATP-dependant proteases leaving the toxin free to act on cellular targets.

The *M. smegmatis vapBC* and *N. gonorrhoeae fitAB* operons have been shown to be subject to auto-regulation through binding of these TA complexes to inverted repeats in their promoter regions (Wilbur *et al.* 2005; Robson *et al.* 2009). The inverted repeat sequence bound by VapBC in the *M. smegmatis vapBC* promoter (IR-1), overlaps the -35 and -10 sequences. Binding of the VapBC complex here interferes with RNA polymerase binding, as seen by a threefold increase in the transcriptional activity of the *vapBC* promoter of *M. smegmatis ΔvapBC* compared to that of wildtype *M. smegmatis* mc^2155 (Robson *et al.* 2009). The FitAB complex (VapBC from *N. gonorrhoeae*) binds an inverted repeat in its promoter with high affinity via a ribbon-helix-helix (RHH) DNA binding motif in FitA (Mattison *et al.* 2006). RHH domains are similar to transcription factors and function to aid sequence-specific DNA binding. Electrophoretic mobility shift assays (EMSAs) with purified *M. tuberculosis* VapBC TA complex, Rv0617/0616 has shown that Rv0617/0616 also binds the inverted repeat in its promoter

region (Abby Sharrock, unpublished data), corroborating the proposed autoregulation function of TA complexes. The first prokaryotic VapBC structure was published in 2009 by Miallau *et al*. involving VapC-5 (Rv0627) from *M. tuberculosis* in complex with a fragment of VapB-5 (Rv0626) (the missing residues were part of the N-terminal region predicted to bind DNA). Recently, another *M. tuberculosis* VapBC structure was solved, Rv0301/0300 (Min *et al.* 2012), supporting the proposed mechanisms for VapCs RNase activity, its inhibition by a cognate antitoxin, and its ability to bind promoter DNA (in this case, via RHH motifs like for the FitAB system). Both the Rv0301/0300 and VapBC-5 TA complexes from *M. tuberculosis* show similarity to the FitAB system from *N. gonorrhoeae* (Min *et al.* 2012). The structural homologue for VapC-5 is FitB, however their partners (VapB-5 and FitA, respectively) do not show the same degree of similarity, and structural comparisons suggest the VapBC interaction to be tighter than that for FitAB (Miallau *et al.* 2009). This strong binding of VapB and VapC reaffirms the need for exacting control of toxic VapC proteins in *M. tuberculosis*.

1.4.3 Targets of TA Toxin Proteins

The diverse range of targets for *E.coli* TA toxins include mRNA, tRNA, ribosomes, DNA replication machinery and cytoskeletal proteins (Yamaguchi & Inouye 2011), and many more are likely to emerge in the future from as yet uncharacterised TA systems. The majority of type II TA toxin proteins (including RelE, HigB, MazF, HicA, MqsR, Doc, HipA and VapC) inhibit translation by targeting the translation machinery. The target with the greatest representation is mRNA, and those proteins which inhibit translation via mRNA cleavage either in a ribosome-dependent or ribosome-independent manner, are called mRNA interferases (Yamaguchi & Inouye 2009). Ribosome-independent mRNA interferases can cleave mRNA directly in the cytoplasm as the RNA does not have to be undergoing translation for cleavage to occur. Ribosome-independent mRNA interferases include the MazF, ChpBK, MqsR, VapC, YhaV and HicA toxins. Ribosome-dependent mRNA Interferases cleave mRNAs at the ribosomal A site and include RelE, YoeB, YafO, YafQ and HigB.

The cellular targets and causes of induction of toxin proteins are highly diverse, suggesting these proteins play diverse roles in bacterial organisms. While a lot is known about the targets

in *E. coli*, there is still much debate about the function and targets of toxins in *M. tuberculosis*. As *M. tuberculosis* has so many TA systems and the number of systems possibly relates to pathogenicity (Pandey & Gerdes 2005), it is proposed TA systems may play a role in the ability of *M. tuberculosis* to persist, effectively dormant, within macrophages for long periods, and have been implicated in the metabolic switch from the replicative state to dormancy under nutrient stress (Albrethsen *et al.* 2013). Identifying the targets of *M. tuberculosis* toxins will provide a better understanding of their function in bacterial physiology and pathogenicity, and therefore the ability to develop novel antimicrobial therapies to target these persistent bacteria.

1.4.4 Physiological Function and Roles of Toxin–Antitoxin Systems

1.4.4.1 The Role of Toxin-Antitoxin Systems in the Stress Response

MazF expression is induced by a number of stresses (nutrient starvation, antibiotics, DNA damage and heat) once the MazE antitoxin has been degraded by either of the ATP-dependent serine proteases, ClpA–ClpP or Lon (Aizenman *et al.* 1996; Christensen *et al.* 2003). This leads to cell growth arrest or "quasi-dormancy", and programmed cell death in *E. coli* (Aizenman *et al.* 1996; Hazan *et al.* 2004; Suzuki *et al.* 2005; Kolodkin-Gal *et al.* 2007). MazF can induce cells into a quasi-dormant state where they are metabolically active, still producing RNA and protein (so long as the encoding mRNA is devoid of the MazF target sequence ACA) but are resistant to antibiotics (Suzuki *et al.* 2005; Yamaguchi & Inouye 2011). Cell growth can be recovered by the production of MazE antitoxin but only up to a point, at which stage *mazEF*-mediated cell death may occur and cannot be rescued (Amitai *et al.* 2004) although this was found to normally occur only during the exponential phase of growth (Kolodkin-Gal & Engelberg-Kulka 2009). The involvement of MazF in programmed cell death is suggested to occur via a quorum sensing mechanism whereby the altruistic suicide of some cells furnishes the remaining population with nutrients during times of stress (Engelberg-Kulka *et al.* 2006). Other groups have not observed this same *mazEF*-mediated cell death phenomenon (Gerdes *et al.* 2005; Van Melderen 2010), instead postulating that MazF induces a reversible bacteriostatic state (Pedersen *et al.* 2002). These differing theories as to the exact

role of MazEF highlight the extent to which our knowledge of these TA systems is limited, especially given that the MazEF system is one of the best characterised TA systems.

The other well characterised TA system is RelBE, which is involved in the stringent response induced under conditions of amino acid starvation. RelB (as with MazF) is also degraded by Lon protease, releasing RelE which inhibits translation, in turn activating RelA to synthesise guanosine tetraphosphate (ppGpp) synthase and therefore, stringent control of RNA synthesis (Figure 1.4) (Gerdes *et al.* 2005). The *mazEF* operon lies downstream of *relA*, so is also considered to be related to the stringent response (Yamaguchi & Inouye 2011). RelE not only causes inhibition of protein synthesis, but also a reduction in colony forming units which can then be revived by expression of RelB (Christensen *et al.* 2001). The ability of RelB to resuscitate RelE-inhibited cells suggests these cells are not dead, but in a state of stasis, corroborating Gerdes' hypothesis that these TA systems are involved in the stress response where they inhibit growth, directing cells into reversible bacteriostasis (Pedersen *et al.* 2002; Gerdes *et al.* 2005).

The two *higBA* loci from *Vibrio cholera* are hypothesised to function as stress response elements during the change in growth rate from fast (in the human intestine) to slow or dormant (in its aquatic environment) (Christensen-Dalsgaard & Gerdes 2006). Both HigB toxins are activated by amino acid starvation resulting in inhibition of protein synthesis and cell growth in *E. coli,* which can be reversed by expression of HigA (Christensen-Dalsgaard & Gerdes 2006; Gupta 2009).

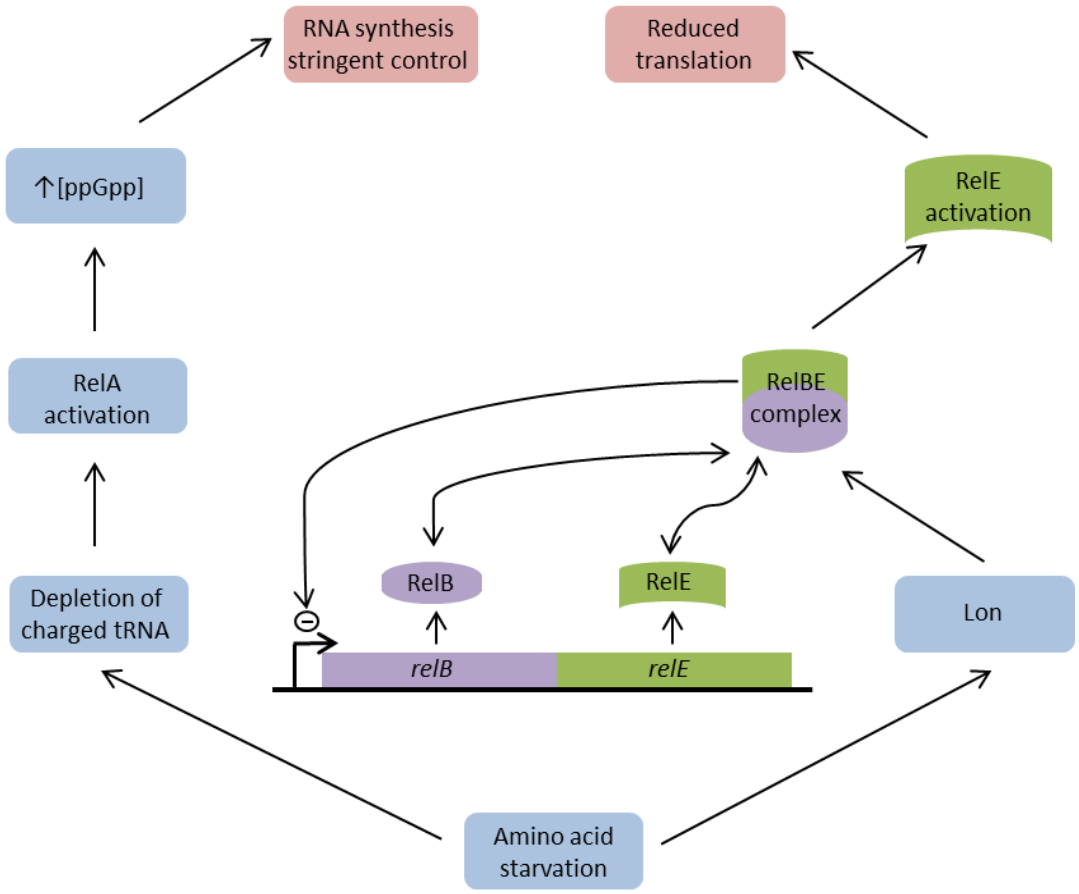


Figure 1.4: Model for *relBE* regulation in the context of the stringent response. RelE activation reduces translation (but does not completely inhibit it) which reduces the level of charged tRNAs, activating RelA to synthesise ppGpp and therefore stringent control of gene expression. Figure adapted from Gerdes *et al*. (2005).

In a recent study, multiple TA systems were also implicated with the switch into dormancy of *M. tuberculosis*, by the detection of TA system members in larger quantities in nutrient starved cultures, strengthening the role for these systems in the metabolic switch from the replicative to dormant state (Albrethsen *et al.* 2013). In total, 15 TA members were shown to be increased (VapC4, VapC27, VapC5, VapB32, VapC13, MazF6, VapC37, ParE2, VapC38, VapC39, VapC19, VapC41, VapC22, RelE2 and VapC44), of which two are antitoxins and 13 toxins, and the majority (11) belong to the VapBC family of TA systems. This was the first demonstration of *M. tuberculosis* TA proteins being induced during the same conditions of stress which may be experienced during latent infection.

1.4.4.2 The Role of Toxin-Antitoxin Systems in Bacterial Persistence

Bacterial persister cells are a small subpopulation of dormant variants that become highly tolerant to antibiotics without undergoing genetic change, and arise primarily in biofilms and in stationary-phase cultures (Lewis 2008). Most of the antimicrobials available for use today require bacteria to be actively growing to cause cell death, so the dormant or non-replicating persistent bacteria are likely to be tolerant to treatment (Gomez & McKinney 2004). This antibiotic-tolerant state has been shown to be induced by TA system activation in *E.coli* (Keren *et al.* 2004), strengthening the view that the TA systems in *M. tuberculosis* may also have this effect, contributing to the extended period which is currently required for effective antibiotic treatment of this disease. HipA of the *E.coli* HipBA TA locus induces growth arrest leading to a persistent multidrug-resistant state once cellular HipA toxin levels reach a certain threshold (Rotem *et al.*). This threshold allows part of the population to enter a dormant state without affecting the growth of the remaining cells. Phosphorylation of the essential translation factor EF-Tu is the proposed mechanism for HipA induced growth arrest (Schumacher *et al.* 2009).

The first direct evidence of a TA operon affecting cell persistence upon deletion was found with *mqsRA,* a member of the RelBE TA family (Brown *et al.* 2009). Deletion of *mqsRA* or *mqsR* alone decreases persister formation, and production of MqsRA increases persistence (Kim & Wood 2010); *mqsR* is also one of the most highly induced genes in *E. coli* persister cells (Shah *et al.* 2006). Transcription of *mqsRA* is activated by amino acid and glucose starvation, and after MsqA degradation, MsqR is proposed to interact with Hha and CspD (Kim & Wood 2010) as well as cleave RNA in a ribosome independent fashion (Christensen-Dalsgaard *et al.* 2010). Induction of *relE*, *higB*, *mazF*, *yafQ*, and *yoeB* have also been implicated in persister cells, however the MqsRA TA system remains the only TA system affecting persistence upon deletion where the antitoxin is a protein (Shah *et al.* 2006; Wang & Wood 2011).

1.4.4.3 The Role of Toxin-Antitoxin Systems in Biofilm Formation

Biofilms are dense, multicellular communities of bacteria formed in liquid environments by attaching to the air-liquid interface, submerged surfaces or to each other (Kolter & Losick

1998). Initially, it was thought that TA systems were not involved in biofilm formation as *mazF* and *relE* toxin knockouts had shown no difference in biofilm formation compared to wildtype strains (Lemos *et al.* 2005), but since then a number of studies have suggested a role for TA systems in biofilm formation. MqsRA was the first TA system to be linked to biofilm formation and the autoinducer-2 quorum sensing system, where MqsR in *E.coli* was induced in biofilm cells and enhanced cell motility, and the deletion of *mqsR* inhibited biofilm formation (González Barrios *et al.* 2006). MqsR has also been linked to biofilm formation via its association with QseB, an important protein in biofilm formation, as QseB levels were reduced in a *mqsR* deletion mutant (González Barrios *et al.* 2006).

Additional evidence that TA pairs affect biofilm formation came from a mutant strain lacking the five TA systems MazF/MazE, RelE/RelB, YoeB/YefM, and YafQ/DinJ and ChpB (of which the toxins all inhibit translation by cleaving RNAs). This mutation had no impact on the stress response of cells (Tsilibaris *et al.* 2007), but biofilm formation was reduced during the first 8 hr before recovering after 24 hr (Kim *et al.* 2009). Transcriptome profiling of these biofilm cultures to determine the mechanism by which the TA systems were affecting biofilm formation found differential expression of a single gene, *yjgK*. Induction of YjgK decreased biofilm formation at 8 hr (by repressing fimbria genes as bacteria use fimbriae for adherence) and increased it at 24 hr, while deleting *yjgK* had the reverse effect corroborating the complex phenotype observed in the five deletion TA systems mutant (Kim *et al.* 2009). Another group confirmed this phenomenon and found the initial biofilm formation defect was mainly due to deleting *mazF* and *yafQ* which decreased cell lysis (Kolodkin-Gal *et al.* 2009). Deletion of *yafQ* showed no effect on biofilm formation but elevated levels of cell death under exposure to some bactericidal antibiotics during biofilm formation, while overproduction of YafQ increased biofilm formation substantially (Harrison *et al.* 2009). These results suggest cell growth inhibition or lysis caused by the TA system toxins affects biofilm formation.

1.4.4.4 Other Roles of Toxin-Antitoxin Systems

TA systems have also been postulated to function in other diverse roles such as DNA damage-independent replication arrest (MazF and RelE), the SOS response (YafON and YafQ–DinJ)

(Yamaguchi & Inouye 2011), and stabilising integrons (HigBA) (Christensen-Dalsgaard & Gerdes 2006), as well as interacting within a network of TA systems. Interaction between TA systems in a network is supported by examples of noncognate toxin-antitoxin associations, even among different families both *in vivo* and *in vitro* (Figure 1.5) (Zhu *et al.* 2010). Zhu *et al.* (2010) showed that expression of *M. tuberculosis* VapB-25 can neutralize the toxicity of *M. tuberculosis* MazF-3 in *E. coli,* but that interactions detected *in vivo* may not be as stable under nonphysiological conditions. The suggestion of cross talk between TA systems is controversial, as two other studies have found no such phenomenon in the *M. tuberculosis* TA systems which they investigated (Ramage *et al.* 2009; Ahidjo *et al.* 2011). These results using *M. tuberculosis* TAs challenge the general consensus of 'one toxin for one antitoxin' rule established from studies in other bacteria (usually *E.coli*), and suggest the possibility of the combined action of multi-TA system toxins functioning so as *M. tuberculosis* can adapt and survive in the diverse environmental conditions encountered.

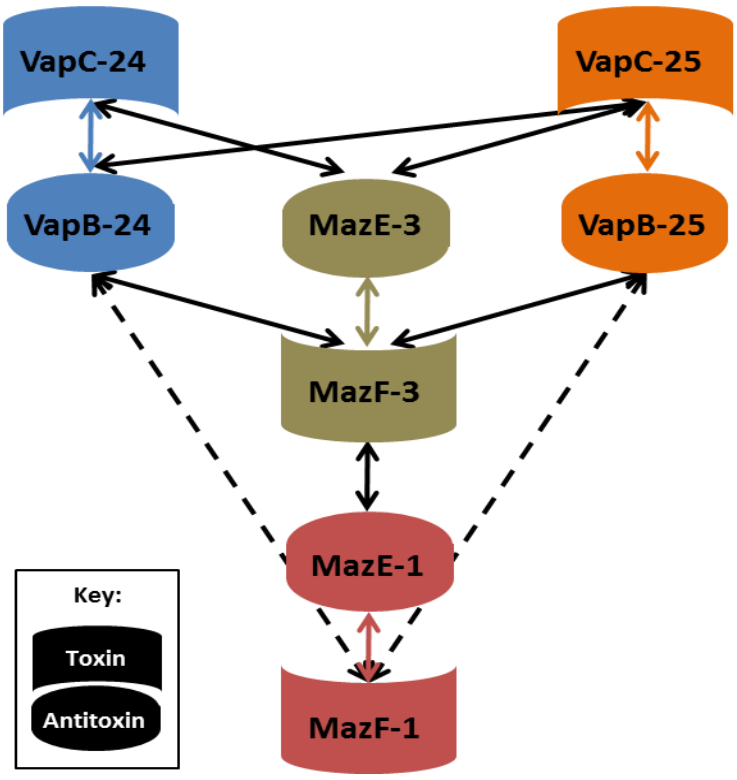


Figure 1.5: Summary of possible TA system networks in *M. tuberculosis*. Coloured lines represent interactions between cognate TA pairs, black lines between non-cognate pairs. Dashed lines indicate weaker interactions based on partial rescue of toxicity. Figure adapted from Zhu *et al.* (2010).

The role TA systems play in response to antibiotic, nutrient, temperature, oxidative or other stressors caused by the extreme environments which they encounter is still under intense investigation, and while there is a wealth of information regarding the function of TA systems in *E. coli*, their role in other microbes, especially *M. tuberculosis*, remains largely unexplored.

1.4.5 VapBC Toxin-Antitoxin Systems

VapBC TAs are virulence associated proteins. The first *vapBC* locus was identified on a *Salmonella Dublin* virulence plasmid (Pullinger & Lax 1992), where inactivation of *vapB* caused the loss of plasmid-associated virulence as well as growth inhibition on selected media. Many examples have since followed from both virulence plasmids, where the locus functions by killing plasmid-free cells and therefore maintaining plasmid survival (such as *mvpAT* in *Shigella fleneri* (Sayeed *et al*. 2000), and chromosomal *vapBC* loci which will be discussed in greater detail in the following sections.

1.4.5.1 PIN-Domain Proteins

VapC toxins belong to the PIN-domain family of proteins. PIN-domain proteins were originally named based on their sequence similarity to the N-terminal domain of a PilT (PilT N-terminus) protein, and are small ribonuclease proteins (approximately 130 amino acids) containing four conserved acidic residues within an RNase-H-like fold (Arcus *et al.* 2004; Mattison *et al.* 2006; Miallau *et al.* 2009; Arcus *et al.* 2011). These conserved amino acids form a negatively charged pocket or active site which co-ordinates the binding of a Mg^{2+} or Mn^{2+} ion and facilitates metal dependent nuclease activity, as shown by structural characterisation of FitB from *Neisseria gonorrhoeae* (Mattison *et al.* 2006), VapC from *Pyrobaculum aerophilum* (Bunker *et al.* 2008) and VapC5 from *M. tuberculosis* (Miallau *et al.* 2009). Figure 1.6 shows the α/β/α stack with a central twisted parallel β sheet of five short strands, typical of PIN-domain protein structures (Arcus *et al.* 2011), and the four conserved acidic residues within.

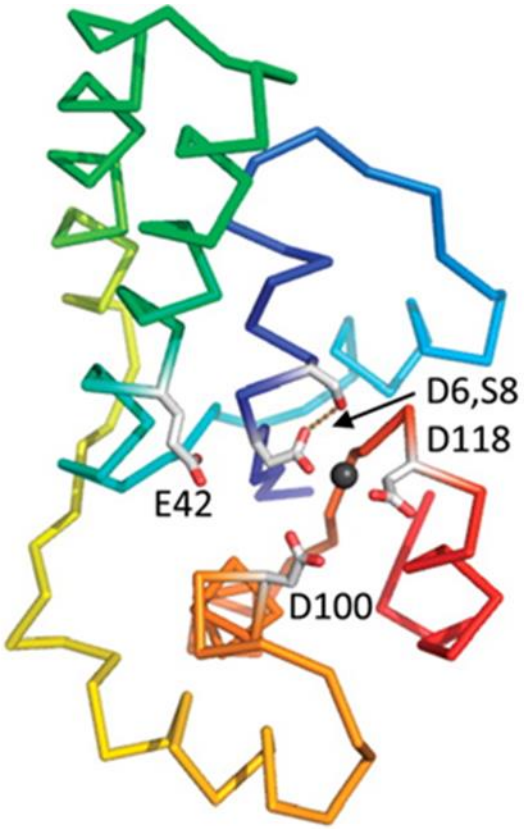


Figure 1.6: The structure of the PIN-domains. A cartoon representation of the α/β/α stack containing a central 5-stranded parallel β sheet typical of PIN-domain protein structures, and the four conserved acidic residues within. The N-terminus to the C-terminus is coloured blue to red and the active site metal 2+ ion is shown in black. Figure from Arcus *et al*. (2011).

1.4.6 Toxin-Antitoxin Systems in *M. tuberculosis*

M. tuberculosis belongs to the *Mycobacterium tuberculosis* complex (MTBC) which includes all *Mycobacterium* species that can cause tuberculosis in humans or other organisms. The number of TA systems in *M. tuberculosis* is greatly expanded with 88 putative candidates, the majority of which are conserved only in the MTBC and absent from other, even closely-related mycobacterial species (Ramage *et al.* 2009). This expansion of TA systems in the MTBC (of which 37 % are located in regions identified as regions of horizontal gene transfer (HGT)) is proposed to have occurred during (or after) the time of divergence from the last common ancestor (Stinear *et al.* 2008; Ramage *et al.* 2009). This sudden expansion of TA systems is consistent with the theories that TA systems in the MTBC were acquired by HGT, and DNA

acquired by HGT played a crucial role MTBC speciation (Becq *et al.* 2007), strengthening the importance of TA systems in the evolution of *M. tuberculosis*.

Ramage *et al*. (2009) tested expression of 78 (of the 88) putative *M. tuberculosis* TA systems in *M. smegmatis,* and found 30 of them to be functional (i.e., expression of the toxin resulted in toxicity to the cells and subsequent expression of the cognate antitoxin neutralised the toxicity). Of the 30 functional TA systems they identified, 20 were VapBC, four MazEF, two RelBE, one HigBA and three were novel TA family systems. These functionality results are inconsistent with those obtained previously for 38 of the same candidates in *E.coli,* where seven of the TA systems deemed functional in *M. smegmatis* were found to not be functional in *E.coli* (Zhu *et al.* 2006; Gupta 2009), highlighting the importance of testing the functionality of TA systems in a closely related host. While functional TA systems have been identified on *M. tuberculosis* chromosomes, not much is known about what their physiological role is in *M. tuberculosis* itself, as most of what we know about their functionality has been gathered from studies of the TA systems in other bacterial species.

1.4.6.1 The RelBE Toxin-Antitoxin System

To date, only three Rel TA modules have been identified in *M. tuberculosis* (*relBE, relFG* and *relJK*), all of which are functional in *M. smegmatis* and whose antitoxins (by themselves) can act on transcription: RelJ as a repressor, and RelB and RelF as activators (Korch *et al.* 2009). All three Rel TA complexes also act to repress expression when combined, consistent with other TA systems. Korch *et al.* (2009) also showed for the first time that *M. tuberculosis* expresses TA modules both in broth culture (all six *rel* genes) and during infection of human macrophages (*relE*, *relF*, and *relK*). This involvement suggests they may be required for survival within macrophages and function in bacteriostasis as opposed to cell death, and therefore are linked to *M. tuberculosis* persistence (Korch *et al.* 2009; Min *et al.* 2012). Initial studies in *E.coli* found RelE associated with the ribosomal A site to initiate endogenous ribonuclease activity of the ribosome which preferentially cleaves mRNA between the second and third bases of the stop codon (with a higher specificity for the stop codon UAG), leading to inhibition of translation elongation and protein synthesis (Christensen & Gerdes 2003;

Pedersen *et al.* 2003; Hurley & Woychik 2009). However a recent contradictory study reported no codon-dependent cleavage by RelE *in vivo*, but preferential cleavage at the beginning of the 5' end of the coding region (Hurley *et al.* 2011). The activity of other members of the RelBE TA family (such as HigB, YoeB, YafQ and YafO) is dependent on their association with the 50S ribosomal subunit, and all have differing cleavage specificities (Hurley & Woychik 2009).

1.4.6.2 The MazEF Toxin-Antitoxin System

Nine MazEF TA systems have been identified in *M. tuberculosis*, four of which have been confirmed as functional in *M. smegmatis* (Pandey & Gerdes 2005; Ramage *et al.* 2009). MazF cleaves the phosphodiester bond of mRNA at the 5'-end, yielding a free 5'-OH group on the 3' cleavage product and a 2'3'-cyclic phosphate on the 5' product (Zhang *et al.* 2005). MazF does not cleave RNA/DNA or RNA/RNA duplex sequences, nor ssDNA (Zhu *et al.* 2009). Activity of MazF does not require Mg^{2+} (like RNase A), which is unlike the other TA endoribonucleases (Zhang *et al.* 2005). MazF toxins recognise different sequences in different organisms ranging from three bases in length for *E. coli* (*ACA) (* indicates cleavage site) (Zhang *et al.* 2005), to a pentad sequence for *Staphylococcus aureus* and *Bacillus subtilis* (U*ACAU) (Zhu *et al.* 2009; Park *et al.* 2011) and *Myxococcus xanthus* (GU*UGC (although the first G can be replaced with an A)) (Nariya & Inouye 2008). The *M. tuberculosis* MazF homologues each have different cleavage site specificities, for example MazF-mt1 = CU*ACC (and to a lesser extent UU*ACA sequences), MazF-mt3 = UU*CCU or CU*CCU, MazF-mt7 = U*CGCU and the consensus cleavage sites for MazF-mt6 are (U/C)U*(A/U)C(U/C) (Zhu *et al.* 2006; Zhu *et al.* 2008). Different cleavage site specificities for each toxin within the same species may allow differential mRNA degradation as a way to adapt to the various environmental conditions *M. tuberculosis* encounters.

1.4.6.3 The VapBC Toxin-Antitoxin System

VapBC TA systems are over-represented in the *M. tuberculosis* genome, accounting for 47 of the 88 systems present (Ramage *et al.* 2009; Ahidjo *et al.* 2011) highlighting the need to determine their biological role. No cross talk between non-cognate *M. tuberculosis* VapB

antitoxins and VapC toxins has been observed for the Rv0300/0301, Rv1561/1560, Rv2829c/2830c and Rv3408/3407 VapBC TA systems (Ramage *et al.* 2009). Toxins from these four VapBC systems inhibit translation (inhibiting growth) and Rv0301 and Rv1561 degrade MS2 RNA (though Rv0301 to a lesser extent). These effects were all reversed by co-expression with a cognate antitoxin supporting the theory that VapC toxins affect translation directly via their ribonuclease RNase activity (Ramage *et al.* 2009). Mg^{2+} dependent ribonuclease activity of another VapC, VapC-5, has also been demonstrated, although a portion of the protein was in complex with its antitoxin so the observed activity was weak compared to other VapC toxins (Miallau *et al.* 2009).

A survey of the toxicity of 10 *M. tuberculosis* VapC proteins showed only Rv0549c, Rv0595c, Rv2549c and Rv2829c were toxic when expressed from a tetracycline-regulated promoter in *M. smegmatis* as well as when conditionally expressed in *M. tuberculosis*, while Rv0627, Rv1953, Rv2010, Rv2546, Rv2548 and Rv3320c had no effect on the growth of *M. tuberculosis* in liquid culture (Ahidjo *et al.* 2011). Toxicity of Rv3320c was observed but only when it was constitutively expressed, while toxicity of Rv2549c only occurred once the level of protein exceeded a certain threshold. The observations that toxicity of Rv2549c correlated with the level of protein expressed, while Rv2456 showed no toxicity and was induced at markedly lower levels than Rv2549c, suggested VapC proteins termed 'non-toxic' may just have been expressed in levels insufficient to exert an effect. This is supported by discrepancies from other studies on the effects of VapC toxin expression on the growth of *E. coli*, *M. smegmatis* and/or *M. tuberculosis* hosts (Gupta 2009; Ramage *et al.* 2009). For example Ahido *et al*. (2011) found Rv0627 was non-toxic in *M. smegmatis* and *M. tuberculosis* while Miallau *et al*. (2009) determined it to be toxic in both organisms. These discrepancies are possibly due to differences in the expression vectors, growth conditions or strains used, or translation initiation signals causing insufficient levels of VapC expression (Gupta 2009; Ramage *et al.* 2009; Ahidjo *et al.* 2011). Therefore, it is likely that the current number of VapC proteins described as 'toxic' is under-represented.

Expression of all 30 of the functional *M. tuberculosis* TA systems has been monitored via qPCR under hypoxic conditions, both in culture and during infection of bone marrow-derived

macrophages, to test if subsets of TA systems respond to different cellular stressors. *Rv2009-Rv2010* and *Rv1955-Rv1956* (a HigBA homologue) were induced during hypoxia while two different VapBC systems, *Rv0549-Rv0550* and *Rv1560-1561,* were induced during macrophage infection suggesting TA systems may indeed be regulated independently, and in response to different environmental stressors (Ramage *et al.* 2009).

Much of the pioneering investigation into mycobacterial VapBC systems was performed in *M. smegmatis* as it has just three TA systems, only one of which is a VapBC (MSMEG_1283/1284). The *M. smegmatis* VapC toxin, $VapC_{MS1284,}$ is the homologue of Rv0624 in *M. tuberculosis* and its function has been investigated by Jo McKenzie in our lab and our collaborators at the University of Otago during the course of this PhD. Robson *et al*. determined that the *M. smegmatis* VapBC complex binds to an inverted repeat in the promoter DNA to regulate its transcription, and *vapB* and *VapC* are transcribed as a single leaderless mRNA (Robson *et al.* 2009). While expression of the VapBC complex had no effect on growth, expression of VapC inhibited growth in the *ΔvapBC* background, and to a lesser extent in the wild type strain where endogenous VapB was present to counteract some of its activity. Growth was restored when conditional expression of VapC ceased, suggesting a bacteriostatic, not bacteriocidal effect of VapC (Robson *et al.* 2009). $VapC_{MS1284}$ was confirmed as an endoribonuclease, specifically cleaving and binding AUA(U/A) ssRNA sequences to inhibit translation *in vitro* and *in vivo* (McKenzie 2011; McKenzie *et al.* 2012b). These AUA(U/A) sequences were found to be abundant in transcripts (and their upstream regions) which had been down regulated in response to overexpression of $VapC_{MS1284}$ in microarray experiments. A large proportion of the down regulated genes had annotated roles in carbon transport and metabolism, suggesting a role for VapC as a regulator of carbon utilisation in response to the demands of the cell, in order to balance the rate of anabolism and catabolism post-transcriptionally (McKenzie *et al.* 2012b).

The biological role of *M. tuberculosis* VapBC systems is largely unknown. The expanded number of VapBC systems encoded in the genome together with evidence for various TA systems involvement in pathogenesis, persistence, cell death and response to stress suggested *M. tuberculosis* VapBC systems may play a role in slowing or regulating metabolism

and cell growth for survival and in tuberculosis persistence. As such, determining the role of VapBC systems in *M. tuberculosis* is crucial to improving our understanding of these systems and how this pathogenic organism can adapt to the diverse array of environmental conditions it encounters.

1.5 Objectives

The observations above defined the main objectives to my doctoral research, as outlined below:

Objective One - Identify conditions in which *M. smegmatis vapB* or *vapC* expression varied to help identify the mechanisms involved in the regulation of VapC, and develop an optimised RNA isolation and RT-qPCR method suitable for the quantification of these target genes.

Objective Two - Clone the 47 VapBC proteins from *M. tuberculosis* into *M. smegmatis* and assess their soluble expression and purification.

Objective Three - Determine the specificity and activity of a number of purified *M. tuberculosis* VapC proteins.

Objective Four - Investigate the structure of a *M. tuberculosis* VapBC complex using X-ray crystallography.

Objective Five - Classify the endemic Rangipo strain of *M. tuberculosis*, and develop a fast and efficient Rangipo specific diagnostic assay.

Chapter three details the development of the Rangipo strain-specific diagnostic assay, while chapters four and five investigate the regulation and function of mycobacterial VapC.

Chapter Two:

Methods

2.1 Methods relating to Chapter Three: Rangipo Diagnostic

2.1.1 Clinical Isolates and *M. tuberculosis* genomic DNA (gDNA)

Crude *M. tuberculosis* gDNA used for typing PCRs (Section 2.1.3.1) and lineage classification (Sections 2.1.3.2 - 2.1.3.5) was extracted in-house from sonicated and heat-killed clinical specimens obtained from the Waikato Hospital Laboratory as for Section 2.1.1.1.
M. tuberculosis gDNA used for developing the Rangipo diagnostic (Section 2.1.3.6) was obtained either from clinical isolates from Waikato Hospital Laboratory (Hamilton, NZ) extracted by Roberto Colangeli (University of Medicine and Dentistry of New Jersey, USA), or crude clinical isolate extracts from labPLUS (Auckland, NZ).

2.1.1.1 *M. tuberculosis* Genomic DNA Extraction for Typing PCRs

Isolates 'A-F' and 'J-S' were extracted using a hexadecyltrimethylammonium bromide (CTAB) extraction protocol as described below. 0.5 ml sterilised cells were resuspended in 50 µl Tris-EDTA-SDS lysis solution and added to 50 µl 10 % SDS before incubating for 30 min at 70 °C, 700 rpm.100 µl 5 M NaCl and 80 µl CTAB/NaCl solution were added, samples mixed, then incubated for 2 min at 80 °C followed by 10 min at 65 °C, 700 rpm. 700 µl chloroform was added and samples were shaken vigorously for 1 min before rotating for 15 min at room temperature to deproteinise. The samples were spun (13,000 rpm, 10 min, room temperature), aqueous top layer transferred to a new tube and an equal amount of isopropanol added. Samples were incubated for 50 min at -20 °C then spun (13,000 rpm, 15 min, room

temperature) before the supernatant was removed. Samples were washed with 1 ml 70 % ethanol, spun (13,000 rpm, 5 min, room temperature), supernatant removed then the pellet air dried. Pellets were resuspended in 500 µl 5 M GITC, pH 7 by rotating for 15 min at room temperature. 250 µl of 7.5 M ammonium acetate and an equal volume of chloroform (750 µl) were added before the samples were shaken well and rotated for 15 min at room temperature. Samples were spun (13,000 rpm, 10 min, room temperature), the top layer transferred to a new tube, and an equal amount of isopropanol added before precipitating for 50 min at -20 °C. Samples were spun (13,000 rpm, 15 min, room temperature), supernatant removed and the pellets air dried before resuspending in 25 µl TE.

Isolate 'T' was extracted using a GITC based extracted protocol as follows. 0.5ml heat killed and sonicated cells were resuspended in 0.5 ml 5 M GITC, pH7 and transferred to a sterile tube containing 0.3 g of 0.1 mm and 2.5 mm zirconia beads. The homogenate was beaten four times for 20, 25, 30 and 35 s respectively in a FastPrep cell disrupter (FP120) (Thermo Savant, USA) at setting 6 with 1 min rests in between. 250 µl of 7.5 M ammonium acetate and an equal volume of phenol/chloroform, pH 8 (750 µl) were added before the samples were shaken and rotated for 15 min at room temperature to deproteinise. The samples were spun (13,000 rpm, 15 min, room temperature), top layer transferred to a new tube and an equal amount of isopropanol added. Samples were precipitated overnight at -20 °C then spun (13,000 rpm, 15 min, room temperature) before the supernatant was removed. Samples were washed with 1 ml 70 % ethanol, spun (13,000 rpm, 5 min, room temperature) and the supernatant removed. Samples were washed again with 1 ml 100 % ethanol, spun (13,000 rpm, 5 min, room temperature), supernatant removed and the pellets air dried before resuspending in 20 µl TE.

2.1.2 Whole Genome Sequencing (WGS)

WGS was performed by Roberto Colangeli and David Alland at the Department of Medicine, University of Medicine and Dentistry of New Jersey (USA) (Colangeli *et al.* 2014). 10 Rangipo clinical isolates from a time range spanning 19 years (A, C, E, F, N, O, R, S, T and U) were sent from New Zealand, and DNA was isolated using a CTAB extraction protocol as described

in (van Embden *et al.* 1993). Genome sequencing, processing and SNP detection was performed as described in Colangeli *et al*. (2014). Briefly, gDNA was sheared and barcoded libraries prepared then processed on the 5500xl SOLiD™ System (Applied Biosystems, USA) as per manufacturer's recommendation. The overall fold coverage for each strain ranged from 392 to 40. Sequence reads were trimmed to remove short and low quality reads and mapped to the *M. tuberculosis* H37Rv genome (NCBI Reference Sequence: NC_000962.3). Analysis was performed using CLC Genomics Workbench (CLCbio, Denmark). Single nucleotide polymorphisms (SNPs) in the subject sequences were identified by comparing to the reference strain genome, H37Rv, and those SNPs with <80 of the reads showing the polymorphism were discarded. Mutations found to be common to all 10 of the study isolates (compared to the reference strain) were deemed as probable Rangipo specific SNPs.

2.1.3 Polymerase Chain Reaction

Primers for all PCRs were obtained from IDT (USA).

2.1.3.1 PCR-Based Genotyping of *M. tuberculosis* targeting GC-rich repeats and IS*6110* Inverted Repeats

Two different approaches based on Kotlowski *et al.* (2004) and Otal *et al.* (1997) were used for preliminary typing of the 15 Rangipo clinical isolates obtained from Waikato Hospital Laboratory to see if there were any differences between them.

The first approach used primers which target frequently repeated 16-bp GC rich sequences (Mtb1 or Mtb2) in combination with specific primers sited within the inverted repeats flanking IS*6110* (IS1 and IS2) (Kotłowski *et al.* 2004). PCR reactions were carried out with Green Taq DNA Polymerase (GenScript, USA) in 25 µl volumes using the following concentrations: 1 x PCR buffer, 0.25 mM deoxynucleotide mix (dATP, dCTP, dTTP & dGTP), 2.5 mM $MgCl_2$, 0.625 U Green Taq DNA Polymerase, 0.2 µM each primer (IS1, IS2 and Mtb1 or IS1, IS2 and Mtb2) and 200 ng crude template DNA. The following cycling conditions were used for amplification:

95 °C	3:00		(min:sec)
95 °C	0:20	┐	
62 °C	0:20	│ x 39	
72 °C	2:00	┘	
72 °C	5:00		

PCR reactions were visualised as in Section 2.2.7.5, except on 1.5 % agarose gels.

The second approach used the same primer above targeting the frequently repeated 16-bp GC rich sequence (Mtb1) in combination with a primer containing three degenerate inosine bases sited within the IS*6110* repeat (CAR2) (Otal *et al.* 1997). PCR reactions and conditions were set up as above.

2.1.3.2 Lineage-specific Large Sequence Polymorphism (LSP) based PCR for *M. tuberculosis* Lineage Classification

PCRs based on LSPs identified by Gagneux *et al.* (2006) were conducted to classify which of the six main *M. tuberculosis* lineages the Rangipo strain belonged to. Primers used were the same as those used by Reed *et al.* (2009) based on regions of difference (RD) with respect to H37Rv. PCR reactions were carried out with *Taq* DNA polymerase (Invitrogen, USA) in 25 µl volumes using the following concentrations: 1 x PCR buffer, 0.2 mM deoxynucleotide mix (dATP, dCTP, dTTP & dGTP), 1.5 mM $MgCl_2$, 1 U *Taq* DNA Polymerase, 0.3 µM each primer and 7 - 25 ng template DNA. The following cycling conditions were used for amplification:

95 °C	3:00		(min:sec)
95 °C	0:45	┐	
58 °C	0:30	│ x 34	
72 °C	0:40	┘	
72 °C	10:00		

PCR reactions for RD9, RD105, RD239 and RD750 were visualised as in Section 2.2.7.5. Those for RD1-15 were visualised as for Section 2.3.1.4 with SYBR Safe™ DNA gel stain (Invitrogen, USA) for recovery, then the DNA was extracted from the gel as in Section 2.3.1.5 prior to sequencing.

2.1.3.3 Euro-American Lineage Classification Based on the *katG*^{463} SNP Using the 'on/off Switch Assay'

MTBC strains with the katG 463 codon 'ctg' rather than 'cgg' are considered as belonging to the Euro-American lineage (Sreevatsan *et al.* 1997; Nahid *et al.* 2010). This SNP was used in an on/off switch assay, in an attempt to speed up and reduce the costs associated with the Euro-American strain classification process used above (Section 2.1.3.2) requiring a RD1-15 PCR, purification and sequencing step. The on/off switch assay (Zhang & Li 2003) uses phosphorothioate-modified primers and *Pfu* polymerase (a high-fidelity DNA polymerase with 3' exonuclease activity) to distinguish between matched and mismatched primers. The forward primer katG-F is common to both Euro-American and non Euro-American genotypes. The reverse primers harbour a 3' end phosphorothioate-modified nucleotide and are specific to Euro-American (katG-R1) or non-Euro-American (katG-R2/3) strains with the katG 463 codon 'ctg' or 'cgg' respectively. Primers and reactions were as described in (Alix *et al.* 2006). Primers were obtained from IDT (USA) both with and without HPLC purification. PCR reactions were visualised as in Section 2.2.7.5.

2.1.3.4 Euro-American Sub-lineage Classification

PCRs were carried out for each of the 10 Euro-American sub-lineages as defined by Gagneux *et al.* (2006) and covered regions of difference (RD) 115, 122, 174, 182, 183, 193, 219, 724, 726 and 761. Here, deletions in each region of difference are specific for a certain Euro-American sub-lineage, so if the deletion is detected the sample is deemed to belong to that sub-lineage and if no deletion is detected it is not. Additional PCR reactions with internal flanking primers (de Jong *et al.* 2009) were required for successful amplification of RD 174, 182 and 219 due to the size of the original amplicon and fragmented nature of template DNA used.

PCR reactions with *Taq* DNA polymerase (Invitrogen, USA) were carried out in 25 µl volumes using the following concentrations: 1 x PCR buffer, 0.2 mM deoxynucleotide mix (dATP, dCTP, dTTP & dGTP), 1.5 mM $MgCl_2$, 2.5 U *Taq* DNA Polymerase, 0.3 µM each primer and 25 ng template DNA. PCR reactions with HOT FIREPol® DNA Polymerase (Solis BioDyne, Estonia) were carried out as for those with *Taq*, but only 1.25 U HFP polymerase was used. The following cycling conditions were used for amplification:

Reactions with *Taq*			**Reactions with HFP**		
95 °C	3:00	(min:sec)	95 °C	15:00	
95 °C	0:30	┐	95 °C	0:20	┐
varied	0:30	x 39	varied	0:30	x 35
72 °C	varied	┘	72 °C	varied	┘
72 °C	7:00		72 °C	7:00	

Table 2.1: PCR set up for region of difference amplifications

PCR	Polymerase	Extension Time (min:sec)	Amplicon size (bp)	
			H37Rv	Deleted
RD115*	HFP	4:00	3,442	835
RD122*	*Taq*	1:30	1,456	285
RD174*	not successful		4,524	874
RD182*	not successful		7,292	812
RD183*	HFP	4:30	3,965	1,048
RD193*	*Taq*	1:30	1,308	807
RD219*	not successful		3,609	717
RD724^	*Taq*	1:30	1,288	1,650
RD726#	HFP	2:05	2,062	313
RD761#	HFP	2:05	1,389	295

All reactions were carried out at 64 °C.

***** Primers from (Tsolaki *et al.* 2004)

^ Primers from (Mostowy *et al.* 2004)

Primers from Gagneux *et al.* (2006)

Table 2.2: PCR set up for internal flanking region of difference amplifications

PCR	Polymerase	Extension Time (min:sec)	Amplicon size (bp)	
			H37Rv	Deleted
RD174-Int*	*Taq*	0:40	533	0
RD182-Int*	*Taq*	0:40	593	0
RD219-Int*	*Taq*	0:40	395	0

All reactions were carried out at 60 °C.

***** Primers used as described in de Jong *et al.* (2009).

PCR reactions were visualised as in Section 2.2.7.5.

2.1.3.5 Latin American Mediterranean (LAM) Strain Classification

The Latin American Mediterranean (LAM) strains belong to a sub-lineage of the Euro-American superlineage. Members of the LAM family are identified via screening for the LAM specific SNP at codon 103 of the gene encoding Ag85C (Rv0129c) by restriction fragment length polymorphism (RFLP) analysis or by LAM-specific IS*6110* PCR (Marais *et al.* 2006; Gibson *et al.* 2008). PCR reactions were carried out with HOT FIREPol® DNA Polymerase (Solis BioDyne, Estonia) in 20 µl volumes using the following concentrations: 1 x PCR buffer, 0.2 mM deoxynucleotide mix (dATP, dCTP, dTTP & dGTP), 3.5 mM $MgCl_2$, 1.5 U HOT FIREPol® DNA Polymerase, 1.25 µM each primer (reactions for the LAM-restricted IS*6110* locus used LAMF, LAMR and XhoI, reactions for *Ag85C*[103] PCR-RFLP used Ag85C103F and Ag85C103R), 1 x Solution S and 10 ng template DNA. The following cycling conditions were used for amplification:

95 °C	15:00		(min:sec)
95 °C	0:30	┐	
62 °C	1:00	x 34	
68 °C	1:00	┘	
72 °C	7:00		

PCR reactions for the LAM-restricted IS*6110* locus were visualised as in Section 2.2.7.5. Those for *Ag85C*[103] PCR-RFLP had 8 µl of the reaction visualised as for Section 2.2.7.5 and the remaining 12 µl digested with 6 U Mn1I (New England Biolabs, USA) in 1 x NEB buffer 4 and 1 x BSA for 4 hr at 37 °C, 900 rpm, then visualised as for Section 2.2.7.5 but on a 4 % gel.

2.1.3.6 Rangipo Diagnostic PCRs

2.1.3.6.1 Amplification Across SNPs

Gene sequences spanning probable Rangipo-specific SNP sites were amplified and sequenced to confirm the presence of the SNP in Rangipo strains. PCR reactions were carried out with *Taq* DNA polymerase (Invitrogen, USA) in 20 µl volumes using the following concentrations: 1 x PCR buffer, 0.2 mM deoxynucleotide mix (dATP, dCTP, dTTP & dGTP), 1.5 mM $MgCl_2$, 2 U *Taq* DNA Polymerase, 0.3 µM each primer and 10 ng template DNA. The following cycling conditions were used for amplification:

95 °C	3:30		(min:sec)
95 °C	0:30	┐	
58 °C	0:30	x 34	
72 °C	0:35	┘	
72 °C	7:00		

10 µl of PCR reactions for Rv0405, Rv1192, Rv1821, Rv2546, Rv3119, Rv3769 and Rv3881c were run on a gel as in Section 2.2.7.5 to check there was just a single product, and then the remainder was ExoISAP treated in preparation for sequencing as in Section 2.1.4. Where multiple products were amplified (Rv0071, Rv2224c and Rv3616c) a two-step PCR was performed as below prior to confirmation of single products and ExoISAP treatment (as above) before sequencing.

Two-step PCR reactions were set up as above but cycled as follows:

95 °C	3:30		(min:sec)
95 °C	0:20	x 34	
65 °C	0:40		
68 °C	7:00		

2.1.3.6.2 Uniplex Rangipo-Specific SNP RFLP Assays

Rangipo-specific SNPs were PCR amplified then digested in a RFLP assay to determine if the RFLP assay could detect the SNP in Rangipo or non-Rangipo strains and therefore distinguish between the two strains. PCR reactions were set up as for Section 2.1.3.6.1 using the two-step cycling protocol with an extension time of 30 s. 10 µl of PCR reactions were run on a gel as in Section 2.2.7.5 to check products had amplified, then the remainder was digested overnight with 1 µl of the respective restriction enzyme and incubation temperature (see Table 2.3 below) prior to analysing on a 3 % agarose gel as for Section 2.2.7.5. Digests were initially performed in their respective buffers, then all those which produced satisfactory results were tested in just the PCR reaction mixture, without the addition of restriction enzyme buffers or additives.

Table 2.3: Uniplex Rangipo-specific SNP RFLP assays

SNP	Restriction Enzyme	NEB buffer	Temperature (°C)	Product sizes (bp)		
				Undigested	Digested	
					Rangipo	Non-Rangipo
Rv0071	ApaI[1]	-	25	221	221	26, 195
Rv0405	HinfI[2]	-	37	222	11, 93, 118	11, 22, 71, 118
Rv0719	RsaI[2]	-	37	119	119	40,79
Rv1821	Sau3AI[3]	4+BSA	37	493	493	149, 344
Rv1854c	BclI[1]	-	37	401	401	130, 271
Rv2224c	AvaI[4]	-	37	146	146	23, 123
Rv2504c	DraIII[5]	-	37	241	241	48, 193
Rv2546	EcoR1[5]	3	37	225	225	24, 201
Rv2716	DraIII[5]	-	37	208	208	27, 208
Rv3119	HaeII[5]	-	37	403	80, 323	403
Rv3513c	HaeIII[2]	-	37	366	12, 49, 150, 155	12, 150, 204
Rv3616c	TaqI[1]	-	65	292	23, 269	292
Rv3769	DraIII[5]*	3+BSA	37	153	153	27, 126
Rv3867	MluI[1]	-	37	144	25, 119	144
Rv3881c	MvaI[1]	-	37	434	48, 130, 256	178, 256

Restriction enzymes were from the following manufacturers at the following concentrations: **1** Boehringer Mannheim (10 U/µl), **2** Roche (10 U/µl), **3** Stratagene (10 U/µl), **4** Boehringer Mannheim (5 U/µl), **5**, NEB (20 U/µl).* Initial overnight digests of the Rv3769 SNP were performed with DraIII, but later digests performed for triplex diagnostic assays used DraIII-HF for only 2 hr.

2.1.3.6.3 Multiplex Rangipo-Specific SNP RFLP Assays

Multiplex Rangipo-specific SNP RFLP PCR assays were performed using two or three SNPs in one reaction to strengthen the ability of the assay to detect and distinguish between Rangipo and non-Rangipo strains. PCR reactions and RFLP assays were initially carried out using the conditions and parameters used for uniplex assays above (Section 2.1.3.6.2), but then with the optimised PCR condition identified using the modified Taguchi methods (below), and with digests performed for a total of 2 hr only with 0.4 µl of each RE. The optimised triplex PCR condition chosen for the Rangipo SNP-RFLP diagnostic assay were; 1 x PCR buffer, 0.1 mM

dNTPs, 2.4 mM $MgCl_2$, 3 U *Taq* DNA Polymerase, 0.1 µM Rv3119 and Rv1821 primers and 0.2 µM Rv2504c primers, and 10 ng template DNA annealed at 68 °C.

2.1.3.6.4 Taguchi Based Optimisation of Multiplex PCRs

Di and triplex PCRs were optimised using modified Taguchi methods based on (Cobb & Ciarkson 1994; Anilkumar *et al.* 2012). Here, the conditions varied were annealing temperature, and the concentrations of *Taq* DNA polymerase, magnesium, primers and dNTPs. Apart from the conditions being optimised, PCR reaction conditions were the same as those used in Section 2.1.3.6.2, but in a 25 µl total volume. Tables showing the orthogonal array conditions for each are shown in Appendix A.1.

2.1.4 ExoISAP (Exonuclease I - Shrimp Alkaline Phosphatase) Clean Up of PCR Products

Single PCR products were cleaned up with ExoISAP treatment prior to sequencing. The Exonuclease I removes residual single-stranded primers and DNA while the Shrimp Alkaline Phosphatase (SAP) removes remaining dNTPs from the PCR that would otherwise interfere with the labelling step of the sequencing process. 0.5 U SAP (Roche Applied Science, Switzerland) and 10 U Exonuclease I (Fermentas, Lithuania) were added to 10 µl PCR product and incubated for 30 min at 37 °C before the two enzymes were inactivated by incubating for 15 min at 80 °C.

2.1.5 Identification of Rangipo-Specific SNPs

After probable Rangipo-specific SNPs were identified through WGS by Roberto Colangeli and David Alland at the Department of Medicine, University of Medicine and Dentistry of New Jersey (USA) (Section 2.1.2), a shortlist of definite Rangipo-specific SNPs was established. SNP positions had been identified relative to the H37Rv genome, so the H37Rv genome was used as the reference strain to align and screen all other TB strains in the TB Database (Reddy *et al.* 2009). This initial screen was to check for the presence of annotated SNPs in other TB strains at the putative Rangipo-specific SNP position, and the level of genetic

variation/number of SNPs present within each putative SNP containing gene. Any putative Rangipo-specific SNPs which were found within other strains in the TB Database, or within genes containing a high number of annotated SNPs (>20) were ruled out for use as a Rangipo-specific marker.

Candidate Rangipo-specific SNPs were further ranked according to their surrounding genes function to trim the list down to the top 10 Rangipo-specific SNP candidates. Genes of interest included v*apB* or v*apC* genes, or genes involved with virulence, persistence or pathogenesis. Regions of approximately 450 bp spanning the SNP were PCR amplified in Rangipo strains as in Section 2.1.3.6.1 then sequenced to confirm the presence of the SNP.

2.1.6 Development of SNP Specific PCR-RFLP Assays

Rangipo-specific SNPs were initially tested singly in PCR-RFLP assays before testing in combination in order to develop the best assay for detecting *M. tuberculosis* Rangipo isolates.

2.1.6.1 Uniplex Assays

SNPs can be detected by RFLP analysis if only one of the two variant bases creates a cleavage site for a restriction enzyme. WatCut (http://watcut.uwaterloo.ca/watcut/watcut/template.php), an on-line SNP-RFLP analysis program, was used to deduce all possible restriction enzymes that would selectively recognise and cut only one of the two base variants at each SNP site. The choice of restriction enzyme used was based on availability and cost, and preference was given to those enymes which required no or few base changes in the primers used for the assay (with any modifications at least 3bp away from the SNP to anchor it), and enzymes which had few additional recognition sequences in the region being amplified surrounding the SNP. Regions surrounding single SNPs were PCR amplified, then digested with the appropriate restriction enzyme as for Section 2.1.3.6.2.

2.1.6.2 Multiplex Assays

Combinations of two or three PCR-RFLP assays were analysed for assays which produced different size bands in Rangipo and non-Rangipo isolates to easily and robustly distinguish between the two. A combination containing at least one Rangipo and one non-Rangipo SNP variant recognition site as a built in control, and restriction enzymes which worked together at the same temperature, for the same time and in PCR buffer alone was chosen for the final Rangipo diagnostic. Assays were performed as for Section 2.1.3.6.3.

2.1.7 Testing the Rangipo-Specific PCR-RFLP Diagnostic on Clinical Isolates

The Rangipo-specific PCR-RFLP diagnostic was initially tested on H37Rv and Rangipo clinical isolate gDNA before testing both known and blind samples from LabPLUS. Blind samples were sent labelled with an identification number only, they were analysed blind with the Rangipo-specific PCR-RFLP diagnostic and the results sent to LabPLUS who then confirmed whether the samples were Rangipo or not according to their MIRU typing results.

2.2 Methods relating to Chapter Four: RT-qPCR

2.2.1 Bacterial Strains and Growth Conditions

Mycobacterial glycerol stocks were streaked onto blood agar, Luria-Bertani (LB), or LB supplemented with 0.05 % Tween 80 (LBT) plates and incubated at 37 °C until single colonies had formed. Single colonies of *M. smegmatis*: mc^26 (wild-type strain); mc^2155 (laboratory strain) and derived strains (4517 (overexpression strain); and JR121 (deletion strain)) were grown in 10 mL of either LB, LBT, brain heart infusion (BHI) or Sauton's medium. The cultures were incubated at 37 °C and either grown planktonically on a shaker at 200 rpm, or as a biofilm under stationary conditions in 50 ml CellStar tissue culture flasks (Greiner Bio-One Inc, Germany). Cultures used for antibiotic stress experiments were grown as above but with the addition of antibiotics at specific concentrations.

M. smegmatis mc^{2}4517 overexpression strains harbouring pYUBMS1283 (VapB) and pYUBMS1283/4 (VapBC) constructs were obtained from Jo McKenzie (The University of Waikato) according to the methods outlined in McKenzie, J. L. (2011) (See Appendix A.2 for a table of bacterial strains and plasmids).

2.2.2 Total Cellular Protein Determination

Total protein was extracted and estimated from mycobacterial cultures by the Lewin method (Meyers *et al.* 1998) as follows: 1 ml samples were centrifuged (10 min, 13,000 g, 4 °C) then the pellets were washed with 1 ml of phosphate-buffered saline (PBS), pH 7.0 without resuspending the cells, and centrifuged again. At this point the pellets were frozen at -20°C for later analysis, or resuspended in 0.1 ml of 1 M sodium hydroxide (NaOH) and incubated at 99 °C for 10 min. The samples were neutralised by adding 0.02 ml of 5 M HCl, and the volumes adjusted to 1 ml by adding 0.88 ml of PBS, pH 7.0. Samples were centrifuged (30 min, 13,000 g, 4 °C) to pellet the debris, and the supernatant was removed for protein determination using the Quant-iT Protein Assay Kit and Qubit fluorometer (Invitrogen, USA). A Qubit working solution was made by diluting the Qubit protein reagent 1:200 in Qubit protein buffer. An aliquot of 190 µl of working solution was added to 10 µl of standard or protein sample in clear 0.5 ml thin wall Axygen® tubes (Corning, USA), vortexed briefly and incubated at room temperature for 15 min before reading. Before reading samples, the fluorometer was calibrated using three protein standards. Successful calibration of the instrument resulted in a greater than 30 fold difference between the first and second standard values, and a greater than 1.4 fold difference between the second and third standard values.

2.2.3 RNA Isolation

A number of different standard laboratory RNA extraction methods were tested including Tri Reagent (MRC, USA), Trizol (Invitrogen, USA), RNeasy (Qiagen, Netherlands), and Acidic Guanidinium Isothyocynate/Phenol/Chloroform isolation. In order to optimize RNA isolation, modifications of the methods were also tested including: multiple culture volumes; various bead quantities; combinations of TRI Reageant and GITC methods; RNA precipitation with

shorter centrifugation times, and RNA washes without 100 % ethanol. In order to compare the different extraction methods, they were performed in duplicate using aliquots of the same culture sample. All samples were bead beaten using a FastPrep cell disrupter (FP120) (Thermo Savant, USA) on setting 6. Isolated RNA was used immediately or stored at −80 °C until use.

2.2.3.1 TRI Reagent and Trizol

Total RNA was isolated using TRI Reagent or Trizol following the manufacturer's instructions, with the additional step of disrupting the cells by bead beating. Briefly, 1.5 ml of culture (approximately between 10^6 and 10^7 cells) was pelleted then homogenised in 1 ml of TRI Reagent or Trizol. The homogenate was beaten with 0.3 g of 0.1 mm and 2.5 mm zirconia beads four times for 20, 25, 30 and 35 s respectively with 1 min rests in between. Either 0.2 ml chloroform or 0.1 ml of 1-bromo-3-chloropropane (BCP) was added and shaken vigorously for 1 min, then left for 5 min at room temperature. After centrifugation (13,000 rpm, 15 min, 4 °C) the upper aqueous phase was carefully transferred into a sterile tube. The RNA was precipitated with 0.5 ml of isopropanol and incubated at room temperature for 10 min, then pelleted by centrifugation (13,000 rpm, 15 min, 4 °C). The precipitated RNA was washed in 1 ml 70 % ethanol and centrifuged (13,000 rpm, 5 min, 4 °C), briefly air-dried, and then solubilised in 35 µl 10 mM Tris pH 7.4, 0.5 mM $MnCl_2$.

2.2.3.2 Acidic Guanidinium Isothiocyanate/Phenol/BCP Method

Total RNA was isolated based on the Acidic Guanidinium Isothyocynate/Phenol/Chloroform (AP-GITC) procedure (Chomczynski & Sacchi 1987) with a few modifications, namely, BCP was used instead of chloroform and two successive phenol-BCP extractions were performed instead of one. A culture sample (1.5 ml) (approximately between 10^6 and 10^7 cells) was added to 6 ml of 5 M GITC, pH7, mixed by inverting and pelleted by centrifugation (3,000 g, 10 min, 4 °C). The pellet was resuspended in 0.5 ml of fresh 5 M GITC, pH7, transferred to a sterile tube containing 0.3 g of 0.1 mm and 2.5 mm zirconia beads and stored at −80 °C until use. Cells were disrupted by bead beating as described above (Section 2.2.3.1) and briefly centrifuged to remove the resulting foam. NaOAc (50 µl of 2 M, pH 4) and 0.5 ml water

saturated phenol were added, vortexed for 15 s then rotated for 10 min at room temperature. After addition of 100 µl BCP, the sample was shaken vigorously for 1 min, incubated for 5 min on ice and centrifuged (13,000 rpm, 15 min, 4 °C). The aqueous phase was re-extracted by repeating the two previous steps in a new tube, scaling up the volumes accordingly to account for any increase in aqueous phase volume. The RNA contained in the aqueous phase was transferred to a new tube and precipitated with an equal volume of isopropanol before being incubated on ice for 1 hr. The RNA was pelleted by centrifugation (13,000 rpm, 30 min, 4 °C), washed in 1 ml 70 % ethanol and centrifuged again (13,000 rpm, 10 min, 4 °C). Finally, the RNA was washed in 1 ml 100 % ethanol, centrifuged (13,000 rpm, 10 min, 4 °C), briefly air-dried and then solubilised in 35 µl 10 mM Tris pH 7.4, 0.5 mM $MnCl_2$.

2.2.3.3 RNeasy

RNA was extracted using the RNeasy Mini Kit (Qiagen, Netherlands) according to the manufacturer's instructions with the exception that the cells were homogenised via the bead beating disruption method, and the RNA was eluted in 10 mM Tris pH 7.4, 0.5 mM $MnCl_2$.
A culture sample (1.5 ml) (approximately between 10^6 and 10^7 cells) was pelleted (13,000 rpm, 5 min, 4 °C) then resuspended in 600 µl Buffer RLT. The cells were disrupted by beating with 0.3 g of 0.1 mm and 2.5 mm zirconia beads as described above (Section 2.2.3.1). RNA was precipitated with 600 µl 70 % ethanol and then 700 µl was transferred to an RNeasy spin column (placed in a 2 ml collection tube) and centrifuged (13,000 rpm, 15 s, 4 °C). Successive 700 µl aliquots were centrifuged in the same column. The column was washed with 700 µl Buffer RW1 and centrifuged (13,000 rpm, 15 s, 4 °C), then 500 µl of Buffer RPE and centrifuged. The column was washed one additional time and centrifuged for longer to ensure complete removal of ethanol (13,000 rpm, 2 min, 4 °C). The column was transferred to a new 2 ml collection tube and centrifuged to remove residual buffer (13,000 rpm, 1 min, 4 °C). Finally, RNA was eluted by transferring the column to a new 1.5 ml tube, adding 35 µl 10 mM Tris pH 7.4, 0.5 mM $MnCl_2$ and centrifuged (13,000 rpm, 1 min, 4 °C). A further column wash was performed using the eluate, to ensure as much concentrated RNA was retrieved from the column as possible.

2.2.4 RNA Purity & Quantification

Total RNA purity and concentration were estimated for all extraction methods by measuring A_{260}/A_{230} and A_{260}/A_{280} ratios with the NanoDrop®ND-1000 Spectrophotometer (NanoDrop Technologies, USA), and by running 800 ng on a 1.5 % agarose gel containing 0.5 µg/ml ethidium bromide (EtBr), to check for intact bands representing the 23S and 16S fragments.

2.2.5 DNase Treatment

RNA was treated with DNase to eliminate free DNA. One unit of RQ1 RNase-free DNase (Promega, USA) was added to 1 µg of RNA in 10 µl 10 mM Tris-0.5 mM $MnCl_2$, pH 7.4 and incubated at 37 °C for 30 min in a Thermomixer (Eppendorf, Germany) at 900 rpm. The DNase was inactivated by the addition of 1 µl RQ1 DNase Stop Solution (Promega, USA) and incubating at 65 °C for 10 min, 900 rpm.

2.2.6 Reverse Transcription

Total RNA was reverse-transcribed using SuperScript III Reverse Transcriptase (RT) (Invitrogen, USA) with random primers in a 20 µl total volume using Invitrogen reagents unless specified. DNase treated RNA (1 µg) was incubated with 100 ng random primers at 65 °C for 5 min, cooled on ice for 1 min followed by the addition of 1 X RT buffer, 5 mM DTT, 0.5 mM dNTP mix (Fermentas, Lithuania) and 140 U SuperScript III RT. The reactions were incubated at 25 °C for 10 min then 50 °C for 1 hr, before they were terminated at 85°C for 5 min and chilled on ice. For each sample, a negative RT (no addition of reverse transcriptase) reaction was performed and used as a negative control in subsequent reverse transcription quantitative real-time PCR (RT-qPCR).

2.2.7 Reverse Transcription Quantitative Real-Time PCR (RT-qPCR)

2.2.7.1 Primer Design

Primers for each gene were designed based on the published *M. smegmatis* mc^{2}155 sequence (GenBank accession number NC_008596) with the aid of Primer 3 (Koressaar & Remm 2007) and Vector NTI (Life Technologies, USA) software. Where possible, primers were designed to have similar properties so that PCRs for different genes could be performed in the same run. Secondary template structure was avoided where possible and primers were obtained from Sigma Aldrich (USA) or IDT (USA).

2.2.7.2 Reaction Set Up

All RT-qPCR reactions were carried out in triplicate 20 µl volumes using either SYBR Green I or SYTO 82 (Invitrogen, USA). A master mix was prepared consisting of 1 X Thermo-Start PCR Buffer (ABgene, UK), 5 mM $MgCl_2$ (ABgene, UK) and 200 µM each dNTP (Fermentas, Lithuania), then DNase treated with 10 U/ml of RQ1 RNase-Free DNase (Roche Applied Science, Switzerland) at 37 °C for 1 hr at 900 rpm, inactivated at 80 °C for 30 min, and stored at -20 °C until use. A PCR master mix was prepared containing the above master mix, a fluorescent nucleic acid dye (0.5 X SYBR Green I or 2 µM SYTO 82) and 36 U/ml Thermo-Start Taq DNA Polymerase (ABgene, UK). For each primer set, a qPCR master mix was prepared containing PCR master mix, 125 nM primers and 10 µl of diluted cDNA. cDNA was diluted in qPCR master mix at a ratio of 0.075 µl per 1 µl of qPCR master mix, except for the 16S reference samples where the diluted cDNA was further diluted to give 0.0001125 µl cDNA per 1 µl of qPCR master mix. Each sample contained 19 µl of the qPCR master mix, 0.25 µl of primer (10 µM) and 0.75 µl of cDNA, and each 16S reference sample contained 19.75 µl of the qPCR master mix, 0.25 µl of primer and 0.00125 µl of cDNA.

2.2.7.3 qPCR Run

Real-time PCR was performed on the Rotor-Gene 6000 (Corbett Research, Australia). The reaction conditions were; 1 cycle of 95 °C for 15 min, followed by 35 cycles of 95 °C for 20 s,

55 °C for 20 s, 68 °C for 30 s, and 80 °C for 10 s with data acquired on the 80 °C step in the green channel (470 - 510 nm) for SYBR Green, and yellow channel (530 - 555 nm) for SYTO 82. Immediately following amplification, a melt curve was performed consisting of 90 s at 72 °C followed by 5 s steps with 0.5 °C increase in temperature per step rising to 99 °C. The identity of the PCR products was verified by the melt curve profile and electrophoresis on a 1 % agarose gel (Section 2.2.7.5). Each RT-qPCR run included a no template control (NTC) per amplicon and a no RT control per sample.

2.2.7.4 Analysis

Initially, the Wilkening & Bader (2004) equation (below) was used to calculate the initial copy number (ICN) for the housekeeping gene and GOI, using quantification cycle (Cq) and efficiency values generated from the Rotor-Gene 6000 software.

$N_0 = 1.12 \times 10^{10}/E\text{^}CN$, where

N_0 = initial copy number or starting concentration

E = PCR amplification efficiency

^ = denotes 'to the power'

CN = the cycle number at the Cq

The ICNs for the GOI were then normalised by dividing by the ICN of the houseeping gene to determine the relative level of expression.

Later, RT-qPCR assays were analysed using the more accurate equation used in LinRegPCR. Here, the PCR efficiency for each individual sample was estimated using LinRegPCR (version 11.1) software based on the slope of the log-transformed fluorescent data of the exponential phase (Ramakers *et al.* 2003). Data were normalised to a reference gene and the resulting mRNA expression level ratios were determined using the efficiency corrected calculation below.

$$R = \frac{AVERAGE\ (N0_{GOI})}{AVERAGE\ (N0_{HK})}$$, where

R = expression ratio of the gene of interest, normalised to the HK

N0 = Nq / (E(MEAN)^Cq) , where

N0 = starting concentration (expressed in arbitrary fluorescence units)

Nq = fluorescence threshold set to determine Cq

E(MEAN) = mean amplicon PCR efficiency

^ = denotes 'to the power'

Cq = number of cycles needed to reach Nq (previously referred to as Ct)

First, the starting concentration (N0) values of technical replicates were averaged (per biological sample) for the GOI samples and the HK samples, and then the ratio of these two averages was calculated to determine the relative expression value.

2.2.7.5 Agarose Gel Electrophoresis

DNA fragments were separated via agarose gel electrophoresis. Samples were mixed with 6 x DNA loading dye prior to loading, and run on a 1 % TAE gel containing 0.3 µg/ml EtBr. Gels were visualised by UV light, and images captured. Band sizes were determined by visual comparison against a 100 bp DNA ladder (Invitrogen, USA, or GenScript, USA).

2.3 Methods relating to Chapter Five: *M. tuberculosis* VapBC Characterisation

2.3.1 General Methods/DNA manipulations

2.3.1.1 Antibiotic stocks

Hygromycin and kanamycin stocks were prepared at 50 mg/ml in MQ water then filter sterilised, and used at a final concentration of 50 µg/ml.

2.3.1.2 *M. tuberculosis* gDNA

M. tuberculosis H37Ra genomic DNA was obtained from the School of Biological Sciences, University of Auckland.

2.3.1.3 Polymerase Chain Reaction (PCR)

M. tuberculosis vapBC operons were amplified from H37Ra genomic DNA then cloned into the vector pYUB28b. The vector was transformed into *E. coli* prior to transformation and protein expression in *M. smegmatis*. All H37Ra *vapBC* sequences were compared with those from H37Rv to ensure the amplified product would be 100 % identical to that from H37Rv. Primer sequences can be found in Appendix A.1.

2.3.1.3.1 Primer Design

All primers for cloning were designed using Geneious Pro (Version 5.02) (BioMatters Ltd, NZ) to include 5' and 3' restriction sites necessary for ligation into pYUB28b including a C-terminus His tag. An extra 3 – 6 bases were included to allow for efficient cleavage (dependent on the restriction enzymes used). An NcoI site was included at the 5' end and where the NcoI sequence was present within the gene, the gene was ordered from GeneArt® (Life Technologies, USA) with the nucleic acid sequence altered to produce the same amino acid sequence but without the internal NcoI site. HindIII was used primarily as the 3' restriction site, but where HindIII was present within the gene NdeI or BamHI were used. Where BamHI was used, an extra two bases were added in front of the BamHI recognition sequence to put the His tag back in-frame. The choice of restriction sites were based on including a C-terminus His tag and removing the N-terminus His tag, limiting the linker distance between the restriction enzyme and T7 sites, and using sites which occurred only once in the vector and no where in the gene. All putative sequences were virtually translated within the pYUB28b vector to ensure both the VapB and VapC protein sequences were correct, and the C-terminal His tag was in-frame.

Primers for amplification of the plasmid by itself contained no flanking restriction sites. All primers were supplied by IDT (USA) and resuspended in 1 x TE.

2.3.1.3.2 Polymerase Chain Reaction Amplification of vapBC Operons for Ligation into pYUB28b

All PCR reactions amplifying inserts for ligation were carried out with the proofreading enzyme Platinum *Pfx*® DNA polymerase (Invitrogen, USA). The annealing temperature was optimised for each primer set by initially performing a gradient PCR spanning 5 °C above and below the calculated Tm of the primers. PCR reactions were carried out in 25 µl volumes using the following concentrations: 1 x *Pfx* amplification buffer, 0.3 mM deoxynucleotide mix (dATP, dCTP, dTTP & dGTP), 1 mM $MgSO_4$, 1 U Platinum *Pfx*® DNA polymerase, 1 µM each primer and 100 ng template DNA. The following cycling conditions were used for amplification:

94 °C	2:00		(min:sec)
94 °C	0:30	┐	
56 – 66 °C*	0:30	│ x 29	
68 °C	1:00	┘	
68 °C	7:00		

*Annealing temperature determined by gradient PCR

For primer sets where no temperature within the gradient produced clean product, the magnesium concentration was further optimised using a selection of $MgSO_4$ concentrations between 0.75 mM and 3 mM. PCR reactions were visualised as in Section 2.3.1.4 below.

2.3.1.3.3 PCR Screening of pYUB28b Transformants

All pYUB28b plasmids purified from *E. coli* (Section 2.3.1.11) were checked for the presence of insert prior to sequencing by performing PCR with T7 primers. PCR reactions were carried out with *Taq* DNA polymerase (Invitrogen) in 25 µl volumes using the following concentrations: 1 x PCR buffer, 0.2 mM deoxynucleotide mix (dATP, dCTP, dTTP & dGTP), 1.5 mM $MgCl_2$, 2.5 U *Taq* DNA Polymerase, 1 µM each primer and 30 - 60 ng template DNA. The following cycling conditions were used for amplification:

94 °C	2:00		(min:sec)
94 °C	0:30	┐	
55 °C	0:30	x 29	
72 °C	0:45	┘	
72 °C	5:00		

PCR reactions were visualised as in Section 2.3.1.4 below.

2.3.1.4 Agarose Gel Electrophoresis

DNA fragments were separated via agarose gel electrophoresis. Samples were mixed with 10 x DNA loading dye prior to loading, and run on a 1 % TAE gel. Agarose gels were post-stained for 20 min with either 0.5 µg/ml EtBr, or 1 x SYBR Safe™ DNA gel stain (Invitrogen, USA) if products were going to be recovered. Gels were visualised by UV light (EtBr) or on a blue light box (SYBR Safe™) if products were going to be recovered, and images captured. Band sizes were determined by visual comparison against a 1kb-Plus DNA ladder (Invitrogen, USA).

2.3.1.5 DNA Purification and Quantification

DNA from agarose gels were separated and visualised with SYBR Safe™ DNA dye as described in Section 2.3.1.4. The desired bands were extracted using a sterile scalpel blade and DNA purified using the QIAquick Gel Extraction Kit (Qiagen, Netherlands) according to the manufacturer's instructions. DNA was eluted in 30 µl of elution buffer.

PCR products in solution from two 50 µl PCR reactions were pooled and purified using a Qiaquick PCR Product Purification Kit (Qiagen, Netherlands) according to manufacturer's instructions.

DNA purity and concentration was estimated by measuring A_{260}/A_{230} nm and A_{260}/A_{280} nm ratios using a NanoDrop®ND-1000 Spectrophotometer (NanoDrop Technologies, USA).

2.3.1.6 Restriction Enzyme Digestion

Restriction enzymes were purchased from Invitrogen (USA), and digests were performed on gel purified PCR products (Section 2.3.1.5) or plasmid DNA (Section 2.3.1.11) singly due to buffer incompatibilities. Less restriction enzyme was used compared to the manufacturer's recommendation in order to reduce costs. NcoI (1 µl) was added to 28 µl of DNA along with the appropriate amount of buffer, BSA and water for a 100 µl reaction according to the manufacturer's instructions. Digests were incubated at 37 °C with shaking at 600 rpm for 3 hrs or overnight, and then purified from solution according to Section 2.3.1.5. The second digest was performed on the purified product with either 2 µl HindIII, BamHI or NdeI, and treated as for the first restriction enzyme digest.

2.3.1.7 pYUB28b Vector Dephosphorylation

Digested pYUB28b vector (30 µl) was dephosphorylated with 5 U Shrimp Alkaline Phosphatase (SAP) and 1 x SAP dephosphorylation buffer (Roche Applied Science, Switzerland) at 37 °C with shaking at 600 rpm for 30 min. The reaction was inactivated by incubating at 65 °C with shaking at 600 rpm for 15 min.

2.3.1.8 DNA Ligation

DNA ligations were set up using 1 U of T4 DNA Ligase (Invitrogen, USA), and a vector to insert ratio of 30 fmol : 90 fmol in a total volume of 10 µl. Ligations were incubated for 6 hrs at room temperature, overnight at 14 °C or over the weekend at 4 °C.

2.3.1.9 Preparation of Electrocompetent *E. coli* TOP10 Cells

A glycerol stock of *E. coli* TOP10 cells was streaked onto a low salt LB agar plate and incubated at 37 °C overnight. An aliquot of low salt LB (10 ml) was inoculated with a single colony and grown overnight at 37 °C, 200 rpm. The overnight culture was used to inoculate 1 L of low salt LB, which was grown at 37 °C, 200 rpm until it reached an OD_{600} of 0.5 - 0.7. All subsequent steps were performed with ice-cold 10 % glycerol, in pre-chilled, sterile containers.

The cells were chilled on ice for at least 20 min, centrifuged (4,000 g, 15 min, 4 °C), the supernatant removed and the pellet resuspended in 500 ml 10 % glycerol. The spin and wash step was repeated two further times, resuspending in 250 ml then 20 ml respectively. The cells were transferred to a 50 ml falcon tube and centrifuged (4,000 g, 15 min, 4 °C) one final time before resuspending in 2 ml 10 % glycerol, flash freezing 50 µl aliquots in liquid nitrogen and storing at -80 °C.

2.3.1.10 Electroporation of *E. coli*

Plasmid DNA (1 µl) or ligation reaction was added to 50 µl freshly thawed (on ice) electrocompetent *E. coli* TOP10 cells. The mixture was placed in a 2 mm electroporation cuvette (BioRad Laboratories, USA), tapped to remove air bubbles and electroporated with a Bio-Rad Gene PulserTM (Bio-Rad Laboratories, USA) at 2.5 kV, 25 µF capacitance and 200 Ω resistance. Transformed cells were recovered immediately in 1 ml of SOC media and incubated at 37 °C shaking at 200 rpm for 30 min. Aliquots of the cells (50 µl and 100 µl) were plated onto low salt LB agar plates containing hygromycin for selection and incubated at 37 °C overnight. Single colonies from these agar plates were used to inoculate 5 ml low salt LB cultures containing hygromycin. These cultures were incubated at 37 °C shaking at 200 rpm overnight and used to isolate plasmid DNA according to the method in Section 2.3.1.11.

2.3.1.11 *E. coli* Plasmid DNA Extraction

E. coli plasmid DNA was extracted from 5 ml overnight cultures using the QIAprep Spin Miniprep Kit (Qiagen, Netherlands) according to manufacturer's instructions, and eluted in 50 µl elution buffer.

2.3.1.12 Sequencing of DNA

DNA was sequenced by the Waikato DNA Sequencing Facility on an Applied Biosystems 3130xl Genetic Analyzer at The University of Waikato.

2.3.1.13 Preparation of Electrocompetent *M. smegmatis*

A glycerol stock of *M. smegmatis* 4517 cells was streaked onto a low salt LBT + kanamycin agar plate and incubated in a humidity chamber at 37 °C for three days. Cultures of 5 ml low salt LBT + kanamycin were inoculated with a single colony and grown for 48 hrs at 37 °C at 200 rpm. The 5 ml culture was used to inoculate 500 ml of 7H9 + ADC + Tween 80 + kanamycin, which was grown at 37 °C at 200 rpm until it reached an OD_{600} of 0.6 - 0.7.
All subsequent steps were performed with ice-cold 10 % glycerol, in multiple pre-chilled, sterile 50 ml falcon tubes rather than large centrifuge bottles (in order to achieve a firmer pellet). The cells were chilled on ice for 1.5 hrs, centrifuged (4,000 g, 20 min, 4 °C), the supernatant removed and then the pellet resuspended in 50 ml 10 % glycerol. The spin and wash step was repeated two further times, resuspending in 50 ml then 5 ml respectively. The cells were pooled and transferred to a single 50 ml falcon tube then centrifuged (4,000 g, 20 min, 4 °C) one final time before resuspending in 500 µl 10 % glycerol, flash freezing 40 µl aliquots in liquid nitrogen and storing at -80 °C.

2.3.1.14 Electroporation of *M. smegmatis*

Plasmid DNA (1 µl) was added to 40 µl freshly thawed (on ice) electrocompetent *M. smegmatis* 4517 cells. The mixture was placed in a 2 mm electroporation cuvette (BioRad Laboratories, USA), tapped to remove air bubbles and electroporated with a Bio-Rad Gene PulserTM (Bio-Rad Laboratories, USA) at 2.5 kV, 25 µF capacitance and 1000 Ω resistance. Transformed cells were recovered immediately in 1 ml of 7H9 media and incubated at 37 °C shaking at 200 rpm for 3 hrs. Aliquots of the cells (50 µl and 100 µl) were plated onto 7H10 agar plates containing hygromycin and kanamycin for selection and incubated at 37 °C for 3 days in a humidity chamber.

2.3.1.15 Glycerol Stocks

Glycerol stocks for long term storage of transformed bacteria at -80 °C were made by adding 0.5 ml of sterile 50% glycerol to 0.5 ml of overnight culture (LB + hygromycin for *E. coli* or LBT + hygromycin and kanamycin for *M. smegmatis* + pYUB28b plasmid).

2.3.2 Protein Expression and Purification

2.3.2.1 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) Protein Analysis

SDS-PAGE gels were cast in a Hoefer mini gel casting system. SDS-PAGE gels consisted of a 5 % acrylamide stacking gel overlaid on a 15 or 16.5 % acrylamide resolving gel (depending on the protein size). All SDS-PAGE gels were made up with 30 % acrylamide with an acrylamide:bisacrylamide ratio of 37.5:1 (Bio-Rad Laboratories, USA), included 0.1 % filtered SDS, and were polymerised by the addition of 0.05 % ammonium persulphate (APS) and TEMED.

Protein samples were mixed in a 3:1 ratio with 4 x SDS loading buffer and denatured for 5 min at 95 °C prior to loading onto the gel. Gels were run with 1 x SDS-PAGE running buffer at constant 70 V until the samples entered the resolving gel, then at 150 V until the dye front reached the end of the gel. For protein MW estimation, 10 µl of Precision Plus Protein ™ Unstained Standard (Bio-Rad Laboratories, USA) was run alongside each gel.

2.3.2.2 Native Polyacrylamide Gel Electrophoresis (Native-PAGE) Protein Analysis

Native-PAGE gels of varying concentrations (depending on the size of the proteins being separated) were cast and run as for SDS-PAGE gels (Section 2.3.2.1), except with no SDS in the gel, running buffer or loading dye.

2.3.2.3 Isoelectric focusing (IEF) SDS-PAGE

The two-dimensional electrophoresis was carried out with immobilized pH gradient-IEF in the first, and SDS-PAGE in the second dimensions.

IEF was performed using the Bio-Rad PROTEAN II IEF Cell and 7 cm linear, pH 3-10 ReadyStrip ™ IPG strips (Bio-Rad Laboratories, USA). Rehydration was carried out at room temperature overnight with either 2 µg or 5 µg of purified protein, plus 45 µl IEF sample buffer

and 80 µl re-hydration buffer per strip. Isoelectrofocusing was performed at 200 V for 10 min, 1500 V for 1.5 hr and 3500 V for 1 hr and 15 min. Following IEF, the strips were reduced then alkylated by soaking and gently rocking for 20 min in 2 ml Equilibration Buffer A, then 2 ml Equilibration Buffer B.

To separate the proteins in the second dimension by SDS-PAGE, the IPG strips were sealed on top of a 1 mm thick 14 % polyacrylamide gel with 0.5 % agarose in 1 x stacking buffer + bromophenol blue. Vertical electrophoresis was carried out at 75 V until the samples had entered the resolving layer, then 25 V for approximately 1 hr. Gels were fixed in 50 % methanol + 2 % phosphoric acid for three 30 min incubations, then stained overnight in Coomassie Brilliant Blue G-250 before de-staining in 10 % methanol. The image was captured using a GS-800 calibrated densitometer (Bio-Rad), and spots detected with the PDQuest v6.0 software package (Bio-Rad).

2.3.2.4 Coomassie Blue Stain for Protein Gel Electrophoresis

Gels were stained by colloidal Coomassie staining using the quick stain method (Wong *et al.* 2000) with Fairbanks staining solutions A-D. Gels were microwaved for 30 s in 50 ml Fairbanks staining solution A, cooled to room temperature while shaking gently, the stain removed then repeated as above with stains B, C and lastly D (de-stain).

2.3.2.5 Quantification of Protein Concentration

The NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, USA) was used to measure absorbance at 280 nm and the Beer-Lambert equation ($A=\varepsilon.c.l$) applied to correlate absorbance with concentration. Here A= absorbance at 280 nm, ε= the theoretical extinction coefficient of the protein, c= concentration, and l=pathlength. Molecular weights of the proteins are listed in Table 5.2. Theoretical extinction coefficients were calculated based on the amino acid composition using Geneious Pro (Version 5.02 Biomatters Ltd, NZ) or the ExPASy ProtParam tool (Gasteiger *et al.* 2005).

2.3.2.6 Concentration of Protein Samples

Vivaspin concentrators (10 kDa molecular weight cut-off) (Sartorius AG, Germany) were used to concentrate protein samples. Protein samples were added to the upper reservoir of 20 ml or 500 µl concentrators and spun at 3,900 g at 4 °C until the desired volume or concentration was reached.

2.3.2.7 Dialysis of Protein Samples

Protein samples were dialysed into target buffer using 6 - 8 kDa molecular weight cut off Spectra Por® dialysis tubing (Spectrum Laboratories, USA). Samples were loaded into pre-wet tubing and the ends sealed using dialysis clips before dialysing into 1 L of gently stirring target buffer at 4 °C. The buffer was replaced after 2 hrs and protein left to completely dialyse overnight.

2.3.2.8 Buffer Exchange

A 5 ml HiTrap Desalting column (GE Healthcare, UK) was used to desalt protein samples into a buffer containing less salt. The column was equilibrated with 25 ml target buffer at 5 ml/min before 1.5 ml of protein sample was added. The column was flushed with 5 ml of buffer and the eluent (containing the desalted protein) collected. A final wash step with 25 ml of water was performed before re-equilibrating the column with 25 ml 20 % ethanol and storing at 4 °C.

2.3.2.9 MALDI-TOF Mass Spectrometry (MS) of Whole Protein

2.3.2.9.1 Sample Preparation

2,5-dihydrozybenzoic acid and 2-hydroxy-5-methoxybenzoic acid (Super-DHB) matrix (5 mg), was mixed with 20 µl acetonitrile and 10 µl 0.1 % T trifluoroacetic acid (TFA) and fully resuspended by 10 min sonication in a sonicating water-bath (Elma, Germany). The matrix was spun (13,000 rpm for 10min) and mixed with sample or calibrant at a 1:1 ratio (0.5 µl:0.5 µl) before spotting onto the AnchorChip™ MALDI-TOF target plate (Bruker Daltonics, USA) then air dried prior to MALDI-TOF MS analysis.

2.3.2.9.2 MALDI-TOF Set Up

MALDI-TOF analysis was performed with the Autoflex™ II MALDI-TOF mass spectrometer (Bruker Daltonics, USA). Analysis were obtained in the reflectron positive ion linear mode with a 450 ns pulse, range of 5 - 20 kDa, accelerated voltage in the ion source set at 20 kV and laser power at around 60%. Spectra for the Protein Standard I (Bruker Daltonics, USA) were collected first, and the spectrometer calibrated to this with an automatic polynomial correction.

Spectra were exported to DataAnalysis™ or FlexAnalysis™ software (Bruker Daltonics, USA), and peaks identified and labelled.

2.3.2.10 Cloning of TB VapBCs into pYUB28b

pYUB28b is an *E. coli* – mycobacterial shuttle vector with hygromycin resistance. The TB *vapBC* operon open reading frames (ORFs) were amplified from the attenuated *M. tuberculosis* strain, H37Ra after confirming there was 100% sequence homology with the H37Rv strain (Section 2.3.1.3). The primers used for amplification were designed to contain an NcoI restriction site in the forward primer and either a BamHI, NdeI or HindIII restriction site in the reverse primer. The amplified products were digested with the appropriate restriction enzymes (Invitrogen, USA) (Section 2.3.1.6), purified using spin columns (Roche Applied Science, Switzerland) (Section 2.3.1.5) and ligated into the pYUB28b shuttle vector between the NcoI and BamHI/NdeI/HindIII restriction sites enabling expression with a C-terminal His-tag (Section 2.3.1.8). Vectors containing the *vapBC* operons were transformed into *E. coli* TOP10 cells (Section 2.3.1.10) and plated onto low salt LB agar plates containing hygromycin for selection. The plasmid constructs were purified then sequenced to confirm correct insertion before being transformed into *M. smegmatis* $mc^2$4517 electrocompetent cells (Section 2.3.1.14) and plated onto 7H10 agar plates containing hygromycin and kanamycin for selection.

2.3.2.11 Small Scale Protein Expression and Purification from *M. smegmatis*

Expression of *M. tuberculosis* VapBCs in *M. smegmatis* were initially confirmed with small scale His-tag binding tests using Ni Sepharose beads. A single transformed colony either from the freshly transformed plate or a glycerol stock streaked plate was used to inoculate a 5 ml PA-0.5G + 0.05 % Tween 80 seeder culture. This culture was grown for 48 hr at 37 °C, 200 rpm then 100 µl was used to seed a 10 ml ZYP-5052 + 0.05 % Tween 80 expression culture which was grown for 90 hr at 37 °C, 200 rpm. Cultures were spun (4,600 rpm, 20 min, 4 °C) and cell pellets resuspended in 1 ml lysis buffer (50 mM phosphate buffer pH 7.4, 200 mM NaCl, 20 mM imidazole) before 3 µl was removed and added to 12 µl lysis buffer and 5 µl 4 x SDS loading dye (named whole-cell sample). The remainder of the cells were sonicated on ice using a fine tip probe for 12 s bursts at setting 3.5 with a Misonix Sonicator (USA) until cells were lysed. The lysate was spun (13,000 rpm, 20 min, 4 °C), the insoluble fraction was resuspended in 200 µl lysis buffer, then 3 µl of soluble and insoluble fractions were added to 12 µl lysis buffer and 5 µl 4 x SDS loading buffer (named L and I respectively).
50 µl of Ni Sepharose (GE Healthcare, UK) was transferred to a 1.5 ml eppendorf tube and washed with 1 ml of lysis buffer, then centrifuged (3,000 rpm for 1 min) to settle the beads. The wash supernatant was discarded and 200 µl of the sample's soluble fraction (L) was added and then incubated at room temperature for 15 min at 1,000 rpm. The beads were spun again (3,000 rpm for 1 min) and supernatant removed before a 15 µl aliquot of the supernatant was added to 5 µl 4 x SDS loading buffer (named FT). The beads were washed with 1 ml of lysis buffer then spun (3,000 rpm for 1 min) and supernatant removed (named W1), before repeating the wash step another two times. An aliquot of W1 (15 µl) was added to 5 µl 4 x SDS loading buffer. 50 µl 4 x SDS loading buffer was added to the beads then 20 µl from all six samples were analysed by SDS-PAGE.

2.3.2.12 Large Scale VapBC Protein Expression and Purification from *M. smegmatis*

Large scale VapBC protein expression cultures were grown as for small scale protein expression cultures in Section 2.3.2.11, except this time 10 ml seeder cultures were used to

inoculate 1 L expression cultures (again, in a 1:100 ratio). Cultures were spun (4,600 rpm, 20 min, 4 °C) and cell pellets either frozen at -80 °C for later use, or resuspended in 25 ml lysis buffer (50 mM phosphate buffer pH 7.4, 200 mM NaCl, 20 mM imidazole) with the addition of an EDTA free protease inhibitor tablet (Roche Applied Science, Switzerland). The cells were sonicated on ice using a large probe for 30 s bursts at setting 7 – 8.5 with 1 min rests in between, on a Misonix Sonicator (USA) until cells were lysed (5 – 8 min). The lysate was spun (13,000 rpm, 20 min, 4 °C) and the insoluble fraction resuspended in 20 ml lysis buffer, then 3 µl of soluble and insoluble fractions were added to 12 µl lysis buffer and 5 µl 4 x SDS loading buffer for later SDS-PAGE analysis. The remainder of the soluble fraction (containing the His-tagged protein) was run through a HisTrap HP column (GE Healthcare, UK) or Ni-NTA Column (MACHEREY-NAGEL, Germany) and purified via Immobilised Metal Affinity Chromatography (IMAC) on an ÄKTA FPLC system as in Section 2.3.2.13. Fractions containing the protein of interest were purified further by size exclusion chromatography (SEC) using a Superdex column (GE Healthcare, UK) as in Section 2.3.2.14. Aliquots of fractions (15 µl) containing the protein of interest from both IMAC and SEC were added to 5 µl of 4 x SDS loading buffer for analysis with SDS-PAGE.

2.3.2.13 Purification of His-Tagged Proteins via Immobilised Metal Affinity Chromatography (IMAC)

A 5 ml HiTrap™ Chelating HP column (GE Healthcare, UK) or 1 ml Protino® Ni-NTA Column (MACHEREY-NAGEL, Germany) was pre-equilibrated with lysis buffer (50 mM phosphate buffer pH 7.4, 200 mM NaCl, 20 mM imidazole) at a flow rate of 1 ml/min. The supernatant from lysed large scale expression cultures was filtered through 1.2 and 0.45 µm Minisart filters (Sartorius AG, Germany) then loaded onto the IMAC column which was then attached to an ÄKTA Prime™, ÄKTA Basic™ or ÄKTA Purifier™ FPLC system. Unbound proteins were removed by washing the column with 15 - 25 ml of lysis buffer at a flow rate of 1 ml/min, before bound proteins were removed using a gradient of 0 – 50 % elution buffer (lysis buffer + 1 M imidazole) over 50 ml or 19 ml (for 5 ml or 1 ml columns respectively) at a flow rate of 1 ml/min. Aliquots of fractions (15 µl) were then analysed by SDS-PAGE (Section 2.3.2.1). The IMAC

column was stripped of Ni^{2+} ions using 10 ml 100 mM EDTA pH 8.0, washed with 20 ml water and re-primed with 5 ml 100 mM $NiCl_2$ between purifications of different proteins.

2.3.2.14 Size Exclusion Chromatography (SEC)

Proteins were purified further by size exclusion chromatography (SEC) using either a preparative HiLoad™ 16/60 Superdex™ 200 pg (S200 16/60) or analytical Superdex™ 200 10/300 (S200 10/300) column (GE Healthcare, UK) which have a separation range of 10 - 600 kDa. Proteins were concentrated to 5 ml or 500 µl respectively, and filtered with a Nanosep® MF 0.2 µm filter (Pall, USA) before loading and running on a pre-equilibrated (50 mM phosphate buffer pH 7.4, 200 mM NaCl) column. Aliquots of fractions (15 µl) were then analysed by SDS-PAGE (Section 2.3.2.1) and if necessary, agarose gel electrophoresis (Section 2.3.1.4) to identify fractions containing protein which was purified along with DNA.

Further SEC of proteins bound to DNA were performed using the S200 10/300 column as above but in a high salt (2 M NaCl) buffer to dissociate the co-purified DNA.

Calibration of the S200 16/60 and 10/300 analytical size exclusion columns were carried out by Jo McKenzie (Proteins & Microbes Laboratory, University of Waikato) using the high & low molecular weight gel filtration calibration kits (GE Healthcare, UK) according to manufacturer's instructions. A calibration curve was prepared for each column by plotting the Kav value (elution volume function) of the protein standards against their log molecular weight, to generate the following equation for determining Kav values for the VapBC proteins:

$$Kav = (Ve - Vo) / (Vc – Vo)$$

Where Ve = elution volume
Vo = column void volume
Vc = geometric column volume

The Kav values for VapBC proteins were substituted into equations (determined previously in the laboratory using formula from calibration curves) in Table 2.4 to calculate the molecular weights of the VapBC complexes.

Table 2.4: SEC calibration curve molecular weight equations.

SEC Column	Equation	Vo (ml)	Vc (ml)
S200 10/300	$MW = e^{((Kav - 1.6833)/ -0.115)}$	8.11	24.0
S200 16/60	$MW = e^{((Kav - 2.1739)/ -0.159)}$	46.26	124.0

2.3.2.15 Tryptic Digest of the VapBC Complex

The purified VapBC complex was subjected to a limited tryptic digest to preferentially digest away the labile VapB protein, leaving the VapC behind. Trypsin from bovine pancreas (0.1 mg) (Sigma Aldrich, USA) was added to 0.7 mg VapBC and incubated for 1 hr at room temperature with gentle agitation on a rocking platform. Trypsin digest reactions were stopped by the addition of an equal amount of trypsin inhibitor from Glycine max (soybean) (Sigma Aldrich, USA), and incubating for 15 min at room temperature with gentle agitation prior to purification via ion exchange chromatography.

2.3.2.15.1 Crude Tryptic Digest without Ion Exchange Chromatography

Where ion exchange was not performed on samples after tryptic digestion to produce pure VapC, a crude tryptic digest was performed resulting in a mixture of VapC, digested VapB, trypsin and trypsin inhibitor. The digest was performed as above, but the concentration of resulting VapC was estimated for an impure protein mixture sample where the rough concentration of VapC (mg/ml) = absorbance $_{280\ nM}$ / spectrometer path length (1 cm)

2.3.2.16 Purification of VapC by Anion Exchange Chromatography

Tryptic digests of purified VapBC complex inhibited with trypsin inhibitor (Section 2.3.2.15) were either concentrated until they reached 1 mg/ml then diluted 1:50 into anion exchange buffer, or dialysed against anion exchange buffer overnight prior to separating the trypsin + inhibitor from VapC via anion exchange chromatography.

The predominantly used anion exchange start buffer was 50 mM phosphate buffer pH 7.4, 100 mM NaCl, although variations on the pH and salt concentration were used for VapC proteins which didn't initially bind with 100 mM salt, or who's net charge (as judged by their predicted pI) was too weak for effective binding at pH 7.4.

A 5 ml HiTrap Q XL Anion Exchange column (GE Healthcare, UK) was pre-equilibrated with 40 ml of start buffer. The digest was filtered through a 0.2 µm Minisart filters (Sartorius AG, Germany) then loaded onto the anion exchange column at a flow rate of 1 - 2 ml/min, which was then attached to an ÄKTA Prime™, ÄKTA Basic™ or ÄKTA Purifier™ FPLC system. Unbound protein and trypsin were removed by washing the column with 25 ml of start buffer at a flow rate of 1 ml/min, before bound proteins were removed using a gradient of 0 – 100 % elution buffer (start buffer + 1 M NaCl) over 50 ml at a flow rate of 1 ml/min. Aliquots of fractions (15 µl) were then analysed by SDS-PAGE (Section 2.3.2.1). The anion exchange column was regenerated by washing with 30 ml water followed by 30 ml of 20 % ethanol.

2.3.2.16.1 Anion bead pH screen

Q Sepharose™ Fast Flow beads (GE Healthcare, UK) were used in a buffer pH screen to find the optimal pH for binding to an anion column for proteins with problematic or unknown pI's. 40 µl of beads were transferred to a 1.5 ml eppendorf tube and washed with 1 ml of buffer (Table 2.5) then centrifuged (3,000 rpm for 4 min) to settle the beads. The wash supernatant was discarded and 20 µl of buffer + 2.5 µl protein was added to the beads before incubating at room temperature for 15 min at 1,000 rpm. The beads were spun again (3,000 rpm for 4 min) and supernatant removed. 15 µl of the supernatant was added to 5 µl 4 x SDS loading buffer. The beads were washed with 200 µl of buffer then spun (3,000 rpm for 4 min) and supernatant discarded. 4 x SDS loading buffer (10 µl) was added to the beads then both bound (bead) and unbound (supernatant) samples were analysed by SDS-PAGE.

Table 2.5: Buffers used in anion exchange chromatography pH screen

Substance	Concentration (mM)	pH
Piperazine	20	4.8
Piperazine	20	5.8
L-Histidine	20	5.5
L-Histidine	20	6.5
Bis-Tris	20	6.0
Bis-Tris	20	7.0
Tris	20	7.4
Phosphate	20	8.0
Tris	20	8.5
Bis-Tris propane	20	8.6
Bis-Tris propane	20	9.6
Ethanolamine	20	9.0
Diethanolamine	50	9.4
Piperazine	20	10.0

2.3.2.17 Purification of VapC by Cation Exchange Chromatography

Tryptic digests of purified VapBC complex were purified by cation exchange chromatography as for Section 2.3.2.16, but using cation exchange buffers (Table 2.6) and a 5 ml HiTrap SP XL Cation Exchange column (GE Healthcare, UK). The cation exchange column was stored in 20 % ethanol containing 0.2 M sodium acetate.

2.3.2.17.1 Cation bead pH screen

SP Sepharose™ Fast Flow beads (GE Healthcare, UK) were used in a buffer pH screen to find the optimal pH for binding to a cation column for proteins with problematic or unknown pI's. Bead aliquots (40 µl) were transferred to a 1.5 ml eppendorf tube and washed with 1 ml of buffer then centrifuged (3,000 rpm for 4 min) to settle the beads. The wash supernatant was discarded and 20 µl of buffer + 2.5 µl protein was added to the beads before incubating at room temperature for 15 min at 1,000 rpm. The beads were spun again (3,000 rpm for 4 min) and supernatant removed. 15 µl of the supernatant was added to 5 µl 4 x SDS loading buffer. The beads were washed with 200 µl of buffer then spun (3,000 rpm for 4 min) and supernatant

discarded. 4 x SDS loading buffer (5 µl) was added to the beads then both bound (bead) and unbound (supernatant) samples were analysed by SDS-PAGE.

Table 2.6: Buffers used in cation exchange chromatography pH screen

Substance	Concentration (mM)	pH
Citric acid	50	3.5
Citric acid	50	4.5
Acetic acid	50	4.4
Acetic acid	50	5.4
MES	50	5.6
MES	50	6.6
Phosphate	50	7.0
Phosphate	50	8.0
BICINE	50	8.8

2.3.2.18 Size Exclusion Chromatography of VapC

SEC using an analytical Superdex™ 75 10/300 GL column (GE Healthcare, UK) (3 – 70 kDa separation range) was performed (Section 2.3.2.14) on VapC resulting from a trypsin digest of VapBC in an attempt to purify away the trypsin from VapC.

2.3.2.19 Small Scale (20 µl) Buffer Screen

A small selection of buffers more suited to using in crystallography experiments than phosphate buffer were chosen for a buffer screen with purified VapC. A protein aliquot (20 µl) was dialysed in 20 µl dialysis buttons against 50 mM Tris, HEPES or MOPS at pH 7.4 + 200 mM NaCl at 4 °C overnight. Buttons were examined under the microscope and scored for precipitate.

2.3.3 Crystallography

2.3.3.1 Initial Crystallisation Trials

Initial crystallisation trials for proteins were set up using the high throughput mosquito® Crystal robotic system (TTP Labtech Ltd, UK), using PEGRx HT, Crystal Screen HT, Index HT and SaltRx HT crystallisation screens (Hampton Research, USA). Aliquots of each screen condition (100 µl) were manually pipetted into either a 96 well Intelli-plate (Hampton Research, USA) or triple drop 96-well plate (TTP Labtech Ltd, UK) before 100 nl of protein and reservoir solutions were robotically added into the crystallisation well. All plates were sealed with ClearSeal™ Film (Hampton Research, USA) and stored in a temperature controlled crystallisation room at 18 °C.

2.3.3.2 Optimisation of Crystallisation Conditions (Fine Screening)

Once promising crystallisation conditions had been identified, a variety of second optimisation steps were employed to produce crystals of the highest possible quality suitable for X-ray analysis. All conditions tested included a negative buffer only control to rule out the possibility of crystals forming due to the purification buffer crystallising with the reagents rather than the protein itself.

2.3.3.2.1 Hanging Drops

VDX 24-well plates (Hampton Research, USA) were pre-greased with Glisseal N then 0.5 ml of fine screen 'mother liquor' (ML) was transferred into the well. In general, the fine screen ML trialled contained a variation on the original concentration of the macromolecule, precipitant, or pH. The appropriate volume of protein (usually 1 µl) and ML was pipetted onto either a 22 mm square siliconised or unsiliconised coverslip (Hampton Research, USA) which was then inverted on top of the greased well.

2.3.3.2.2 Microseeding

Streak seeding

A parent crystal of inadequate quality was stroked with a cat's whisker (which had previously been cleaned with ethanol and dried) to pick up invisible seeds. Those seeds were then introduced into a fresh source drop of ML by streaking the whisker straight through it. A variation of this streak seeding was also performed which involved streaking the seeds through three drops serially.

Dilution seeding

A 2 µl hanging drop containing a parent crystal of inadequate quality was resuspended in 55 µl fresh ML and then serially diluted 1/10, 1/100 or 1/1,000 into the new ML condition. 0.5 ml of this seeding mixture was used as the ML for laying down hanging drops as for Section 2.3.3.2.1.

Pulverized seeding

As for dilution seeding above, but resuspending the drop in only 10 µl of fresh ML, and using the much more vigorous resuspension method of vortexing.

2.3.3.2.3 Additive Screens

Crystals were grown in ML containing a set concentration of an additive as for Section 2.3.3.2.1. Additives and concentrations chosen for testing were based on those included in the Hampton Research Additive Screen HT (see Appendix A.8 for conditions tested).

2.3.3.2.4 Tryptic and DNase Digests

Trypsin digest screens

Trypsin (from bovine pancreas) (Sigma Aldrich, U.S.A) was diluted in varying volumes of SEC buffer so that 1 µl could be added to the same volume of protein to give 1:125, 1:250, 1:500, 1:1,000, 1:5,000 and 1:50,000 molar ratios of Trypsin to protein. The various trypsin/protein dilutions were made up one at a time then used to lay down crystals immediately (as for Section 2.3.3.2.1) so as to limit the activity of the trypsin on the protein prior to beginning the crystallisation process.

DNase digest screens

DNase I (from bovine pancreas) (Sigma Aldrich, U.S.A) was diluted in varying volumes of nuclease free water so that 1 µl could be added to the same volume of protein to give 1:500, 1:1,000, 1:2,000 and 1:5,000 ratios of DNase to protein. The various DNase/protein dilutions were made up one at a time then used to lay down crystals immediately (as for Section 2.3.3.2.1) so as to limit the activity of the DNase on the protein prior to beginning the crystallisation process, or after 10 min incubation at room temperature.

Duplicates were set up and 10 mM $MnCl_2$ was added to the ML for one of each of the duplicates to promote DNase activity. Additional controls were used to rule out the possibility of crystals forming due to the DNase or $MnCl_2$ crystallising in the ML.

DNase control: 1 µl of DNase was used in place of the 1 µl of protein

DNase + $MnCl_2$ control: 1 µl of DNase was used in place of the 1 µl of protein, and the ML contained 10 mM $MnCl_2$

2.3.3.2.5 Protein Concentration, Age and Drop Size Trials

A variety of protein concentrations were trialled, along with drop sizes and protein to ML drop ratios. Each protein purification yielded differing amounts of protein, so crystal trials were tested with protein ranging from 4 mg/ml to 19 mg/ml. Drop sizes of (protein + ML) 0.6 µl + 0.6 µl, 1 µl + 1 µl, 2 µl + 2 µl and 1.5 µl + 1 µl, and drop ratios of 1:1, 2:1 and 2:3 were all trialled, as was pre-equilibrating the protein with an equal volume of water before adding to the ML drop.

2.3.3.2.6 Liquid Bridge

As for a hanging drop (Section 2.3.3.2.1), but the protein and ML drops were pipetted onto the coverslip with a gap in between, then a cat's whisker was used to 'join' the two drops so that free diffusion could occur between them and induce crystallisation.

2.3.3.3 Heavy Metal Soaks

Prior to testing crystals at the synchrotron, a mercury soak was performed to incorporate the heavy metal into the crystal at the single cysteine position. ML (250 µl) containing 2 mM mercury(II) acetate was transferred into a pre-greased well. 2 µl of ML/mercury(II) acetate was spotted onto a coverslip, five crystals were transferred into the drop then sealed above the well for 3 hrs before freezing down ready for testing.

2.3.3.4 Testing of Crystals by X-Ray Diffraction

Protein crystals were tested at The Australian Synchrotron on the high-throughput Macromolecular Crystallography Beamline, MX1. Here, single crystals were robotically loaded then detected with an ADSC Quantum 210r Detector (Area Detector Systems Corporation, USA), and analysed using multiple wavelength anomalous dispersion (MAD).

2.3.4 VapC Ribonuclease Activity

An assay testing for ribonuclease activity of VapC protein was originally developed by Jo McKenzie (Proteins & Microbes Laboratory, University of Waikato) for testing $VapC_{MS1284}$ from *M. smegmatis* and $VapC_{PAE0151}$ and $VapC_{PAE2754}$ from *Pyrobaculum aerophilum* (McKenzie *et al.* 2012a).

2.3.4.1 Urea Denaturing Polyacrylamide Gel Electrophoresis (Urea Denaturing-PAGE)

Urea denaturing-PAGE gels were cast in a Hoefer mini gel casting system and consisted of a 10 or 20 % acrylamide resolving gel (depending on the RNAs size). Gels were made with 30 % acrylamide with an acrylamide:bisacrylamide ratio of 37.5:1 (Bio-Rad Laboratories, USA), included 6 M urea, and were polymerised by the addition of 0.05 % ammonium persulphate (APS) and TEMED.

RNA samples were mixed in a 1:1 ratio with 2 x formamide loading dye and denatured for 5 min at 70 °C prior to loading onto the gel. Gels were pre-run in 1 x TBE buffer at 70 V for 30

min and wells flushed to remove any leached urea before samples were loaded. Gels were run at constant 150 V until the xylene cyanol dye front reached the end of the gel.

For RNA band size estimation, 10 µl of prepared Low Range ssRNA ladder (New England Biolabs, USA) was run alongside each gel. Gels were stained for 40 min with SYBR® Green II RNA Stain (Invitrogen, U.S.A) and visualised under UV light.

2.3.4.2 Isolation of Total RNA from *M. smegmatis*

Total RNA was isolated from *M. smegmatis* as for Section 2.2.3.2.

2.3.4.3 The Pentaprobe System

The pentaprobe system is made as six overlapping dsDNA plasmid molecules and covers every combination of five bases (Kwan *et al.* 2003). Each single strand of pentaprobe was cloned under a T7 promoter in pcDNA3 in Joel McKay's laboratory, School of Molecular Bioscience, University of Sydney.

The plasmids containing pentaprobe strands in the forward direction are termed A922, A923, A924, A925, A926 and A927 and in the reverse direction A928, A929, A930, A931, A932 and A933. A922 is complementary to A928, A923 to A929 etc. All inserts were cloned into pcDNA3 using KpnI and XbaI restriction sites except A926 and A932 which, due to cloning difficulties were cloned into pcDNA3 using blunt-end cloning and the EcoRV restriction site. This resulted in a larger gap or "linker" before the start of pentaprobe sequence compared with the inserts cloned using KpnI and XbaI restriction sites. RNA fragments covering every combination of five bases, are transcribed from this set of 12 plasmids (sequences in Appendix B.3).

The set of 12 pentaprobe inserts were produced by Jo McKenzie (Proteins & Microbes Laboratory, University of Waikato) from plasmids obtained from Joel McKay, School of Molecular Bioscience, University of Sydney. Briefly, pentaprobe plasmids were electroporated into *E. coli* DH5α electrocompetent cells, linearised with ApaI then filled with Klenow Enzyme (Roche Applied Science, Switzerland) and purified. Pentaprobe inserts were then amplified

using the *Pfx*® DNA polymerase system (Invitrogen, USA) with the forward primer flanking the T7 promoter in pcDNA3 and reverse flanking the end of the pentaprobe insert, and gel purified.

2.3.4.3.1 In Vitro Transcription of Pentaprobe Inserts

RNA was transcribed from Jo McKenzie's pentaprobe PCR product using the T7 MEGAscript® kit (Ambion, USA). Reactions of 1 µg DNA template, 2 µl each rNTP (75 mM stock), 2 µl T7 polymerase enzyme mix, 2 µl 10 x reaction buffer and nuclease free water in a total volume of 20 µl were incubated at 37 °C for 4 hr (time determined by time course transcription reaction) then DNase treated with 1 µl TURBO™ DNase for 15 min at 37 °C, 600 rpm.

Transcribed RNA was purified by sodium acetate and ethanol precipitation as for Section 2.3.4.5.1. Secondary structure of RNA molecules were predicted using mfold (Zuker 2003).

2.3.4.3.2 Design of 932 Pentaprobe Oligonucleotides

932 Pentaprobe RNA oligonucleotides were designed by Jo McKenzie. Nine 30 – 36 base oligonucleotides were designed to cover bases 54 - 242 of pentaprobe 932 RNA so that their secondary structure mimicked the corresponding region on the whole 932 pentaprobe RNA molecule.

2.3.4.4 VapC Ribonuclease Activity Assay

Ribonuclease activity of VapC is monitored across four time points, as well as over three different negative controls: RNA only to ensure there is no RNase contamination of the assay buffer; with EDTA included in the buffer to rule out the possibility of metal independant RNases also being purified during the purification process; and with the VapBC complex to show lack of VapC ribonuclease activity when in complex with VapB. Total RNA from *M. smegmatis* is used initially as the target substrate to test for ribonuclease activity, and once activity is confirmed, specific RNA's are used for determining the cut sites of the VapC proteins.

Assay Reactions (Time Course)

Four 10 µl reactions containing 1 µg purified VapC protein, 1 µg purified RNA, 6 µl assay buffer (20 mM phosphate buffer pH 7.4, 20 mM NaCl, 10 mM $MgCl_2$*), and nuclease free water were incubated at 37 °C for the appropriate time (time course).

* NB. Oligonucleotide assays analysed by MALDI-TOF MS were carried out in 20 mM ammonium phosphate buffer pH 7.4, 20 mM NaCl, 10 mM $MgCl_2$.

Negative Controls

RNA Only Negative Control (0 hr and 1 hr)

Assay reactions were set up as above but no VapC was added to the reaction. Two reactions were set up, 0 hr (which was stopped immediately) and 1 hr which was incubated at 37 °C for 1 hr (or the assay duration).

EDTA Negative Control (1 hr)

One assay reaction was set up for 1 hr (or the assay duration) as above but using assay buffer containing 20 mM EDTA.

VapBC Negative Control (1 hr)

One assay reaction was set up as above but purified VapBC complex was added to the reaction and incubated at 37 °C for 1 hr (or the assay duration) instead of purified VapC

Assay reactions were stopped by the addition of 10 µl 2 x formamide loading dye, or for analysis by mass spectrometry; assays were heat inactivated at 70 °C for 10 min.

2.3.4.5 MALDI-TOF Mass Spectrometry of RNA Oligonucleotides

2.3.4.5.1 Sample Purification

932 RNA oligonucleotides were ordered HPLC purified from either Sigma Aldrich (USA) or IDT (USA).

RNA from ribonuclease assays (Section 2.3.4.4) was precipitated to remove contaminants and de-salted via ammonium acetate and ethanol precipitation ready for MALDI-TOF MS analysis. Each 10 µl reaction was precipitated either on ice for 1 hr or at -20 °C overnight, with 8 µl 5 M ammonium actetate and three times the volume (54 µl) of 100 % ethanol. Samples were spun (13,000 rpm, 20 min, 4 °C) and all the supernatant carefully removed without disturbing the pellet. The pellet was washed with 50 µl 70 % ethanol and spun as above. After carefully removing all of the supernatant, the pellet was air dried briefly then resuspended in 5 µl nuclease free water and stored at -80 °C.

2.3.4.5.2 Sample Preparation

Matrix for MALDI-TOF MS of RNA oligonucleotides was made by vortexing 5 mg of 3-hydroxypicolinic acid (3-HPA), 10 µl 2.5 M diammonium citrate, 125 µl of acetonitrile (ACN) and 365 µl of nuclease free water until dissolved, then centrifuging for 5 min at 13,000 rpm. Matrix (1 µl) was spotted onto an Anchorchip™ target plate (Bruker Daltonics, USA) and left to air dry. An aliquot of sample or oligonucleotide calibration standard (1 µl) (Bruker Daltonics, USA, or made in the lab) was spotted on top of the dried matrix and left to air dry.

2.3.4.5.3 Construction of Oligonucleotide Calibration Standard

For oligonucleotide calibration standard mixture made in the lab, the three oligonucleotides specified in the Oligonucleotide Calibration Standard (#206200) (Bruker Daltonics, USA) were purchased HPLC purified from IDT (USA) (oligo sequences can be found in Appendix A.1). The oligos were made up in UltraPure water then mixed together to give a calibration mixture containing the prescribed amount of each oligo as specified by Bruker, or at a customised ratio which was empirically determined.

Bruker mix: oligo1 = 0.25 pmol/µl; oligo2 = 1.25 pmol/µl; oligo3 = 5 pmol/µl
Custom mix: oligo1 = 0.83 pmol/µl; oligo2 = 1.66 pmol/µl; oligo3 = 5 pmol/µl

Aliquots of oligonucleotide calibration standard mix (5 µl) were stored at -80 °C until needed.

2.3.4.5.4 MALDI-TOF target cleaning

The MALDI-TOF target plate was cleaned when full to remove all traces of previously analysed samples. Deionised water and lint-free Kimwipes were used in all steps. Old matrix was removed from the plate by wiping with water before wiping with 50 % methanol then ACN. The plate was sonicated upside down in a container with 50 % methanol for 10 min in a sonicating water-bath (Elma, Germany) then rinsed with water. Acetone was poured over the plate to remove the water then left to air dry.

2.3.4.5.5 MALDI-TOF Set Up

MALDI-TOF analysis was performed with the Autoflex™ II MALDI-TOF mass spectrometer (Bruker Daltonics, USA). Analysis were obtained in the positive ion linear mode with a 150 ns pulse, range of 2 - 12 kDa, accelerated voltage in the ion source set at 20 kV and laser power at around 50%. Spectra for the oligonucleotide calibration standard (Bruker Daltonics, USA) were collected first, and the spectrometer calibrated to this with an automatic polynomial correction.

Spectra were exported to DataAnalysis™ or FlexAnalysis™ software (Bruker Daltonics, USA), and peaks identified and labelled.

2.3.4.6 Determination of VapC Cut Sites

The mass list and corresponding intensities were exported from the MALDI-TOF analysis software into a comma separated excel file then processed using in-house software developed by Tony Smith and Vic Arcus (The University of Waikato) to determine possible cut sites. The software identifies all one and two point cuts in the oligonucleotide and calculates the putative mass for each resulting fragment, factoring in a 5' phosphate on the 3' cleavage product. Comparisons are made to the actual peak masses from the MALDI-TOF data, and those fragments resulting from single cuts which match the actual peak masses most closely are considered as a putative VapC target cut sites. The intensity cut off was set to greater than 800, so all peaks above an intensity of 800 were picked for analysis.

Chapter Three:

Rangipo Diagnostic

3.1 Introduction

3.1.1 *M. tuberculosis* in a New Zealand Context

Unless otherwise indicated, the following statistics were obtained from (Lim & Heffernan 2013).

3.1.1.1 Notifications of Disease

Tuberculosis notification rates in New Zealand have remained relatively stable over the last five years, with the 2012 figure totalling 6.6 per 100,000 population. Although this figure is much lower than the global average of 122 incident cases per 100,000, rates in New Zealand are still higher than those in other developed countries such as Australia, Canada and America (5.7, 4.0 and 3.2 per 100,000 respectively) (WHO 2013). The overall rates were higher in males than in females and in Asian and Middle Eastern/Latin American/African (MELAA) ethnic groups, with the dominant risk factors being born outside New Zealand or residence with a person born outside of New Zealand, and living in socio-economically deprived areas.

The highest notification rates in 2012 were from the Asian (41.4 per 100,000 population) and Middle Eastern/Latin American/African (MELAA) (31.8 per 100,000) ethnic groups, followed by the Pacific 12.4 per 100,000), Maori (5.4 per 100,000) and European or Other (0.9 per 100,000) ethnic groups. While the percentage of cases born in the Pacific Islands and Africa and the Middle East regions has decreased significantly (26 - 30 %) over the past five years, there has been an increasing trend in notification rates for the Asian ethnic group, where cases born in the Southern and Central Asia region (predominantly from India) had the highest notification rate, followed by the South-East Asia region (most commonly from the Philippines). Notably, the majority of New Zealand born cases were from the Maori ethnic group (50 %),

while just 29 % accounted for the European or Other ethnic groups. Consistant with previous observations that TB incidence is associated with household crowding (Baker *et al.* 2008), a disproportionate number of TB cases reside in the most socioeconomically deprived areas of New Zealand (over 60 %), an observation which was even more apparent in New Zealand born cases.

The clinical characteristics of TB from foreign born and New Zealand born cases was strikingly different in the period between 2008 and 2012. While foreign born cases tended to present with pulmonary or extra-pulmonary TB in approximately equal numbers, 80 % of cases born in New Zealand were pulmonary. The common sites of infection for the extra-pulmonary cases of TB were the lymph nodes (excluding abdominal), followed by pleural and intra-abdominal (excluding renal), and there were approximately equal numbers of tuberculous meningitis and miliary TB reported during this timeframe.

3.1.1.2 Treatment

The vast majority of new TB cases were reported to have received treatment (98.7 %) in 2012, with the median time to the start of treatment from the onset of symptoms being two months for the 52 % of cases for which the interval could be calculated for. This interval between onset of symptoms and treatment was approximately the same for pulmonary cases, where the delay poses a risk to public health due to transmission of the disease from person to person.

The majority of cases in 2012 (85.4 %) were fully susceptible to all five antimicrobials routinely tested for (isoniazid, rifampicin, ethambutol, pyrazinamide and streptomycin), and just four cases were multi drug resistant (MDR) (resistant to at least isoniazid and rifampicin). In the last five years, the average number of MDR culture-positive TB cases was 1.2 %, the majority of which (~94 %) were from foreign born individuals (predominantly from the Asian ethnic group) assumed to have acquired their MDR-TB overseas. Only one case of extensively drug resistant TB (XDR-TB) which is MDR-TB plus resistance to any fluoroquinolone and at least one of the following second-line drugs: amikacin, capreomycin or kanamycin, has been identified in New Zealand thus far, occurring in 2010. TB cases which had previously been

treated were less likely to be fully susceptible to the antimicrobials tested, significantly more resistant to rifampicin and streptomycin and more likely to be MDR-TB compared with new cases of TB.

Of the 279 new cases of TB in 2012, the fatality rate remained low with just four cases (all of which were over 60 years of age), as did the HIV co-infection rate (three cases of co-infection). The number of TB reactivation cases also remained low in 2012 with 15 reactivation notifications.

3.1.1.3 Confirmation and Classification of TB Cases

Laboratory confirmation of new TB cases is by: positive culture for MTBC, positive microscopic examination for acid-fast bacilli when a culture has not been or cannot be obtained, demonstration of MTBC nucleic acid directly from specimens, or histology strongly suggestive of tuberculosis when there is a strong clinical probability (Jacobs 2012). Figures for 2012 indicate that approximately 85 % of new TB cases were laboratory-confirmed, the majority of which were by culture of MTBC (97.7 % were *M. tuberculosis*, 1.8 % *M. bovis*, and 0.45 % MTBC) from a clinical specimen. Approximately 70 % of all new TB cases in New Zealand are diagnosed by presentation to a health practitioner, while immigrant or refugee screening and contact tracing detect 11.2 % and 7.2 % of cases respectively.

Presently, all molecular typing of human TB cases in New Zealand is carried out by LabPLUS consisting of primary (12 loci) and secondary (a further 12 loci) Mycobacterial Interspersed Repetitive Units (MIRU) analysis alone. Prior to this, primary typing was by MIRU 12 and secondary typing was by IS*6100* Restriction Fragment Length Polymorphism (RFLP) at LabPLUS, and before this, mycobacteriology reference laboratories throughout the country performed RFLP typing alone to type TB isolates. Of the typing results available for new TB cases in 2012, approximately two thirds had a unique strain type while the remainder had non-unique molecular types, assigned to 44 separate molecular clusters (five of which were new). Over the past five years, the median cluster size was two, with most clusters (~ 90 %) containing less than five cases. Interestingly, non-unique molecular types were much more

prevalent than unique strain types among Maori (29.3 % vs. 5.1 % for unique) and Pacific (28.5 % vs. 7.5 %) ethnic groups, and for cases born in New Zealand (40.8 % vs. 11.6 %). The opposite trend was observed for the Asian/MELAA ethnic group where the proportion of cases was lower amonst those with non-unique molecular types.

In 2012, there were three outbreaks and 22 associated cases of *M. tuberculosis*. The outbreaks comprised of two, six and 14 cases, and involved exposure to an index case in a private home, during church gatherings and in private homes, and in a work environment and in private homes respectively. Large outbreaks are responsible for spikes in the rates of TB in New Zealand, such as in 1999 where two outbreaks (one of which was the Rangipo strain) contributed to the highest incidence recorded over a 25 year period of 12.6 cases per 100,000 (De Zoysa 2001).

3.1.2 Rangipo

The Rangipo strain of *M. tuberculosis* is the main endemic strain in New Zealand as well as the largest single *M. tuberculosis* strain cluster (Sexton *et al.* 2008; Bower 2013). It predominantly circulates in the Waikato and Bay Of Plenty regions, and has been responsible for a number of outbreaks over previous decades. The Rangipo strain is highly transmissible, and previous analysis of Rangipo outbreaks found the majority of cases were of Maori or Pacific Islands ethnicity under 30 years of age, and tended to have gang affiliations or contact with persons who had been incarcerated in prison (De Zoysa 2001; Health 2012; Karalus 2013). The Rangipo strain was thought to originate from an index case incarcerated at Rangipo prison, and was thus called 'Rangipo', but isolates from cases preceding this have since been identified.

The number of cases within a Rangipo cluster is generally much larger than clusters from other *M. tuberculosis* strains (Sexton *et al.* 2008; Lim & Heffernan 2013) demonstrating its high transmissibility. As well as this, index cases in Rangipo outbreaks are highly infectious (McElnay *et al.* 2004) and while no virulence testing has been undertaken in the Rangipo strain, it is regarded as being a virulent *M. tuberculosis* strain (Karalus 2013). It has been

suggested the apparent high transmissibility of the endemic Rangipo strain may be instead due to lifestyle and environmental factors within this population, causing them to be more susceptible to transmission, such as high density housing and social or work gatherings, alcohol and drug abuse and socioeconomic deprivation. This proposal is unlikely however, given the general theory that one person infected with tuberculosis must infect 20 people in order to cause one active case of disease (Sutherland 1976), as the number of Rangipo infected individuals in outbreaks developing active disease is much higher (Karalus 2013). In a recent investigation of a Rangipo outbreak in Hawkes Bay, there were 94 infected cases identified and 14 cases of active disease, approximately three times more active cases than would be expected from a 'normal' tuberculosis infection (McElnay *et al.* 2004; Calder 2013). Other *M. tuberculosis* strains such as those belonging to the Beijing lineage, have also been noted as having an increased ability to transmit within particular populations, and generate active disease (Reed *et al.* 2009).

3.1.3 *M. tuberculosis* Typing Methods

Molecular typing of *M. tuberculosis* strains is essential for identifying and linking source cases to contacts, as well as identifying specific strains (such as drug resistant isolates) or tracking strains as they spread throughout the population. This insight allows researchers to identify and stop the spread of clustered cases, observe how particular strains are evolving, and in the case of drug resistant isolates, prevent the emergence of more MDR or XDR strains from developing by ensuring the optimal treatment is provided and adhered to.

Because mycobacteria exhibit very low DNA sequence diversity, the Multilocus Sequence Typing (MLST) method used routinely to subdivide microbes on the basis of neutral sequence diversity is not possible as the level of resolution is too low for the subdivision of these and other genetically monomorphic bacteria. The investigation of mycobacterial genetic diversity up until now has therefore relied on non-sequencing based methods such as Restriction Fragment Length Polymorphism (RFLP) typing, and PCR based methods such as Clustered Regulatory Short Palindromic Repeats (CRISPR) and Variable Number Tandem Repeats (VNTR) genotyping (Figure 3.1).

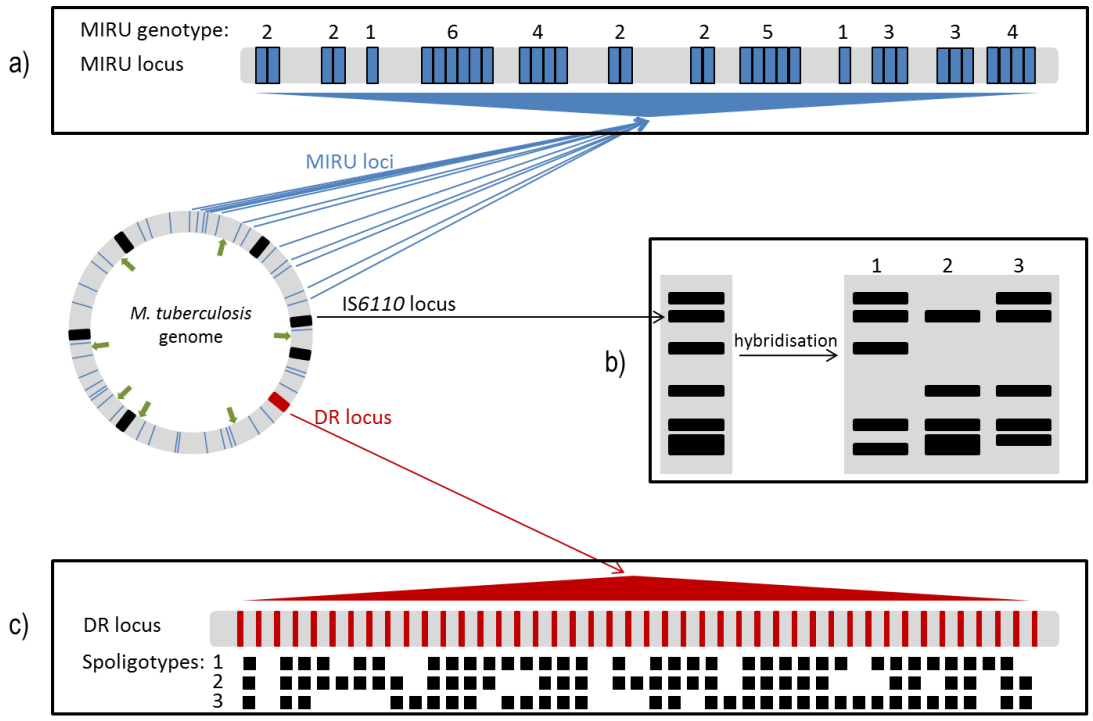


Figure 3.1: Schematic comparison of the three common *M. tuberculosis* typing methods. a) VNTR typing (or MIRU in mycobacteria) uses PCR amplification to ascertain the variable copy number of specific tandemly repeated sequences within the genome. There are 41 MIRU loci (shown in blue on the *M. tuberculosis* genome), of which 12, 15 or 24 are selected for genotyping. a) Shows the 12 tandemly repeated sequences used in MIRU 12 analysis and the number of tandem repeat copies present within each one for the representative strain. The MIRU 12 genotype is represented as 12 characters (221642251334), where each character corresponds to the number of repeats within each of the 12 MIRU loci. b) IS*6110* RFLP typing. *M. tuberculosis* genomes vary in the number and position of IS*6110* copies (black boxes) which they possess. Restriction enzymes cleave DNA at specific sites (green arrowheads on the *M. tuberculosis* genome) then the fragments are separated by size by gel electrophoresis and those containing IS*6110* insertion sequence hybridise to a specific radioactive probe thereby producing a characteristic banding pattern or fingerprint for each isolate. c) Spoligotyping patterns differentiate three *M. tuberculosis* strains by comparing the presence (black boxes) or absence (white boxes) of the 43 spacers located between direct repeats in the DR locus.

IS*6110* RFLP typing is a southern blot based assay which detects the mobile IS*6110* insertion sequence found exclusively within the MTBC (Thierry *et al.* 1990), where strains are described according to the number and position of IS*6110* copies which they possess. IS*6110* fingerprinting was the 'gold standard' method for discriminating *M. tuberculosis* isolates for over a decade, but is technically demanding, time-consuming (as it requires several weeks for culturing the *M. tuberculosis* isolates), difficult to interpret and reproduce between laboratories, and does not always discriminate low-IS*6110*-copy (</= five copies) (Smittipat *et al.* 2005) isolates.

CRISPR based genotyping in MTBC is referred to as spoligotyping (spacer oligonucleotide typing). Spoligotyping began in 1997 and is a rapid PCR-based genotyping method targeting the well-conserved 36 bp direct repeat (DR) locus of MTBC bacteria. Amplification of 'spacers' (43 short unique regions) contained between the DRs is followed by hybridisation onto a membrane with oligonucleotides complementary to each spacer, and variation in the presence of spacer sequences among strains results in a characteristic spot pattern which is then translated into a numerical code of 15 digits (Kamerbeek *et al.* 1997). Spoligotyping is fast (as genotyping is performed on DNA directly isolated from clinical samples so bacteria do not need to be cultured first), cheap, and the binary data is represented digitally enabling comparisons between laboratories. However, spoligotyping does not differentiate well within large strain families (eg. Beijing) (Driscoll 2009), nor distinguish between unrelated strains possessing identical patterns as a result of independent evolutionary events which caused deletion of the same spacer sequences (Flores *et al.* 2007). While spoligotyping has the lowest discrimination power to differentiate strains, it has proven to be a very useful tool to add to the discriminatory power of IS*6110* RFLP or MIRU-VNTR typing when used in combination, especially for strains harbouring low numbers of IS*6110* copies (Dale *et al.* 2003; Supply *et al.* 2006).

VNTR based genotyping in MTBC is referred to as Mycobacterial Interspersed Repetitive Units (MIRU)-VNTR typing. MIRU-VNTR is a PCR based assay used to detect and compare the strain-specific numbers of regions within *M. tuberculosis* DNA (called MIRU loci) which are prone to possessing short tandemly repeated sequences. There are a total of 41 MIRU loci, of which 24 have been selected for use in genotyping. MIRU results are represented as either 12, 15 or 24 digit numbers (depending on the depth of coverage required), where each digit corresponds to the number of repeats present at each one of the 12, 15 or 24 MIRU loci, listed in a standard order. Unlike IS*6110* RFLP or spoligotyping, MIRU-VNTR can be used by itself to differentiate between *M. tuberculosis* isolates that have low copy numbers of or are without IS*6110* elements (Supply *et al.* 2000). Initially, assays using just 12 MIRU loci were used, but this was less discriminatory than IS*6110* RFLP even in combination with spoligotyping, so more recently a set of 24 MIRU-VNTR loci with a significantly higher discriminatory power than the original 12 locus system, and close to that of IS*6110* fingerprinting, was developed (Supply *et al.* 2006). Interestingly, studies in some populations have suggested a lower discriminatory

power of the 24 loci MIRU-VNTR typing method than IS*6110* RFLP for strains belonging to particular phylogenetic lineages, especially the Beijing strain (Roetzer *et al.* 2011). A pared down discriminatory subset of 15 loci has since been defined, and was proposed as the new gold standard (in combination with spoligotyping) for fast, simple, routine epidemiological discrimination of *M. tuberculosis* isolates (Supply *et al.* 2006; Oelemann *et al.* 2007), and is the typing method used currently in the majority of laboratories around the world.

Recently, a fully automated cartridge-based diagnostic was developed for the detection of MTBC DNA, and mutations within the rifampin resistance-determining region of the *rpoB* gene (as a proxy for MDR-TB). This Xpert MTB/RIF test is based on the Cepheid GeneXpert system, and uses six molecular beacons multiplexed within the same qPCR reaction, producing results in less than two hours directly from sputum (Helb *et al.* 2010). The ability of the Xpert MTB/RIF test to quickly detect smear-negative cases is especially useful for the detection of rifampicin resistant TB in people displaying low numbers of bacilli, and has similar accuracy as culture for detecting TB (Boehme *et al.* 2010). This rapid, automated diagnostic minimises handling and safety risks, and requires little technical training to operate, increasing the likelihood of timely treatment for rifampicin resistant cases (although standard drug-susceptibility testing is still required to determine the complete drug-resistance profile).

Rangipo isolates from 2002 onwards used in this research were typed as identical by IS*6100* RFLP typing at Waikato Hospital (Table 3.1), and isolates A, C, F, N, O, and S were also also sent to LabPLUS for MIRU 12 and 24 typing to determine if any differences could be detected between them. All six isolates were typed as identical: 233325153324 by MIRU 12 typing, and 341444223362 by MIRU 24* (Table 3.2).

Table 3.1: Rangipo isolates used in this study.

Year	RFLP type	Specimen type	Isolate Code
1991	N/A	Sputum	A
1992	N/A	Sputum	B
1992	N/A	Sputum	C*
1992	N/A	Gastric Aspirate	D
1992	N/A	CSF	E
1996	N/A	CSF	F
2002	13/008RP	Sputum	J
2004	13/008RP	Pleural Biopsy	K
2006	13/008RP	Sputum	L
2006	13/008RP	Sputum	M
2007	13/008RP	Sputum	N
2008	13/008RP	Sputum	O
2008	13/008RP	Sputum	P
2009	13/008RP	Sputum	R*
2010	13/008RP	Sputum	S

* Isolates are from the same patient, separated by 17 years. N/A = Not Available

Table 3.2: MIRU 12 and MIRU 24 Loci for Rangipo isolates A, C, F, N, O, and S.

MIRU 12 Loci												
Loci name	2	4	10	16	20	23	24	26	27	31	39	40
Rangipo	2	3	3	3	2	5	1	5	3	3	2	4
MIRU 24 Loci												
Loci name	Mtub04	ETRC	Mtub21	QUB11b	ETRA	Mtub29	Mtub30	ETRB	Mtub34	Mtub39	QUB26	QUB4156
Rangipo	3	4	1	4	4	4	2	2	3	3	6	2

3.1.3.1 Molecular Typing and Phylogenetics

While spoligotyping and MIRU-VNTR are used routinely for local molecular epidemiological applications, tracing tuberculosis transmission and distinguishing new cases from those which are relapsed, some studies have proposed its use as a tool for global phylogenetic and population genetic analyses of the MTBC (Brudey *et al.* 2006; Wirth *et al.* 2008). The use of

spoligotyping and MIRU-VNTR in evolutionary studies is questionable however, as a later study comparing a robust MTBC phylogeny obtained from multilocus sequence analysis (MLSA) of 108 MTBC strains with that inferred from spoligotyping or MIRU-VNTR found while the phylogenetic accuracy of MIRU-VNTR was better than spoligotyping, both methods were unable to detect all strain lineages (Comas *et al.* 2009). It was determined that the limited number of loci and markers used in spoligotyping and MIRU-VNTR evolved rapidly, tending to converge, and it was this level of homoplasy in the markers which deemed these typing methods inappropriate for defining deep phylogenetic relationships (Filliol *et al.* 2006; Comas *et al.* 2009). Comas *et al*. (2009) also found that the MLSA derived MTBC phylogeny corresponded with one which had been determined previously based on LSPs (Gagneux *et al.* 2006), just as Ford *et al.* (2012) found the reconstructed phylogeny from 55 whole genomes resolved the same main lineages as those derived from the 45 SNP position phylogeny (Filliol *et al.* 2006). Comas *et al*. (2009) thereby proposed strain assignment to the main MTBC lineages should be based on phylogenetically robust markers like SNPs or LSPs, while MIRU-VNTR loci be used for epidemiological discrimination purposes, as currently there is not a sole method which fulfils the requirements for both local (molecular epidemiology such as strain typing/transmission) and global (phylogeography/large scale evolutionary analyses) strain classification which is affordable for routine application.

3.1.3.2 Molecular Typing and Whole Genome Sequencing (WGS)

WGS is a relatively new tool used for bacterial genotyping but is currently cost prohibitive to use routinely for typing bacterial strains. WGS has been used to investigate the spread of *M. tuberculosis* and how it is evolving, both long and short term, as well as it's geographic distribution and evolutionary response to antimicrobial therapies (Filliol *et al.* 2006; Ioerger *et al.* 2010; Ford *et al.* 2011; Gardy *et al.* 2011; Saunders *et al.* 2011).

WGS has advantages over other typing methods at a global, local, individual and bacterial level. WGS allows the detection of *M. tuberculosis* global genetic diversity with unparalleled resolution, for example, two Beijing isolates with the same genotype according to both IS*6110* RFLP and MIRU-VNTR patterns were found to differ at 131 individual sites by WGS (one of

which was a large deletion), highlighting the limited ability of the current typing methods to distinguish between closely related strains (Niemann *et al.* 2009; Ford *et al.* 2012).
At a local level, WGS has improved investigations into: the acquisition and spread of drug resistant strains, identifying that MDR and XDR strains from the same region are not just emerging once and undergoing clonal expansion and transmission, but evolving independently from each other (Ioerger *et al.* 2010; Ford *et al.* 2012); as well as outbreak and contact tracing. A TB outbreak in Canada was defined by MIRU-VNTR analysis as a single clonal outbreak, with identical MIRU-VNTR patterns for all isolates and no identifiable source case (even after contact tracing). WGS was used and revealed there were in fact two distinct transmission chains. This WGS based phylogenetic analysis in combination with a social network analysis allowed the researchers to identify the route through which the infection had spread, where the high resolution 24 loci MIRU-VNTR method could not (Gardy *et al.* 2011).
WGS has been useful for identifying diversity at the individual level too. While Mtb infection is generally considered to be clonal, the emergence of variants leading to a greater than expected level of heterogeneity within the infecting bacterial population challenges this dogma (García de Viedma *et al.* 2004; Martin *et al.* 2007). Heterogeneity within the bacterial population could arise from simultaneous infection by multiple strains, reinfection by a new strain, or within host evolution of an individual strain such as the development of drug resistance (Ford *et al.* 2012). The fine resolution that WGS offers has allowed investigation of genetic heterogeneity in greater detail, revealing that exposure to antibiotics creates a strong bottleneck during the development of drug resistance (Saunders *et al.* 2011), and the mutation rate during latent and active disease is approximately equal quelling the assumption that Mtb does not aquire mutations during latency (Ford *et al.* 2011; Colangeli *et al.* 2014). WGS has also provided valuable information regarding diversity at the bacterial level, allowing the identification of nonsynonymous to synonymous genetic change ratios (dN/dS) which indicate that the Mtb genome is under purifying selection (Hershberg *et al.* 2008; Comas *et al.* 2010).

The ability of WGS to fulfil the requirement for both molecular epidemiology fine typing and large scale evolutionary analyses means as the cost reduces, it is likely to become the next gold standard in routine molecular typing, and will be of particular benefit for *M. tuberculosis* and other monomorphic pathogens where limited sequence diversity has prevented strain

differentiation via other readily available sequence based typing methods. However limitations such as the more challenging detection of repetitive regions, genomic duplications and genomic rearrangements do exist with WGS (Ford *et al.* 2012), so these challenges must be kept in mind when using WGS to mine the wealth of information within bacterial strain sequences. Once WGS becomes readily affordible, researchers will be able to trace the transmission and evolution of strains of interest in real time (in particular those which are more virulent or resistant to antimicrobials) which will result in better control, treatment and understanding of this pathogenic organism.

3.1.3.3 Molecular Typing and Large Sequence Polymorphism (LSP) or Single Nucleotide Polymorphism (SNP) analysis

Detection of LSPs or regions of difference (RDs) representing well-characterised unique-event deletions are effective, simple and robust methods for assigning strains to the main Mtb lineages (Gagneux *et al.* 2006; Alland *et al.* 2007; Gagneux & Small 2007; Reed *et al.* 2009). Each of the six major Mtb lineages is defined by just one LSP present in all isolates within that particular lineage (Reed *et al.* 2009). Because Mtb exhibits very low sequence diversity, there are few SNPs within the genome and therefore only a very remote chance of the same SNP existing in independent lineages. This means SNPs, like LSPs, can also be used for population analysis, however using LSPs to genotype isolates is simplier and more efficient (Reed *et al.* 2009).

3.1.4 *M. tuberculosis* Lineage Classification

It is proposed that *M. tuberculosis* originated in East Africa and spread throughout the world, diversifying as it went, resulting in *M. tuberculosis* lineages possibly adapted to different human host populations (Gagneux *et al.* 2006).

Early phylogenetic studies classified Mtb strains using several different nomenclatures, but the strain classification system proposed by Gagneux *et al.* (2006) which defines the global population structure of *M. tuberculosis* into six main lineages, and 15 sublineages is now used. It was initially thought each of the six main lineages was restricted to patients from a particular

geographical area, so as such, the main lineages were named: (1) Indo-Oceanic, (2) East Asian (or W/Beijing), (3) East African/Indian, (4) Euro-American, (5) West African 1 and (6) West African 2. Lineages corresponding to earlier reported strain groupings include: Indo-Oceanic, containing strains previously referred to as 'ancestral' because they contain the TbD1 region which is deleted in 'modern' Mtb strains (Brosch *et al.* 2002); West-African 1 and 2, previously referred to as *Mycobacterium africanum* (Mostowy *et al.* 2004); and Euro-American strains which were previously referred to as principal genetic groups 2 and 3 (Sreevatsan *et al.* 1997).

Lineage specific LSP PCRs are used to classify isolates into the six main *M. tuberculosis* lineages (Figure 3.2). Here, PCR is used to detect either the presence or absence of the lineage specific LSP markers, where RD239, RD105, and RD750 deletions are used to specifically classify isolates belonging to the Indo-Oceanic (1), East Asian (2) and East African/Indian (3) strain lineages respectively (Gagneux *et al.* 2006), and deletion of RD711 and RD702 characterises the West African 1 (5) and West African 2 (6) lineages respectively. Deletion of RD9 confirms the presence of all atypical MTBC infections, as RD9 is absent from all members of the MTBC other than *M. tuberculosis* and *M. canetti* (Brosch *et al.* 2002; Reed *et al.* 2009). The Euro-American (4) lineage could not be defined by a LSP so is instead defined by a seven base pair deletion in the *pks1-15* gene (Gagneux *et al.* 2006).

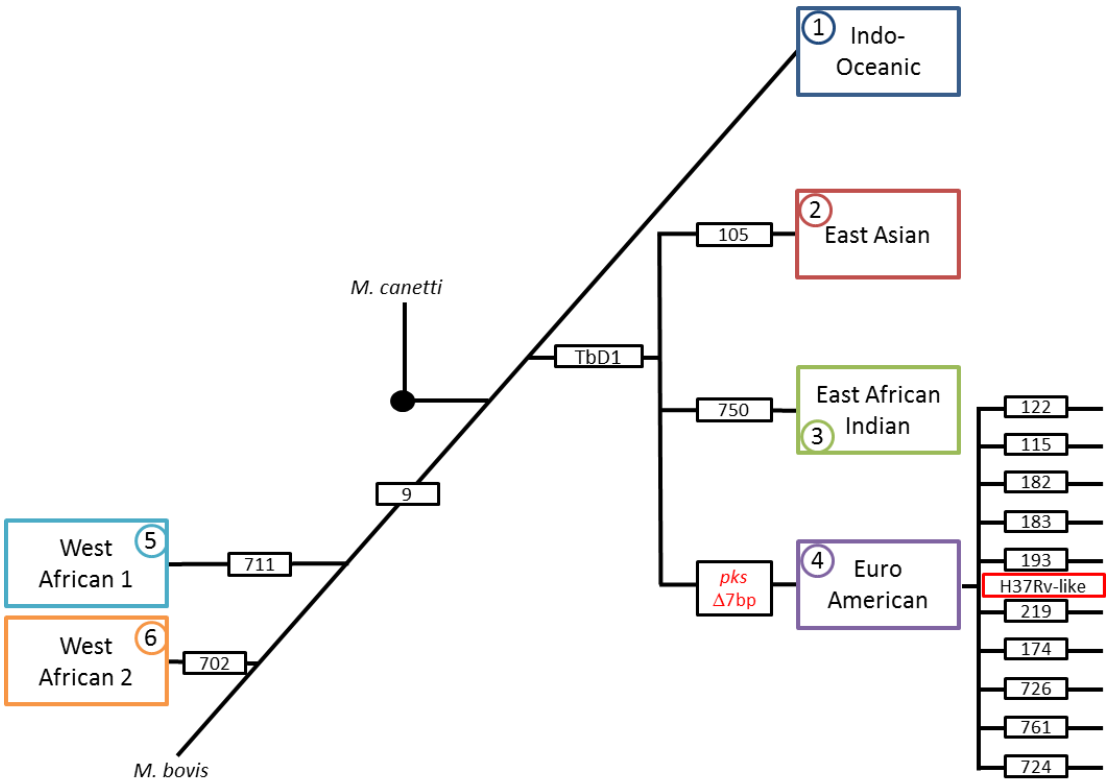


Figure 3.2: Schematic representation of the global phylogeny of *M. tuberculosis* as defined by LSPs. Coloured boxes indicate the six main lineages, while black boxes represent the lineage-defining LSPs or regions of difference which characterise each of the groups

Different *M. tuberculosis* lineages and sublineages have shown varying abilities to cause infection and disease (de Jong *et al.* 2008; Reed *et al.* 2009; Kato-Maeda *et al.* 2010; Kato-Maeda *et al.* 2012). For example, a study in The Gambia found while transmission rates were similar between strains from different lineages, pathogenicity (or progression to disease) was variable with rates ranging from 1 % to 5.6 % for strains from lineage 6 and 2 respectively, and from 1.2 % to 3.9 % for strains from three lineage 4 sublineages (de Jong *et al.* 2008). No significant differences between lineages and their correlation with particular clinical phenotypes has been observed (Reed *et al.* 2009).

3.1.4.1 Euro-American Sublineage Classification

The Euro-American lineage is the most recent 'modern' (containing strains with a deleted TbD1 region) Mtb lineage, as it includes strains belonging to principal genetic groups 2 and 3 (Sreevatsan *et al.* 1997). It is also the predominant lineage of *M. tuberculosis* in many parts of

the world, including North and South America, as well as several European countries and New Zealand (Gagneux & Small 2007; Bower 2013). Euro-American strains can be further classified into into: 11 distinct sublineages according to deletions of a region of difference (RD); RD115, RD122, RD174, RD182, RD183, RD193, RD219, RD724, RD726 and RD761, and the remaining strains with no deletion (called H37Rv-like) (Gagneux *et al.* 2006); or five lineages (T, Haarlem, Latin American-Mediterranean (LAM), S and X) and 34 sublineages by spoligotyping (Rindi *et al.* 2012); or 35 sublineages by analysis with multiple-biomarker tensors (Ozcaglar *et al.* 2011). While certain deletions are associated with a particular spoligotype lineage (deletion RD115 to T and LAM, RD174 to LAM, RD182 to Haarlem and RD219 to T for example), RD-defined sublineages and spoligotype groupings often do not correspond. This disagreement between LSP and spoligotype analysis is likely due to the high degree of homoplasy in the direct repeat locus resulting in convergent evolution and identical spoligotypes, which renders spoligotyping unsuitable for determining the phylogenetic relationships within the Euro-American lineage (Rindi *et al.* 2012).

While diagnostic tools have been recently developed to distinguish the apparently more drug resistant and high virulence 'Beijing' strains plaguing other parts of the world (Lopez *et al.* 2003; Manca *et al.* 2004; Leung *et al.* 2011; Zhang *et al.* 2014), no such work has been performed for the virulent Rangipo strain endemic in New Zealand. The high transmissibility and virulence of the Rangipo strain of *M. tuberculosis* together with the extended period it has been circulating for in the Waikato and Bay Of Plenty regions, is why this particular strain needs to be classified and a rapid diagnostic test for it developed.

3.1.5 Objectives

Given the high transmissibility and virulence of the *M. tuberculosis* Rangipo strain within New Zealand, we attempted to classify the Rangipo strain and develop a fast and efficient Rangipo specific diagnostic assay which could be implemented into the current hospital diagnostic setting.

3.2 Results and Discussion

3.2.1 PCR-Based Genotyping of *M. tuberculosis* targeting GC-rich repeats and IS*6110* Inverted Repeats

Primers targeting frequently repeated 16-bp GC rich sequences (Mtb1 or Mtb2) were used in combination with either specific primers (IS1 and IS2), or a single degenerate primer (CAR2) sited at the inverted repeats (IR) flanking IS*6110,* to investigate the degree of differentiation across 15 suspected clinical *M. tuberculosis* Rangipo isolates isolated over a 20 year period. IS*6110* typing only became routine at Waikato hospital in 2002, so only nine out of these 15 isolates (J – S) were typed and identified as belonging to the Rangipo strain (13/008RP) (see Table 3.1). Figure 3.3 is a theoretical illustration of the genotyping method used, and Figure 3.4 shows the banding pattern obtained for each isolate using either specific or degenerate primers to amplify regions between IS*6110* copies and the frequently repeated 16-bp GC rich sequences. The banding patterns across all 15 Rangipo isolates are similar (supporting the previous typing results identifying most of these as belonging to the Rangipo strain), but show some subtle differences indicating there is sequence diversity among the isolates. Isolates C and R, which were from the same patient isolated 17 years apart, show clear differences in banding for the Mtb2 + IS1 & IS2 genotyping. This sequence diversity suggests a robust genotyping method needed to be developed to ensure the correct identification of Rangipo isolates.

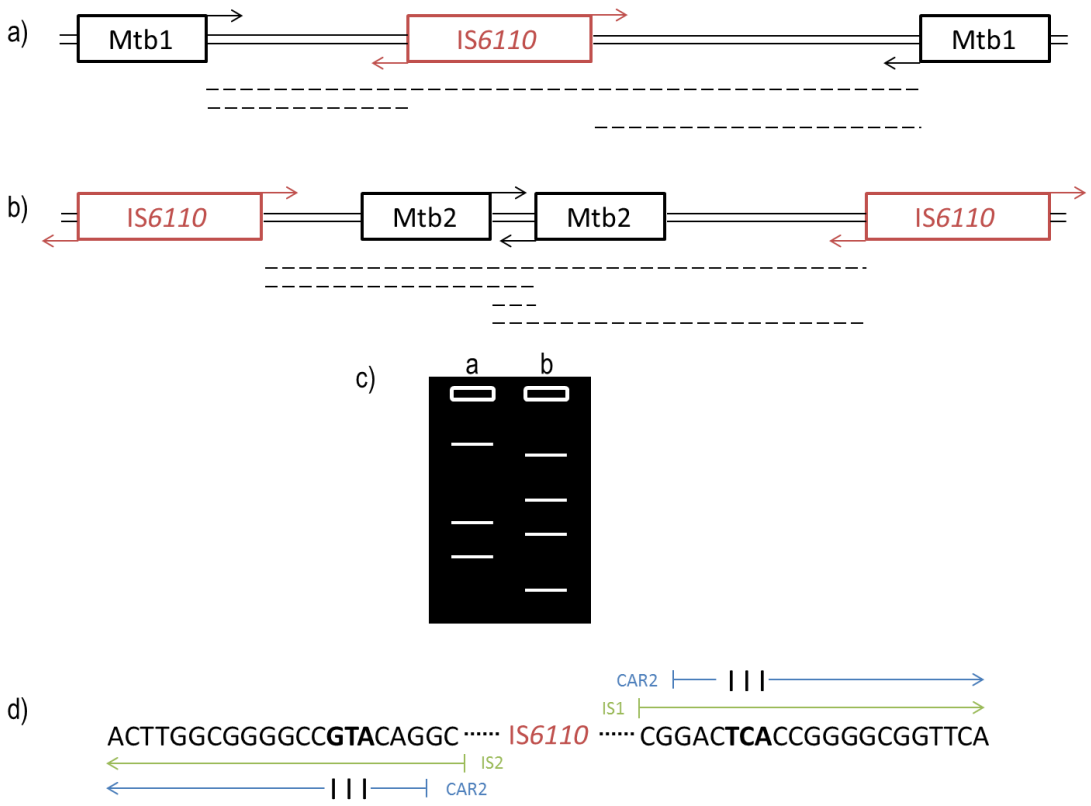


Figure 3.3: Theoretical illustration of genotyping by PCR amplification between copies of IS*6110* and Mtb1 (a) or Mtb2 (b). Depiction of agarose gel (c) shows the resulting banding pattern for the above theoretical PCR reactions, while the enlargement of the IR region (d) shows the variation within the IR and the differences between oligonucleotides. Mtb1 and Mtb2 are oligonucleotides for common, frequently repeated 16-bp GC rich sequences found within the *M. tuberculosis* genome. The IR flanking IS*6110* differs by three nucleotides (indicated in bold). Oligonucleotides specific for each IR sequence (IS1 and IS2) were used together, or else a single degenerate oligonucleotide containing inosines (CAR2) was used in the amplification reaction. Figure adapted from Kotłowski *et al*. (2004).

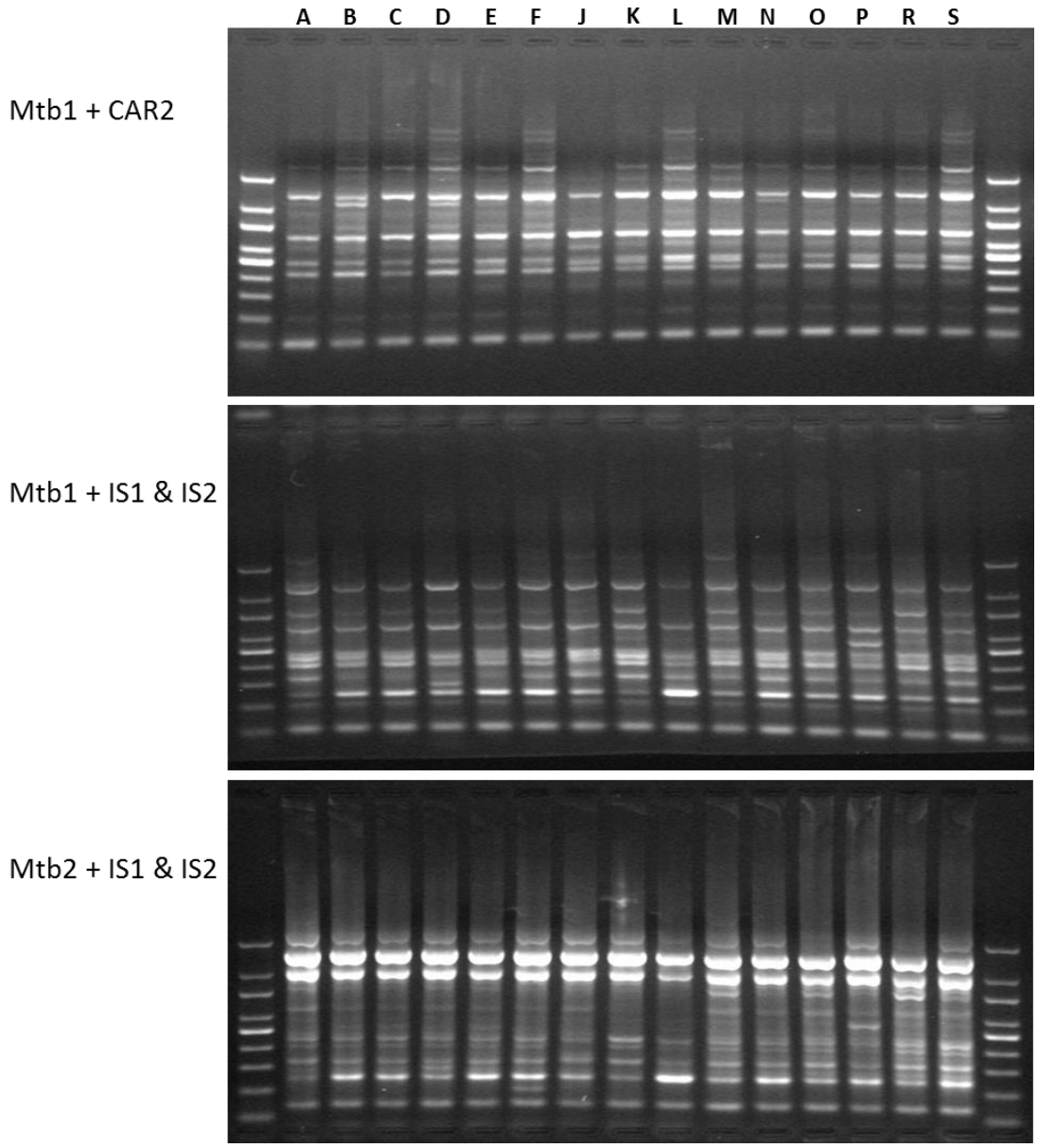


Figure 3.4: Genotyping banding patterns of 15 Rangipo clinical isolates based on amplification between IS*6110* and frequently occurring GC rich sequences. Mtb1 and Mtb2 are oligonucleotides for two frequently repeated 16-bp GC rich sequences. PCR amplification was performed with either Mtb1 or Mtb2 in combination with either two specific oligonucleotides (IS1 and IS2) or a single degenerate oligonucleotide (CAR2) for the IS*6110* IR. Letters at the top represent the 15 isolates tested.

3.2.2 Rangipo lineage classification

While the Rangipo strain had already been typed by RFLP and MIRU methods, it was unknown which *M. tuberculosis* lineage it belonged to. A series of PCRs based on LSPs identified by Gagneux *et al.* (2006) were performed to determine which of the six main *M.*

tuberculosis lineages the Rangipo strain belonged to, followed by sublineage classification PCRs.

3.2.2.1 Lineage-specific Large Scale Polymorphism (LSP) based PCR for *M. tuberculosis* Strain Classification

Multiplex PCRs including two reverse primers - one internal to the deleted region and the other immediately downstream of the LSP, were used to detect lineage-defining LSPs and provide a clear visual assignment of strain lineage based on the size of the PCR products produced. The presence of a band at the 'deleted' size for each RD PCR confirms the isolate belongs to that lineage, while a band of the 'intact' size suggests it does not. All nine Rangipo isolates tested were found to be intact for the four LSP PCRs which classify strains into five of the six lineages (see Table 3.3), suggesting they may belong to the Euro-American lineage (Figure 3.5). Euro-American lineage classification was determined by amplifying across the *pks1-15* gene and sequencing to confirm a representation of the isolates (C, F, S, T and U) contained the seven bp deletion characteristic of the Euro-American lineage (Gagneux *et al.* 2006) (Figure 3.6).

Table 3.3: Expected product sizes for LSP PCRs

Lineage	LSP	Product Size (bp)	
		Intact	Deleted
Indo-Oceanic	RD239	531	344
East African/Indian	RD750	556	371
East Asian (or West Beijing)	RD105	547	330
West African 1 / West African 2 (Atypical MTBC)	RD9	501	368
Euro-American	*pks1-15*	561*	N/A

* Euro-American lineage classification is defined by a 7 bp deletion in the *pks1-15* gene, determined by DNA sequencing of the of *pks1-15* PCR fragment. N/A = Not applicable.

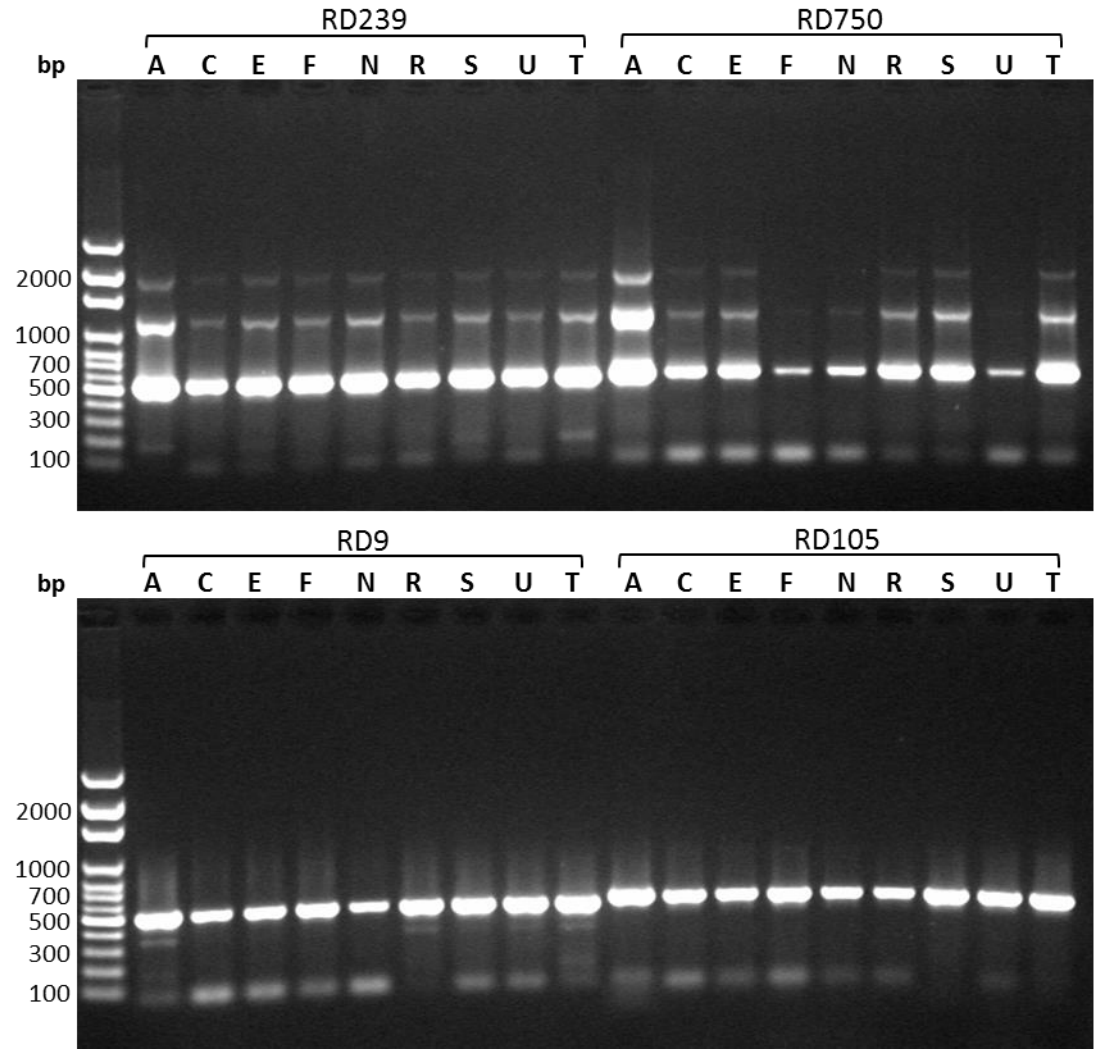


Figure 3.5: Lineage classification of nine Rangipo isolates using lineage specific LSP based PCRs. Letters across the top represent the nine isolates tested and RD numbers refer to the LSP tested. All four LSP PCRs show bands at the 'intact' size of 531, 556, 501 and 547 bp suggesting isolates do not belong to the Indo-Oceanic (RD239), East Africa/Indian (RD750), Atypical MTBC (including West African 1&2) (RD9) or East Asian (RD105) lineages respectively.

Some non-specific banding was observed in the RD239 and RD750 PCRs, but it was of lesser intensity and larger than that of the correct 'intact' product. Personal communication with Michael Reed (Reed *et al.* 2009) confirmed they also saw the same two larger non-specific bands in their 'intact' PCRs, while any isolates belonging to these lineages produced a very clear single band of the deleted size.

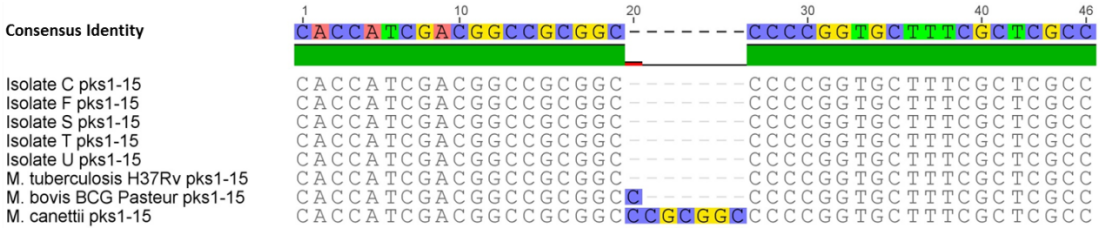


Figure 3.6: Alignment of Rangipo isolates over the pks1-15 deletion region confirming classification within the Euro-American lineage. All Rangipo isolates (C – U) contain the 7 bp pks1-15 deletion characteristic of the Euro-American lineage as shown by alignment with *M. tuberculosis* H37Rv (Euro-American), and the two control *M. bovis* BCG and *M. canettii* (non Euro-American) sequences.

3.2.2.2 Euro-American Lineage Classification Based on the *katG*^{463} SNP Using the 'on/off Switch Assay'

Another marker used to define Euro-American lineage is the ctg to cgg substitution at codon 463 of *katG,* which encodes the catalase-peroxidase enzyme (Sreevatsan *et al.* 1997). Attempts were made at using an on/off switch assay (Section 2.1.3.3) to distinguish between Euro-American and non Euro-American genotypes, to reduce the time and costs involved in the above Euro-American strain classification method. The assay uses a forward primer common to both genotypes, while the two reverse primers are specific for the Euro-American (katG-R1) or non-Euro-American (katG-R2/3) strains with the *katG* 463 codon 'ctg' or 'cgg' respectively. Both reverse primers harbour a 3' end phosphorothioate-modified nucleotide so when PCR is carried out with a proofreading polymerase, a product is only produced when the matched primer for the *katG* 463 codon is present, and not the mismatched primer. Unfortunately, a band was produced when either of the reverse primers were used, on both isolates tested (S and T) (Figure 3.7). HPLC purified 3' end phosphorothioate-modified nucleotide reverse primers were then used in case the original primers contained a mixed population of oligonucleotides both with and without the modification which the polymerase could act on and produce a product, but both reverse primers still produced a product. Even though HPLC purification results in a purer oligo than a standard desalted purification, purity is only guaranteed at ~85 % so it is possible that there was still a mixed population of oligonucleotides present which is why a fragment was produced for both reverse primers.

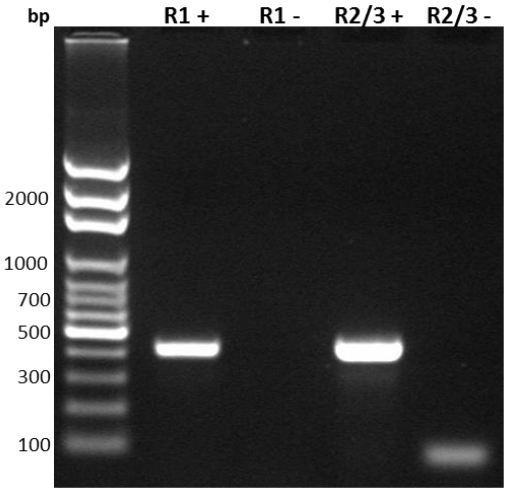


Figure 3.7: *katG* 463 codon phosphorothioate-modified PCR for Euro-American lineage classification. Rangipo isolate S (belonging to the *M. tuberculosis* Euro-American lineage) containing the *katG* 463 codon 'ctg', should amplify only with the matched reverse primer (R1) in this *katG* 463 codon phosphorothioate-modified PCR, and not with the mismatched reverse primer (R2/3). A product of the expected size (403 bp) was observed for both the matched (R1 +) and mismatched (R2/3 +) reactions above (no product for the no template negative controls (-)), indicating the *katG* 463 codon phosphorothioate-modified on/off switch PCR was unsuccessful and therefore is not an appropriate method for detecting strains belonging to the Euro-American lineage.

3.2.2.3 Euro-American Sublineage Classification

Euro-American sublineage classification PCRs were initially performed solely on isolate T, and then a representative sample of PCRs were repeated with isolates A, C, F, N, O, R, S and H37Rv to confirm the same results were obtained for all Rangipo isolates tested. All 10 region of difference (RD) PCRs defining each of the 10 Euro-American sublineages resulted in an 'intact' sized PCR product (see Table 2.1 and Table 2.2 for intact and deleted sizes). This suggests the Rangipo strain of *M. tuberculosis* does not belong to any of the sublineages currently defined by region of difference deletions. Figure 3.8a shows the 'intact' product obtained for RD 115, 122, 183, 193, 724, 726 and 761 PCRs on isolate T. Initial PCRs for the three remaining RD PCRs (174, 182 and 219) were unsuccessful due to the large size of the original amplicon and fragmented nature of the template DNA used (due to sonication), so were successfully repeated with internal flanking primers as seen in Figure 3.8b. Figure 3.8c shows the 'intact' product obtained for RD 182 and 219 PCRs on Rangipo isolates A, C, F, N, O, R, S as well as H37Rv, confirming the results obtained initially for Rangipo isolate T.

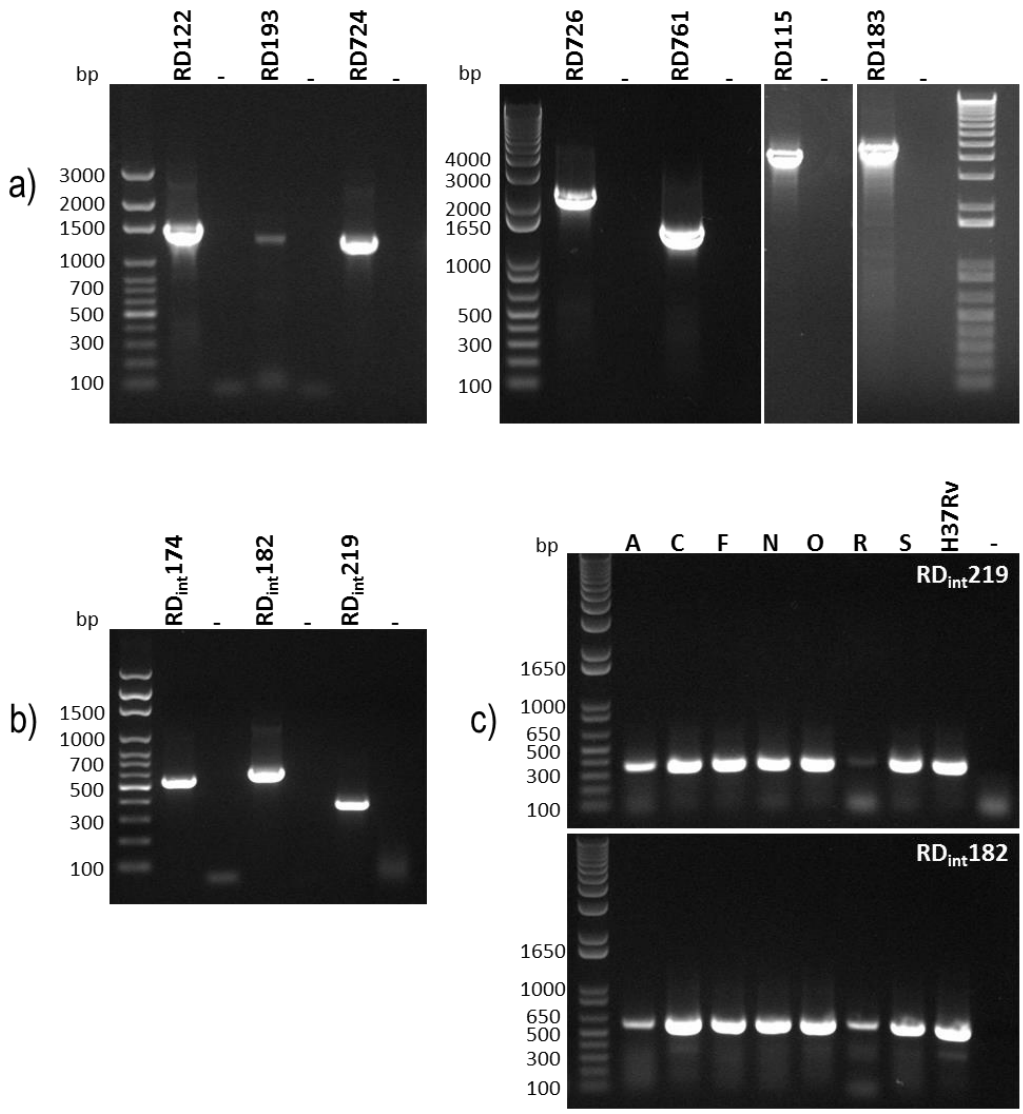


Figure 3.8: Region of Difference (RD) Euro-American sublineage classification PCRs. a) RD and b) RD internal (RDint) PCRs performed on Rangipo isolate T. Numbers across the top show the RD PCR performed, followed by the no template negative control (-). c) shows RDint219 (top) and RDint182 (bottom) PCRs performed on Rangipo isolates A, C, F, N, O, R and S as well as H37Rv and a no template negative control (-). All PCRs show products of the 'intact' RD size (see Table 2.1 and Table 2.2 for sizes) (Gagneux *et al*. 2006; de Jong *et al*. 2009) confirming the Rangipo isolates do not belong within any of the currently defined Euro-American sublineages.

3.2.2.4 Latin American Mediterranean (LAM) Strain Classification

Preliminary results from WGS suggested the Rangipo isolates may belong to the Euro-American LAM sublineage. Strains belonging to the Euro-American LAM lineage are found worldwide, and are characterised by an IS*6110* insertion at position 932204 of the H37RV

genome (Sampson *et al.* 1999) as well as by the LAM lineage specific SNP at codon 103 of Rv0129c which encodes Ag85C (Gibson *et al.* 2008).

Rangipo isolates A, E, S and T were subjected to LAM-specific IS*6110* PCR (Section 2.1.3.5) and produced a 141 bp fragment (Figure 3.9a), not a 205 bp fragment, indicating no LAM-specific IS*6110* element is present and they therefore do not belong to the LAM lineage (Marais *et al.* 2006). A second LAM-specific screen was also performed on isolates A, C, F, N, O, R and S as well as H37Rv to confirm the Rangipo strain did not belong to the LAM lineage. RFLP analysis using the LAM-specific SNP at codon 103 of the gene encoding Ag85C (Rv0129c) showed an absence of the SNP due to the 519 bp PCR product being cut into three fragments (365, 96 and 58 bp) instead of two (461 and 58 bp) (Figure 3.9b), as is seen in LAM strains containing the SNP (Gibson *et al.* 2008).

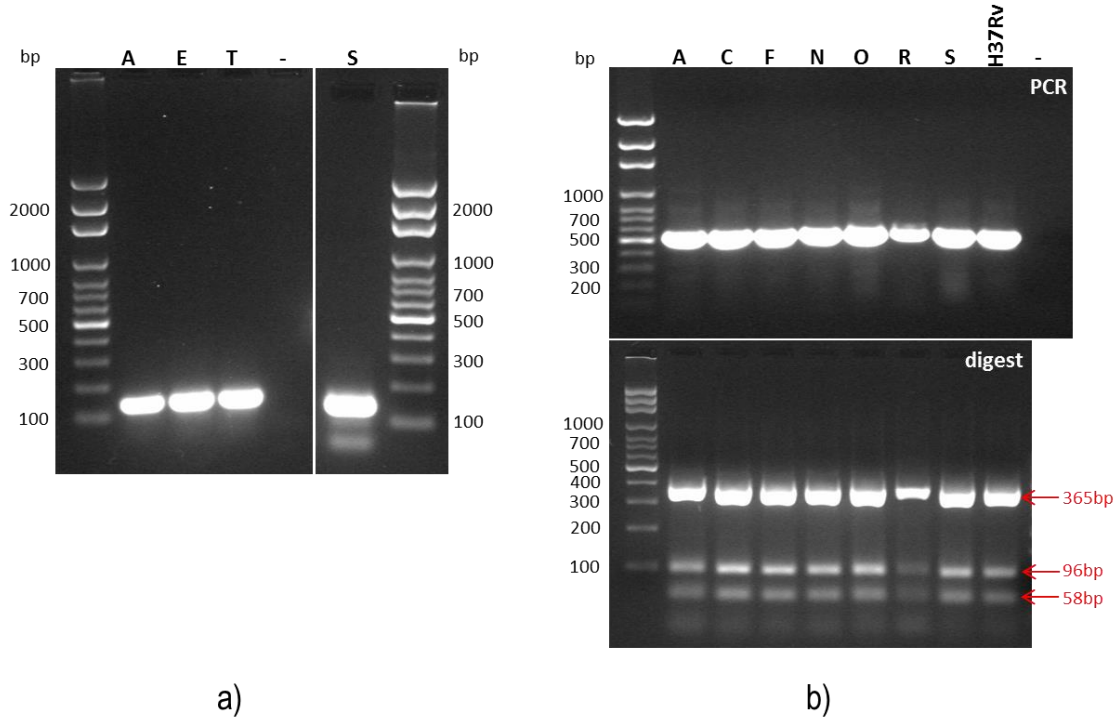


Figure 3.9: LAM-specific IS*6110* PCR (a) and Ag85C RFLP analysis (b) of Rangipo isolates. a) Rangipo isolates A, E, T and S were amplified with LAMF, LAMR and XhoI primers resulting in an approximately 141bp product indicating the absence of the LAM specific IS*6110* insertion element. b) Rangipo isolates A, C, F, N, O, R and S as well as H37Rv were amplified with Ag85C103F and Ag85C103R primers in a LAM specific Ag85C103 SNP RFLP. The resulting 519 bp products (top gel) were digested with MnlI then run on a 4 % agarose gel (bottom gel). The presence of the 3 bands at 365, 96 and 58 bp indicates a lack of the AG85C103 SNP and confirms Rangipo is not a LAM strain.

The Rangipo strain was successfully determined as belonging to the Euro-American lineage of *M. tuberculosis* (Figure 3.6), but did not match to any of the classified Euro-American sublineages, suggesting it belongs to an as-yet unclassified Euro-American sublineage.

3.2.3 Rangipo-Specific SNPs

3.2.3.1 Whole Genome Sequencing SNP Results

WGS of 10 Rangipo isolates was employed to identify SNPs common to all of the Rangipo isolates sequenced, in order to design a robust Rangipo specific diagnostic. WGS was carried out by our collaborators at the Department of Medicine, University of Medicine and Dentistry of New Jersey (USA) as per (Colangeli *et al.* 2014), and identified 247 putative Rangipo-specific SNPs.

3.2.3.2 Identification of Rangipo-Specific SNPs

All 247 SNPs were cross checked against all mycobacterial strains available in the TB Database (TBDB) (http://www.tbdb.org) and Broad Institute database (http://www.broadinstitute.org/scientific-community/data) to check for the presence of annotated SNPs in other strains at the putative Rangipo-specific SNP position: nine were found not to be SNPs relative to the H37Rv genome reference strain, 80 were present in at least one other strain, 59 were located in the same position as a different SNP in at least one other strain and 99 were not currently annotated in any other strain in the databases (see Table 3.4). A list of mycobacterial strains (and their lineage classification) used for the SNP analysis is included in Appendix A.2. Any putative Rangipo-specific SNPs which were found within other strains, or within genes containing a high number of annotated SNPs (and therefore in a region of high genetic variation) were ruled out for use as a Rangipo-specific marker before a shortlist of Rangipo-specific SNPs was established. The shortlist of Rangipo SNPs included those in genes containing no (or few) annotated SNPs in other mycobacterial strains, or those which fell within genes with interesting functions such as virulence, persistence/dormancy, pathogenesis or within *vapC* genes. There were five SNPs which fell within genes containing no annoted SNPs (Rv0071, Rv2546, Rv3119, Rv3573c and Rv3769),

and two which were within *vapC* toxin genes (Rv1953 and Rv2546). The full list of SNPs identified by WGS and annotations is incuded on the compact disk - Appendix D.

Table 3.4: Classification of 247 putative Rangipo-specific SNPs

Classification	# of SNPs
Present in H37Rv and ∴ not a SNP	9
Present in at least one other strain	80
Located in same position as a different SNP	59
Not in any other strains and ∴ Rangipo-specific	99
Total SNPs identified by WGS	247

3.2.3.3 Amplification and Sequencing to Confirm SNPs

A representative sample (10) of the 99 Rangipo-specific SNPs identified above were PCR amplified from Rangipo isolate S, ExoISAP treated then sequenced as for Section 2.1.3.6.1 to confirm presence of the SNP. The presence of each SNP (relative to H37Rv) was confirmed by aligning the sequence with that from the H37Rv genome (see Appendix C.3).

3.2.4 Development of a Rangipo RFLP Diagnostic Assay

SNPs can be detected by RFLP analysis if only one of the two variant bases creates a cleavage site for a restriction enzyme. Rangipo-specific SNPs were screened in WatCut to identify REs which would selectively recognise only one of the two base variants at each SNP site as for Section 2.1.6.1. A number of the preferentially ranked SNPs were ruled out from being used in the diagnostic at this stage because no suitable RE could be identified to distinguish between the variants for those SNPs. The suitability of REs was based on: availability; cost; the number of base changes required for the recognition sequence; and the number of RE recognition sites within the amplification region in addition to that at the SNP position, to allow for clearly distinguishable cleavage fragments. An example of the WatCut SNP-RFLP analysis is shown in Figure 3.10, where HaeII has been chosen as the RE to differentially cleave samples with the SNP variant 'G' for the Rv3119 SNP. HaeII was chosen because it requires no changes to be made in the primer sequence in order for the RE to

recognise the sequence. This allows for flexibility in primer design when amplifying the region surrounding the SNP, as one of the primers does not need to be positioned in this region like it would be if HincII was used for example, where a T must be changed to a G. HaeII was chosen over the other seven REs which also require no base changes for recognition due to its availability already in the lab.

Rv3119 (#1) GCTCCCCACCGTGCGGAGGCATTTG [A/G] CGCCGCGCGTATGTGTATCGAGTTG

Enzyme	recognizes	cleaves at	SNP site / mutations	Changes
BanI	G^GYRCC	0	..GTGCGGAGGCATTTG G CGCCGCGCGTATGTG..	0
BbeI	GGCGC^C	4	..GTGCGGAGGCATTTG G CGCCGCGCGTATGTG..	0
EheI	GGC^GCC	2	..GTGCGGAGGCATTTG G CGCCGCGCGTATGTG..	0
HaeII	RGCGC^Y	4	..GTGCGGAGGCATTTG G CGCCGCGCGTATGTG..	0
HgaI	GACGCN5^	9	..GTGCGGAGGCATTTG A CGCCGCGCGTATGTG..	0
KasI	G^GCGCC	0	..GTGCGGAGGCATTTG G CGCCGCGCGTATGTG..	0
NarI	GG^CGCC	1	..GTGCGGAGGCATTTG G CGCCGCGCGTATGTG..	0
NlaIV	GGN^NCC	2	..GTGCGGAGGCATTTG G CGCCGCGCGTATGTG..	0
HincII	GTY^RAC	-1	..GTGCGGAGGCAGTTG A CGCCGCGCGTATGTG..	1
Hpy8I	GTN^NAC	-1	..GTGCGGAGGCAGTTG A CGCCGCGCGTATGTG..	1
AflIII	A^CRYGT	1	..GTGCGGAGGCATTTG A CGCGTCGCGTATGTG..	2
AjuI	GAAN7TTGGN7^	8	..GTGAAGAGGCATTTG G CGCCGCGCGTATGTG..	2
AscI	GG^CGCGCC	1	..GTGCGGAGGCATTTG G CGCGCCGCGTATGTG..	2
BdaI	TGAN6TCAN10^	20	..GTGCGGAGGCATTTG A CGCCGCTCATATGTG..	2
BssECI	C^CN2GG	-4	..GTGCGGAGGCCTTG G CGCCGCGCGTATGTG..	2
BssHII	G^CGCGC	1	..GTGCGGAGGCATTTG G CGCGCCGCGTATGTG..	2
BstXI	CCAN5^NTGG	-3	..GTGCCCAGGCATTTG G CGCCGCGCGTATGTG..	2
Cac8I	GCN^NGC	3	..GTGCGGAGGCATTTG G CGCGCCGCGTATGTG..	2

Figure 3.10: WatCut SNP-RFLP analysis of the Rangipo-specific SNP in Rv3119. The sequence surrounding the 2 SNP variants A/G (indicated by the square brackets) for Rv3119 was analysed in WatCut, which returns a list of REs able to selectively recognise only one of the two base variants. The SNP variant recognised by each RE is shown in blue text while any changes required for the RE recognition sequence are shown in red. The RE HaeII (highlighted in green) was chosen as the Rv3119 RFLP RE because it requires no changes to be made to the recognition sequence unlike others such as HincII (higlighted in purple), and because it was already available in the lab.

3.2.4.1 Uniplex Rangipo-Specific SNP RFLP Assays

Initially, SNP-RFLP assays were conducted singularly to check products amplified for both Rangipo and non-Rangipo samples, and REs selectively cut the products as expected in PCR buffer at 37 °C. Table 3.5 shows details of the theoretically suitable SNP-RFLP assays, and summarises the success of each applied uniplex assay. Figure 3.11 shows SNP containing PCR products both before and after RE digestion, for H37Rv as the non-Rangipo control and isolate S as the Rangipo control.

Table 3.5: Summary of SNP-RFLP designs and applied outcomes.

SNP	Substitution (S/NS)	# annotated SNPs in gene	gene function	RE	# base changes required	SNP base variant recognised	strain differentially cut	outcome of physical uniplex assay
Rv0071	NS	0	Deleted in Beijing isolates (RD105). Orthologs = RNA-directed DNA polymerase.	ApaI	2	**G**/T	NR	x
Rv0405	NS	39	pks6, pathogenesis, virulence[1]	HinfI	2	**C**/T	NR	~
Rv0719	NS	3	50S ribosomal unit protein	RsaI	0	**T**/C	NR	✓
Rv1821	S	6	secA2, pathogenesis, blocks apoptosis of infected macrophages, encodes a component of a virulence related secretion system[2]	Sau3AI	0	**G**/A	NR	✓
Rv1854c	NS	3	ndh	BclI	0	**T**/G	NR	✓
Rv2224c	NS	2	caeA, pathogenesis, post-translational protein modification, response to host immune response, required for survival and virulence in mice[3]	AvaI	1	**C**/A	NR	x
Rv2504c	NS	1	scoA	DraIII	0	**G**/T	NR	✓
Rv2546	NS	0	vapC18	EcoR1	2	**G**/A	NR	x
Rv2716	NS	3	conserved hypothetical protein	DraIII	2	**G**/C	NR	~
Rv3119	NS	0	moaE1	HaeII	0	A/**G**	R	✓
Rv3513c	NS	0	fadD18	HaeIII	0	T/**G**	R	x
Rv3616c	NS	7	espA, pathogenesis, interacts with ESX-1 genes	TaqI	1	C/**A**	R	x
Rv3769	NS	0	hypothetical protein	DraIII	2	**C**/T	NR	✓
Rv3867	NS	1	conserved hypothetical protein	MluI	1	G/**C**	R	x
Rv3881c	NS	8	espB, pathogenesis, inhibits phagosome maturation[4]	MvaI	0	A/**C**	R	x

Bold type indicates which of the two SNP base variants is recognised, NR = non-Rangipo, R = Rangipo, x = SNP-RFLP assay not successful, ~ = SNP-RFLP assay moderately successful, ✓ = SNP-RFLP assay successful. Notes on gene functions were primarily gathered from the TBDB' Functional Annotations' or from other sources as indicated

[1] (Hisert *et al.* 2004), [2] (Baena & Porcelli 2009), [3] (Lun & Bishai 2007), [4] (Xu *et al.* 2007)

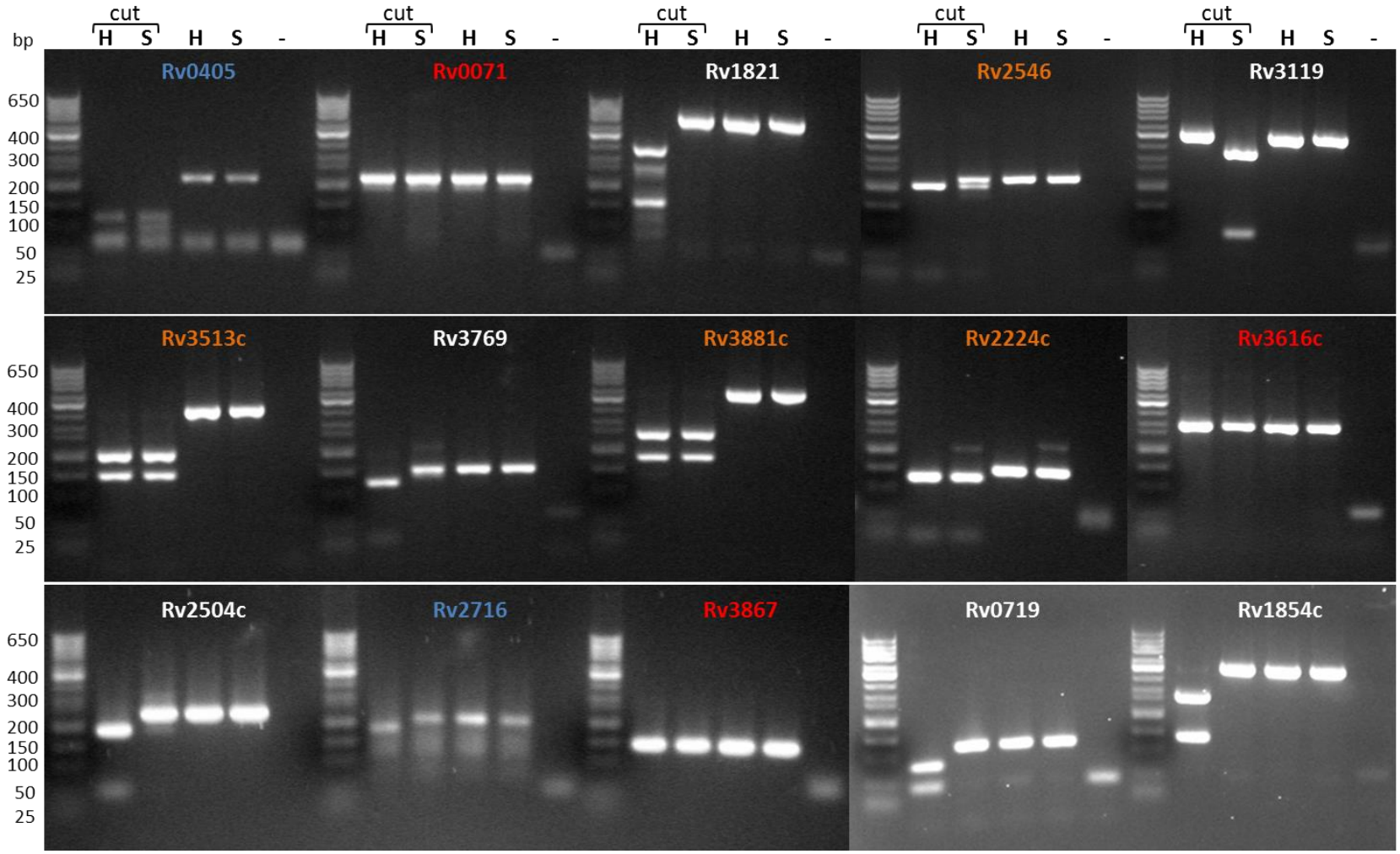


Figure 3.11: SNP-RFLP uniplex assays. SNP-RFLP assays showing resulting fragments for Rangipo (isolate S (S)) and non-Rangipo (H37Rv (H)) samples, both before and after (cut) RE digestion. Differentially digested products were obtained from SNP-RFLP assays for Rv1821, Rv3119, Rv3769, Rv2504c, Rv0719 and Rv1854c while the rest were either not cut (indicated in red text), had both SNP variants cut (orange), or the digest patterns for each variant were not clearly distinguishable from each other (blue).

SNPs which were not cut as expected were Rv0071, Rv2224c, Rv2546, Rv3513c, Rv3867 and Rv3881c. Rv0071, Rv3616c and Rv3867 were not cut for either variant, Rv0071 and Rv3616c require digestion with ApaI at 25 °C and TaqI at 65 °C respectively, instead of 37 °C so are not ideal SNPs to include in the multiplex diagnostic, and the MluI used to digest Rv3867 was expired, which may be why it did not digest the product. Both SNP variants were cut for Rv2224c and Rv2546 (Rv2546 only partially so both 'cut' and 'un-cut' products were observed) indicating some degeneracy in the RE recognition sequences. REs cut non-SNP specific RE recognition sites for the Rv3513c and Rv3881c products, but not for the Rangipo SNP specific RE recognition site in Rangipo isolate S. These SNP-RFLP assays required no base changes in order for the RE to recognise the SNP, so theoretically if the SNP is present (as it should be in isolate S) cleavage should occur here as well as at the other recognition sequences within the amplification product. The absence of cleavage at the SNP specific RE recognition sequence initially suggested the absence of these SNPs in isolate S, but while the Rv3513c SNP had not been confirmed by PCR and sequencing, the Rv3881c SNP had, (Section 3.2.3.3) which ruled this possibility out. Another analysis program was used to theoretically digest the Rv3881c product in case there was an error in the SNP-RFLP design but the same size fragments and cut sites were observed as those obtained from using the Geneious software. MvaI cuts CC/WGG sequences (where W stands for A or T, and / designates the cleavage site), of which there are two in the Rv3881c product for isolate S; CC/**A**GG for the SNP specific recognition site and CC/**T**GG for the non-SNP specific site. No evidence for preferential cleavage of T over A sites could be found in the literature for MvaI in non-methylated DNA so the reason for MvaI not cleaving at both the Rv3513c and Rv3881c SNP specific recognition sites remains unknown. The above six SNPs which were not cut as expected in the SNP-RFLP assays were not carried through to the following stage of diagnostic development.

SNPs which were cut as expected but produced fragments hard to visually distinguish were those in Rv0405 and Rv2716. The Rv0405 SNP produced cut fragments of 118, 93 and 11 bp for Rangipo and 118, 71, 22 and 11 bp for non-Rangipo samples which were too close to easily distinguish from each other, while the Rv2716 SNP did not amplify cleanly under the standard diagnostic conditions (probably due to the higher Tm of the primers; 68 °C compared

to ~ 60 °C). Both Rv0405 and Rv2716 SNPs were not carried through to the following stage of diagnostic development as they produced fragments which were not easily distinguishable and therefore not suitable for the diagnostic.

SNPs which were cut producing clearly distinguishable fragments of the expected sizes were Rv3119 (which cuts Rangipo the strain) and Rv0719, Rv1821, Rv1854c, Rv2504c and Rv3769 which cut non-Rangipo strains.

3.2.4.2 Multiplex Rangipo-Specific SNP RFLP Assays

Duplex and triplex SNP-RFLP assays were theoretically digested to ascertain which combinations would produce fragments that easily distinguished between Rangipo and non-Rangipo strains. Figure 3.12 shows theoretical digest results for the Rv1821, Rv2504c and Rv3119 triplex SNP-RFLP assay. Promising combinations (see Table 3.6) were then performed experimentally to confirm that the specific REs cut the duplex or triplex products as expected (Figure 3.13).

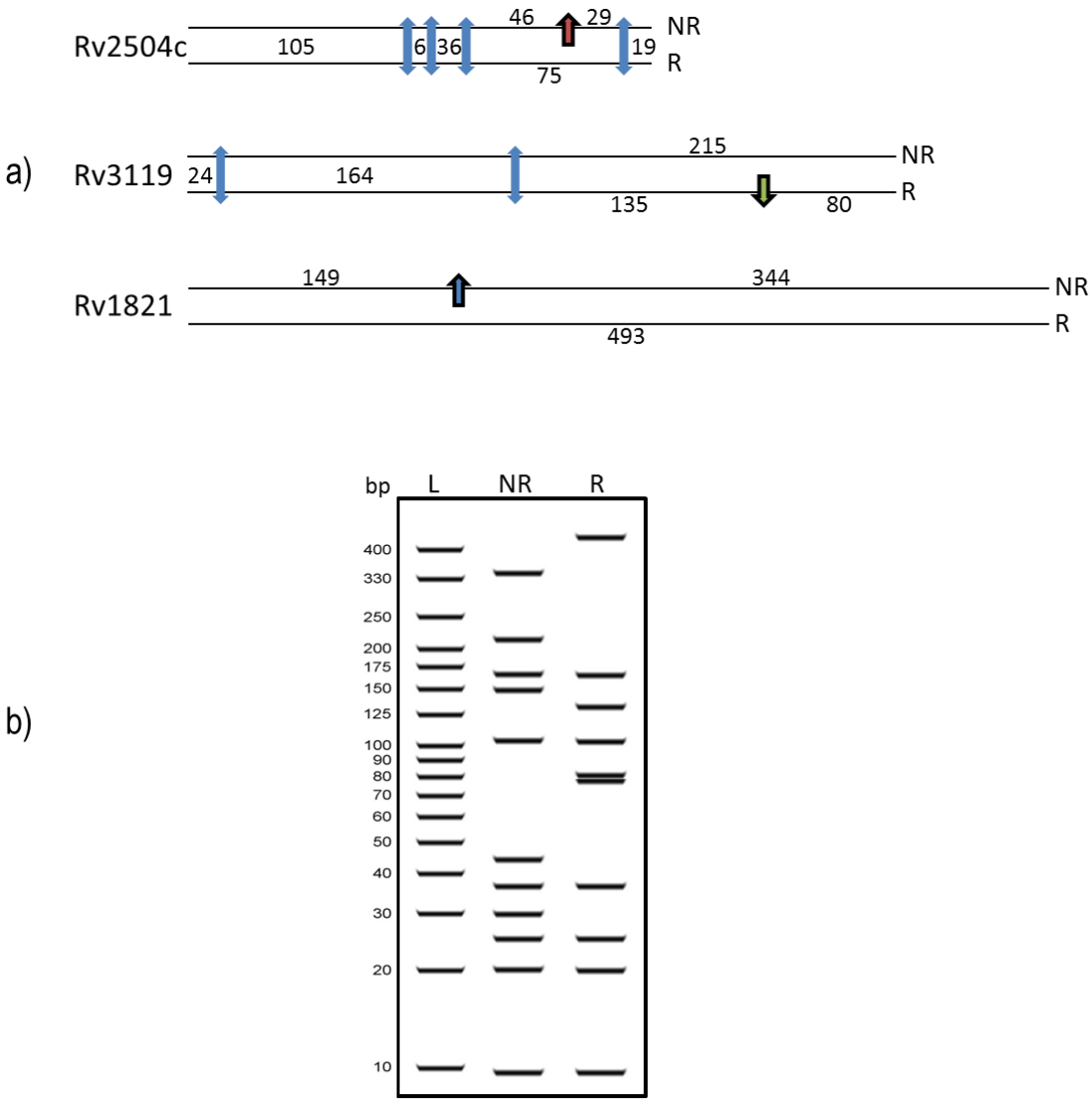


Figure 3.12: Theoretical Rv1821, Rv2504c and Rv3119 triplex SNP-RFLP assay. a) Schematic showing positions of RE recognition sequences (arrows) and the resulting fragment sizes (in bp) once each of the three amplicons is digested. Single arrows targeting only one of the two strains (N = non-Rangipo, R = Rangipo) represent the RE recognition sequence for the SNP within that gene, where blue = Sau3AI, green = HaeII and red = DraIII. b) Theoretical banding patterns of fragments resulting from the triplex SNP-RFLP assay. L = ladder, NR = non-Rangipo and R = Rangipo.

Table 3.6: Promising SNP-RFLP combinations and fragment sizes.

SNP combination	identifier	un-cut size (bp)	cut size - R (bp)	cut size - NR (bp)
Rv0719, Rv3119	1	119, 403	80, 119, 323	40, 79, 403
Rv2504c, Rv3119	2	241, 403	80, 241, 323	48, 193, 403
Rv1821, Rv3119	3	493, 403	25, 80, 135, 164, 493	25, 149, 164, 215, 344
Rv0719, Rv1821, Rv3119	4	119, 493, 403	25, 80, 119, 135, 164, 204, 289	25, 40, 55, 79, 149, 164, 215, 289
Rv1821, Rv2504c, Rv3119	5	493, 241, 403	6, 19, 24, 36, 75, 80, 105, 135, 164, 493	6, 19, 24, 29, 36, 46, 105, 149, 164, 215, 344

Identifier refers to the SNP combinations displayed in Figure 3.13 below, cut size refers to the size of predicted fragments after RE digestion, R = Rangipo, NR = non-Rangipo.

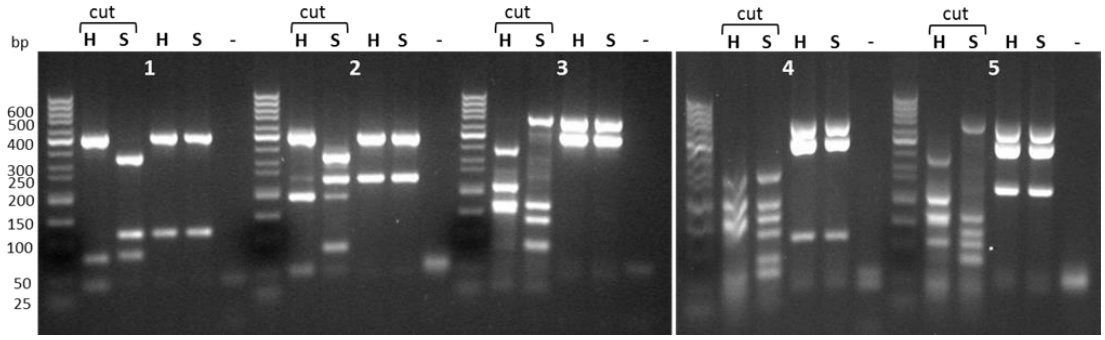


Figure 3.13: SNP-RFLP duplex and triplex assays. SNP-RFLP assays showing resulting fragments for Rangipo (isolate S (S)) and non-Rangipo (H37Rv (H)) samples, both before and after (cut) RE digestion. Expected fragment sizes are shown in Table 3.6 above for each duplex (assays 1 - 3) and triplex (assays 4 & 5) reaction. SNP combinations are as follows: 1 = Rv0719/Rv3119; 2 = Rv2504c/Rv3119; 3 = Rv1821/Rv3119; 4 = Rv0719/Rv1821/Rv3119 and 5 = Rv1821/Rv2504c/Rv3119. Differentially digested products were obtained from SNP-RFLP assays 1, 2, 3 and 5, while fragments produced for assay 4 were not clearly distinguishable from each other.

Fragments which are similar in size often appear as one band, while fragments which are small (less than 25 bp) are not usually observed on the agarose gels, therefore, the optimal assay produces vastly different sized fragments and/or a different number of fragments between the two SNP variants which are easily distinguishable by eye. All of the assays in Figure 3.13 above show distinct fragment banding for the two strains, except for assay 4 where the band sizes are too similar to separate them easily by eye on the gel. Assay 5 is therefore the best triplex SNP-RFLP assay to use for the Rangipo diagnostic, as the fragment banding is easily distinguishable between Rangipo and non-Rangipo strains, contains a RE recognition site specific for both Rangipo (Rv3119) and non-Rangipo (Rv2504c and Rv1821) SNP variants as

a built in control, and uses REs which function together at the same temperature, for the same duration and in PCR buffer alone.

3.2.4.3 Optimisation of RE Digest in the SNP-RFLP Assays

All previous RE digests for SNP-RFLP assays had been performed overnight at 37 °C which meant it took at least one day to obtain the results from the diagnostic assay. A RE digest timecourse assay was performed to determine the minimum amount of time the products needed to be digested for to achieve full digestion, in order to reduce the turnaround time and thereby improve the diagnostic assay. The triplex assays were digested for 2 hr, 4 hr or overnight, this time using DraIII-HF in place of DraIII as the DraIII had been discontinued. The digests appeared smeary and showed non-specific cleavage products even after 2 hr (Figure 3.14a). While the DraIII-HF supposedly has reduced star activity compared to its DraIII predecessor, it appears to be exhibiting a lot of star activity under these triplex conditions where PCR buffer is used instead of RE buffer, and where 3 µl of REs are used in a 13 µl total volume resulting in a high glycerol concentration (11.5 %). A series of triplex assays were digested for 2 hr with reducing concentrations of the REs to eliminate the star activity of the DraIII-HF, until suitable fragment banding was observed (Figure 3.14b), like that observed previously when DraIII was used instead of DraIII-HF (see Figure 3.13). All future digests were performed with 0.4 µl of each RE for 2 hr, which reduced the total time taken for the Rangipo SNP-RFLP assay to approximately 4.5 hr.

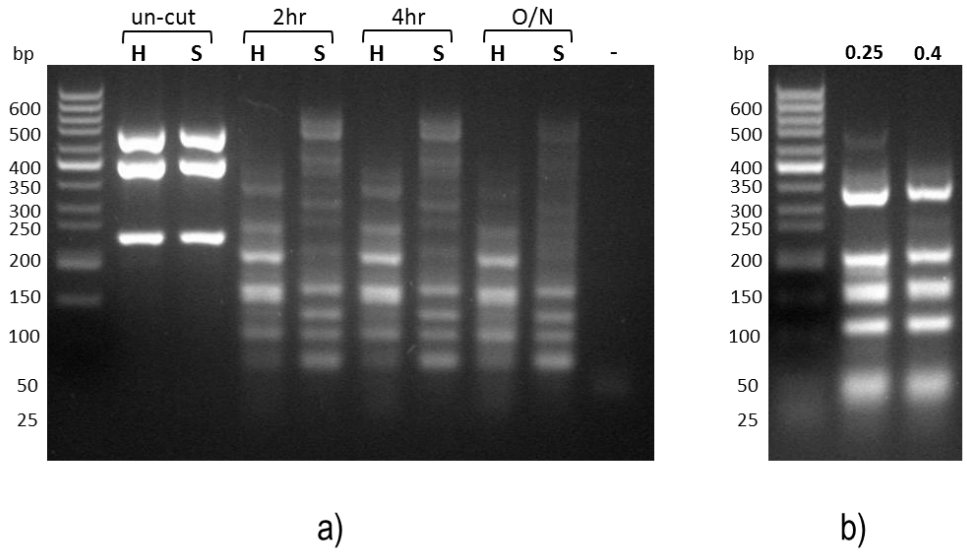


Figure 3.14: SNP-RFLP triplex RE digest timecourse assays using DraIII-HF instead of DraIII a) Initial assays were performed with 1 µl of each RE for 2 hr, 4 hr or overnight (O/N) resulting in non-specific digestion of both the non-Rangipo (H37Rv (H)) and Rangipo (isolate S (S)) control samples. b) Optimised assays on the H37Rv non-Rangipo control were digested for 2 hr only with either 0.25 or 0.4 µl of each RE resulting in the correct banding pattern of fragments at 344, 215, 164+149, 105, and 46+36+29+24 bp (where numbers on either side of the + appear as one band on the gel).

An attempt was made to reduce the cost of the diagnostic by replacing the RE Sau3AI with one of its cheaper isoschizomers, MboI in the triplex digest. When used in combination with the other two RE used in the SNP-RFLP assay (HaeII and DraIII-HF), MboI didn't cut the Rangipo and non-Rangipo samples as efficiently, and also appeared to cut the non-Rangipo sample incorrectly so Sau3AI continued to be used.

3.2.4.4 Taguchi Based Optimisation of Multiplex PCRs

A Taguchi based optimisation was performed on the triplex PCR to determine the optimal conditions to then be used in the Rangipo SNP-RFLP assay. Optimisations were performed with both Rangipo (isolate C) and non-Rangipo (H37Rv) DNA to find the condition which produced each of the three products in relatively equal amounts for both strains. Figure 3.15 shows condition #14 amplified and was digested well for both strains, so this PCR condition was used for the Rangipo triplex SNP-RFLP diagnostic assay; 1 x PCR buffer, 0.1 mM dNTPs, 2.4 mM $MgCl_2$, 3 U *Taq* DNA Polymerase, 0.1 µM Rv3119 and Rv1821 primers and 0.2 µM Rv2504c primers, and 10 ng template DNA annealed at 68 °C.

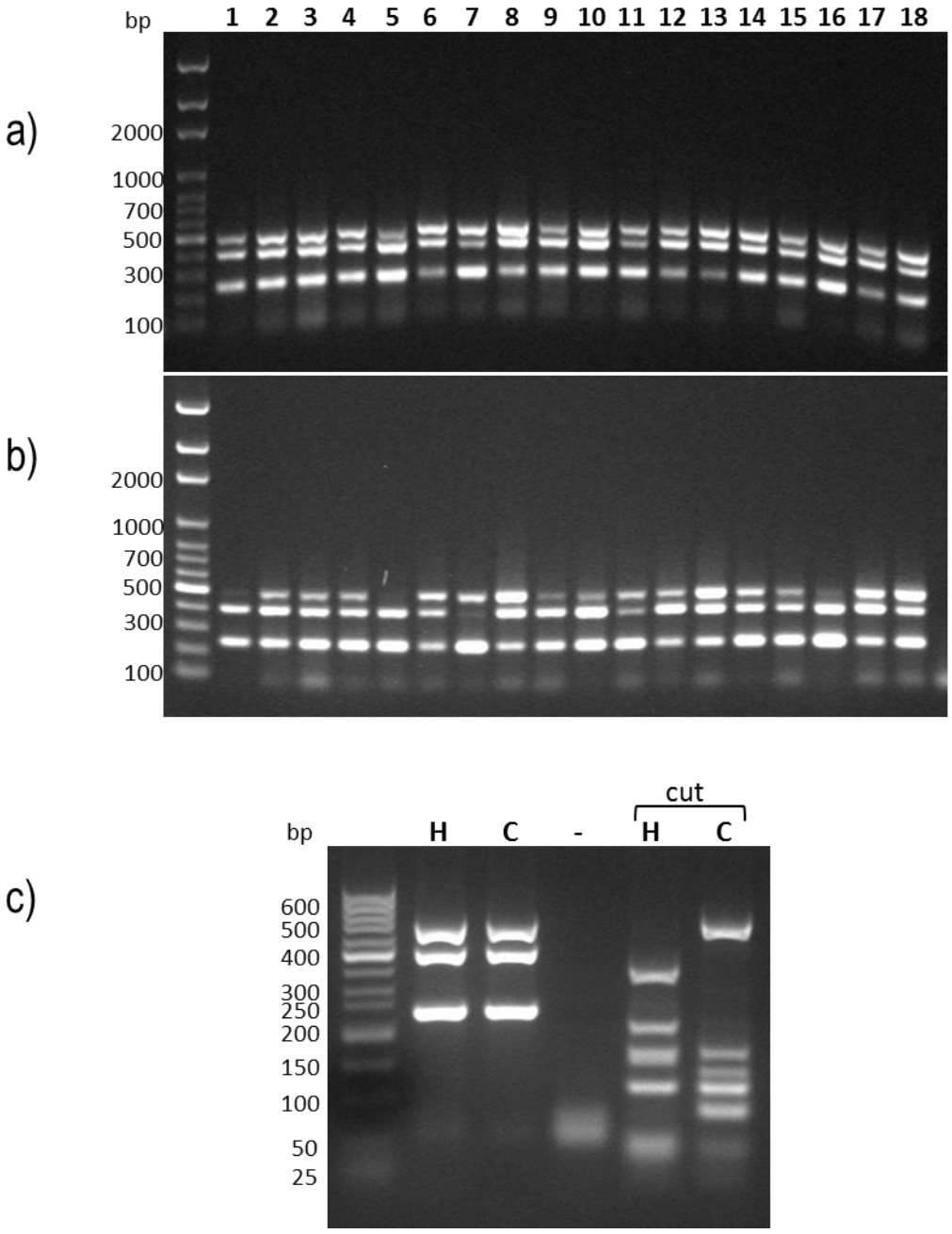


Figure 3.15: Taguchi optimisation of SNP-RFLP triplex PCR using a) non-Rangipo DNA and b) Rangipo DNA. The banding pattern obtained from the 18 reactions of the orthogonal array. Numbers across the top correspond to the 18 reactions in the orthogonal array (see Table A.3 in Appendix A.1). The 3 bands represent each of the 3 amplification products for Rv1821 (493 bp), Rv3119 (403 bp) and Rv2504c (241 bp). c) shows both Rangipo (isolate C) and non-Rangipo (H37Rv) samples were successfully amplified then digested using the #14 conditions from the orthogonal array, - depicts the negative control.

3.2.5 Screening Clinical Isolates with the Rangipo Specific PCR/RE Diagnostic

Clinical *M. tuberculosis* isolates from LabPLUS of known MIRU type were initially screened with the Rangipo SNP-RFLP diagnostic to confirm it correctly identifies all Rangipo samples without any false positives or negatives. Figure 3.16 shows no non-Rangipo samples were identified as Rangipo, and all Rangipo samples were correctly identified while one sample (#12) assigned to the Rangipo strain by MIRU typing was found to be unique by the SNP-RFLP diagnostic. Upon discussion with LabPLUS, it was ascertained that sample #12 was from a case which likely contracted TB outside of New Zealand, and the strain was not actually Rangipo in the strictest sense, but probably a related strain. This illustrates the limitation of MIRU typing for distinguishing closely related strains (even with MIRU 24 typing), and re-affirms why WGS is becoming the method of choice for pathogen typing. The figure also shows that this Rangipo SNP-RFLP diagnostic assay groups samples in identical groups as the MIRU typing does, as well as differentiating isolates belonging to the Otara (062) cluster (responsible for the third largest cluster endemic to New Zealand).

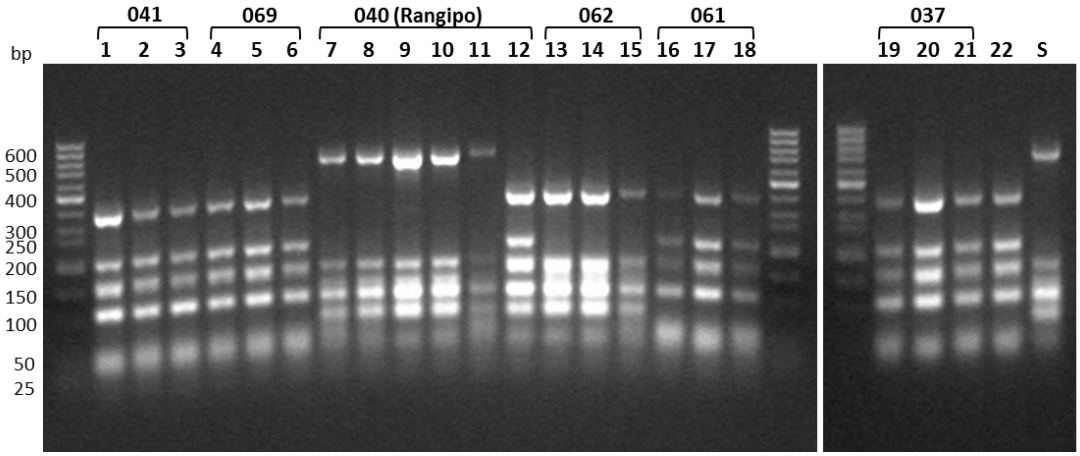


Figure 3.16: Rangipo SNP-RFLP diagnostic screening of clinical isolates of known *M. tuberculosis* MIRU type. Samples 1 - 22 were screened with the Rangipo SNP-RFLP diagnostic assay. The three digit number at the top of samples indicates which 12 MIRU loci and therefore strain they belong to (040 = Rangipo, 062 = Otara, 061 = Southern Cross) according to MIRU typing. Sample 22 = H37Rv and S = the Rangipo isolate S positive control. The banding patterns show that no non-Rangipo samples share the same banding as samples from the Rangipo strain, and all Rangipo samples were correctly identified as Rangipo by their banding pattern, except for sample 12 which was grouped as Rangipo by MIRU typing but showed a unique MIRU 24 pattern so is likely to infact be a related strain to and not Rangipo.

Batches of 'blind' clinical samples (labelled with an identification number only) were then screened with the Rangipo SNP-RFLP diagnostic and the results sent to LabPLUS for confirmation according to their MIRU typing results. Figure 3.17a shows the banding patterns obtained from the diagnostic for all blind samples, which correctly identified all 4 Rangipo and the 40 non-Rangipo samples. Rangipo isolates A, C, F, N, O, R and S were also screened with the Rangipo SNP-RFLP diagnostic assay confirming the same banding pattern (Figure 3.17b).

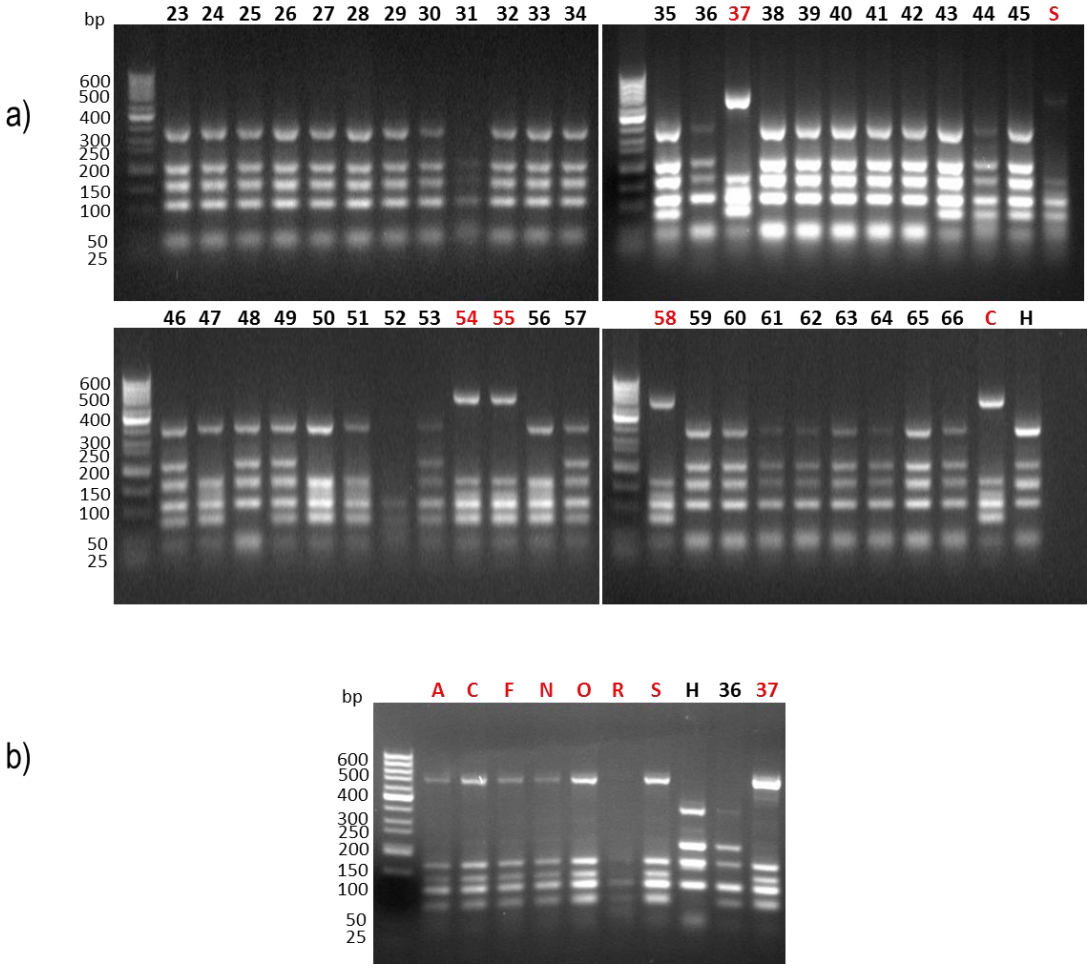


Figure 3.17: Rangipo SNP-RFLP diagnostic screening of 'blind' clinical isolates of unknown *M. tuberculosis* MIRU type, and Rangipo isolates A, C, F, N, O, R and S. a) blind *M. tuberculosis* samples 23 - 66 were screened with the Rangipo SNP-RFLP diagnostic assay revealing 4 Rangipo (shown in red text) and 40 non-Rangipo samples when compared with the Rangipo (isolates C & S) and non-Rangipo (H37Rv (H)) controls. b) Rangipo isolates A, C, F, N, O, R and S were screened along with a H37Rv non-Rangipo control (H), and repeats of blind samples 36 and 37 which were not clear in a) (above). All Rangipo samples were correctly identified, as confirmed by MIRU typing results from LabPLUS.

3.3 Conclusions

The Rangipo strain of *M. tuberculosis* is responsible for the largest cluster of tuberculosis infections within New Zealand. This strain has been circulating readily in the Waikato and Bay of Plenty regions for a quarter of a century, and its control has proved difficult owing to its high transmissibility and virulence, as well as a lack of a rapid and reliable test to identify the strain.

During the course of this research, the Rangipo strain was identified through LSP and RD PCR analysis, as belonging to an as yet unclassified Euro-American sublineage of *M. tuberculosis*. Robust strain classification is essential for revealing how *M. tuberculosis* has evolved and may continue to do so (Achtman 2008), and how strains can differentially effect the pathology of the disease (de Jong *et al.* 2008; Thwaites *et al.* 2008).

10 Rangipo isolates were fully sequenced and MIRU typed by collaborators, and during the course of this research Rangipo specific SNPs were identified and used to develop a Rangipo strain specific diagnostic assay. This RFLP based diagnostic is able to quickly and sensitively distinguish Rangipo isolates from other *M. tuberculosis* strains (4.25 hr once DNA has been obtained), with high specificity and at little cost. While the samples tested here involved DNA from fully grown cultures, it is expected that the diagnostic would also work well using DNA from partially grown cultures and even directly from sputum (for culture positive samples), given the extreme sensitivity of PCR based methods which are already used in TB detection tools such as the Xpert MTB/RIF system. The median time to start treatment from the onset of symptoms is approximately 2 months in New Zealand (Lim & Heffernan 2013). This delay before treatment starting is precariously long in the context of the Rangipo strain due to its high transmissiblity and virulent nature, and the majority of Rangipo infections being pulmonary. The ability of the Rangipo diagnostic assay to reduce the diagnosis turnaround time will be immensely beneficial in disrupting the chain of transmission of this highly transmissible strain, which is instrumental in controlling tuberculosis infection (Barnes & Cave 2003).

Chapter Four:

Expression of *vapB* and *vapC* in *Mycobacterium smegmatis* using RT-qPCR

4.1 Introduction

Mycobacterial VapBC systems have been linked to pathogenesis, persistence, cell death and response to stress (Gerdes *et al.* 2005; Korch *et al.* 2009; Ramage *et al.* 2009; Min *et al.* 2012). Here, reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) is used to attempt to identify conditions in which the *M. smegmatis vapB* or *vapC* genes are up or downregulated, in order to gain a better understanding of the biological roles and mechanisms involved in the regulation of the numerous VapBC systems in *M. tuberculosis*. Quantitative real-time polymerase chain reaction (qPCR) is used extensively to quantitate RNA and DNA. RT-qPCR is an extremely sensitive and robust method for measuring the expression of genes of interest (GOI), validating microarray experiments, and monitoring biomarkers. RT-qPCR is especially useful for detecting lowly expressed mRNA transcripts, or where there are small changes in expression levels (Bustin 2000). As such, it has become known as the gold standard in quantification of mRNA and validation of microarray data (Canales *et al.* 2006), but due to the variability in materials and protocols used, and limited understanding of the

principals involved in the data analysis, it is not a standard assay. The possibility that VapBC systems may play a role in the ability of *M. tuberculosis* to survive in the varied and hostile conditions/environments it encounters, suggests growing *M. smegmatis* under conditions of stress may be an ideal way to elicit a change in the expression of *vapB* or *vapC*. Many microarray and RT-qPCR studies performed in *M. smegmatis* have shown a multitude of stressors can effect the expression of particular genes, including temperature shock, nutrient starvation, oxidative stress, and antibiotic stress (Fernandes *et al.* 1999; Smeulders *et al.* 1999; Milano *et al.* 2001; Shires & Steyn 2001; Singh & Singh 2009). The expression of these particular genes is regulated by the bacteria, in order for it to adapt to the particular stress condition(s) it is experiencing.

4.1.1 The Principles and Pitfalls of Real-Time Reverse-Transcription PCR (RT-qPCR)

In conventional PCR, only the end product can be quantified (by running it on an agarose gel) which is time consuming and does not reliably reflect differences in the starting quantity of the template material. Real-time PCR differs from conventional PCR in that the increase in amplification product is measured in real time by using a fluorescent reporter molecule, with product quantification after every cycle. The ability to isolate just the exponential phase of amplification for analysis where component concentrations are not limiting, allows information about the initial quantity of template DNA to be calculated. In the case of RT-qPCR, where RNA is first converted into cDNA by the reverse-transcription reaction, the difference in mRNA expression levels of particular GOI can be determined. RT-qPCR has traditionally been used to detect lowly expressed mRNA transcripts and validate microarray data, but with the recent introduction of RNA sequencing (RNA-Seq) for mapping and quantifying entire transcriptomes, RT-qPCR will become an even more important tool for validating the differentially expressed transcripts identified (Fang & Cui 2011).

There are two main methods for detecting amplicons in RT-qPCR: gene-specific fluorescent probes which depend on Förster Resonance Energy Transfer (FRET) technology to produce a fluorescence signal via the uncoupling of a fluorogenic dye molecule and its quencher, and

non-sequence specific fluorescent intercalating double stranded (ds) DNA binding dyes. The most common probe based system is TaqMan which is sequence specific, requires little optimization but is expensive. Fluorogenic intercalating dyes are inexpensive and have much a higher fluorescence when bound to dsDNA compared to the unbound dye (Vitzthum *et al.* 1999), but are not sequence specific so require thorough optimisation to distinguish specific from non-specific products. SYBR Green I has been used as the default fluorogenic dye until recently, when other dyes such as SYTO 82 were found to have a much better detection limit than SYBR Green. This is because unlike SYBR Green I, they don't inhibit PCR and do not preferentially bind to GC rich sequences (Gudnason *et al.* 2007). This makes SYTO 82 a valuable tool in detecting the lowly expressed TA genes in the GC rich (67 % (Schneider *et al.* 2006)) *M. smegmatis* bacterium.

There is no universally accepted method for data normalisation which accounts for all variables possible during the course of a RT-qPCR experiment. Therefore, selection of a normalliser is dependent upon the studies aims, and levels of acceptable tolerance (Huggett *et al.* 2005). The main strategies are: normalisation to sample size, total RNA quantity, or the common practice of normalising to an internal reference or housekeeping (HK) gene. These are not mutually exclusive and it is recommended that sample sizes are matched, quality RNA of a similar quantity is used for the RT reactions, and an endogenous control (reference gene) is also measured (Huggett *et al.* 2005).

There are two main methods of gene quantification: absolute and relative. Absolute quantification uses a standard curve for each reaction run and determines the copy number or concentration of the target transcripts based on their quantification (or threshold) cycle (Cq) values. While this method can correct for differences in primer efficiencies, there is a large reduction in the number of samples that can be analysed per run. Relative quantification determines variation in mRNA levels of a target gene relative to a reference/HK gene, and is expressed as a relative fold change in target expression values. There are many mathematical models available to calculate relative expression, some take into account the amplification efficiency of the PCR reaction and correct for this while others do not, instead assuming an 'ideal' efficiency where the PCR product doubles every cycle during the exponential phase.

Models which correct for efficiency are highly recommended and result in a more accurate estimation of the relative expression ratio, as not all PCRs have an ideal amplification efficiency, and small efficiency differences between genes can result in a large over- or under-estimation of the 'real' initial mRNA amount (Pfaffl 2004).

Both of the above quantification strategies require normalisation of target gene expression to a reference or housekeeping gene. This is to account for sample to sample variation (such as yield and integrity of RNA, RT efficiency, cDNA loading variation and PCR efficiency), which should affect the normaliser and the GOI equally, and ensures that all the steps required to obtain the final PCR measurement are controlled for within the same sample (Nolan *et al.* 2006). The ideal housekeeping gene must be stably expressed under all the experimental conditions used, express at a similar level relative to target gene levels (Pfaffl 2004), and not have related pseudogenes. Housekeeping gene expression can vary considerably between cell type, stage of growth, and experimental conditions (Thellin *et al.* 1999; Bustin 2000) so the chosen housekeeping gene(s) needs to be experimentally validated. Recent studies identifying suitable housekeeping genes suggest that the ideal and universal control gene does not exist, so where possible, multiple housekeeping genes should be used for reliable normalisation of RT-qPCR data (Pfaffl *et al.* 2002; Vandesompele *et al.* 2002; Andersen *et al.* 2004).

4.1.1.1 Minimising Technical Variation

The difference in mRNA expression levels between samples is not just due to biological variation, it is also a result of experimentally induced or technical variation. In order to determine an accurate account of the changes occurring at the biological level, technical variation must be kept to a minimum. Technical variation may be introduced at multiple stages along the RT-qPCR process, and includes factors such as differences in the quality and amount of starting material, overall transcriptional activity, and enzymatic efficiencies in the RT and RT-qPCR processes (Vandesompele *et al.* 2002; Pfaffl 2004). To minimise technical variation, the materials and protocols employed to isolate, quantitate and reverse transcribe the RNA must be kept consistent, and also be able to produce material of sufficient quantity

and quality for the downstream qPCR application. Any remaining differences in the amount and quality of starting material can then be normalised for by using a suitably selected housekeeping gene, to more competently assess the true biological variation.

4.1.1.2 Analysis and Quantitation of Results

A typical RT-qPCR amplification consists of an exponential phase (where the amount of amplification product approximately doubles per cycle), followed by a plateau phase once reagents begin to deplete. Initially (in the exponential phase), fluorescence remains at background levels with no detectable increase in signal from the amplified product. At a particular point in the reaction, enough product is accumulated to produce a detectable fluorescent signal above the background noise, and the cycle number at which this occurs is called the quantification cycle (Cq) (also referred to as the threshold cycle (Ct), crossing point (Cp), and take-off point (TOP)). The linear phase is reached when reaction components become limited and the reaction no longer progresses exponentially (mainly due to product renaturation competing with primer binding (Dharmaraj), and the point at which this begins varies greatly. Because the Cq is obtained in the exponential phase where reaction components are not limited and amplification is extremely reproducible, this methodology can be employed to calculate the starting quantity of template. In a typical reaction, a higher starting template concentration will result in an earlier or lower Cq, as relatively fewer amplification cycles are needed for the amplification product to rise above the background fluorescence signal.

The two main strategies commonly used to quantify results obtained by real-time RT-qPCR are the standard curve method (for both absolute and relative quantitation), and the comparative Ct method (for relative quantitation only). The standard curve method uses a curve constructed from RNA of known concentration, as a reference to extrapolate the titre of template mRNA in samples. While this method is relatively simple and robust (Cikos *et al.* 2007), it requires dilution curves to be run for both the standard and reference genes which takes up valuable space during each run. An alternative model was proposed which utilised the observed correlation between PCR efficiency and Cq to estimate the initial copy number of template

molecules without having a standard of known concentration (Wilkening & Bader 2004). Here, the average number of copies of PCR product at the Cq for 16 PCRs using a known amount of starting material, was found to be $1.12 \times 10^{10} \pm 3.66 \times 10^{9}$. This value was then substituted into the equation in order to calculate the approximate initial copy number of unknown samples where the PCR efficiency is known (see Equation 1).

Equation 1

$N_0 = 1.12 \times 10^{10}/E\text{^}CN$, where

N_0 = initial copy number or starting concentration

E = PCR amplification efficiency

^ = denotes 'to the power'

CN = the cycle number at the Cq

For relative quantitation, the comparative Ct method (also called the $\Delta\Delta$Ct method) (Livak & Schmittgen 2001) also removes the need for a standard curve, by comparing the Cq values (referred to as Ct values when the equation was created) of the samples of interest with a control (such as a non-treated sample or RNA from a wt cell) after they have been normalised to a HK gene. The original comparative Ct method however, assumes that the amplification efficiencies of both the HK gene and the target GOI are approximately equal and are at maximum (ie.100 %) (see Equation 2), which is often not the case. A modified method from the original calculation was developed by Pfaffl (2001) which factors in the efficiency of each of the amplification reactions (see Equation 3), and gives a more accurate estimation of the initial mRNA concentration (Pfaffl 2001; Pfaffl 2004).

Equation 2

$\Delta\Delta Ct = \Delta Ct\ (sample) - \Delta Ct\ (control)$, where

$\Delta Ct = Ct\ (GOI) - Ct\ (HK)$, therefore

$\Delta\Delta Ct = (Ct\ (HK\ sample) - Ct\ (GOI\ sample)) - (Ct\ (HK\ control) - Ct\ (GOI\ control))$, where

Ct (*GOI)* = Ct value of gene of interest in the sample or control

Ct (*HK)* = Ct value of housekeeping gene in the sample or control

The ratio of the GOI in the sample relative to in the control is calculated as follows:

$R = 2^{\Delta\Delta CT}$.

Equation 3

$$R = \frac{(E_{GOI})^{\Delta Ct\ GOI\ (MEAN\ control\ -\ MEAN\ sample)}}{(E_{HK})^{\Delta Ct\ HK\ (MEAN\ control\ -\ MEAN\ sample)}}$$, where

$\Delta Ct = MEAN\ Ct\ (control) - MEAN\ Ct\ (sample)$, where

R = expression ratio of the gene of interest relative to the control, normalised to the HK

E = PCR amplification efficiencies of gene of interest or HK gene

^ = denotes 'to the power'

Ct control = Ct value for 'control' gene of interest or HK gene

Ct sample = Ct value for 'sample' gene of interest or HK gene

There are many programs for analysing RT-qPCR data, both off the shelf and instrument specific, each using different methodologies to interpret the resulting data. The plethora of instruments and methods for analysing the resulting RT-qPCR data makes interpreting and comparing results from the literature difficult. One of the earlier analysis programs published (2003) is LinRegPCR, which has been used extensively since, with improvements made along the way (Ramakers *et al.* 2003; Ruijter *et al.* 2009). The LinRegPCR program performs a baseline correction and determines a window-of-linearity for the data, then uses linear regression analysis to calculate the PCR efficiency for each sample from the slope of the resulting regression line. The starting concentration (expressed in arbitrary fluorescence units) for each sample is then calculated (based on Pfaffl's equation) using PCR efficiency and Cq

values, which can then be used to determine the normalised, relative change in expression levels (see Equation 4). This calculation is very similar to Equation 1 derived from the computation model, except it is not based on a set value obtained from the averaging the number of copies of PCR product at the Cq from a random group of 16 genes, but instead more accurately utilises the Nq values for the actual amplicons in question.

Equation 4

: $R = \frac{AVERAGE\ (N0_{GOI})}{AVERAGE\ (N0_{HK})}$, where

R = expression ratio of the gene of interest, normalised to the HK

N0 = Nq / (E(MEAN)^Cq) , where

N0 = starting concentration (expressed in arbitrary fluorescence units)

Nq = fluorescence threshold set to determine Cq

E(MEAN) = mean amplicon PCR efficiency

^ = denotes 'to the power'

Cq = number of cycles needed to reach Nq (previously referred to as Ct)

Statistical software has been developed to help the user identify the most appropriate housekeeping genes for normalisation of their experimental data set. GeNorm uses the average pairwise variation of an individual gene to other genes to determine their expression stability, and progressively eliminates the least stabily expressed ones (Vandesompele *et al.* 2002). NormFinder also assigns expression stability values to single candidate reference genes, but uses a model-based approach, and factors in the possible variations within and between sample groups (Andersen *et al.* 2004). The BestKeeper algorithm combines all reference genes into an index, and calculates the correlation between it and each of the reference genes (based on their raw Cq values) utilising the coefficient of variance, standard deviation and coefficient of correlation variables (Pfaffl *et al.* 2004). While there is no consensus on which software is the best, GeNorm is the most widely used, and often, multiple algorithms are compared in order to identify which genes are consistently ranked as the most stable.

Where experimental design and treatment(s) effect housekeeping gene expression, multiple endogenous housekeeping genes, or an introduced RNA standard may be required for normalisation. Whilst spiking in a synthetic RNA (usually with the ends identical to the target gene so as the same primer sequence can be used for both the synthetic standard and target) can be useful in normalising reaction efficiencies, they are costly and easily degraded, making them a less satisfactory option than using multiple housekeeping genes. GeNorm and BestKeeper are two algorithms which take into account the the geometic mean of multiple housekeeping genes, to reduce the effect of treatment(s) on these housekeeping genes, and thereby strengthen the validity of the resultant data (Vandesompele *et al.* 2002; Pfaffl *et al.* 2004).

The vast array of sample preparation methods, reagents, platforms, normalisation and analysis strategies which are available to users, combined often with a limited understanding of the principals involved in RT-qPCR analysis, makes interpretation of the data difficult and results sometimes impossible to reproduce. A unified approach to RT-qPCR is therefore required for it to become a standard assay.

4.1.2 Isolation of Mycobacterial RNA

Intact mycobacterial mRNA is difficult to extract due to the physical strength and mycolic acid content of the cell wall, the short half-life of RNA (Lewin 1990), and the low content of RNA per unit of DNA (Verma *et al.* 1999). The functional half-life of the bulk of *E. coli* mRNA is only 2.4 minutes (Regnier & Arraiano 2000). Degradation (by both endo and exonucleases) of the transcribed mRNA is thought to begin within one minute of the start of transcription (Lewin 1990), so in order to extract high yields of intact mycobacterial RNA the cells must be rapidly disrupted, releasing their RNA directly into chaotropic buffers (such as guanidinium isothiocyanate to inhibit any endogenous RNases (Chomczynski & Sacchi 1987)) and preserved (Dietrich *et al.* 2000).

RNA is somewhat unstable and easily degraded by inadequate sample handling or storage once extracted (Schoor *et al.* 2003; Bustin *et al.* 2005; Perez-Novo *et al.* 2005), and RNA

quality and yield can be vastly different between two extraction methods, even if just the method of homogenisation changes (Fleige *et al.* 2006). This is also true for the purity of RNA, it has been shown that levels of DNA carried over during the RNA extraction process varies greatly between extraction methods (Schagat *et al.* 2008). Starting with RNA of poor purity or integrity may affect the results of downstream applications such as gene expression evaluation which can be labour-intensive, time-consuming and highly expensive (Raeymaekers 1993; Imbeaud *et al.* 2005). Traditionally, determination of RNA purity was assessed via analysis of the A260/A280 ratio and its integrity analysed via rRNA bands on an agarose gel. The recent introduction of micro-fluidic capillary electrophoresis instruments such as the Agilent Bioanalyzer (Agilent Technologies) have allowed RNA quality, quantity and integrity to be measured much faster and using only picogram amounts of RNA.

The first mycobacterial RNA extraction method (Katoch & Cox 1986) resulted in poor yields of RNA prone to degradation. There have been many different methods employed since utilising rapid lysis techniques such as sonication (Patel *et al.* 1991), the French press (Kinger *et al.* 1993) and bead beating (Hurley *et al.* 1987; Mangan *et al.* 1997), but the problem of obtaining intact RNA free of contaminating DNA still remains. A variety of factors such as yield variations, processing requirements, and sample availability of cells or tissues must be taken into account when choosing an extraction method as they can all influence the ability to obtain intact, high-quality RNA that is free of RNases, inhibitors, proteins and genomic DNA (Bustin & Nolan 2004).

Each isolation method extracts the total pool of RNA, of which only ~ 1 - 5 % is mRNA in bacteria (McGrath *et al.* 2008). An isolation method's efficiency or preference for extracting non-messenger RNAs (i.e., transfer RNA and ribosomal RNA which make up ~ 15 % and 80 % of the total RNA in bacteria respectively (Gonzalez-y-Merchand *et al.* 1998; Clark & Pazdernik 2013)) may lead to a difference in the amount of mRNA isolated in relation to the amount of total RNA present in a cell.

4.1.3 Objectives

The aims of this study were to identify conditions in which *M. smegmatis vapB* or *vapC* expression varied in order to help identify the mechanisms involved in the regulation of VapC, and to develop an optimised RNA isolation and RT-qPCR method suitable for the quantification of these target genes.

The biochemical function of VapC from *M. smegmatis* and activity of VapC proteins from *M. tuberculosis* was unknown at the beginning of my research.

4.2 Results and Discussion

4.2.1 Comparison of Protein Concentration and Optical Density

Tween 80 acts as a carbon source for mycobacteria (Saito *et al.* 1983; Wayne 1994) and inhibits the growth of some microorganisms, thereby reducing the possibility of contamination (Singh & Reyrat 2009). Detergents such as Tween 80 are also used to reduce clumping in bacterial cultures, which are most probably caused by exposed lipids of the cell surface establishing hydrophobic interactions (Singh & Reyrat 2009). Tween 80 was initially used here to reduce clumping in the wild type (wt) mc^26 cultures (as mc^2155 is a non-clumping daughter strain), but during the course of later environmental stress experiments, possible differences in gene expression between cultures grown in media with and without Tween 80 were observed. This difference in expression may be due to the ability of detergents such as Tween 80 to promote aerobic growth by improving the transfer of oxygen (Tang *et al.* 2009). To limit extrinsic variation, cultures were grown in media without Tween 80, and cell growth was estimated in relation to the OD 600 values of cultures grown in media with Tween 80 using the total cellular protein method.

The total cellular protein method (Meyers *et al.* 1998) was used to successfully estimate cell growth along with optical density measurements in *M. smegmatis* cultures grown both with and without Tween 80. *M. smegmatis* cultures were grown in LB or LBT and sampled over a time course to measure absorbance at OD 600 and total cellular protein concentration. Figure 4.1

shows a comparison of the protein and OD methods for monitoring the growth of *M. smegmatis*.

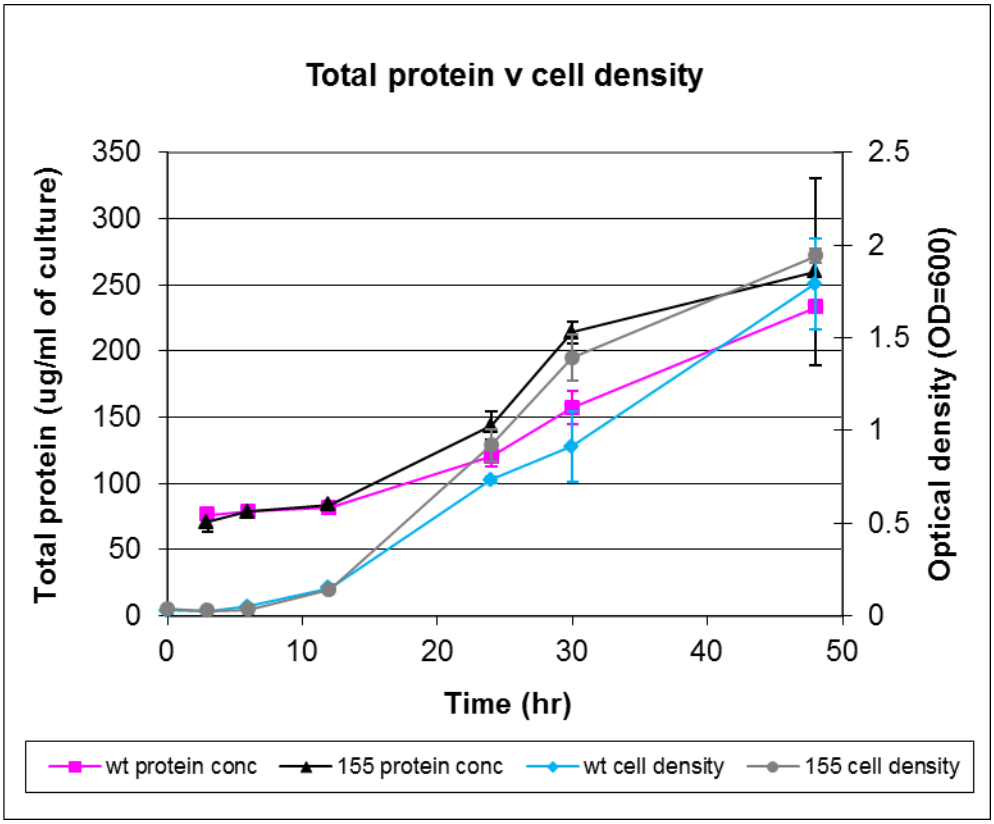


Figure 4.1: Growth curves of *M. smegmatis* cultures in LBT. Curves were plotted from total protein yield (pink and black), and OD 600 (blue and grey) data for both wt and mc2155 cultures. Each data point represents the average of two measurements. Error bars show the standard deviations of the means.

4.2.2 Optimisation of Total RNA Isolation

Several total RNA isolation procedures were compared in combination with DNA removal methods to determine an optimal protocol for isolating sufficient quantities of sound quality, pure RNA from *M. smegmatis*. Here we show how the different isolation methods produce total RNA varying in integrity, concentration, and residual DNA levels.

4.2.2.1 Comparison of RNA Isolation Methods

Methods investigated included variations of Trizol, TRI Reagent, and RNeasy supplier protocols, and modifications of the Acidic Guanididium Isothiocynate/Phenol/Chlorofom (AP-

GITC) procedure (Chomczynski & Sacchi 1987) including the use of BCP instead of chloroform, and performing two successive phenol-BCP extractions instead of one. Due to the recalcitrant nature of mycobacterial cell walls, all extraction methods involved initial bead beating to rapidly disrupt the cells. The quality of the extracted RNA was assessed by spectrophotometry using the NanoDrop, and by analysis of the 23S to 16S ribosomal RNA band ratio on a 1 % non-denaturing agarose gel. The absorbance ratios (A_{260}/A_{280} and A_{260}/A_{230}) for pure RNA at 230, 260 and 280 nm should be close to 2.0. Any deviation from this value indicates the presence of contaminants that absorb strongly at or near these wavelengths. Sharp 23S and 16S rRNA bands which display a ratio of close to 2:1, without the presence of degraded RNA (visable as smears or small fragments) indicates the RNA is intact and of sound quality (Sambrook *et al.* 1989).

Absorbance readings showed that the modified AP-GITC and RNeasy methods produced the most pure RNA (Table 4.1), as well as intact RNA when assessed on an agarose gel (Figure 4.2) for both the concentrated and dilute cell culture samples. The concentration of the total RNA extracted with the modified AP-GITC method was slightly lower than that for RNeasy (Table 4.1). This minor loss of RNA probably results from the multiple phase separation steps during the extraction method. Total RNA produced using the RNeasy method showed little 5S or other small or degraded RNAs. This is because RNAs less than 200 nucleotides (accounting for approximately 15 - 20 % of total RNA) do not bind the RNeasy column quantitatively and are selectively excluded, thereby enriching mRNA recovery (RNeasy® Mini Handbook 2012).

The TRI Reagent method extracted less total RNA compared to the modified AP-GITC and RNeasy methods (Table 4.1), as well as produced RNA of an inferior quality as demonstrated by the lower absorbance ratios (Table 4.1) and degraded 23S product (Figure 4.2). The low A_{260}/A_{230} ratio is likely due to the carry over of phenol from the TRI Reagent, which absorbs in the UV both at 230 nm and approximately 270 nm (260/280 and 260/230 Ratios 2008), and the production of degraded RNA is most likely due to the speed with which the cells are initially lysed and preserved. The modified AP-GITC method lyses cells immediately when the culture is added directly into GITC, whereas the TRI reagent method requires the cells to be pelleted (in order to remove media) prior to the addition of lysis solution. The RNeasy method also

requires the cells to be pelleted prior to adding the lysis solution, but the lack of degradation observed with this method may be due to the RNeasy lysis solution possibly acting faster than the TRI reagent, or the TRI reagent possibly preferentially isolating smaller RNA fragments.

Table 4.1: Comparison of RNA concentration and absorbance values extracted by three different methods

Method	Culture (OD 600)	Absorbance		Concentration (ng/µl)	Average concentration (ng/µl)
		A_{260}/A_{280}	A_{260}/A_{230}		
RNeasy	0.5	1.93	1.97	311	252
		2.13	1.97	193	
	1.0	1.06	1.96	520	642
		2.24	1.96	765	
TRI Reagent	0.5	0.94	1.89	160	185
		1.09	1.93	210	
	1.0	0.77	1.86	426	482
		1.37	1.94	537	
Modified AP-GITC	0.5	1.98	1.95	311	290
		1.73	1.95	268	
	1.0	1.98	1.94	506	556
		2.07	1.97	606	

Total RNA extractions were performed at two different OD's (OD 600 = 0.5 and 1.0), on two biological replicates

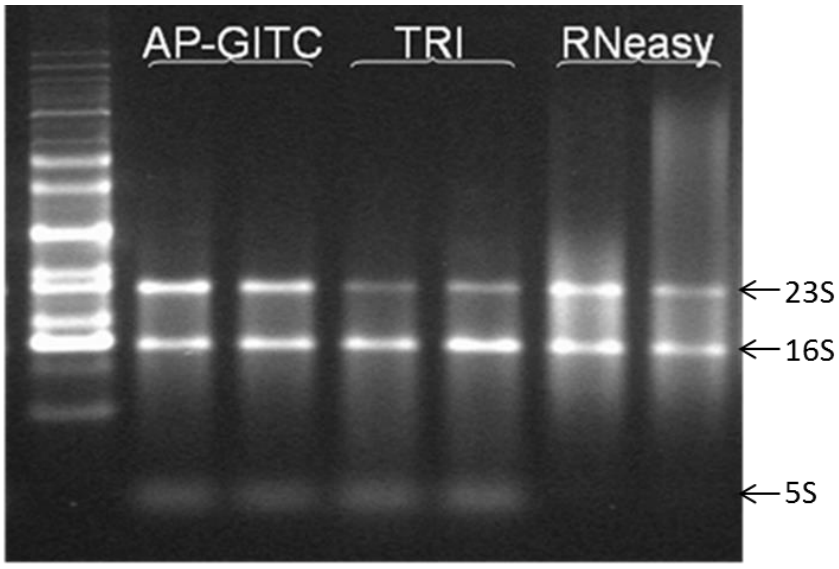


Figure 4.2: Non-denaturing agarose gel electrophoresis of *M. smegmatis* total RNA comparing RNA extraction methods. 800 ng of total RNA extracted by AP-GITC, TRI Reagant and RNeasy methods was separated on a 1% SB agarose gel to show the difference in integrity of resulting 23S and 16S products. Each extraction method was tested on cultures at two different OD's (OD 600 = 1.0 and 0.5 respectively).

4.2.2.2 Effect of DNase Treatment

While the modified AP-GITC and RNeasy total RNA extraction methods produced RNA of sound quality, the RNeasy extracted RNA co-purified with a large amount of genomic DNA. Figure 4.2 shows the co-purified DNA as a large molecular weight smear, and RT experiments confirm the presence of DNA by the visualisation of products in the noRT controls after RT-qPCR (Figure 4.3, lanes 2 and 4). Residual DNA in the RNeasy isolated RNA was also present after attempting on-column DNase digestion (data not shown).

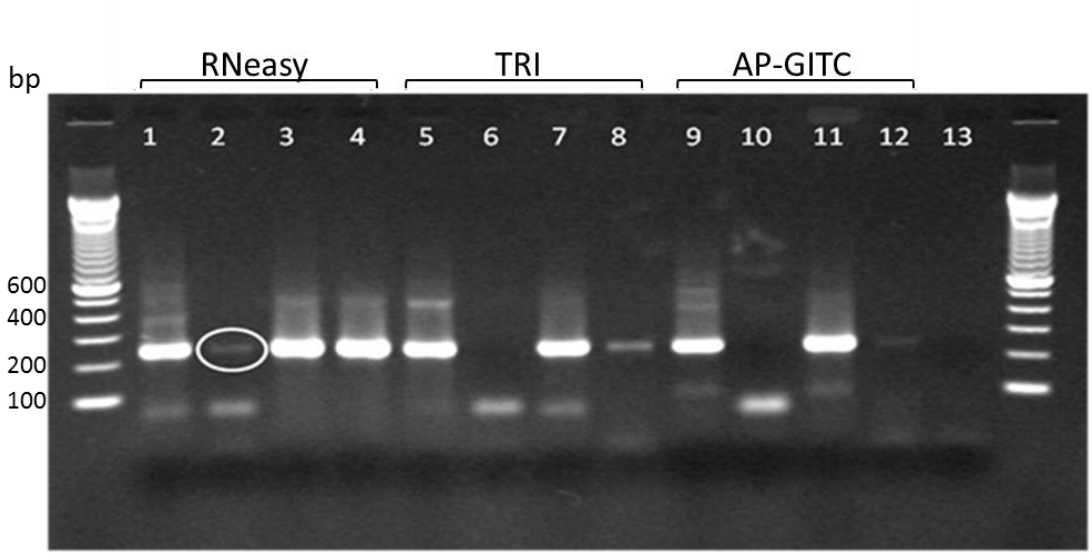


Figure 4.3: RT-qPCR of Reverse Transcription (RT) controls from the three RNA extraction methods. RNA (both with and without DNase treatment) was reverse transcribed and the resulting cDNA PCR amplified (with VapC toxin primers), and run on a 1 % SB agarose gel (stained with 0.5 µg/ml EtBr). Lanes: 1-4 = RNeasy, 5-8 = TRI Reagent, 9-12 = modified AP-GITC, and 13 = negative control. Lanes: 1,5,9 = +DNase, +RT; 2,6,10 = +DNase, -RT; 3,7,11 = -DNase, +RT; and 4,8,12 = -DNase, -RT. Amplification of the PIN toxin product (231 bp) was observed for all RNA extraction methods when RT was included in the cDNA preparations. Amplification of carry over DNA was observed for all -RT control reactions where the RNA was not DNase treated (lanes 4,8,12). The circled amplification product in lane 2 shows RNeasy was the only extraction method containing residual DNA after DNase treatment.

RNA extracted using the modified AP-GITC method contained far less contaminating DNA compared to the RNeasy extracted RNA (as visualised by the intensity of bands in the -DNase and -RT controls in Figure 4.3 (lanes 4 and 12)), which was able to be removed by DNase treatment prior to cDNA synthesis (Figure 4.3 lane 10). RNA extracted using the TRI Reagent method also contained less contaminating DNA than RNeasy extracted RNA, as a product of medium intensity was observed in the -DNase and -RT control (Figure 4.3 lane 8), which was also able to be removed by DNase treatment (Figure 4.3 lane 6).

BCP was used in place of chloroform in the modified AP-GITC method as it is less toxic and less volatile than chloroform (Chomczynski & Mackey 1995), and possibly resulted in the removal of slightly more chromosomal DNA from the RNA (observations from our laboratory and Chomczynski and Mackey (1995)). This is likely due to the ability of BCP to produce a more compact interphase than chloroform, so therefore less residual DNA from the interphase is carried over during the phase separation step.

During the course of this research I observed better performance from the RQ1 RNase-free DNase when the reactions were shaken during incubations rather than kept still.

4.2.3 Optimisation of RT-qPCR

RT-qPCR experiments were optimised on *M. smegmatis* cDNA using either SYBR Green I or SYTO 82 as the dsDNA binding dye. While both dyes produced satisfactory amplification plots and melt curves, and no obvious differences were observed in primer dimer formation across the genes being studied, the range of amplification efficiencies for genes of interest (GOI) and reference genes varied slightly more with the SYBR Green I chemistry. The greater range observed in efficiencies between our reference genes and GOI, as well as the documented inhibition of PCR and preferential binding of SYBR Green I to GC rich sequences, is the reason Syto 82 was deemed a more appropriate dye for the quantification of gene expression levels in this study even though the amplification efficiencies were slightly lower. A 'homemade' dye master mix was used over a ready made product to reduce the costs involved with the assays, and keep the reaction set up flexible so reagent concentrations could be varied if needed. Interestingly, it was noticed that the SYTO 82 colour reduces over time (even when stored at -20 °C), and possibly caused a reduction in the levels of fluorescence detected. It was for this reason that batches of master mix were stored at -20 °C without the dye, and the dye was added immediately prior to use.

RT-qPCR products were between 150 and 250bp, with melting temperatures ranging from 89.7 to 93 °C using SYTO 82 chemistry (see Appendix B.1). SYBR Green I reactions produced

products with melting temperatures approximately 0.7 °C higher than those using SYTO 82 dye.

Optimal template and primer concentrations yielding the lowest Cq, highest amplification efficiency and least amount of nonspecific amplification were established for amplicons by performing RT-qPCR trials. A template concentration trial was performed for the amplicon with the lowest efficiency (*vapB*), using 0.65, 0.75 and 0.85 µl cDNA per 20 µl reaction, which determined using 0.75 µl of template resulted in the greatest amplification efficiency with no formation of primer dimers. Variations of forward or reverse primer concentrations at 125, 250 and 500 nM were tested, resulting in fairly consistent amplification efficiencies and Cq's, although the 125 nM concentrations produced samples with little primer dimer, and the least baseline errors compared to the higher concentrations when analysed in LinRegPCR.

Amplification of primer dimers during RT-qPCR may mask the actual results when using a dsDNA binding dye. Melt curves for RT-qPCRs were analysed to check for contamination from primer dimers or other non-specific amplification products in the samples. The formation of primer dimers were avoided where possible with good primer design, but were apparent in the melt curves and gel electrophoresis for some of the primer sets used, especially in the no-template controls and reactions with limited/low abundance target. Accurate quantification of the target dsDNA product was ensured by eliminating fluorescence induced by primer dimers or unspecific minor products. This was achieved by acquiring the fluorescence data at 80 °C instead of during the 68 °C extension phase, as most primer dimers have completely denatured at this elevated temperature while the target products are still intact (Klein *et al.* 2000; Pfaffl 2004; Mehndiratta *et al.* 2008).

The amplification efficiencies of the GOI were low (approximately 1.6 and 1.7 for *vapB* and *vapC* respectively), as is often the case when detecting low abundance mRNA targets in a GC-rich organism such as *M. tuberculosis* (Grassi *et al.* 2006), and varied. Because *M. tuberculosis* is GC-rich, strong RNA secondary structures are often created which can prevent full-length RT-PCR products due to the RT enzyme stopping or falling off the template when it hits loop structures (Mathews *et al.* 1999). The high GC content (66.6 % and 60.79 % for

MSMEG_1283 (*vapB*) and MSMEG_1284 (*vapC*) respectively) and low free energy (ΔG) values predicted for the GOI RNA (-50.49 kcal/mol and -78.32 kcal/mol for MSMEG_1283 and MSMEG_1284 respectively, calculated in mfold (Zuker 2003) at 55 °C, the same temperature experienced in the RT-PCR denaturation/annealing step), suggests RNA secondary structure may be limiting the RT-PCR transcription of these already lowly expressed GOI transcripts. Due to the variation in efficiencies between amplicons, the efficiency corrected analysis computations (Equation 1 or Equation 2 in Section 4.1.1.2) were used to more accurately calculate the relative levels of expression. New primers were designed in an attempt to improve the resulting PCR efficiencies, but the difference in efficiencies was negligable. Of the methods reported to improve PCR efficiency and specificity (using additives such as DMSO, formamide and ammonium sulphate among others (Kovarova & Draber 2000; Grassi *et al.* 2006)), formamide was trialled and found to be somewhat successful in improving the efficiency of the *vapB* amplicon. Whilst a slight improvement was made for the *vapB* amplicon, the use of formamide was discontinued because the resulting change in efficiencies was not universal, which skewed the change in expression value relative to the *vapC* amplicon, making it even more difficult to detect a 'real' change in expression. Increasing the concentration of DNA polymerase in the reactions to 50 U/ml from 36 U/ml was also unsuccessful in improving the efficiencies of the amplicons.

4.2.4 Selection of a Suitable Housekeeping Gene for Relative Quantification

Because intersample variation affects data analysis, expression results needed to be normalised to an endogenous housekeeping gene whose expression remained unchanged under the experimental conditions tested. Candidate housekeeping genes were selected and tested for stability of expression under some or all of the following experimental conditions: antibiotic (rifampicin, streptomycin, tetracycline), oxidative (H_2O_2), temperature, or age/nutritional stress; other stressors (methylglyoxal, 2,4-Dinitrophenol (DNP), clotrimazole, UV); and growth conditions (planktonic v biofilm) or between strains (deletion and overexpression strains). The selection of housekeeping genes was initially based on those most commonly used in the literature such as 16S rRNA which is ubiquitously expressed in

cells and tissues, and whose transcription is generally resistant to experimental conditions (Thellin *et al.* 1999; Bustin 2000; Suzuki *et al.* 2000), and then on those which had been recorded in the literature as suitable internal controls (*ftsZ* (Delogu *et al.* 2006), *rpoB* (Gutierrez *et al.* 2005), *sigA* (Geiman *et al.* 2006), were identified from microarray studies as being similarly expressed in the conditions tested (*oxcA* (Ojha & Hatfull 2007)), or were trialled based on work done previously in the lab (unpublished) (guanylate kinase, Rv0930, *rpsB, sigH and metC*).

16S was the most stably expressed transcript under all of the experimental conditions tested, however because *M. smegmatis* posseses two 16S gene copies, it amplified much earlier than the lowly expressed GOI (approximately 15 - 20 cycles earlier). The remainder of HK genes tested were either poorly expressed (and therefore amplified after the GOI), or showed variable levels of expression under the spectrum of experimental conditions tested (see Appendix B.2). In RT-qPCR, it is important to have the reference gene amplifying within the same range as the GOI, so that both genes are experiencing the same conditions and RT-qPCR kinetics during amplification (Pfaffl 2004). To achieve this, 16S cDNA was diluted 1:99 to enable amplification within a similar range as the GOI. A five-fold dilution series of cDNA was run for each amplicon as a control, to check for inhibition in the RT-qPCR and ensure amplification was in the linear range. Dilutions of cDNA produced fluorescent curves with equally distant quantification cycles (Cq's) with the same slope (Figure 4.4), and approximately equal amplification efficiencies. This indicates that the reactions were not being inhibited, and therefore that diluting the 16S cDNA (or not diluting the GOI cDNA) would not distort the amplification efficiencies and gene expression values.

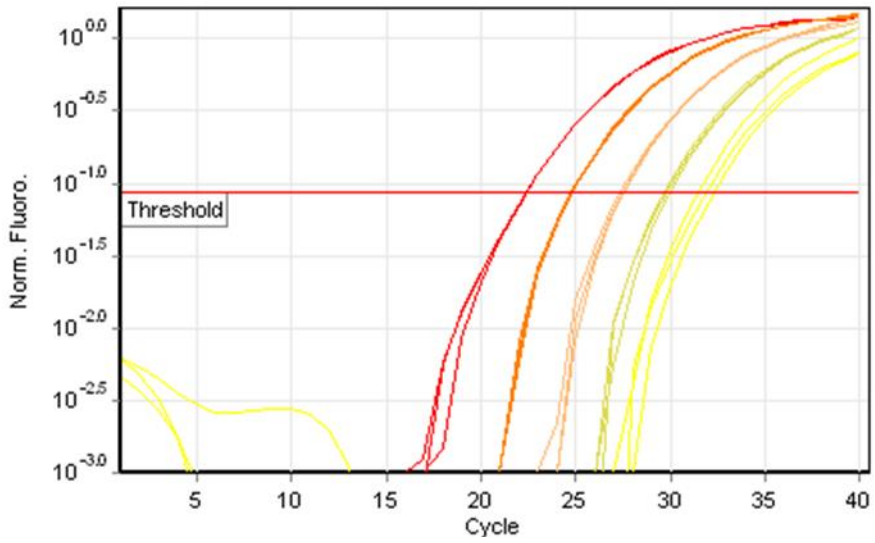


Figure 4.4: Representative real-time PCR plots following amplification of an *M. smegmatis* cDNA dilution series. Each sample was subjected to five-fold serial dilution (from red down to yellow) and then analyzed by real-time PCR in triplicate. Graph shows fluorescence intensity as a function of the number of amplification cycles. Primers were specific for the vapC gene.

As with many of the so called housekeeping genes used in the literature (Vandesompele *et al.* 2002; Liu & Slininger 2007), the majority of candidate housekeeping genes tested here were inconsistently expressed under the environmental stimuli, and were therefore deemed unsuitable as internal controls. The recommendation that a panel of housekeeping genes belonging to different functional and abundance classes should be used for accurate normalisation (Vandesompele *et al.* 2002) is valid, but under the conditions tested here, only the expression of one housekeeping gene (16S) remained constant, so using a panel of suitable genes was not possible. While the use of multiple housekeeping genes is likely to provide statistically more significant results in RT-qPCR (especially for lowly expressed transcripts) (Vandesompele *et al.* 2002), the use of a single gene for normalisation in RT-qPCR is an acceptable approach, so long as it has been validated (Bustin 2000; Suzuki *et al.* 2000; Erickson *et al.* 2007). 16S ribosomal ribonucleic acid (rRNA) genes (*rrs* genes) have frequently been rejected as normalisers by some researchers due to their high transcript levels compared to the GOI (sometimes 100 fold higher than the GOI), even when their transcription remained constant (Frost & Nilsen 2003; Robinson *et al.* 2007; Ritz *et al.* 2009). This difference in the kinetic interactions during qPCR between the highly expressed *rrs* genes and (often) lowly expressed GOI can be avoided however, by diluting the cDNA used for the 16S sample to enable its amplification within the same range as the GOI, as was performed here. Because

just a single normalising housekeeping gene was used here, any small resulting changes in GOI expression levels (fold changes close to the often cited two-fold cut-off for differential expression (Yang *et al.* 2002)) would need to be treated with caution to account for any unperceivable differences in expression that may be occurring for the single housekeeping gene.

4.2.5 VapBC Gene Expression Levels in *M. smegmatis*

Initial experiments used the *M. smegmatis* mc^24517 overexpression strain containing VapB (pYUBMS1283) or VapBC (pYUBMS1283/4) constructs, to confirm the RT-qPCR assay could detect the *vapB* and *vapC* transcripts, as well as the change in expression levels resulting from their overexpression. These experiments were performed before the RT-qPCR assay was fully optimised so whilst the exact normalised expression levels of *vapB* and *vapC* can not be read from the data, the overall picture of expression can be obtained from the cycle numbers where fluorescence levels are detected above background (represented as Cq values). In a typical reaction, a higher starting template concentration will result in an earlier or lower Cq, as relatively fewer amplification cycles are needed for the amplification product to rise above the background fluorescence signal. Table 4.2 shows how equal amounts of the *vapB* amplicon are detected for both the VapB and VapBC overexpression strains, while the expected increase in *vapC* for the VapBC overexpression strain is represented by the lower Cq value of 16.16, compared to 29.52 for the VapB only overexpression strain.

Table 4.2: Cq values of *vapB* and *vapC* amplicons in the VapB and VapBC overexpression strains.

Overexpression strain	Cq *vapB*	Cq *vapC*
VapB (pYUBMS1283)	14.37	29.52
VapBC (pYUBMS1283/4)	14.46	16.16

As well as the difficulties identifying a reliable housekeeping gene for robust data analysis, interpretation of results was also complicated by the use of different analysis methods and LinRegPCR versions. Initially, analysis was performed using LinRegPCR version 7.4 and

relative expression levels calculated using Equation 1 (Section 4.1.1.2), but later experiments used the improved LinRegPCR version 11.1 which included the following features: estimation of baseline fluorescence, different approaches for determining the window-of-linearity and fluorescence threshold, definition of sample groups per amplicon, calculation of the mean PCR efficiency per amplicon group, calculation of starting concentrations based on mean efficiency and individual Ct values, and reports on data quality for each sample, and Equation 4 (Section 4.1.1.2) to determine relative levels of expression. The biggest improvement with the later versions of LinRegPCR was the change to using the mean PCR efficiency (per amplicon group) in determining initial fluorescence, as using the efficiency per sample is too variable to provide robust, reproducible results (Cikos *et al.* 2007; Karlen *et al.* 2007).

Given the very low levels of expression, the low amplification efficiencies, and the use of a single housekeeping gene (all of which may result in inaccurate normalisation, masking any subtle variations in expression levels), any small resulting differences in expression would need to be carefully examined to ensure they were real, and not as a result of compensation during analysis or statistical variation. Many of the 'stress' conditions which were tested for effects on expression of *vapB* or *vapC*, occurred as a result of searching for the *vapC* phenotype which was unknown at the time. As it was, no significant and consistent changes in the expression of *vapB* or *vapC* were observed under any of the conditions tested which included: antibiotic (rifampicin, streptomycin, tetracycline), oxidative (H_2O_2), temperature, or age/nutritional stress; other stressors (methylglyoxal, 2,4-Dinitrophenol (DNP), clotrimazole, UV); and growth conditions (planktonic v biofilm) or between strains (deletion and overexpression strains) (see Table 4.3). This is demonstrated by the stress experiment using wt *M. smegmatis* shaken in LB, where 1 mM DNP is added for the final 15 minutes incubation period (Figure 4.5). Here, the levels of *vapB* are greater than those of *vapC* both with and without DNP present in the media (approximately five and three-fold respectively), but in a replicate experiment, *vapB* and *vapC* levels were determined to be equal, and the increase in expression observed here with DNP (approximately 2.8 fold for *vapB* and 1.6 fold for *vapC*) was reversed (data not shown). This is representative of the other stress experiments in general, where: approximately equal levels of *vapB* and *vapC* were expressed, and any

changes in *vapB* or *vapC* expression observed due to conditions of stress were inconsistent, and unable to be reliably repeated in replicate experiments.

Table 4.3: Conditions tested in vapB and vapC RT-qPCR experiments.

Type	Treatment	Concentration	Duration
Growth	7H9 v bhi	-	24 hr, 3 d, 5 d
	Sautons +/- RPMI 1640 Medium (GIBCO)	25 %	24 hr, 5 d, 6 d
	+/- Tween 80	0.05 %	1 hr, 18 hr
	biofilm v planktonic	-	7 d
Temperature	0 °C or 42 °C	-	30 min
	48 °C	-	15 min, 90 min
Antibiotic	+/- Tween 80, +/- streptomycin	0.05%, 2 µg/µl	6 hr, 24 hr
	streptomycin	2 µg/µl	30 min, 1 hr
	streptomycin for 6 hr, followed by rifampicin for 30 min	10 µg/ml, 10 µg/ml	6 hr, 30 min
	streptomycin in planktonic v biofilm	5 µg/ml	1 hr
	rifampicin	1.25 µg/ml	40 hr
	rifampicin	8 µg/ml	30 min
Other	substitution of supernatant from 4, 5 or 6 d cultures with 'aged' supernatant (from 1 month old cultures left at RT)	-	10 min
	+/- substitution of supernatant from 3 d culture with 'aged' supernatant (from 1 month old cultures left at RT), +/- streptomycin	-, 8 µg/ml	30 min
	EtOH, tetracycline	50 %, 25 µg/ml	35 min
	H_2O_2	17.64 µM	30 min, 1 hr
	DNP	1 mM	15 min, 6 hr
	Clotrimazole	24 µg/ml	15 min
	DNP, clotrimazole	1 mM, 24 µg/ml	15 min

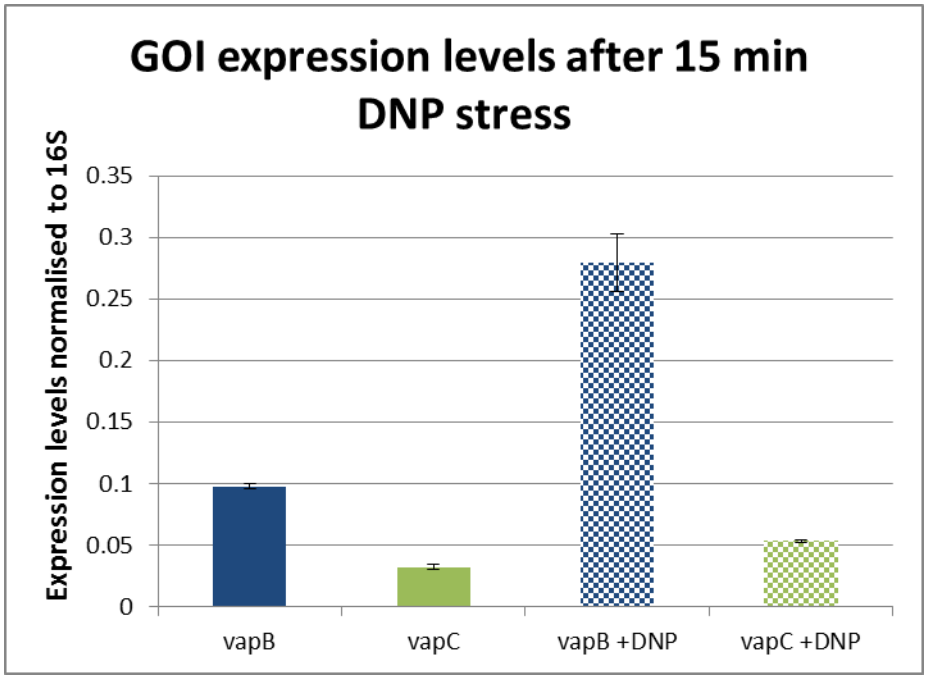


Figure 4.5: Relative *M. smegmatis* wt *vapB* and *vapC* expression levels in response to 15 min DNP stress. RNA isolated from wt *M. smegmatis* cultures inubated +/- 1 mM DNP for 15 minutes was subjected to RT-qPCR to determine the expression levels of *vapB* and *vapC* transcripts. Expression of *vapB* is greater than that of *vapC*, and expression of both appear to increase marginally after DNP treatment by approximately 2.8 and 1.65 fold respectively. Expression levels (normalised to 16S) were determined using LinRegPCR version 11.1, and are expressed in arbitrary fluorescence units.

The fact that no reproducible large increase in *vapC* expression was detected under any of the stress conditions tested here intimates the importance of tightly regulating this toxic protein, and suggests that regulation of toxic VapC may be occurring at the proteomic level, and not at a transcriptional level. Strengthening the proposed requisite for tight regulation of this toxin, were the unfruitful attempts by others in the lab to express VapC protein by itself (without its associated antitoxin to neutralise the toxicity), repeatedly resulting in the occurance of a stop codon (McKenzie 2011), and the absence of VapC protein despite the presence of mRNA transcript in the *vapB* deletion strain (Robson *et al.* 2009). The lack of correlation between proteomic and transcriptomic analyses has been observed elsewhere, where it was postulated to occur as a result of differring protein turnover and/or RNA stability (Scherl *et al.* 2005; Ritz *et al.* 2009).

The negative RT-qPCR result observed here was the impetus for looking at modelling changes in protein levels compared with RNA levels to explain how the production of VapC could be

being regulated (discussed in Chapter Six), and why we now hypothesise it is the rate of translation which is important in regulating the level of VapC toxin, not the difference in levels of *vapB* and *vapC* transcription themselves.

4.3 Conclusions

This research presents an optimised RNA isolation and RT-qPCR method suitable for the quantification of the *vapB* and *vapC* target genes, and determined there was no differential expression of *vapC* under the conditions tested.

As pointed out in the above sections, there were many issues identified with the RT-qPCR process as it was in 2008 when this data was collected. I am aware these (and other) issues have been highlighted by others since, and vast developments have been made within the field of RT-qPCR. I support the changes proposed in the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines (Bustin *et al.* 2009), which addresses these issues and will result in more reliable RT-qPCR results, presented in a standard and easily interpreted format. As with the vast array of subsequent reports which found many HK genes were unacceptable and often produced unreliable results, this research also found the majority of HK genes tested here were not suitable (so the results were rejected), and any expression change ratios obtained with the single 16S HK gene were treated with caution. Even if we were to go back and implement the recommended MIQE controls (specifically, using a panel of HK genes for normalisation), we are confident this wouldn't change our finding that no consistent change in the mRNA expression levels of *vapB* and *vapC* occurs, especially given since this observation was made there have been no entries in the TB Database showing any confirmed significant changes. This negative result is very important, as together with the modelling data it supports our current hypothesis that it is not the transcriptional differences between *vapB* and *vapC* which are important in the regulation of VapC, but the overall rate of VapBC translation itself.

Chapter Five:

Mycobacterium tuberculosis VapBC Characterisation

5.1 Introduction

This chapter focuses on characterisation of the VapBC proteins from *M. tuberculosis*, in particular, the specificity and ribonuclease activity of the VapC proteins. The biological role of VapBC proteins in *M. tuberculosis* is poorly understood, but their link to regulating cell growth in response to environmental stress is an inviting intimation that they may play an important part in the pathogenesis of its disease.

5.1.1 *M. tuberculosis* VapC Ribonuclease Activity

Recently, the structures have been determined for two *M. tuberculosis* VapBC complexes, VapBC-5 (Rv0626/0627) (Miallau *et al.* 2009) and Rv0301/0300 (Min *et al.* 2012). The toxins from these structures are consistent with other PIN-domain proteins, forming an α/β/α sandwich consisting of five parallel beta strands flanked by alpha helices, and containing four conserved acidic residues (Asp26, Glu57, Asp115 and Asp135 for VapC-5, and Asp9, Glu43, Asp99, and Asp117 for Rv0301). This conformation forces the conserved acidic residues together which coordinate an Mg^{2+} ion and form the putative RNase active site (Miallau *et al.* 2009; Arcus *et al.* 2011; Min *et al.* 2012). Both of these *M. tuberculosis* VapBC complexes

show overall structural similarity to VapBC complexes from other organisms such as FitAB from *N. gonorrhoeae* (Figure 5.1), *Shigella flexneri* VapBC, and *Rickettsia felis* VapBC2 (Min *et al.* 2012), illustrating tight binding between VapB and VapC as a tetramer of VapBC heterodimers. While the *M. tuberculosis* toxin proteins show limited conformational differences to their closest structural homologue FitB, their antitoxin partners have notably different conformations (Miallau *et al.* 2009; Min *et al.* 2012).

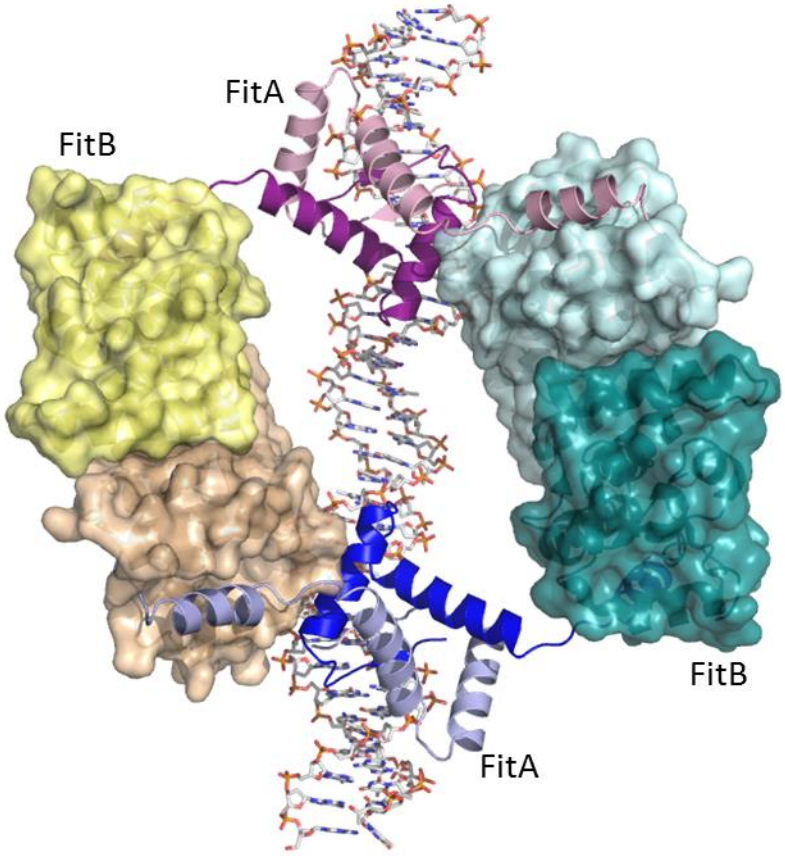


Figure 5.1. Tetramer of FitAB heterodimers from *N. gonorrhoeae* bound to IR from its promoter DNA. Pink and mauve, and dark and light blue chains denote two FitA dimers each binding to one half of the DNA inverted repeat. Surface diagrams depict FitB dimers. Each FitA monomer is bound to a FitB monomer, forming heterodimers. Figure from Arcus *et al*. (2011).

Ribonuclease activity has been demonstrated for toxins from many of the other TA systems, few VapC proteins have been fully characterised. Those which have been characterised also demonstrate ribonuclease activity (Daines *et al.* 2007; Hurley & Woychik 2009; Ramage *et al.* 2009; McKenzie *et al.* 2012b). Of the two *M. tuberculosis* VapC proteins characterised thus

far, Mg^{2+} dependent ribonuclease activity has been demonstrated for VapC-5 (but its catalytic mechanism or sequence specificity was not determined) (Miallau *et al.* 2009), while only suggested for Rv0301 from its structural characterisation (Min *et al.* 2012). During the course of my PhD, a possible biochemical function of VapC from *M. smegmatis* was determined, as was confirmation of ribonuclease activity for two further *M. tuberculosis* VapC proteins, Rv0065 and Rv0617, which were cloned in this thesis and published in Ahidjo *et al*. (2011). Rv0065 and Rv0617 are Mg^{2+} dependent, sequence-selective ribonucleases which target ssRNA rather than dsRNA, ssDNA or dsDNA (and unlike VapC from *M. smegmatis*, they do not show ss or dsRNA binding), whose activity is selectively inhibited by their cognate antitoxins (Rv0064 and Rv0616 respectively) (Ahidjo *et al.* 2011; McKenzie *et al.* 2012a). Both *M. tuberculosis* VapC proteins were found to have the same sequence specificity targeting GC-rich 4mers (i.e., GGCG, GCCG and GGGC) but with a certain level of redundancy, suggesting that the biologically relevant target may combine RNA sequence and secondary structure (McKenzie *et al.* 2012a) and raising the possibility that both VapC toxins could target different secondary structural motifs. VapC from enteric bacteria have recently been shown to specifically cleave initiator $tRNA^{fMet}$ between the anticodon stem and loop (Winther & Gerdes 2011). However experiments with mycobacterial VapCs ($VapC_{MS1284}$, $VapC_{Rv0065}$ or $VapC_{Rv0617}$) in our lab have not displayed any such activity, despite the $VapC_{MS1284}$ cut site being present in $tRNA^{fMet}$, strengthening the idea that the optimal substrate combines both sequence and RNA secondary structure (McKenzie *et al.* 2012a; McKenzie *et al.* 2012b).

5.1.2 Protein Cloning, Expression, Purification & Crystallisation

5.1.2.1 Protein Expression in *M. smegmatis*

Previous work in our lab showed expressing mycobacterial proteins in a mycobacterial host could overcome the problem of insoluble protein expression which occurred when expressed in *E. coli* (McKenzie 2011). Expression of the *M. smegmatis* VapBC proteins in *M. smegmatis* also ensured recognition of the one base pair overlap between *vapB* and *vapC* genes in the operon enabling expression of VapC in complex with VapB. This overlap was not recognised when expressed in *E. coli,* instead resulting in insoluble VapB protein only (McKenzie 2011).

The non-pathogenic mycobacterial species *M. smegmatis* has a relatively fast doubling time of approximately three hours (Snapper *et al.* 1990), as well as mycobacterial chaperones to aid protein folding, making it an ideal host for expression of mycobacterial proteins. The *M. smegmatis* overexpression strain mc^24517 with mycobacterial expression vector pYUB28b (allowing a C–terminal His–tag) can be cultured in autoinduction media using methods established for *E. coli,* which reduces the cost of culturing by removing the need for albumin dextrose catalase supplement (ADC) enrichment and optical density monitoring/induction traditionally used for mycobacterial cultures (Bashiri *et al.* 2007). Autoinduction media was also chosen over growing in standard media and inducing expression with IPTG, because *M. smegmatis* cultures grew well under these conditions but did not produce VapBC protein whereas they did when autoinduction media was used (McKenzie 2011). Autoinduction media functions due to the bacteria being able to utilise a variety of carbon sources (including latose), but initially glucose. While glucose is in the seeder media, it is limited in the autoinduction media so once glucose from the seeder media is exhausted, the bacteria switches to using lactose. Lactose enters the cell and induces expression of T7 polymerase, and therefore the production of VapBC protein.

Soluble expression of VapC proteins is difficult due to their toxicity (Mattison *et al.* 2006; Miallau *et al.* 2009; Ramage *et al.* 2009). Previous attempts in the lab to express *M. smegmatis* VapB and VapBC in *M. smegmatis* were fruitful, however expression of the toxic VapC protein by itself (without its associated antitoxin to neutralise the toxicity) repeatedly resulted in a two base pair insertion causing a stop codon downstream, and no protein expression (McKenzie 2011). This insertion is likely due to basal expression of VapC from the leaky T7 promoter prior to induction, causing the bacteria to mutate the plasmid before expression is induced. While expression systems are available in *E.coli* which supress basal expression from leaky promoters therefore enabling expression of toxic proteins (Saida *et al.* 2006), no such system exists for expression in mycobacteria.

Expression of VapC as a maltose binding protein (MBP) fusion in *E. coli* has been shown to increase the solubility of VapC, and expression of *M. tuberculosis* VapC proteins as MBP fusions has been successful in other labs (Ramage *et al.* 2009), however expression of the *M.*

smegmatis MBP-VapC fusion was not successful in our lab as it formed a soluble aggregate which could not be refolded (McKenzie 2011).

Soluble *M. smegmatis* VapC can be expressed in *M. smegmatis* when in complex with VapB and various strategies have been employed previously to disrupt the interaction between the two proteins and produce functional VapC. Disruption with 8 M urea, 2 M NaCl and buffers ranging in pH from 3.5 to 7 were all unsuccessful, and while 6 M guanidium and 0.1 % and 0.5 % SDS did disrupt the VapBC complex, they presumably denatured the VapC which could not be successfully refolded (McKenzie 2011). It was proposed that the strong interaction between VapB and VapC in the VapBC complex from *M. smegmatis* is the same as that for FitAB from *Neisseria gonorrhoeae* where the hydrophobic surfaces of FitA and FitB proteins associate tightly (Wilbur *et al.* 2005; Mattison *et al.* 2006). The strong interaction in the *M. smegmatis* VapBC complex was disrupted by taking advantage of the antitoxin being more susceptible to proteolytic degradation than the toxin, a common feature of TA systems (Buts *et al.* 2005; Gerdes *et al.* 2005). The protease trypsin was used to digest away VapB, leaving behind functional VapC, and trypsin was removed by anion exchange chromatography (McKenzie 2011). Dynamic light scattering determined the molecular weight of the purified *M. smegmatis* VapBC complex to be 152 kDa, suggestive of a tetramer of VapBC heterodimers bound to DNA (McKenzie 2011). This is consistent with the FitAB complex which has a molecular weight of 121 kDa and forms a tetramer of FitAB heterodimers when bound to an inverted repeat in the *fitAB* promoter (Mattison *et al.* 2006).

5.1.3 Objectives

Given the different systems and contradictions presented in the literature, I attempted to identify all of the unannotated *M. tuberculosis* VapB proteins, then clone, express and purify all 47 VapBC proteins from *M. tuberculosis* into *M. smegmatis*. Due to the ribonuclease activity described previously for VapC proteins from other organisms, I then set out to determine the specificity and activity of a number of the purified *M. tuberculosis* VapC proteins.

The biochemical function of VapC from *M. smegmatis* and activity of VapC proteins from *M. tuberculosis* was unknown at the beginning of my research.

5.2 Results and Discussion

5.2.1 Cloning of *M. tuberculosis* VapBC's into *M. smegmatis*

The first step in characterising the 47 *M. tuberculosis* VapBC proteins was to locate the many unannotated antitoxins. Here, BLAST was used with the upstream region of VapC to identify VapB (a hypothetical antitoxin approximately 250 bp in size) and the VapBC overlap (see Figure 5.2). Once the antitoxins were identified, each *M. tuberculosis* VapBC operon was successfully cloned into the *E. coli* - mycobacteria shuttle vector pYUB28b between the NcoI and HindIII, BamHI or NdeI restriction sites enabling expression of protein with a C-terminal His tag. The modified pYUB1049 vector pYUB28b was used rather than the traditional *E. coli* - mycobacteria shuttle vector pYUB1049, as it allows in-frame C-terminal His tag expression without having to add additional bases to the forward primer, and does not require digestion and gel purification to remove a large (approximately 1,000 bp) multiple cloning site segment before inserting the gene of interest (Bashiri *et al.* 2010). A C-terminal tagged VapC protein was required so as both the VapB and VapC could be expressed in an operon (as VapCs are generally toxic and unable to be expressed by themselves), and the VapC could be recovered once VapB had been removed. A summary of the restriction sites used for each operon are shown in Table 5.1 below, and full gene sequences and cloning information can be found on the compact disk - Appendix D.

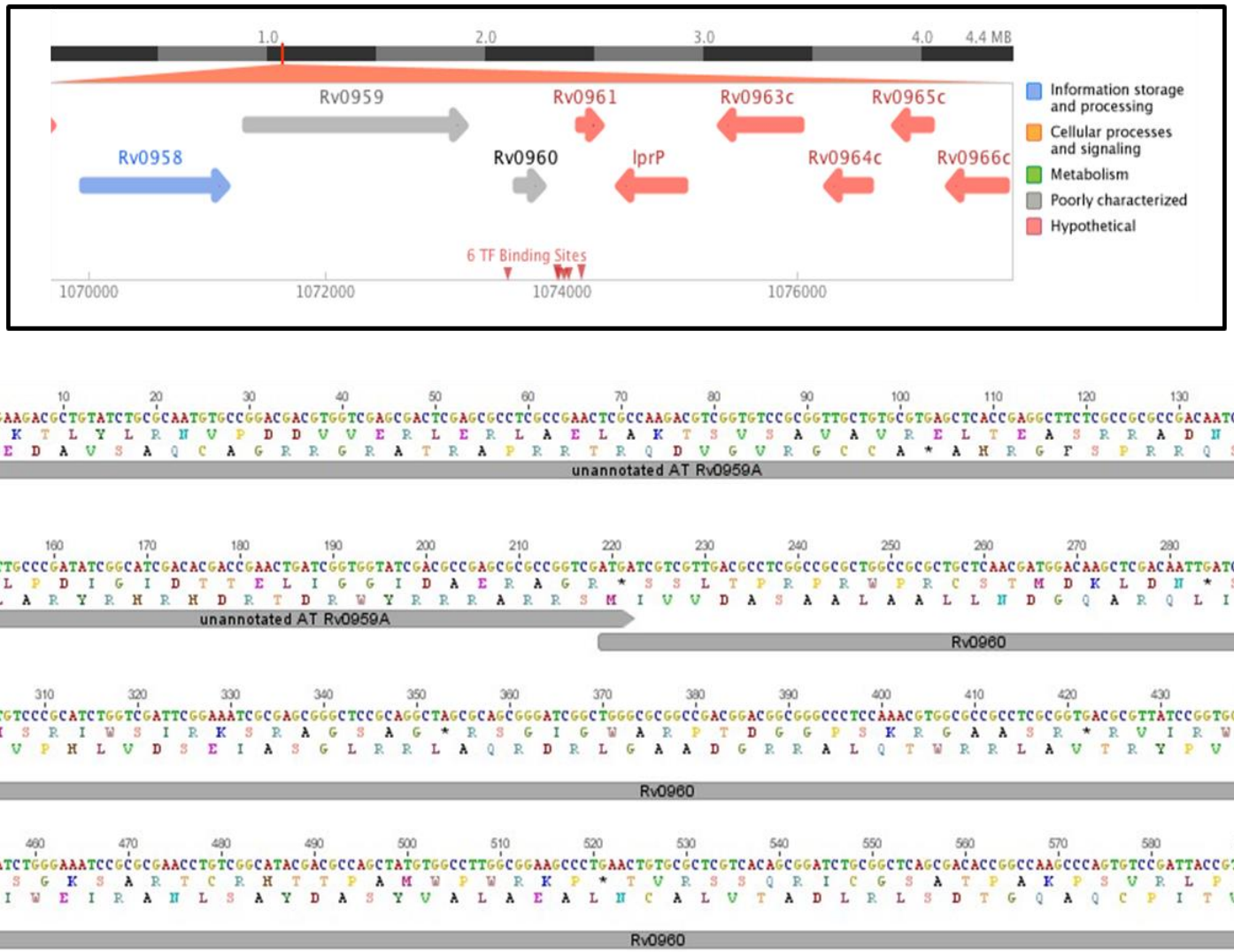


Figure 5.2: Identification of unannotated antitoxins. Top: TBDB image of the genomic region surrounding Rv0960, a VapC toxin, showing the absence of an annotated cognate antitoxin. Bottom: Nucleotide and amino acid sequences for Rv0960 and its upstream region, showing the overlapping unannotated assumed cognate antitoxin Rv0959A identified through BLAST analysis, which is translated in a different frame.

Table 5.1: Summary of restriction sites used to clone each *M. tuberculosis* VapBC operon into pYUB28b.

VapBC Operon	Position in H37Rv genome, strand direction, and size (bp)	Restriction sites included in primers (Fwd/Rev) for ligation into pYUB28b
Rv0064/0065	71589 - 71828 (+) (240 bp) / 71821 - 72222 (+) (402 bp)	NcoI/HindIII
Rv0229c	274306 - 274986 (-) (681 bp)	GeneArt*
Rv0239/0240	289104 - 289337 (+) (234 bp) / 289345 - 289782 (+) (438 bp)	NcoI/HindIII
Rv0277a/0277c	333160 - 333417 (-) (258 bp) / 332708 - 333136 (-) (429 bp)	NcoI/HindIII
Rv0300/0301	363826 - 364047 (+) (222 bp) / 364044 - 364469 (+) (426 bp)	NcoI/HindIII
Rv0549/0550	640228 - 640641 (-) (414 bp) / 640638 - 640904 (-) (267 bp)	GeneArt*
Rv0582/0581	677922 - 678329 (+) (408 bp) / 677710 - 677925 (+) (216 bp)	NcoI/HindIII
Rv0595/0596c	694839 - 695231 (-) (393 bp) / 695228 - 695485 (-) (258 bp)	NcoI/HindIII
Rv0598c/0599	697154 - 697567 (-) (414 bp) / 697564 - 697800 (-) (237 bp)	NcoI/HindIII
Rv0609/0608	703486 - 703887 (+) (402 bp) / 703244 - 703489 (+) (246 bp)	NcoI/HindIII
Rv0617/0616	711006 - 711407 (+) (402 bp) / 710782 - 711009 (+) (228 bp)	NcoI/HindIII
Rv0627/0626	718282 - 718689 (+) (408 bp) / 718025 - 718285 (+) (261 bp)	NcoI/HindIII
Rv0656/0657	752984 - 753367 (-) (384 bp) / 753462 - 753617 (-) (156 bp)	NcoI/HindIII
Rv0661c/0662c	755335 - 755772 (-) (438 bp) / 755769 - 756137 (-) (369 bp)	NcoI/HindIII
Rv0665/0664	758801 - 759139 (+) (339 bp) / 758532 - 758804 (+) (273 bp)	NcoI/HindIII
Rv0749/0748	841228 - 841656 (+) (429 bp) / 840947 - 841204 (+) (258 bp)	NcoI/HindIII
Rv0960/0959a	1073545 - 1073928 (+) (384 bp) / 1073327 - 1073548 (+) (222 bp)	NcoI/HindIII
Rv1114/1113	1239610 - 1239984 (+) (375 bp) / 1239416 - 1239613 (+) (198 bp)	NcoI/HindIII
Rv1242/1241	1384535 - 1384966 (+) (432 bp) / 1384278 - 1384538 (+) (261 bp)	NcoI/HindIII
Rv1397c/1398c	1574112 - 1574513 (-) (402 bp) / 1574510 - 1574767 (-) (258 bp)	NcoI/HindIII

Rv1561/1560	1764979 - 1765383 (+) (405 bp) / 1764755 - 1764973 (+) (219 bp)	NcoI/HindIII
Rv1720/1721c	1947030 - 1947419 (-) (390 bp) / 1947416 - 1947643 (-) (228 bp)	GeneArt*
Rv1741/1740	1967917 - 1968165 (+) (249 bp) / 1967705 - 1967917 (+) (213 bp)	NcoI/HindIII
Rv1838c/1839c	2087257 - 2087652 (-) (396 bp) / 2087649 - 2087912 (-) (264 bp)	NcoI/HindIII
Rv1953/1952	2200938 - 2201249 (+) (312 bp) / 2200726 - 2200941 (+) (216 bp)	NcoI/HindIII
Rv1962c/1962a	2204866 - 2205273 (-) (408 bp) / 2205277 - 2205549 (-) (273 bp)	GeneArt*
Rv1982c/1982a	2225413 - 2225832 (-) (420 bp) / 2225841 - 2226101 (-) (261bp)	NcoI/BamHI
Rv2010/2009	2258273 - 2258671 (+) (399 bp) / 2258030 - 2258272 (+) (243 bp)	NcoI/HindIII
Rv2103c/2104c	2364086 - 2364520 (-) (435 bp) / 2364527 - 2364781 (-) (255 bp)	NcoI/BamHI
Rv2494/2493	2808310 - 2808735 (+) (426 bp) / 2808083 - 2808304 (+) (222 bp)	NcoI/HindIII
Rv2527/2526	2851315 - 2851716 (+) (402 bp) / 2851091 - 2851318 (+) (228 bp)	NcoI/HindIII
Rv2530c/2530a	2854267 - 2854686 (-) (420 bp) / 2854683 - 2854907 (-) (225 bp)	NcoI/HindIII
Rv2546/2545	2868154 - 2868567 (+) (414 bp) / 2867783 - 2868061 (+) (279 bp)	GeneArt*
Rv2548/2547	2868860 - 2869237 (+) (378 bp) / 2868606 - 2868863 (+) (258 bp)	NcoI/NdeI
Rv2549c/2550c	2869727 - 2870122 (-) (396 bp) / 2870119 - 2870364 (-) (246 bp)	GeneArt*
Rv2596/2595	2925734 - 2926138 (+) (405 bp) / 2925492 - 2925737 (+) (246 bp)	NcoI/HindIII
Rv2602/2601	2930344 - 2930784 (+) (441 bp) / 2930070 - 2930357 (+) (288 bp)	NcoI/HindIII
Rv2757/2758c	3070170 - 3070586 (-) (417 bp) / 3070583 - 3070849 (-) (267 bp)	GeneArt** + BamHI
Rv2759c/2760c	3070875 - 3071270 (-) (396 bp) / 3071267 - 3071536 (-) (270 bp)	NcoI/HindIII
Rv2829c/2830c	3136620 - 3137012 (-) (393 bp) / 3137009 - 3137224 (-) (216 bp)	NcoI/HindIII
Rv2863/2862a	3174992 - 3175372 (+) (381 bp) / 3174801 - 3174995 (+) (195 bp)	NcoI/HindIII
Rv2872/2871	3183382 - 3183825 (+) (444 bp) / 3183138 - 3183395 (+) (258 bp)	NcoI/HindIII

Rv3320c/3321	3707642 - 3708070 (-) (429 bp) / 3708074 - 3708316 (-) (243 bp)	NcoI/BamHI
Rv3384c/3385c	3799243 - 3799635 (-) (393 bp) / 3799635 - 3799943 (-) (309 bp)	NcoI/HindIII
Rv3408/3407	3826548 - 3826958 (+) (411 bp) / 3826252 - 3826551 (+) (300 bp)	NcoI/HindIII
Rv3697c/3697a	4139805 - 4140242 (-) (438 bp) / 4140239 - 4140463 (-) (225 bp)	NcoI/HindIII

Positions in H37Rv genome relative to those published in TBDB. * Operons containing the NcoI sequence were ordered from GeneArt (with the internal NcoI site removed), and ligated into pYUB28b using NcoI and HindIII. ** Rv2757/2758 contained both NcoI and HindIII sequences so was ordered from GeneArt with the NcoI site removed, then amplified with a 3' primer containing BamHI to put the BamHI site in frame.

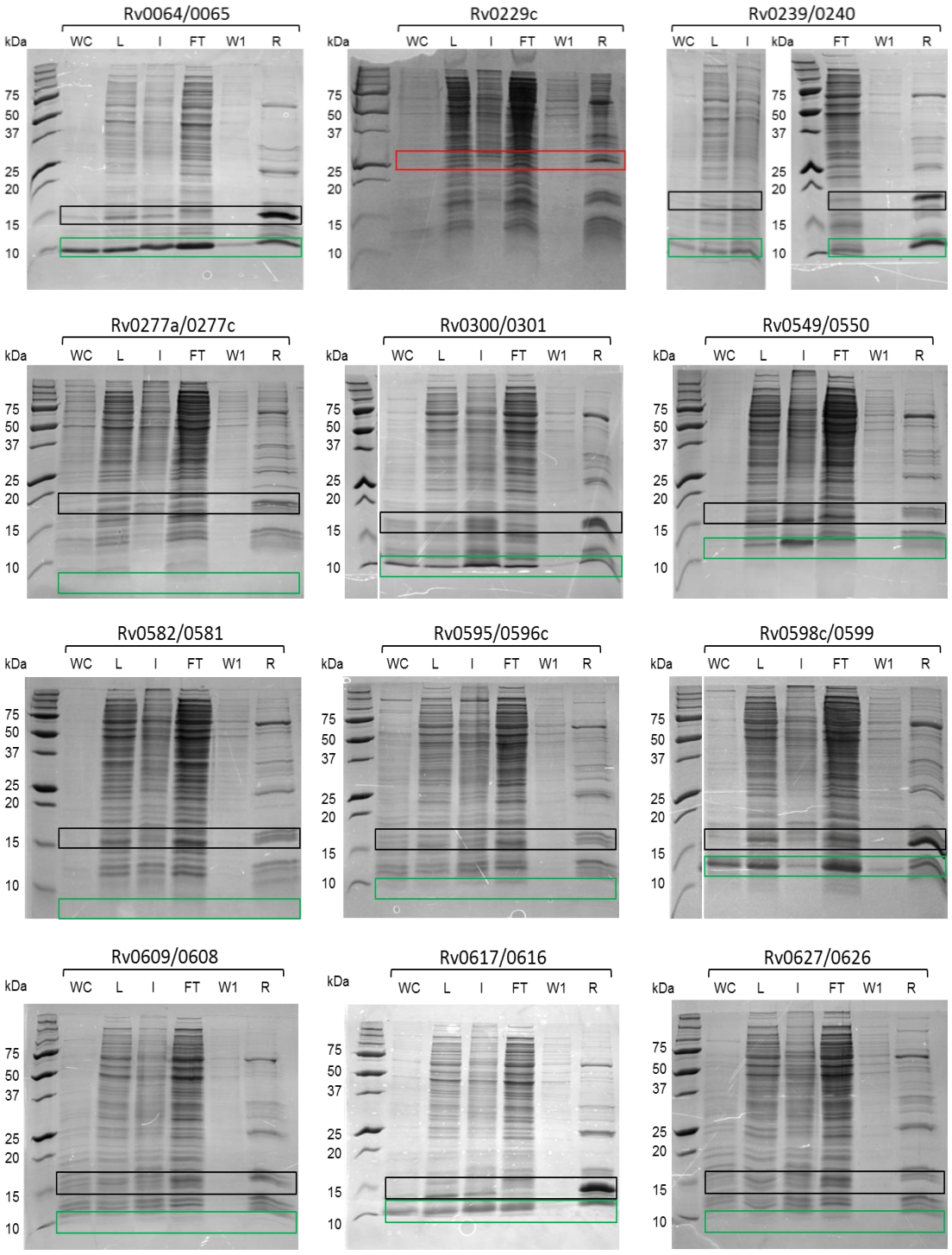
Purification of plasmids from mycobacteria is difficult due to the waxy nature of the mycobacterial cell wall which makes it hard to lyse cells whilst still retaining plasmid integrity. Therefore, ligation reactions were initially transformed into electrocompetent *E. coli* cells to facilitate plasmid purification and screening via sequencing, prior to transforming into *M. smegmatis* $mc^2$4517 cells. TOP10 cells were used rather than other *E. coli* strains due to their high transformation efficiency and the ability to select transformants using hygromycin B.

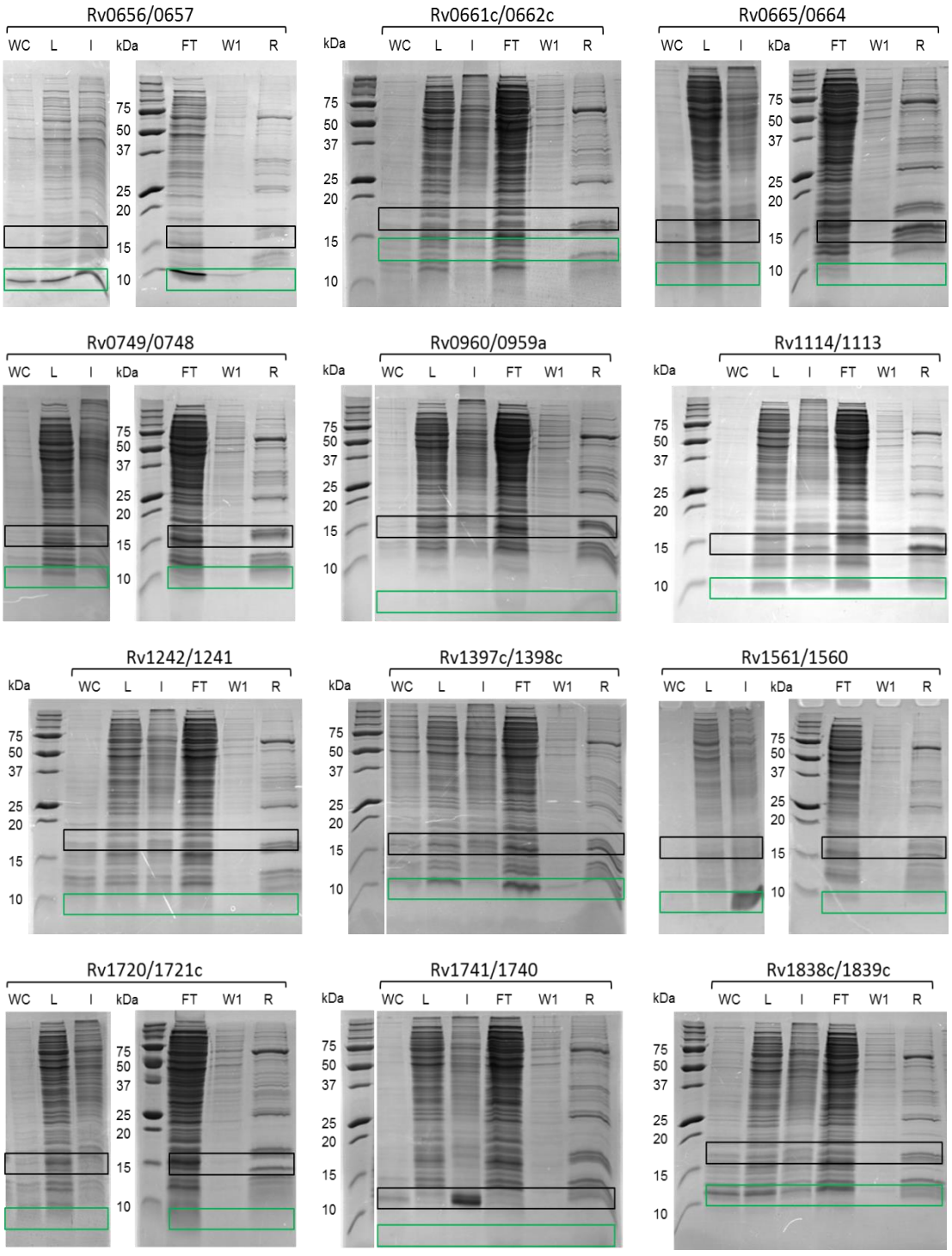
Successful insertion of the *vapBC* genes was confirmed by extracting the plasmid from positive *E. coli* transformants (identified through colony PCR screening) and sequencing with primers flanking the pYUB28b multiple cloning site (T7 forward and reverse).

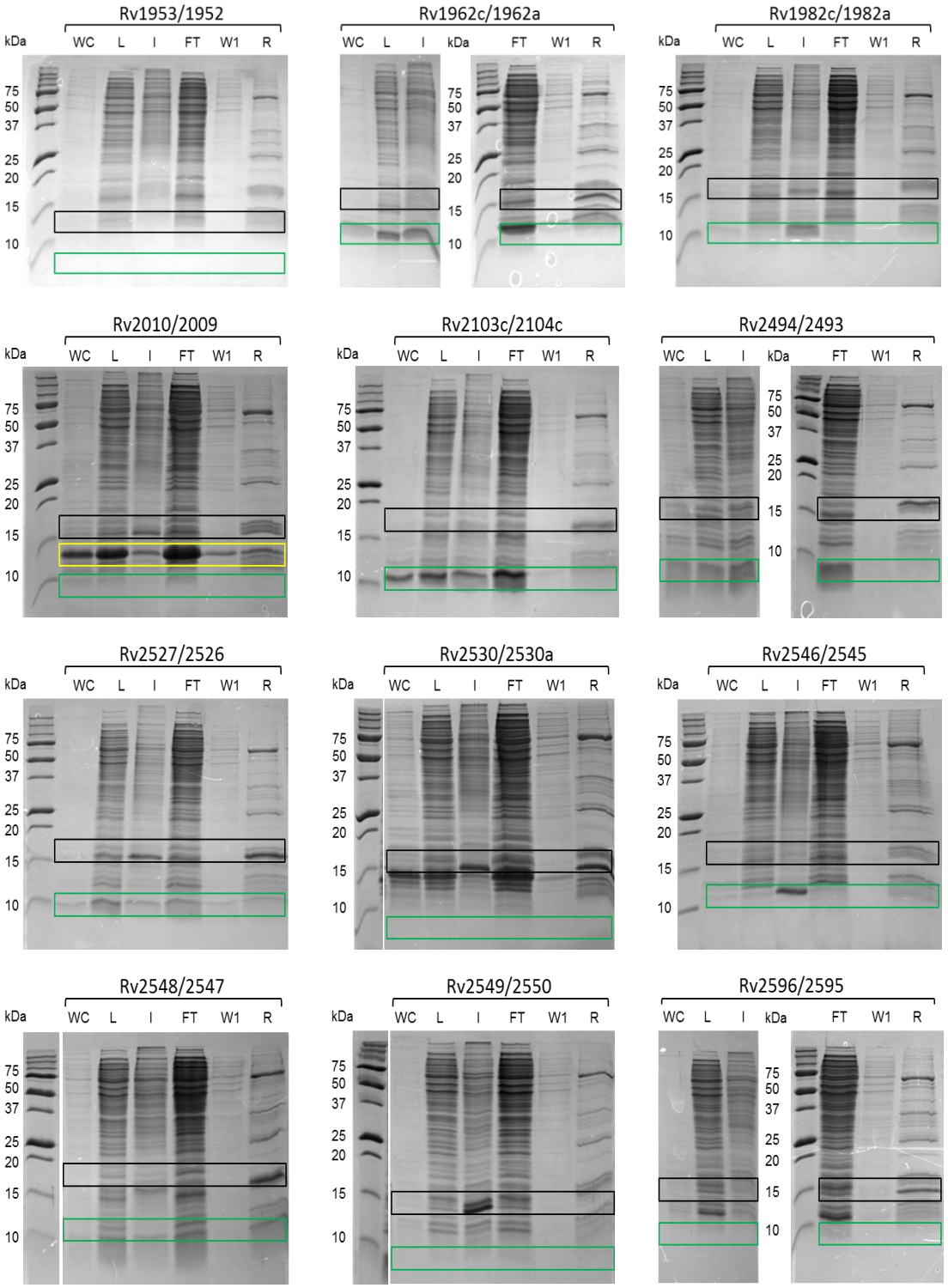
5.2.2 Expression and Purification of *M. tuberculosis* VapBC's

5.2.2.1 Small Scale Expression Trials of *M. tuberculosis* VapBC's

Once all 47 of the *M. tuberculosis* VapBC operons were successfully cloned into *M. smegmatis*, small scale protein expression tests were performed in duplicate and used to evaluate VapBC expression and his-tag binding affinity (Figure 5.3). Protein expression of *M. tuberculosis* VapBC's in *M. smegmatis* using the standard expression protocol (Section 2.3.2.11) was moderately successful, as summarised in Table 5.2.


Rv0064/0065
Rv0229c
Rv0239/0240
Rv0277a/0277c
Rv0300/0301
Rv0549/0550
Rv0582/0581
Rv0595/0596c
Rv0598c/0599
Rv0609/0608
Rv0617/0616
Rv0627/0626
kDa
WC
L
I
FT
W1
R
75
50
37
25
20
15
10


Rv0656/0657
WC L I kDa FT W1 R
75 50 37 25 20 15 10
Rv0661c/0662c
kDa WC L I FT W1 R
75 50 37 25 20 15 10
Rv0665/0664
WC L I kDa FT W1 R
75 50 37 25 20 15 10
Rv0749/0748
WC L I kDa FT W1 R
75 50 37 25 20 15 10
Rv0960/0959a
kDa WC L I FT W1 R
75 50 37 25 20 15 10
Rv1114/1113
kDa WC L I FT W1 R
75 50 37 25 20 15 10
Rv1242/1241
kDa WC L I FT W1 R
75 50 37 25 20 15 10
Rv1397c/1398c
kDa WC L I FT W1 R
75 50 37 25 20 15 10
Rv1561/1560
WC L I kDa FT W1 R
75 50 37 25 20 15 10
Rv1720/1721c
WC L I kDa FT W1 R
75 50 37 25 20 15 10
Rv1741/1740
kDa WC L I FT W1 R
75 50 37 25 20 15 10
Rv1838c/1839c
kDa WC L I FT W1 R
75 50 37 25 20 15 10


Rv1953/1952
Rv1962c/1962a
Rv1982c/1982a
Rv2010/2009
Rv2103c/2104c
Rv2494/2493
Rv2527/2526
Rv2530/2530a
Rv2546/2545
Rv2548/2547
Rv2549/2550
Rv2596/2595
kDa
WC
L
I
FT
W1
R
75
50
37
25
20
15
10

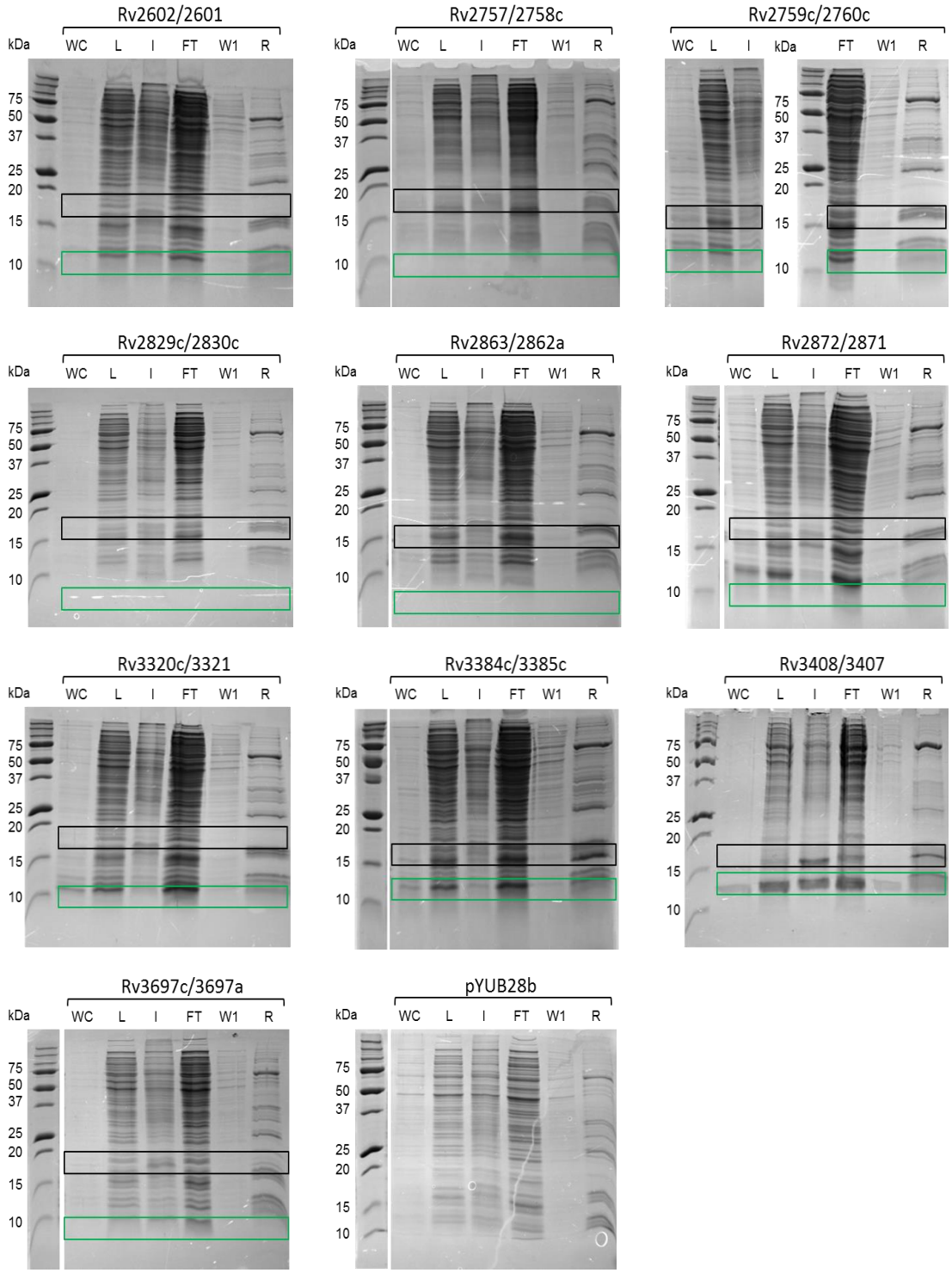


Figure 5.3: Small scale expression screens of *M. tuberculosis* VapBC (pYUB28b constructs) in *M. smegmatis*. Expected band sizes of His-tag bound and expressed VapC (black box), VapB (green box), fused VapBC (red box) and unknown protein (yellow box) are displayed. Labels: whole cell fraction (WC), soluble fraction loaded onto Ni^{2+} resin (L), insoluble fraction (I),flow through (FT), wash one (W1), Ni^{2+} resin with protein bound after wash steps (R). pYUB28b sample shows the empty vector control.

Table 5.2: Summary of small scale expression tests and his-tag binding affinity.

Operon	VapB				VapC			
		expected size (kDa)	expressed insolubly	expressed solubly		expected size (kDa)	expressed insolubly	expressed solubly
Rv0064/0065	Rv0064	8.7	✓	✓	Rv0065	15.8	x	✓
Rv0229c*	-	-	-	-	Rv0229c	26.8	x	x
Rv0239/0240	Rv0239	11.5	x	✓	Rv0240	17.9	x	✓
Rv0277a/0277c	Rv0277a	6.0	x	x	Rv0277c	17.4	x	x
Rv0300/0301	Rv0300	8.1	✓	✓	Rv0301	17.2	x	✓
Rv0549/0550	Rv0550	9.5	✓	x	Rv0549c	17.7	✓	x
Rv0582/0581	Rv0581	7.6	x	x	Rv0582	15.9	x	x
Rv0595/0596c	Rv0596c	9.7	x	x	Rv0595c	15.6	x	x
Rv0598c/0599	Rv0599	8.2	x	✓	Rv0598c	17.2	x	✓
Rv0609/0608	Rv0608	8.8	x	x	Rv0609	16	x	x
Rv0617/0616	Rv0616	8.2	✓	✓	Rv0617	15.5	x	✓
Rv0623/0624**	Rv0623	9.1	x	x	Rv0624	17.3	x	x
Rv0627/0626	Rv0626	9.5	x	x	Rv0627	15.9	x	x
Rv0656/0657	Rv0657	11.1	✓	x	Rv0656c	16.9	x	x
Rv0661c/0662c	Rv6062c	14	x	x	Rv0661	16.8	x	x
Rv0665/0664	Rv0064	9.2	x	x	Rv0665	13.1	x	✓
Rv0749/0748	Rv0748	9	x	x	Rv0749	17.4	x	x
Rv0960/0959a	Rv0959a	7.9	x	x	Rv0960	15.4	x	x
Rv1114/1113	Rv1113	6.9	✓	✓	Rv1114	15	✓	✓
Rv1242/1241	Rv1241	9.8	x	x	Rv1242	17.4	x	x
Rv1397c/1398c	Rv1398c	9.4	x	x	Rv1397c	16.5	x	✓
Rv1561/1560	Rv1560	8.3	✓	x	Rv1561	16.2	x	x
Rv1720/1721c	Rv1721c	7.9	x	x	Rv1720	15.4	x	x
Rv1741/1740	Rv1740	7.6	x	x	Rv1741	10.4	✓	x
Rv1838c/1839c	Rv1839c	10.4	x	✓	Rv1838c	16.2	x	x
Rv1953/1952	Rv1952	7.8	x	x	Rv1953	13.1	x	x
Rv1962c/1962a	Rv1962a	9.9	✓	✓	Rv1962c	16	x	✓
Rv1982c/1982a	Rv1982a	9.6	✓	x	Rv1982c	17.1	✓	x
Rv2010/2009	Rv2009	9.0	✓	✓	Rv2010	16.3	✓	x
Rv2103c/2104c	Rv2104c	9.3	✓	x	Rv2103c	18.1	x	✓
Rv2494/2493	Rv2493	8.0	✓	x	Rv2494	16.9	x	✓

Rv2527/2526	Rv2526	8.2	✓	✓	Rv2527	16.7	✓	✓
Rv2530c/2530a	Rv2530a	7.9	x	x	Rv2530	16.3	✓	✓
Rv2546/2545	Rv2545	10.3	✓	x	Rv2546	16.6	x	x
Rv2548/2547	Rv2547	9.9	✓	✓	Rv2548	17.4	✓	✓
Rv2549c/2550c	Rv2550c	9.1	x	x	Rv2549c	16.1	✓	x
Rv2596/2595	Rv2595	9.2	x	x	Rv2596	15.9	x	✓
Rv2602/2601	Rv2601	10.4	x	x	Rv2602	17.6	x	x
Rv2757/2758c	Rv2758c	9.4	x	x	Rv2757	18.2	x	x
Rv2759c/2760c	Rv2760c	10.1	x	x	Rv2759c	15.9	x	x
Rv2829c/2830c	Rv2830c	7.5	x	x	Rv2829c	16.1	x	x
Rv2863/2862a	Rv2862a	7.4	x	x	Rv2863	15.8	x	x
Rv2872/2871	Rv2871	9.1	x	x	Rv2872	18.1	x	x
Rv3320c/3321	Rv3321	8.8	x	x	Rv3320c	18.2	x	x
Rv3384c/3385c	Rv3385c	11.0	x	✓	Rv3384c	15.8	x	✓
Rv3408/3407	Rv3407	10.9	✓	✓	Rv3408	16.2	✓	✓
Rv3697c/3697a	Rv3697a	8.4	x	x	Rv3697c	17.8	x	x

VapBC operons in red bold type indicate those which expressed soluble VapC protein. * indicates VapB fused to VapC, ** indicates operon cloned previously in the laboratory.

No VapB or VapC expression, either soluble or insoluble, was observed for the following operons; Rv0229c, Rv0277a/0277c, Rv0582/0581, Rv0595/0596c, Rv0609/0608, Rv0627/0626, Rv0661c/0662c, Rv0749/0748, Rv0960/0959a, Rv1242/1241, Rv1720/1721c, Rv1953/1952, Rv2602/2601, Rv2757/2758c, Rv2759c/2760c, Rv2829c/2830c, Rv2863/2862a, Rv2872/2871, Rv3320c/3321 or Rv3697c/3697a.

Insoluble expression of VapB and VapC was observed for Rv0549/0550, Rv1982c/1982a and Rv2010/2009. Insoluble expression of VapB only was observed for Rv0656/0657, Rv1561/1560 and Rv2546/2545. Insoluble expression of VapC only was observed for Rv1741/1740 and Rv2549c/2550c.

Soluble expression of VapB and VapC was observed for; Rv0064/0065, Rv0239/0240, Rv0300/0301, Rv0598c/0599, Rv0617/0616, Rv1114/1113, Rv1962c/1962a, Rv2527/2526, Rv2548/2547, Rv3384c/3385c and Rv3408/3407. Soluble expression of VapB only was

observed for Rv1838c/1839c. Soluble expression of VapC only was observed for Rv0665/0664, Rv1397c/1398c, Rv2103c/2104c, Rv2494/2493, Rv2530c/2530a and Rv2596/2595.

The 20 VapBC operons which displayed no expression of either VapB or VapC under standard conditions could be screened for protein expression under reduced incubation temperatures and/or lysis buffer screens to attempt expression of these proteins in the future.

Rv2010/2009 showed a small amount of insolubly expressed VapB and VapC, as well as a large amount of soluble VapB (as seen by the intense band in the Flow Through sample). Due to the insoluble expression of VapC, no His tag was present in the soluble fraction for the soluble VapB protein to be able to bind to the resin. This was also the case for Rv1838c/1839c where a small amount of Rv1839c was solubly expressed, but with no soluble expression of Rv1838c, therefore there was no His-tag for the VapB to bind the resin.

Rv1962c/1962a and Rv3384c/3385c both showed soluble expression of VapB and VapC but no binding of VapB to the resin (as seen by VapB bands in the Flow Through but not Resin samples). This suggests that these VapBC complexes may be bound less tightly than the other VapBC complexes, and the VapB portion is dissociating from VapC more readily.

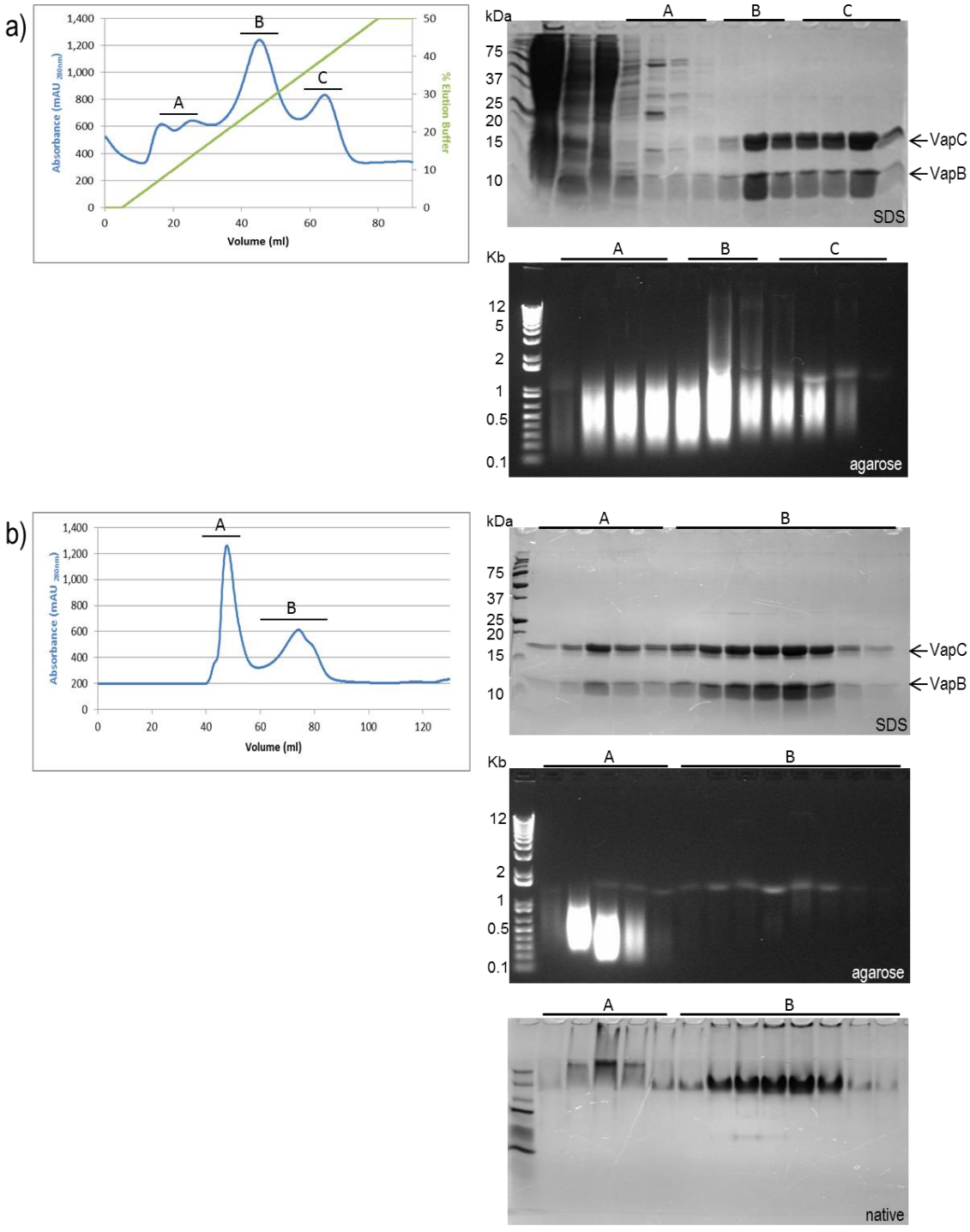
Out of the 47 *M. tuberculosis* VapBC's, 17 were picked for large scale expression studies based on their soluble expression of VapC in the small scale protein expression tests. Only 10 of those 17 *M. tuberculosis* VapBC's were purified during the course of this research, three were purified by others in the lab (Rv0064/0065, Rv0300/0301 and Rv0617/0616) and the remaining four (Rv0665/0664, Rv1114/1113, Rv1397c/1398c and Rv2596/2595) are awaiting purification.

5.2.2.2 Large Scale Expression and Purification of Soluble *M. tuberculosis* VapBC's

Having established that 17 of the 47 *M. tuberculosis* VapCs could be expressed solubly, 10 of these small scale expression screens yielding soluble VapC were then scaled up and the resulting proteins successfully purified by IMAC as in Section 2.3.2.13. The following chromatograms and SDS-PAGE gels are representative of the resulting peaks and fractions observed from IMAC purifications, and show proteins bound to the column non-specifically eluted early in the gradient, whereas VapBC eluted later at ~250 - 300 mM imidazole, SDS-PAGE gels show the column load, flow through and insoluble purification fractions, followed by the resulting IMAC fractions. Fractions identified by IMAC as containing just the VapBC protein complex, were further purified by SEC (Section 2.3.2.14) and typically eluted as a double peak. Purified VapBC samples were also run on an agarose or native PAGE gel when the resulting chromatogram peaks indicated they may have co-purified with DNA. Molecular weights for each SEC purified VapBC complex were calculated as described in Section 2.3.2.14 and are summarised in Table 5.3.

Rv0239/0240

Rv0239/0240 VapBC was successfully purified using both IMAC (Figure 5.4a) and SEC (Figure 5.4b and c). SDS-PAGE and agarose gels of SEC fractions showed DNA was purified along with VapBC for both chromatogram peaks, although the first peak contained the majority of the DNA (Figure 5.4b). A further SEC purification using 2 M NaCl successfully purified the VapBC protein away from the DNA, as seen by the absence of protein in peak A (present only in peak B) on the SDS-PAGE gel, and the absence of DNA in peak B (present only in peak A) on the agarose gel in Figure 5.4c. Native (non-denaturing) PAGE gels also confirm both the presence of DNA (as seen by the difference in size of the VapBC protein bands for A and B fractions in Figure 5.4b), and its removal after SEC with 2 M NaCl (as seen by the absence of protein for peak A fractions, and presence in peak B fractions only in Figure 5.4c). The calculated MWs of the purified VapBC based on elution volume (for SEC peaks B) are 94.89 or 121.74 kDa (after further purification in 2 M NaCl), which equates to approximately a trimer (or 3.23 x the MW of VapBC) or tetramer (or 4.14 x the MW of VapBC) of Rv0239/0240 VapBC complexes respectively.


a)
Absorbance (mAU 280nm)
% Elution Buffer
Volume (ml)
A
B
C
kDa
75
37
25
20
15
10
VapC
VapB
SDS
Kb
12
5
2
1
0.5
0.1
agarose
b)
native

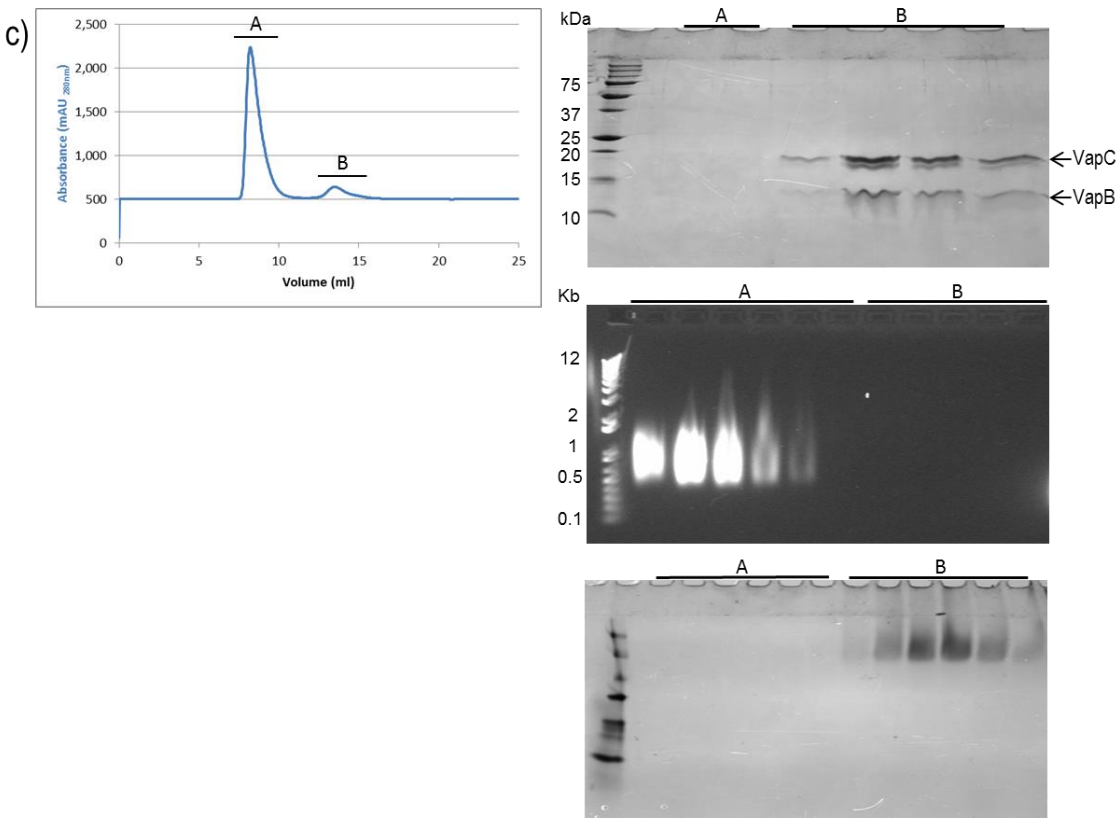


Figure 5.4: Purification of Rv0239/0249 via a) IMAC, b) SEC and c) 2 M NaCl SEC. Chromatograms depict the UV absorbance (blue) and for IMAC, the elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts a) the column load, flow through and insoluble purification fractions followed by the resulting IMAC fractions, or b) and c), the SEC fractions alone. Fractions and their elution position are shown by the black bar. Right hand side, middle (a), b) and c)) = agarose gels showing the presence of DNA in purified fractions, right hand side, bottom (b) and c)) = native PAGE gels confirming the separation of VapBC from co-purified DNA in chromatogram peaks A and B.

Rv0598c/0599

Rv0598c/0599 VapBC was successfully purified using both IMAC and SEC (Figure 5.5). The calculated MW of the purified VapBC based on elution volume (for SEC peak B) is 64.82 kDa which equates to approximately a trimer (or 2.55 x the MW of VapBC) of Rv0598c/0599 VapBC complexes.

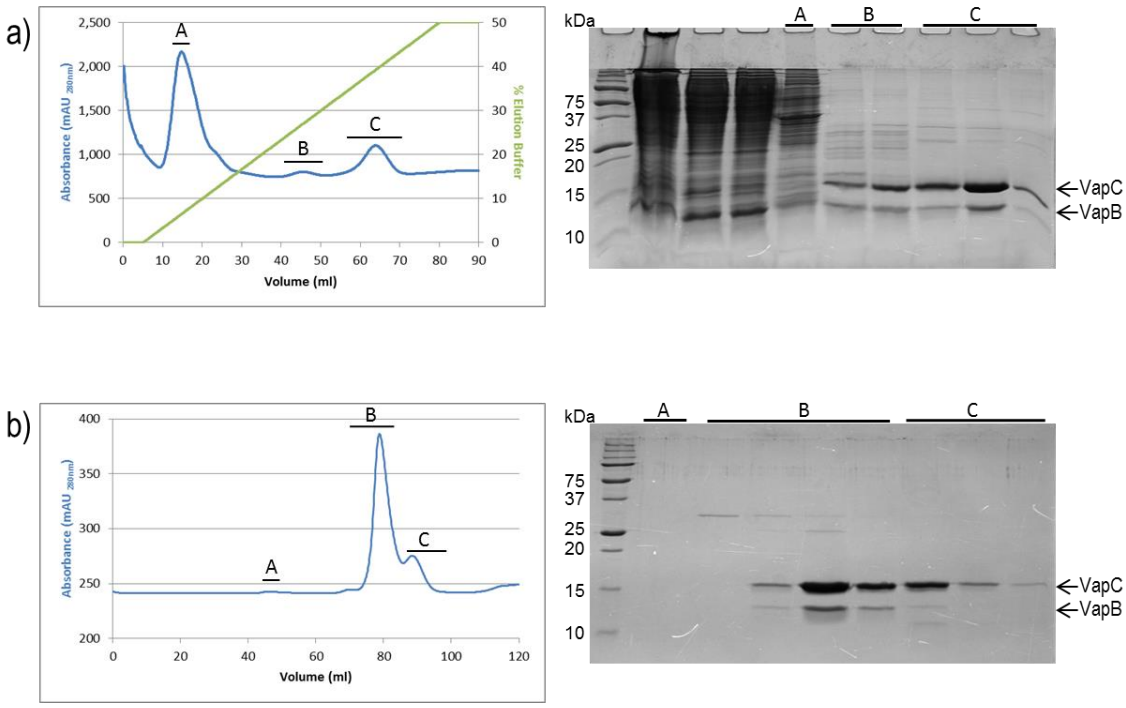


Figure 5.5: Purification of Rv0598c/0599 via a) IMAC and b) SEC. Chromatograms depict the UV absorbance (blue) and for IMAC, the elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts a) the column load, flow through and insoluble purification fractions followed by the resulting IMAC fractions, or b), the SEC fractions alone. Fractions and their elution position are shown by the black bar.

Rv1962c/1962a

Rv1962c/1962a VapBC was successfully purified using both IMAC and SEC (Figure 5.6). The calculated MW of the purified VapBC based on elution volume is 92.98 kDa which equates to approximately a tetramer (or 3.59 x the MW of VapBC) of Rv1962c/1962a VapBC complexes.

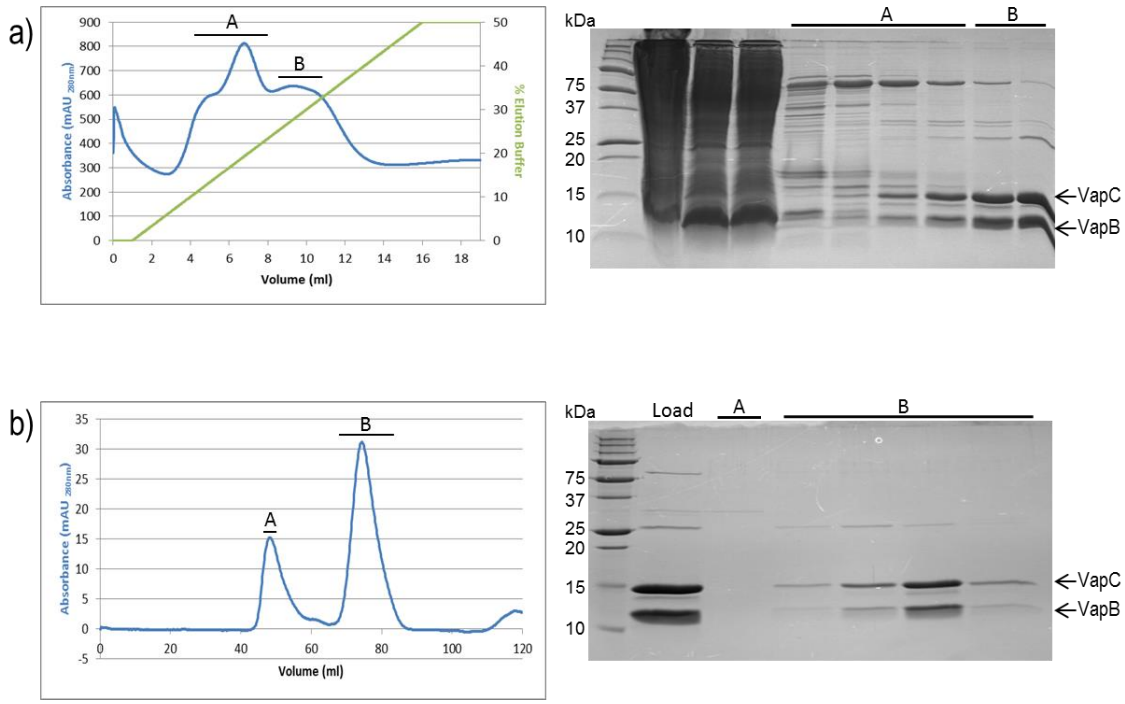


Figure 5.6: Purification of Rv1962c/1962a via a) IMAC and b) SEC. Chromatograms depict the UV absorbance (blue) and for IMAC, the elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts a) the column load, flow through and insoluble purification fractions followed by the resulting IMAC fractions, or b), the SEC fractions alone. Fractions and their elution position are shown by the black bar.

Rv2103c/2104c

Rv2103c/2104c VapBC was successfully purified via IMAC (Figure 5.7). SEC was not performed as the resulting protein from IMAC appeared clean, containing only one other faint band at approximately 27.4 kDa which is the size of the Rv2103c/2104c complex.

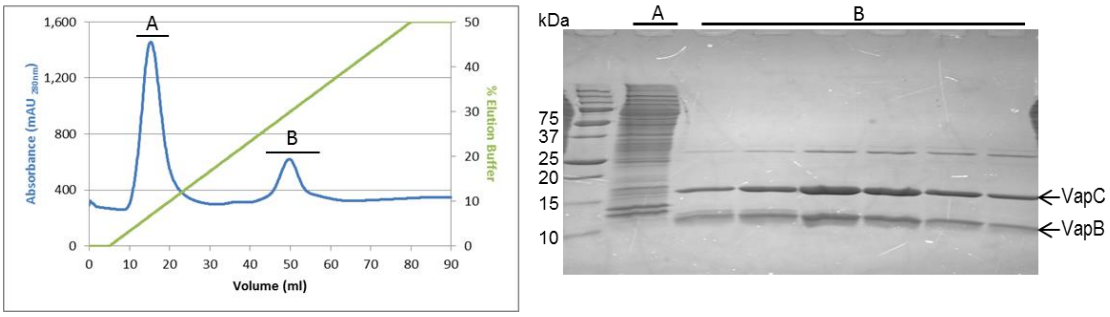


Figure 5.7: Purification of Rv2103c/2104c via IMAC. Chromatogram depicts the UV absorbance (blue) and elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts the column load, flow through and insoluble purification fractions followed by the resulting IMAC fractions. Fractions and their elution position are shown by the black bar.

Rv2494/2493

Only a small amount of Rv2494/2493 VapBC was purified using IMAC and SEC (Figure 5.8). The calculated MW of this purified VapBC based on elution volume is 149.27 kDa, equating to approximately a hexamer (or 6 x the MW of VapBC) of Rv2494/2493 VapBC complexes. However, this MW may be incorrect as it was calculated using the elution volume from the SEC trace below, which is imperfect due to there being so little protein purified. Further purification, possibly under different expression conditions, should be undertaken in the future to purify a greater amount of protein (which elutes with a sharp SEC peak) suitable for calculating an accurate MW value for the protein complex.

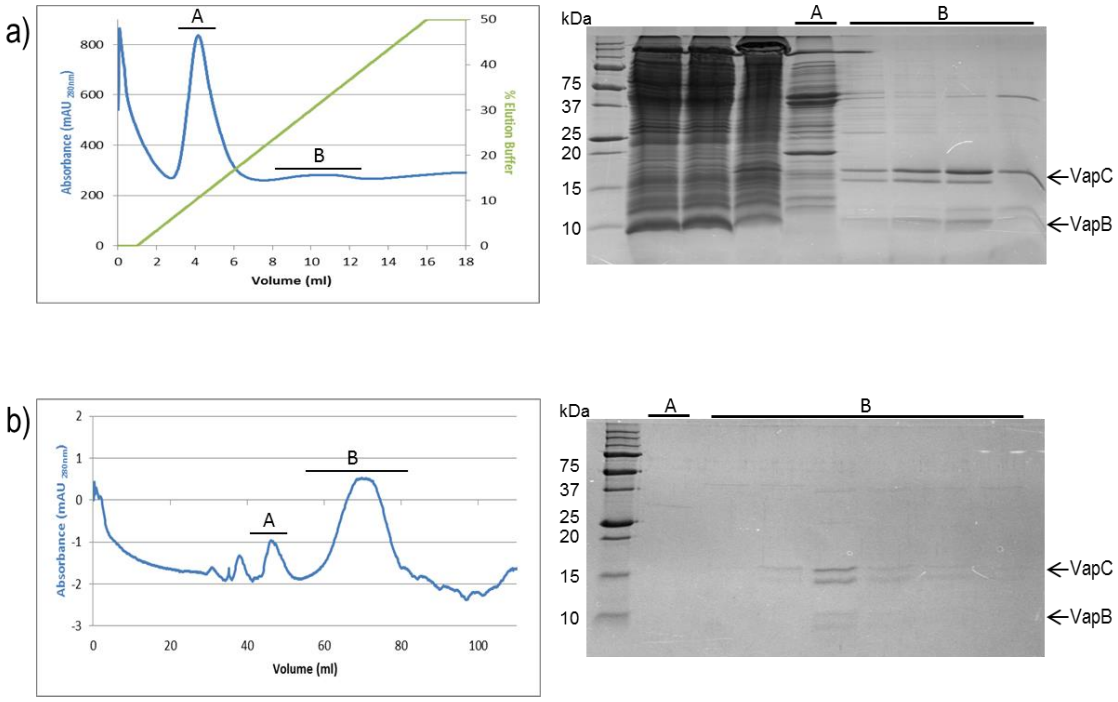


Figure 5.8: Purification of Rv2494/2493 via a) IMAC and b) SEC. Chromatograms depict the UV absorbance (blue) and for IMAC, the elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts a) the column load, flow through and insoluble purification fractions followed by the resulting IMAC fractions, or b), the SEC fractions alone. Fractions and their elution position are shown by the black bar.

Rv2527/2526

Rv2527/2526 VapBC was successfully purified using both IMAC and SEC (Figure 5.9). The calculated MW of the purified VapBC based on elution volume is 86.67 kDa which equates to approximately a tetramer (or 3.48 x the MW of VapBC) of Rv2527/2526 VapBC complexes.

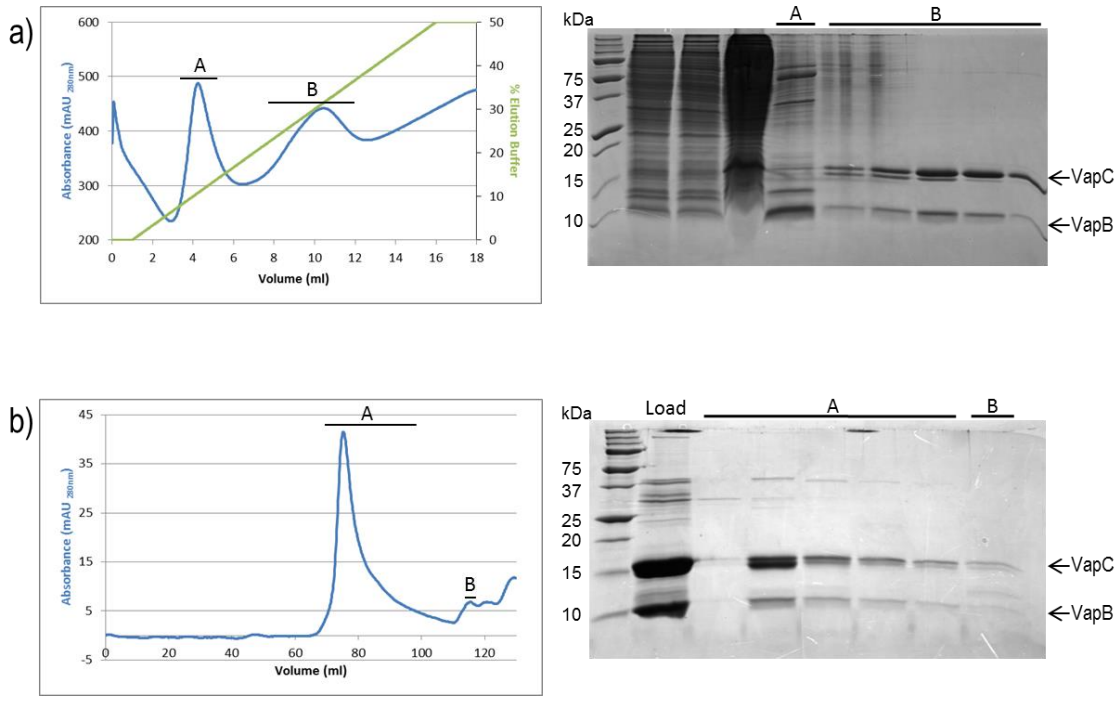


Figure 5.9: Purification of Rv2527/2526 via a) IMAC and b) SEC. Chromatograms depict the UV absorbance (blue) and for IMAC, the elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts a) the column load, flow through and insoluble purification fractions followed by the resulting IMAC fractions, or b), the SEC fractions alone. Fractions and their elution position are shown by the black bar.

Rv2530c/2530a

Rv2530c/2530a VapBC was successfully purified using both IMAC and SEC (Figure 5.10). VapB was not observed on the 16.5 % SDS-PAGE gels in Figure 5.10a and b, nor on a 20 % SDS-PAGE gel (data not shown), but its presence was confirmed by MALDI-TOF MS (Section 2.3.2.9) of the VapBC protein complex (data not shown) and when SEC fractions were run on a native PAGE gel (Figure 5.10b). The fact that the VapB protein could be visualised on a native PAGE gel and that Rv2530a VapB has the lowest pI of all the VapB (and VapC) proteins (3.9), suggests there is something about the protein in its denatured state which prohibits it from migrating through an SDS-PAGE gel in the standard manner. The calculated molecular weights of the purified VapBC based on elution volume for peaks A and B are 663.54 and 119.25 kDa, which equate to approximately 27.42 or 4.93 x the MW of Rv2530c/2530a VapBC complexes respectively. The MW of peak A is unusually large, suggestive of the presence of DNA, but SEC fractions run on an agarose gel showed little DNA, of which there was an equivalent amount in fractions from both A and B peaks (data not

shown). The large MW for peak A may be due to an oligomer with a larger hydrodynamic volume, or a larger oligomeric species.

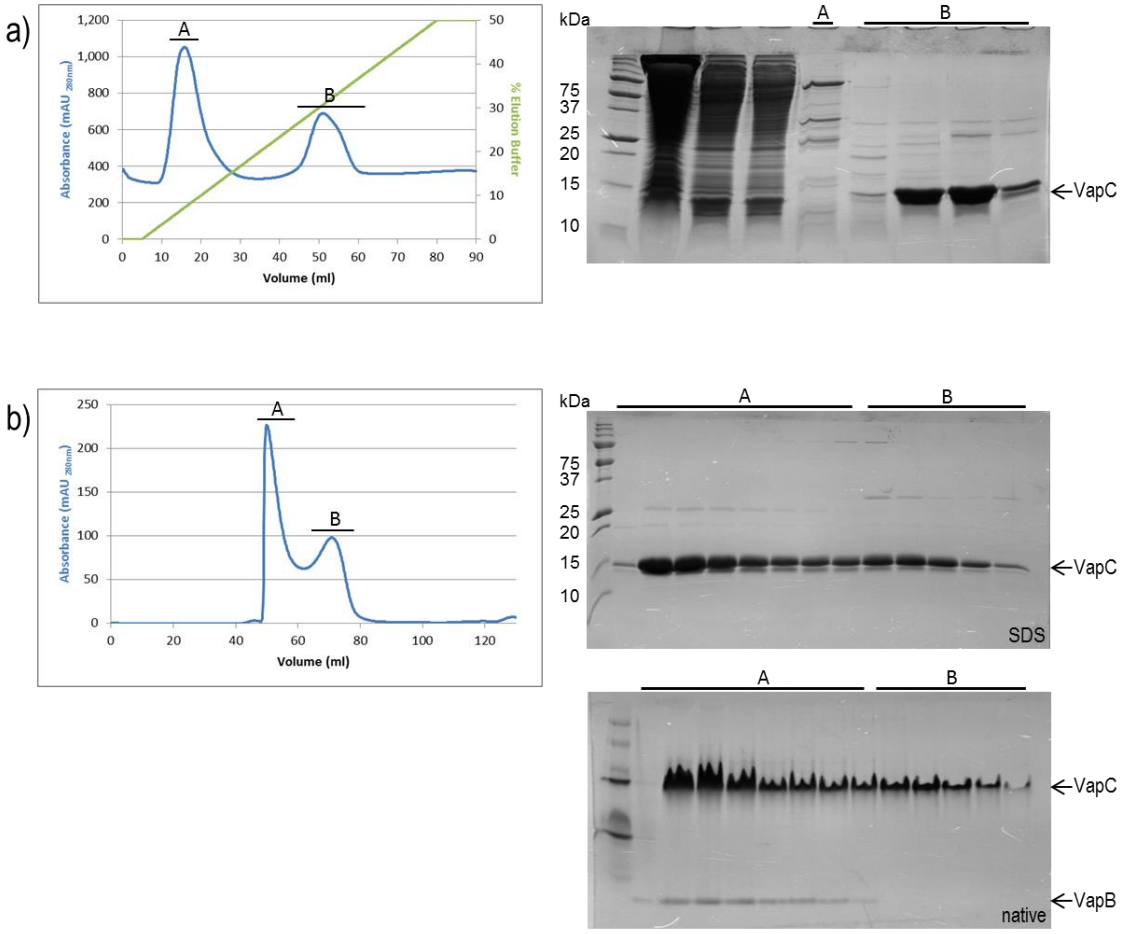


Figure 5.10: Purification of Rv2530c/2530a via a) IMAC and b) SEC. Chromatograms depict the UV absorbance (blue) and for IMAC, the elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts a) the column load, flow through and insoluble purification fractions followed by the resulting IMAC fractions, or b), the SEC fractions alone. Fractions and their elution position are shown by the black bar. Right hand side, bottom, b) = native PAGE gel confirming the presence of VapB in the most concentrated fractions.

Rv2548/2547

Rv2548/2547 VapBC was successfully purified using both IMAC and SEC (Figure 5.11). SDS-PAGE and agarose gels of SEC fractions showed DNA (and a small amount of protein) was present in peak A, whilst the majority of protein was in peak B (Figure 5.11b). The calculated MW of the purified VapBC from peak B (based on elution volume) is 98.73 kDa which equates to approximately a tetramer (or 3.61 x the MW of VapBC) of Rv2548/2547 VapBC complexes.

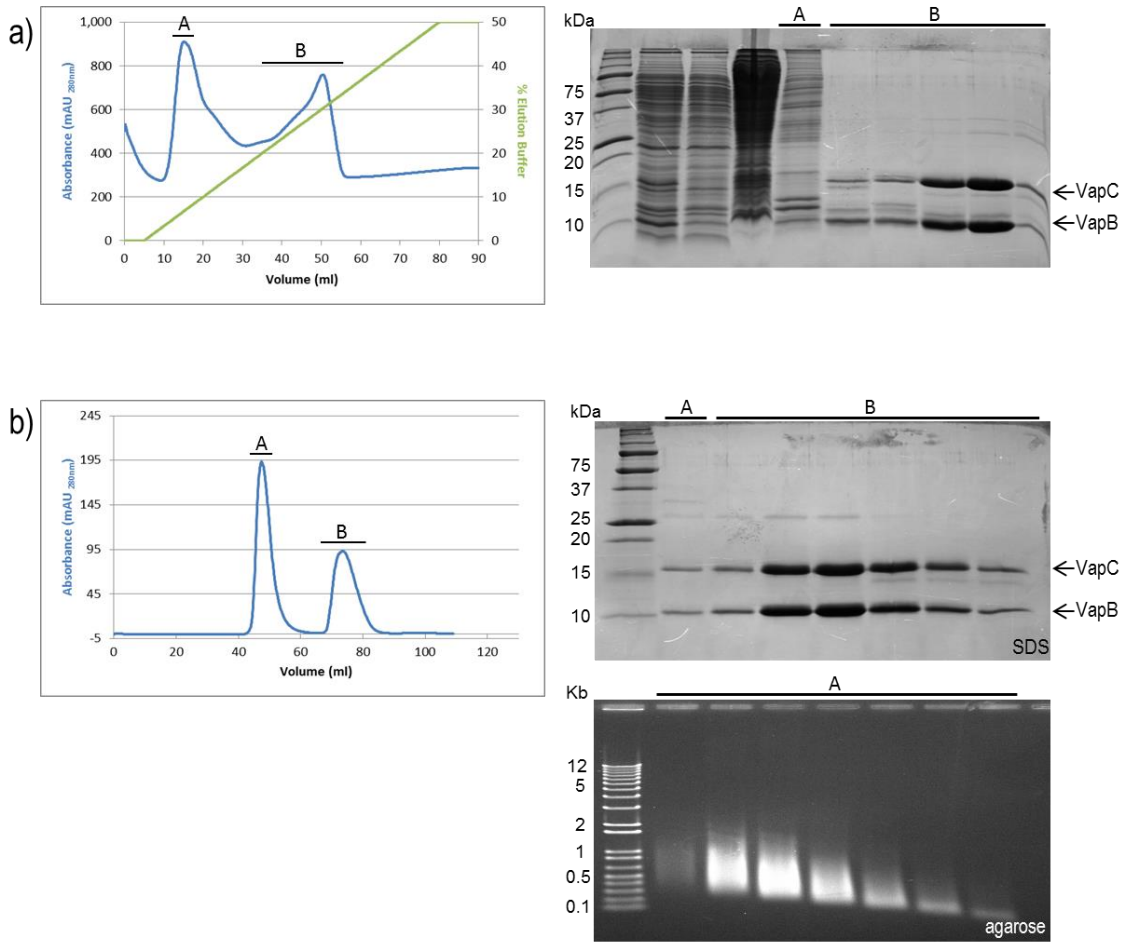


Figure 5.11: Purification of Rv2548/2547 via a) IMAC and b) SEC. Chromatograms depict the UV absorbance (blue) and for IMAC, the elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts a) the column load, flow through and insoluble purification fractions followed by the resulting IMAC fractions, or b), the SEC fractions alone. Fractions and their elution position are shown by the black bar. Right hand side, bottom b) = agarose gel showing the presence of DNA in purified peak A fractions.

Rv3384c/3385c

Rv3384c/3385c VapBC was successfully purified via IMAC (peak C, Figure 5.12). SEC was not performed as the resulting protein from IMAC peak C appeared clean, containing only one other faint band at approximately 26.84 kDa which is the size of the Rv3384c/3385c complex. VapBC protein was observed in earlier IMAC chromatogram peaks (A and B) but these also contained DNA as seen in the agarose gel, Figure 5.12.

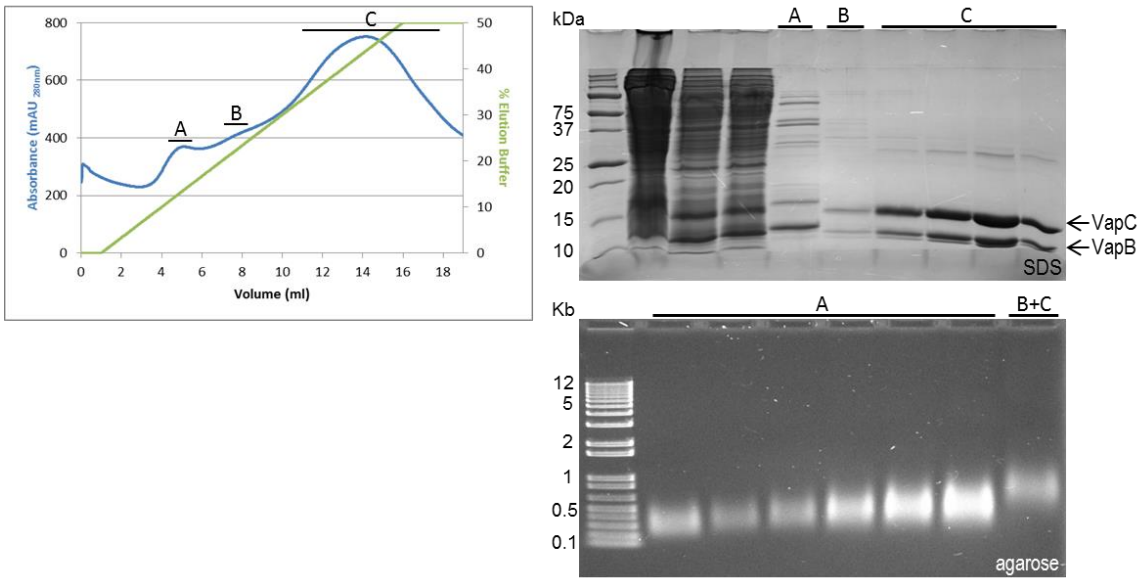


Figure 5.12: Purification of Rv3384c/3385c via IMAC. Chromatogram depicts the UV absorbance (blue) and elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts the column load, flow through and insoluble purification fractions followed by the resulting IMAC fractions. Fractions and their elution position are shown by the black bar. Right hand side, bottom = agarose gel showing the presence of DNA in purified peak fractions.

Rv3408/3407

Only a small amount of Rv3408/3407 VapBC was purified via IMAC and SEC (Figure 5.13). The calculated MW of the purified VapBC from peak A (based on elution volume) `is 152.69 kDa, equating to approximately a hexamer (or 5.63 x the MW of VapBC) of Rv3408/3407 VapBC complexes. Protein is also present along the broad shoulder (labelled B) coming off peak A, with molecular weights ranging from 152.69 down to 26.22 kDa. The true MW of this VapBC complex is likely to be in the middle of this range, and would be more accurately determined if a larger amount of protein were purified.

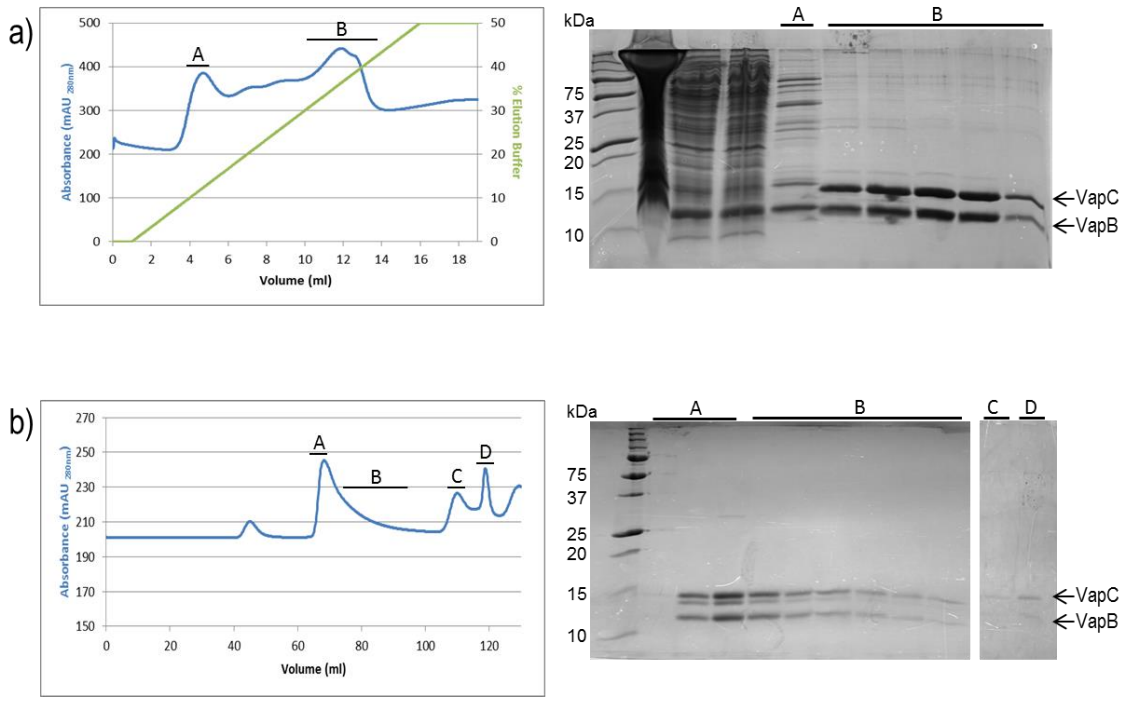


Figure 5.13: Purification of Rv3408/3407 via a) IMAC and b) SEC. Chromatograms depict the UV absorbance (blue) and for IMAC, the elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts a) the column load, flow through and insoluble purification fractions followed by the resulting IMAC fractions, or b), the SEC fractions alone. Fractions and their elution position are shown by the black bar.

Table 5.3 denotes the elution volume of each purified VapBC protein and its corresponding MW as calculated using the equation $MW = e^{((Kav - 2.1739)/ -0.159)}$ from Section 2.3.2.14.

Table 5.3: Calculated MW of purified VapBC proteins after SEC.

Protein	Elution volume (ml)	Kav	MW (kDa)	Predicted VapBC MW	MW / Predicted MW
Rv0239/0240 Rv0239/0240*	47.09 / 73.6 13.50*	0.01 / 0.35 0.34*	810 / 95 122*	29.40	27.56 / 3.23 4.14*
Rv0598c/0599	78.31 / 88	0.41 / 0.54	65 / 30	25.38	2.55 / 1.17
Rv1962c/1962a	74.35	0.36	93	25.90	3.59
Rv2103c/2104c	N/A				
Rv2494/2493	68.00	0.28	149	24.89	6.00
Rv2527/2526	74.72	0.37	87	24.89	3.48
Rv2530c/2530a	49.56 / 70.78	0.04 / 0.32	664 / 119	24.20	27.42 / 4.93
Rv2548/2547	73.61	0.35	99	27.33	3.61
Rv3384c/3385c	N/A				
Rv3408/3407	68.22	0.28	153	27.13	5.63

Predicted MWs were calculated using ProtParam (http://web.expasy.org/cgi-bin/protparam/protparam). * indicates Rv0239/0240 protein purified after 2 M NaCl SEC purification on the S200 10/300 column. SEC was not performed for Rv2103c/2104c or Rv3384c/3385c.

The calculated molecular weights suggest these *M. tuberculosis* VapBCs are large protein complexes, the majority of which may be trimeric or tetrameric. These results are in accordance with the VapBC structures from *M. smegmatis* and *M. tuberculosis* (Rv0300/0301 and VapBC-5 (Rv0626/0627)), which are tetramers of heterodimers (Miallau *et al.* 2009; McKenzie 2011; Min *et al.* 2012) in an arrangement similar to the *N. gonorrhoeae* FitAB structure (Mattison *et al.* 2006).

Of the 10 purified VapBC complexes above, only eight were carried through to the following stage (VapC purification, Section 5.2.3), as not enough VapBC protein was produced during the purification process for Rv2494/2493 and Rv2527/2526 (which precipitated readily at both 4 and 18 °C during concentration and dialysis).

The order with which proteins were purified/focussed on was based on VapC toxicity results and the requirement of VapBCs during various stages of growth which were published during the course of this research (Beste *et al.* 2009; Gupta 2009; Ramage *et al.* 2009; Ahidjo *et al.* 2011). Toxicity studies were performed in *M. smegmatis*, *M. tuberculosis* or *E.coli*, while the TraSH analysis was performed on BCG and identified the genes required or selected for

(either positively or negatively) at slow, fast or the switch from slow to fast growth rates and vice-versa (see Table 5.4). Results published during the course of this research identified VapC proteins of potential interest for the following reasons: Rv2527 and Rv2530c were prioritised due to their classification in the group of genes most potent in disrupting the bacteriums latency and dormancy program. These genes were identified through predicted gene regulatory networks and computational gene deletion experiments, from microarray gene expression profiles of bacilli adaptation and survival in latency, simulated by hypoxia (Magombedze & Mulder 2013). Rv0595c, Rv0598c, Rv0627, Rv1838c, Rv2103c, Rv2494, Rv2530c, Rv2548, Rv2602, Rv2829c and Rv3320c proteins were found to be increased in nutrient starved cultures using proteomic profiling, implicating a metabolic role for the *M. tuberculosis* toxin proteins in the switch to dormancy (Albrethsen *et al.* 2013).

The *M. tuberculosis* VapBC proteins Rv0064/0065, Rv0300/0301 and Rv0617/0616 were all purified by others in the laboratory; Rv0665/0664, Rv1114/1113, Rv1397c/1398c and Rv2596/2595 were not purified during the course of this research and are earmarked for future work.

Table 5.4: Basis for preferential purification of VapBC proteins.

Protein	VapC Toxicity studies			Genetic Requirements for Fast and Slow Growth - TraSH analysis (BCG)	
	Ahidjo *et al.* (*M.tb* & *M.sm*)	Gupta *et al.* (*E.coli*)	Ramage *et al.* (*M.sm*)	VapB	VapC
Rv0064/0065	NT	-	-		
Rv0229c	NT	NT	NT		Pos selected at fast GR
Rv0239/0240	NT	NT	-		
Rv0277a/0277c	NT	NT	+		
Rv0300/0301	NT	slight	+		
Rv0549/0550	+	slight	+		Pos selected at fast GR
Rv0582/0581	NT	NT	+		
Rv0595/0596c	+	+	-	Neg selected at fast GR	
Rv0598c/0599	NT	NT	-	Req slow-fast GR switch	
Rv0609/0608	NT	NT	+		
Rv0617/0616	NT	NT	-		
Rv0627/0626	-	slight	-		
Rv0656/0657	NT	-	-	Req slow-fast GR switch	
Rv0661c/0662c	NT	-	-		
Rv0665/0664	NT	slight	-		Pos selected slow-fast
Rv0749/0748	NT	NT	+		
Rv0960/0959a	NT	-	-		
Rv1114/1113	NT	NT	+		
Rv1242/1241	NT	NT	+		Pos selected slow-fast
Rv1397c/1398c	NT	slight	-		
Rv1561/1560	NT	+	+		
Rv1720/1721c	NT	-	-		
Rv1741/1740	NT	NT	-		Req slow-fast GR switch
Rv1838c/1839c	NT	-	-		
Rv1953/1952	-'ve control	slight	-	Pos selected at slow GR	Neg selected at slow GR
Rv1962c/1962a	NT	NT	+		Neg selected at fast GR + pos selected slow-fast
Rv1982c/1982a	NT	NT	-		
Rv2010/2009	-	-	+		
Rv2103c/2104c	NT	NT	+ ^		Neg selected at slow GR + pos selected slow-fast

Rv2494/2493	NT	NT	-	Pos selected slow-fast	Neg selected at slow GR
Rv2527/2526	NT	slight	-		
Rv2530c/2530a	NT	NT	+		
Rv2546/2545	-	-	-	Pos selected slow-fast	
Rv2548/2547	-	-	+		
Rv2549c/2550c	+	+	-		
Rv2596/2595	NT	NT	-		
Rv2602/2601	NT	NT	+		
Rv2757/2758c	NT	slight	+	Req slow-fast GR switch	
Rv2759c/2760c	NT	NT	-		Req slow-fast GR switch
Rv2829c/2830c	+	-	+		
Rv2863/2862a	NT	slight	-		Pos selected fast-slow
Rv2872/2871	NT	NT	+		
Rv3320c/3321	+*	NT	-	Pos selected at slow GR + pos selected slow-fast	Pos selected slow-fast
Rv3384c/3385c	NT	NT	+		Req slow-fast GR switch
Rv3408/3407	NT	NT	+		Req slow-fast GR switch
Rv3697c/3697a	NT	NT	NT		Pos selected at slow GR

VapC toxicity studies show VapC proteins which were toxic for cell growth in the organism stated, NT = not tested, + = toxic, +* = toxic but only when constitutively expressed, - = not toxic (Gupta 2009; Ramage *et al.* 2009; Ahidjo *et al.* 2011). Toxicity could be rescued with each cognate VapB protein but rescue of Rv2103c toxicity with Rv2104c VapB was not tested (^). TraSH analysis identified genes required for (Req) or positively (Pos) or negatively (Neg) selected for, for fast, slow or the switch to/from fast to slow growth rates (GR) (Beste *et al.* 2009).

5.2.3 Purification of *M. tuberculosis* VapCs

The next stage in obtaining VapC proteins for functional analysis was separating and purifying the VapCs from the purified VapBC complexes. Previous work (McKenzie 2011) showed that functional VapC could be recovered from the *M. smegmatis* VapBC complex by performing a tryptic digest on the protein complex. This method takes advantage of the differences in stability of the VapB and VapC proteins, and the greater susceptibility of VapB to proteolytic degradation which is a trait common to TA systems (Gerdes *et al.* 2005; Miallau *et al.* 2009).

5.2.3.1 Purification of VapC Following Trypsin Digestion of the VapBC Complex

VapB proteins were digested from the VapBC complexes by treating the complex with 0.1 mg trypsin per 0.7 mg protein for 1 hr then stopping the reaction by the addition of an equal amount of trypsin inhibitor for 15 min. Pure VapC was required for RNase activity analysis, so anion exchange chromatography was employed to purify VapC away from trypsin, inhibitor and any remaining VapB degradation products.

For VapCs with isoelectric points unsuitable for anion exchange, attempts were made using size exclusion chromatography, Immobilised Metal Affinity Chromatography or cation exchange chromatography to purify the VapC protein.

5.2.3.1.1 Successful anion exchange

Rv2530c

VapC (pI 7.2) was successfully purified by anion exchange chromatography in 50 mM phosphate buffer pH 8, 100 mM NaCl. VapC elutes from the column at 200 - 300 mM NaCl (Figure 5.14). Higher molecular weight bands are also present with the VapC samples on the SDS-PAGE gel, and as with the SEC results for this protein (Figure 5.10), the fractions appear to consist of two different populations. Fractions from peak B include a faint band at approximately 30 kDa which is possibly a dimer of VapC proteins (corresponding to a molecular weight of ~32 kDa), whilst those from A include faint bands at approximately 18 and 24 kDa.

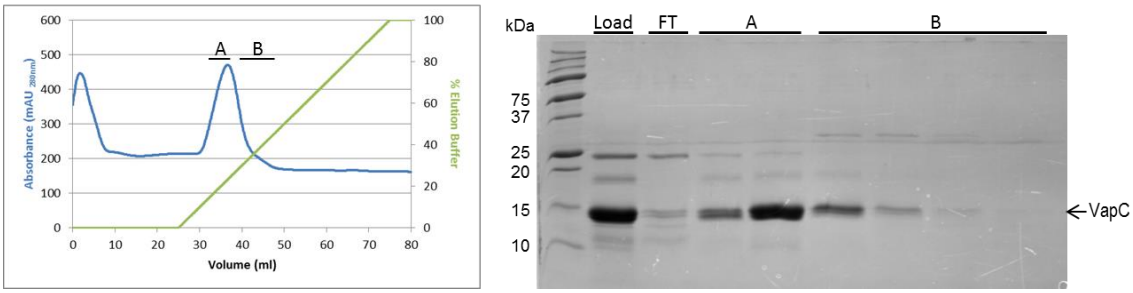


Figure 5.14: Purification of Rv2530c via anion exchange. Chromatograms depict the UV absorbance (blue) and elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts the column load, flow through (FT) and resulting anion exchange fractions (elution positions are shown by the black bar) containing VapC at approximately 16.3 kDa.

Rv2548

VapC (pI 5.7) was successfully purified by anion exchange chromatography in 50 mM phosphate buffer pH 7.4, 100 mM NaCl. VapC elutes from the column at 200 - 300 mM NaCl (Figure 5.15). Two smaller molecular weight bands are present with the VapC samples on the SDS-PAGE gel, as well as a large low molecular weight smear. The two bands appear to be truncated VapC products which have been degraded by trypsin, and the smear degraded VapB and VapC. MALDI-TOF MS (data not shown) revealed VapC at 17.4 kDa which correlates to VapC including the C-terminal His-tag, and multiple smaller species of VapC ranging from 14.4 to 16 kDa. These smaller species account for VapC minus the C-terminal His-tag as well as portions of the linker region which would have molecular weights of 16.6 and 13.7 kDa respectively (Protparam), and other VapC degradation products. The low molecular weight smear seen on the SDS-PAGE gels is represented by multiple peaks ranging from 7.4 to 10.8 kDa. None of these peaks are at 9.9 (the calculated molecular weight of VapB) indicating that VapB has been degraded during the tryptic digest leaving behind only VapC and various degradation products.

A further tryptic digest was performed with less trypsin (a trypsin to protein ratio of 0.1 mg : 1 mg repectively) in an attempt to stop the cleavage of VapC. This still produced the three VapC bands observed in Figure 5.15 and also failed to completely digest VapB, leaving behind enough intact VapB to hinder RNase activity of the VapC (data not shown), suggesting the conformation of this VapC makes it more susceptible to degradation by trypsin. All subsequent digests were performed with the standard trypsin to protein ratio.

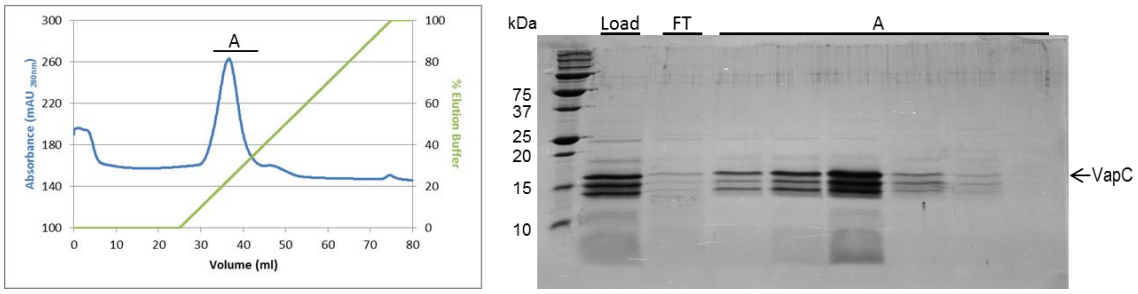


Figure 5.15: Purification of Rv2548 via anion exchange. Chromatograms depict the UV absorbance (blue) and elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts the column load, flow through (FT) and resulting anion exchange fractions (elution positions are shown by the black bar) containing VapC at approximately 17.4 kDa and degraded VapC products at lower molecular masses.

5.2.3.1.2 Unsuccessful anion exchange

Rv0240

VapC (pI 6.8) did not bind the anion exchange column in 50 mM phosphate buffer pH 7.4, 100 mM NaCl, or when NaCl was dropped to 50 mM and buffer at pH 8.2, instead it remained in the flow through (data not shown). The lack of binding suggests the actual pI of the protein in its native state is higher than 6.8, probably due to an abundance of basic amino acid residues concentrated on the outside of the protein structure.

A cation bead screen to determine ion exchange conditions where the protein would bind showed Rv0240 successfully bound the cation beads at approximately pH 4.5 and below (Figure 5.16a). An anion bead screen was unsuccesfull as the protein did not bind the beads properly, although a small amount of binding to the beads and a corresponding decrease in the amount of protein in the supernatant samples was observed above pH 8.5 (Figure 5.16b). This lack of good binding to the anion beads is probably due to the conditions never reaching a high enough pH to deprotonate the outer residues and allow binding to the positively charged beads. The pI of the protein in its native state is possibly higher than its predicted pI due to an abundance of basic amino acids such as arginine concentrated on the outside of the protein structure. Arginine (of which Rv0240 has 16 of its 145 aa) has a high pka value (12) so requires a higher pH for deprotonation to occur and create an overall negative net charge on the protein so it can therefore bind to a positively charged anion column. From these results we can assume the actual pI of Rv0240 in its native state is between approximately 5.4 and

8.6, and would require pH conditions of greater than 10 for anion exchange purification which would be fruitless as the protein is likely to denature in this pH range.

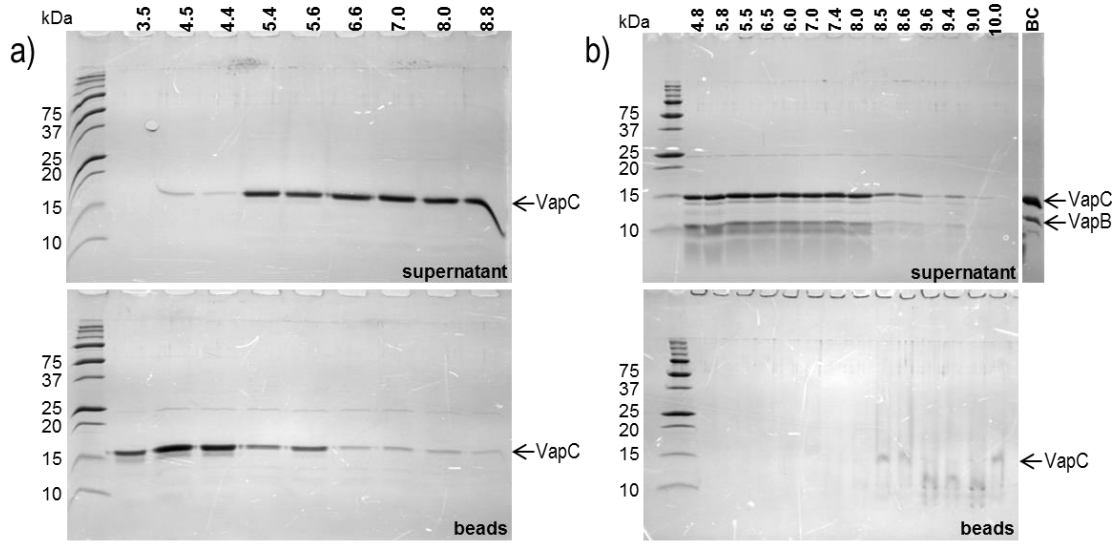


Figure 5.16: Rv0240 ion exchange affinity tests on a) cation beads, and b) anion beads. The SDS-PAGE gels depict varying ion exchange pH conditions across the gel, and the resulting protein bound to the bead (bottom gels) and corresponding supernatant fractions (top gels). Numbers at the top of gels refer to the pH of the buffers trialled (note, ion exchange buffer conditions can be found in Sections 2.3.2.16 and 2.3.2.17), BC refers to VapBC protein run to confirm no VapB was present in the VapC samples.

A full cation exchange purification was performed in 20 mM citric acid pH 3.5, 50 mM NaCl, Figure 5.17 shows Rv0240 of approximately 17.9 kDa eluting at approximately 700 mM NaCl (peak C), after what appears to be trypsin (which has a molecular weight of 23.3 kDa) in peak B. Whilst purification of Rv0240 away from trypsin appeared to be successful via cation exchange in citric acid pH 3.5, the protein quickly degraded in this buffer at low pH and was unusable. Cation exchange on a fresh preparation of trypsin treated Rv0240 in 20 mM acetic acid pH 4.4, 50 mM NaCl was trialled, but was unable to cleanly separate trypsin (with a pI of 10.1 – 10.5) from the VapC protein (data not shown). Future cation exchange purifications may be successful in citric acid pH 3.5 if the fractions are collected immediately into an equal volume of neutralising buffer.

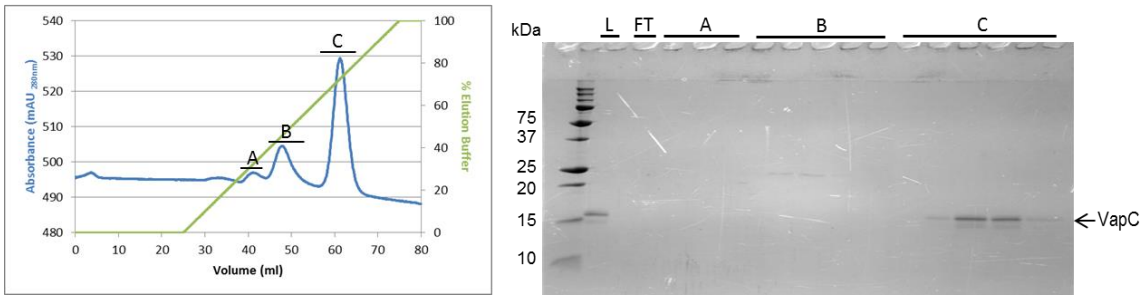


Figure 5.17: Purification of Rv0240 via cation exchange in 20 mM citric acid pH 3.5, 50 mM NaCl. Chromatogram depicts the UV absorbance (blue) and elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts the column load (L), flow through (FT) and resulting cation exchange fractions (elution positions are shown by the black bar). VapC is eluted in peak C (approximately 17.9 kDa), separate from trypsin which elutes in peak B (approximately 23.3 kDa).

Rv0598c

VapC (pI 8.5) did not bind the anion exchange column in 50 mM phosphate buffer pH 9.4, 100 mM NaCl, instead it remained in the flow through (data not shown). It was unlikely that the phosphate buffer was buffering properly at this pH as it is too far away from any of its three pKa's of 2.15, 7.2 or 12.32. It was also unlikely that the protein would remain stable or biologically active at a pH closer to the highest pKa value where anion exchange would be possible, so VapB was digested with a limited tryptic digest (Section 2.3.2.15.1) without purifying the VapC afterwards.

Rv1962c

VapC (pI 6.1) did not bind the anion exchange column 50 mM phosphate buffer pH 7.4, 100 mM NaCl, instead it remained in the flow through (data not shown), and so would be a good candidate for an anion bead screen in future experiments.

Rv2103c

VapC (pI 7.2) did not bind the anion exchange column in 50 mM phosphate buffer pH 8.2, 100 mM NaCl, or when NaCl was dropped to 50 mM, instead it remained in the flow through (data not shown).

Cation exchange was performed at a pH below the predicted pI for VapC (7.2) in case Rv2013c wasn't binding the anion column because its actual pI in solution was higher than the

predicted pI. SEC of Rv2130c in 50 mM MES pH 5.6, 200 mM NaCl was performed prior to cation exchange where it eluted from the column at the correct size of approximately 18.1 kDa (Figure 5.18a), however no binding of 2103c was observed during cation exchange in 50 mM MES pH 5.6, 100 mM NaCl (data not shown). IEF was performed as in Section 2.3.2.3 to determine the actual pI of the Rv2103c VapC protein. The specific pI of a protein is the pH at which the protein has no net charge and therefore stops migrating through a pH gradient in an electric field. Figure 5.18b shows VapC at pH 7.3 and trypsin inhibitor at pH 4.7. Both of these proteins have migrated through the IEF gel close to their predicted pI's of 7.2 and 4.5 respectively. This result suggests the VapC protein should have bound the cation column at pH 5.6, and possibly could have bound the anion column at pH 8.2. However, during IEF the protein is in its denatured state, so during ion exchange when it's in its native state, it may harbour a vastly different overall net charge and therefore pI. This may explain why the VapC protein didn't bind the ion exchange columns at the pH range tested, so future experimentation could be conducted using a wider pH range in an attempt to obtain binding.

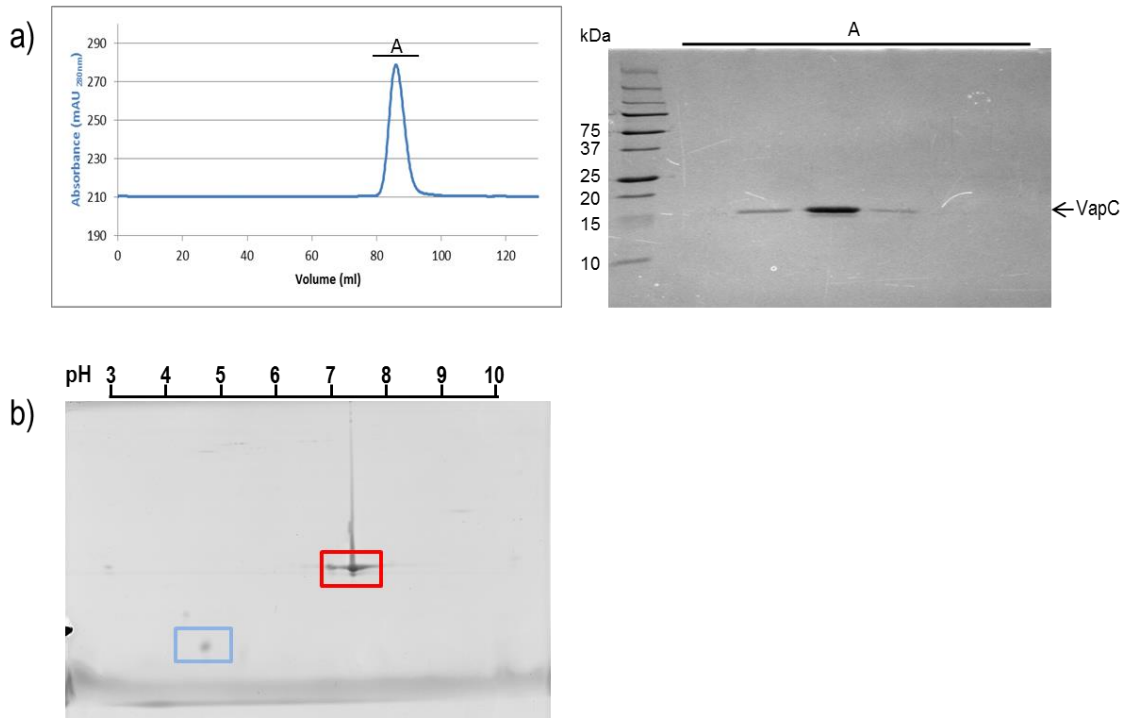


Figure 5.18: Rv2103c a) SEC in MES and b) IEF gel. a) SEC chromatogram depicts the UV absorbance of the VapC protein. The corresponding SDS-PAGE gel depicts the SEC fractions with their elution position shown by the black bar. b) Numbers at the top of the IEF gel represent the pH gradient across the gel while coloured boxes indicate the pH at which proteins have been immobilised. The red box indicates Rv2103c is immobilised at pH 7.3, and blue box indicates trypsin inhibitor at pH 4.7.

Rv3384c

The pI of VapC (10.6) was too close to that of trypsin (10.1 - 10.5) to purify the protein using anion exchange chromatography. A small scale limited tryptic digest and His-tag affinity binding test using Ni Sepharose beads was performed to see if the VapC protein still contained its His-tag after the digest and could therefore be purified by this process. Figure 5.19a shows binding of VapC to the nickel resin after a limited tryptic digest (Resin sample), although some has been lost (when compared to the VapBC control sample). This suggests the standard limited tryptic digest partially removes the His-tag from VapC as well as degrades the VapB protein. A full scale limited tryptic digest and purification via IMAC using a HisTrap HP column was then performed but with limited success. Figure 5.19b shows only a small amount of VapC (15.8 kDa) in peak B with very little present in the Load or Flow Through samples when compared to the trypsin band at approximately 23.3 kDa, suggesting the remainder of VapC has been degraded by the trypsin. This is supported by the presence of two very faint smaller bands at approximately 14 and 13 kDa which are likely to be VapC minus the C-terminal His-tag and linker region (which would have a predicted molecular weight of 14.3 kDa (Gasteiger *et al.* 2005), and further cleaved VapC products. Another full scale limited tryptic digest and IMAC purification was performed with digestion for only 45 minutes (compared to one hour previously), to limit the cleavage of VapC and its His-tag and increase the amount of purified VapC recovered after the process. More VapC was recovered in the fractions this time (data not shown), although once these fractions were concentrated to approximately 1 mg/ml, a small amount of VapB could still be seen at 11 kDa when run on an SDS-PAGE gel (Figure 5.19b, BC lane), and no RNase activity was observed when the sample was tested on total RNA (data not shown).

The molecular weights of VapC and trypsin are close (15.8 and 23.3 kDa respectively), but size exclusion chromatography was performed on VapC resulting from a standard limited trypsin digest of VapBC in a final attempt to purify the VapC protein away from trypsin. The sample was analysed on an analytical S75 10/300 gel filtration column with a separation range of 3 – 70 kDa as in Section 2.3.2.18, but no separation of trypsin and VapC bands were observed (data not shown).

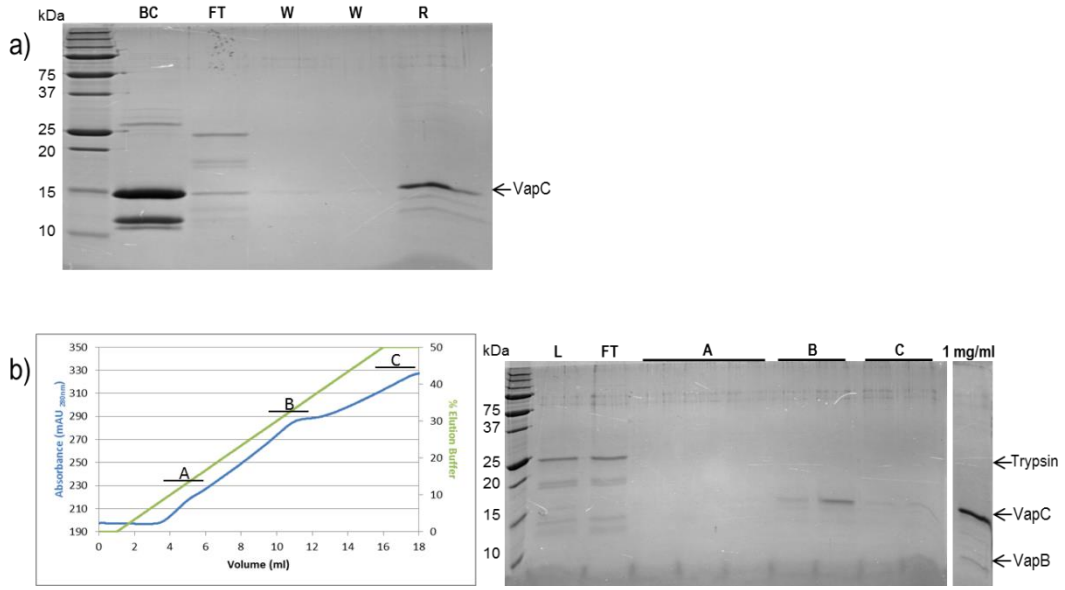


Figure 5.19: Rv3384c IMAC purification from a) small scale affinity test, and b) full scale HisTrap HP column. a) SDS-PAGE gel depicts the untreated VapBC control (BC), flow through (FT), wash (W), and Ni resin with Rv3384c bound after wash steps (R). b) Chromatogram depicts the UV absorbance (blue) and elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts the column load (L), flow through (FT) and resulting IMAC fractions (elution positions are shown by black bar). The gel slice on the right shows VapC protein (concentrated to 1 mg/ml) resulting from the 45 min tryptic digest and IMAC purification also contains VapB.

Rv3408

VapC (pI 6.7) did not bind the anion exchange column in 50 mM phosphate buffer pH 8, 100 mM NaCl, nor did it appear in the flow through or load fractions (data not shown) It is likely that both the VapB and VapC proteins have been degraded during the standard limited tryptic digest protocol, which suggests the Rv3408/3407 complex is less stable than other VapBCs. Limited tryptic digests were performed with a lesser trypsin to protein ratio and for a shorter duration by others in the laboratory resulting in digested VapB and intact active VapC which could be separated by anion exchange using the standard protocol (Chelsea Vickers, unpublished data).

5.2.4 Identification of RNase Contamination during Limited Tryptic Digests

5.2.4.1 Identification of Contaminating RNase in Trypsin Stocks

The possibility of RNase contamination arising during the limited tryptic digest process was noticed during RNase activity testing of Rv0598c. This VapC was unable to be purified away from trypsin in Section 5.2.3.1.2, so crude VapC without anion exchange was tested for RNase activity and demonstrated unusually fast RNase activity. During the total RNA RNase activity assay, a trypsin control consisting of RNA + trypsin and trypsin inhibitor in assay buffer and water completely degraded the RNA, indicating a contaminating RNase was present. As a consequence of these results, various RNase activity assay controls were carried out to ascertain which reagent or labware the RNase contamination was coming from, as such high levels of RNase activity hadn't been noticed previously in any of the other *M. tuberculosis* VapC assays. Total RNA RNase activity assays were set up without any VapC protein added, instead testing equal volumes of water, buffer and trypsin + trypsin inhibitor aliquots, as well as various tubes and spatulas used to make up the reagents. There was no contamination observed in the water, buffer aliquots or tubes used (Figure 5.20a) whilst samples including trypsin + trypsin inhibitor still exhibited RNase activity. New trypsin + trypsin inhibitor reagents were made up using existing laboratory spatulas or gamma irradiated loops and tested in standard RNase activity buffer as well as buffer including EDTA. The fresh aliquots of trypsin + trypsin inhibitor exhibited RNase activity regardless of how they were prepared, and whether they were in buffer containing EDTA or not which indicated the contaminating RNase activity was not metal dependent. To confirm if the contaminant was in the trypsin or trypsin inhibitor, each reagent was added individually to assays samples and tested. Figure 5.20b shows all samples containing trypsin were affected whilst those containing trypsin inhibitor were not. An un-opened vial of TPCK trypsin was also tested in Figure 5.20b and found to be contaminated indicating the contamination was inherent in, and not introduced into, the trypsin stocks.

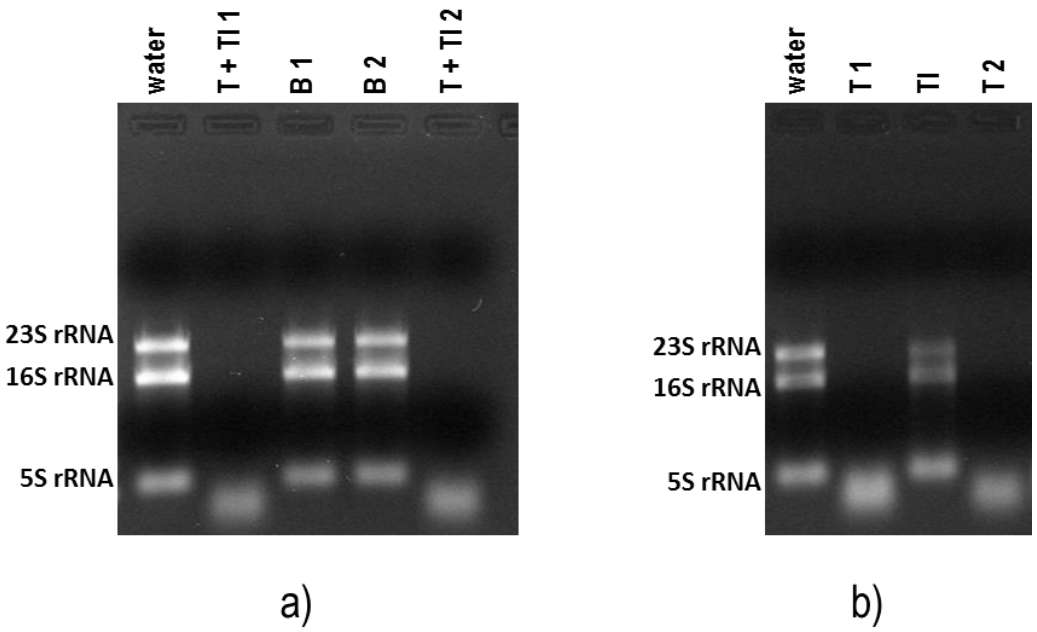


Figure 5.20: Total RNA RNase assays identifying RNase contamination of trypsin stocks. RNase assays were set up minus VapC protein and included the addition of 0.7 µl of each of the following reagents: a) nuclease free water, trypsin + trypsin inhibitor prepared in baked tubes (T + TI 1), assay buffer stock 1 (B 1), assay buffer stock 2 (B 2) and trypsin + trypsin inhibitor prepared in non-baked tubes (T + TI 2); or b) nuclease free water, trypsin (T 1), trypsin inhibitor (TI) and TPCK treated trypsin from an unopened vial (T 2). Figure a) shows only samples containing trypsin + trypsin inhibitor show RNase activity, and figure b) that RNase activity occurs in the trypsin not trypsin inhibitor samples.

5.2.4.2 Removal of Contaminating RNase via Anion Exchange Chromatography

Attempts to source RNase-free trypsin or another protease harbouring a pI dissimilar to that of the VapC proteins to enable purification via ion exchange chromatography were unsuccessful. Therefore, due to a contaminating RNase being introduced into the VapC samples during the limited tryptic digest stage, further analysis was required to elucidate if the contaminant could be removed during the ion exchange process. A *M. tuberculosis* VapC which had been successfully purified by anion exchange previously, Rv2548, was purified as for Sections 2.3.2.13 - 2.3.2.15 and split into two aliquots. Only one of the aliquots was further purified via anion exchange chromatography before their RNase activities were tested. An identical anion exchange was also performed on a mock trypsin + trypsin inhibitor control (the same as that for Rv2548 but without any VapC protein) to clarify which anion exchange fraction contained

the contaminating RNase. Fractions from both of the anion exchange purifications were split into five groups: FT (flow through), A, B (fractions containing VapC), C and D as illustrated in Figure 5.21a. The B samples were concentrated to 1 mg/ml from 12 ml down to 1.2 ml, and this concentration factor of 10 was used as a guide for concentrating the other four samples to volumes reflecting a similar concentration relative to VapC, because there is no actual protein concentration to be physically measured for these samples.

RNase activity assays were set up on total RNA as for Section 2.3.4.4 using fractions from the two anion exchange purifications at their relative 1mg/ml concentrations, and the crude Rv2548 without anion exchange purification. Figure 5.21b demonstrates fast RNase activity of crude Rv2548 compared to anion exchange purified Rv2548, indicating the contaminating RNase from trypsin has been removed from the VapC sample by anion exchange purification. The figure also shows the contaminating RNase appears to not bind the anion exchange column, instead ending up in the flow through sample as demonstrated by RNase activity of the anion exchange purified Rv2548 and mock trypsin FT samples. The RNase activity is much more intense in the Rv2548 FT sample compared to that from mock trypsin, most probably due to the combined effect of the contaminant and VapC.

These results confirm a contaminating RNase is introduced to the VapC samples during the limited tryptic digest process, which is subsequently removed during anion exchange purification. This indicates crude VapC samples resulting from tryptic digests will be contaminated and unfit for testing in RNase activity assays unless further purified by anion exchange, and trypsin digested protein purification by other means such as cation exchange, SEC or IMAC would need to be verified first to ensure the contaminant is also removed during these procedures.

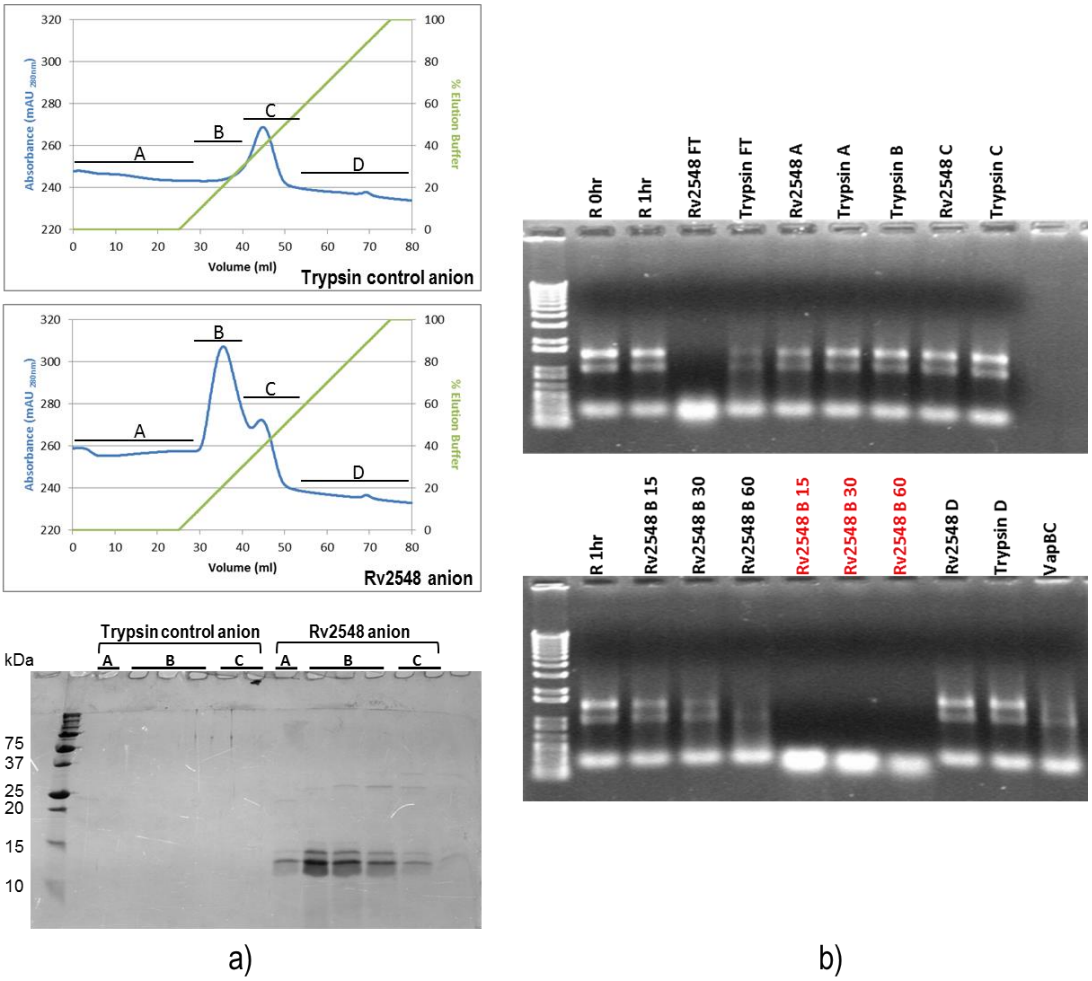


Figure 5.21: Purification of VapC via anion exchange removes the RNase contamination acquired from trypsin. a) Purification of Rv2548 and a control trypsin + trypsin inhibitor sample via anion exchange. Chromatograms depict the UV absorbance (blue) and elution profile of the protein complex as the concentration of elution buffer increases (green). The corresponding SDS-PAGE gel depicts the flow through (FT) and resulting anion exchange fractions (elution positions are shown by the black bar) containing VapC at approximately 17.4 kDa for the Rv2548 anion samples, and nothing for the trypsin control anion samples. b) Total RNA RNase assays on samples with and without anion exchange purification. RNase assays were set up using *M. smegmatis* total RNA as a substrate. Negative controls are shown by R 0hr and 1hr (RNA only) and VapBC, while time points represent the assay time course (60 min if not specified). Rv2548 in black = protein purified via anion exchange or red = crude (without anion exchange purification), Trypsin = mock trypsin + trypsin inhibitor control purified via anion exchange. The contaminating RNase is removed via anion exchange purification as observed by activity in the Rv2548 FT sample, and the reduction in activity of the purified Rv2548 samples compared to crude assay time point samples.

5.2.5 VapC Ribonuclease Activity

Once the *M. tuberculosis* Rv2548 and Rv2530c VapC proteins had been successfully purified, the next step was to test them for general ribonuclease activity. Multiple VapC toxin proteins have been shown to cleave ribosomal RNA *in vitro* (Daines *et al.* 2007; Hurley & Woychik 2009; Ramage *et al.* 2009; McKenzie *et al.* 2012b), and recently *in vivo* activity has also been demonstrated with *E. coli* MazF targeting 16S rRNA within the 30S ribosomal subunit (Vesper *et al.* 2011). Total RNA screens for general ribonuclease activity of *M. tuberculosis* VapC toxins *in vitro* were followed with determination of the VapC ribonuclease sequence-specificity, using screens developed previously in our laboratory for *M. smegmatis* ($VapC_{MS1284}$) and *Pyrobaculum aerophilum* ($VapC_{PAE2754}$ and $VapC_{PAE0151}$) (McKenzie 2011). Here, pentaprobe RNA was used as a substrate for VapC, and MALDI-TOF MS employed to analyze the resulting cleavage products and thus determine the RNA cleavage sites (Kwan *et al.* 2003).

Previous work on *M. smegmatis* and *P. aerophilum* VapCs demonstrated these endoribonucleases only cleaved ssRNA and did not degrade alternative substrates such as dsDNA, ssDNA or dsRNA (McKenzie 2011).

5.2.5.1 General RNase Activity Screening with *M. smegmatis* Total RNA

Rv2548 and Rv2530c both displayed general ribonuclease activity against *M. smegmatis* total RNA, degrading 23S rRNA and 16S rRNA to smaller products during the time course assay (Figure 5.22a and b). Both *M. tuberculosis* VapCs also demonstrated Mg^{2+} dependent activity as when EDTA was included in the assay buffer, ribonuclease activity was abolished. Inclusion of multiple control samples indicated no ribonuclease contamination of buffers or protein preparations was present as no RNA degradation was observed in these samples. RNA 0 hr and 1 hr samples indicate reagents were free of contaminants, the VapBC control sample implies the purification process up until the VapBC stage didn't introduce any RNase contaminants as well as shows that VapB inhibits the ribonuclease activity of VapC, and the EDTA sample shows no non metal-dependent RNases such as ribonuclease A were present. Ribonuclease activity of Rv2548/7 VapC and VapBC proteins was determined after storage at -80 °C to test for retention of RNase activity. Figure 5.22c and d shows that freezing the protein

does not appear to affect the VapC's RNase activity, nor the ability of the VapBC complex to inhibit RNase activity, on both total *M. smegmatis* RNA and pentaprobe 932 RNA.

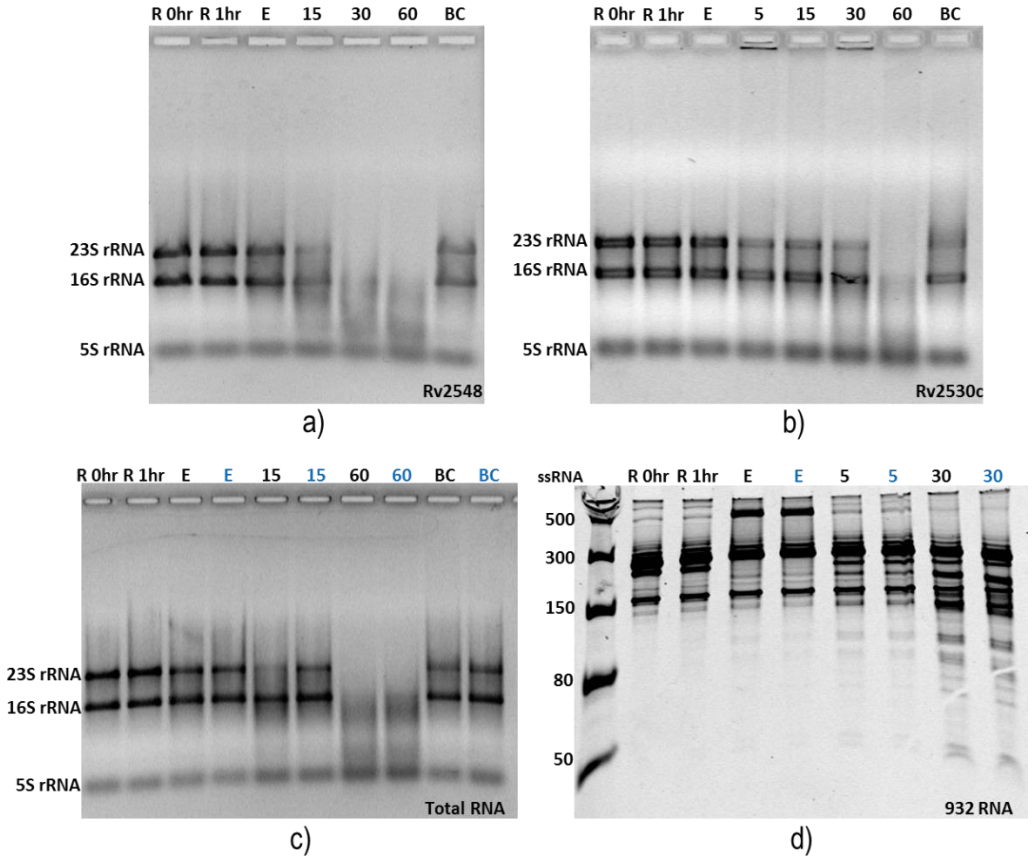


Figure 5.22: Time course assays showing ribonuclease activity of *M. tuberculosis* a) $VapC_{Rv2548}$ and b) $VapC_{Rv2530c}$ using *M. smegmatis* total RNA as a substrate, and frozen $VapC_{Rv2548}$ on c) *M. smegmatis* total RNA and d) pentaprobe 932 RNA as substrates. Negative controls are shown by R 0hr and 1hr (RNA only), E (EDTA) and BC (VapBC), time points represent the assay time course, black indicates non-frozen protein and blue indicates freeze/thawed protein.

5.2.5.2 VapC RNase Activity Assays on Pentaprobe RNA

After the VapC proteins had displayed general ribonuclease activity on the total RNA substrate, they were tested against pentaprobe RNA's (Kwan *et al.* 2003) to try and determine their specificity. Pentaprobe RNA's (containing every combination of five bases) were used as substrates in VapC ribonuclease activity assays, allowing screening of up to pentad recognition sequences. All six pentaprobe RNA's (from the forward direction) as well as 932 (the reverse complement of 926) were tested as substrates for VapC RNase activity. 932 was also included because previous assays had found $VapC_{MS1284}$ cleaved pentaprobe RNA 926

and 932 the most efficiently out of the pentaprobe RNAs (McKenzie 2011). Figure 5.23 and Figure 5.24 show $VapC_{Rv2548}$ and $VapC_{Rv2530c}$ display Mg^{2+}-dependent ribonuclease activity against a variety of pentaprobe RNA molecules. Sequence specificity is shown by the different banding patterns produced by the cleaved RNA fragments for each of the pentaprobe substrates. Both $VapC_{Rv2548}$ and $VapC_{Rv2530c}$ show similar cutting efficiencies to each other against all of the pentaprobe substrates. The lack of one pentaprobe substrate being cut preferentially over the others suggests that none of these are the 'ideal' substrate for either of the VapC proteins, possibly due to unsuitable RNA sequence and/or secondary structure. The larger size of RNA substrates 926 and 932 compared to the others is due to the different restriction enzymes used in the cloning process.

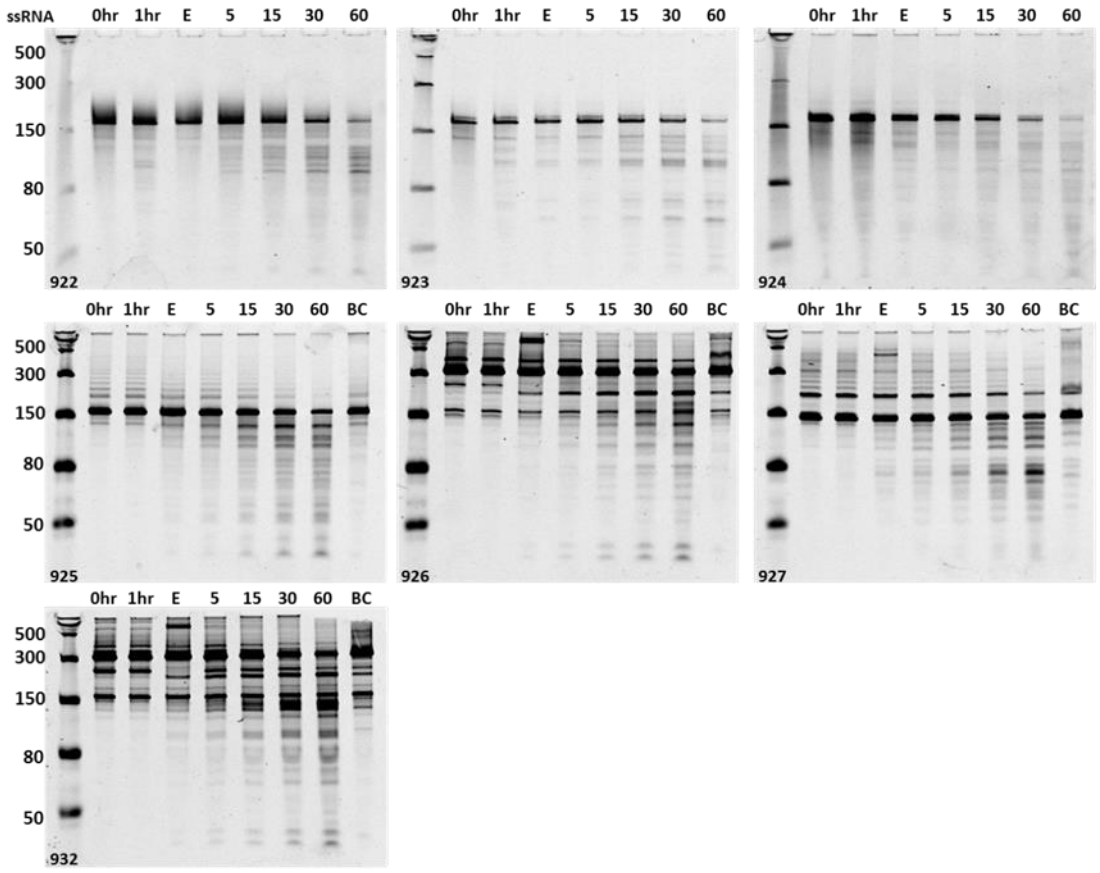


Figure 5.23: Ribonuclease Activity of $VapC_{Rv2548}$ against pentaprobe RNA substrates. Numbers at bottom left corner of gels denote the pentaprobe insert (932 is the reverse complement of 926). 0 hr and 1 hr = RNA only negative controls, numbers = the time at which the sample was taken in the assay (minutes), E = EDTA negative control, BC = VapBC. Numbers down the side of the 10 % urea-denaturing gels denote the low range ssRNA ladder base sizes. The multiple bands observed for 926, 927 and 932 RNA are due to the larger amount of secondary structure in these RNA molecules compared with the others.

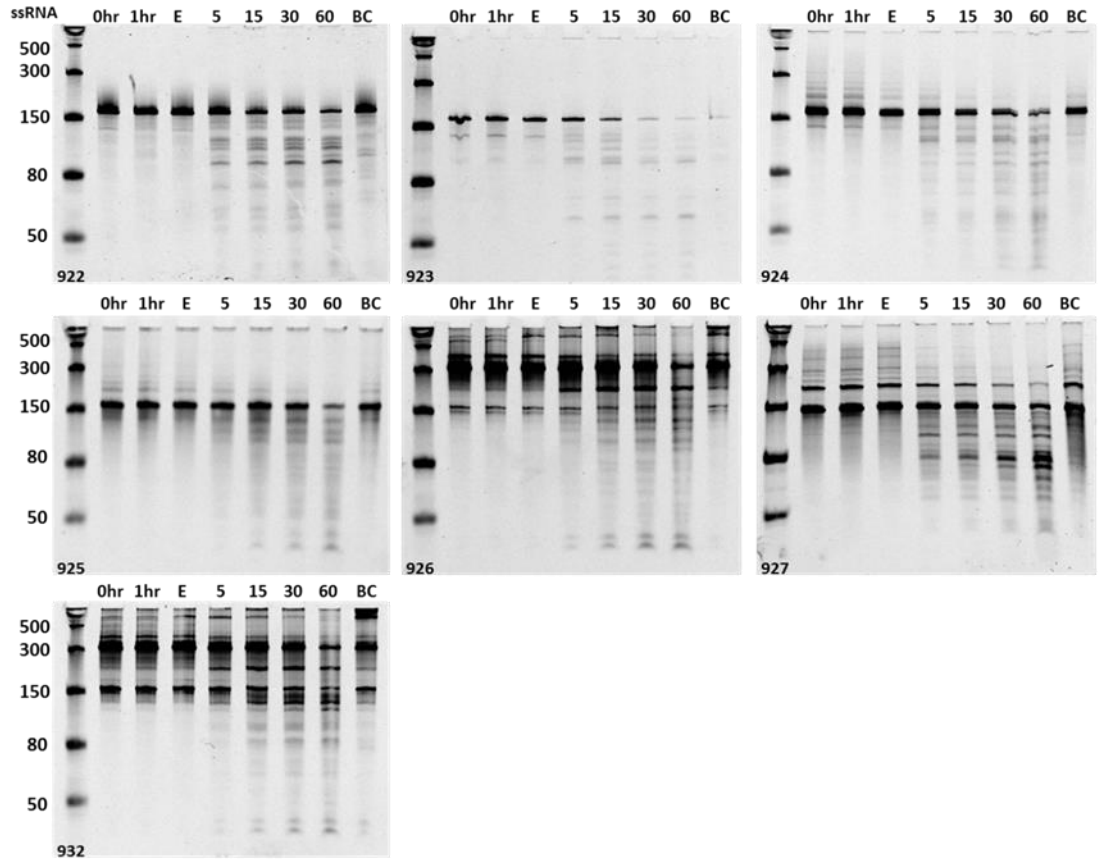


Figure 5.24: Ribonuclease Activity of VapC$_{Rv2530c}$ against pentaprobe RNA substrates. Numbers at bottom left corner of gels denote the pentaprobe insert (932 is the reverse complement of 926). 0 hr and 1 hr = RNA only negative controls, numbers = the time at which the sample was taken in the assay (minutes), E = EDTA negative control, BC = VapBC. Numbers down the side of the 10 % urea-denaturing gels denote the low range ssRNA ladder base sizes. The multiple bands observed for 926, 927 and 932 RNA are due to the larger amount of secondary structure in these RNA molecules compared with the others.

Based on the gels in Figure 5.23 and Figure 5.24, the recognition sequences of VapC$_{Rv2548}$ and VapC$_{Rv2530c}$ are likely to be the same, as the fragment patterns of the cut RNA for each protein look similar. The same banding pattern was also observed for two other *M. tuberculosis* VapCs (VapC$_{Rv0065}$ and VapC$_{Rv0617}$) previously (McKenzie *et al.* 2012a), suggesting the possibility that *M. tuberculosis* VapCs target the same recognition sequence. Previous work found VapC$_{PAE0151}$ and VapC$_{PAE2754}$ were also sequence-specific and targeted the same sequence as each other for the 924 pentaprobe substrate, but this sequence was different to that which the *M. tuberculosis* VapCs targeted on the same 924 substrate (McKenzie *et al.* 2012a). The sequence-specificity of VapC$_{PAE0151}$, VapC$_{PAE2754}$, VapC$_{Rv0065}$, VapC$_{Rv0617}$, VapC$_{Rv2548}$ and VapC$_{Rv2530c}$ also differ from that of VapC$_{MSMEG_1284}$, which does not cut the 924 Pentaprobe RNA (McKenzie *et al.* 2012a).

Interestingly, as with *P. aerophilum* VapCs, no binding to the RNA was seen with the *M. tuberculosis* VapCs, contrary to that observed with VapC from *M. smegmatis*. $VapC_{MS1284}$ appeared to bind the RNA which it cleaved (evidenced by smearing on the gels), but not to RNA it did not cleave, nor when EDTA was present, or when VapB was in complex with VapC (McKenzie 2011; McKenzie *et al.* 2012b). Therefore, while binding and cleavage may be separable properties of VapC in *M. smegmatis*, this is not the case with VapC proteins from *M. tuberculosis* or *P. aerophilum*.

5.2.5.3 Determination of Rv2530c VapC Cut Sites

$VapC_{Rv2548}$ and $VapC_{Rv2530c}$ appeared to cleave all pentaprobe RNA substrates with equal efficiency, so overlapping 932 RNA oligonucleotides (with similar secondary structure to the original 932 RNA molecule) were chosen as substrates for determining sequence specificity of $VapC_{Rv2530c,}$ as these had been used previously in determining the sequence specificity for $VapC_{MS1284}$, $VapC_{PAE2754}$ and $VapC_{PAE0151.}$ 932 RNA oligonucleotide design and sequences can be found in Appendices B.3 and B.4. VapC ribonuclease activity assays on 932 RNA oligos were performed, purified and analysed with MALDI-TOF MS to identify the cleaved fragments, then target recognition sites were determined using in-house software which identifies all one and two point cuts in the oligonucleotide as in Sections 2.3.4.4 - 2.3.4.6.

5.2.5.3.1 MALDI-TOF Mass Spectrometry of 932 RNA Oligonucleotide Ribonuclease Assays with $VapC_{Rv2530c}$

Raw data can be found on the compact disc - Appendix D.

Mass spectra display the mass to charge (*m/z*) values of the ions plotted against their intensity to portray the number of components in the sample, their size, and their relative abundance in the gas phase (this is not usually representative of abundance in solution). In MALDI-TOF, ions are nearly always singly charged species (so the *m/z* measurement is equal to the mass of the ionized molecule), although they can be accompanied by salt adducts, doubly charged

ions (at approximately half the *m/z* value), or traces of dimeric species at approximately twice the *m/z* value.

All 932 RNA oligonucleotides were cut by $VapC_{Rv2530c}$ over the assay time course, some better than others. Spectra show no degradation of oligos in the RNA only negative controls (0 hr and 1 hr) but the EDTA samples were not useful as negative controls as these showed some degree of activity. Oligonucleotide fragments four bases or less could not be detected due to the suppression of matrix peaks in this range (matrix suppression was 1000 Da). Multiple 'noise' peaks in the lower *m/z* range were observed for oligo 2, 3 and 8 controls indicating these oligos contain multiple species formed during the manufacturing process, even though they were HPLC purified. These 'noise' peaks can mask the true cleavage peaks formed during RNase assays unless they are of sufficient intensity to rise above the noise, so are not ideal substrates for determining sequence specificity of $VapC_{Rv2530c}$. Figure 5.25 - Figure 5.32 display the mass spectra results for $VapC_{Rv2530c}$ ribonuclease activity on 932 RNA oligonucleotides 1 - 8.

Oligo 1 was not cut well. Only one small peak was observed in the spectra (*m/z* 7631.8) after 15 minutes which, without an additional 'partnered' peak (adding up to the total mass of the oligo), suggests it was not produced as a result of just a single cleavage event, so may not be a primary/major cut site. Two additional peaks of low intensity (*m/z* 5752.7 and 8263.7) were not produced until 60 minutes, again suggesting these were not primary target cut sites due to the time taken for these to be formed.

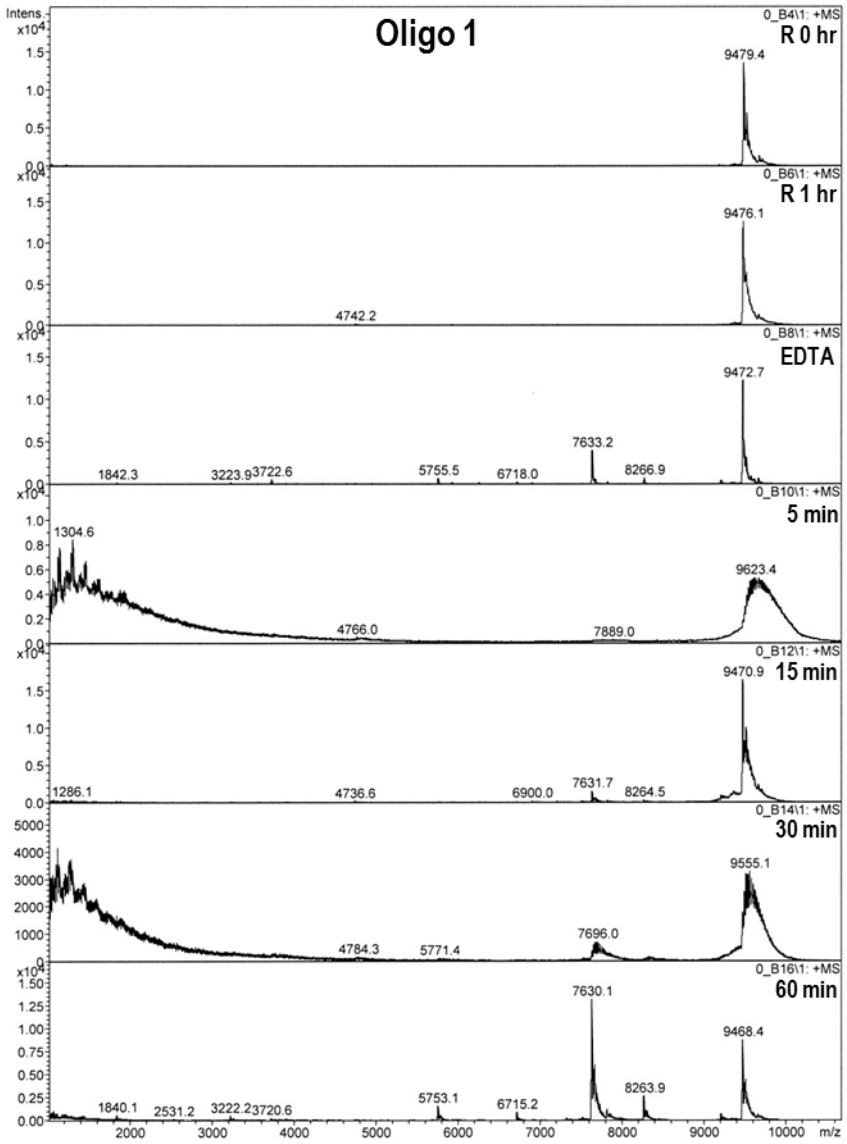


Figure 5.25: 932 RNA Oligo 1 VapC$_{Rv2530c}$ MALDI-MS results. R 0 hr and 1 hr negative refer to assay reactions with no VapC protein; EDTA, inclusion of 20 mM EDTA in the reaction buffer, and time course (15 to 60 minute) assay samples with VapCRv2530c. Data analysed by Bruker Data Analysis.

As mentioned above, the data for oligo 2 were noisy so peaks resulting from cleavage of the oligonucleotide were hard to distinguish from the background. As with the peaks produced for the oligo 1 assays, the only peaks produced for oligo 2 at 30 and 60 minutes (with intensities which were above the background (*m/z* 4512 and 5146)) were un-partnered so unlikely to be products from a single cleavage event, and therefore not from a primary target cut site.

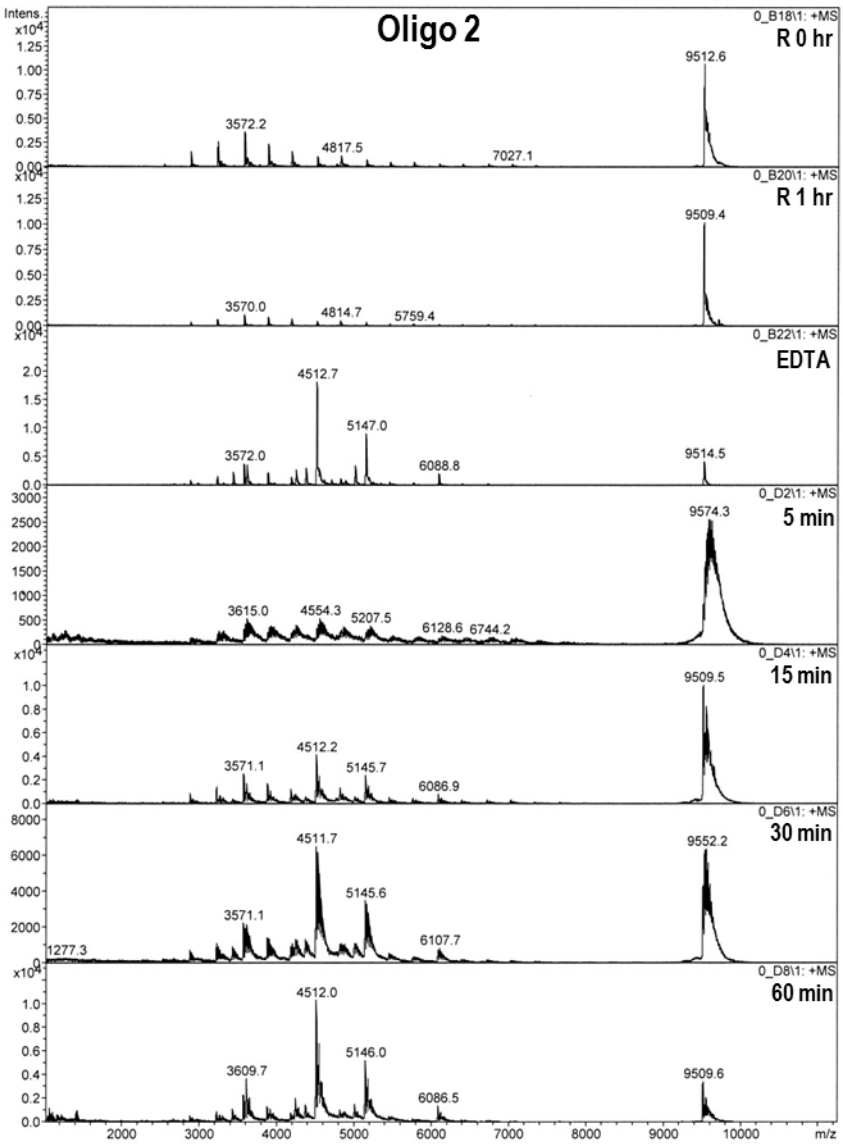


Figure 5.26: 932 RNA Oligo 2 VapC$_{Rv2530c}$ MALDI-MS results. R 0 hr and 1 hr negative refer to assay reactions with no VapC protein; EDTA, inclusion of 20 mM EDTA in the reaction buffer, and time course (15 to 60 minute) assay samples with VapCRv2530c. Data analysed by Bruker Data Analysis.

Oligo 3 was also noisy so peaks resulting from cleavage of the oligonucleotide were hard to distinguish from the background. Peaks with intensities above background were visible at 15 minutes (*m/z* 2669, 3304, 4268 and 6530), although only peaks of *m/z* 4268 and 6530 were partnered totalling the whole oligo mass and therefore likely to be products from a single cleavage event.

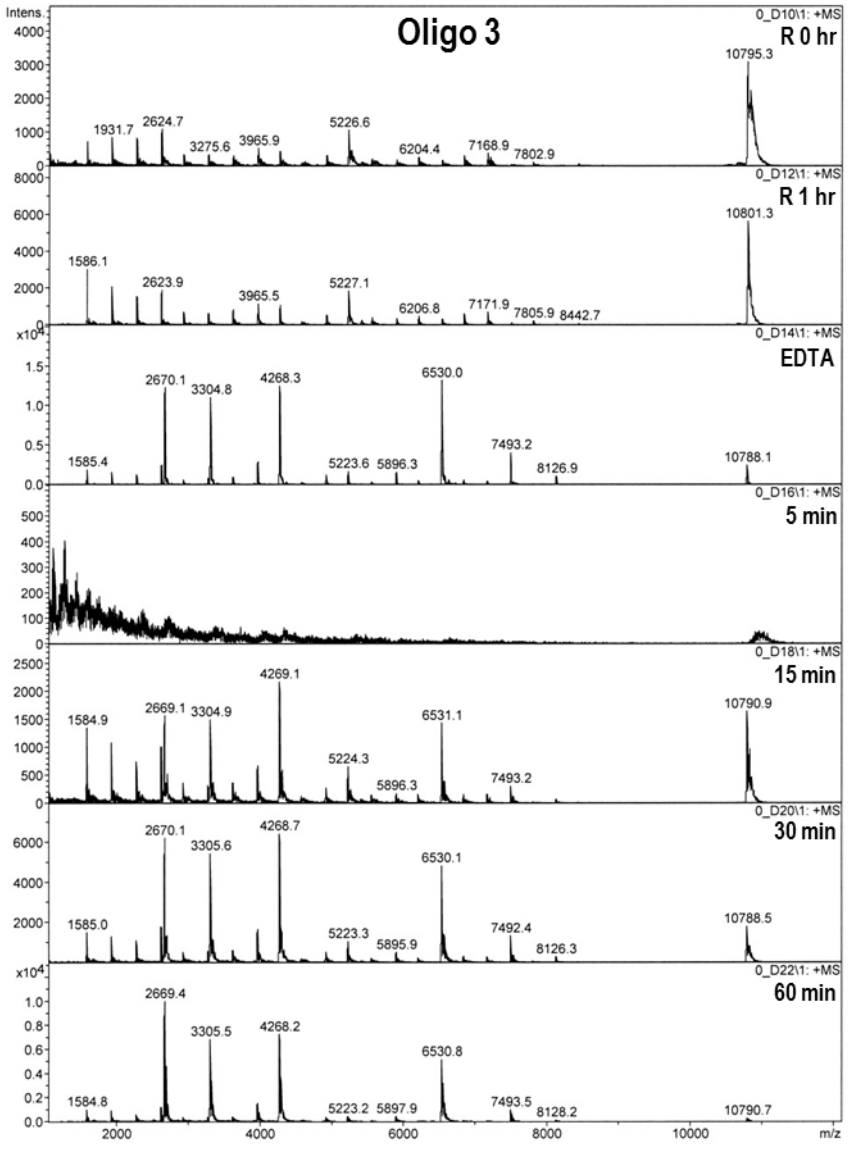


Figure 5.27: 932 RNA Oligo 3 VapC$_{Rv2530c}$ MALDI-MS results. R 0 hr and 1 hr negative refer to assay reactions with no VapC protein; EDTA, inclusion of 20 mM EDTA in the reaction buffer, and time course (15 to 60 minute) assay samples with VapCRv2530c. Data analysed by Bruker Data Analysis.

The majority of oligo 4 was degraded by 60 minutes, although the resulting major peaks formed (*m/z* 2299, 3914 and 5065) were all of lower masses which couldn't be formed from the 11531 MW oligo with just a single cut. Therefore, these were likely to be products formed from multiple subordinate cut sites.

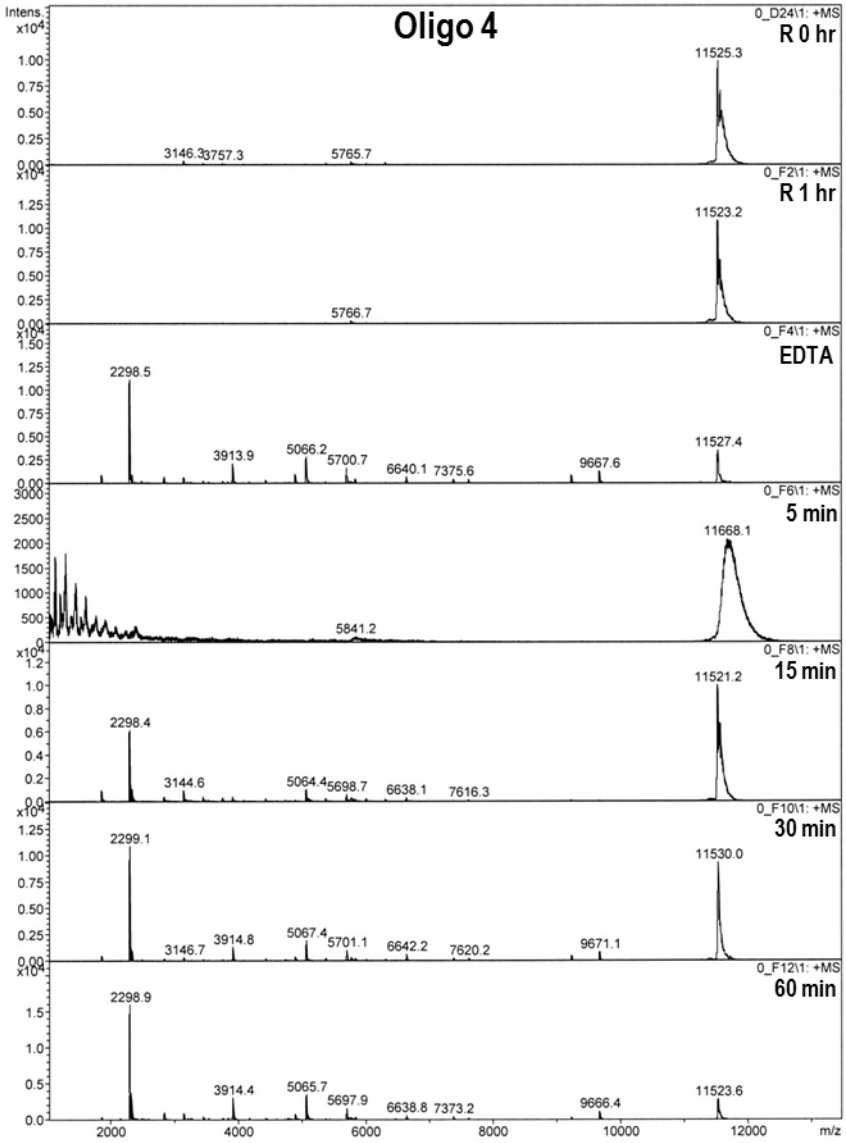


Figure 5.28: 932 RNA Oligo 4 VapC$_{Rv2530c}$ MALDI-MS results. R 0 hr and 1 hr negative refer to assay reactions with no VapC protein; EDTA, inclusion of 20 mM EDTA in the reaction buffer, and time course (15 to 60 minute) assay samples with VapCRv2530c. Data analysed by Bruker Data Analysis.

Oligo 5 was the best substrate tested for $VapC_{Rv2530c}$ as it was cut the fastest (cleavage fragments were visible after five minutes) at what appeared to be two separate major single cut sites, each producing two 'partnered' fragments (*m/z* 6774.6 + 3181.9 and *m/z* 3473 + 6483.7) which added up to the mass of the whole RNA oligo (*m/z* 9957). Two additional peaks were also visible (*m/z* 8364.6 + 1593.2) but these were not as abundant. Cleavage products of 932 RNA oligo 5 observed in the 5 minute assay sample were also present at 15, 30 and 60 minutes, but no additional cleavage fragments were observed at these time points. By 60 minutes, the majority of the RNA oligo was degraded.

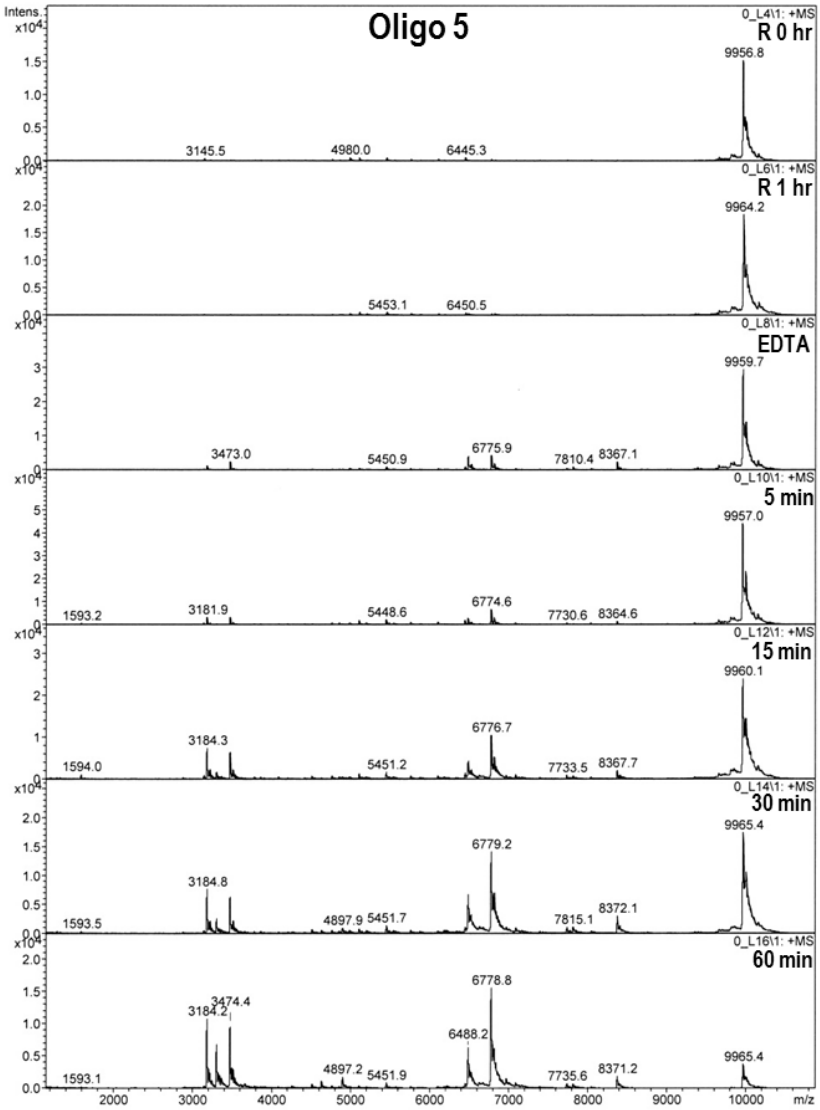


Figure 5.29: 932 RNA Oligo 5 $VapC_{Rv2530c}$ MALDI-MS results. R 0 hr and 1 hr negative refer to assay reactions with no VapC protein; EDTA, inclusion of 20 mM EDTA in the reaction buffer, and time course (15 to 60 minute) assay samples with VapCRv2530c. Data analysed by Bruker Data Analysis.

Oligo 6 was cut by Rv2530c but the peaks were hard to distinguish from background until 60 minutes due to the noise, by which time products from initial cleavage events may have been cut further masking the major cleavage event fragments.

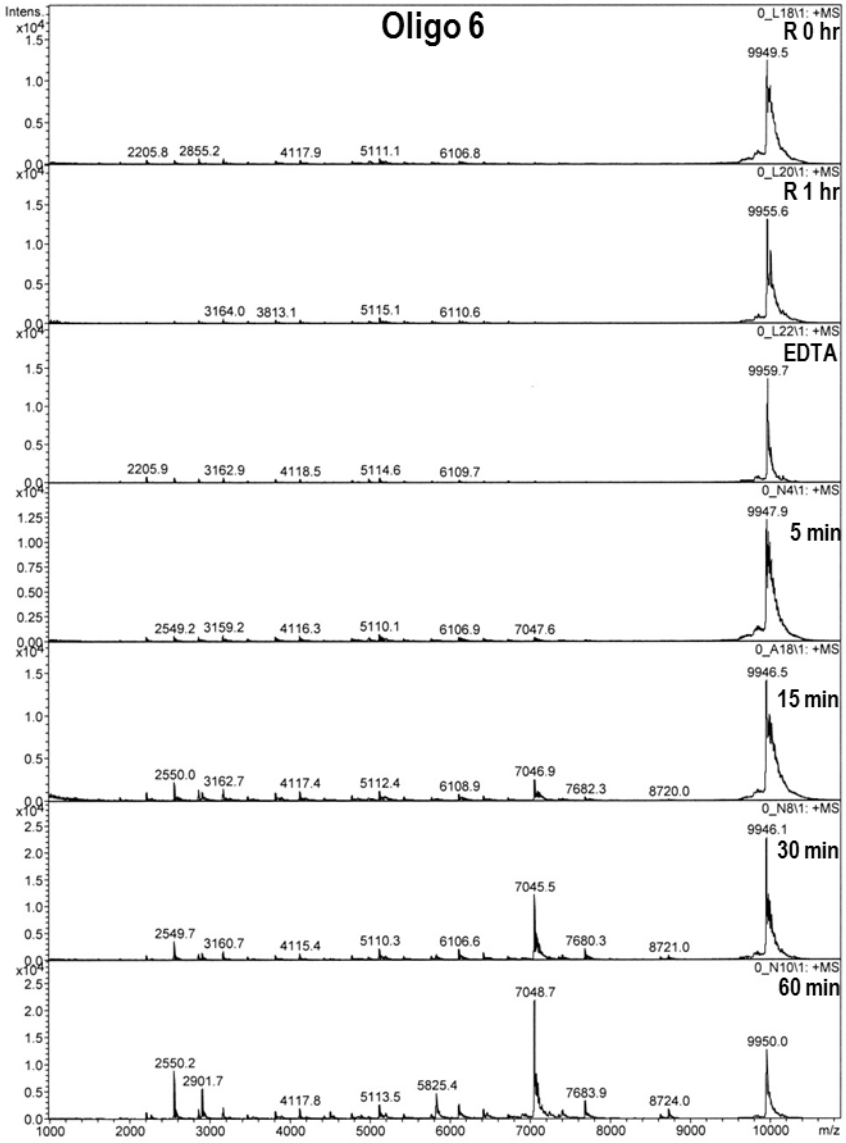


Figure 5.30: 932 RNA Oligo 6 VapC$_{Rv2530c}$ MALDI-MS results. R 0 hr and 1 hr negative refer to assay reactions with no VapC protein; EDTA, inclusion of 20 mM EDTA in the reaction buffer, and time course (15 to 60 minute) assay samples with VapCRv2530c. Data analysed by Bruker Data Analysis.

Oligo 7 was cut marginally by 15 minutes at two separate single cut sites, each producing partnered fragments (*m/z* 4110.6 + 5758.3 and 4747 + 5121.7). These peaks were low in intensity, and given the whole oligo was still largely intact were unlikely to be products from a major optimal cut site.

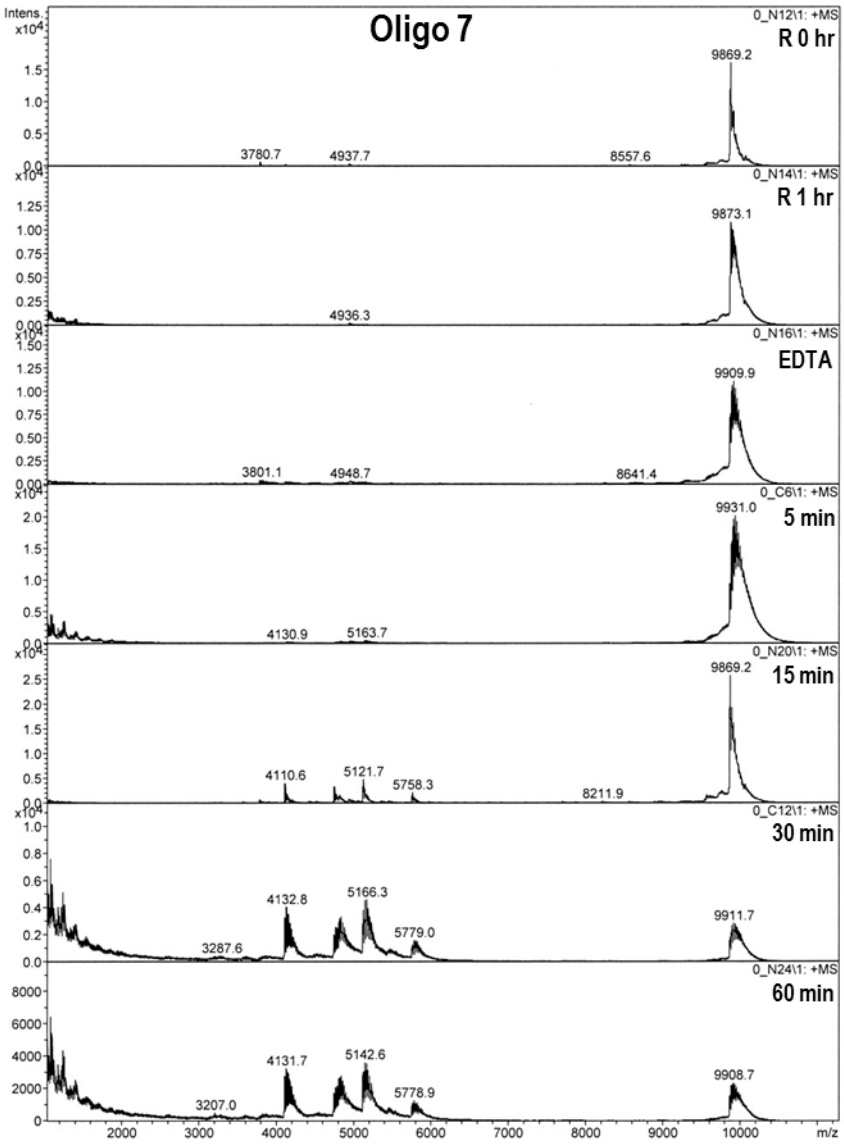


Figure 5.31: 932 RNA Oligo 7 VapC$_{Rv2530c}$ MALDI-MS results. R 0 hr and 1 hr negative refer to assay reactions with no VapC protein; EDTA, inclusion of 20 mM EDTA in the reaction buffer, and time course (15 to 60 minute) assay samples with VapCRv2530c. Data analysed by Bruker Data Analysis.

Oligo 8 was also cut but peaks were hard to distinguish from background until 60 minutes due to the noise. A large majority of the oligo was degraded by 60 minutes, although the resulting major peaks formed (*m/z* 2550.8 and 2917.9) were all of lower masses which couldn't be formed from the whole oligo with just a single cut. These were likely to be products formed from multiple subordinate cut sites.

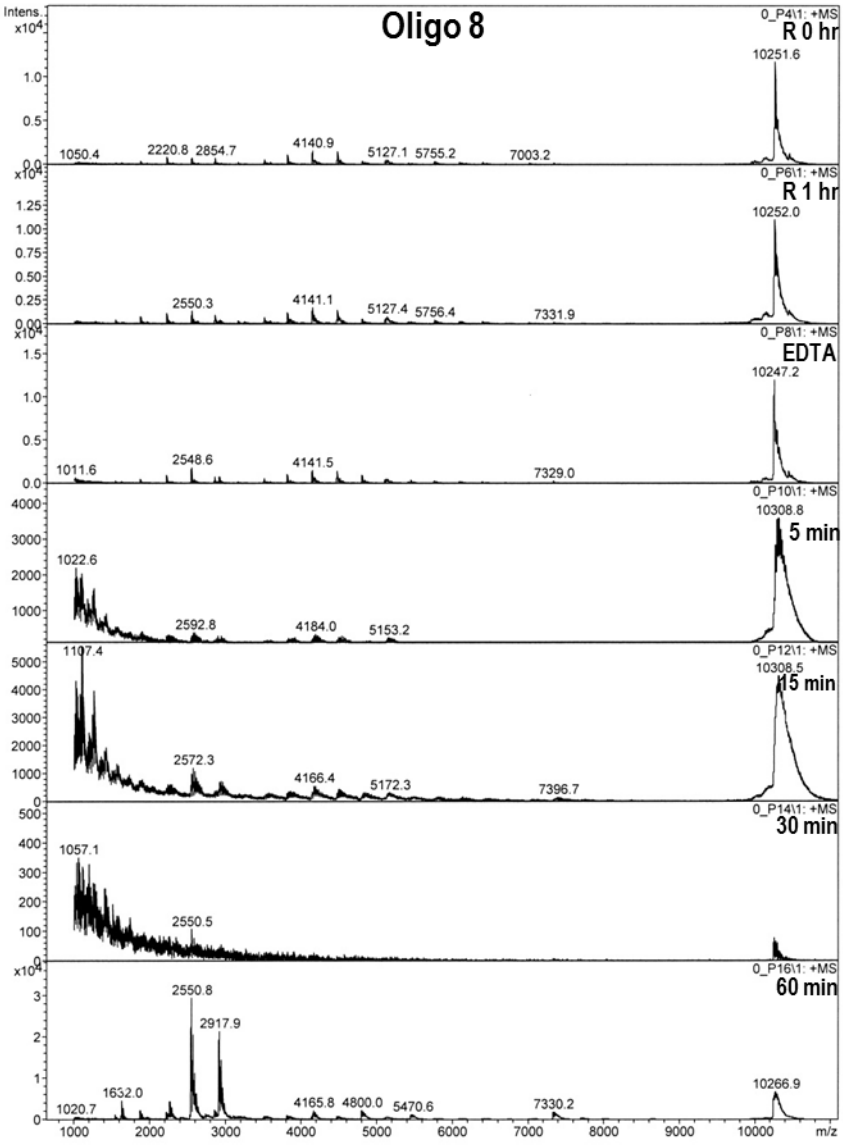


Figure 5.32: 932 RNA Oligo 8 VapC$_{Rv2530c}$ MALDI-MS results. R 0 hr and 1 hr negative refer to assay reactions with no VapC protein; EDTA, inclusion of 20 mM EDTA in the reaction buffer, and time course (15 to 60 minute) assay samples with VapCRv2530c. Data analysed by Bruker Data Analysis.

Resulting peaks in the mass spectra from $VapC_{Rv2530c}$ ribonuclease activity on 932 oligonucleotide substrates (above) and from those performed previously for $VapC_{Rv0065}$ and $VapC_{Rv0617}$ were almost identical. Therefore, the three *M. tuberculosis* VapC proteins appear to be targeting the same sequence. Spectra peaks for identical 932 oligonucleotide substrates from *M. tuberculosis* VapC proteins ($VapC_{Rv2530c,}$ $VapC_{Rv0065}$ and $VapC_{Rv0617}$) are different to those obtained for *P. aerophilum* ($VapC_{PAE0151}$, $VapC_{PAE2754}$) indicating that these species target different sequences.

5.2.5.3.2 Analysis of 932 RNA Oligonucleotide Ribonuclease Assays and $VapC_{Rv2530c}$ Cut Site Determination

Mongo Oligonucleotide Calculator V2.06 (Rozenski 1999) (which allows the addition of 5' and 3' modifications to RNA oligonucleotides) was used to determine theoretical masses of RNA oligonucleotide fragments to compare to masses from the MS spectra. It was determined that $VapC_{Rv2530c}$ cleaved RNA in the same manner as $VapC_{PAE2754}$, $VapC_{PAE0151}$, $VapC_{Rv0065}$, and $VapC_{Rv0617}$ which all cleave the 3' phosphodiester bond of RNA to produce 3' hydroxyl and 5' phosphate cleavage products, consistent with other metal-dependent ribonuclease enzymes (Tadokoro & Kanaya 2009; McKenzie *et al.* 2012a). This is in contrast to the reported mechanism of the MazF toxin, as well as that for $VapC_{pMYSH6000}$ from *S. flexneri* and $VapC_{LT2}$ from *Salmonella enterica* which cleave the phosphodiester bond yielding a free 5' OH group on the 3' cleavage product, and 2'-3' cyclic phosphate on the 3' end of the 5' cleavage product, similar to RNase H (Muñoz-Gómez *et al.* 2005; Zhang *et al.* 2005; Winther & Gerdes 2011).

The mass list and corresponding intensities from the spectra were exported into a comma separated excel file then processed using in-house software to determine possible cut sites. The software determines the difference between the predicted mass of the oligonucleotide and the actual peak mass (as measured by MALDI-TOF MS) for all one and two point cuts, including a 5' phosphate on the 3' cleavage product. Different parameters can be selected by setting a threshold cut-off for intensity values (so only peaks above a certain intensity level will be included for analysis), and by selecting fragments with the closest matching predicted and

actual masses. Any products containing less than four bases do not fly well resulting in nonsense MW values, therefore their size is disregarded during analysis.

Logic dictates that if the optimal cut site is present in a substrate, it would be cut preferentially over other subordinate cut sites almost immediately, and produce just two fragments. If no 'optimal' cut site were present, then it is likely that multiple fragments could be produced from many subordinate cut sites over the duration of the ribonuclease assays. Initially, results for all one point cuts were analysed to check they accounted for the major peaks in the MALDI-TOF MS spectra and labelled as probable major cut sites, and if no major peaks could be matched to single cut sites then all two point cuts were analysed and labelled as probable minor cut sites. Any spectra peaks unable to be matched were deemed to be products from more than two cuts and therefore not from optimal cut sites.

Once all major single and two point cut possibilities were identified, the orientation of the fragments and therefore the sequence of the cut site was determined. Results from most major cuts produced candidates with only one possible fragment orientation combination (eg, all candidates having fragment x on the left and y on the right), but where two different fragment orientation combinations were present (x/y and y/x), the orientation with the greatest number of candidates was picked, and if there was an equal number of candidates for each combination then both cut sites were proposed as putative sites (cut sites summarised in Table 5.5).

The intensity threshold was set low to 800 (or higher if back ground noise was present), so as to include peaks from the spectra which may be being produced from a major cut site and simultaneously cleaved at a subordinate site which would be keeping their relative intensity low.

Table 5.5: Summary of major cut sites for $VapC_{Rv2530c}$ on 932 RNA oligonucleotides 1 – 8.

Oligo	Major cut site target sequence	Time cut site detected
Oligo 1	None	N/A
Oligo 2	Complex activity on this substrate	N/A
Oligo 3	AU*AU or GC*CG	15 minutes
Oligo 4	None	N/A
Oligo 5	UA*GG and UA*GG	5 minutes
Oligo 6	Complex activity on this substrate	N/A
Oligo 7	UC*UA	15 minutes
Oligo 8	Complex activity on this substrate	N/A

* denotes cleavage site.

For oligo 5, all four major peaks could be matched to two single cuts for the 5 minute assay (*m/z* 6774.6 + 3181.9 and *m/z* 3473 + 6483.7). Both of these single cuts occur at the same sequence, UA*GG (* denotes cleavage site). This major cut site sequence occurs only twice in all of the oligo substrates, both of which are present in oligo 5. This appears to be the most optimal substrate tested for $VapC_{Rv2530c}$ as it is cut the fastest producing the most conclusive cut site. Two minor peaks were also able to be matched to a single cut in the 15 minute assay (*m/z* 8367.7 + 1592.7), AU*CA.

For oligos 3 and 7, only two of the major peaks could be matched to a single cut for each of the 15 minute assays (*m/z* 4268 + 6530, and *m/z* 4747 + 5121.7 respectively), and the orientation of fragments could not be conclusively determined for oligo 3 so both candidates are postulated as major cut sites (see Table 5.5). No major peaks from oligos 1 and 4 could be matched to cut sites from a single cut, but one major peak (m/z 7631.8) from oligo 1 was found in a two point cut for the 15 minute assay with the minor cut site sequence of UC*AC. No major peaks from spectra for oligos 2, 6 or 8 could be matched to cut sites from single or two point cuts for oligos 2, 6 and 8 once the intensity threshold was raised above the background noise. It was assumed oligos 2, 6 and 8 did not contain any major cut sites for $VapC_{Rv2530c}$.

As oligo 5 appeared to contain the most optimal cut site for $VapC_{Rv2530c}$ (UA*GG), the MALDI-TOF MS data analysis method focussing on only single cuts described above, was applied to

oligo 5 spectra data obtained previously for $VapC_{Rv0065}$ and $VapC_{Rv0617}$ to see if these *M. tuberculosis* VapC proteins may in fact target this same sequence, and not GC-rich 4mers as previously thought (McKenzie 2011; McKenzie *et al.* 2012a). MS data for these assays had been obtained during a period of poor laser performance from the MALDI-TOF MS, and produced spectra not ideal to work from. Spectra showed substantial noise with broad peaks, but some major peaks could be matched to single cuts for the 30 minute assays. Of the four ambiguous single cut sites, UA*GG appeared twice which is the same as that obtained for $VapC_{Rv2530c}$. This suggests that UA*GG is the preferred cut site for the three *M. tuberculosis* VapC proteins tested thus far. Repeat oligo 5 RNase assays should be performed again on $VapC_{Rv0065}$ and $VapC_{Rv0617}$ once the laser is fully functional, to confirm UA*GG is the preferred cut site for these two VapC RNases as well as for $VapC_{Rv2530c}$. The reason different cut sites were obtained for the two VapCs from the same data is likely due to the analysis technique applied. Here, only one point cuts were considered (two point cuts were only included if no one point cuts were identified), and only the predominant orientation was included, rather than taking all one and two point cuts and every combination of their orientation, as was done in the previous analysis (McKenzie 2011). This analysis technique seems a more robust and logical method for determining the major or preferred cut site, as if the optimal cut site was present in a substrate, it should be cut preferentially over other suboptimal sites and produce just two fragments.

The *M. tuberculosis* VapC RNases appear to have a different recognition sequence to VapCs from both *P. aerophilum* (targeting G rich sequences) and *M. smegmatis* (targeting AU sequences) (McKenzie *et al.* 2012a; McKenzie *et al.* 2012b). Confirming exactly which target sequence residues are crucial for the RNase activity of the *M. tuberculosis* VapCs, will be instrumental in determining their optimal target sequence, and therefore the pathways and roles they may perform. One way of achieving this could be performing a microarray to identify genes which are downregulated in response to expression of a particular *M. tuberculosis* VapC in its VapBC knockout, as was the case with VapC from *M. smegmatis* (McKenzie *et al.* 2012b).

5.2.5.3.3 Bioinformatic Analysis of the $VapC_{Rv2530c}$ Target Sequence

The *M. tuberculosis* H37Rv genome was analysed to determine the frequency of the $VapC_{Rv2530c}$ target sequence, TAGG (UAGG). This analysis included a 100 bp region upstream of all genes, because cleavage of the stable stem loop structure of 5' untranslated regions (UTR) could leave the mRNA more susceptible to general nuclease degradation, as well as destabilise mRNA secondary structure, and possibly improving the translation efficiency of these genes (Rocha *et al.* 1999; Gu *et al.* 2010). TAGG is present in 2,435 of the 4,111 protein coding genes (or their upstream region) (see Table 5.6) and is the 40th rarest quartet (out of 256 possible quartets) in the H37Rv genome, appearing a total of 5,171 times (the least common quartet is TTAA which appears 1,271 times, and the most common is GCCG which occurs 79,400 times).

Table 5.6: Prevalence of $VapC_{Rv2530c}$ target sequence in genes and their upstream regions in the *M. tuberculosis* H37Rv genome.

Region	Number of genes or upstream sequences with TAGG	Number of times TAGG occurs
Upstream (U)	709	781
Gene (G)	2067	4260
U + G	2435	5041

Genes and upstream sequences refer to the number of genes or upstream regions the VapC target sequence is present in. Times TAGG occurs refers to the total number of times the VapC target sequence is present in the genes or upstream region. Total number of H37Rv protein coding genes analysed was 4,111, including a 100bp region upstream of each.

Bioinformatic analysis of the *M. tuberculosis* H37Rv genome revealed an abundance of hypothetical proteins and genes involved in metabolic processes containing the TAGG $VapC_{Rv2530c}$ target sequence. This was observed when the number of target sequences contained in either the gene and/or upstream region was ranked from greatest to fewest, and occurred in ever higher frequency when ranked according to the number of target sequences relative to size of the gene (Table 5.7), so genes with an equal number of TAGG sequences will appear earlier in the rankings if they are smaller. The majority of genes with a greater relative number of target sequences were annotated as hypothetical proteins, but also included those annotated as transposases, TA genes and other genes associated with horizontal gene

transfer (Table 5.8). An intriguing explanation for why $VapC_{Rv2530c}$ is targeting genes involved in HGT could be that it is also functioning as a selfish element – targeting and eliminating the other selfish elements in order to survive the genome war within. The number of TA genes appearing in the top 40 results once the data has been ranked according to relative size is also very interesting (four out of 40), especially since only 88 putative TA genes have been identified in the *M. tuberculosis* H37Rv genome. A possible explanation for why one toxin may target another toxin or antitoxin is to initiate a transcriptional regulation cascade when a rapid change in the proteomic response and metabolic state of the cell is required during conditions of stress. Identifying which genes are being targeted by $VapC_{Rv2530c}$ *in vivo* (possibly by microarray analysis comparing genes up or down regulated in response to its expression, or RNA-seq), would greatly help in the understanding of the biochemical function and target sequence of $VapC_{Rv2530c}$.

Table 5.7: Prevalence of *M. tuberculosis* H37Rv annotated genes containing the $VapC_{Rv2530c}$ target sequence in the top 40 rankings.

Annotation	Upstream (U)	Gene (G)	U + G	Relative size
Hypothetical	15/40	10/40	9/40	23/40
Transposase	0/40	6/40	6/40	6/40
Toxin/Antitoxin	0/40	0/40	0/40	4/40

Numbers reflect the top 40 results for the greatest prevalence of TAGG according to their ranking (U, G, U+G or relative size). Genes (G) and upstream (U) sequences refer to the number of genes and/or upstream regions the VapC target sequence is present in; relative size refers to genes ranked (in ascending order) according to the number of target sequences relative to the size of the gene (smaller genes with the same number of target sequences as a larger gene will appear earlier in the ranking).

Table 5.8: Top 20 *M. tuberculosis* H37Rv genes according to number of $VapC_{Rv2530c}$ target sequences present in the gene and upstream region, relative to gene size.

rank	annotation	position	size (bp)	size/count	U	G	U+G
1	hypothetical protein	3081028	105	53	2	0	2
2	hypothetical protein (Rv1950c)	2200005	192	64	0	3	3
3	hypothetical protein	3585673	273	68	2	2	4
4	hypothetical protein	1784305	216	72	0	3	3
5	hypothetical protein	986677	150	75	2	0	2
6	hypothetical protein	545191	231	77	0	3	3
7	glutaredoxin-like protein NrdH	3414730	240	80	0	3	3
8	antitoxin (Rv0456B - MazE1)	547347	174	87	1	1	2
9	hypothetical protein	3841894	363	91	1	3	4
10	phosphoribosyl-ATP pyrophosphatase	2380671	282	94	3	0	3
11	transcriptional regulator (Rv1473A)	1662387	192	96	2	0	2
12	hypothetical protein	3019824	390	98	0	4	4
13	immunogenic protein MPT64	2223351	687	98	1	6	7
14	hypothetical protein (rv2401A)	2698052	204	102	1	1	2
15	hypothetical protein	3594190	102	102	1	0	1
16	phiRv1 phage protein (Rv1579c)	1783316	315	105	1	2	3
17	toxin (Rv2142c - ParE2)	2402201	318	106	0	3	3
18	hypothetical protein	530299	429	107	2	2	4
19	hypothetical protein	1301253	108	108	1	0	1
20	hypothetical protein	2055688	432	108	0	4	4

The top 20 genes are ranked (in ascending order) according to the number of TAGG target sequences present in the combined gene (G) and upstream (U) regions (G+U) relative to gene size (size/count). 'Position' refers to the position of the TAGG sequence in the H37Rv genome and 'size', the number of bp for each gene. Gene names in red show annotated toxin or antitoxin genes, their gene number, and which TA family they belong to.

5.2.6 Crystallisation of Rv2530a/c

The three-dimensional structure of a protein can reveal details about its molecular organisation and interactions with other proteins, as well as its function. Purified Rv2530a/c and Rv2530c proteins were subjected to crystallisation trials in an attempt to gain further insight into their molecular functions. Initial crystallisation trials with purified VapBC proteins was challenging due to the low yield and buffer requirements of these proteins. Many of the crystals obtained

from initial robot crystallisation screens were phosphate salt crystals due to the low solubility of particular compounds in the protein purification buffer (50 mM sodium phosphate pH 8, 200 mM NaCl). In an attempt to reduce crystals forming from the purification buffer, VapBC was dialysed overnight into 50 mM HEPES, MOPS or Tris at pH 7.4 + 200 mM NaCl and observed for precipitation. Heavy precipitation of the VapBC complex was observed upon dialysis into HEPES and MOPS buffers, and a little precipitation was observed in Tris buffer, which indicated the protein was not as stable in these buffers. The crystallisation screens below continued to be performed in the initial sodium phosphate protein purification buffer.

5.2.6.1 Initial Crystallisation Trials

Rv2530a/c VapC and VapBC was purified by IMAC and SEC in 50 mM sodium phosphate buffer pH 7.4, 200 mM NaCl then subjected to robot crystallisation screens (Section 2.3.3.1) at concentrations ranging from 7 to 14.5 mg/ml. VapBC SEC fractions were pooled covering either the whole protein containing population, or split into two populations covering fractions 8 - 14, and 17 – 23 which appeared to have different banding patterns to each other on an SDS-PAGE gel (as seen in Figure 5.10). A number of promising conditions producing either crystals or needles were identified for each of the protein samples although more were obtained for the VapBC complex than VapC by itself (Appendix C.2). Once conditions containing reagents unavailable in the lab, or those containing lithium, calcium, magnesium or high concentrations of ammonium were removed due to the likelihood that the crystals formed would be phosphate salt crystals formed in the presence of the purification buffer, only two candidate conditions remained. Crystals from the candidate conditions 0.2 M ammonium acetate, 0.1 M Bis-tris pH 5.5, 45 % (+/-) -2-Methyl-2,4-pentanediol (MPD) and 0.5 M ammonium sulfate, 0.1 M HEPES pH 7.5, 30 % (+/-) MPD were tested for X-ray diffraction on the Home-Source at the University of Auckland which confirmed that they were protein crystals, then followed up with fine screen testing. No cryoprotectant was required for testing these crystals as the MPD in the crystallisation condition fulfilled this role adequately.

5.2.6.2 Optimisation of Crystallisation Conditions (Fine Screening)

There are various strategies available for optimising crystal formation, the most common being fine-tuning precipitant and protein concentrations, adjusting the pH, and screening additives. All of these fine screening procedures were attempted, as well as many other less common crystal optimisation techniques in order to produce crystals of as higher as possible quality that are suitable for X-ray analysis.

5.2.6.2.1 0.2 M ammonium acetate, 0.1 M BIS-TRIS pH 5.5, 45 % (+/-) MPD

An initial fine screen for Rv2530a/c SEC 8 – 14 (8.13 mg/ml) of 0.1 - 0.5 M ammonium acetate spanning a Bis-tris pH range of 5.3 – 6 produced crystals by day 23 in all 0.2 M ammonium acetate drops. Variation in Bis-tris pH made no difference to whether crystals formed or not, most likely due to the pH being dominated by that of the 0.2 M ammonium acetate (Figure 5.33a).

A finer ammonium acetate screen was performed along with varying MPD (25, 35 or 45 %) and Bis-tris (0.05, 0.1 or 0.15 M) concentrations on both Rv2530a/c SEC 8 – 14 and 17 – 23 five day old protein samples at 6.8 and 14.5 mg/ml respectively. Crystals of small, very thin plates were observed by day 14 for the SEC 17 – 23 protein for all Bis-tris and ammonium acetate concentrations when 45 % MPD was included (Figure 5.33b). No crystals were observed for the SEC 8 – 14 protein under any condition, even by day 38. The inability to reproduce crystals for the SEC 8-14 sample in this fine screen could be due to the lower concentration for this protein purification (6.8 mg/ml compared to 8.13 mg/ml used in the initial fine screen above), as crystals for this condition had been observed previously for RV2530a/c (whole) at 7.03 mg/ml by day 45, Rv2530c at 8.73 mg/ml by day eight, and Rv2530a/c SEC 8 – 14 at 8.13 by day 14. A range of crystals from SEC 17-22 (Figure 5.33b) were tested for X-ray diffraction at the Australian Synchrotron but diffracted to a poor resolution (approximately 8 Å). Identical screens were repeated for the Rv2530a/c SEC 17 – 23 protein when it was five weeks old but no crystals had appeared by day 33 suggesting the protein may have succumbed to proteolytic degradation after five weeks of storage at 4 °C.

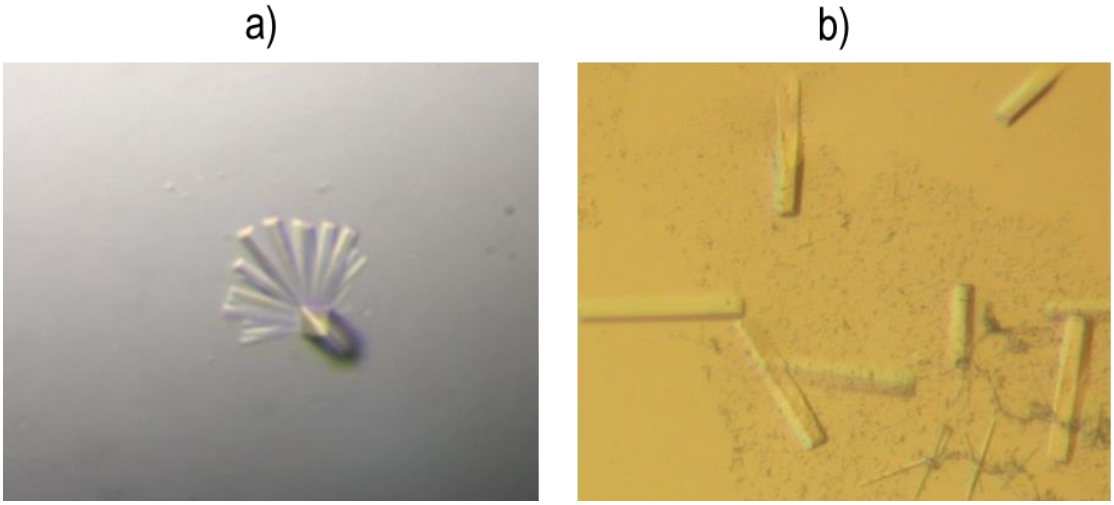


Figure 5.33: Crystals from fine screens in a) 0.2 M ammonium acetate, 0.1 M Bis-tris pH 5.3, 45 % (+/-) MPD from Rv2530a/c SEC fractions 8 – 14 (8.13 mg/ml) and b) 0.2 M ammonium acetate, 0.1 M Bis-tris pH 5.5, 45 % (+/-) MPD from Rv2530a/c SEC fractions 17 – 23 (14.5 mg/ml).

Crystals formed in 0.2 M ammonium acetate, 0.1 M Bis-tris pH 5.5, 45 % (+/-) MPD had been obtained for the whole Rv2530a/c protein, Rv2530a/c SEC fractions 8 – 14 and 17 – 23, and Rv2530c by itself so, for consistency, whole Rv2530a/c (containing pooled SEC fractions 8 – 23) was used for all further fine screen trials.

5.2.6.2.2 0.2 M ammonium acetate/sulphate, 0.1 M Bis-tris pH 5.5, 45 % (+/-) MPD and 0.5 M ammonium sulphate, 0.1 M HEPES pH 7.5, 30 % (+/-) MPD

A robot crystallisation screen (Section 2.3.3.1) was performed with eight day old Rv2530a/c protein (12.92 mg/ml) dialysed into anion exchange buffer (50 mM sodium phosphate pH 8, 100 mM NaCl). By day eight, delicate crystals were observed in 0.5 M ammonium sulphate, 0.1 M HEPES pH 7.5, 30 % (+/-) MPD (Figure 5.34a), and promising needles had formed in the original 0.2 M ammonium acetate, 0.1 M Bis-tris pH 5.5, 45 % (+/-) MPD condition. A fine screen was laid down around 0.5 M ammonium sulphate, 0.1 M HEPES pH 7.5, 30 % (+/-) MPD when the protein was 15 days old but no crystals had appeared by day 13, suggesting the protein may have begun to aggregate, or suffered from proteolysis when laid down at 15 days.

In an attempt to resurrect the ageing protein, at 27 days, proteolytic digest screens in 0.2 M ammonium acetate, 0.1 M Bis-tris pH 5.5, 45 % (+/-) MPD were performed with trypsin added to the protein immediately prior to laying down at 1:125 - 1:50,000 ratios of trypsin to protein. Identical screens were also performed but with ammonium sulphate used instead of the ammonium acetate seeing as crystals had formed in the robot screen previously in similar conditions with ammonium sulphate. Crystals were observed by day five in both ammonium sulphate (1:1,000) and ammonium acetate (1:500 and 1:1,000) conditions, although the rods were larger in drops with ammonium sulphate as seen in Figure 5.34b and c. No crystals had formed previously once the protein was over eight days old in the fine screens above without trypsin, suggesting trypsin is required to make the conformational changes necessary for crystallisation of the protein once it ages and begins to degrade.

Because the addition of trypsin to aged protein successfully produced crystals above, a proteolytic digest screen was also performed around the other successful crystallisation condition (0.5 M ammonium sulphate, 0.1 M HEPES pH 7.5, 30 % (+/-) MPD) but only needles developed (for the 45 % MPD drops with trypsin at 1:5,000 and 1:50,000). The 0.2 M ammonium acetate/sulphate, 0.1 M Bis-tris pH 5.5, 45 % (+/-) MPD condition was used for all further fine screen trials.

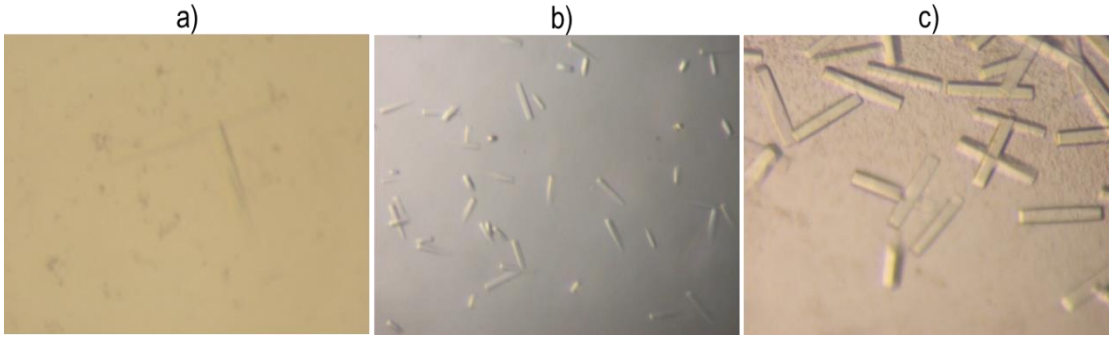


Figure 5.34: Robot and fine screen crystals from 12.92 mg/ml Rv2530a/c protein at a) 8 days old in 0.5 M ammonium sulphate, 0.1 M HEPES pH 7.5, 30 % (+/-) MPD , and at 27 days old with trypsin added at b) 1:1,000 in 0.2 M ammonium acetate, 0.1 M Bis-tris pH 5.5, 45 % (+/-) MPD and c) 1:1,000 in 0.2 M ammonium sulphate, 0.1 M Bis-tris pH 5.5, 45 % (+/-) MPD.

5.2.6.2.3 Additive Screens and Drop Ratios

An additive screen with 36 day old protein was laid down to see if any additives would stabilise or improve crystal conformity in an aged protein, as well as trialling varying ratios of protein to ML (1:1, 2:1 and 2:3). Crystals had grown by day six for many of the additive conditions, with L-Proline, butanol, glycerol and urea looking the most promising (Figure 5.35), while only a protein to ML ratio of 1:1 produced crystals. In an attempt to reduce the number of microcrystals forming, microseeding and liquid bridge diffusion was performed (Section 2.3.3.2) but crystals were only formed with a liquid bridge diffusion drop, and these were small and no improvement on those from the standard drop set up.

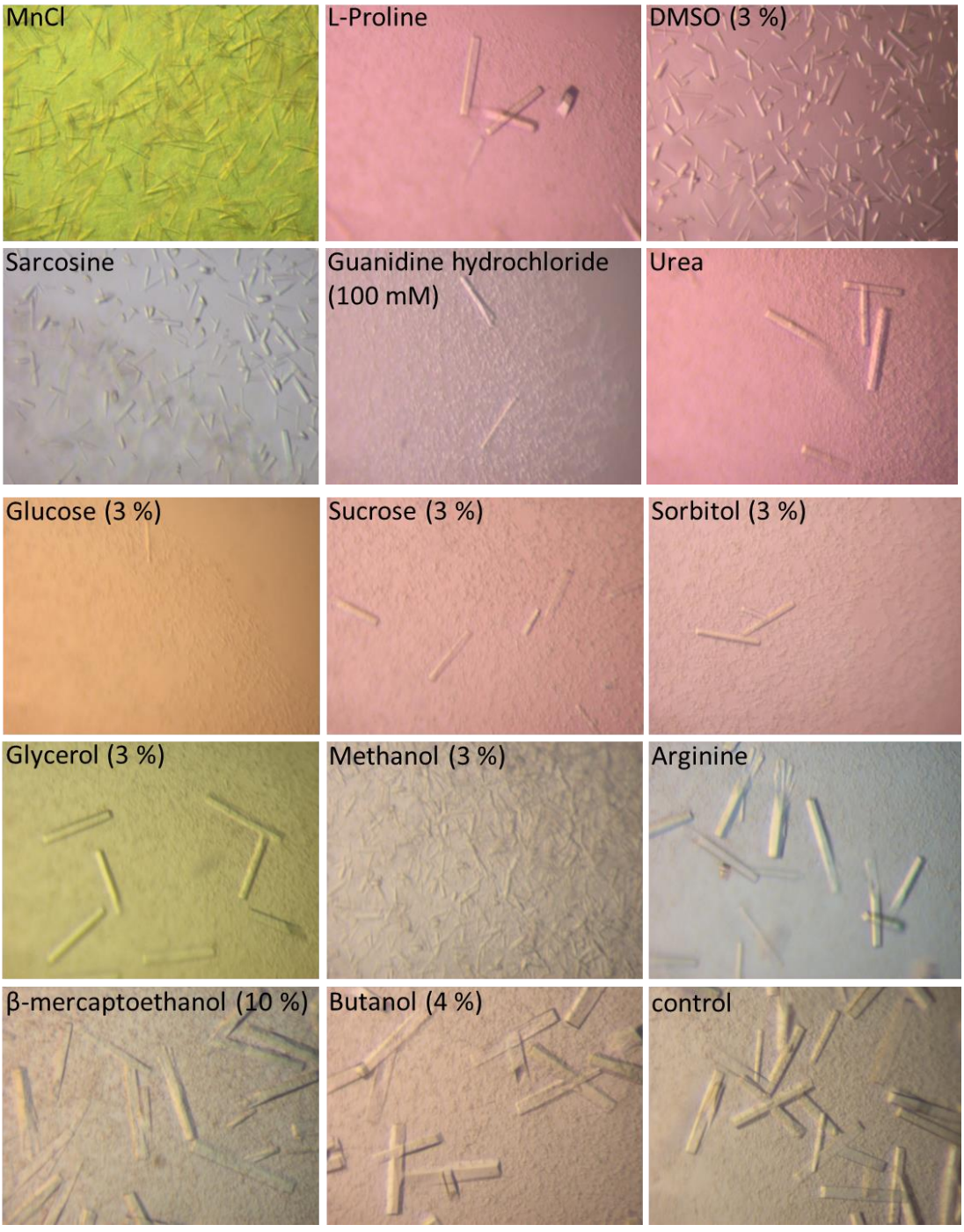


Figure 5.35: Additive screen crystals at day six in 0.2 M ammonium sulphate, 0.1 M Bis-tris pH 5.5, 45 % (+/-) MPD with 36 day old 12.92 mg/ml Rv2530a/c protein. Additives used at [10 mM] unless specified.

Positive improvements made to crystal formation from the above fine screens were applied to screens using freshly purified Rv2530a/c protein at 11.05 mg/ml. Day old protein was subjected to 0.2 M ammonium acetate/sulphate, 0.1 M Bis-tris pH 5.3, 45 % (+/-) MPD screens with the additives L-Proline, butanol, glycerol or urea, and varying the drop sizes. By day three, rods formed in all ammonium sulphate conditions while smaller rods were only observed in the control and L-Proline drops for ammonium acetate. By day seven, the crystals in ammonium sulphate drops had begun to decline and 'feather' around the edges, and X-ray diffraction at the Australian Synchrotron determined the best crystal (from the control drop) only diffracted to approximately 6 Å. Further fine screens around the most promising condition (0.2 M ammonium sulphate, 0.1 M Bis-tris pH 5.3, 45 % (+/-) MPD +/- glycerol) comparing drop sizes and pre-equilibration of protein with water before the addion of ML did not significantly improve the diffraction.

5.2.6.2.4 Crystal Dehydration and DNase Treatment

No great improvement in diffraction had been made in fine screens thus far varying ionic strength, slowing growth rate (drop dilutions/varied drop ratios) or adding co-factors to assist the protein in assuming a more rigid conformation, so screens aimed at dehydrating the crystals or removing any bound DNA from the protein with DNase were performed to try and improve crystal packing and therefore diffraction. Dehydrating crystals in a condition with such a high MPD concentration (45 %) is problematic because the highly volatile solvent competes with protein for water, so just dehydrating the drop actually results in the crystals becoming more hydrated. Instead of dehydrating the whole drop, a 2.5 – 25 % PEG 400 to PEG 10,000 additive screen was performed in an attempt to dehydrate the crystals themselves using one day old Rv2530a/c protein at 17, 13 and 11 mg/ml, but the needles and tiny crystal rods formed were inferior to those in the control drops. DNase was added to protein immediately prior to laying down drops in an attempt to stabilise crystal formation in case there was any DNA present causing structural heterogeneity. DNase additive drops (at 1:500, 1:1,000 and 1:2,000 DNase to protein) were laid down using one day old Rv2530a/c at 16 mg/ml both with and without the addition of manganese chloride (as DNase requires calcium, magnesium or manganese for activity, and previous crystallisation screens showed that the first two additives

form phosphate salt crystals). Phosphate salt crystals were apparent in the manganese chloride control drop (where no protein was added) so only drops without the addition of manganese chloride were followed up. Crystal rods were observed in drops for all DNase ratios by day 14. Interestingly, DNase I is having an effect on crystal formation even though DNase I activity has been shown to be almost negligible in buffers containing no divalent cations (Guéroult *et al.* 2010), and has a large reduction in activity in NaCl concentrations over 30 mM (Latham 2001) as is the case in the crystalisation buffer.

Screens replacing Bis-tris with alternate buffers yielded no crystals when Tris was used, and feathered and inferior crystals with Bis-tris propane. As it appeared better crystals formed as the protein aged, a screen of three day old Rv2530a/c was laid down with DNase (at 1:2,000 and 1:5,000) immediately as well as after 10 minutes incubation at room temperature. Protein was diluted to 17, 13, 11 and 9 mg/ml for the screen in water as well as in purification buffer (50 mM sodium phosphate pH 8, 200 mM NaCl) to compare crystal integrity in each of the diluents. Figure 5.36 shows protein diluted in water formed superior crystals to those from protein diluted in buffer indicating either the phosphate or salt has a detrimental effect on crystal formation in this crystallisation condition. Superior crystals were also produced in drops with protein concentrations of 9 – 13 mg/ml compared to 17 mg/ml, and without the 10 minute incubation when DNase was added. X-ray diffraction at the Australian Synchrotron indicated crystals in the control drops were slightly better than those with DNase treatment as they diffracted to a greater resolution (5.5 – 9 Å compared to 8 - 10 Å). Data to 5.5 Å (Figure 5.37) were collected for one crystal taken from a control drop laid down at 9 mg/ml (Figure 5.36a). The dataset was indexed in iMosflm giving mosaicity of 0.72 and defining the space group as I4. Indexed data was then scaled and truncated using Aimless which suggested resolution of 5.5 Å with an rmerge value of 0.8 in the outer shell. The space group was redefined as $I4_122$ with mosaicity of 0.67, which together with the MW calculated from SEC is suggestive of tetrameric symmetry. Given the resolution of the dataset, it was not feasible to solve the structure of Rv2530a/c with this dataset.

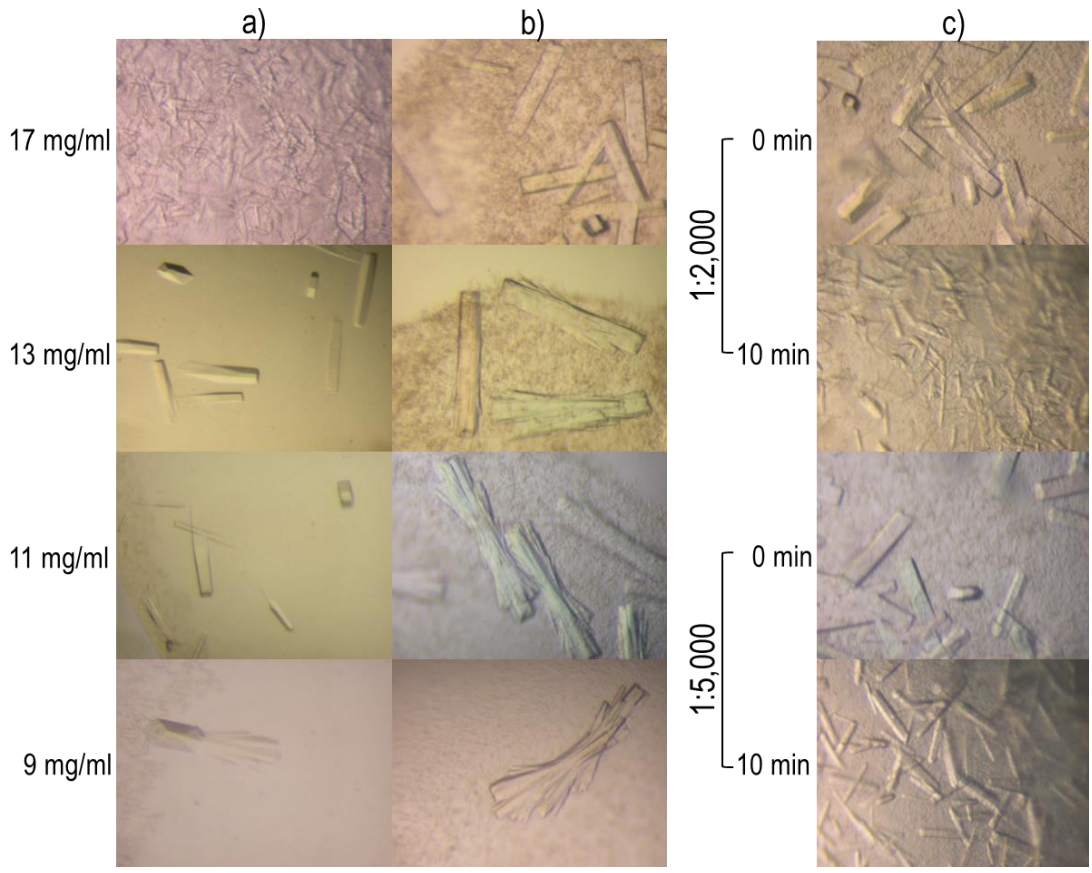


Figure 5.36: Fine screen crystals with three day old Rv2530a/c at day 12 in 0.2 M ammonium sulphate, 0.1 M Bis-tris pH 5.3, 45 % (+/-) MPD a) controls with protein diluted in water, b) protein diluted in buffer and c) + DNase at 1:2,000 or 1:5,000 ratio of DNase to protein +/- a 10 min incubation before drops laid down.

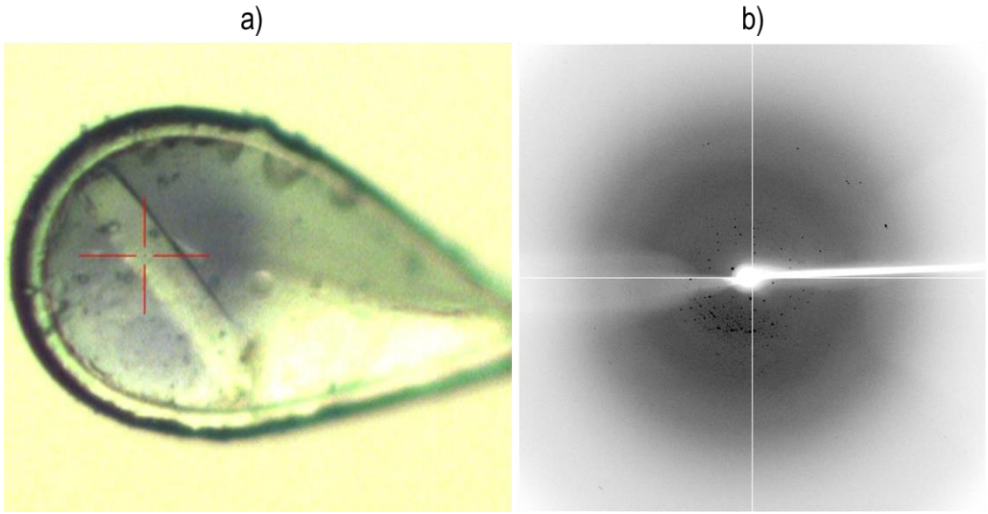


Figure 5.37: Rv2530a/c crystal and X-ray diffraction to 5.5 Å a) Rv2530a/c crystal from Figure 5.36a used to collect dataset, b) X-ray diffraction of one frame obtained during data collection at the Australian Synchrotron (plate resolution 2.37 Å).

Crystallisation attempts of Rv2530c in the future should focus on reducing the structural flexibility and therefore structural heterogeneity of the protein. Removal of the His-tag would reduce the structural flexibility, but as it is C-terminal it is difficult to do, especially given there are no protease sites between the protein and His-tag. Other approaches for reducing the structural heterogeneity of the protein could involve screening ligands (and inhibitors), different additives and co-factors which may help the protein assume a more rigid conformation. Another strategy which may improve the diffraction quality of the protein crystal is slowing the rate of crystal growth (Heras & Martin 2005). While attempting to slow the rate of crystal growth, one could investigate the use of MPD as the precipitant in the crystallisation condition. MPD is a small polyalcohol which can act as a precipitant via competition for water, and also functions as a cryoprotectant. MPD is mildly denaturing and promotes hydration of proteins via association with hydrophobic protein surface patches, suggesting it may stabilise the protein by filling up cavities on the protein surface and indirectly facilitate proper crystallisation (Anand *et al.* 2002). Once the rate of crystallisation was slowed, possibly by laying crystals down at a different temperature (especially given MPDs additional role as a cryoprotectant), packing of the voids by MPD may also improve resulting in improved diffraction of the crystal.

5.3 Conclusions

The role of VapBC systems in *M. tuberculosis* has been the subject of great interest over recent years, due to their expanded number in the genome and links with the regulation of cell growth in response to environmental stress. Their ability to regulate or slow growth under adverse conditions for presumed survival advantages possibly leading to latency, persistence or bacteriostasis, make them ideal candidates for developing new treatments to target these persistent or non-replicating bacteria. The finding that the three *M. tuberculosis* VapC toxins tested thus far display ribonuclease activity and may target the same mRNA sequence is intriguing. This raises the possibility that all 47 VapCs in *M. tuberculosis* are RNases and target the same sequence, although maintaining such a large surfeit of genes to perform the same function would be excessive and seemingly costly to the organism in terms of energy and maintenance. Alternatively, perhaps they belong to sub-groups which each target a different sequence. If multiple VapCs share the same target, they may still be regulated by, or under the

control of different activators, which are triggered in response to different environmental stimuli. This would allow the bacteria to respond and adapt to the diverse range of conditions which they encounter. The ability of an organism to have groups of VapCs with different targets, could allow it to exude both gross and fine control over the cells metabolism. One set of RNases could target a ubiquitous cleavage site in response to critical stressors effecting a large and fast change in the cells transcriptional profile, while another set may respond to less adverse stressors with a more gentle effect, targeting infrequent mRNA sites and therefore specific messages. Having this combined approach may explain why *M. tuberculosis* has maintained 47 VapBC loci in its genome, as it must be able to survive and regulate its growth in many environments, and under a diverse range of stressors (supported by the observation that the majority of TA systems are only found in virulent mycobacteria of the MTBC (Ramage *et al.* 2009)). It is also possible that it is a combination of sequence and secondary structure which is being targeted by the various VapCs as, while target sequences were preferentially cut, other suboptimal sites were also cut, although to a lesser degree. While multiple VapCs may target the same sequence, other factors are likely to be involved in the regulation of this efficient RNase. Research being undertaken currently in our lab indicates AmtR (a global repressor in the nitrogen regulation system of a closely related organism to *M. tuberculosis* (Jakoby *et al*. 2000)) may play a role in the regulation of VapBC in *M. smegmatis* (Chelsea Vickers, unpublished data), suggesting VapCs may not only be involved in the regulation of glycerol or carbon utilisation, but of other metabolic substrates as well such as nitrogen. Another way in which the VapCs may be differentially regulated is via targeted degradation of their cognate antitoxin by different proteases. The method by which VapB degradation occurs, and the environmental triggers which initiate activation of these proteases to act on the VapB protein are poorly understood, so it is possible that particular conditions of stress activate specific proteases which only target a certain VapB (or subset of VapBs), leading to targeted VapC RNase activity. There is also a possibility that not all 47 of the *M. tuberculosis* VapCs function as an RNase, although this is becoming less likely as each new VapC is characterised and shown to display RNase activity.

While the exact biological function of VapBCs in *M. tuberculosis* has not been elucidated, it is clear they play some role in regulating growth in response to changes in the environments the

organism encounters. The ability to adapt to these changes allows the bacterium to evade host responses and the effects of antibiotics, and persist for long periods. This, together with their suggested role in virulence (Arcus *et al.* 2005; Magombedze & Mulder 2013), make VapBCs ideal candidates to target when investigating the treatment for, and pathogenesis of, *M. tuberculosis*.

Chapter Six

Conclusions

Given that the two major impediments to effective TB eradication are the time to diagnosis and duration of treatment, the overarching intention of this research was to develop a rapid strain specific *M. tuberculosis* diagnostic assay, and investigate the role *M. tuberculosis* toxin-antitoxins play in the pathogens survival and persistence. To do this, we focussed on the VapBC system of toxin-antitoxins, and proposed three objectives: one, to scrutinise the highly transmissible and virulent Rangipo strain of *M. tuberculosis* prevalent within New Zealand, and develop a fast and efficient Rangipo specific diagnostic assay which could be implemented into the current hospital diagnostic setting; two, to identify conditions where *vapB* or *vapC* expression was differentially expressed (in *M. smegmatis*) in order to help identify the mechanisms involved in the regulation of VapC, as well as to develop an optimised RNA isolation and RT-qPCR method suitable for the quantification of these target genes expression. And lastly, to identify all unannotated *M. tuberculosis* VapBC proteins, then clone, express and purify all 47 into *M. smegmatis*, and to determine the specificity and activity of a number of the purified *M. tuberculosis* VapC proteins.

6.1 Rangipo Diagnostic

Control of the *M. tuberculosis* Rangipo strain, which has been circulating in the Waikato and Bay of Plenty regions for a quarter of a century and is responsible for the largest cluster of tuberculosis infections within New Zealand, has proved difficult owing to its virulence and high transmissibility. Whole genome sequencing of 10 Rangipo isolates revealed that two of the 99 Rangipo specific SNPs identified were within *vapC* toxin genes, while LSP and RD PCR analysis determined the Rangipo strain belongs to an as yet unclassified *M. tuberculosis* Euro-American sublineage. Robust discrimination between strains of pathogenic bacteria is

essential for controlling infection, as well as gaining insight into the differences in transmissibility, virulence and pathology of disease between strains, which will lead to better targeted microbial therapies. A pivotal factor in the effective control of tuberculosis is the development of rapid and robust methods for *M. tuberculosis* infection diagnosis. An affordable, rapid and reliable diagnostic RFLP assay based on the Rangipo specific SNPs has been developed to quickly identify the Rangipo strain. Fast diagnosis of this highly transmissible strain will reduce the delay in treatment of its predominantly pulmonary infection, in turn limiting its spread, reducing the associated contact tracing and treatment costs, and improving patient outcomes.

6.2 Expression of *vapB* and *vapC* in *M. smegmatis* using RT-qPCR

A modified AP-GITC RNA isolation method was successfully developed, and used in RT-qPCR to determine that there was no differential expression of vapC under the conditions of stress tested. Housekeeping genes belonging to a range of functional and abundance classes were trialled for use as RT-qPCR normalisers in the stress experiments, but almost all were deemed unsuitable due to inconsistent or poor expression. A single housekeeping gene (16S) was found to be consistently expressed, but its expression was much higher than that for the genes of interest due to the presence of two copies. 16S was used successfully as a normalising housekeeping gene however, by diluting the amount of cDNA used in the housekeeper samples during RT-qPCR. Because just a single housekeeping gene was used to normalise the data (when multiple housekeeping genes have been shown to produce more reliable results), the resultant expression data was treated cautiously. The fact that no consistent change in the mRNA expression levels of *vapB* and *vapC* occurred, prompted the use of modelling to investigate how the levels of toxic VapC are being controlled. This negative expression data together with the modelling data, form the basis for our current hypothesis that it is not the transcriptional differences which are important in the regulation of VapC, but post-transcriptional factors.

Modelling of changes in protein and mRNA levels suggests transcription of the VapBC operon results in a ratio of VapB to VapC of approximately 10 : 1. This is due to the ribosome only occasionally slipping back across the VapB/C overlap to transcribe the toxic VapC, instead of falling off once it reaches the end of VapB. The equilibrium 'norm', is that VapC is always in a complex with VapB and is very rarely free, so under 'normal' conditions, the majority of the toxic VapC is held in a benign complex with its antitoxin, and only a very small amount is free to exert its toxic effect. VapB is much more susceptible to proteolytic degradation than VapC, so over time the VapB is degraded at a faster rate than VapC which pulls the equilibrium backwards, leading to an increase in free VapC. Therefore, under 'normal' conditions VapBC must be constantly translated so that the equilibrium holds, avoiding the release of free VapC into the cell. As translation slows, be it due to encountering external stressors or the growth rate reaching stationary phase, the equilibrium is forced backwards resulting in a spike of free VapC, which only settles back to low steady state levels once the VapBC translation steadies again (see Figure 6.1) (Arcus 2013 (unpublished); (Fasani & Savageau 2013).

The VapC regulation model proposes there is a very slight amount of free VapC present under normal conditions. This is strengthened by the subtle increase in growth rate observed for the *vapBC* deletion mutant, Δ*vapBC*, where no trace amounts of free VapC are present to exert their toxicity and slow growth (McKenzie *et al.* 2012b). Microarray data has shown that genes involved in carbon transport and metabolism are downregulated when VapC is overexpressed (Robson *et al.* 2009), and it was observed that glycerol uptake for the *vapBC* deletion mutant was approximately twice that of its wildtype counterpart, while the amount of biomass produced (relative to glycerol uptake) was halved (McKenzie *et al.* 2012b). This suggests glycerol uptake in Δ*vapBC* is inefficient, and results in unregulated growth due to the absence of VapBC to control glycerol metabolism. Taken together with the modelling data, this suggests that VapC is required for the regulation, or fine tuning of carbon and possibly nitrogen (Arcus 2013 unpublished) metabolism in response to growth rate.

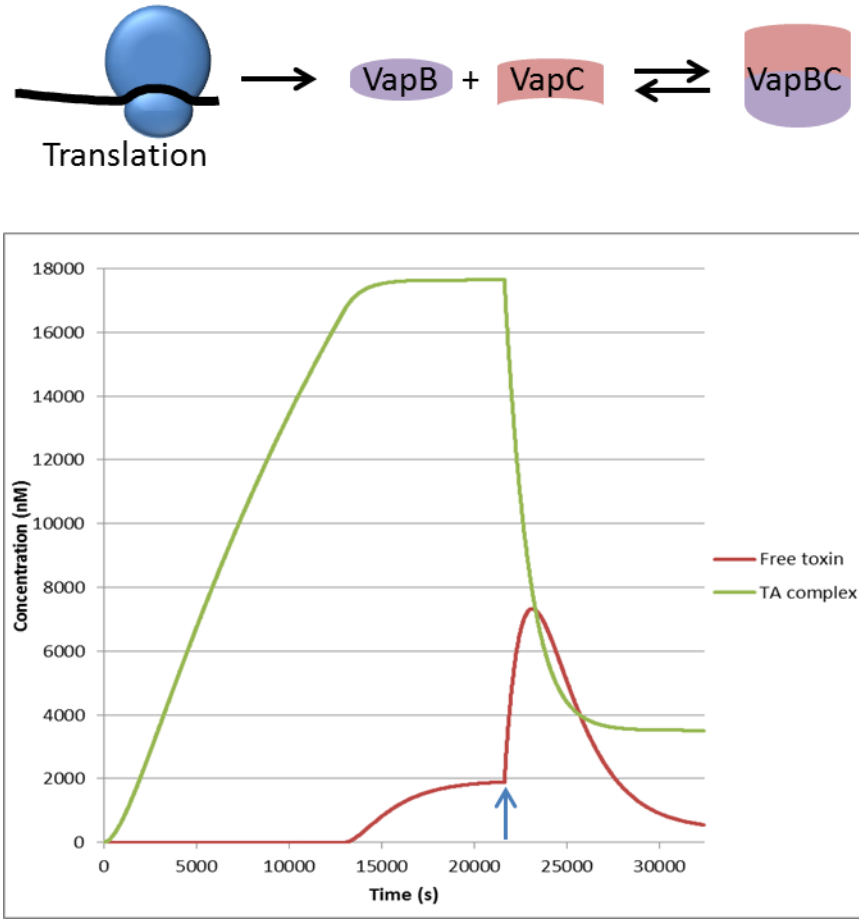


Figure 6.1: Top: The VapBC equilibrium. Translation results in VapB and VapC which form the benign VapBC complex, but if VapB is degraded, the equilibrium shifts to the left, resulting in free VapC toxin. Bottom: Model depicting the regulation of VapC toxin. Over time, the toxin-antitoxin (TA) complex concentration (green) increases to a steady state, prior to the (much lower) free toxin levels (red). If the translation rate is changed to a lower rate (at the blue arrow), there is a temporary spike in free toxin levels before they return to their lower concentration. Figure from unpublished results from Arcus (2013).

6.3 *M. tuberculosis* VapBC Characterisation

The large number of VapBC systems in *M. tuberculosis* and their links to the regulation of cell growth in response to environmental stress, have prompted questions as to the role VapBC systems may play in *M. tuberculosis*, and make them interesting candidates for the development of new *M. tuberculosis* treatments. Small scale protein expression tests identified 17 of the 47 *M. tuberculosis* VapBC's as expressing soluble VapC protein, 13 of which were put through large scale expression studies during the course of this research (three of those were purified by others in the lab). Because the VapC protein from *M. smegmatis* ($VapC_{MS1284}$) had previously been shown to exhibit ribonuclease activity (McKenzie 2011), the *M.*

tuberculosis VapC proteins $VapC_{Rv2548}$ and $VapC_{Rv2530c}$ purified here, were also tested for RNase activity. Both VapCs displayed general Mg^{2+}-dependent RNase activity against *M. smegmatis* total RNA, as well as the same specific activity (with similar cutting efficiencies to each other) against all of the pentaprobe substrates. This was also the case for the two other *M. tuberculosis* VapCs ($VapC_{Rv0065}$ and $VapC_{Rv0617}$), determined by others in the lab (McKenzie *et al.* 2012a). This same banding pattern suggests that *M. tuberculosis* VapCs may target the same recognition sequence, which is different to that observed for $VapC_{PAE0151}$ and $VapC_{PAE2754}$ from *P. aerophilum*, and $VapC_{MS1284}$ from *M. smegmatis* (McKenzie *et al.* 2012a). The lack of one pentaprobe substrate being cut preferentially over the others observed here for $VapC_{Rv2548}$ and $VapC_{Rv2530c}$, suggests that none of these are the 'ideal' substrate for either of the VapC proteins, possibly due to unsuitable RNA sequence and/or secondary structure. MALDI-TOF MS using overlapping 932 RNA oligonucleotide substrates was used to determine the sequence specificity of $VapC_{Rv2530c}$, in order to gain a better understanding of the targets, and therefore the possible pathways and functions the *M. tuberculosis* VapC proteins are involved with. The most optimal cut site for $VapC_{Rv2530c}$ as determined by MALDI-TOF MS, was UA*GG. When this same analysis method for determining the optimal cut site was applied to historical MS data for $VapC_{Rv0065}$ and $VapC_{Rv0617}$, UA*GG was also the most likely preferred cut site, suggesting that UA*GG is the preferred cut site for the three *M. tuberculosis* VapC proteins tested thus far. Previously, GC-rich 4mers were thought to be the preferred cut sites for $VapC_{Rv0065}$ and $VapC_{Rv0617}$ (McKenzie 2011; McKenzie *et al.* 2012a), but the analysis method used then appears to be less robust than that which has generated the UA*GG sequence. It must be noted however, that both analyses have relied on imperfect MALDI-TOF MS spectra obtained from a period of poor laser performance, so repeat assays should be performed again on $VapC_{Rv0065}$ and $VapC_{Rv0617}$ to confirm UA*GG is the preferred cut site for these two VapC RNases, as well as for $VapC_{Rv2530c}$. The fact that the three *M. tuberculosis* VapC toxins tested thus far display identical ribonuclease activity raises the possibility that all 47 VapCs are RNases and are targeting the same sequence. The resulting cost to the organism of maintaining such an excessive number of genes in order to seemingly perform the same function would appear extreme. Alternatively; the 47 VapCs may belong to sub-groups, each targeting a different sequence and/or secondary structure allowing *M. tuberculosis* to exude both gross and fine control over the cells metabolism; or they may in fact share the

same target, but are regulated by different activators triggered in response to different environmental stimuli, in order to respond and adapt to the diverse range of conditions which *M. tuberculosis* encounters. While multiple VapCs may target the same sequence, other factors are likely to be involved in the regulation of this efficient RNase, such as AmtR, suggesting that as with the VapC in *M. smegmatis, M. tuberculosis* VapCs may be involved in the regulation of a wide range of metabolic substrates. Bioinformatic analysis of the *M. tuberculosis* H37Rv genome revealed an abundance of hypothetical proteins and genes involved in metabolic processes containing the TAGG $VapC_{Rv2530c}$ target sequence. Once ranked relative to gene size, genes containing the highest number of TAGG sequences were annotated as hypothetical proteins, transposases, TA genes and other genes associated with horizontal gene transfer. $VapC_{Rv2530c}$ could be targeting HGT elements because it is also functioning as a selfish element - to eliminate the other selfish elements in order to survive the genome war within, and targeting TAs in order to initiate a transcriptional regulation cascade when a rapid change in the proteomic response and metabolic state of the cell is required, such as during conditions of stress.

6.4 Future Research

The three *M. tuberculosis* VapC RNases tested thus far, appear to have an identical recognition sequence. As well as attempting to purify the four remaining solubly expressed VapCs and testing for RNase activity, MALDI-TOF MS of RNA oligos should be performed again on $VapC_{Rv0065}$ and $VapC_{Rv0617}$ as well as $VapC_{Rv2548}$ once the laser is fully functional, to obtain better data and more conclusively confirm they share the same preferred cut site as $VapC_{Rv2530c}$. Because this cut site is different to that for VapCs from both *P. aerophilum* (targeting G rich sequences) and *M. smegmatis* (targeting AT sequences) (McKenzie *et al.* 2012a; McKenzie *et al.* 2012b), confirming exactly which target sequence residues are crucial for the RNase activity of the *M. tuberculosis* VapCs, will be instrumental in determining their optimal target sequence and therefore their biochemical function. One way of achieving this would be performing a microarray or RNA-Seq to identify genes which are downregulated in response to expression of a particular *M. tuberculosis* VapC in its VapBC knockout, as was the case with VapC from *M. smegmatis* (McKenzie *et al.* 2012b).

The exact biological function of *M. tuberculosis* VapBCs is yet to be elucidated, but their ability to regulate or slow growth under adverse conditions for presumed survival advantages possibly leading to dormancy, persistence or bacteriostasis, make them ideal candidates for developing new treatments to target the persistent or non-replicating *M. tuberculosis* bacterium.

References

Achtman, M. 2008, 'Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens', *Annu Rev Microbiol*, vol. 62, no. pp. 53-70.

Ahidjo, B. A.*, et al.* 2011, 'VapC Toxins from *Mycobacterium tuberculosis* Are Ribonucleases that Differentially Inhibit Growth and Are Neutralized by Cognate VapB Antitoxins', *PLoS ONE*, vol. 6, no. 6, pp. e21738.

Aizenman, E.*, et al.* 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*, vol. 93, no. 12, pp. 6059-6063.

Albrethsen, J.*, et al.* 2013, 'Proteomic Profiling of *Mycobacterium tuberculosis* Identifies Nutrient-Starvation-Responsive Toxin-Antitoxin Systems', *Mol Cell Proteomics*, vol. 12, no. 5, pp. 1180-91.

Alix, E.*, et al.* 2006, 'Identification of a Haarlem Genotype-Specific Single Nucleotide Polymorphism in the mgtC Virulence Gene of *Mycobacterium tuberculosis*', *Journal of Clinical Microbiology*, vol. 44, no. 6, pp. 2093-2098.

Alland, D.*, et al.* 2007, 'Role of Large Sequence Polymorphisms (LSPs) in Generating Genomic Diversity among Clinical Isolates of *Mycobacterium tuberculosis* and the Utility of LSPs in Phylogenetic Analysis', *J. Clin. Microbiol.*, vol. 45, no. 1, pp. 39-46.

Amitai, S.*, et al.* 2004, 'MazF-Mediated Cell Death in *Escherichia coli*: a Point of No Return', *Journal of Bacteriology*, vol. 186, no. 24, pp. 8295-8300.

Anand, K.*, et al.* 2002, 'An Overview on 2-methyl-2,4-pentanediol in Crystallization and in Crystals of Biological Macromolecules', *Acta Crystallographica Section D*, vol. 58, no. 10 Part 1, pp. 1722-1728.

Andersen, C. L.*, et al.* 2004, 'Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: a Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets', *Cancer Res*, vol. 64, no. 15, pp. 5245-50.

Anilkumar, A. K.*, et al.* 2012, 'Standardization and Evaluation of a Tetraplex Polymerase Chain Reaction to Detect and Differentiate *Mycobacterium tuberculosis* Complex and Nontuberculous Mycobacteria—a Retrospective Study on Pulmonary TB Patients', *Diagnostic microbiology and infectious disease*, vol. 72, no. 3, pp. 239-247.

Arcus, V. L.*, et al.* 2004, 'Distant Structural Homology Leads to the Functional Characterization of an Archaeal PIN Domain as an Exonuclease', *Journal of Biological Chemistry*, vol. 279, no. 16, pp. 16471-16478.

Arcus, V. L.*, et al.* 2011, 'The PIN-Domain Ribonucleases and the Prokaryotic VapBC Toxin–Antitoxin Array', *Protein Engineering Design and Selection*, vol. 24, no. 1-2, pp. 33-40.

Arcus, V. L.*, et al.* 2005, 'The PIN-Domain Toxin-Antitoxin Array in Mycobacteria', *Trends in Microbiology*, vol. 13, no. 8, pp. 360-365.

Baena, A. & Porcelli, S. A. 2009, 'Evasion and Subversion of Antigen Presentation by *Mycobacterium tuberculosis*', *Tissue Antigens*, vol. 74, no. 3, pp. 189-204.

Baker, M.*, et al.* 2008, 'Tuberculosis Associated with Household Crowding in a Developed Country', *Journal of Epidemiology and Community Health*, vol. 62, no. 8, pp. 715-721.

Barnes, P. F. & Cave, M. D. 2003, 'Molecular Epidemiology of Tuberculosis', *New England Journal of Medicine*, vol. 349, no. 12, pp. 1149-1156.
Bashiri, G.*, et al.* 2010, 'Metabolic Engineering of Cofactor F_{420} Production in *Mycobacterium smegmatis*', *PLoS ONE*, vol. 5, no. 12, pp. e15803.
Bashiri, G.*, et al.* 2007, 'Expression, Purification and Crystallization of Native and Selenomethionine Labeled *Mycobacterium tuberculosis* FGD1 (Rv0407) Using a *Mycobacterium smegmatis* Expression System', *Protein Expression and Purification*, vol. 54, no. 1, pp. 38-44.
Becq, J.*, et al.* 2007, 'Contribution of Horizontally Acquired Genomic Islands to the Evolution of the Tubercle Bacilli', *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1861-1871.
Beste, D. J. V.*, et al.* 2009, 'The Genetic Requirements for Fast and Slow Growth in Mycobacteria', *PLoS ONE*, vol. 4, no. 4, pp. e5349.
Bifani, P.*, et al.* 2000, 'Molecular Characterization of *Mycobacterium tuberculosis* H37Rv/Ra Variants: Distinguishing the Mycobacterial Laboratory Strain', *J Clin Microbiol*, vol. 38, no. 9, pp. 3200-4.
Boehme, C. C.*, et al.* 2010, 'Rapid Molecular Detection of Tuberculosis and Rifampin Resistance', *New England Journal of Medicine*, vol. 363, no. 11, pp. 1005-1015.
Brennan, P. J. 2003, 'Structure, Function, and Biogenesis of the Cell Wall of *Mycobacterium tuberculosis*', *Tuberculosis (Edinburgh, Scotland)*, vol. 83, no. 1, pp. 91-97.
Brosch, R.*, et al.* 2002, 'A New Evolutionary Scenario for the *Mycobacterium tuberculosis* Complex', *Proceedings of the National Academy of Sciences*, vol. 99, no. 6, pp. 3684-3689.
Brown-Elliott, B. A. & Wallace, R. J., Jr. 2002, 'Clinical and Taxonomic Status of Pathogenic Nonpigmented or Late-Pigmenting Rapidly Growing Mycobacteria', *Clin Microbiol Rev*, vol. 15, no. 4, pp. 716-46.
Brown, B. L.*, et al.* 2009, 'Three Dimensional Structure of the MqsR:MqsA Complex: A Novel TA Pair Comprised of a Toxin Homologous to RelE and an Antitoxin with Unique Properties', *PLoS Pathog*, vol. 5, no. 12, pp. e1000706.
Brudey, K.*, et al.* 2006, '*Mycobacterium tuberculosis* Complex Genetic Diversity: Mining the Fourth International Spoligotyping Database (SpolDB4) for Classification, Population Genetics and Epidemiology', *BMC Microbiol*, vol. 6, no. pp. 23.
Bunker, R. D.*, et al.* 2008, 'Crystal Structure of PAE0151 from *Pyrobaculum aerophilum*, a PIN-Domain (VapC) Protein from a Toxin-Antitoxin Operon', *Proteins: Structure, Function, and Bioinformatics*, vol. 72, no. 1, pp. 510-518.
Bustin, S. & Nolan, T. 2004, 'Template Handling, Preparation, and Quantification ', in S. Bustin (ed.), *A-Z of quantitative PCR* International University Line (IUL) La Jolla, CA, USA, pp. 141-191.
Bustin, S. A. 2000, 'Absolute Quantification of mRNA Using Real-Time Reverse Transcription Polymerase Chain Reaction Assays', *J Mol Endocrinol*, vol. 25, no. 2, pp. 169-193.
Bustin, S. A.*, et al.* 2009, 'The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments', *Clinical Chemistry*, vol. 55, no. 4, pp. 611-622.
Bustin, S. A.*, et al.* 2005, 'Quantitative real-time RT-PCR - a perspective', *Journal of Molecular Endocrinology*, vol. 34, no. 3, pp. 597-601.
Buts, L.*, et al.* 2005, 'Toxin–Antitoxin Modules as Bacterial Metabolic Stress Managers', *Trends in Biochemical Sciences*, vol. 30, no. 12, pp. 672-679.

Calder, L. 2013, 'Tuberculosis Outbreak Among Hawke's Bay Maori', *2013 Australasian Tuberculosis Conference*, Auckland, New Zealand, pp. 22.
Canales, R. D.*, et al.* 2006, 'Evaluation of DNA Microarray Results with Quantitative Gene Expression Platforms', *Nat Biotechnol*, vol. 24, no. 9, pp. 1115-22.
Chomczynski, P. & Mackey, K. 1995, 'Substitution of Chloroform by bromo-chloropropane in the Single-Step Method of RNA Isolation', *Anal Biochem*, vol. 225, no. 1, pp. 163-4.
Chomczynski, P. & Sacchi, N. 1987, 'Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction', *Analytical Biochemistry*, vol. 162, no. 1, pp. 156-159.
Christensen-Dalsgaard, M. & Gerdes, K. 2006, 'Two higBA Loci in the *Vibrio cholerae* Superintegron Encode mRNA Cleaving Enzymes and can Stabilize Plasmids', *Molecular Microbiology*, vol. 62, no. 2, pp. 397-411.
Christensen-Dalsgaard, M.*, et al.* 2010, 'Three New RelE-Homologous mRNA Interferases of *Escherichia coli* Differentially Induced by Environmental Stresses', *Molecular Microbiology*, vol. 75, no. 2, pp. 333-348.
Christensen, S. K. & Gerdes, K. 2003, 'RelE Toxins from Bacteria and Archaea Cleave mRNAs on Translating Ribosomes, Which are Rescued by tmRNA', *Molecular Microbiology*, vol. 48, no. 5, pp. 1389-1400.
Christensen, S. K.*, et al.* 2001, 'RelE, a Global Inhibitor of Translation, is Activated During Nutritional Stress', *Proceedings of the National Academy of Sciences*, vol. 98, no. 25, pp. 14328-14333.
Christensen, S. K.*, et al.* 2003, 'Toxin–Antitoxin Loci as Stress-Response-Elements: ChpAK/MazF and ChpBK Cleave Translated RNAs and are Counteracted by tmRNA', *Journal of Molecular Biology*, vol. 332, no. 4, pp. 809-819.
Cikos, S.*, et al.* 2007, 'Relative Ruantification of mRNA: Comparison of Methods Currently Used for Real-Time PCR Data Analysis', *BMC Mol Biol*, vol. 8, no. pp. 113.
Clark, D. P. & Pazdernik, N. J. 2013, *Molecular Biology*, Academic Press,
Cobb, B. D. & Ciarkson, J. M. 1994, 'A Simple Procedure for Optimising the Polymerase Chain Reaction (PCR) Using Modified Taguchi Methods', *Nucleic Acids Research*, vol. 22, no. 18, pp. 3801-3805.
Colangeli, R.*, et al.* 2014, 'Whole Genome Sequencing of *Mycobacterium tuberculosis* Reveals Slow Growth and Low Mutation Rates during Latent Infections in Humans', *PLoS ONE*, vol. 9, no. 3, pp. e91024.
Cole, S. T.*, et al.* 1998, 'Deciphering the Biology of *Mycobacterium tuberculosis* from the Complete Genome Sequence', *Nature*, vol. 393, no. 6685, pp. 537-544.
Comas, I.*, et al.* 2010, 'Human T Cell Epitopes of *Mycobacterium tuberculosis* are Evolutionarily Hyperconserved', *Nat Genet*, vol. 42, no. 6, pp. 498-503.
Comas, I.*, et al.* 2009, 'Genotyping of Genetically Monomorphic Bacteria: DNA Sequencing in *Mycobacterium tuberculosis* Highlights the Limitations of Current Methodologies', *PLoS ONE*, vol. 4, no. 11, pp. e7815.
Coscolla, M.*, et al.* 2013, 'Novel *Mycobacterium tuberculosis* Complex Isolate from a Wild Chimpanzee', *Emerging infectious diseases*, vol. 19, no. 6, pp. 969 - 976.
Daines, D. A.*, et al.* 2007, 'VapC-1 of Nontypeable *Haemophilus influenzae* Is a Ribonuclease', *Journal of Bacteriology*, vol. 189, no. 14, pp. 5041-5048.

Dale, J. W.*, et al.* 2003, 'Evolutionary Relationships among Strains of *Mycobacterium tuberculosis* with Few Copies of IS*6110*', *Journal of Bacteriology*, vol. 185, no. 8, pp. 2555-2562.

Danelishvili, L.*, et al.* 2003, '*Mycobacterium tuberculosis* Infection Causes Different Levels of Apoptosis and Necrosis in Human Macrophages and Alveolar Epithelial Cells', *Cellular Microbiology*, vol. 5, no. 9, pp. 649-660.

de Jong, B. C.*, et al.* 2009, 'Use of Spoligotyping and Large Sequence Polymorphisms To Study the Population Structure of the *Mycobacterium tuberculosis* Complex in a Cohort Study of Consecutive Smear-Positive Tuberculosis Cases in The Gambia', *Journal of Clinical Microbiology*, vol. 47, no. 4, pp. 994-1001.

de Jong, B. C.*, et al.* 2008, 'Progression to Active Tuberculosis, but Not Transmission, Varies by *Mycobacterium tuberculosis* Lineage in The Gambia', *Journal of Infectious Diseases*, vol. 198, no. 7, pp. 1037-1043.

De Zoysa, R. 2001, *A Prolonged Outbreak of Tuberculosis in the North Island,*

Delogu, G.*, et al.* 2006, 'The hbhA Gene of *Mycobacterium tuberculosis* is Specifically Upregulated in the Lungs but Not in the Spleens of Aerogenically Infected Mice', *Infection and Immunity*, vol. 74, no. 5, pp. 3006-3011.

Dietrich, G.*, et al.* 2000, 'Isolation of RNA From Mycobacteria Grown Under *in vitro* and *in vivo* Conditions', *FEMS Microbiology Letters*, vol. 186, no. 2, pp. 177-180.

Driscoll, J. 2009, 'Spoligotyping for Molecular Epidemiology of the *Mycobacterium tuberculosis* Complex', in D. A. Caugant (ed.), *Molecular Epidemiology of Microorganisms*, Humana Press, pp. 117-128.

Engelberg-Kulka, H.*, et al.* 2006, 'Bacterial Programmed Cell Death and Multicellular Behavior in Bacteria', *PLoS Genet*, vol. 2, no. 10, pp. e135.

Erickson, H. S.*, et al.* 2007, 'Assessment of Normalization Strategies for Quantitative RT-PCR Using Microdissected Tissue Samples', *Lab Invest*, vol. 87, no. 9, pp. 951-62.

Fang, Z. & Cui, X. 2011, 'Design and Validation Issues in RNA-seq Experiments', *Brief Bioinform*, vol. 12, no. 3, pp. 280-7.

Farnia, P.*, et al.* 2010, 'Growth and Cell-Division in Extensive (XDR) and Extremely Drug Resistant (XXDR) Tuberculosis Strains: Transmission and Atomic Force Observation', *Int J Clin Exp Med*, vol. 3, no. 4, pp. 308-14.

Fasani, R. A. & Savageau, M. A. 2013, 'Molecular Mechanisms of Multiple Toxin-Antitoxin Systems are Coordinated to Govern the Persister Phenotype', *Proc Natl Acad Sci U S A*, vol. 110, no. 27, pp. E2528-37.

Fernandes, N. D.*, et al.* 1999, 'A Mycobacterial Extracytoplasmic Sigma Factor Involved in Survival following Heat Shock and Oxidative Stress', *Journal of Bacteriology*, vol. 181, no. 14, pp. 4266-4274.

Feuerriegel, S.*, et al.* 2010, 'Thr202Ala in thyA Is a Marker for the Latin American Mediterranean Lineage of the *Mycobacterium tuberculosis* Complex Rather than Para-Aminosalicylic Acid Resistance', *Antimicrobial Agents and Chemotherapy*, vol. 54, no. 11, pp. 4794-4798.

Filliol, I.*, et al.* 2006, 'Global Phylogeny of *Mycobacterium tuberculosis* Based on Single Nucleotide Polymorphism (SNP) Analysis: Insights into Tuberculosis Evolution, Phylogenetic Accuracy of Other DNA Fingerprinting Systems, and Recommendations for a Minimal Standard SNP Set', *Journal of Bacteriology*, vol. 188, no. 2, pp. 759-772.

Fineran, P. C.*, et al.* 2009, 'The Phage Abortive Infection System, ToxIN, Functions as a Protein–RNA Toxin–Antitoxin Pair', *Proceedings of the National Academy of Sciences*, vol. 106, no. 3, pp. 894-899.
Fleige, S.*, et al.* 2006, 'Comparison of Relative mRNA Quantification Models and the Impact of RNA Integrity in Quantitative Real-Time RT-PCR', *Biotechnology Letters*, vol. 28, no. 19, pp. 1601-1613.
Flores, L.*, et al.* 2007, 'Large Sequence Polymorphisms Classify *Mycobacterium tuberculosis* Strains with Ancestral Spoligotyping Patterns', *J. Clin. Microbiol.*, vol. 45, no. 10, pp. 3393-3395.
Ford, C.*, et al.* 2012, '*Mycobacterium tuberculosis* – Heterogeneity Revealed Through Whole Genome Sequencing', *Tuberculosis (Edinburgh, Scotland)*, vol. 92, no. 3, pp. 194-201.
Ford, C. B.*, et al.* 2011, 'Use of Whole Genome Sequencing to Estimate the Mutation Rate of *Mycobacterium tuberculosis* During Latent Infection', *Nat Genet*, vol. 43, no. 5, pp. 482-6.
Forrellad, M. A.*, et al.* 2013, 'Virulence Factors of the *Mycobacterium tuberculosis* Complex', *Virulence*, vol. 4, no. 1, pp. 3-66.
Fozo, E. M.*, et al.* 2010, 'Abundance of Type I Toxin–Antitoxin Systems in Bacteria: Searches for New Candidates and Discovery of Novel Families', *Nucleic Acids Research*, vol. 38, no. 11, pp. 3743-3759.
Frost, P. & Nilsen, F. 2003, 'Validation of Reference Genes for Transcription Profiling in the Salmon Louse, *Lepeophtheirus salmonis*, by Quantitative Real-Time PCR', *Veterinary Parasitology*, vol. 118, no. 1–2, pp. 169-174.
Gagneux, S. 2012, 'Host-Pathogen Coevolution in Human Tuberculosis', *Philos Trans R Soc Lond B Biol Sci*, vol. 367, no. 1590, pp. 850-9.
Gagneux, S.*, et al.* 2006, 'Variable Host–Pathogen Compatibility in *Mycobacterium tuberculosis*', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 8, pp. 2869-2873.
Gagneux, S. & Small, P. M. 2007, 'Global Phylogeography of *Mycobacterium tuberculosis* and Implications for Tuberculosis Product Development', *The Lancet Infectious Diseases*, vol. 7, no. 5, pp. 328-337.
García de Viedma, D.*, et al.* 2004, 'Analysis of Clonal Composition of *Mycobacterium tuberculosis* Isolates in Primary Infections in Children', *Journal of Clinical Microbiology*, vol. 42, no. 8, pp. 3415-3418.
Gardy, J. L.*, et al.* 2011, 'Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak', *New England Journal of Medicine*, vol. 364, no. 8, pp. 730-739.
Gasteiger, E.*, et al.* 2005, 'Protein Identification and Analysis Tools on the ExPASy Server', in J. M. Walker (ed.), *The Proteomics Protocols Handbook*, Humana Press, pp. 571-607.
Geiman, D. E.*, et al.* 2006, 'Differential Gene Expression in Response to Exposure to Antimycobacterial Agents and Other Stress Conditions Among Seven *Mycobacterium tuberculosis* whiB-Like Genes', *Antimicrob Agents Chemother*, vol. 50, no. 8, pp. 2836-41.
Gerdes, K. 2012, *Prokaryotic Toxin-Antitoxins*, Springer,
Gerdes, K.*, et al.* 2005, 'Prokaryotic Toxin-Antitoxin Stress Response Loci', *Nat Rev Micro*, vol. 3, no. 5, pp. 371-382.

Gerdes, K.*, et al.* 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*, vol. 83, no. 10, pp. 3116-3120.
Gibson, A. L.*, et al.* 2008, 'Application of Sensitive and Specific Molecular Methods To Uncover Global Dissemination of the Major RDRio Sublineage of the Latin American-Mediterranean *Mycobacterium tuberculosis* Spoligotype Family', *Journal of Clinical Microbiology*, vol. 46, no. 4, pp. 1259-1267.
Girardi, E.*, et al.* 2000, 'Impact of the HIV Epidemic on the Spread of Other Diseases: the Case of Tuberculosis', *AIDS*, vol. 14 Suppl 3, no. pp. S47-56.
Gomez, J. E. & McKinney, J. D. 2004, '*M. tuberculosis* Persistence, Latency, and Drug Tolerance', *Tuberculosis (Edinburgh, Scotland)*, vol. 84, no. 1, pp. 29-44.
Gonzalez-y-Merchand, J. A.*, et al.* 1998, 'Roles of Multiple Promoters in Transcription of Ribosomal DNA: Effects of Growth Conditions on Precursor rRNA Synthesis in Mycobacteria', *J Bacteriol*, vol. 180, no. 21, pp. 5756-61.
González Barrios, A. F.*, et al.* 2006, 'Autoinducer 2 Controls Biofilm Formation in *Escherichia coli* through a Novel Motility Quorum-Sensing Regulator (MqsR, B3022)', *Journal of Bacteriology*, vol. 188, no. 1, pp. 305-316.
Goulard, C.*, et al.* 2010, 'The *Yersinia pestis* Chromosome Encodes Active Addiction Toxins', *J Bacteriol*, vol. 192, no. 14, pp. 3669-77.
Grassi, M.*, et al.* 2006, 'An Improved, Real-Time PCR Assay for the Detection of GC-Rich and Low Abundance Templates of *Mycobacterium tuberculosis*', *Journal of Microbiological Methods*, vol. 64, no. 3, pp. 406-410.
Greub, G. & Raoult, D. 2004, 'Microorganisms Resistant to Free-Living Amoebae', *Clinical Microbiology Reviews*, vol. 17, no. 2, pp. 413-433.
Gu, W.*, et al.* 2010, 'A Universal Trend of Reduced mRNA Stability Near the Translation-Initiation Site in Prokaryotes and Eukaryotes', *PLoS computational biology*, vol. 6, no. 2, pp. e1000664.
Gudnason, H.*, et al.* 2007, 'Comparison of Multiple DNA Dyes for Real-Time PCR: Effects of Dye Concentration and Sequence Composition on DNA Amplification and Melting Temperature', *Nucleic Acids Research*, vol. 35, no. 19, pp.
Guéroult, M.*, et al.* 2010, 'How Cations Can Assist DNase I in DNA Binding and Hydrolysis', *PLoS Comput Biol*, vol. 6, no. 11, pp. e1001000.
Gupta, A. 2009, 'Killing Activity and Rescue Function of Genome-Wide Toxin–Antitoxin Loci of *Mycobacterium tuberculosis*', *Fems Microbiology Letters*, vol. 290, no. 1, pp. 45-53.
Gutacker, M. M.*, et al.* 2006, 'Single-Nucleotide Polymorphism-Based Population Genetic Analysis of *Mycobacterium tuberculosis* Strains from 4 Geographic Sites', *J Infect Dis*, vol. 193, no. 1, pp. 121-8.
Gutierrez, M. C.*, et al.* 2005, 'Ancient Origin and Gene Mosaicism of the Progenitor of *Mycobacterium tuberculosis*', *PLoS Pathog*, vol. 1, no. 1, pp. e5.
Harries, A. D. & Dye, C. 2006, 'Tuberculosis', *Ann Trop Med Parasitol*, vol. 100, no. 5-6, pp. 415-31.
Harrison, J. J.*, et al.* 2009, 'The Chromosomal Toxin Gene yafQ Is a Determinant of Multidrug Tolerance for *Escherichia coli* Growing in a Biofilm', *ANTIMICROBIAL AGENTS AND CHEMOTHERAPY*, vol. 53, no. 6, pp. 2253-2258.
Hayes, F. & Van Melderen, L. 2011, 'Toxins-Antitoxins: Diversity, Evolution and Function', *Critical reviews in biochemistry and molecular biology*, vol. 46, no. 5, pp. 386-408.

Hazan, R.*, et al.* 2004, '*Escherichia coli* mazEF-Mediated Cell Death Is Triggered by Various Stressful Conditions', *Journal of Bacteriology*, vol. 186, no. 11, pp. 3663-3669.
Health, M. O. o. 2012, *Tuberculosis outbreak*, Public Health Report Hawkes Bay.
Helb, D.*, et al.* 2010, 'Rapid Detection of *Mycobacterium tuberculosis* and Rifampin Resistance by Use of On-Demand, Near-Patient Technology', *Journal of Clinical Microbiology*, vol. 48, no. 1, pp. 229-237.
Heras, B. & Martin, J. L. 2005, 'Post-Crystallization Treatments for Improving Diffraction Quality of Protein Crystals', *Acta Crystallographica Section D*, vol. 61, no. 9, pp. 1173-1180.
Hershberg, R.*, et al.* 2008, 'High Functional Diversity in *Mycobacterium tuberculosis* Driven by Genetic Drift and Human Demography', *PLoS Biol*, vol. 6, no. 12, pp. e311.
Hirsh, A. E.*, et al.* 2004, 'Stable Association Between Strains of *Mycobacterium tuberculosis* and Their Human Host Populations', *Proc Natl Acad Sci U S A*, vol. 101, no. 14, pp. 4871-6.
Hisert, K. B.*, et al.* 2004, 'Identification of *Mycobacterium tuberculosis* Counterimmune (cim) Mutants in Immunodeficient Mice by Differential Screening', *Infection and Immunity*, vol. 72, no. 9, pp. 5315-5321.
Huggett, J.*, et al.* 2005, 'Real-Time RT-PCR Normalisation; Strategies and Considerations', *Genes and Immunity*, vol. 6, no. 4, pp. 279-284.
Hurley, J. M.*, et al.* 2011, 'Bacterial Toxin RelE Mediates Frequent Codon-independent mRNA Cleavage from the 5' End of Coding Regions in Vivo', *Journal of Biological Chemistry*, vol. 286, no. 17, pp. 14770-14778.
Hurley, J. M. & Woychik, N. A. 2009, 'Bacterial Toxin HigB Associates with Ribosomes and Mediates Translation-dependent mRNA Cleavage at A-rich Sites', *Journal of Biological Chemistry*, vol. 284, no. 28, pp. 18605-18613.
Hurley, S. S.*, et al.* 1987, 'Rapid Lysis Technique for Mycobacterial Species', *J Clin Microbiol*, vol. 25, no. 11, pp. 2227-9.
Imbeaud, S.*, et al.* 2005, 'Towards Standardization of RNA Quality Assessment Using User-Independent Classifiers of Microcapillary Electrophoresis Traces', *Nucleic Acids Research*, vol. 33, no. 6, pp.
Ioerger, T. R.*, et al.* 2010, 'The Non-Clonality of Drug Resistance in Beijing-Genotype Isolates of *Mycobacterium tuberculosis* from the Western Cape of South Africa', *Bmc Genomics*, vol. 11, no. pp. 670.
Jacobs, M. 2012, *Communicable Disease Control Manual 2012*, Ministry of Health, Wellington, New Zealand.
Jakoby, M.*, et al.* 2000, 'AmtR, a Global Repressor in the Nitrogen Regulation System of *Corynebacterium glutamicum*', *Molecular Microbiology*, vol. 37, no. 4, pp. 964-977.
Jørgensen, M. G.*, et al.* 2009, 'HicA of *Escherichia coli* Defines a Novel Family of Translation-Independent mRNA Interferases in Bacteria and Archaea', *Journal of Bacteriology*, vol. 191, no. 4, pp. 1191-1199.
Kamerbeek, J.*, et al.* 1997, 'Simultaneous Detection and Strain Differentiation of *Mycobacterium tuberculosis* for Diagnosis and Epidemiology', *Journal of Clinical Microbiology*, vol. 35, no. 4, pp. 907-914.
Karlen, Y.*, et al.* 2007, 'Statistical Significance of Quantitative PCR', *BMC Bioinformatics*, vol. 8, no. pp. 131.

Kato-Maeda, M.*, et al.* 2010, 'Differences Among Sublineages of the East-Asian Lineage of *Mycobacterium tuberculosis* in Genotypic Clustering', *The International Journal of Tuberculosis and Lung Disease*, vol. 14, no. 5, pp. 538-544.
Kato-Maeda, M.*, et al.* 2012, 'Beijing Sublineages of *Mycobacterium tuberculosis* Differ in Pathogenicity in the Guinea Pig', *Clin Vaccine Immunol*, vol. 19, no. 8, pp. 1227-37.
Katoch, V. M. & Cox, R. A. 1986, 'Step-Wise Isolation of RNA and DNA from Mycobacteria', *Int J Lepr Other Mycobact Dis*, vol. 54, no. 3, pp. 409-15.
Keane, J.*, et al.* 1997, 'Infection by *Mycobacterium tuberculosis* Promotes Human Alveolar Macrophage Apoptosis', *Infection and Immunity*, vol. 65, no. 1, pp. 298-304.
Keane, J.*, et al.* 2000, 'Virulent *Mycobacterium tuberculosis* Strains Evade Apoptosis of Infected Alveolar Macrophages', *J Immunol*, vol. 164, no. 4, pp. 2016-20.
Keren, I.*, et al.* 2004, 'Specialized Persister Cells and the Mechanism of Multidrug Tolerance in *Escherichia coli*', *Journal of Bacteriology*, vol. 186, no. 24, pp. 8172-8180.
Kim, Y.*, et al.* 2009, 'Toxin-Antitoxin Systems in *Escherichia coli* Influence Biofilm Formation through YjgK (TabA) and Fimbriae', *Journal of Bacteriology*, vol. 191, no. 4, pp. 1258-1267.
Kim, Y. & Wood, T. K. 2010, 'Toxins Hha and CspD and Small RNA Regulator Hfq are Involved in Persister Cell Formation Through MqsR in *Escherichia coli*', *Biochemical and Biophysical Research Communications*, vol. 391, no. 1, pp. 209-213.
Kinger, A. K.*, et al.* 1993, 'A Method for the Isolation of Pure Intact RNA from Mycobacteria', *Biotechniques*, vol. 14, no. 5, pp. 724-5.
Klein, D.*, et al.* 2000, 'Assessment of Ribozyme Cleavage Efficiency Using Reverse Transcriptase Real-Time PCR', *Mol Biotechnol*, vol. 14, no. 3, pp. 189-95.
Kolodkin-Gal, I. & Engelberg-Kulka, H. 2009, 'The Stationary-Phase Sigma Factor σS Is Responsible for the Resistance of *Escherichia coli* Stationary-Phase Cells to mazEF-Mediated Cell Death', *Journal of Bacteriology*, vol. 191, no. 9, pp. 3177-3182.
Kolodkin-Gal, I.*, et al.* 2007, 'A Linear Pentapeptide is a Quorum-Sensing Factor Required for mazEF-Mediated Cell Death in *Escherichia coli*', *Science*, vol. 318, no. 5850, pp. 652-655.
Kolodkin-Gal, I.*, et al.* 2009, 'A Differential Effect of *E. coli* Toxin-Antitoxin Systems on Cell Death in Liquid Media and Biofilm Formation', *PLoS ONE*, vol. 4, no. 8, pp. e6785.
Kolter, R. & Losick, R. 1998, 'One for All and All for One', *Science*, vol. 280, no. 5361, pp. 226-227.
Korch, S. B.*, et al.* 2009, 'Three *Mycobacterium tuberculosis* Rel Toxin-Antitoxin Modules Inhibit Mycobacterial Growth and are Expressed in Infected Human Macrophages', *Journal of Bacteriology*, vol. 191, no. 5, pp. 1618-1630.
Korch, S. B.*, et al.* 2003, 'Characterization of the hipA7 Allele of *Escherichia coli* and Evidence that High Persistence is Governed by (p)ppGpp Synthesis', *Molecular Microbiology*, vol. 50, no. 4, pp. 1199-1213.
Korch, S. B. & Hill, T. M. 2006, 'Ectopic Overexpression of Wild-Type and Mutant hipA Genes in *Escherichia coli*: Effects on Macromolecular Synthesis and Persister Formation', *Journal of Bacteriology*, vol. 188, no. 11, pp. 3826-3836.
Koressaar, T. & Remm, M. 2007, 'Enhancements and Modifications of Primer Design Program Primer3', *Bioinformatics*, vol. 23, no. 10, pp. 1289-1291.

Kotłowski, R.*, et al.* 2004, 'PCR-Based Genotyping of *Mycobacterium tuberculosis* with New GC-Rich Repeated Sequences and IS*6110* Inverted Repeats Used as Primers', *Journal of Clinical Microbiology*, vol. 42, no. 1, pp. 372-377.
Kovarova, M. & Draber, P. 2000, 'New Specificity and Yield Enhancer of Polymerase Chain Reactions', *Nucleic Acids Res*, vol. 28, no. 13, pp. E70.
Kwan, A. H. Y.*, et al.* 2003, 'Pentaprobe: a Comprehensive Sequence for the One-Step Detection of DNA-Binding Activities', *Nucleic Acids Research*, vol. 31, no. 20, pp. e124-e124.
Lemos, J. A. C.*, et al.* 2005, 'Characteristics of *Streptococcus mutans* Strains Lacking the MazEF and RelBE Toxin–Antitoxin Modules', *Fems Microbiology Letters*, vol. 253, no. 2, pp. 251-257.
Leplae, R.*, et al.* 2011, 'Diversity of Bacterial Type II Toxin–Antitoxin Systems: a Comprehensive Search and Functional Analysis of Novel Families', *Nucleic Acids Research*, vol. 39, no. 13, pp. 5513-5525.
Leung, E. T. Y.*, et al.* 2011, 'Rapid and Simultaneous Detection of *Mycobacterium tuberculosis* Complex and Beijing/W Genotype in Sputum by an Optimized DNA Extraction Protocol and a Novel Multiplex Real-Time PCR', *Journal of Clinical Microbiology*, vol. 49, no. 7, pp. 2509-2515.
Lewin, B. 1990, *Genes IV*, Oxford University Press, Oxford.
Lewis, K. 2008, 'Multidrug Tolerance of Biofilms and Persister Cells', in T. Romeo (ed.), *Bacterial Biofilms*, Springer Berlin Heidelberg, pp. 107-131.
Lim, E. & Heffernan, H. 2013, *Tuberculosis in New Zealand: Annual Report 2012*, Institute of Environmental Science and Research Ltd (ESR), Porirua, New Zealand.
Liu, Z. L. & Slininger, P. J. 2007, 'Universal External RNA Controls for Microbial Gene Expression Analysis Using Microarray and qRT-PCR', *Journal of Microbiological Methods*, vol. 68, no. 3, pp. 486-496.
Livak, K. J. & Schmittgen, T. D. 2001, 'Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2(-Delta Delta C(T)) Method ', *Methods*, vol. 25, no. 4, pp. 402-408.
Lopez, B.*, et al.* 2003, 'A Marked Difference in Pathogenesis and Immune Response Induced by Different *Mycobacterium tuberculosis* Genotypes', *Clinical & Experimental Immunology*, vol. 133, no. 1, pp. 30-37.
Lun, S. & Bishai, W. R. 2007, 'Characterization of a Novel Cell Wall-anchored Protein with Carboxylesterase Activity Required for Virulence in *Mycobacterium tuberculosis*', *Journal of Biological Chemistry*, vol. 282, no. 25, pp. 18348-18356.
Magombedze, G. & Mulder, N. 2013, 'Understanding TB Latency Using Computational and Dynamic Modelling Procedures', *Infect Genet Evol*, vol. 13, no. pp. 267-83.
Maisonneuve, E.*, et al.* 2011, 'Bacterial Persistence by RNA Endonucleases', *Proceedings of the National Academy of Sciences*, vol. 108, no. 32, pp. 13206-13211.
Manca, C.*, et al.* 2004, 'Differential Monocyte Activation Underlies Strain-Specific *Mycobacterium tuberculosis* Pathogenesis', *Infection and Immunity*, vol. 72, no. 9, pp. 5511-5514.
Mangan, J. A.*, et al.* 1997, 'An Effective Method of RNA Extraction from Bacteria Refractory to Disruption, Including Mycobacteria', *Nucl. Acids Res.*, vol. 25, no. 3, pp. 675-676.

Marais, B. J.*, et al.* 2006, 'Beijing and Haarlem Genotypes Are Overrepresented among Children with Drug-Resistant Tuberculosis in the Western Cape Province of South Africa', *Journal of Clinical Microbiology*, vol. 44, no. 10, pp. 3539-3543.
Martin, A.*, et al.* 2007, 'Rapid Clonal Analysis of Recurrent Tuberculosis by Direct MIRU-VNTR Typing on Stored Isolates', *BMC Microbiol*, vol. 7, no. pp. 73.
Mathews, D. H.*, et al.* 1999, 'Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure', *J Mol Biol*, vol. 288, no. 5, pp. 911-40.
Mattison, K.*, et al.* 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*, vol. 281, no. 49, pp. 37942-37951.
McElnay, C.*, et al.* 2004, 'A Community and Workplace Outbreak of Tuberculosis in Hawke's Bay in 2002', *N Z Med J*, vol. 117, no. 1200, pp. U1019.
McGrath, K. C.*, et al.* 2008, 'Isolation and Analysis of mRNA from Environmental Microbial Communities', *J Microbiol Methods*, vol. 75, no. 2, pp. 172-6.
McKenzie, J. L. 2011, The Biochemistry of VapBC Toxin-Antitoxins, Thesis, University of Waikato.
McKenzie, J. L.*, et al.* 2012a, 'Determination of Ribonuclease Sequence-Specificity Using Pentaprobes and Mass Spectrometry', *RNA*, vol. 18, no. 6, pp. 1267-1278.
McKenzie, J. L.*, et al.* 2012b, 'A VapBC Toxin-Antitoxin Module Is a Posttranscriptional Regulator of Metabolic Flux in Mycobacteria', *Journal of Bacteriology*, vol. 194, no. 9, pp. 2189-2204.
Mehndiratta, M.*, et al.* 2008, 'Fluorescence Acquisition During Hybridization Phase in Quantitative Real-Time PCR Improves Specificity and Signal-to-Noise Ratio', *Biotechniques*, vol. 45, no. 6, pp. 625-634.
Meyers, P. R.*, et al.* 1998, 'Novel Method for Rapid Measurement of Growth of Mycobacteria in Detergent-Free Media', *Journal of Clinical Microbiology*, vol. 36, no. 9, pp. 2752-2754.
Miallau, L.*, et al.* 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*, vol. 284, no. 1, pp. 276-283.
Milano, A.*, et al.* 2001, 'Transcriptional Regulation of furA and katG Upon Oxidative Stress in *Mycobacterium smegmatis*', *J Bacteriol*, vol. 183, no. 23, pp. 6801-6.
Min, A. B.*, et al.* 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*, vol. 21, no. 11, pp. 1754-1767.
Mine, N.*, et al.* 2009, 'The Decay of the Chromosomally Encoded ccdO157 Toxin–Antitoxin System in the *Escherichia coli* Species', *Genetics*, vol. 181, no. 4, pp. 1557-1566.
Mitchell, H. L.*, et al.* 2010, '*Treponema denticola* Biofilm-Induced Expression of a Bacteriophage, Toxin–Antitoxin Systems and Transposases', *Microbiology*, vol. 156, no. 3, pp. 774-788.
Mostowy, S.*, et al.* 2004, 'Genomic Analysis Distinguishes *Mycobacterium africanum*', *Journal of Clinical Microbiology*, vol. 42, no. 8, pp. 3594-3599.
Muñoz-Gómez, A. J.*, et al.* 2005, 'RNase/Anti-RNase Activities of the Bacterial parD Toxin-Antitoxin System', *Journal of Bacteriology*, vol. 187, no. 9, pp. 3151-3157.

Nahid, P.*, et al.* 2010, 'Influence of *M. tuberculosis* Lineage Variability within a Clinical Trial for Pulmonary Tuberculosis', *PLoS ONE*, vol. 5, no. 5, pp. e10753.
Nariya, H. & Inouye, M. 2008, 'MazF, an mRNA Interferase, Mediates Programmed Cell Death during Multicellular Myxococcus Development', *Cell*, vol. 132, no. 1, pp. 55-66.
Niemann, S.*, et al.* 2009, 'Genomic Diversity among Drug Sensitive and Multidrug Resistant Isolates of *Mycobacterium tuberculosis* with Identical DNA Fingerprints', *PLoS ONE*, vol. 4, no. 10, pp. e7407.
Nolan, T.*, et al.* 2006, 'Quantification of mRNA Using Real-Time RT-PCR', *Nature Protocols*, vol. 1, no. 3, pp. 1559-1582.
Ntolosi, B. A.*, et al.* 2001, 'Growth Phase-Associated Changes in Protein Expression in *Mycobacterium smegmatis* Identify a New Low Molecular Weight Heat Shock Protein', *Tuberculosis*, vol. 81, no. 4, pp. 279-289.
Oelemann, M. C.*, et al.* 2007, 'Assessment of an Optimized Mycobacterial Interspersed Repetitive- Unit-Variable-Number Tandem-Repeat Typing System Combined with Spoligotyping for Population-Based Molecular Epidemiology Studies of Tuberculosis', *Journal of Clinical Microbiology*, vol. 45, no. 3, pp. 691-697.
Ojha, A. & Hatfull, G. F. 2007, 'The Role of Iron in *Mycobacterium smegmatis* Biofilm Formation: the Exochelin Siderophore is Essential in Limiting Iron Conditions for Biofilm Formation but Not for Planktonic Growth', *Molecular Microbiology*, vol. 66, no. 2, pp. 468-483.
Otal, I.*, et al.* 1997, 'Use of a PCR Method Based on IS*6110* Polymorphism for Typing *Mycobacterium tuberculosis* Strains from BACTEC Cultures', *Journal of Clinical Microbiology*, vol. 35, no. 1, pp. 273-277.
Ozcaglar, C.*, et al.* 2011, 'Sublineage Structure Analysis of *Mycobacterium tuberculosis* Complex Strains Using Multiple-Biomarker Tensors', *Bmc Genomics*, vol. 12, no. 2, pp. 1-26.
Pandey, D. P. & Gerdes, K. 2005, 'Toxin–Antitoxin Loci are Highly Abundant in Free-Living but Lost From Host-Associated Prokaryotes', *Nucleic Acids Research*, vol. 33, no. 3, pp. 966-976.
Park, J.-H.*, et al.* 2011, '*Bacillus subtilis* MazF-bs (EndoA) is a UACAU-Specific mRNA Interferase', *FEBS letters*, vol. 585, no. 15, pp. 2526-2532.
Patel, B. K. R.*, et al.* 1991, 'Extraction and Characterization of mRNA from Mycobacteria: Implication for Virulence Gene Identification', *Journal of Microbiological Methods*, vol. 13, no. 2, pp. 99-111.
Pedersen, K.*, et al.* 2002, 'Rapid Induction and Reversal of a Bacteriostatic Condition by Controlled Expression of Toxins and Antitoxins', *Molecular Microbiology*, vol. 45, no. 2, pp. 501-510.
Pedersen, K.*, et al.* 2003, 'The Bacterial Toxin RelE Displays Codon-Specific Cleavage of mRNAs in the Ribosomal A Site', *Cell*, vol. 112, no. 1, pp. 131-40.
Perez-Novo, C. A.*, et al.* 2005, 'Impact of RNA Quality on Reference Gene Expression Stability', *Biotechniques*, vol. 39, no. 1, pp. 52-56.
Pfaffl, M. 2004, 'Quantification Strategies in Real-Time PCR', in S. Bustin (ed.), *A-Z of quantitative PCR*, International University Line La Jolla, CA, pp. 87-112.
Pfaffl, M. W. 2001, 'A New Mathematical Model for Relative Quantification in Real-Time RT-PCR', *Nucleic Acids Res*, vol. 29, no. 9, pp. e45.

Pfaffl, M. W.*, et al.* 2002, 'Relative Expression Software Tool (REST(C)) for Group-Wise Comparison and Statistical Analysis of Relative Expression Results in Real-Time PCR', *Nucl. Acids Res.*, vol. 30, no. 9, pp. e36-.
Pfaffl, M. W.*, et al.* 2004, 'Determination of Stable Housekeeping Genes, Differentially Regulated Target Genes and Sample Integrity: BestKeeper--Excel-Based Tool Using Pair-Wise Correlations', *Biotechnol Lett*, vol. 26, no. 6, pp. 509-15.
Pullinger, G. D. & Lax, A. J. 1992, 'A *Salmonella dublin* Virulence Plasmid Locus that Affects Bacterial Growth Under Nutrient-Limited Conditions', *Molecular Microbiology*, vol. 6, no. 12, pp. 1631-1643.
Raeymaekers, L. 1993, 'Quantitative PCR: Theoretical Considerations with Practical Implications', *Anal Biochem*, vol. 214, no. 2, pp. 582-5.
Ramage, H. R.*, et al.* 2009, 'Comprehensive Functional Analysis of *Mycobacterium tuberculosis* Toxin-Antitoxin Systems: Implications for Pathogenesis, Stress Responses, and Evolution', *PLoS Genet*, vol. 5, no. 12, pp. e1000767.
Ramakers, C.*, et al.* 2003, 'Assumption-Free Analysis of Quantitative Real-Time Polymerase Chain Reaction (PCR) Data', *Neuroscience Letters*, vol. 339, no. 1, pp. 62-66.
Reddy, T. B. K.*, et al.* 2009, 'TB Database: an Integrated Platform for Tuberculosis Research', *Nucleic Acids Research*, vol. 37, no. suppl 1, pp. D499-D508.
Reed, M. B.*, et al.* 2009, 'Major *Mycobacterium tuberculosis* Lineages Associate with Patient Country of Origin', *J. Clin. Microbiol.*, vol. 47, no. 4, pp. 1119-1128.
Regnier, P. & Arraiano, C. M. 2000, 'Degradation of mRNA in Bacteria: Emergence of Ubiquitous Features', *Bioessays*, vol. 22, no. 3, pp. 235-44.
Rindi, L.*, et al.* 2012, 'Large Sequence Polymorphisms of the Euro-American lineage of *Mycobacterium tuberculosis*: A phylogenetic Reconstruction and Evidence for Convergent Evolution in the DR Locus', *Infection, Genetics and Evolution*, vol. 12, no. 7, pp. 1551-1557.
Ritz, M.*, et al.* 2009, 'Determination of rpoA as the Most Suitable Internal Control to Study Stress Response in *C. jejuni* by RT-qPCR and Application to Oxidative Stress', *Journal of Microbiological Methods*, vol. 76, no. 2, pp. 196-200.
Robinson, T. L.*, et al.* 2007, 'Validation of Candidate Bovine Reference Genes for Use with Real-Time PCR', *Vet Immunol Immunopathol*, vol. 115, no. 1-2, pp. 160-5.
Robson, J.*, et al.* 2009, 'The vapBC Operon from *Mycobacterium smegmatis* Is An Autoregulated Toxin–Antitoxin Module That Controls Growth via Inhibition of Translation', *Journal of Molecular Biology*, vol. 390, no. 3, pp. 353-367.
Rocha, E. P. C.*, et al.* 1999, 'Translation in *Bacillus subtilis*: Roles and Trends of Initiation and Termination, Insights from a Genome Analysis', *Nucleic Acids Research*, vol. 27, no. 17, pp. 3567-3576.
Roetzer, A.*, et al.* 2011, 'Evaluation of *Mycobacterium tuberculosis* Typing Methods in a 4-Year Study in Schleswig-Holstein, Northern Germany', *Journal of Clinical Microbiology*, vol. 49, no. 12, pp. 4173-4178.
Rotem, E.*, et al.* 'Regulation of Phenotypic Variability by a Threshold-Based Mechanism Underlies Bacterial Persistence', *Proceedings of the National Academy of Sciences*, vol. 107, no. 28, pp. 12541-12546.
Ruijter, J. M.*, et al.* 2009, 'Amplification Efficiency: Linking Baseline and Bias in the Analysis of Quantitative PCR Data', *Nucl. Acids Res.*, vol. 37, no. 6, pp. e45.

Russell, D. G. 2007, 'Who Puts the Tubercle in Tuberculosis?', *Nat Rev Microbiol*, vol. 5, no. 1, pp. 39-47.
Saida, F.*, et al.* 2006, 'Expression of Highly Toxic Genes in *E. coli*: Special Strategies and Genetic Tools', *Current Protein & Peptide Science*, vol. 7, no. 1, pp. 47-56.
Saito, H.*, et al.* 1983, 'Mycobacteriocins Produced by Rapidly Growing Mycobacteria are Tween-Hydrolyzing Esterases', *J Bacteriol*, vol. 153, no. 3, pp. 1294-300.
Sambrook, J.*, et al.* 1989, *Molecular cloning : a laboratory manual* Cold Spring Harbor Laboratory Press, New York.
Sampson, S. L.*, et al.* 1999, 'Disruption of Coding Regions by IS*6110* Insertion in *Mycobacterium tuberculosis*', *Tubercle and Lung Disease*, vol. 79, no. 6, pp. 349-359.
Saunders, N. J.*, et al.* 2011, 'Deep Resequencing of Serial Sputum Isolates of *Mycobacterium tuberculosis* During Therapeutic Failure due to Poor Compliance Reveals Stepwise Mutation of Key Resistance Genes on an Otherwise Stable Genetic Background', *The Journal of infection*, vol. 62, no. 3, pp. 212-217.
Sayeed, S.*, et al.* 2000, 'The Stability Region of the Large Virulence Plasmid of *Shigella flexneri* Encodes an Efficient Postsegregational Killing System', *Journal of Bacteriology*, vol. 182, no. 9, pp. 2416-2421.
Scherl, A.*, et al.* 2005, 'Correlation of Proteomic and Transcriptomic Profiles of *Staphylococcus aureus* During the Post-Exponential Phase of Growth', *Journal of Microbiological Methods*, vol. 60, no. 2, pp. 247-257.
Schmidt, O.*, et al.* 2007, 'prlF and yhaV Encode a New Toxin–Antitoxin System in *Escherichia coli*', *Journal of Molecular Biology*, vol. 372, no. 4, pp. 894-905.
Schneider, K. L.*, et al.* 2006, 'The UCSC Archaeal Genome Browser', *Nucleic Acids Research*, vol. 34, no. suppl 1, pp. D407-D410.
Schoor, O.*, et al.* 2003, 'Moderate Degradation Does Not Preclude Microarray Analysis of Small Amounts of RNA', *Biotechniques*, vol. 35, no. 6, pp. 1192-6, 1198-201.
Schumacher, M. A.*, et al.* 2009, 'Molecular Mechanisms of HipA-Mediated Multidrug Tolerance and Its Neutralization by HipB', *Science*, vol. 323, no. 5912, pp. 396-401.
Sevin, E. & Barloy-Hubler, F. 2007, 'RASTA-Bacteria: a Web-Based Tool for Identifying Toxin-Antitoxin Loci in Prokaryotes', *Genome Biology*, vol. 8, no. 8, pp. R155.
Sexton, K.*, et al.* 2008, *Five Years of Molecular Typing of Mycobacterium tuberculosis Isolates in New Zealand, 2003 to 2007*, Institute of Environmental Science and Research Limited ("ESR"), Porirua, New Zealand.
Shah, D.*, et al.* 2006, 'Persisters: a Distinct Physiological State of *E-coli*', *Bmc Microbiology*, vol. 6, no. pp. 53.
Shao, Y.*, et al.* 2011, 'TADB: a Web-Based Resource for Type 2 Toxin–Antitoxin Loci in Bacteria and Archaea', *Nucleic Acids Research*, vol. 39, no. suppl 1, pp. D606-D611.
Shires, K. & Steyn, L. 2001, 'The Cold-Shock Stress Response in *Mycobacterium smegmatis* Induces the Expression of a Histone-Like Protein', *Mol Microbiol*, vol. 39, no. 4, pp. 994-1009.
Shleeva, M. O.*, et al.* 2011, 'Dormant Ovoid Cells of *Mycobacterium tuberculosis* are Formed in Response to Gradual External Acidification', *Tuberculosis (Edinb)*, vol. 91, no. 2, pp. 146-54.
Short, F. L.*, et al.* 2013, 'Selectivity and Self-Assembly in the Control of a Bacterial Toxin by an Antitoxic Noncoding RNA Pseudoknot', *Proceedings of the National Academy of Sciences*, vol. 110, no. 3, pp. E241-E249.

Singh, A. K. & Reyrat, J. M. 2009, 'Laboratory Maintenance of *Mycobacterium smegmatis*', *Curr Protoc Microbiol*, vol. Chapter 10, no. pp. Unit10C 1.
Singh, A. K. & Singh, B. N. 2009, 'Differential Expression of sigH Paralogs During Growth and Under Different Stress Conditions in *Mycobacterium smegmatis*', *J Bacteriol*, vol. 191, no. 8, pp. 2888-93.
Smeulders, M. J.*, et al.* 1999, 'Adaptation of *Mycobacterium smegmatis* to Stationary Phase', *Journal of Bacteriology*, vol. 181, no. 1, pp. 270-283.
Smith, N. H.*, et al.* 2006a, 'Bottlenecks and Broomsticks: the Molecular Evolution of *Mycobacterium bovis*', *Nat Rev Microbiol*, vol. 4, no. 9, pp. 670-81.
Smith, N. H.*, et al.* 2006b, 'Ecotypes of the *Mycobacterium tuberculosis* Complex', *J Theor Biol*, vol. 239, no. 2, pp. 220-5.
Smith, T. 1898, 'A Comparative Study of Bovine Tubercle Bacilli and of Human Bacilli from Sputum', *J Exp Med*, vol. 3, no. 4-5, pp. 451-511.
Smittipat, N.*, et al.* 2005, 'Polymorphism of Variable-Number Tandem Repeats at Multiple Loci in *Mycobacterium tuberculosis*', *Journal of Clinical Microbiology*, vol. 43, no. 10, pp. 5034-5043.
Snapper, S. B.*, et al.* 1990, 'Isolation and Characterization of Efficient Plasmid Transformation Mutants of *Mycobacterium-Smegmatis*', *Molecular Microbiology*, vol. 4, no. 11, pp. 1911-1919.
Sreevatsan, S.*, et al.* 1997, 'Restricted Structural Gene Polymorphism in the *Mycobacterium tuberculosis* Complex Indicates Evolutionarily Recent Global Dissemination', *Proceedings of the National Academy of Sciences*, vol. 94, no. 18, pp. 9869-9874.
Stinear, T. P.*, et al.* 2008, 'Insights from the Complete Genome Sequence of *Mycobacterium marinum* on the Evolution of *Mycobacterium tuberculosis*', *Genome Research*, vol. 18, no. 5, pp. 729-741.
Supply, P.*, et al.* 2006, 'Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of *Mycobacterium tuberculosis*', *Journal of Clinical Microbiology*, vol. 44, no. 12, pp. 4498-4510.
Supply, P.*, et al.* 2000, 'Variable Human Minisatellite-Like Regions in the *Mycobacterium tuberculosis* Genome', *Molecular Microbiology*, vol. 36, no. 3, pp. 762-771.
Supply, P.*, et al.* 2003, 'Linkage Disequilibrium Between Minisatellite Loci Supports Clonal Evolution of *Mycobacterium tuberculosis* in a High Tuberculosis Incidence Area', *Mol Microbiol*, vol. 47, no. 2, pp. 529-38.
Sutherland, I. 1976, 'Recent Studies in the Epidemiology of Tuberculosis, Based on the Risk of Being Infected with Tubercle Bacilli', *Adv Tuberc Res*, vol. 19, no. pp. 1-63.
Suzuki, M.*, et al.* 2005, 'Single Protein Production in Living Cells Facilitated by an mRNA Interferase', *Molecular Cell*, vol. 18, no. 2, pp. 253-261.
Suzuki, T.*, et al.* 2000, 'Control Selection for RNA Quantitation', *Biotechniques*, vol. 29, no. 2, pp. 332-7.
Szekeres, S.*, et al.* 2007, 'Chromosomal Toxin–Antitoxin Loci can Diminish Large-Scale Genome Reductions in the Absence of Selection', *Molecular Microbiology*, vol. 63, no. 6, pp. 1588-1605.
Tadokoro, T. & Kanaya, S. 2009, 'Ribonuclease H: Molecular Diversities, Substrate Binding Domains, and Catalytic Mechanism of the Prokaryotic Enzymes', *FEBS Journal*, vol. 276, no. 6, pp. 1482-1493.

Talaat, A. M.*, et al.* 1999, 'Pathogenicity of *Mycobacterium fortuitum* and *Mycobacterium smegmatis* to Goldfish, *Carassius auratus*', *Vet Microbiol*, vol. 66, no. 2, pp. 151-64.
Tang, Y. J.*, et al.* 2009, 'Central Metabolism in *Mycobacterium smegmatis* During the Transition from O_2-rich to O_2-Poor Conditions as Studied by Isotopomer-Assisted Metabolite Analysis', *Biotechnol Lett*, vol. 31, no. 8, pp. 1233-40.
Thellin, O.*, et al.* 1999, 'Housekeeping Genes as Internal Standards: Use and Limits', *Journal of Biotechnology*, vol. 75, no. 2-3, pp. 291-295.
Thierry, D.*, et al.* 1990, 'Characterization of a *Mycobacterium tuberculosis* Insertion Sequence, IS*6110*, and its Application in Diagnosis', *Journal of Clinical Microbiology*, vol. 28, no. 12, pp. 2668-2673.
Thwaites, G.*, et al.* 2008, 'Relationship between *Mycobacterium tuberculosis* Genotype and the Clinical Phenotype of Pulmonary and Meningeal Tuberculosis', *Journal of Clinical Microbiology*, vol. 46, no. 4, pp. 1363-1368.
Tian, Q. B.*, et al.* 1996, 'Gene Product Identification and Promoter Analysis of *hig* Locus of Plasmid Rts1', *Biochemical and Biophysical Research Communications*, vol. 225, no. 2, pp. 679-684.
Tsilibaris, V.*, et al.* 2007, 'What is the Benefit to *Escherichia coli* of Having Multiple Toxin-Antitoxin Systems in its Genome?', *Journal of Bacteriology*, vol. 189, no. 17, pp. 6101-6108.
Tsolaki, A. G.*, et al.* 2004, 'Functional and Evolutionary Genomics of *Mycobacterium tuberculosis*: Insights From Genomic Deletions in 100 Strains', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 14, pp. 4865-4870.
van Embden, J. D.*, et al.* 1993, 'Strain Identification of *Mycobacterium tuberculosis* by DNA Fingerprinting: Recommendations for a Standardized Methodology', *Journal of Clinical Microbiology*, vol. 31, no. 2, pp. 406-409.
Van Melderen, L. 2010, 'Toxin–Antitoxin Systems: Why So Many, What For?', *Current Opinion in Microbiology*, vol. 13, no. 6, pp. 781-785.
Van Melderen, L. & Saavedra De Bast, M. 2009, 'Bacterial Toxin–Antitoxin Systems: More Than Selfish Entities?', *PLoS Genet*, vol. 5, no. 3, pp. e1000437.
Vandesompele, J.*, et al.* 2002, 'Accurate Normalization of Real-Time Quantitative RT-PCR Data by Geometric Averaging of Multiple Internal Control Genes', *Genome Biology*, vol. 3, no. 7, pp. 1-12.
Velayati, A. A.*, et al.* 2009, 'Totally Drug-Resistant Tuberculosis Strains: Evidence of Adaptation at the Cellular Level', *Eur Respir J*, vol. 34, no. 5, pp. 1202-3.
Verma, A.*, et al.* 1999, '*Mycobacterium tuberculosis* rrn Promoters: Differential Usage and Growth Rate-Dependent Control', *Journal of Bacteriology*, vol. 181, no. 14, pp. 4326-4333.
Vesper, O.*, et al.* 2011, 'Selective Translation of Leaderless mRNAs by Specialized Ribosomes Generated by MazF in *Escherichia coli*', *Cell*, vol. 147, no. 1, pp. 147-157.
Vitzthum, F.*, et al.* 1999, 'A Quantitative Fluorescence-Based Microplate Assay for the Determination of Double-Stranded DNA Using SYBR Green I and a Standard Ultraviolet Transilluminator Gel Imaging System', *Anal Biochem*, vol. 276, no. 1, pp. 59-64.

Wang, X. & Wood, T. K. 2011, 'Toxin-Antitoxin Systems Influence Biofilm and Persister Cell Formation and the General Stress Response', *Applied and Environmental Microbiology*, vol. 77, no. 16, pp. 5577-5583.
Wang, X. M.*, et al.* 2008, 'IS*1096*-Mediated DNA Rearrangements Play a Key Role in Genome Evolution of *Mycobacterium smegmatis*', *Tuberculosis (Edinb)*, vol. 88, no. 5, pp. 399-409.
Wayne, L. G. 1994, 'Cultivation of *Mycobacterium tuberculosis* for Research Purposes', in B. R. Bloom (ed.), *Tuberculosis: Pathogenesis, Protection, and Control.*, ASM Press, Washington, D.C., pp. 73-83.
Wilbur, J. S.*, et al.* 2005, '*Neisseria gonorrhoeae* FitA Interacts with FitB To Bind DNA through Its Ribbon−Helix−Helix Motif', *Biochemistry*, vol. 44, no. 37, pp. 12515-12524.
Wilkening, S. & Bader, A. 2004, 'Quantitative Real-Time Polymerase Chain Reaction: Methodical Analysis and Mathematical Model', *J Biomol Tech*, vol. 15, no. 2, pp. 107-111.
Winther, K. S. & Gerdes, K. 2011, 'Enteric Virulence Associated Protein VapC Inhibits Translation by Cleavage of Initiator tRNA', *Proceedings of the National Academy of Sciences*, vol. 108, no. 18, pp. 7403-7407.
Wirth, T.*, et al.* 2008, 'Origin, Spread and Demography of the *Mycobacterium tuberculosis* Complex', *PLoS Pathog*, vol. 4, no. 9, pp. e1000160.
Wong, C.*, et al.* 2000, 'Heating Greatly Speeds Coomassie Blue Staining and Destaining', *Bio Techniques*, vol. 28, no. 3, pp. 426-432.
Xu, J.*, et al.* 2007, 'A Unique Mycobacterium ESX-1 Protein Co-Secretes with CFP-10/ESAT-6 and is Necessary for Inhibiting Phagosome Maturation', *Molecular Microbiology*, vol. 66, no. 3, pp. 787-800.
Yamaguchi, Y. & Inouye, M. 2009, 'Chapter 12 mRNA Interferases, Sequence-Specific Endoribonucleases from the Toxin–Antitoxin Systems', in C. Ciaran (ed.), *Progress in Molecular Biology and Translational Science*, Academic Press, pp. 467-500.
Yamaguchi, Y. & Inouye, M. 2011, 'Regulation of Growth and Death in *Escherichia coli* by Toxin–Antitoxin Systems', *Nat Rev Micro*, vol. 9, no. 11, pp. 779-790.
Yang, I. V.*, et al.* 2002, 'Within the Fold: Assessing Differential Expression Measures and Reproducibility in Microarray Assays', *Genome Biol*, vol. 3, no. 11, pp. research0062.
Zhang, J. & Li, K. 2003, 'Single-Base Discrimination Mediated by Proofreading 3' Phosphorothioate-Modified Primers', *Molecular Biotechnology*, vol. 25, no. 3, pp. 223-227.
Zhang, L.*, et al.* 2014, 'A Novel Method of Identifying *Mycobacterium tuberculosis* Beijing Strains by Detecting SNPs in Rv0444c and Rv2629', *Current Microbiology*, vol. 68, no. 3, pp. 381-386.
Zhang, M.*, et al.* 1999, 'Enhanced Capacity of a Widespread Strain of *Mycobacterium tuberculosis* to Grow in Human Macrophages', *J Infect Dis*, vol. 179, no. 5, pp. 1213-7.
Zhang, Y.*, et al.* 2009, 'Characterization of YafO, an *Escherichia coli* Toxin', *Journal of Biological Chemistry*, vol. 284, no. 38, pp. 25522-25531.
Zhang, Y. L.*, et al.* 2005, 'Insights into the mRNA Cleavage Mechanism by MazF, an mRNA Interferase', *Journal of Biological Chemistry*, vol. 280, no. 5, pp. 3143-3150.
Zhang, Y. L.*, et al.* 2003, 'MazF Cleaves Cellular mRNAs Specifically at ACA to Block Protein Synthesis in *Escherichia coli*', *Molecular Cell*, vol. 12, no. 4, pp. 913-923.

Zhu, L.*, et al.* 2009, '*Staphylococcus aureus* MazF Specifically Cleaves a Pentad Sequence, UACAU, Which Is Unusually Abundant in the mRNA for Pathogenic Adhesive Factor SraP', *Journal of Bacteriology*, vol. 191, no. 10, pp. 3248-3255.
Zhu, L.*, et al.* 2008, 'The mRNA Interferases, MazF-mt3 and MazF-mt7 from *Mycobacterium tuberculosis* Target Unique Pentad Sequences in Single-Stranded RNA', *Molecular Microbiology*, vol. 69, no. 3, pp. 559-569.
Zhu, L.*, et al.* 2010, 'Noncognate *Mycobacterium tuberculosis* Toxin-Antitoxins Can Physically and Functionally Interact', *Journal of Biological Chemistry*, vol. 285, no. 51, pp. 39732-39738.
Zhu, L.*, et al.* 2006, 'Characterization of mRNA Interferases from *Mycobacterium tuberculosis*', *Journal of Biological Chemistry*, vol. 281, no. 27, pp. 18638-18643.
Zuber, B.*, et al.* 2008, 'Direct Visualization of the Outer Membrane of Mycobacteria and Corynebacteria in their Native State', *J Bacteriol*, vol. 190, no. 16, pp. 5672-80.
Zuker, M. 2003, 'Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction', *Nucleic Acids Research*, vol. 31, no. 13, pp. 3406-3415.

Appendices

Appendix A: Reagents

A.1 Primers Used in This Study

Table A.1: A list of primers and their sequence, used in Chapters Three, Four and Five.

Primer	Sequence
Chapter Three:	
Mtb1	CCGGCGGGGCCGGCGG
Mtb2	CGGCGGCAACGGCGGC
IS1	CGGACTCACCGGGGCGGTTCA
IS2	CGGACATGCCGGGGCGGTTCA
CAR2	GACIIICCGGGGCGGTTCA
Lineage Classification Primers	
RD9 Fwd	GTGACGGTATCGTCGAGCAG
RD9 Rev_{Int}	CGCGCCAACTATGTCTACGG
RD9 Rev_{Del}	GCTCGAGCTAGACCTGCAC
RD105 Fwd	GGTCATATCACGCGTTCGTG
RD105 Rev_{Int}	TGCGGTCAAAGCACGCCTTG
RD105 Rev_{Del}	GGTGGCCCAGAAACCACCA
RD239 Fwd	CGTAGACTGCTCGCATGACC
RD239 Rev_{Int}	CAGTGAGATCCCAAATGCTGC
RD239 Rev_{Del}	CGACGCAATCTGACCGACAG
RD750 Fwd	GTCAACTGCCGATGGCTGAC
RD750 Rev_{Int}	CGTCAGCGATGATCACCTCG
RD750 Rev_{Del}	GTGAACTAGGTCGAGCATCG
katG-F	GCCTTGGGCTCCAGCACG
katG-R1	CAGCCTTAAGAGCCAGATCCPST
katG-R2/3	GCCTTAAGAGCCAGATCCPSG
Euro-American Sublineage Primers	
RD115 Fwd	TTCGGGGGCCAGGTCGTTGAT
RD115 Rev	CCGCGGGAATGCCTGCTGTTAT
RD122 Fwd	GCCGGCACCGCTAATCGCTACTT
RD122 Rev	TTCTCTTGGGCATGATCATCCTTTCGTTA
RD174 Fwd	GGTGCCCTCCCGCAGAACTGTG
RD174 Rev	AGCGCGATCGCAGCGGTGAA
RD-Int 174 Fwd	GCTCGGCGACACAGGGTTGGT
RD-Int 174 Rev	GCCGCCGGTGGTGTCTGTTTC
RD182 Fwd	TGTTATACGCCCTGTCGGCGGTCACCAT

RD182 Rev	GCGCATCCCGCCGGCGTTGGTT
RD-Int 182 Fwd	GATGACTACCCCACCCGACA
RD-Int 182 Rev	AGCACCACACCGAGCACCT
RD183 Fwd	CCGAACCGCCCGCATCAAG
RD183 Rev	ACGACGGCCGAAACCACAGGAA
RD193 Fwd	CCGACTACGCCTGGCGCTAAACC
RD193 Rev	GTAGGGGCCACCCGGATTGTCAC
RD219 Fwd	GCGTCGGTGCCCAGTTGC
RD219 Rev	GCGACGGCGGTTTTGATGC
RD-Int 219 Fwd	GGATGCGACTCATCGTGGAC
RD-Int 219 Rev	GAGCAGGATCTTGGCGTCCT
RD724 Fwd	CCATGCGATTTGACTTCCGATTGA
RD724 Rev	ATATACCGTGCCGCGACTTGCTCT
RD726 Fwd	GCCCGGGCGGATGCTGTT
RD726 Rev	CGGCGGCGCGTTTGTCA
RD761 Fwd	GCCGGCGTGCTCAATGCTCAG
RD761 Rev	CCTAGGCCGGCGACGAAGTGC
Ag85C103F	CTGGCCGTTGCTCTGCGAGGGCTGATACCA
Ag85C103R	CGATCTCGCGCTGCGGCCACGACATT
LAMF	TAGCCCACCACCACAGCTTC
LAMR	ACCACCCTGCCTAACCAATTC
XhoI	TTCAACCATCGCCGCCTCTAC
SNP Confirmation Primers	
Rv0071 Fwd	ATCACGGTGTCGGTGGACCC
Rv0071 Rev	ATCAGGTCCGCGTGGTCGAT
Rv0405 Fwd	TTCGTCTGATGCACTGCCGC
Rv0405 Rev	GCAGTTCGCCACCGGTTGAT
Rv1192 Fwd	TCGTATCCCGGCTTCCCGAC
Rv1192 Rev	TACCCGAGGCGACGCAAGAT
Rv1821 Fwd	GTGCAGGCCCTGATCAACCG
Rv1821 Rev	CGCCCGGCCATTTGAGTTGA
Rv2224c Fwd	TGGCGTTGGTTCTTGTGGGC
Rv2224c Rev	TCCAGGTCCTTGGCGACGTT
Rv2546 Fwd	TGTACTCGGCGAACTCCGCT
Rv2546 Rev	CTCCGTAGGTTGGCCGGTGA
Rv3119 Fwd	GCCTACCCTTACTGTCGGCTCA
Rv3119 Rev	CCCATTCAGCACCGGTCGAG
Rv3616c Fwd	CGTTTCCGGGTGATGGCTGG
Rv3616c Rev	ACCCCGTGAGCTTGTCCCAA
Rv3769 Fwd	AAGGAGCTTGGAGCACGGGT
Rv3769 Rev	CGAGCTGGCCGACTTTGGTC
Rv3881c Fwd	ATGACGCAGTCGCAGACCGT
Rv3881c Rev	CCTTGGTCGCCCGTTTCGAG
Rangipo Diagnostic Primers	
Rv0405 RFLP Fwd	GGCGGCAGCTGCCGTTTTGTGGAGT

Rv0405 RFLP Rev	TCTGGCCACTCGAATTACTTCACGG
Rv0071 Fwd	see SNP Confirmation Primers for sequence
Rv0071 RFLP Rev	AACCCGCCCGGTTTGGGGATCTTGCGGGCC
Rv0719 RFLP Fwd	ACCCTGGTGTCCAACCTGGTC
Rv0719 RFLP Rev	TACCCCAGCGCAAACTCCAGA
Rv1821 RFLP Fwd	TGCGGAATGACGGGAACCGC
Rv1821 RFLP Rev	GACCGCGCAGCTGGTTGTCT
Rv1854c RFLP Fwd	CTGCAGCGGTCGGGTTGACTAC
Rv1854c RFLP Rev	GTAGGTGTGACCGAGCAATTCCG
Rv2224c Fwd	see SNP Confirmation Primers for sequence
Rv2224c RFLP Rev	CTTGCCGCATAACGCGCCCCCG
Rv2504c RFLP Fwd	GACCGACTTCGCACTGGTGCAT
Rv2504c RFLP Rev	CACCGTCTCCCTCTCGATCTTCT
Rv2546 RFLP Fwd	TGTTCTGCGTCGACACCAGCG
Rv2546 RFLP Rev	CTTGGCATGCGCGCGTAAAGAATT
Rv2716 RFLP Fwd	CGGACACCCGACCGTGGGAGCCACC
Rv2716 RFLP Rev	CGGAAAGTCGGCGGGGTCGGCGG
Rv3119 Fwd	see SNP Confirmation Primers for sequence
Rv3119 Rev	see SNP Confirmation Primers for sequence
Rv3513c RFLP Fwd	TGGCGGCATCTTTAAGTGAGAAC
Rv3513c RFLP Rev	CCTTCAAAGCGGCCTCGACC
Rv3616c RFLP Fwd	ACCTACATCCCGGTCGTCGGGCTCG
Rv3616c Rev	see SNP Confirmation Primers for sequence
Rv3769 Fwd	see SNP Confirmation Primers for sequence
Rv3769 RFLP Rev	TTTGGTCGTCACGCGGTCCAGTCGCACGG
Rv3867 RFLP Fwd	CTCACCGAACGCGACGAGGAGCC
Rv3867 RFLP Rev	CTGCGCTTCGGACATGCTGGACG
Rv3881c RFLP Fwd	ATGACGCAGTCGCAGACC
Rv3881c RFLP Rev	TGGTCGCCCGTTTCGAGCTT
Chapter Four:	
16S Fwd	CAGCTCGTGTCGTGAGATGT
16S Rev	AGACCGGCTTTGAAAGGATT
Msmeg_0930 Fwd	TCGACACCACACTCAGATCG
Msmeg_0930 Fwd	GCGTTGACCGATTCGTCAT
ftsZ Fwd	ACTACCTCGCGGTCATCAAG
ftsZ Rev	AGCTCCTCGATGTCGTCCT
guanylate kinase Fwd	GGCGTGGACTACACCTTTGT
guanylate kinase Rev	GGTCGACCTCGATCAGTACC
metC Fwd	ACGTGTCGTCGGTGAACTA
metC Rev	ATGTTGGCGACGTGACTGT
oxcA Fwd	GGTTCAGCGGTATGGAACTC
oxcA Rev	CCGAAGGCCTTGATCATGTA
rpoB Fwd	GTCGACGAGTGCAAAGACAA
rpoB Rev	TGGTCTCGTCGAAGTACACG
rpsB Fwd	CAACGGCATCTACATCATCG

rpsB Rev	ACGGTGGAGAAGTTGGTGAG
sigA Fwd	CCAAGGGCTACAAGTTCTCG
sigA Rev	TGGATCTCCAGCACCTTCTC
sigH Fwd	ACCTCAAGGCGTGGCTCTAC
sigH Rev	TGCCAGTCGGTGATCTCGTC
vapB Fwd	TGGCTCTAAGCATCAAACACC
vapB Rev	TTCCTCACGAAGCGGAAC
vapC Fwd	CTGACGATCCTGTTCGGACT
vapC Rev	AGGAAAAACAGTCGCCGTAA
Chapter Five:	
Cloning Primers	
Rv0064A Fwd	GTTACCATGGCTACCATTCAAGTTCGGGA
Rv0065 Rev	GGCAAGCTTCCGAACGAGTTTGATTTCG
Rv0229c Fwd	TAGTCCATGGTGAGACAGCCGCGCCGG
Rv0229c Rev	TATAAGCTTGTCGATCGTGCCCGCTGGC
Rv0239 Fwd	ATGTCCATGGCTCGAACGCAAGTCCAG
Rv0240c Rev	ATTAAGCTTGCCATCCGACGTTATCGGGT
Rv0277A Fwd	AATGCCATGGCCGAAGACGCCCTTCG
Rv0277c Rev	CAATAAGCTTGATTGGCGGGCGAATATGGCG
Rv0300 Fwd	TAATCCATGGCTGATGTACTGATTCGGGA
Rv0301 Rev	TATAAGCTTAGCGGAAGGCGGGCG
Rv0550 Fwd- GeneArt	ATCTATGCATATGTTGTTGAGCCGGCGCA
Rv0549 Rev- GeneArt	CATAAGCTTCAGCCGATGGCGTGAGC
Rv0581 Fwd	ATTACCATGGACAAGACGACGGTCTACC
Rv0582 Rev	ATTAAGCTTCGGAATGACGGTGAAGCGC
Rv0596c Fwd	CATTCCATGGCTGCTACGATACCCGCTC
Rv0595 Rev	TATAAGCTTGATCGTTATGACCTCGACGTCGG
Rv0599 Fwd	AGAACCATGGCGGCTGTTGTGGATGC
Rv0598c Rev	TATTAAAGCTTCGCGGCGACAACGACGTG
Rv0608 Fwd	ATTACCATGGCGTTGAACATCAAAGAT
Rv0609 Rev	TATAAGCTTCCGCCGATCCAGTGC
Rv0616 Fwd	TAATCCATGGTGCGCACTACCATCGA
Rv0617 Rev	TTAAAGCTTGGTGGTCGTTGGAATGAG
Rv0626 Fwd	TACTCCATGGCTGAGGTGGCCTCG
Rv0627 Rev	AGAAAGCTTGACCCGAATGATCTCCACA
Rv0657 Fwd	ATACCCATGGCGGTGACCCAGATCGA
Rv0656c Rev	ATCAAGCTTGTCGTCGGCGCTGAC
Rv0662 Fwd	CCAACCATGGCCCTCCCTAATACGCGTGC
Rv0661c Rev	TATAAGCTTGCGGTCGCCGAGCAGGCTGA
Rv0664 Fwd	TCATCCATGGAAAAGTCACGGTGCCAC
Rv0665 Rev	TATAAGCTTCACCTTGATCTCCGCGGTTGC
Rv0748 Fwd	AATACCATGGCCACCACGGTGTCAATCT
Rv0749 Rev	AATAAGCTTGAGCGGCGGGCGAATGTGGCG
Rv0959A Fwd	TCACCATGGCGACGCTGTATCTGCGC
Rv0960 Rev	TCAAAGCTTCCTGGGCACAACGGTAATCGGA

Rv1113 Fwd	ATCACCATGGCGACGACGGTGACCGTT
Rv1114 Rev	TTAAAGCTTGGACACTTCCTCAGCCAGCGCAA
Rv1241 Fwd	TTCACCATGGCCACCACCTTGACGCT
Rv1242 Rev	TACAAGCTTCGCGCAACGGGTCGGTCC
Rv1398c Fwd	AACTCCATGGCGCGGACCAACATCTAC
Rv1397c Rev	CATAAGCTTTGCCCGCGGTCGGTGCCC
Rv1560 Fwd	CATTCCATGGTGTATCGTTGGTGTATGTCAC
Rv1561 Rev	TTAAAGCTTCCGAAACGATGCGGCCTGC
Rv1720	GeneArt
Rv1721	GeneArt
Rv1740 Fwd	CTACCATGGAATTGGCGGCTCGAATGGG
Rv1741 Rev	CATAAGCTTGCCGGCGAGAACCGCGCGAAT
Rv1839c Fwd	AATCCCATGGCTAAGCGCCTGCAGGTTC
Rv1838 Rev	TATAAGCTTGGCCAGACGCTTGATCCCCGGAT
Rv1952 Fwd	TTAACCAATGGTGATCCGCAACCTTCCC
Rv1953 Rev	CCGAAGCTTAAACCAGCTTATACCAGAATATAG
Rv1962c	GeneArt
Rv1962A	GeneArt
Rv1982A Fwd	AATGCCATGGCGCTGAATATCAAAGACC
Rv1982c Rev	TATCGGATCCGCGACGCCTGGCCAGT
Rv2009 Fwd	TTAGCCATGGTGTATAGTGGTGTTGTGT
Rv2010 Rev	TTAAAGCTTGAACAACGGCTCGGTGC
Rv2104 Fwd	ATTGCCATGGCGACGACTGTCACATTGG
Rv2103 Rev	TAATGGATCCGCCAACAGCGCGGG
Rv2493 Fwd	TAAGCCATGGTGCGAACGACACTTGACC
Rv2494 Rev	CATAAGCTTGAGGATGGTCAGCAGCTCC
Rv2526 Fwd	CTTACCATGGCCGTAAAGAGGACCACGA
Rv2527 Rev	TATAAGCTTGCGCCCGCGAGAGAGTT
Rv2530A Fwd	ATTACCATGGCCACCACGTTGCAGATTGAT
Rv2530c Rev	TATAAGCTTCAACACCTCGACGAGGCC
Rv2545 Fwd - GeneArt	ATTGCCATGGTGAGCACCACCATCGTTG
Rv2546 Rev - GeneArt	TATAAGCTTAAGGGTCCCGCGCGGCA
Rv2547 Fwd	ATTGCCATGGCTGGTGATGTCATGCGGA
Rv2548 Rev	ATTGCATATGGTACGGCGGCTGCAGATC
Rv2549	GeneArt
Rv2550	GeneArt
Rv2595 Fwd	TATGCCATGGCAACAACCATCGATGTCG
Rv2596 Rev	CATAAGCTTGGTCACCAGCTCAACCTCG
Rv2601 Fwd	ATTACCATGGCGACCACGCTCGACCTGC
Rv2602 Rev	TATAAGCTTCTTGGCCAGGAGCCGCAGC
Rv2758c Fwd	GeneArt + CTAGCCATGGCTCGCGGATATGCATTAG
Rv2757 Rev	GeneArt + TATTGGATCCGGCGCTGTCCCTGGG
Rv2760 Fwd	CCAGCCATGGCCCTCAATATCAAGAGCC
Rv2759 Rev	ATAAAGCTTGGTGCACGCGGGCCGGATA
Rv2830 Fwd	TAATCCATGGCCGCTACGGAGGTGAAGG

Rv2829c Rev	TATAAGCTTCCAGACGGTGACCGGTCGT
Rv2862A Fwd	AATGCCATGGCCCTCAGTAACTGGCTGC
Rv2863 Rev	AATAAGCTTTGATCGGAATGCGCTGGCC
Rv2871 Fwd	TATGCCATGGTGCGCACGACGATCCGTA
Rv2872 Rev	CTCGAAGCTTTAGATGGGTCTGACCGTCCAAC
Rv3321 Fwd	ATTACCATGGCCACCACGTTGTCCATCG
Rv3320 Rev	TATTGGATCCAGCGTTGCCAGGTGCTGT
Rv3385c Fwd	TTATCCATGGCGCCGACCGCTTGTGC
Rv3384c Rev	ATTAAGCTTCGACGGGGCGATCACGGCC
Rv3407 Fwd	TTAGCCATGGCTGCTACCGTTGGGCTTG
Rv3408 Rev	TATAAGCTTCCGGACTGCGCCGGGTGAG
Rv3697A Fwd	ATTGCCATGGCCACTACGATCGACCTCG
Rv3697c Rev	TATAAGCTTGCCGGAGAAGGGGTCGCGA
T7 Fwd	TAATACGACTCACTATAGGG
T7 Rev	TAGTTATTGCTCAGCGGTGG
MALDI-TOF MS Oligonucleotide Calibration Standard Oligos	
Oligo 1 (12mer)	ACGTACGTACGT
Oligo 2 (20mer)	ACGTACGTACGTACGTACGT
Oligo 3 (30mer)	ACGTACGTACGTACGTACGTACGTACGTAC
932 RNA Oligonucleotide Sequences	
Oligo 1	GGAAUUCUGCAGAUCUCCAACCUCACACCA
Oligo 2	CCUCACACCACAUUACACUGGGGUGAUAUA
Oligo 3	GGGGUGAUAUAAUAUGAUGCCGGGCGGGGGCCG
Oligo 4	GGGCCGUAACGUAGCACUACAUGACCCGAUACGCUA
Oligo 5	CCGAUACGCUAGGUGGGAUUAGGCAUCACAC
Oligo 6	AGGCAUCACACUGGCGGCCGCUCGAGCAUGC
Oligo 7	UAGAGGGCCCUAUUCUAUAGUGUCACCUAAA
Oligo 8	CGAGCAUGCAUCUAGCUAAAUGCUAGAGCUCG

[PS] Indicates phosphorothioate modification

Table A.2: Orthogonal array used for the optimisation of diplex PCRs

Reaction	T (°C)	Taq (U)	$MgCl_2$ (mM)	dNTP (mM)	P1 (µM)	P2 (µM)
1	64	2.5	1.5	0.1	0.1	0.1
2	64	2.5	2	0.15	0.2	0.2
3	64	2.5	2.5	0.2	0.3	0.3
4	64	3	1.5	0.1	0.2	0.2
5	64	3	2	0.15	0.3	0.3
6	64	3	2.5	0.2	0.1	0.1
7	66	3.5	1.5	0.15	0.1	0.3
8	66	3.5	2	0.2	0.2	0.1
9	66	3.5	2.5	0.1	0.3	0.2
10	66	2.5	1.5	0.15	0.3	0.2
11	66	2.5	2	0.2	0.1	0.3
12	66	2.5	2.5	0.1	0.2	0.1
13	68	3	1.5	0.1	0.3	0.1
14	68	3	2	0.2	0.1	0.2
15	68	3	2.5	0.15	0.2	0.3
16	68	3.5	1.5	0.1	0.1	0.3
17	68	3.5	2	0.2	0.3	0.1
18	68	3.5	2.5	0.15	0.2	0.2
19 (-'ve)	66	3	2	0.15	0.2	0.2

The orthogonal array used for the optimisation of diplex PCRs based on modified Taguchi methods. The different shaded cells in each column represent 3 different concentrations chosen for each variable and the 19 rows represent 19 different reactions. T = annealing temperature; P1 = primer Rv3119; P2 = primer Rv3769.

Table A.3: Orthogonal array used for the optimisation of triplex PCRs

Reaction	T (°C)	Taq (U)	$MgCl_2$ (mM)	dNTP (mM)	P1 (µM)	P2 (µM)	P3 (µM)
1	62	2.5	1.5	0.1	0.1	0.1	0.1
2	62	2.5	2.4	0.15	0.2	0.2	0.2
3	62	2.5	3.3	0.2	0.3	0.3	0.3
4	62	3	1.5	0.1	0.2	0.2	0.3
5	62	3	2.4	0.15	0.3	0.3	0.1
6	62	3	3.3	0.2	0.1	0.1	0.2
7	64.4	3.5	1.5	0.15	0.1	0.3	0.2
8	64.4	3.5	2.4	0.2	0.2	0.1	0.3
9	64.4	3.5	3.3	0.1	0.3	0.2	0.1
10	64.4	2.5	1.5	0.2	0.3	0.2	0.2
11	64.4	2.5	2.4	0.1	0.1	0.3	0.3
12	64.4	2.5	3.3	0.15	0.2	0.1	0.1
13	68	3	1.5	0.15	0.3	0.1	0.3
14	68	3	2.4	0.2	0.1	0.2	0.1
15	68	3	3.3	0.1	0.2	0.3	0.2
16	68	3.5	1.5	0.2	0.2	0.3	0.1
17	68	3.5	2.4	0.1	0.3	0.1	0.2
18	68	3.5	3.3	0.15	0.1	0.2	0.3
19 (-'ve)	64.4	3	2.4	0.15	0.2	0.2	0.2

The orthogonal array used for the optimisation of triplex PCRs based on modified Taguchi methods. The different shaded cells in each column represent 3 different concentrations chosen for each variable and the 19 rows represent 19 different reactions. T = annealing temperature; P1 = Rv3119 primers; P2 = Rv2504c primers; P3 = Rv1821 primers.

A.2 Plasmids & Bacterial Strains Used in This Study

Table A.4: Plasmids & bacterial strains used in Chapters Three, Four and Five

Bacterial Strain	Description
Escherichia coli	
DH10B (TOP10)	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*dlac*ZΔM15 Δ*lac*X74 deoR *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L *end*A1 *nup*G λ-
Mycobacterium smegmatis	
mc^26	*M. smegmatis* wild-type strain
mc^2155	Electrocompetent lab strain of *M. smegmatis*
mc^24517	*M. smegmatis* expression strain with T7 RNA polymerase; Km^r
Plasmids	
pYUB1049	*E. coli* mycobacterium shuttle vector with T7 promoter encoding both C- and N-terminal His-tags; Hyg^r
pYUB28b	*E. coli* mycobacterium shuttle vector with T7 promoter, MCS pET28b, encoding both C- and N-terminal His-tags; Hyg^r
pMind	*M. smegmatis* tetracycline inducible expression vector; Km^r Hyg^r

Hyg^r = hygromycin B resistance, Km^r = kanamycin resistance

Table A.5: Strains screened in the Rangipo-specific SNP identification process and their respective MTBC lineage

First round of screening (TB Database)	
Strain	**MTBC Lineage**
M. tuberculosis T83	Lineage 1*
M. tuberculosis CDC1551	Lineage 4*
M. tuberculosis SG1	Lineage 3*
M. tuberculosis T67	Lineage 2
M. tuberculosis T85	Lineage 2
M. tuberculosis 00_1695	Lineage 2
M. tuberculosis 98_1833	Lineage 2
M. tuberculosis M4100A	Lineage 2
M. tuberculosis 91_0079	Lineage 3
M. tuberculosis K49	Lineage 3
M. tuberculosis GM_1503	Lineage 4
M. tuberculosis 4783_04	Lineage 4
M. tuberculosis K37	Lineage 4
M. tuberculosis T17	Lineage 1
M. tuberculosis T92	Lineage 1

M. tuberculosis 95_0545	Lineage 1
M. tuberculosis K21	Lineage 1
M. tuberculosis K67	Lineage 1
M. tuberculosis K93	Lineage 1
M. africanum 11821_03	Lineage 5
M. africanum 5444_04	Lineage 5
M. africanum 4141_04	Lineage 6
M. africanum GM_0981	Lineage 6
M. canetti K116	Outgroup
M. tuberculosis H37Rv	Lineage 4
M. tuberculosis F11	Lineage 4
Additional strains screened in second round of screening (Broad Institute)	
Strain	**MTBC Lineage**
M. tuberculosis Haarlem	Lineage 4
M. tuberculosis KZN 4207 (DS)	Lineage 4**
M. tuberculosis KZN 1435 (MDR)	Lineage 4**
M. tuberculosis KZN 605 (XDR)	Lineage 4**
M. tuberculosis C	Lineage 4
M. tuberculosis 98-R604 INH-RIF-EM	N/A
M. tuberculosis W-148	Lineage 2 ^
M. tuberculosis 02_1987	Lineage 2
M. tuberculosis EAS054	Lineage 1
M. tuberculosis 94_M4241A	Lineage 2
M. tuberculosis T46	Lineage 1
M. africanum CPHL_A	Lineage 5
M. africanum K85	Lineage 6

* (Hershberg *et al.* 2008), ** (Feuerriegel *et al.* 2010), ^ This strain is a variant of the Beijing family which belongs to the East-Asian Lineage, therefore, it belongs to MTBC Lineage 2

A.3 Buffers, Media and Solutions

All concentrations in % are percentages in weight per volume (w/v) unless otherwise indicated.

Table A.6: General buffers and solutions used in this study

Reagent	Description
CTAB/NaCl solution	10 % CTAB, 0.7 M NaCl
5 x DNA Loading Dye	0.05 % bromophenol blue, 0.25 % xylene cyanol, 30 % (*v/v*) glycerol
Equilibration Buffer A	50 mM Tris-HCl pH 6.8, 6 M Urea, 30 % *(v/v)* glycerol, 2 % SDS, 0.25 % DTT
Equilibration Buffer B	50 mM Tris-HCl pH 6.8, 6 M Urea, 30 % *(v/v)* glycerol, 2 % SDS, 2.5 % iodacetamide
Fairbanks Staining Solution A	0.05 % coomassie blue R-250, 25 % (*v/v*) isopropanol, 10 % (*v/v*) acetic acid
Fairbanks Staining Solution B	0.005 % coomassie blue R-250 10 % (*v/v*) isopropanol, 10 % (*v/v*) acetic acid
Fairbanks Staining Solution C	0.002 % coomassie blue, 10 % (*v/v*) acetic acid
Fairbanks Staining Solution D	10 % (*v/v*) acetic acid
GITC	295.4 g guanidine thiocynate, 2.5 g N-lauroyl sarcosine, 3.9 g tri-sodium citrate, 3.6 ml 2-meracaptoethanol, 280 ml DEPC H_2O (total volume 500 ml), pH 7.0
Native Gel Loading Buffer	0.3 M Tris-HCl, 1 % bromophenol blue, pH 6.8
10 x PBS	0.2 M phosphate (0.038 M NaH_2PO_4, 0.162 M Na_2HPO_4, 1.5 M NaCl, pH 7.4
1 x PBS	100 ml 10 x PBS + 900 ml H_2O
Resolving Buffer	1.5 M Tris-HCl, pH 8.8
4 x SDS loading buffer	200 mM Tris-HCl pH 6.8, 8 % SDS, 40 % (*v/v*) glycerol, 0.4 % bromophenol blue, 400 mM 2-meracaptoethanol
SDS running buffer	25 mM Tris-HCl pH 6.8, 0.1 % SDS, 190 mM glycine
Stacking Buffer	0.5 M Tris-HCl, pH 6.8
10 x TAE	400 mM Tris-acetate, 20 mM EDTA
1 x TAE	100 ml 10 x TAE + 900 ml H_2O
10 x TBE	0.89 M Tris-HCl, 0.89 M boric acid, 20mM 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
Tris-EDTA-SDS lysis solution	0.1 M Tris, 0.05 M EDTA, 1 % SDS, 0.5 M NaCl
Tris-Glycine Running Buffer	25 mM Tris-HCl, 250 mM glycine, pH 8.5

Table A.7: Media used in this study

Media	Description
7H9	0.47 g 7H9 powder, 0.2 % (*v/v*) glycerol in 90 ml H_2O, autoclave then add 10 ml ADC enrichment and 0.05 % (*v/v*) Tween 80
7H10-agar	1.9 g 7H10 powder, 0.5 % (*v/v*) glycerol in 90 ml H_2O, autoclave then add 10 ml ADC enrichment and 0.05 % (*v/v*) Tween 80 at 50 °C
Blood-agar	10 % sheep blood (Life Technologies), 90 % sterile Columbia Blood Agar (44 g/L Columbia Blood Agar Base)
Hartmans de Bont	Components added in order given: 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, pH 7.0. Autoclave, then add 1 x phosphate stock **
LB-agar	1 % bactotryptone, 0.5 % yeast extract, 1 % NaCl, 15 g/L agar, pH 8.0
LB	1 % bactotryptone, 0.5 % yeast extract, 1 % NaCl, pH 8.0
LBT-agar	1 % bactotryptone, 0.5 % yeast extract, 1 % NaCl, 15 g/L agar, 0.05 % (*v/v*) Tween 80, pH 8.0
LBT	1 % bactotryptone, 0.5 % yeast extract, 1 % NaCl 0.05 % (v*/v*) Tween 80, pH 8.0
Low salt LB-agar	1 % bactotryptone, 0.5 % yeast extract, 0.5 % NaCl, 15 g/L agar 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***, 100 µg/ml each of 17 amino acids (no Cys, Tyr or Met). Individual components autoclaved to sterilise
Sauton's	4 g l-asparagine, 0.5 g KH2PO4, 0.5 g MgSO4, 50 mg FeNH4 citrate, 2.4 g citric acid, and 60 ml glycerol per litre
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, 50 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***

*100 x trace metals stock for Hartmans de Bont media: EDTA [0.1 g], $MgCl_2.6H_2O$ [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.

**100 x phosphate stock for Hartmans de Bont media: K_2HPO_4 [15.5g], $NaH_2PO_4.2H_2O$ [11.1g] to 100 ml with dH_2O.

*** 1000 x metals mix made up from the sterile stock solutions of each component to give the following concentrations: 50 µM $FeCl_3$ in 0.12 M HCl (filter sterile), 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 A.8: Additives used in crystal additive screens

Additive	Final Concentration
Cobalt(II) chloride hexahydrate	10 mM
Copper(II) chloride dihydrate	10 mM
Manganese(II) chloride tetrahydrate	10 mM
Zinc chloride	10 mM
Potassium chloride	100 mM
Sodium chloride	200 mM
Potassium sodium tartrate tetrahydrate	100 mM
Sodium citrate tribasic dihydrate	100 mM
Cesium chloride	100 mM
Sodium malonate pH 7.0	100 mM
L-Proline	10 mM
Dimethyl sulfoxide	3 %
Glycine	100 mM
Sarcosine	10 mM
Guanidine hydrochloride	100 mM
Urea	10 mM
Ethylenediaminetetraacetic acid disodium salt dihydrate	10 mM
Polyethylene glycol 3,350	1 %
D-(+)-Glucose monohydrate	3 %
Sucrose	3 %
D-Sorbitol	3 %
Glycerol	3 %
Polyethylene glycol 400	5 %
Ethanol	3 %
2-Propanol	3 %
Methanol	3 %
Acetonitrile	4 %
Acetone	4 %
1-Butanol	0.7 %
2-Mercaptoethanol	10 %
Arginine	10 mM

Appendix B: Gene & Protein Information

B.1 *M. smegmatis vapB*, *vapC* & 16S Genomic Information

Amplicon gene sequences with forward and reverse primer sequences shown in red, followed by the amplicons size and melt temperature (Tm) as determined by RT-qPCR using SYTO82 dye.

MS_1283 VapB (255bp)

ATGGCTCTAAGCATCAAACACCCGGAAGCCGACCGGCTGGCCAGAGAGCTAGCGGCA
CGTACGGGGGAGACACTGACCGAGGCGGTGGTGATGGCACTGCGGGAGCGACTGGCG
CGCACTGTCGGCCGCACTCAGGTCGTTCCGCTTCGTGAGGAACTCGCCGCGATCCGT
CGTCGCTGCGCAGCCCTACCGGTGTTGGACGACCGGACAGCCGAATCGATCCTGGGC
TACGACGACCGCGGTCTGCCGTCCTGA

Size:155bp
Tm: 92.7 - 93 °C

MS_1284 VapC (390bp)

ATGGTTATCGACACTTCTGCCCTCGTTGCCATCTTGACCGACGAACCCGACGCCGAG
TTGTTGGAGGGGGCGGTGGCTGACGATCCTGTTCGGACTATGTCCACCGCGTCGTAT
CTGGAGACAGCGATCGTGATCGAAAGTCGCTTCGGCGAGCCAGGTGGTCGCGAACTC
GATCTGTGGTTGCATCGCGCATCAGTGGCACTGGTCGCTGTGGATGCCGATCAGGCC
GACGCCGCCCGGTTGGCTTACCGGAGATACGGCAAGGGGCGCCATCGTGCAGGCCTG
AATTACGGCGACTGTTTTTCCTATGCGCTCGCCAAGGTCAGCGGTCAGCCTTTGTTG
TTCAAGGGCGAGGCTTTTCGGCTCACCGACGTCGCTGCGGTCCACTGA

Size:231bp
Tm: 91.8 - 92 °C

rrsA&B 16S (1528bp)

TTTTTGTTTGGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCTTAACA
CATGCAAGTCGAACGGAAAGGCCCTTTCGGGGGTACTCGAGTGGCGAACGGGTGAGT
AACACGTGGGTGATCTGCCCTGCACTTTGGGATAAGCCTGGGAAACTGGGTCTAATA
CCGAATACACCCTGCTGGTCGCATGGCCTGGTAGGGGAAAGCTTTTGCGGTGTGGGA
TGGGCCCGCGGCCTATCAGCTTGTTGGTGGGGTGATGGCCTACCAAGGCGACGACGG
GTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGATACGGCCCAGACTCC
TACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACG
CCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCACAGACGAAGCGC
AAGTGACGGTATGTGCAGAAGAAGGACCGGCCAACTACGTGCCAGCAGCCGCGGTAA
TACGTAGGGTCCGAGCGTTGTCCGGAATTACTGGGCGTAAAGAGCTCGTAGGTGGTT
TGTCGCGTTGTTCGTGAAAACTCACAGCTTAACTGTGGGCGTGCGGGCGATACGGGC

AGACTAGAGTACTGCAGGGGAGACTGGAATTCCTGGTGTAGCGGTGGAATGCGCAGA
TATCAGGAGGAACACCGGTGGCGAAGGCGGGTCTCTGGGCAGTAACTGACGCTGAGG
AGCGAAAGCGTGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACG
GTGGGTACTAGGTGTGGGTTTCCTTCCTTGGGATCCGTGCCGTAGCTAACGCATTAA
GTACCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGC
CCGCACAAGCGGCGGAGCATGTGGATTAATTCGATGCAACGCGAAGAACCTTACCTG
GGTTTGACATGCACAGGACGCCGGCAGAGATGTCGGTTCCCTTGTGGCCTGTGTGCA
GGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAAC
GAGCGCAACCCTTGTCTCATGTTGCCAGCACGTTATGGTGGGGACTCGTGAGAGACT
GCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGCCCCTTATGT
CCAGGGCTTCACACATGCTACAATGGCCGGTACAAAGGGCTGCGATGCCGTGAGGTG
GAGCGAATCCTTTCAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAACTCGACCCCG
TGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAACGCTGCGGTGAATACGTTCCCG
GGCCTTGTACACACCGCCCGTCACGTCATGAAAGTCGGTAACACCCGAAGCCGGTGG
CCTAACCCTTGTGGAGGGAGCCGTCGAAGGTGGGATCGGCGATTGGGACGAAGTCGT
AACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTTCT

Size:234bp
Tm: 89.7 °C

B.2 Housekeeping Gene Information

Table B.1: HK genes tested, and the reason for their rejection as housekeepers.

HK Gene	Reason for Rejection
Msmeg_0930	poor expression
ftsZ	variable expression
guanylate kinase	poor expression
metC	poor expression
oxcA	poor & variable expression
rpoB	variable expression
rpsB	variable expression & multiple products
sigA	variable expression
sigH	variable expression

B.3 Pentaprobe Sequences + Flanking Vector Sequence

Sequences represent the RNA sequence that will be transcribed from the T7 promoter (excludes the T7 promoter). Yellow highlighting corresponds to the pentaprobe sequence.

922 + Flanking Sequences
AGACCCAAGCTTGGTACCGGAATTCTACGAATTTTTCTTTTGTTTATTTCCTTTCGCTTTGC
TTCTCTTCCCTTCGGTTCTGTTCCGTTTTACCTTGTCTTGCCTTATCTTACTTTATCTAGAG
GGCCCTATTCTATAGTGTCACCTAAATGCTAGAGCTCGCT

923 + Flanking Sequences
AGACCCAAGCTTGGTACTATCTTACTTTAGTTTCATTTAATTGTGTTGTACTCTCCTCTGCG
TTCACTTAGCTTAACTTGGTTTGGCTTGATTTGACTTCAGTTGCGCTCTATTCTATCTAGAG
GGCCCTATTCTATAGTGTCACCTAAATGCTAGAGCTCGCT

924 + Flanking Sequences
AGACCCAAGCTTGGTACCGCTCTATTCTACTGTCCTGTGCATTCAATCGTTGAGTTCGATCT
AGTCTCGTCTAACCCTCCCCTGCTCCGCTGGTCTGGCCTCGCCTATCCTACCCATTCTAGAG
GGCCCTATTCTATAGTGTCACCTAAATGCTAGAGCTCGCT

925 + Flanking Sequences
AGACCCAAGCTTGGTACTATCCTACCCATTGGGCTCATCTGATCCATCCGGTCCCGTCCACT
CGGCTATGTTATGCTGTATTGCAGTCGTGTCGCGTCGAGCTGCCCTAATCCCACCTCTAGAG
GGCCCTATTCTATAGTGTCACCTAAATGCTAGAGCTCGCT

926 + Flanking Sequences
AGACCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCT
GCAGATCCTAATCCCACCTAGCGTATCGGGTCATGTAGTGCTACGTTACGGCCCCCGCCCGG
CATCATATTATATCACCCCAGTGTAATGTGGTGTGAGGTTGGAGCATCACACTGGCGGCCGC
TCGAGCATGCATCTAGAGGGCCCTATTCTATAGTGTCACCTAAATGCTAGAGCTCGCT

927 + Flanking Sequences
AGACCCAAGCTTGGTACGTGAGGTTGGAGTCCGACCTGGAATCTCAGCCTGACGTGCCATGC
GGTGCGATGTCACGCCGCGCCACGGTATAGTATGGTACGGGATCCCGTCTAGAGGGCCCTAT
TCTATAGTGTCACCTAAATGCTAGAGCTCGCT

932 + Flanking Sequences (932 = reverse complement to 926)
AGACCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCT
GCAGATCTCCAACCTCACACCACATTACACTGGGGTGATATAATATGATGCCGGGCGGGGGC
CGTAACGTAGCACTACATGACCCGATACGCTAGGTGGGATTAGGCATCACACTGGCGGCCGC
TCGAGCATGCATCTAGAGGGCCCTATTCTATAGTGTCACCTAAATGCTAGAGCTCGCT

B.4 932 RNA Oligonucleotide Design

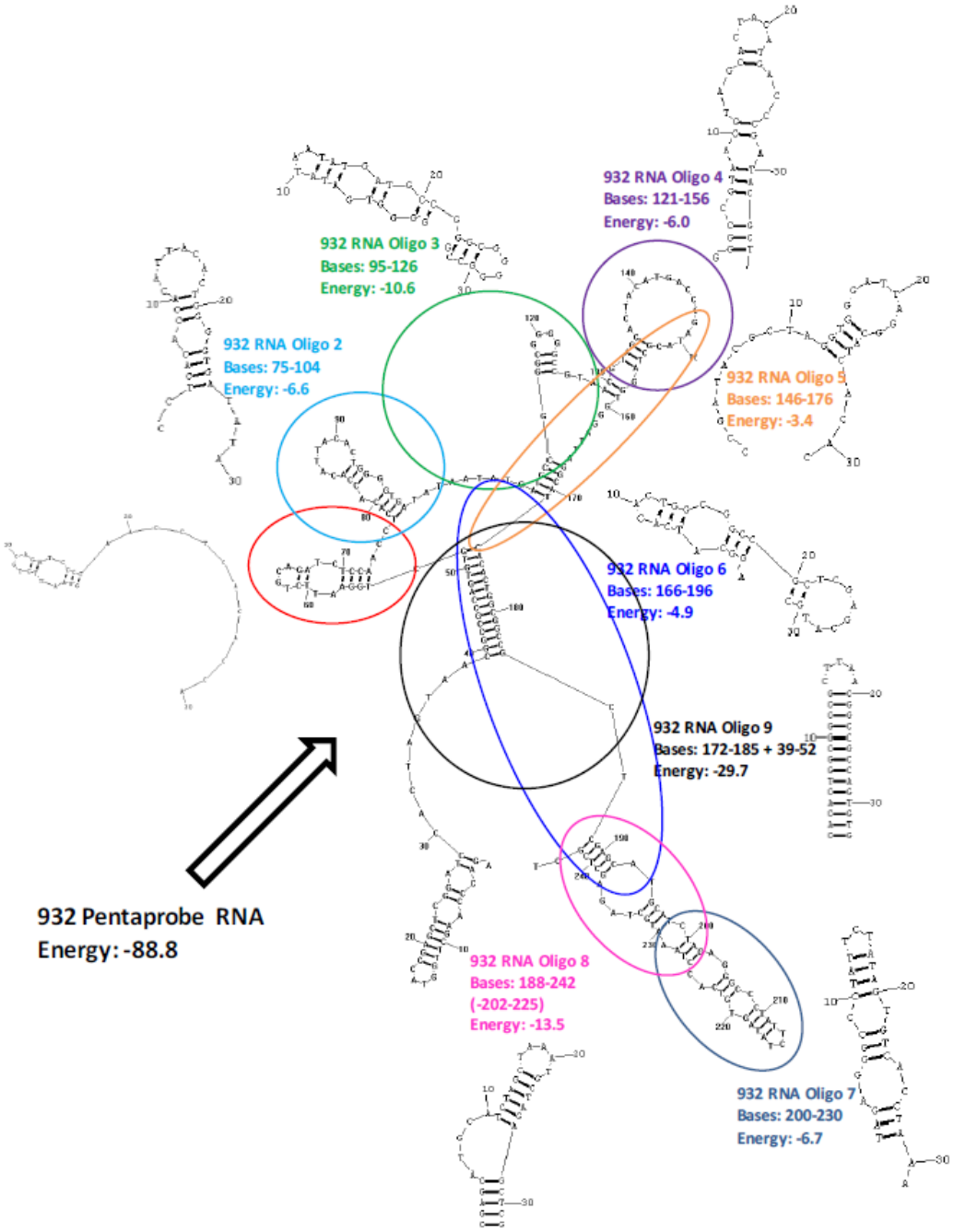


Figure B.1: Design and predicted secondary structure of 932 RNA oligonucleotides 1 - 9. Figure taken from McKenzie 2011. 'Energy' refers to ΔG values for each of the oligos.

Appendix C: Results

C.1 Robot Crystallisation Screen Results

Table C.1: Promising robot screen crystalisation conditions

Well #	Condition	Rv2530 c	Rv2530 a/c	Rv2530 a/c SEC 8 - 14	Rv2530 a/c SEC 17 - 23
C = A1	0.02 M Calcium chloride dihydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 30 % (+/-)-2-Methyl-2,4-pentanediol		✓	✓	✓
C = A12	0.2 M Magnesium chloride hexahydrate, 0.1 M HEPES sodium pH 7.5, 30 % propanol				✓
C = B2	0.2 M Calcium chloride dihydrate, 0.1 M HEPES sodium pH 7.5, 28 % Polyethylene glycol 400				✓
C = B12	0.2 M Calcium chloride dihydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 20 % 2-Propanol	✓		✓	✓
C = C1	0.1 M Imidazole pH 6.5, 1.0 M sodium acetate trihydrate				✓
C = C10	0.1 M Sodium acetate trihydrate pH 4.6, 2.0 M Sodium formate				✓
C = D8	0.2 M Magnesium formate dihydrate				✓
C = D9	0.2 M Zinc acetate dihydrate, 0.1 M sodium cacodylate trihydrate pH 6.5, 18 % PEG 8000			✓	
C = E1	2.0 M Sodium chloride, 10 % PEG 6000				P
C = E5	2.0 M Ammonium sulphate, 5 % propanol				P
C = E11	0.01 M Cobalt(II) chloride hexahydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 1.0 M 1,6-Hexanediol		✓	✓	
C = G3	0.01 M Zinc sulfate heptahydrate, 0.1 M MES monohydrate pH 6.5, 25% Polyethylene glycol monomethyl ether 550	✓	✓	✓	
C = G5	0.5 M Ammonium sulfate, 0.1 M HEPES pH 7.5, 30% (+/-)-2-Methyl-2,4-pentanediol		✓		
C = G7	0.1 M HEPES pH 7.5, 20% Jeffamine M-600		✓		
I = B2	0.1 M BIS-TRIS pH 6.5, 0.5 M Magnesium formate dihydrate				✓
I = D12	0.2 M Calcium chloride dihydrate, 0.1 M BIS-TRIS pH 5.5, 45% (+/-)-2-Methyl-2,4-pentanediol		✓		

I = E1	0.2 M Calcium chloride dihydrate, 0.1 M BIS-TRIS pH 6.5, 45 % (+/-)-2-Methyl-2,4-pentanediol	✓	✓	✓	✓
I = E2	0.2 M Ammonium acetate, 0.1 M BIS-TRIS pH 5.5, 45 % (+/-)-2-Methyl-2,4-pentanediol	✓		✓	
I = E3	0.2 M Ammonium acetate, 0.1 M BIS-TRIS pH 6.5, 45% (+/-)-2-Methyl-2,4-pentanediol		P		
I = F3	5 % Tacsimate pH 7.0, 0.1 M HEPES pH 7.0, 10 % Polyethylene glycol monomethyl ether 5,000				✓
I = H9	0.05 M Zinc acetate dihydrate, 20% Polyethylene glycol 3,350		✓		
S = C9	1.2 M DL-Malic acid pH 7.0, 0.1 M BIS-TRIS propane pH 7.0	P			
S = E11	0.5 M Succinic acid pH 7.0, 0.1 M BIS-TRIS propane pH 7.0				P
S = F8	0.8 M Lithium sulfate monohydrate, 0.1 M BIS-TRIS propane pH 7.0	✓			
S = G2	1.0 M Magnesium sulphate hydrate, 0.1 M BIS-TRIS propane pH 7.0				✓
S = G3	1.0 M Magnesium sulfate hydrate, 0.1 M Tris pH 8.5			✓	
S = G5	1.8 M Magnesium sulfate hydrate, 0.1 M BIS-TRIS propane pH 7.0			✓	✓
S = G6	1.8 M Magnesium sulfate hydrate, 0.1 M Tris				✓
S = H1	0.6 M Potassium sodium tartrate tetrahydrate, 0.1 M BIS-TRIS propane pH 7.0				P
P = A4	0.1 M Sodium acetate trihydrate pH 4.5, 30% Polyethylene glycol 300		P		
P = A8	0.1 M MES monohydrate pH 6.0, 22 % Polyethylene glycol 400				✓
P = B12	0.1 M Tris pH 8.0, 30 % PEG MME 2000				P
P = D7	0.1 M Sodium acetate trihydrate pH 4.5, 10 % PEG 10000				P
P = E3	0.05 M Calcium chloride dihydrate, 0.1 M MES monohydrate pH 6.0, 45% PEG 200	P		✓	
P = E7	0.2 M Ammonium acetate, 0.1 M Sodium citrate tribasic dihydrate pH 5.5, 24 % PEG 400				P
P = E12	0.15 M DL-Malic acid pH 7.0, 0.1 M Imidazole pH 7.0, 22% Polyethylene glycol monomethyl ether 550		✓		
P = F5	0.2 M L-Proline, 0.1 M HEPES pH 7.5, 24 % Polyethylene glycol 1,500				✓
P = F7	0.1 M Sodium chloride, 0.1 M BIS-TRIS propane pH 9.0, 25 % PEG 1500				P
P = G12	10 % 2-Propanol, 0.1 M Sodium acetate trihydrate pH 4.0, 22 % Polyethylene glycol 6,000				✓

Table indicates robot crystalisation screen conditions which either produced crystals (✓) or showed promise (P) by day 49. Letters in bold indicate which crystallisation plate the well belonged to: **C**rystal, **I**ndex, **S**alt or **P**EG.

C.2 Confirmation of the top 10 Rangipo SNPs

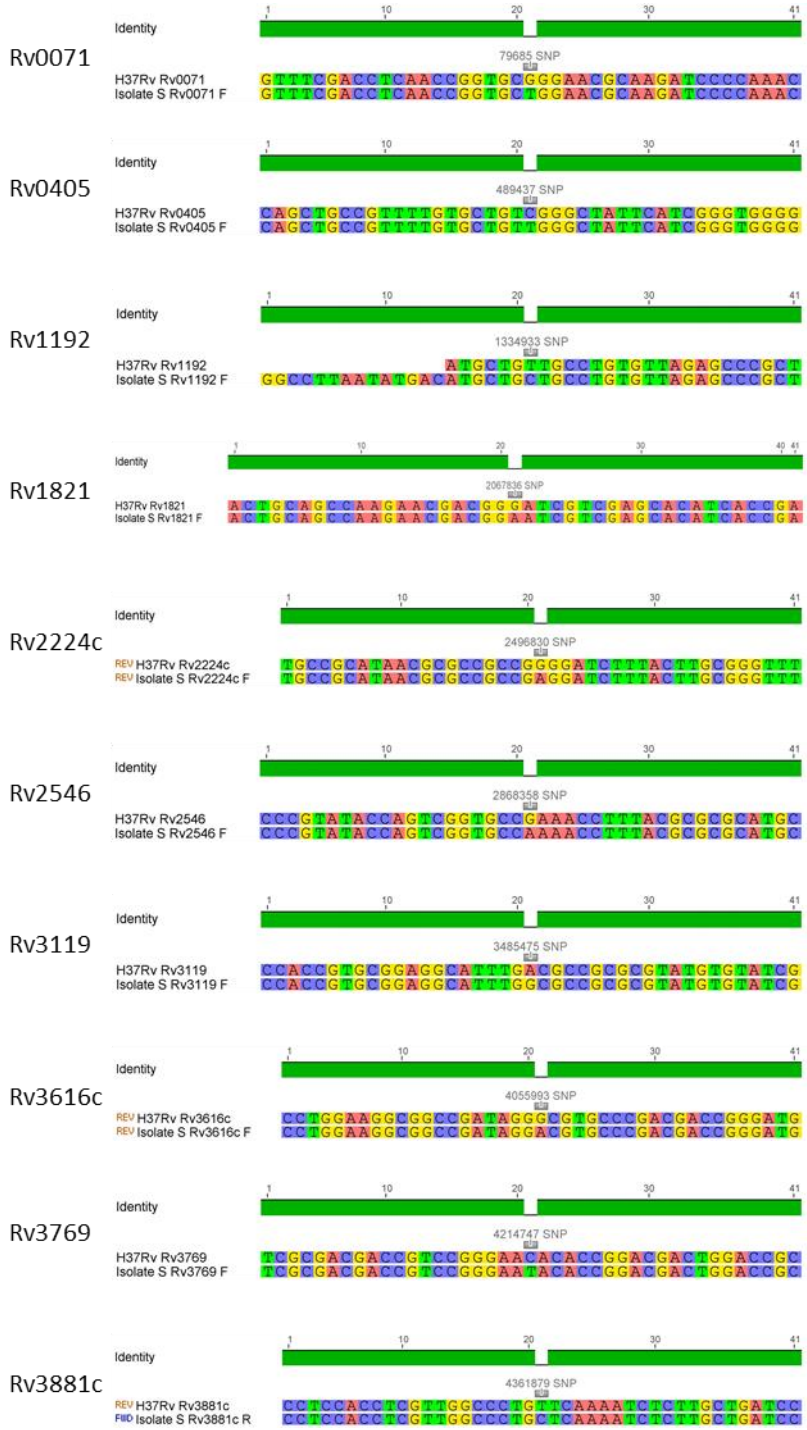


Figure C.1: Sequencing results of the top 10 SNPs PCR amplified from Rangipo isolate S, aligned against H37Rv to confirm the presence of each SNP.

.

Appendix D: Raw Data

Raw data for Chapters Three, Four and Five can be found on compact disc at the back of this thesis. Contents include:

Chapter Three - WGS identification of putative Rangipo specific SNPs

- Excel table of 247 putative Rangipo specific SNPs, their positions relative to the H37Rv genome, and findings from the database trawling

Chapter Four - *M. smegmatis* wt *vapB* and *vapC* RT-qPCR data post DNP stress (Figure 4.5)

- Rotor-Gene 6000 file: Syto 82 2009-04-16 (1) 2009-07-21 (1).rex
- LinRegPCR file: Linreg output.xls

Note, during the course of this research the nomenclature of vap genes changed from PIN to vap, so in the raw data, PINT = vapC and PINAT = vapB.

Chapter Five - TB VapBC cloning information, and MALDI-TOF MS RNA Oligos Data

- 47 TB VapBC gene sequences with cloning information
- $VapC_{Rv2530c}$ MALDI-TOF MS raw data and analysis files

Appendix E: Publications