



THE UNIVERSITY OF  
**WAIKATO**  
*Te Whare Wānanga o Waikato*

Research Commons

<http://researchcommons.waikato.ac.nz/>

## Research Commons at the University of Waikato

### Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

# The Biochemistry of VapBC Toxin-Antitoxins

A thesis submitted in fulfilment  
of the requirements for the degree  
of  
**Doctor of Philosophy in Biological Sciences**  
at  
**The University of Waikato**  
by  
**Joanna Leigh McKenzie**

---

The University of Waikato  
2011



THE UNIVERSITY OF  
**WAIKATO**  
*Te Whare Wānanga o Waikato*

---

## Abstract

The VapBC toxin-antitoxin (TA) systems were first identified in 2005 and little is known about their contemporary biological function, despite the fact that VapBC TAs are the largest TA family and are widespread in bacteria and archaea (Arcus, Rainey, & Turner, 2005; Gerdes, Christensen, & Lobner-Olsen, 2005). *Mycobacterium tuberculosis* has a surprisingly large repertoire of 45 VapBC TAs. In contrast *Mycobacterium smegmatis*, a model organism for *M. tuberculosis*, contains only one *vapBC* operon, thereby making it an ideal system to uncover the possible role(s) that VapBC proteins play in mycobacteria. This thesis describes the functional characterisation of VapC from *M. smegmatis* and two homologues from *Pyrobaculum aerophilum*, along with biophysical characterisation of the VapBC complex from *M. smegmatis*. The VapBC proteins from *M. smegmatis* form a tight complex in a 1:1 ratio and interactions between the proteins result in a tetramer of VapBC heterodimers. VapB is susceptible to proteolytic degradation when not bound to DNA thus hinting at a mechanism for VapC activation.

VapC proteins from *P. aerophilum* (VapC<sub>P<sub>AE</sub>2754</sub> and VapC<sub>P<sub>AE</sub>0151</sub>) and VapC from *M. smegmatis* display Mg<sup>2+</sup>/Mn<sup>2+</sup> dependent, sequence-specific ribonuclease activity. VapC from *M. smegmatis* targets the AU rich sequences AUAU and AUAA. Whereas VapC<sub>P<sub>AE</sub>2754</sub> and VapC<sub>P<sub>AE</sub>0151</sub> from *P. aerophilum* both target GGUG and GGGG sequences. These sequences are present in over half the mRNA transcripts encoded in the *P. aerophilum* genome, making them potent toxins. VapC ribonuclease activity is inhibited when VapC is bound to VapB. When VapC is released from the VapBC complex it cleaves cohorts of mRNA transcripts thereby reducing protein synthesis for this cohort of genes. Microarray analysis revealed that the majority of transcripts downregulated in response to VapC expression in *M. smegmatis* are involved in carbon utilisation and transport (Robson, 2010). Bioinformatics shows that the target sequence is overrepresented in the downregulated transcripts. The majority of the downregulated genes are in operons, which suggests a mechanism for destabilising mRNA transcripts to regulate specific metabolic processes.

---

---

## Acknowledgements

First and foremost I would like to thank my supervisor, Associate Professor Vic Arcus. Thank you for the opportunities you have given me throughout my PhD and your guidance and patience for my never ending questions.

Secondly I would like to thank Dr Ray Cursons, Dr Dick Wilkins, Dr Greg Cook and Dr Jennifer Robson. Dick and Ray, thank you for all your help trouble-shooting experiments, Greg, for helpful and enthusiastic discussions on toxin-antitoxins and Jen, for being my other half on this project, it has been great working alongside you. Thanks to Professor Kenn Gerdes (Newcastle University, UK) for giving me the opportunity to work in his lab and thanks to Kristoffer Winther and others in the Gerdes and Errington lab for the help given in my time there.

To everyone, past and present in the Proteins and Microbes lab at Waikato University I could not have done this with you: Dr Jo H, Dr Svend H, Dr Emma D, Emma S, Emma L, Marisa, Mat, Dr Judith B, Dr Ian M, Line, Ashley, Ali and John. You have all made the last few years fly by and it has been fantastic to work with such a great group. Jo H and Ali, thanks for critical reading of this thesis, Judith for keeping my sugar levels up with your AMAZING baking and to all the other PhD students in the lab, thanks for reminding me I was not alone on the rollercoaster that is completing a PhD.

Thank you to the University of Waikato for my doctoral scholarship, the Health Research Council for funding and to the New Zealand Federation of Graduate Women and the New Zealand Post Graduate Study Abroad Award Fund for enabling me to visit Professor Gerdes lab at the University of Newcastle.

And of course thank you to my family and friends for your love and support. To Mum, Barry and Dad thank you so much for always believing in me and telling me I could be anything I wanted to, I bet you never thought that a scientist it would be! Finally Michael, thanks for your unwavering support, I could not have done this without you.

---



---

# Table of Contents

<b>Abstract</b> .....	<b>ii</b>
<b>Acknowledgements</b> .....	<b>iii</b>
<b>Table of Contents</b> .....	<b>iv</b>
<b>List of Figures</b> .....	<b>xiii</b>
<b>List of Tables</b> .....	<b>xvii</b>
<b>List of Abbreviations</b> .....	<b>xviii</b>
<b>Chapter One: Antibiotic Resistance, Persistence and the Burden of Infectious Disease</b> .....	<b>1</b>
1.1    Toxin-Antitoxin Systems .....	2
1.1.1    Discovery and Function of Toxin-Antitoxin Systems .....	2
1.1.2    Hypothetical Functions of Toxin-Antitoxin Systems .....	8
1.1.3    The Function of Toxin-Antitoxin Systems in <i>Escherichia coli</i> .....	10
1.1.4    The Role of Toxin-Antitoxin Systems in the Stress Response .....	10
1.1.5    The Role of Toxin-Antitoxin Systems in Bacterial Persistence.....	14
1.1.6    The Role of Toxin-Antitoxin Systems in Stabilising Integrons.....	15
1.1.7    The Role of Toxin-Antitoxin Systems in Biofilm Formation.....	16
1.1.8    Cellular Targets & Mode of Action of Toxin Proteins .....	17
1.1.9    Regulation of Toxin-Antitoxin Operon Expression.....	23
1.1.9.1    Stoichiometry of Toxin-Antitoxin Complexes .....	23
1.2    The VapBC Toxin-Antitoxin Family .....	26
1.2.1    PIN-Domain Proteins .....	27
1.2.2    Characterised VapBC Toxin-Antitoxin Systems .....	28
1.3    Toxin-Antitoxin Systems in the <i>Mycobacteria</i> species .....	38
1.3.1    Toxin-Antitoxin Systems Present in <i>Mycobacterium tuberculosis</i> .....	38
1.3.2    The VapBC Toxin-Antitoxin System from <i>Mycobacterium smegmatis</i> ...	43

---

1.4	Objectives.....	45
<b>Chapter Two: Materials &amp; Methods.....</b>		<b>43</b>
2.1	General Materials & Methods.....	43
2.1.1	DNA Manipulations.....	43
2.1.1.1	<i>Mycobacterium smegmatis</i> Genomic DNA.....	43
2.1.1.2	<i>Escherichia coli</i> Plasmid DNA Extraction.....	43
2.1.1.3	Alkaline Lysis Plasmid Preparation Method.....	43
2.1.1.4	Agarose Gel Electrophoresis.....	44
2.1.1.5	DNA Extraction from an Agarose Gel.....	44
2.1.1.6	DNA Quantification.....	44
2.1.1.7	Polymerase Chain Reaction (PCR).....	45
2.1.1.8	Purification of PCR Products from Solution.....	46
2.1.1.9	Restriction Enzyme Digestion.....	46
2.1.1.10	DNA Ligation.....	47
2.1.1.11	DNA Transformation.....	47
2.1.1.12	Preparation of Electrocompetent <i>Escherichia coli</i> .....	47
2.1.1.13	Electroporation of <i>Escherichia coli</i> .....	47
2.1.1.14	Preparation of Electrocompetent <i>Mycobacterium smegmatis</i> .....	48
2.1.1.15	Electroporation of <i>Mycobacterium smegmatis</i> .....	48
2.1.2	Glycerol Stocks.....	49
2.1.3	Protein Purification and Manipulation.....	49
2.1.3.1	SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) Protein Analysis.....	49
2.1.3.2	Native Polyacrylamide Gel Electrophoresis (Native-PAGE) Protein Analysis.....	50
2.1.3.3	Coomassie Blue Stain for Protein Gel Electrophoresis.....	50

---

2.1.3.4	Silver Staining for Protein Gel Electrophoresis.....	50
2.1.3.5	Measurement of Protein Concentration .....	50
2.1.3.6	Concentration of Protein Samples .....	51
2.1.3.7	Dialysis of Protein Samples.....	51
2.1.3.8	Fast Performance Liquid Chromatography (FPLC) .....	51
2.1.3.9	Immobilised Metal Affinity Chromatography (IMAC).....	52
2.1.3.10	Purification of His-Tagged Proteins via IMAC.....	52
2.1.3.11	Size Exclusion Chromatography .....	52
2.1.3.12	Calibration of Size Exclusion Column S200 10/300.....	53
2.2	Methods relating to Chapter Three: Protein Expression, Purification & Characterisation.....	43
2.2.1	Cloning of <i>vapB</i> , <i>vapC</i> and <i>vapBC</i> into pYUB1049 .....	54
2.2.2	Protein Expression and Purification using <i>Mycobacterium smegmatis</i> as a Host .....	54
2.2.2.1	Small Scale Protein Expression and Purification from <i>Mycobacterium smegmatis</i> .....	54
2.2.2.2	Large Scale Protein Expression and Purification from <i>Mycobacterium smegmatis</i> .....	55
2.2.3	Cloning of <i>vapC</i> <sub>MS1284</sub> into pMAL-c2x MBP-fusion vector.....	56
2.2.4	Protein Expression and Purification of a VapC-MBP Fusion Protein ..	57
2.2.4.1	Small Scale MBP-VapC Protein Expression and Purification Tests. ....	57
2.2.4.2	Large Scale MBP-VapC Expression and Purification.....	58
2.2.5	Amylose Affinity Chromatography .....	59
2.2.5.1	Purification of MBP-Tagged Proteins via Amylose Affinity .....	59
2.2.5.2	MBP-Affinity Tag Cleavage by Factor Xa Protease .....	59
2.2.6	Denaturing Purification of VapC from <i>Escherichia coli</i> .....	60

---

2.2.7	Protein Refolding .....	60
2.2.7.1	Refolding Screen Solutions .....	60
2.2.7.2	Small Scale (20 $\mu$ l) Refolding Screen .....	60
2.2.7.3	Scale up (200 $\mu$ l) of Protein Refolding.....	61
2.2.8	Disruption of VapBC Complex from <i>Mycobacterium smegmatis</i> .....	61
2.2.8.1	VapBC Disruption Screens.....	62
2.2.9	Tryptic Digest of the VapBC Complex from <i>Mycobacterium smegmatis</i> .....	63
2.2.9.1	Trypsin Digest of VapBC Complex.....	63
2.2.9.2	Trypsin Inhibitors .....	63
2.2.10	Purification of VapC from <i>Mycobacterium smegmatis</i> .....	63
2.2.10.1	Purification of VapC by Anion Exchange Chromatography .....	63
2.2.10.2	Size Exclusion Chromatography of VapC.....	64
2.2.11	Dynamic Light Scattering .....	64
2.2.12	VapB Peptides.....	65
2.2.12.1	Disruption of the VapBC complex .....	65
2.2.12.2	Binding of VapB Peptides to VapC.....	65
2.2.13	DNA binding VapBC.....	65
2.2.13.1	DNA Oligonucleotide Design.....	65
2.2.13.2	Annealing of DNA Oligonucleotides .....	66
2.2.13.3	DNase Treatment of the VapBC Complex .....	66
2.2.13.4	RNase A Treatment of the VapBC Complex.....	67
2.2.13.5	DNA Exchange .....	67
2.2.13.6	Visualisation of DNA Bound to VapBC.....	68
2.2.13.7	Exonuclease Treatment of VapBC Complex.....	69

---

2.2.14	Initial Crystallisation Trials.....	69
2.2.15	Optimisation of Crystallisation Conditions (Fine Screening).....	69
2.2.15.1	Sitting Drops in Crystalclear Strips .....	70
2.2.15.2	Hanging Drops.....	70
2.2.16	Microseeding.....	70
2.2.17	Testing of Crystals by X-Ray Diffraction.....	70
2.2.18	MALDI TOF Mass Spectrometry of Whole Protein.....	71
2.2.18.1	Sample Preparation .....	71
2.2.18.2	MALDI-TOF Set Up .....	71
2.2.18.3	Analysis of Mass Spectra.....	71
2.3	Methods Relating to Chapter Four: Mechanism of Growth Inhibition by VapC in <i>Mycobacterium smegmatis</i> .....	72
2.3.1	Strains and plasmids used .....	72
2.3.2	Conditional expression of VapC in <i>Mycobacterium smegmatis</i> .....	72
2.3.3	Pulse-Chase Methodology .....	72
2.3.4	Methionine Uptake Methodology .....	73
2.3.5	VapC Antibody Production.....	74
2.3.6	Total Protein Determination.....	74
2.3.7	Western Blot.....	75
2.4	Methods Relating to Chapter Five: VapC Ribonuclease Activity .....	75
2.4.1	VapB and VapC Protein Purification.....	75
2.4.2	RNA Manipulation.....	76
2.4.2.1	Urea Denaturing Polyacrylamide Gel Electrophoresis Protein Analysis (Urea Denaturing-PAGE) .....	76
2.4.2.2	Visualisation of RNA.....	76
2.4.2.3	RNA Quantification.....	77

---

2.4.3	Production of Rotavirus RNA.....	77
2.4.3.1	Incorporation of <sup>33</sup> P-UTP into Rotavirus RNA.....	77
2.4.4	Urea-Denaturing Polyacrylamide Sequencing Gel .....	78
2.4.5	Isolation of Total RNA from <i>Mycobacterium smegmatis</i> .....	78
2.4.6	Two Base Combination RNA Oligonucleotide.....	79
2.4.7	Fluorogenic Substrate.....	80
2.4.7.1	Fluorometer Setup.....	80
2.4.8	The Pentaprobe System.....	81
2.4.8.1	Pentaprobe Plasmid Preparation .....	81
2.4.8.2	PCR Amplification of Pentaprobe Inserts .....	82
2.4.8.3	In Vitro Transcription of Pentaprobe Inserts .....	82
2.4.8.4	Producing Pentaprobe Double-Stranded RNA .....	83
2.4.8.5	Producing Pentaprobe Single Stranded DNA by Asymmetric PCR.. .....	83
2.4.8.6	Design of 932 Pentaprobe Oligonucleotides .....	83
2.4.9	Acridine Orange Nucleic Acid Staining .....	84
2.4.10	VapC Ribonuclease Activity Assay General Method.....	84
2.4.10.1	<i>Mycobacterium smegmatis</i> VapC Ribonuclease Assay.....	84
2.4.10.2	<i>Pyrobaculum aerophilum</i> VapC Ribonuclease Assay.....	85
2.4.11	MALDI TOF Mass Spectrometry of RNA Oligonucleotides .....	86
2.4.11.1	Sample Purification.....	86
2.4.11.2	Sample Preparation .....	87
2.4.11.3	MALDI-TOF Set Up .....	87
2.4.11.4	Analysis of Mass Spectra.....	88
<b>Chapter Three: Protein Expression, Purification &amp; Characterisation.....</b>		<b>89</b>
3.1	Introduction.....	89

---

3.1.1	Protein Expression in <i>Mycobacterium smegmatis</i> .....	89
3.1.2	Protein Expression Using an MBP-Fusion Protein in <i>Escherichia coli</i> .....	90
3.1.3	Protein Refolding .....	91
3.1.4	Auto-regulation of the VapBC Family.....	91
3.1.5	Protein Crystallisation.....	93
3.2	Results.....	94
3.2.1	Cloning of <i>vapB</i> , <i>vapC</i> and <i>vapBC</i> into pYUB1049 .....	94
3.2.2	Small Scale Expression Tests of VapB, VapC and VapBC.....	94
3.2.3	VapB Expression & Purification in <i>Mycobacterium smegmatis</i> .....	95
3.2.4	VapBC Expression & Purification in <i>Mycobacterium smegmatis</i> .....	96
3.2.5	Determination of the Molecular Weight of the VapBC Complex .....	98
3.2.6	Overexpression of VapB, VapC and VapBC does not affect growth of <i>Mycobacterium smegmatis</i> .....	100
3.2.7	Mutation of the VapC Expression Construct .....	101
3.2.8	Denaturing Purification of VapC .....	102
3.2.9	Refolding of VapC .....	103
3.2.10	Cloning of <i>vapC</i> into pMAL-c2x.....	103
3.2.11	Expression and Purification of a MBP-VapC Fusion .....	104
3.2.12	Isolation of VapC .....	108
3.2.13	Purification of VapC Following Trypsin Digestion of the VapBC Complex .....	111
3.2.14	VapB Peptide Inhibition of VapC.....	113
3.2.15	VapBC DNA Binding.....	113
3.2.16	Exonuclease Treatment of VapBC.....	116
3.2.17	Crystallisation screens.....	118
3.3	Discussion .....	120

---

3.3.1	Expression of VapB, VapC and VapBC and Isolation of VapC.....	120
3.3.2	Characterisation of the VapBC Complex from <i>Mycobacterium smegmatis</i> .....	124
<b>Chapter Four: Mechanism of Growth Inhibition by VapC in <i>Mycobacterium smegmatis</i> .....</b>		<b>129</b>
4.1	Introduction.....	129
4.2	Results.....	130
4.2.1	Pulse-Chase Methionine, Uridine and Thymidine Incorporation .....	130
4.2.2	Methionine Uptake to Measure Translation Inhibition.....	133
4.3	Discussion.....	138
<b>Chapter Five: VapC Ribonuclease Activity.....</b>		<b>141</b>
5.1	Introduction.....	141
5.1.1	Ribonuclease Activity of Toxin Proteins.....	141
5.1.2	Ribonuclease Activity of Eukaryotic PIN-Domain Proteins .....	142
5.1.3	Choice of Substrate for Ribonuclease Activity Assays.....	143
5.2	Results.....	144
5.2.1	Rotavirus RNA Ribonuclease Activity Assays.....	144
5.2.2	Total RNA from <i>Mycobacterium smegmatis</i> Ribonuclease Activity Assays .....	145
5.2.3	Fluorogenic RNA Substrate .....	147
5.2.4	Two Base Combination RNA Oligonucleotide Ribonuclease Activity Assays .....	147
5.2.5	Pentaprobe RNA Ribonuclease Activity Assays .....	148
5.2.6	Other possible VapC Substrates (Double-Stranded DNA & RNA and Single-Stranded DNA).....	153
5.2.7	Determination of the VapC Recognition Sequence .....	154
5.2.8	Bioinformatic Analysis of the VapC <sub>MS1284</sub> Target Sequence.....	169

---

5.2.9 Bioinformatic Analysis of the VapC <sub>PAE2754</sub> and VapC <sub>PAE0151</sub> Target Sequences .....	174
5.3 Discussion .....	174
5.3.1 Identification and Analysis of the Recognition Sites for VapC <sub>MS1284</sub> , VapC <sub>PAE2754</sub> and VapC <sub>PAE0151</sub> .....	174
5.3.2 The Role of VapC <sub>MS1284</sub> and Prevalence of the VapC <sub>MS1284</sub> Target Sequence in the Genome of <i>Mycobacterium smegmatis</i> .....	179
5.3.3 Comparison of VapC Ribonuclease Activity .....	180
5.3.4 Predicted Catalytic Mechanism of VapC .....	182
<b>Chapter Six: Discussion.....</b>	<b>186</b>
6.1 The Role of <i>Mycobacterium smegmatis</i> VapBC Proteins as Metabolic Managers .....	187
6.2 VapC from <i>Pyrobaculum aerophilum</i> .....	191
6.3 Conclusions .....	186
<b>Appendices .....</b>	<b>194</b>
Appendix A: Reagents .....	194
A.1 Primers, Plasmids & Bacterial Strains Used in This Study.....	194
A.2 Buffers and Solutions .....	195
A.3 Growth Media .....	198
A.3.1 Solid Media .....	198
A.3.2 Liquid Media .....	198
Appendix B: Gene & Protein Information .....	200
B.1 <i>Mycobacterium smegmatis</i> VapB and VapC Genomic & Protein Information .....	200
B.2. <i>Pyrobaculum</i> Protein Information .....	203
B.3. Pentaprobe Sequences + Flanking Vector Sequence .....	204
Appendix C: Raw Data.....	205
Appendix D: Publications .....	206
<b>References .....</b>	<b>207</b>

---

---

## List of Figures

Figure 1.1. Schematic diagram of a generalised toxin-antitoxin system. ....	4
Figure 1.2. Toxin-Antitoxin Plasmid Addiction. ....	5
Figure 1.3 Distribution of 60 toxin-antitoxin (TA) loci in the <i>M. tuberculosis</i> H37Rv genome. ....	8
Figure 1.4. Schematic diagram of the model proposed by the Engelburg-Kulka group for the <i>E. coli mazEF</i> mediated cell death pathway. ....	11
Figure 1.5. Model for <i>relBE</i> regulation in the context of the stringent response. ....	14
Figure 1.6. Toxins with RNase T1 folds. ....	20
Figure 1.7. The CcdB toxin is a DNA gyrase poison. ....	22
Figure 1.8. Ribbon Diagram of MazEF from <i>E. coli</i> . ....	25
Figure 1.9. The Hidden Markov Model which defines the PIN-domain family of proteins. ....	28
Figure 1.10. Ribbon diagram of FitAB heterodimers. ....	32
Figure 1.11. Ribbon Diagram of FitAB bound to the IR36 DNA fragment. ....	33
Figure 1.12. View of the four conserved acidic residues hypothesised to be the active site of VapC <sub>PAE0151</sub> . ....	35
Figure 1.13 VapC <sub>PAE0151</sub> and VapC <sub>PAE2754</sub> in their various oligomeric states. ....	37
Figure 1.14. Structure of VapBC-5 from <i>M. tuberculosis</i> . ....	42
Figure 3.1 <i>In vitro</i> DNA binding of purified VapBC complex. ....	93
Figure 3.2. Small scale expression screen of VapB, VapC and VapBC. ....	95
Figure 3.3. VapB IMAC purification and corresponding 15% SDS-PAGE gel. ....	96
Figure 3.4. VapBC IMAC purification and corresponding 15% SDS-PAGE gel. ....	97
Figure 3.5. VapBC size exclusion purification and corresponding 15% SDS-PAGE gel. ....	98
Figure 3.6. Calibration curve for the S200 10 300 size exclusion column. ....	99
Figure 3.7. Growth of <i>M. smegmatis</i> mc <sup>2</sup> 4517 expression strains over a 96 hour period. ....	100
Figure 3.8 VapC gene sequence and corresponding amino acids. ....	101

---

Figure 3.9. VapC denaturing IMAC purification and corresponding 15% SDS-PAGE gel.....	102
Figure 3.10 MBP-VapC amylose purification and corresponding 10% SDS-PAGE gel.....	105
Figure 3.11 Factor Xa cleavage of the MBP-VapC fusion. ....	106
Figure 3.12 MBP-VapC size exclusion purification and corresponding 10% SDS-PAGE gel. ....	107
Figure 3.13. Disruption of the VapBC complex with 6 M Guanidium, 8 M Urea, 0.1 and 0.5% SDS (15% SDS-PAGE Gel). ....	109
Figure 3.14. Trypsin Digest of VapBC. ....	110
Figure 3.15. Anion exchange purification of a trypsin digest of VapBC and corresponding 15% SDS-PAGE gel.....	111
Figure 3.16. Size exclusion chromatography of VapC and corresponding 15% SDS-PAGE gel. ....	112
Figure 3.17 DNase and RNase treatment of the VapBC Complex. ....	114
Figure 3.18. VapBC DNA exchange with biotin labelled dsDNA oligonucleotide. ....	116
Figure 3.19. Exonuclease treatment of the VapBC complex. ....	117
Figure 3.20. Exonuclease treated VapBC crystals. ....	119
Figure 3.21 FitB from <i>N. gonorrhoeae</i> with FitA peptide bound. ....	122
Figure 3.22. Ribbon diagram of FitA (mauve) peptide bound to FitB (green). C-terminal arginine of FitA located in the acidic pocket/active site of FitB, inhibiting its activity. ....	123
Figure 3.23. Tetramer of FitAB heterodimers from <i>N. gonorrhoeae</i> bound to IR36 DNA fragment. ....	125
Figure 4.1. Western blot showing VapC protein from mc <sup>2</sup> 155 containing pMind, pMind-VapC after induction with tetracycline and pMind-VapC without induction For the $\Delta vapBC$ strain, E, empty vector control, +Tc, induced and -Tc, uninduced VapC expression is shown. ....	131
Figure 4.2. Effect of VapC on (a) translation, (b) transcription and (c) DNA replication in <i>M. smegmatis</i> . ....	133

---

Figure 4.3. Growth curves of <i>M. smegmatis</i> $\Delta vapBC$ , $\Delta vapBC$ harbouring pMind, pMind- <i>vapC</i> and pMind- <i>vapBC</i> in HdB minimal media post-induction with tetracycline.....	134
Figure 4.4. Time course of VapC expression.....	135
Figure 4.5. Effect of VapC on translation.....	137
Figure 5.1. Ribonuclease activity of PAE0151 and PAE2754 using rotavirus RNA as a substrate.....	144
Figure 5.2. VapC ribonuclease activity of VapC <sub>PAE2754</sub> and VapC <sub>PAE0151</sub> from <i>P. aerophilum</i> and VapC <sub>MS1284</sub> from <i>M. smegmatis</i> using <i>M. smegmatis</i> total RNA as a substrate.....	146
Figure 5.3. Ribonuclease Activity of VapC <sub>PAE2754</sub> from <i>P. aerophilum</i> against pentaprobe RNA substrates.....	149
Figure 5.4. Ribonuclease Activity of VapC <sub>PAE0151</sub> from <i>P. aerophilum</i> against pentaprobe RNA substrates.....	150
Figure 5.5. VapC <sub>PAE2754</sub> and VapC <sub>PAE0151</sub> ribonuclease assay using pentaprobe 925 RNA as a substrate.....	151
Figure 5.6. Ribonuclease Activity of VapC <sub>MS1284</sub> from <i>M. smegmatis</i> against pentaprobe RNA substrates.....	152
Figure 5.7. VapC <sub>PAE2754</sub> , VapC <sub>PAE0151</sub> and VapC <sub>MS1284</sub> assays with dsDNA, ssDNA and dsRNA as substrates.....	154
Figure 5.8. Design of 932 RNA oligonucleotides 1-9.....	155
Figure 5.9. VapC <sub>MS1284</sub> ribonuclease activity against 932 RNA oligonucleotides....	156
Figure 5.10. VapC <sub>PAE2754</sub> and VapC <sub>PAE0151</sub> ribonuclease activity against 932 RNA oligonucleotides.....	157
Figure 5.11. VapC <sub>PAE2754</sub> and VapC <sub>PAE0151</sub> ribonuclease activity assays using 932 RNA oligo 5 as a substrate.....	158
Figure 5.12. Ammonium acetate and ethanol precipitation of MS1284 ribonuclease activity assays using 932 RNA oligos 1 and 5 as substrates.....	159
Figure 5.13. 932 RNA Oligo 5 VapC <sub>PAE2754</sub> MALDI-MS results..	161
Figure 5.14. 932 RNA Oligo 5, VapC <sub>PAE0151</sub> MALDI-MS results.....	162

---

Figure 5.15. 932 RNA Oligo 5 VapC <sub>PAE2754</sub> and VapC <sub>PAE0151</sub> cleavage sites and corresponding masses.....	163
Figure 5.16. 932 RNA Oligo 3, VapC <sub>PAE2754</sub> and VapC <sub>PAE0151</sub> MALDI-MS results.....	164
Figure 5.17. 932 RNA Oligo 3 VapC <sub>PAE2754</sub> and VapC <sub>PAE0151</sub> cleavage sites and corresponding masses.....	165
Figure 5.18. 932 RNA Oligo 3, VapC <sub>MS1284</sub> MALDI-MS results.....	166
Figure 5.19. 932 RNA Oligo 3 VapC <sub>MS1284</sub> cleavage sites and corresponding masses. .....	167
Figure 5.20. 932 RNA Oligo 4 VapC <sub>MS1284</sub> MALDI-MS results.....	168
Figure 5.21. 932 RNA Oligo 4 VapC <sub>MS1284</sub> cleavage sites and corresponding masses. .....	169
Figure 5.22. Analysis of VapC <sub>MS1284</sub> cleavage sites. ....	169
Figure 6.1. Model of mRNA stability. ....	189
Figure 6.2 Model for the role of VapC proteins as metabolic managers in <i>M. smegmatis</i> . ....	190

---

---

## List of Tables

Table 1.1. The nine toxin families, their targets, activities and the cellular process affected by toxin expression.....	6
Table 3.1. Variations of the inverted repeat sequence found in the <i>vapBC</i> promoter region that VapBC binds.....	127
Table 5.1. Free energy values at 37 °C for each pentaprobe RNA molecule as calculated by RNA Structure Version 4.6.....	152
Table 5.2. Prevalence of VapC target sequences in genes and their upstream regions in the <i>M. smegmatis</i> genome.....	170
Table 5.3. Prevalence of VapC target sequences in genes within the <i>M. smegmatis</i> genome .....	171
Table 5.4. Analysis of VapC <sub>MS1284</sub> target sequence in genes downregulated by VapC expression.....	173
Table A.1. List of Primers used in the study.....	194
Table A.2. List of bacterial strains and plasmids used in the study .....	194

---

---

## List of Abbreviations

SI (Système 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
Amp	Ampicillin
ATP	adenosine triphosphate
bp	base pair(s)
BLAST	basic local alignment search tool
C	cytosine
CAM	chloramphenicol
ccd	coupled cell division
CMC	critical micelle concentration
C-terminal	carboxy terminus
Da	Daltons
DEPC	diethyl pyruvate carbonate
dGMP	deoxy-guanine monophosphate
DLS	dynamic light scattering
DNase	deoxyribonuclease
dNMP	deoxynucleotide monophosphate
dNTP	deoxynucleotide triphosphate
doc	death on curing
ds	double-stranded
DTT	dithiothreitol
EDF	extracellular death factor
EDTA	ethylene diamine tetraacetic acid (disodium salt)
EF	elongation factor
EGTA	ethylene glycol tetraacetic acid

---

em.	emission
EMSA	electrophoretic mobility shift assay
ex.	excitation
fit	fast intracellular trafficking
FPLC	fast performance liquid chromatography
g	times the force of gravity
G	guanine
GC	<i>Neisseria gonorrhoeae</i> gonococcus
GFP	green fluorescent protein
GITC	guanidium isothiocyanate
GSH	glutathione
GSSG	glutathione disulfide
GST	glutathione S-transferase
GTE	glucose tris EDTA
HdB	hartmans de bont
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
HGT	horizontal gene transfer
hip	high persistence
His-tag	poly-histidine tag
HIV	human immunodeficiency virus
HPA	hydroxypicolinic acid
Hp Box	hydrophobic residues
HTH	helix turn helix
IMAC	immobilised metal affinity chromatography
IPTG	isopropylthio- $\beta$ -D-galactosidase
IR	inverted repeat
kb	kilobase
kDa	kilo dalton
kV	kilo volt
LB	luria bertani
LDAO	lauryl dimethyl amine oxide

---

LHH	looped helix helix
MAD	multi-wavelength anomalous diffraction
MALDI	matrix assisted laser desorption ionisation
mAU	milli-absorbance units
MB	megabase
MBP	maltose binding protein
MDR	multidrug resistant
MES	2-(N-morpholino)-ethanesulfonic acid
MPD	2-methyl 2, 4-pentanediol
mRNA	messenger RNA
MTBC	<i>Mycobacterium tuberculosis</i> complex
MS	mass spectrometry
N	A, T, C or G bases
Native-PAGE	non-denaturing PAGE
ND	not determined
NMD	nonsense mediated decay
NMR	nuclear magnetic resonance
N-terminal	amino terminus
NTHi	non-typeable <i>Haemophilus influenzae</i>
rNTP	ribonucleotide triphosphate
NYN	Nedd4-BP1, YacP nuclease
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
par	partitioning
PBS	phosphate buffered saline
PCD	programmed cell death
PCR	polymerase chain reaction
PCS	programmed cell stasis
PDB	protein data bank
PEG	polyethylene glycol

---

pI	isoelectric point
PIN	PilT N-terminal domain
PK	polynucleotide kinase
ppGpp	guanosine tetraphosphate
PSK	post segregational killing
RHH	ribbon helix helix
rpm	revolutions per minute
RNase	ribonuclease
RNAi	RNA interference
rNTP	ribonucleotide
rRNA	ribosomal RNA
rTEV	recombinant tobacco etch virus
SDS	sodium dodecyl sulphate
ss	single stranded
T	thymine
TA	toxin antitoxin
TAE	tris-acetate-EDTA
TB	tuberculosis
TBS	tris buffered saline
TBS-T	tris buffered saline + tween
TE	tris EDTA buffer
TEMED	N, N, N, N,-tetramethylethylenediamine
TOF	time of flight
TRX	thioredoxin
Tyr	Tyrosine
U	uracil
UTR	untranslated region
UV	ultra violet
Vap	virulence associated protein
v/v	volume per volume
WHO	world health organisation

---

WT	wild type
w/v	weight per volume
w/w	weight per weight
XDR	extensively drug resistant

---

## Chapter One: Antibiotic Resistance, Persistence and the Burden of Infectious Disease

Infectious diseases can be caused by a wide range of bacteria, resulting in non-symptomatic to life-threatening illnesses. In developed countries like the United States, bacterial infections are a leading cause of death in children and the elderly (Howard & Keisser, 1994). Hospitalised patients and those with chronic diseases are at especially high risk of bacterial infection (Murray & Rosenthal, 1998). Common bacterial infections cause pneumonia, ear infections, diarrhoea, urinary tract infections, and skin disorders. More serious illness caused by bacterial infections, include tuberculosis, influenza, cholera, gastroenteritis and haemolytic uremic syndrome caused by *Escherichia coli* 0157:H7.

Tuberculosis (TB) is one of the most devastating diseases worldwide. *Mycobacterium tuberculosis*, the causative agent of this disease claims approximately 1.7 million lives per annum (WHO, 2007). In 1993 the World Health Organisation (WHO) declared TB a global emergency as an estimated one third of the world's population carries the organism and around 9 to 10 million new cases of TB are reported each year (WHO, 2007). One of the main problems in the treatment of tuberculosis infection is the capacity of *M. tuberculosis* to enter a dormant stage that is less susceptible to antibiotics, dictating long treatment regimes (Connolly, Edelstein, & Ramakrishnan, 2007). The problem of TB has recently been compounded by the dramatic increase in multi-drug resistant (MDR) and extensively-drug resistant (XDR) strains of *M. tuberculosis* and the increased susceptibility of HIV-infected individuals (WHO, 2007).

As an intracellular pathogen *M. tuberculosis* encounters a variety of stressful environmental conditions during host aerosolisation, phagocytosis, active growth, latency and reactivation. The many toxin-antitoxin (TA) systems found in the *M. tuberculosis* genome have been proposed as stress response elements and this

---

suggests that they play a role in the persistent lifestyle of *M. tuberculosis* (Gerdes, *et al.*, 2005; Ramage, Connolly, & Cox, 2009).

## 1.1 Toxin-Antitoxin Systems

### 1.1.1 Discovery and Function of Toxin-Antitoxin Systems

TA loci were first identified on prokaryotic plasmids involved in the prevention of post segregational plasmid loss (Gerdes, Rasmussen, & Molin, 1986). TA loci are found on a wide range of bacterial/archaeal chromosomes (Pandey & Gerdes, 2005) and plasmids yet little is known about their physiological function.

Many bacterial pathogens possess TA systems that are often expanded in number in their genomes, including *Vibrio cholerae*, *Streptococcus pneumoniae*, *E. coli* CFT073, *E. coli* 0157 and *M. tuberculosis* (Fivian-Hughes & Davis, 2010; Pandey & Gerdes, 2005). TA loci have been linked to antibiotic response, persistence and dormancy in pathogenic bacteria (Agarwal, Mishra, Bhatnagar, & Bhatnagar, 2010; Christensen-Dalsgaard, Jorgensen, & Gerdes, 2010; Christensen, Pedersen, Hansen, & Gerdes, 2003; Fu, Donegan, Memmi, & Cheung, 2007; Gerdes, 2000; Sat, *et al.*, 2001). Persister cells are the small population of dormant cells amongst a population that is otherwise growing normally. This bet-hedging strategy enables the bacteria to survive in rapidly fluctuating environments. TA systems play a role in bacterial persistence by determining the onset and duration of the persistent/bacteriostatic state (Kim & Wood, 2009; Lewis, 2007; Rotem, *et al.*, 2010). Hence the main role suggested for TA systems is the control of growth under stress conditions (Gerdes, 2000; Gerdes, 2007). Some bacteria contain a large number of TA systems from different families and members of some TA families have even been implicated in programmed cell death (PCD) (Amitai, Yassin, & Engelberg-Kulka, 2004; Engelberg-Kulka, Amitai, Kolodkin-Gal, & Hazan, 2006).

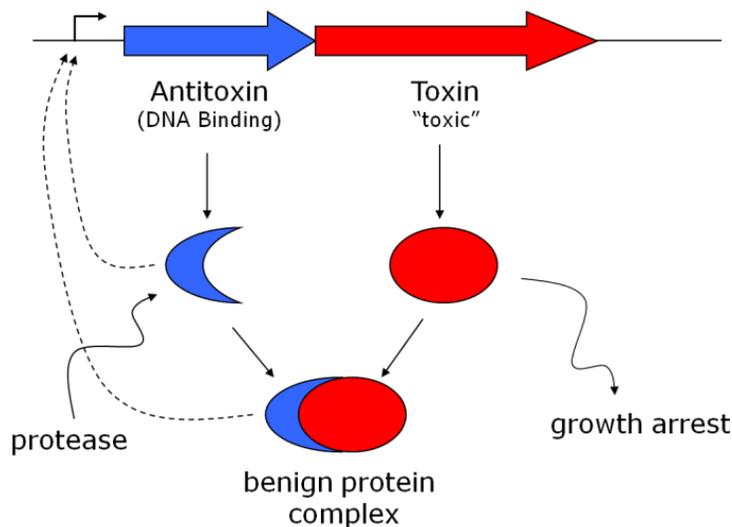
There are three types of TA systems; Type I where the antitoxin is an untranslated, antisense RNA, Type II where the antitoxin is a protein (Fozo, Hemm, & Storz, 2008; Gerdes, Gulyaev, Franch, Pedersen, & Mikkelsen, 1997) and Type III where the

---

antitoxin is an RNA molecule that binds to the toxin (Fineran, *et al.*, 2009). The antitoxin of Type I systems shows complementarity to the toxin mRNA and regulates toxin expression by inhibiting the toxin's translation. Toxins of Type I systems are generally small hydrophobic proteins that damage cell membranes (Gerdes, *et al.*, 1997). Type III TA systems have recently been characterised, the toxin is a protein and the antitoxin a small RNA molecule that binds to the toxin to inhibit its activity (Fineran, *et al.*, 2009)

This chapter will focus on Type II TA systems and outlines the hypothetical functions of these, including the role of TA systems in the stress response, bacterial persistence, stabilisation of integrons and biofilm formation. In addition the cellular targets and mode of action of toxin proteins such as ribosome-independent and ribosome-dependent inhibition of translation and inhibition of DNA replication will be outlined. The regulation of TA operon expression by varying stoichiometry of the TA complex will be discussed and also the VapBC TA family and TA families within the *M. tuberculosis* genome.

TA pairs are arranged as a bicistronic operon (Anantharaman & Aravind, 2003). Figure 1.1 depicts a schematic diagram of a generalised TA system. The toxin gene encodes a stable toxin and the antitoxin gene, a labile protein, that inhibits the activity of the toxin. The antitoxin binds to DNA via its N-terminus and interacts with specific sequences within the promoter of the TA operon (Gerdes, *et al.*, 1986). The toxin generally acts as a co-repressor that does not bind to DNA itself but increases the DNA binding affinity of the antitoxin. Thus, transcription is inhibited by binding of the TA complex to its promoter. When both proteins are produced by the cell, the antitoxin binds to the toxin inhibiting its toxic activity. The antitoxin is susceptible to degradation by cellular proteases induced under stress conditions which causes the cellular levels of the antitoxin to drop, leaving the toxin free to slow growth and/or kill the cell (Hayes, 2003).

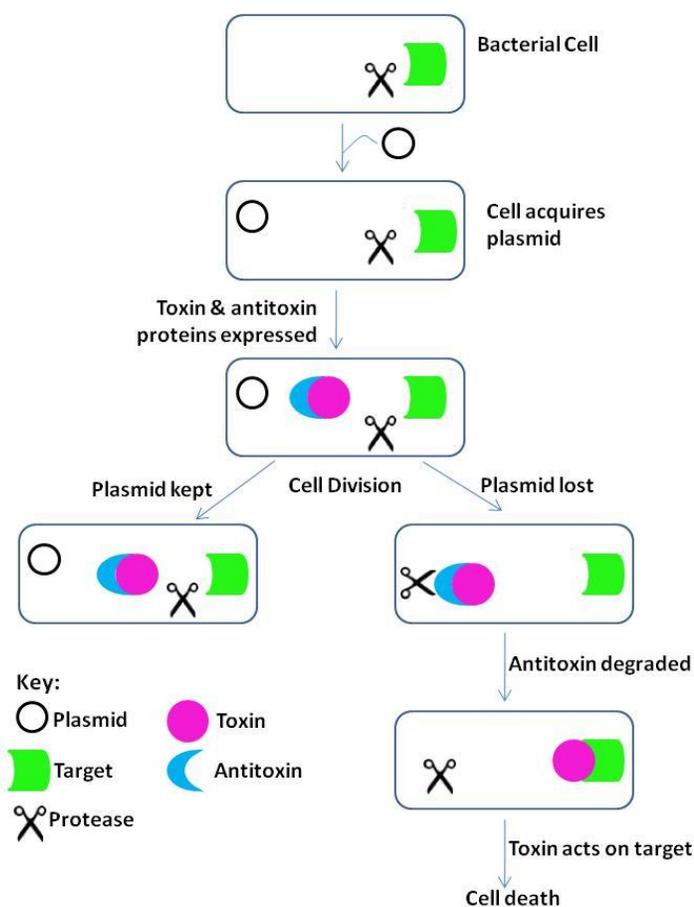


**Figure 1.1. Schematic diagram of a generalised toxin-antitoxin system.** The antitoxin binds to the promoter region of the TA operon, this binding is enhanced by the toxin (dotted arrows). The toxin gene encodes a toxic protein that often leads to growth arrest but when in complex with the antitoxin forms a benign protein complex. The antitoxin is more susceptible to proteolytic degradation.

TA systems were first identified as prokaryotic plasmid elements involved in the prevention of segregational plasmid loss (Buts, Lah, Dao-Thi, Wyns, & Loris, 2005; Gerdes, *et al.*, 1986). If the plasmid is lost during cell division the daughter cell lacking the plasmid is left with just the TA proteins. Over time the labile antitoxin is degraded and without translation of the antitoxin, the toxin persists and kills the cell (Figure 1.2). For this reason TA systems are often referred to as chromosomal or plasmid addiction molecules as the cells become “addicted” to antitoxin production and therefore maintenance of the TA genes. Plasmid encoded TA systems might also function in plasmid-plasmid competition (Cooper & Heinemann, 2000). A conjugative plasmid with a TA system has been shown to outcompete the same plasmid devoid of the TA system, thereby eliminating competitor plasmids in bacterial progeny by post segregational killing (PSK) and increasing the relative fitness of the host (Cooper & Heinemann, 2000). Plasmid TA systems can also be involved in the control of plasmid copy number thereby providing stability (Pimentel, Madine, & de la Cueva-Mendez, 2005).

Bacterial chromosomes as well as plasmids have TA systems; these are homologous to those on plasmids but are thought to play a different role to plasmid TA systems. As mentioned above plasmid TA systems provide a competitive advantage to competing plasmids whereas chromosomal TA systems have likely evolved a different function upon integration into the chromosome. The general school of

thought is that chromosomal TA systems play a role in the physiology of the cell in response to environmental cues resulting in either growth arrest and/or cell death (Buts, *et al.*, 2005). Chromosomal TA systems can also provide an anti-addiction function against plasmids (Saavedra De Bast, Mine, & Van Melderen, 2008). TA systems are present across a range of unrelated bacteria including many bacterial pathogens. TA systems are also common in the genomes of bacteria whose environmental conditions fluctuate and are often absent in related organisms that have a constant environment (Pandey & Gerdes, 2005).



**Figure 1.2. Toxin-Antitoxin Plasmid Addiction.** A plasmid containing a TA operon is introduced to the bacterial cell, leading to low level expression of both toxin and antitoxin, the antitoxin inhibiting the toxic effect of the toxin. If the plasmid is lost upon segregation the antitoxin which is more susceptible to proteolytic degradation is degraded and antitoxin levels are not replenished. The toxin acts upon its target to inhibit growth and kill the cell.

The toxins encoded by TA loci can be separated into nine families based on toxin sequence similarity: RelE, MazF, HipA, Doc, VapC, CcdB, ParE,  $\zeta$  and HicA. The targets of the toxins encoded by TA loci are similar for some groups and vary for others (Table 1.1). RelE, MazF, HipA, VapC and Doc toxin families all act to inhibit translation; RelE and Doc induce ribosome-dependent cleavage of mRNAs

(Christensen & Gerdes, 2003; Garcia-Pino, *et al.*, 2008; Gazit & Sauer, 1999), MazF is a sequence-specific endoribonuclease (Zhang, Zhang, Hoeflich, *et al.*, 2003) and VapC has endoribonuclease activity (Daines, Wu, & Yuan, 2007; Ramage, *et al.*, 2009). CcdB and ParE inhibit DNA gyrase by acting as gyrase poisons (Doa-Thi, *et al.*, 2005; Jiang, Pogliano, Heliniski, & Konieczny, 2002; Simic, De Jonge, Loris, Vesnaver, & Lah, 2009).  $\zeta$  is a phosphotransferase although its target is unknown (Meinhart, Alonso, Strater, & Saenger, 2003) and HipA is a protein kinase that acts on elongation factor EF-Tu (Korch & Hill, 2006). Table 1.1 provides a summary of TA families and their targets. The largest family of TA loci is the VapBC family.

Toxin	Target	Activity	Cellular Process
CcdB	DNA Gyrase	Generates double-stranded DNA breaks	Replication
RelE	Translating Ribosome	Induces mRNA cleavage	Translation
MazF	mRNAs	Endoribonuclease	Translation
ParE	DNA Gyrase	Generates double-stranded DNA breaks	Replication
Doc	Translating Ribosome	Induces mRNA cleavage	Translation
VapC	mRNAs	Endoribonuclease	Translation
$\zeta$	ND	Phosphotransferase	ND
HipA	EF-Tu	Protein Kinase	Translation
HicA	ND	mRNA cleavage	Translation

**Table 1.1. The nine toxin families, their targets, activities and the cellular process affected by toxin expression.** Table adapted from Van Melderren & Bast (2009). Abbreviations: ND, not determined; ccd, coupled cell division; par, partitioning; doc, death on curing; vap, virulence associated protein; hip, high persistence. RelE includes the HigB, YhaV, YoeB and YafQ subfamilies that show weak but significant sequence similarity.

Variations of TA operon characteristics are seen for the three-component TA system ( $\omega$ - $\epsilon$ - $\zeta$ ) in which separate proteins have antitoxin and auto-regulation properties. At a high level of expression the  $\zeta$  toxin inhibits replication, transcription and translation but the specific target is unknown (Gerdes, *et al.*, 1986). A further variation is seen for the genetic organisation of the HipBA, HicBA and MqsRA TA systems which differ from other TA systems in that the toxin gene is upstream of the antitoxin gene in the operon (Black, Kelly, Mardis, & Moyed, 1991; Korch & Hill, 2006; Makarova, Grishin, & Koonin, 2006).

---

All TA systems are regulated at the transcriptional level by their antitoxins which bind to TA loci promoters through their N-terminal domains. There are several classes of DNA binding proteins and more than one can be found in each antitoxin family: RelB antitoxins contain MetJ, Arc, CopG and ribbon-helix-helix (RHH) domains; YefM and Phd antitoxins contain a protein fold with DNA binding activity; MazE antitoxins contain AbrB transcription factor and looped-helix-helix (LHH) domains; ParE antitoxins contain MetJ, Arc and CopG domains; CcdB and  $\omega$  antitoxins contain MetJ, Arc and CopG domains; VapB antitoxins contain MetJ, Arc, CopG, helix-turn-helix (HTH), RHH, AbrB and YefM domains (Gerdes, *et al.*, 2005).

In 2005 Pandey & Gerdes undertook an extensive bioinformatic search of seven TA systems across the completely sequenced genomes (at the time) of 126 bacteria. The two largest TA families were VapBC and RelBE. These families were represented across bacteria and archaea, whereas MazEF, ParDE, HigBA (a member of the RelBE family) and Ccd were all constrained to bacteria (Pandey & Gerdes, 2005). VapBC TA systems were found to be the most common TA systems across archaea. All TA systems were present in both Gram negative and Gram positive bacteria with the exception of the Ccd TA system which is only found in Gram negative bacteria.

Members of a given TA family can be found at multiple locations in the genome of multiple bacterial species, suggesting that they arose by horizontal gene transfer (HGT) (Tsilibaris, Maenhaut-Michel, Mine, & Van Meldern, 2007). Further evidence is found in that many TA systems are often associated with either mobile genetic elements or on occasion, pathogenicity islands (Pandey & Gerdes, 2005). The phylogeny of TA systems is not consistent with bacterial phylogeny (Magnuson, 2007; Pandey & Gerdes, 2005) and their distribution differs greatly between isolates of the same bacterial species, implying that they are mobile (Magnuson, 2007; Mine, Guglielmini, Wilbaux, & Van Melderren, 2009).

Although different TA systems have different cellular targets, Gerdes proposed that they arose from a common ancestral gene as many toxins share weak structural or

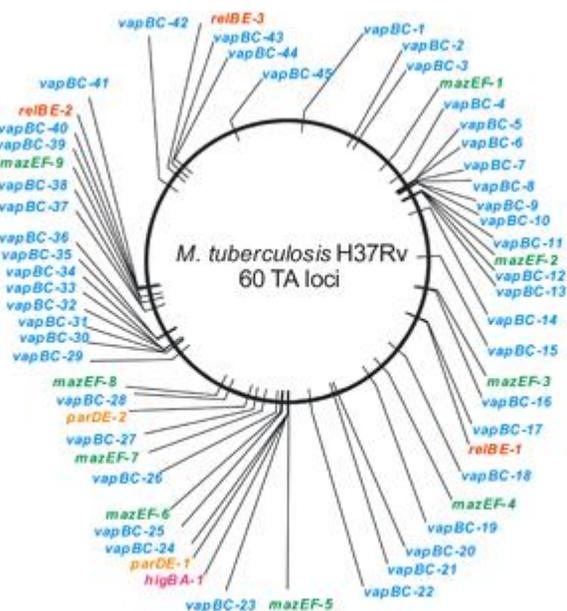
sequence similarity. For example *ccd/pem* and *parD* show weak sequence similarity (Gerdes, 2000), subsequent phylogenetic analyses by Anantharaman and Aravind showed a relationship between RelE and both ParE and HigB (Anantharaman & Aravind, 2003) suggesting that the RelE fold can be adapted to other targets.

The *M. tuberculosis* H37Rv genome has at least 60 TA loci (Figure 1.3) yet the related bacterium *Mycobacterium leprae* has none (Pandey & Gerdes, 2005). This poses the question why does *M. tuberculosis* have so many TA systems? It appears that host-associated organisms, whether parasitic or symbiotic, do not keep their TA systems when they live in a constant environment, whereas TA systems are beneficial to free living organisms that often face environmental change. This trend is seen in spirochetes; the free living

organisms *Leptospira interrogans* has five TA systems and *Treponema denticola* has 33 TA systems whereas the obligate parasitic spirochetes *Borrelia burgdorferi* and *Treponema pallidum* have none (Pandey & Gerdes, 2005). *Nitrosomonas europaea* has 45 TA systems encoded in its genome and is the only organism to have homologues of seven of the TA families (Gerdes, *et al.*, 2005).

### 1.1.2 Hypothetical Functions of Toxin-Antitoxin Systems

There are multiple mechanisms for the activation of TA systems and multiple mechanisms of action, which makes it difficult to assign a general physiological role to TA systems. It is possible that different TA systems have different functions. For



**Figure 1.3 Distribution of 60 toxin-antitoxin (TA) loci in the *M. tuberculosis* H37Rv genome.** Gerdes *et al.* (2005) identified the chromosomal locations of 3 *relBE*, 1 *higBA*, 2 *parDE*, 9 *mazEF* and 45 *vapBC* TA loci using exhaustive BLASTP and TBLASTN searches. Figure from Gerdes *et al.* (2005).

---

example, the *hipBA* TA locus has been implicated in bacterial persistence (Korch & Hill, 2006; Schumacher, *et al.*, 2009) whereas the *higBA* locus has been shown to be important for stabilisation of the superintegron of the *V. cholerae* chromosome (Christensen-Dalsgaard & Gerdes, 2006). The role of plasmid TA systems has been well documented, whereas there have been nine hypothetical functions proposed for chromosomal TA systems (Magnuson, 2007). These are: (1) Chromosomal TA systems are junk, acquired from plasmids or by HGT but are lost at a slow rate; (2) Stabilisation of genomic parasites (Christensen-Dalsgaard & Gerdes, 2006); (3) Selfish alleles: non-addictive alleles cannot replace the addictive TA alleles (Cooper & Heinemann, 2000); (4) Gene regulation, many toxins encode sequence or site specific endoribonucleases (Pedersen, *et al.*, 2003; Zhang, Zhang, Hoeflich, *et al.*, 2003; Zhu, *et al.*, 2009); (5) Growth Control, under stress conditions the toxin acts to inhibit growth in a bacteriostatic manner, this condition can be reversed upon removal of the stress (Boe, Gerdes, & Molin, 1987; Buts, *et al.*, 2005; Christensen, Mikkelsen, Pedersen, & Gerdes, 2001; Hazan, Sat, & Engelberg-Kulka, 2004); (6) Persisters, the *hipBA* locus has been implicated in the formation of persister cells, a subfraction of the population that has slow growth and high antibiotic resistance (Schumacher, *et al.*, 2009); (7) Programmed cell death one: early cell death mediated by the toxin results in more nutrients being available to the population (Lewis, 2000); (8) Programmed cell death two: later cell death results in nutrients becoming available to the population (Aizenman, Engelberg-Kulka, & Glaser, 1996; Engelberg-Kulka, *et al.*, 2006; Kolodkin-Gal & Engelberg-Kulka, 2006); (9) Antiphage, as bacteriophage can disrupt host transcription and translation, activation of TA systems could limit phage production (Hazan & Engelberg-Kulka, 2004; Hazan, Sat, Reches, & Engelberg-Kulka, 2001)

Given these many possible functions for TA systems, it is possible that different TA systems have evolved to perform different functions in different organisms. The possible functions of TA systems are explored in more detail below.

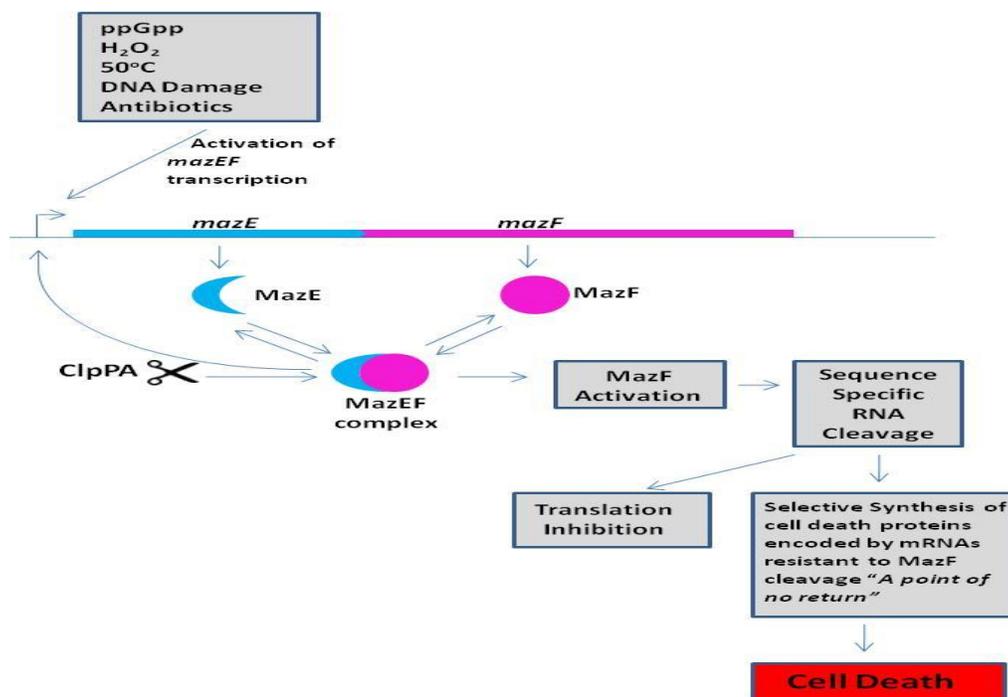
### 1.1.3 The Function of Toxin-Antitoxin Systems in *Escherichia coli*

The *E. coli* genome encodes at least seven well known TA families; *mazEF*, *relBE*, *chpBI-chpBK*, *yefM-yoeB*, *dinJ-yafQ* and *hipBA*. Other TA systems have been identified such as *prlF-yhaV*, *mqsRA*, *yafNO*, *kid-kis*, *yeeU-yeeV*, *yafW-ykfI* and *yfiZ-yjF* (Yamaguchi & Inouye, 2009), some of which require further characterisation to determine their function. Five of the earlier characterised TA systems *mazEF*, *relBE*, *chpBI-chpBK*, *yefM-yoeB* and *dinJ-yafQ* have been deleted from the *E. coli* K12 chromosome (Tsilibaris, *et al.*, 2007) to create an *E. coli*  $\Delta 5$  strain to investigate the role of TA systems in the general stress response of *E. coli*. Tsilibaris *et al.* (2007) subjected the *E. coli*  $\Delta 5$  strain to a variety of stress conditions and determined that growth inhibition or recovery was not dependent on any of the five TA systems. However, other TA loci present in *E. coli* could be involved in the stress response. The function of TA systems in bacteria and especially in *E. coli* is a subject of much debate; Gerdes hypothesised that they are involved in the stress response (Gerdes, *et al.*, 2005) and act to inhibit growth so the cell enters a reversible bacteriostasis, whereas Engelberg-Kulka hypothesises that they are involved in PCD via a quorum sensing mechanism, an altruistic suicide where the death of some of the population provides nutrients to the rest of the population (Engelberg-Kulka, *et al.*, 2006).

### 1.1.4 The Role of Toxin-Antitoxin Systems in the Stress Response

MazF expression is induced under various stressors leading to inhibition of growth or “quasi-dormancy” in *E. coli*. These quasi-dormant cells can produce protein if the mRNA is devoid of the MazF target sequence ACA (Yamaguchi & Inouye, 2009). An *E. coli* *mazEF* knockout infected with bacteriophage P1 was found to produce significantly more phage particles than the wild-type, leading to the conclusion that MazEF is involved in the exclusion of the P1 phage from the bacterial population (i.e. it inhibits the spread of the phage). MazF removes the infected or tainted fraction of cells from the population to protect the healthy uninfected cells (Hazan, *et al.*, 2004; Yamaguchi & Inouye, 2009).

The Engelburg-Kulka group stress that there is a point of no return, where MazEF mediates programmed cell death that cannot be rescued by MazE antitoxin production (Figure 1.4) (Amitai, *et al.*, 2004), but this *mazEF*-mediated cell death was only activated during logarithmic growth. During stationery phase, MazF toxin induction did not result in cell death (Kolodkin-Gal & Engelberg-Kulka, 2009).



**Figure 1.4. Schematic diagram of the model proposed by the Engelburg-Kulka group for the *E. coli mazEF* mediated cell death pathway.** MazF is a labile protein degraded by the ClpPA serine protease, MazF and MazE are coexpressed and interact to negatively autoregulate their transcription. Under conditions of stress (nutrient starvation, H<sub>2</sub>O<sub>2</sub>, temperature, DNA damage and antibiotics) *mazEF* expression is inhibited causing a decrease in the cellular concentration of MazE leaving MazF to cleave RNA. This results in a reduction in translation and after a “point of no return” the selective synthesis of cell death proteins encoded by mRNAs resistant to MazF cleavage results in cell death. Figure adapted from Engelburg-Kulka, Hazan & Amitai (2005).

Alternatively, the Gerdes group suggest, that instead of inducing cell death, MazF induces a reversible bacteriostatic state (Pedersen, Christensen, & Gerdes, 2002). The Engelburg-Kulka group suggest that physiological conditions that lead to an increase in guanosine tetraphosphate (ppGpp) levels reduces the synthesis of the MazE antitoxin (Aizenman, *et al.*, 1996), alongside degradation of the labile antitoxin by the ClpAP protease, and unleashes the endoribonuclease activity of MazF to execute programmed cell death (Amitai, *et al.*, 2004) (Figure 1.4). ppGpp is a global regulator

of transcription during the stringent response and acts to modulate transcription in response to changing nutritional demands (Gerdes, *et al.*, 2005). The Gerdes group examined the transcription pattern of *mazEF* during amino acid starvation and no TA dependent killing of cells was observed (Christensen, *et al.*, 2003). They also showed that the toxicity of MazF can be rescued by MazE antitoxin production within six hours post MazF induction. (Christensen, *et al.*, 2003; Pedersen, *et al.*, 2002). *E. coli* cells modified to over-express MazF have been shown to retain some transcriptional and translational ability for four days, despite growth arrest (Suzuki, Zhang, Liu, Woychik, & Inouye, 2005).

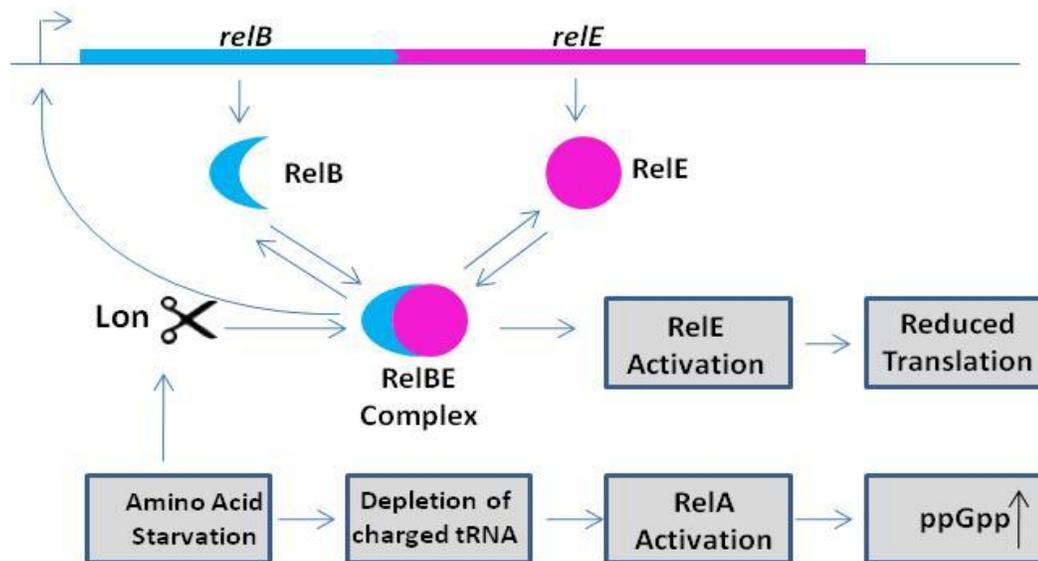
The Engelberg-Kulka group hypothesise that *E. coli mazEF*-mediated cell death is a population phenomenon (dependent on culture density) that requires a quorum sensing factor; a linear pentapeptide NNWNN, that they call the extracellular death factor (EDF). Each of the amino acids present in the peptide are essential for activity (Kolodkin-Gal, Hazan, Gaathon, Carmeli, & Engelberg-Kulka, 2007). MazF activation by proteolytic degradation of MazF under stress (Figure 1.4) leads to an increase in EDF, thereby increasing cell death, but the EDF response is dependent on the protease ClpAP (Kolodkin-Gal & Engelberg-Kulka, 2008). The *mazEF* locus is required for both EDF production and response, along with genes encoding *zwf* (glucose-6-phosphate dehydrogenase, the possible precursor of the EDF) and the protease ClpXP (Kolodkin-Gal & Engelberg-Kulka, 2008; Kolodkin-Gal, *et al.*, 2007). It should be noted that attempts to replicate MazF PCD by other groups have been unsuccessful and these results remain controversial (Christensen, *et al.*, 2003; Tsilibaris, *et al.*, 2007). The conflicting evidence indicates that the physiological role of the MazEF TA system is still poorly understood.

*Myxococcus xanthus*, a Gram-negative soil-dwelling bacterium, has a single *mazF* gene that is regulated by a serine/threonine protein kinase cascade (Nariya & Inouye, 2008). *M. xanthus* has a multi-cellular fruiting body such that, during development, 80% of the cells undergo obligatory cell lysis and deletion of the *mazF* gene drastically reduces this to between 5 and 18% of the wild-type (Nariya & Inouye,

---

2008). Although there is no MazE antitoxin, MrpC, a developmental regulator, can function as an antitoxin for MazF and is also a transcriptional activator. MrpC that is not in complex with MazF is phosphorylated thereby inhibiting its action as a transcriptional regulator of MazF and other developmental genes (Nariya & Inouye, 2008). It is unknown how 20% of the population survive autolysis. This is the first unambiguous example that MazF plays an essential role in PCD.

The RelBE TA family is a well characterised group of toxin-antitoxins. During normal growth RelB is expressed in excess over RelE (Gotfredsen & Gerdes, 1998), but RelB is degraded by Lon proteases, activating RelE (Christensen, *et al.*, 2004; Christensen, *et al.*, 2001). Activation of RelE by amino acid starvation inhibits translation (not completely) this reduces the level of charged tRNAs and activates RelA to synthesise ppGpp and stringent control of gene expression ensues (Gerdes, *et al.*, 2005). Amino acid changes that result in a metabolically unstable RelB leads to hyperactivation of RelE, independent of ppGpp and a complete inhibition of translation during amino acid starvation and a delayed relaxed response (Christensen & Gerdes, 2004). Depletion of charged tRNA is halted, meaning RelA is no longer active and as a result the concentration of ppGpp is restored to pre-starvation levels and rRNA synthesis resumes (i.e. there is a continuation of RNA synthesis after a lag period following the onset of starvation) rather than a reduction in translation as seen for wild-type RelB (Figure 1.5) (Christensen-Dalsgaard, *et al.*, 2010; Christensen & Gerdes, 2004; Christensen, *et al.*, 2001).



**Figure 1.5. Model for *relBE* regulation in the context of the stringent response.** RelE activation reduces translation (not completely inhibited). This reduces the level of charged tRNAs and activates RelA leading to stringent control of gene expression through ppGpp. Figure adapted from Gerdes *et al.* (2005)

Induction of RelE leads to inhibition of protein synthesis and a reduction in colony forming units (Christensen, *et al.*, 2001). However, expression of RelB will resuscitate RelE-inhibited cells, indicating that cells are in stasis, unable to form colonies, rather than dead (Pedersen, *et al.*, 2002). RelB directly inhibits RelE by displacing a helix from the catalytic position in the active site (Li, Zhang, Inouye, & Ikura, 2009). Gerdes and colleagues have shown several lines of independent evidence that RelE is responsible for the bacteriostatic condition: (1) RelE has been shown to inhibit translation both *in vivo* and *in vitro*, (2) mutation of R81A in RelE reduces its toxic effect *in vivo*, and (3) expression of RelE reverses inhibition of translation and cell growth (Christensen, *et al.*, 2001; Pedersen, *et al.*, 2002).

### 1.1.5 The Role of Toxin-Antitoxin Systems in Bacterial Persistence

Persister cells can be described as a small population of dormant cells amongst a population that is growing normally. Single cell measurements have shown that HipBA determines the onset and duration of a bacteriostatic or persistent state (Rotem, *et al.*, 2010). The toxin, HipA is a validated biofilm tolerance factor whose overexpression leads to multi-drug tolerance in *E. coli*, although the mechanism is

---

unknown (Korch & Hill, 2006). A mutation in HipA (HipA7) has a high persistence phenotype in *E. coli* due to the weakened interaction between HipA and its antitoxin, HipB, unleashing the kinase activity of HipA (Scherrer & Moyed, 1988; Schumacher, *et al.*, 2009). The cellular level of HipA must reach a threshold for growth inhibition to occur, and the level of HipA and HipB determine the duration of growth inhibition (Rotem, *et al.*, 2010). The threshold provides a mechanism in which differences in gene expression between cells in a population results in some entering a dormant state while others grow normally.

The *mqsRA* TA locus (also known as YgiUT) is upregulated in persister cells (Yamaguchi, Parkhill, & Inouye, 2009). It is a member of the RelBE TA family; although it shows no sequence similarity to other TA families. The structure of MqsR is similar to that of RelE (Brown, *et al.*, 2009). *E. coli mqsRA* transcription is activated by amino acid and glucose starvation (Christensen-Dalsgaard, *et al.*, 2010) and MqsR cleaves RNA, independent of translation (Christensen-Dalsgaard, *et al.*, 2010; Yamaguchi, *et al.*, 2009). Overexpression of MqsR increases the formation of persister cells (Kim & Wood, 2009), presumably due to its RNase activity and interaction of MqsA with the CspD toxin (Kim & Wood, 2009).

### **1.1.6 The Role of Toxin-Antitoxin Systems in Stabilising Integrons**

The *Vibrio cholerae* genome is made up of two chromosomes of 3.0 and 1.1 MB. The smaller of these, chromosome II, harbours a large integron termed a superintegron. Genes on the superintegron have been identified to have adaptive functions associated with pathogenicity, antibiotic resistance, immunity proteins and restriction endonucleases (Rowe-Magnus, Guerout, Biskri, Bouige, & Mazel, 2003). Two *higBA* loci are located on the superintegron of *V. cholerae* and these both encode functional toxins; HigB-1 and HigB-2 respectively. Expression of these toxins inhibits growth in *E. coli* by inhibition of protein synthesis, which can be counteracted by expression of HigA-1 and HigA-2 (Christensen-Dalsgaard & Gerdes, 2006). HigB appears to be bacteriostatic not bactericidal and both systems are activated by amino acid starvation in *V. cholerae* (Budde, Davis, Yuan, & Waldor, 2007; Christensen-Dalsgaard &

---

Gerdes, 2006). Christensen-Dalsgaard and Gerdes (2006) hypothesise that the two *higBA* loci function as stress response elements to shift from rapid cell growth (in the human intestine) to slow growth or dormancy in its aquatic environment, or that the two *higBA* loci could maintain the genetic stability of the *V. cholerae* superintegron as they efficiently stabilised a test plasmid in *E. coli* (Christensen-Dalsgaard & Gerdes, 2006). There are 74 homologues of *higBA* are present in the genomes of both Gram positive and Gram negative organisms (Pandey & Gerdes, 2005).

The *mosA* and *mosT* genes form a TA system that promotes the stability of SXT, an integrative and conjugative genetic element found in *V. cholerae* isolates that confers multiple antibiotic resistance (Wozniak & Waldor, 2009). Induction of MosT retarded cell growth, which could be counteracted by MosA. MosA is also responsible for auto-regulation of the *mosA-mosT* operon, which are both characteristics of TA loci. Expression of MosT under normal growth conditions is low. However, activation of proteins that promote excision of SXT de-represses the *mosAT* operon (Wozniak & Waldor, 2009). MosT is proposed to maintain SXT due to its ability to slow/stop cell growth upon SXT excision so cell division won't occur while the SXT island is extrachromosomal (Wozniak & Waldor, 2009). The SXT island will copy and transfer before re-integrating prior to cell division (Wozniak & Waldor, 2009).

### **1.1.7 The Role of Toxin-Antitoxin Systems in Biofilm Formation**

The YbaJ-Hha TA system in *E. coli* can control biofilm formation by limiting Hha-mediated fimbriae production and can also control cell death, deletion of *ybaJ-hha* reduces biofilm mass by 81% in minimal media and 50% in complex media (Barrios, Zuo, Ren, & Wood, 2006). It does so by controlling translation via the availability of rare codon tRNAs, which reduces fimbriae production and activates prophage lytic genes (Garcia-Contreras, Zhang, Kim, & Wood, 2008). YbaJ neutralises the toxicity of Hha but surprisingly Hha, not YbaJ, is shown to regulate its transcription via negative feedback (Garcia-Contreras, *et al.*, 2008). Deletion of *ybaJ-hha* in *E. coli* affects biofilm formation, motility and plasmid conjugation

(Barrios, *et al.*, 2006). Further investigation is needed to determine the interactions between YbaJ and Hha.

### 1.1.8 Cellular Targets & Mode of Action of Toxin Proteins

#### *Inhibition of Translation*

The most common cellular target of toxin proteins is translation, leading to growth arrest. Toxin proteins can associate with the ribosome to induce ribonuclease activity or exhibit ribonuclease activity themselves while associated with the ribosome. Alternatively ribonuclease activity by the toxin can be ribosome-independent, where the mRNA is cleaved directly.

#### *Ribosome-Independent Inhibition of Translation*

Toxins that exhibit ribosome-independent inhibition of translation by cleaving cellular mRNAs are often termed mRNA interferases. These proteins can cleave mRNA in the cytoplasm as the RNA does not have to be undergoing translation for cleavage to occur.

*In vivo*, MazF and ChpBK proteins inhibit translation by their ability to cleave mRNA (Christensen, *et al.*, 2003; Pedersen, *et al.*, 2002). MazF from *E. coli* cleaves mRNA between A and C residues at ACA sequences, although cleavage before the first A has also been observed (Zhang, Zhang, Hara, Kato, & Inouye, 2005; Zhang, Zhang, Hoeflich, *et al.*, 2003). Whereas MazF from *S. aureus* targets the pentad sequence U\*ACAU (\* indicates cleavage site) (Zhu, *et al.*, 2009) not the U rich sequences VU\*UV' as originally thought (Fu, *et al.*, 2007). The difference in these proposed target sequences of MazF from *S. aureus* is due to the RNA substrate used in the different experiments (Fu, *et al.*, 2007; Zhu, *et al.*, 2009). Homology searches identified *mazF* genes on the *M. tuberculosis* chromosome using *mazF* from *E. coli* (Pandey & Gerdes, 2005) and the cleavage sites of MazF homologues were subsequently determined. MazF-mt1 cleaves CU\*ACC and to a lesser extent UU\*ACA sequences (Zhu, *et al.*, 2006) whereas MazF-mt6 cleaves in U rich regions; the consensus cleavage sites for MazF-mt6 were defined as (U/C)U\*(A/U)C(U/C)

(Zhu, *et al.*, 2006). MazF-mt3 cleaves at UU\*CCU or CU\*CCU sequences and MazF-mt7 at U\*CGCU (Zhu, *et al.*, 2008). MazF does not cleave RNA/DNA or RNA/RNA duplex sequences (Zhang, Zhang, Hoeflich, *et al.*, 2003; Zhu, *et al.*, 2009). MazF from *M. xanthus* preferentially cleaves at GU\*UGC, however the first G can be replaced with an A (Nariya & Inouye, 2008). Induction of ChpBK (a member of the MazEF TA family) in *E. coli* reduces translation by 60% (Christensen, *et al.*, 2003; Zhang, Zhu, Zhang, & Inouye, 2005). ChpBK cleaves mRNA at the 5' or 3' side of the A residue in ACY sequences, where Y is U, A or G (Zhang, Zhu, *et al.*, 2005). The endoribonuclease activity of ChpBK appears to be slower than that of MazF, which can also be seen in the extent that the two proteins inhibit translation (Zhang, Zhu, *et al.*, 2005). Zhang, Zhu *et al.* (2005) partially attribute this to the difference in the isoelectric points (pIs) of the two proteins (ChpBK has an acidic pI of 5.2 compared to the basic pI of 8.3 for MazF) and hypothesise different physiological roles for the two proteins.

MazF and ChpBK cleave 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, Zhang, *et al.*, 2005). Unlike other endoribonucleases MazF does not require Mg<sup>2+</sup> for activity and in this respect it has been likened to RNase A alongside its requirement for a 2' OH group at the cleavage site (Zhang, Zhang, *et al.*, 2005). This also explains why MazF does not cleave ssDNA.

Another member of the MazEF family is the *parDE* system present on plasmid R1 and R100 from *E. coli* and on the chromosome of *Bacillus anthracis* which also encodes the toxin *pemK* (also known as *kid*). The toxin Kid/PemK cleaves RNA in a ribosome independent manner, similar to that of MazF (Agarwal, *et al.*, 2010). The Kid/PemK toxin dimerises to give two RNA binding sites, but only one of the two RNA binding sites is occupied at a time. Kid from *E. coli* preferentially cleaves 5' or 3' to the A residue at UAH sequences where H is C, A or U in ssRNA, although the same sequences in dsRNA are also efficiently cleaved (Munoz-Gomez, Lemonnier, Santos-Sierra, Berzal-Herranz, & Diaz-Orejas, 2005). PemK from *B. anthracis* has

---

base specificity for the pyrimidines (C/U) (Agarwal, *et al.*, 2010) and has the ability to interact with free RNA and ribosomes, although how it associates with ribosomes is unknown. The mechanism of catalysis for both proteins has been well studied. A catalytic base deprotonates the 2' OH of uracil which then facilitates nucleophilic attack on phosphorous. Transphosphorylation then terminates once the catalytic acid donates a H atom to the adenosine 5' oxygen atom (Kamphuis, *et al.*, 2006). The target of the Kid toxin from *E. coli* is the ColEI RNA II primer that is required for ColEI initiation of replication. EndoA from *Bacillus subtilis* is 25% identical to MazF from *E. coli* and targets UAC sequences similar to PemK from *E. coli* (Pellegrini, Mathy, Gogos, Shapiro, & Condon, 2005).

Other TA systems that cleave free mRNA to inhibit translation include MqsR that cleaves RNA at GCU sequences independent of translation (Christensen-Dalsgaard, *et al.*, 2010). This is the first example of a RelE homologue as an mRNA interferase (Christensen-Dalsgaard, *et al.*, 2010). HicA induces cell stasis by inhibiting the global rate of translation presumably due to its ability to induce translation-independent mRNA cleavage (Jorgensen, Pandey, Jaskolska, & Gerdes, 2009). Whether HicA itself cleaves the mRNA or if it induces cleavage by activating an RNase is unknown.

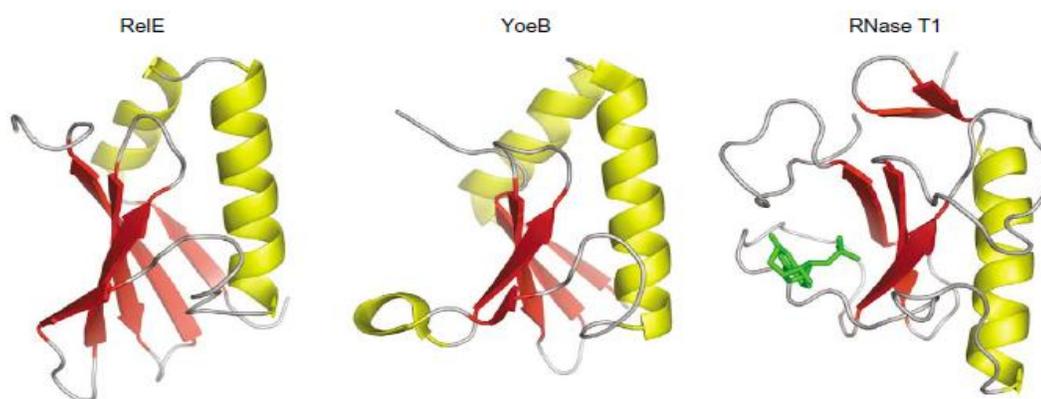
Cleavage sites of MazF homologues in *S. aureus* differ in their recognition sequence to that of their Gram negative counterparts (Fu, *et al.*, 2007; Zhu, *et al.*, 2009). There is also a difference in recognition sequences between homologues within the same species (Zhu, *et al.*, 2008; Zhu, *et al.*, 2006). Differential mRNA degradation by the different toxin genes could be a regulatory mechanism for adaptation to different stresses/environmental conditions.

### ***Ribosome-Dependent Inhibition of Translation***

RelE proteins from both bacteria and archaea cleave mRNA at the ribosomal A site (Pedersen, *et al.*, 2003). RelE proteins are characterised by a microbial RNase fold that resembles that of RNase T1 from *Aspergillus oryzae* (Figure 1.6) (Buts, *et al.*,

2005). Members of the RelBE family generally inhibit translation by ribosome-dependent mechanisms.

RelE associates with ribosomes and inhibits protein synthesis by cleaving specific mRNA codons in the ribosomal A site (Christensen & Gerdes, 2003; Pedersen, *et al.*, 2003). RelE usually initiates cleavage between the second and third bases in the codon, and has a higher specificity for the stop codon UAG. It also cleaves in the E site after peptide release (Pedersen, *et al.*, 2003).



**Figure 1.6. Toxins with RNase T1 folds.** RelE and YoeB both contain the conserved ribonuclease T1 fold. The three proteins are shown with  $\alpha$  helices in yellow and  $\beta$  strands in red. 2'guanosine monophosphate is shown as green sticks bound in the active site of RNase T1. Figure from Buts *et al.* (2005).

RelE has an incomplete RNase T1 fold that lacks the catalytic amino-acid triad present in homologous RNases (Figure 1.6) (Li, *et al.*, 2009). The crystal structure of *E. coli* RelE bound to the 70S ribosome from *Thermus thermophilus* has recently been determined (Neubauer, *et al.*, 2009) and confirms that RelE occupies the A site and induces cleavage of the mRNA after the second base by re-orientating and activating mRNA for 2'OH dependent hydrolysis (Neubauer, *et al.*, 2009). Stacking of the A site bases with conserved residues in RelE and 16S rRNA (Neubauer, *et al.*, 2009) explains the requirement of the ribosome for catalysis and the subtle sequence specificity of RelE.

HigB, YoeB, YafQ and YafO are members of the RelBE TA family and their activity is dependent on association with the 50S ribosomal subunit (Hurley & Woychik,

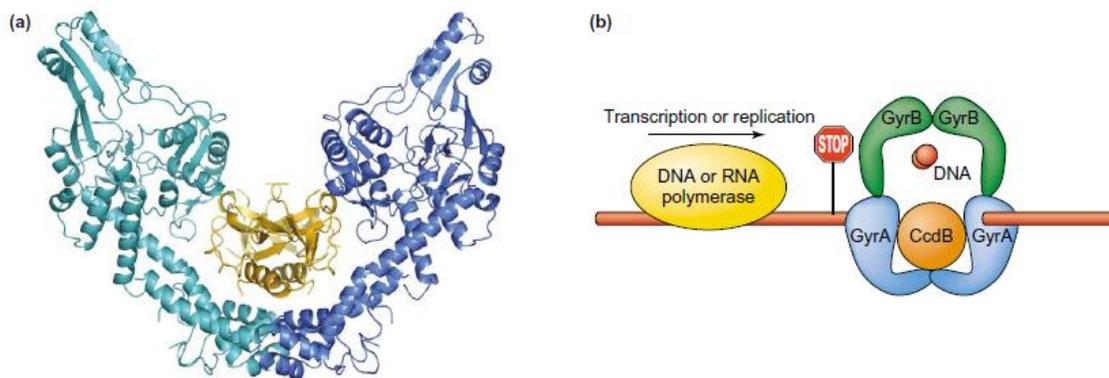
2009). HigB cleaves at AAA triplet sequences both in-frame and out-of-frame, and also cleaves at approximately 20% of AA sequences and was even seen to occasionally cleave at single A bases (Hurley & Woychik, 2009). Whereas YafQ cleaves in a sequence-specific, codon-dependent manner at  $5'AAA-G/A^3'$  sequences (Prysak, *et al.*, 2009) and YoeB cleaves at stop codons (Zhang & Inouye, 2009). HigB and YafQ are responsible for cleavage of RNA rather than initiating ribosome-mediated cleavage (Hurley & Woychik, 2009; Prysak, *et al.*, 2009). The exact mechanism of how YafO inhibits protein synthesis is unknown but is hypothesised to induce mRNA cleavage at the 3' end of the mRNA region that is unprotected by the 70S ribosome (Zhang, Yamaguchi, & Inouye, 2009).

The molecular mechanism of Doc action is different to that of other well characterised toxins from other TA systems. Expression of Doc results in growth arrest but it does not cleave mRNA (Liu, Zhang, Inouye, & Woychik, 2008). Induction of Doc shows significant similarities to the action of the aminoglycoside antibiotic hygromycin B, suggesting that Doc interacts with the 30S ribosomal subunit blocking elongation factor binding. Thereby inhibiting elongation of translation and stalling the ribosomes, causing an increase in mRNA stability as seen upon Doc expression (Liu, *et al.*, 2008).

### ***Inhibition of DNA Replication***

The target of the CcdB and ParD toxins is an essential type II topoisomerase in *E. coli*, DNA gyrase (Bernard, *et al.*, 1993). DNA gyrase is an essential enzyme that introduces negative supercoils in DNA using ATP and has double-stranded break and rejoining activity to relieve the constraints from replication and transcription (Couturier, Bahassi, & Van Melderen, 1998). In *E. coli* DNA gyrase consists of two subunits, GyrA containing the catalytic domain and GyrB containing the ATP hydrolysis site (Simic, *et al.*, 2009). CcdB both traps DNA gyrase on DNA and inhibits DNA gyrase forming a complex with free GyrA subunits (Maki, Takiguchi, Miki, & Horiuchi, 1992). The crystal structure of CcdB bound to the GyrA dimerisation domain shows that CcdB binds to GyrA when in its open conformation

(Figure 1.7a) (Doa-Thi, *et al.*, 2005). The CcdB toxin blocks DNA replication by forming a CcdB-DNA gyrase complex that blocks DNA and RNA polymerase (Figure 1.7b) and also converts supercoiled DNA to a singly cleaved linear form (Couturier, *et al.*, 1998), leading to cell death. The toxin ParE prevents initiation of replication at *oriC* (Ruiz-Echevarria, Gimenez-Gallego, Sabariego-Jareno, & Diaz-Orejas, 1995) by inhibiting DNA gyrase and converting DNA to a singly cleaved linear form in a mechanism similar to that of CcdB (Jiang, *et al.*, 2002).



**Figure 1.7. The CcdB toxin is a DNA gyrase poison.** (a) Structure of CcdB with GyrA59 (the N-terminal catalytic domain of gyrase subunit). Two CcdB monomers are in orange and yellow and the two GyrA59 monomers in blue and cyan. (b) CcdB mechanism of action. Gyrase creates a break in the DNA through which a second DNA segment can be transferred while the first segment remains covalently attached to gyrase. During this process CcdB attaches itself to gyrase, thereby blocking DNA and RNA polymerase. Figure from Buts *et al.* (2005).

### ***Other Modes of Action of Toxin Proteins***

HipA is a protein kinase that phosphorylates EF-Tu to block aminoacyl tRNA binding (Correia, *et al.*, 2006). The structure shows a specialised kinase fold that has a high affinity for ATP (Schumacher, *et al.*, 2009). Other cellular targets of HipA are likely due to its ability to arrest both transcription and translation (Korch & Hill, 2006).

A MazF homologue in *M. tuberculosis*, Rv1495 interacts with DNA topoisomerase I, the single topoisomerase gene (MtbTopA) on the *M. tuberculosis* chromosome. MazF physically interacts with MtbTopA, resulting in inhibition of MazF endoribonuclease activity and MtbTopA topoisomerase activity (Huang & He, 2010). Bacterial topoisomerase I prevents supercoiling of DNA during transcription and may be

important for adaptation to stressful environments (Huang & He, 2010). Two different mechanisms of growth inhibition have been proposed for Rv1495 (1) mRNA interferase activity inhibits growth by cleaving mRNA and (2) the physical interaction of Rv1495 with MtbTopA negatively regulates relaxation of supercoiled DNA (Huang & He, 2010).

### **1.1.9 Regulation of Toxin-Antitoxin Operon Expression**

All TA operons examined auto-regulated at the transcriptional level by the antitoxin which binds to TA operon promoters through their N-terminal domains. Members of all nine TA families have been analysed for transcriptional control. There is a common pattern of regulation in all the two-component TA systems (the three-component system  $\omega$ - $\epsilon$ - $\zeta$  differs in regulation as  $\omega$  is the transcriptional repressor and  $\epsilon$  the toxin inhibitor) (Camacho, *et al.*, 2002; Meinhart, *et al.*, 2003; Zielenkiewicz & Ceglowski, 2005). Antitoxins bind through DNA binding motifs in their N-termini to palindromic or repeat regions in the promoter DNA of the operon to repress transcription (Aizenman, *et al.*, 1996; Garcia-Pino, *et al.*, 2010; Gotfredsen & Gerdes, 1998; Ruiz-Echevarria, Gimenez-Gallego, Sabariego-Jareno, & Diaz-Orejas, 1991; Schumacher, *et al.*, 2009; Van Melderen, *et al.*, 1996). In most cases the toxin acts as a co-repressor, strengthening the interaction between the antitoxin and DNA (Christensen, *et al.*, 2001; Garcia-Pino, *et al.*, 2008; Kamada, Hanaoka, & Burley, 2003; Monti, *et al.*, 2007; Zhang, Zhang, & Inouye, 2003). The ParDE TA system is the exception to this, the antitoxin ParD is sufficient to repress transcription of the *parDE* operon (Davis, Heliniski, & Roberts, 1992; Eberl, Givskov, & Schwab, 1992)

#### **1.1.9.1 Stoichiometry of Toxin-Antitoxin Complexes**

Toxin-antitoxin proteins form complexes of varying stoichiometry. Antitoxin proteins have dual functions (1) to bind to the toxin to directly or indirectly inhibit its activity and (2) to bind to palindromic or inverted repeat sequences in the TA operon promoter DNA to auto-regulate its transcription. Binding is enhanced by the toxin therefore the ratio of toxin to antitoxin can regulate TA operon transcription.

### ***Toxin-Antitoxin Complexes of Varying Stoichiometry***

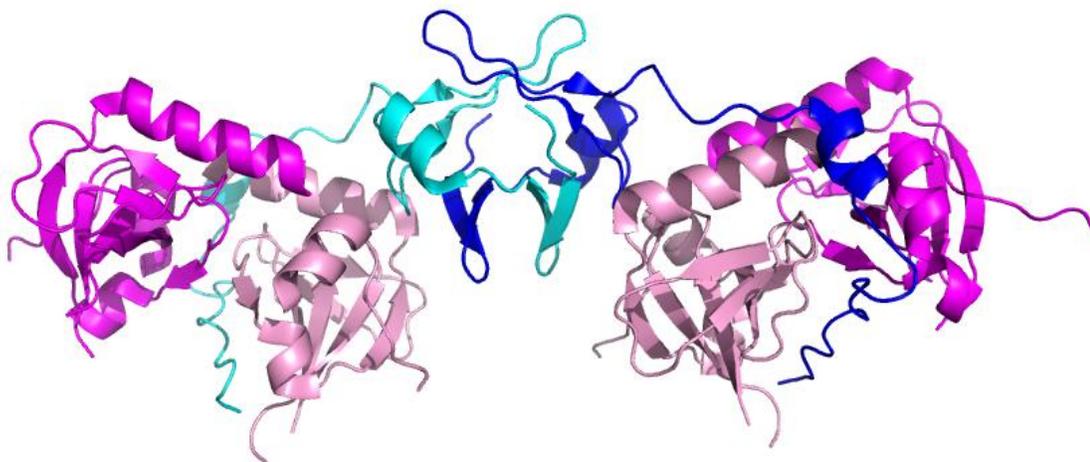
CcdA and CcdB form complexes of different stoichiometry depending on the ratio of each protein. CcdB can activate or inhibit DNA binding of CcdA by the stoichiometry of the CcdAB complex (Van Melderen, *et al.*, 1996). Therefore, the level of CcdB in the cell can regulate expression of the *ccd* operon (Afif, Allali, Couturier, & Van Meldern, 2001). A similar situation is seen with Kid-Kis; the Kid<sub>2</sub>Kis<sub>2</sub>Kid<sub>2</sub> heterohexamer is the most abundant species when Kid is in excess to Kis and weakly interacts with the *parD* operon, but at higher concentrations of Kis a Kid<sub>2</sub>Kis<sub>2</sub> tetramer (and up to heterodecamers) form that strongly interacts with the *parD* operon (Monti, *et al.*, 2007). The stoichiometry of the Phd-Doc TA family also activates or inhibits transcription of the operon. At toxin-antitoxin ratios below one the Phd<sub>2</sub>Doc complex represses transcription but at higher toxin-antitoxin ratios derepression occurs resulting in Phd.Doc<sub>2</sub> that inefficiently represses the operon (Garcia-Pino, *et al.*, 2010; Gazit & Sauer, 1999; Lehnerr & Yarmolinsky, 1995).

RelB recognises and binds to four hexad repeats in the palindromic *relBE* operator, (termed *relO*), by a RHH motif in the N-terminus of the protein (Overgaard, Borch, & Gerdes, 2009). RelE promotes de-repression of *relBE* transcription, as it is the RelB:RelE ratio that controls the transcription rate not the level of RelE (Overgaard, Borch, Jorgensen, & Gerdes, 2008). The repressor complex consists of two RelB<sub>2</sub>.RelE heterotrimers bound to *relO*. Binding of RelB<sub>2</sub> alone is very weak and RelB:RelE ratios lower than 2:1 do not enhance binding of RelB<sub>2</sub> to *relO*. When RelE is in excess it interacts with the TA-operator complex via the unoccupied antitoxin C-termini binding sites and destabilises the TA-operator complex (Overgaard, *et al.*, 2009). This model assumes that RelE has both a high and low affinity binding site for RelB. It is assumed that there is a gradual disruption of the repressor complex with an increasing concentration of RelE, a pseudo-titration effect where one molecule of RelB<sub>2</sub>.RelE is released per RelE molecule (Overgaard, *et al.*, 2008). This theory does contrast with the crystal structure of an archael RelBE complex which shows a RelB<sub>2</sub>.RelE<sub>2</sub> stoichiometry (Francuski & Saenger, 2009). Further studies have shown that the RelB-RelE complex is a (RelB.RelE)<sub>2</sub> heterotetramer but additional studies

also support a 2:1 stoichiometry of RelB:RelE. Gerdes explains this: “TA stoichiometries are highly dynamic and the interconversion between heterotrimer and heterotetramer species allows for increased capacity of the antitoxin to neutralise its cognate toxin and also provides a mechanism for TA ratio-dependent derepression of their cognate promoters” (Overgaard, *et al.*, 2009).

### ***Hexameric Toxin-Antitoxin Complexes***

The structure of MazEF from *E. coli* revealed a 2:4 heterohexamer composed of three alternating toxin and antitoxin homodimers i.e. MazF<sub>2</sub>-MazE<sub>2</sub>-MazF<sub>2</sub> (Kamada, *et al.*, 2003). The intrinsically unstructured C-terminus of MazE wraps around the MazF homodimer and crosses the edge of the MazF dimer interface (Figure 1.8). Each MazE monomer recognizes one MazF homodimer, via interactions with surface regions on the MazF homodimer (Figure 1.8).



**Figure 1.8. Ribbon Diagram of MazEF from *E. coli*.** One MazE homodimer (light blue and dark blue) binds to two MazF homodimers (magenta and pale pink). Viewed perpendicular to the two fold crystallographic axis, relating each half of the MazE homodimer to the other. Interactions between MazE and MazF are through the C-terminus of MazE which wraps around the MazF homodimer, crossing the dimer interface. PDB # 1UB4.

### ***Tetrameric Toxin-Antitoxin Complexes***

It is possible tetrameric TA complexes do have varying stoichiometry, or that they only form tetramers to regulate transcription of the TA operon. Two HipA molecules sandwich two HipB molecules bound to its promoter DNA to form a tetramer (Schumacher, *et al.*, 2009). HipB does not bind in the HipA active site to inhibit its

activity; it prevents a conformational change in HipA by locking it into an open conformation by interacting with both the C and N-terminal domains of HipA (Schumacher, *et al.*, 2009). RelB and RelE from the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3 form a heterotetramer (RelB-RelE)<sub>2</sub> (Shinohara, *et al.*, 2010), whereas RelBE from *E. coli* forms complexes of varying stoichiometry (Overgaard, *et al.*, 2008). It is possible that *P. horikoshii* RelBE also forms complexes of varying stoichiometry at different concentrations of the two proteins. *P. horikoshii* RelB inhibits RelE by steric hindrance, preventing RelE entering the ribosome (Shinohara, *et al.*, 2010). DinJ-YafQ from *E. coli* also form a tetramer of two DinJ and two YafQ protein molecules but how DinJ inhibits YafQ activity is unknown (Motiejunaite, Armalyte, Markuckas, & Suziedeliene, 2007).

Other tetrameric TA systems include the MqsRA complex (Brown, *et al.*, 2009) and ParDE (Johnson, Strom, & Heliniski, 1996). MqsA does not bind into the active site to inhibit the toxic activity of MqsR but binds to the *mqsRA* promoter and three other promoters to regulate their transcription (Brown, *et al.*, 2009). In contrast ParD alone is sufficient for repression of the *parDE* operon (Davis, *et al.*, 1992; Eberl, *et al.*, 1992).

### ***Trimeric Toxin-Antitoxin Complexes***

YefM-YoeB shows different stoichiometry to RelBE, even though it belongs to the RelBE family. YefM and YoeB enhance transcription by forming a heterotrimer complex (YefM<sub>2</sub>-YoeB) (Cherny, Rockah, & Gazit, 2005; Kamada & Hanaoka, 2005). The C-terminus of YefM is unstructured in one YefM homodimer, whereas the C-terminus in the other homodimer adopts a helical conformation which conceals the ribonuclease fold of YoeB and inhibits its activity (Kedzierska, Lian, & Hayes, 2007).

## **1.2 The VapBC Toxin-Antitoxin Family**

The VapBC (virulence associated protein) TA family are defined by their toxic components, which belong to the PilT N-terminal domain (PIN domain) family of

---

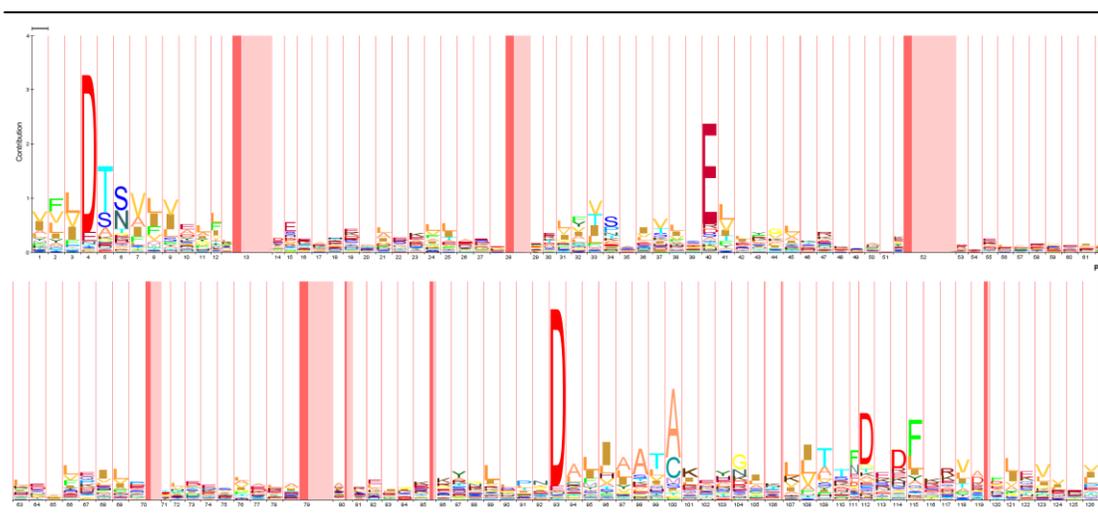
proteins. The *vapB* gene is found upstream of *vapC* in an operon and encodes a DNA binding protein that also binds to VapC and inhibits its nucleolytic activity. The VapBC complex is auto-regulatory and binds to repeat sequences in its promoter DNA through VapB (Robson, McKenzie, Cursons, Cook, & Arcus, 2009)\*. As for many other TA systems the DNA binding affinity of VapB is strengthened by VapC. VapC proteins act to inhibit translation (Robson, *et al.*, 2009)\*, presumably by their ability to cleave mRNA. The VapBC family is the largest family of TA systems and are found across a range of bacteria and archaea. *M. tuberculosis* has numerous VapBC homologues, more than any other prokaryote (Arcus, *et al.*, 2005; Makarova, Wolf, & Koonin, 2009; Pandey & Gerdes, 2005).

### 1.2.1 PIN-Domain Proteins

The Pfam database (Finn, *et al.*, 2010) lists 3457 proteins belonging to the PIN-domain family (PF01850) from 490 different species including eukaryotes, bacteria and archaea. In eukaryotes PIN domain containing proteins have been associated with RNAi and nonsense-mediated RNA degradation (Clissold & Ponting, 2000; Fatica, Tollervey, & Dlakic, 2004; Lamanna & Karbstein, 2009) by virtue of their ribonuclease activity (Anantharaman & Aravind, 2003). PIN-domains are found in nearly half of all sequenced prokaryotes and are generally associated with VapBC TA operons, by virtue of their “toxic” ribonuclease activity (Arcus, *et al.*, 2005). The low overall sequence conservation for this large family is evident in the Hidden Markov Model (HMM) in Figure 1.9 with small stacks of letters at a majority of positions signifying accommodation of a range of amino acids along most of the sequence (Figure 1.9). The PIN domain family is characterised by a triad of acidic residues at positions 4, 40 and 93, and a fourth less well conserved acidic residue at position 112 (Figure 1.9), that are hypothesised to constitute the active site. Another feature for the family is the presence of a polar residue (Thr, Ser or Asn) at position  $i+1$  or  $i+2$  following the first conserved aspartic acid (positions 5 and 6 in Figure 1.9).

\* The purification of the VapBC complex reported in this thesis was used by collaborators in this DNA binding assay and translation inhibition experiments reported in this thesis are published in this paper. Two published papers (Bunker *et al.* 2008 and Robson *et al.* 2009) reporting on the results in this thesis can be found in Appendix D.

The catalytic mechanism of PIN domain proteins is currently unknown. In the related FLAP nucleases the acidic residues coordinate a metal ion ( $Mg^{2+}$ ) and in T4 RNase H, the  $Mg^{2+}$  ion is hypothesised to activate a water molecule leading to nucleophilic attack of the phosphodiester bond. The structural homology of PIN domain proteins to members of the 5'-3' nuclease family including T4 RNase H, *Taq* polymerase and the FLAP endonucleases (Arcus, *et al.*, 2005) strengthens the hypothesis that they have nuclease activity and are involved in RNA degradation.



**Figure 1.9. The Hidden Markov Model which defines the PIN-domain family of proteins.** The height of each letter is proportional to the amount of “information” it provides about the respective position in the PIN-domain family. The width of each column is also an indication of the importance of this position in defining the family. Sequence regions that contain insertions (and therefore almost no information about the family) are shown as dark and light pink columns. Taken from PFM Finn *et al.* (2010).

## 1.2.2 Characterised VapBC Toxin-Antitoxin Systems

### *NtrPR* from *Sinorhizobium meliloti*

The *ntrPR* operon of *Sinorhizobium meliloti* (a symbiotic soil bacterium), is a member of the VapBC TA family. The *ntrP* gene overlaps *ntrR* and the operon is negatively regulated by binding of the NtrPR complex (Bodogai, *et al.*, 2006). The N-terminus of NtrP constitutes the DNA binding domain and the C-terminus has conserved hydrophobic residues (Hp box) that are suggested to interact with NtrR. NtrP binds to direct repeat sequences in the *ntrPR* promoter DNA and the second

repeat overlaps the transcriptional start site thus the repressor activity of NtrPR inhibits the binding of RNA polymerase (Bodogai, *et al.*, 2006).

Expression of NtrR in *E. coli* leads to a reduction in colony formation and cell growth, although the inhibition seen is not as great as RelE or MazF (Bodogai, *et al.*, 2006). The alarmone, ppGpp has no direct regulatory role in controlling NtrPR expression and although the cellular target of NtrR is unknown, indications are that it inhibits translation (Bodogai, *et al.*, 2006). Although it was originally thought that NtrPR regulated nitrogen fixation genes (Olah, *et al.*, 2001), further investigation revealed that it regulates not only nitrogen fixation genes but others involved in metabolic processes (Puskas, *et al.*, 2004). Therefore, it is hypothesised that NtrPR adjusts metabolic processes under symbiosis and other stress conditions (Puskas, *et al.*, 2004).

#### ***VapBCs from Haemophilus influenzae, Sulfolobus solfataricus, Leptospira interrogans and Enterobacteria***

Non-typeable *Haemophilus influenzae* (NTHi) is a pathogenic bacteria that enters a bacteriostatic state in the middle ear during an otitis media infection (Daines, *et al.*, 2007). The NTHi genome has two *vapBC* operons, *vapBC-1* and *vapBC-2*. Expression of the *vapBC-1* operon displays an inverse relationship to culture density (Daines, *et al.*, 2007). VapC-1 is a ribonuclease that cleaves free RNA *in vitro* independently of the ribosome and does not degrade double-stranded or single-stranded DNA (Daines, *et al.*, 2007). VapB-1 forms homodimers and is hypothesised to form multimers with VapC-1, although this has not been confirmed. VapBC-2 has not been characterised.

The *Sulfolobus solfataricus* genome has at least 26 *vapBC* loci. Thermal stress has been shown to trigger high transcription levels of the *vapBC-22* operon. A mutation in the VapC-22 toxin resulted in no obvious phenotype but approximately 100 open reading frames were differentially transcribed two fold or more (Cooper, Daugherty,

Tachdijan, Blum, & Kelly, 2009). It was therefore hypothesised that disruption of the VapBC complex leads to an increased susceptibility to thermal stress.

A *vapBC* locus, as well as the *chpIK* and *mazEF* loci, is present on the chromosome of the pathogen *Leptospira interrogans* (Zhang, *et al.*, 2004). Expression of VapC from *L. interrogans* in *E. coli* results in inhibition of growth which is counteracted by expression of its cognate antitoxin VapB. The presence of *vapBC* on an unstable plasmid prevents plasmid loss indicating that the VapBC proteins act as a typical TA system. VapB binds to two inverted repeat sequences in its promoter to auto-regulate its transcription (Zhang, *et al.*, 2004). The cellular target of VapBC from *L. interrogans* is currently unknown.

VapC toxins from the *vapBC* loci from *Salmonella* LT2 (*vapBC<sub>LT2</sub>*) and the *Shigella* plasmid pMYSH6000 (*vapBC<sub>pMYSH</sub>*) inhibit growth in *E. coli* by causing a bacteriostatic state, that is counteracted by expression of their cognate VapB antitoxins. Both VapC proteins act to inhibit translation and transcription and both *vapBC* loci are induced by amino acid starvation and chloramphenicol. Winther & Gerdes show that induction of VapC<sub>LT2</sub> and VapC<sub>pMYSH</sub> induces mRNA cleavage at stop codons between the second and third codon bases via activation of YoeB (Winther & Gerdes, 2009). The authors' findings are unusual as other VapC proteins have been shown to have nuclease activity directly (Arcus, Backbro, Roos, Daniel, & Baker, 2004; Daines, *et al.*, 2007; Miallau, *et al.*, 2009).

### ***FitAB* from *Neisseria gonorrhoeae***

The *fitAB* (fast intracellular trafficking) locus is a VapBC TA system present on the chromosome of the sexually transmitted pathogen *Neisseria gonorrhoeae* (GC, gonococcus). The *fitAB* locus was first identified as mutations in this locus resulted in GC traversing epithelial cells at a much faster rate than the wild-type and also resulted in a faster replication rate of the bacterium within epithelial cells (Wilbur, *et al.*, 2005). GC can be asymptomatic but still culturable and transmissible thereby acting as a carrier for the disease.

FitAB is hypothesised to slow intracellular trafficking and replication of GC (Mattison, Wilbur, So, & Brennan, 2006). The model that Mattison *et al.* (2006) proposes is as follows: FitAB binds to its promoter DNA when GC is in an extracellular environment; this sequesters the FitAB complex and represses transcription of the *fitAB* locus. Upon invasion into epithelial cells FitAB is released from the DNA and dissociates to slow GC replication by an unknown mechanism.

### ***FitA DNA Binding***

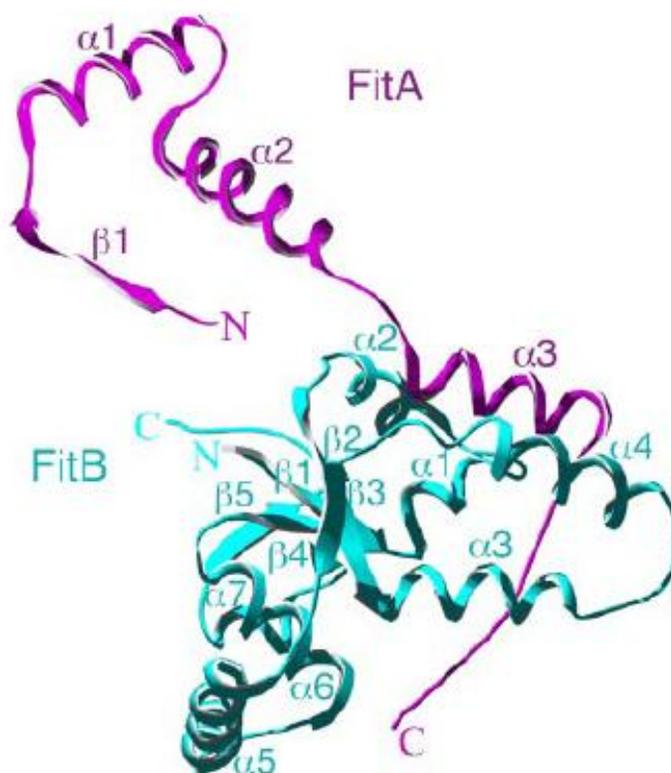
FitA forms a homodimer in the absence of FitB (Wilbur, *et al.*, 2005) and binds to repeat sequences in its promoter DNA, which covers the putative -10 promoter sequence for the *fitAB* locus. FitA binds to this sequence with relatively weak affinity but this is improved 38 fold when bound to FitB (Wilbur, *et al.*, 2005). The FitAB complex will bind to DNA fragments with either half of the inverted repeat scrambled but not when both are scrambled. (Mattison, *et al.*, 2006).

An arginine at position seven is highly conserved among FitA homologues (Wilbur, *et al.*, 2005). Mutation of this arginine to alanine in FitA (FitAR7A) did not destabilise the protein but did result in an eight fold decrease in the FitA DNA binding affinity, indicating that this residue is important for the interaction with DNA (Wilbur, *et al.*, 2005). This mutation did not affect the ability of FitA to interact with FitB.

Interestingly the FitA binding sequence TGCTATCA occurs 14 times in the GC genome and the palindromic sequence TGATATCA occurs four times. These sequences mainly occur in intergenic regions or ORFs, only two are found in a promoter region and this is upstream of *fitAB* (Wilbur, *et al.*, 2005).

### ***Structure of FitAB bound to DNA***

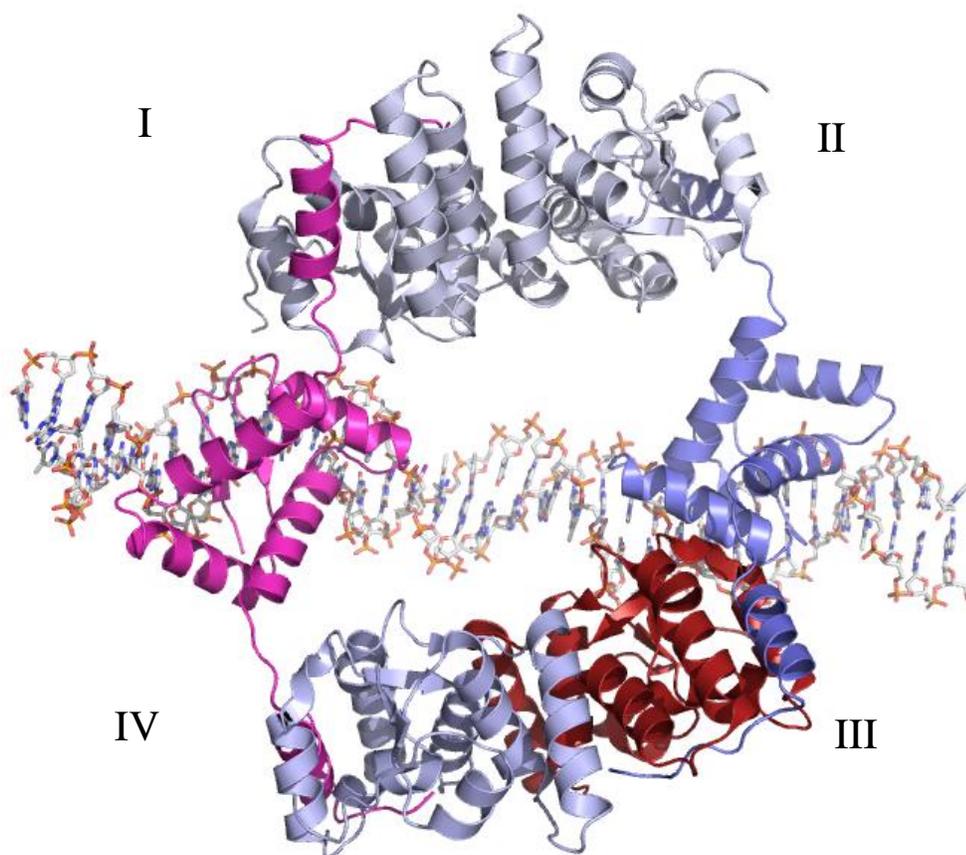
The structure of FitAB bound to a DNA fragment was solved in 2006 by Mattison *et al.* The FitAB complex when not bound to DNA did not crystallise; upon binding the FitAB complex to IR36 DNA (IR36 is a dsDNA fragment including four base pairs either side of the inverted repeat sequences) it crystallised. The first 45 residues of FitA are homologous to RHH DNA binding proteins and the structure confirms this (Figure 1.10). FitB is an  $\alpha/\beta/\alpha$  fold that consists of a central five-stranded parallel  $\beta$  sheet with four  $\alpha$  helices on one side and three  $\alpha$  helices on the other (Figure 1.10). The FitA and FitB proteins associate tightly to form heterodimers. The heterodimerisation interface is formed by contacts between  $\alpha 3$  in the extended C-terminal region of FitA with helices  $\alpha 1$ ,  $\alpha 2$  and  $\alpha 4$  of FitB (Figure 1.10). FitA fills a large hydrophobic groove on FitB.



**Figure 1.10. Ribbon diagram of FitAB heterodimers.** FitA in magenta and FitB in cyan. N and C-termini are labelled along with  $\alpha$  helices and  $\beta$  sheets for both proteins. Figure from Mattison *et al.* (2006) PDB # 2H10

Four FitAB heterodimers associate to form a tetramer of FitAB heterodimers bound to DNA (Figure 1.11). FitAB heterodimers I and IV are tightly associated by FitA dimerisation and bind to one of the inverted repeat sequences, and the same for FitAB heterodimers II and III. FitB dimerisation then generates the hetero-octamer.

The  $\beta$  strands of each FitA monomer form two antiparallel  $\beta$  sheets that are involved in recognition of the inverted repeat sequences on the DNA. All contacts between FitA and DNA are mediated by the  $\beta$  sheet (Figure 1.11). The highly conserved Arg<sup>7</sup> of FitA is one of the main contacts between FitA with DNA (Mattison, *et al.*, 2006).



**Figure 1.11. Ribbon Diagram of FitAB bound to the IR36 DNA fragment.** FitA dimers are coloured in magenta and blue. Fit B proteins in silver, light blue and maroon. Maroon FitB demonstrates the FitB dimer interface. Four FitAB heterodimers are labelled I to IV. From PDB file 2H10.

The FitB homodimer forms by interactions between  $\alpha 3$  of one monomer with  $\alpha 5$  from the other monomer. FitB shows structural similarity to PAE2754, a PIN domain from

*Pyrobaculum aerophilum*. In FitB the four conserved residues across PIN domain proteins, (Asp<sup>5</sup>, Glu<sup>42</sup>, Asp<sup>104</sup> and Asp<sup>122</sup>) are at the C-terminal ends of the  $\beta$  strands and form a negatively charged pocket near the centre of the FitB molecule. No nucleic acid cleavage of ssDNA, dsDNA, ssRNA or dsDNA by the FitAB complex was observed (Mattison, *et al.*, 2006). This is because Arg<sup>68</sup>, present on the C-terminus of FitA, sticks into the acidic pocket that constitutes the active site of FitB where it interacts with the carboxyl groups of Asp<sup>5</sup>, Glu<sup>42</sup> and Asp<sup>104</sup> of FitB and forms strong electrostatic interactions that are not easily displaced. FitB may bind Mg<sup>2+</sup> when FitA is not present as has been shown for the PIN domain PAE2754 (Arcus, *et al.*, 2004; Mattison, *et al.*, 2006).

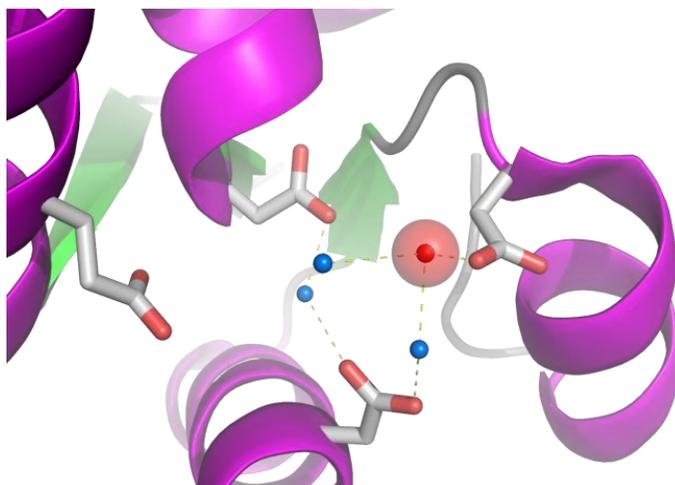
The cellular target and the role of FitAB in GC virulence are currently unknown.

#### ***VapBC from Pyrobaculum aerophilum***

Like the *M. tuberculosis* genome the number of *vapBC* loci in the hyperthermophilic crenarchaeon, *P. aerophilum* genome is expanded. *P. aerophilum* contains 12 *vapBC* loci and two of these have been studied as thermophilic homologues of the VapBC proteins from *M. tuberculosis* (Bunker, McKenzie, Baker, & Arcus, 2008).

The *pae2754* and *pae0151* genes encode PIN-domain proteins found downstream of *pae2755* and *pae0152* respectively, which are related to the MetJ/Arc DNA binding protein family that contain a RHH motif (Bunker, 2005). VapC<sub>PAE0151</sub> and VapB<sub>PAE0152</sub> form a complex with a stoichiometry similar to that of FitAB (Mattison, *et al.*, 2006); a heterotetramer, two VapC<sub>PAE0151</sub> and two VapB<sub>PAE0152</sub> molecules (Bunker, *et al.*, 2008).

In VapC<sub>PAE0151</sub> a manganese ion is coordinated to Asp<sup>118</sup> and bonded via water to Asp<sup>100</sup>; two other water molecules are also present (Figure 1.12). The manganese however is only partially occupied and has an incomplete coordination sphere suggesting other factors may have a role in metal ion coordination (Bunker, *et al.*, 2008).



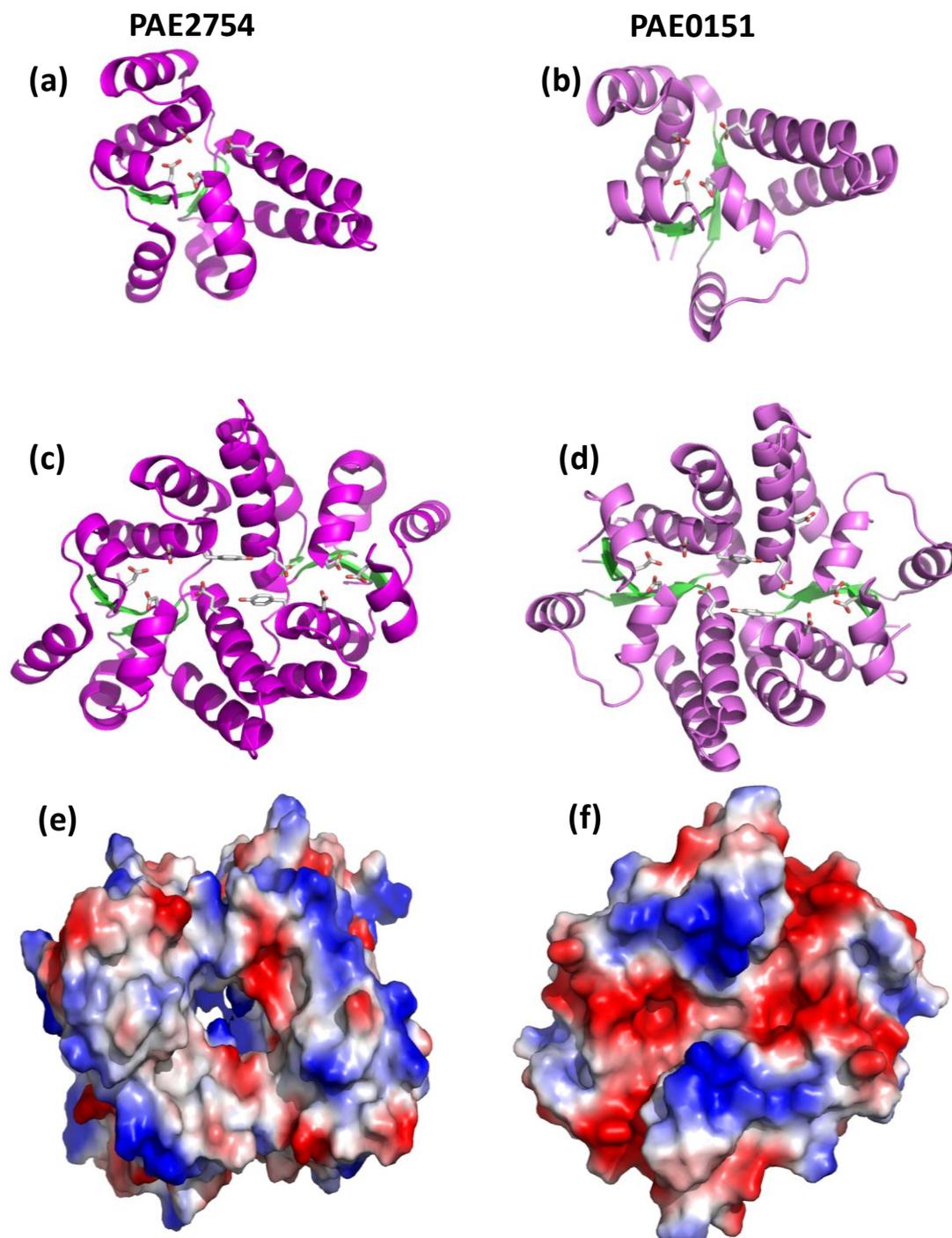
**Figure 1.12. View of the four conserved acidic residues hypothesised to be the active site of VapC<sub>P<sub>AE0151</sub></sub>**. Manganese ion shown in red, coordinated directly to Asp<sup>118</sup> and coordinated via a water molecule (blue sphere) to Asp<sup>100</sup>. Two other water molecules are also coordinated with the metal ion. From PDB file 2FE1.

The VapC<sub>P<sub>AE2754</sub> (Figure 1.13a) and VapC<sub>P<sub>AE0151</sub> (Figure 1.13b) monomers form an  $\alpha/\beta/\alpha$  stack with a central twisted parallel  $\beta$  sheet of five short strands (Arcus, *et al.*, 2004). VapC<sub>P<sub>AE2754</sub> is a tetramer, (a dimer of dimers) as seen in Figure 1.13e below, whereas VapC<sub>P<sub>AE0151</sub> forms a dimer (Figure 1.13d and f). The dimer interface for both proteins is an extensive region dominated by hydrophobic interactions. The tetramer interface of VapC<sub>P<sub>AE2754</sub> is small where the C-terminus of helix  $\alpha$ 2 interacts with the N-terminus of helix  $\alpha$ 6 from another monomer.</sub></sub></sub></sub></sub>

The four conserved acidic residues characteristic of PIN domains Asp<sup>8</sup>, Glu<sup>38</sup>, Asp<sup>92</sup> and Asp<sup>110</sup> in VapC<sub>P<sub>AE2754</sub> and Asp<sup>6</sup>, Glu<sup>42</sup>, Asp<sup>100</sup> and Asp<sup>118</sup> in VapC<sub>P<sub>AE0151</sub> are at the centre of the protein molecule (Figure 1.13a and b). The four acidic residues form a negatively charged pocket at the C-terminal end of the  $\beta$  sheet and N-termini of helices  $\alpha$ 2 and  $\alpha$ 6 in the middle of each protein molecule (Figure 1.13a and b).</sub></sub>

In both VapC<sub>P<sub>AE2754</sub> and VapC<sub>P<sub>AE0151</sub> dimers the two active site pockets are separated by a structure formed by a conserved tyrosine residue (Figure 1.13c and d). The two tyrosine residues lie opposite each other with their aromatic rings in parallel as in Figure 1.13c and d. VapC<sub>P<sub>AE2754</sub> tetramer formation results in a tunnel in which the four active site pockets lie in the interior with restricted access to the openings in the tetramer (Figure 1.13e).</sub></sub></sub>

Although structural comparisons revealed the fold of VapC<sub>PAE2754</sub> to be closely related to protein domains that bind ADP or FAD, the four conserved acidic residues in the active site of the protein show similarity to the T4 RNase H family of exonucleases (Arcus, *et al.*, 2004). Exonuclease activity tests of VapC<sub>PAE2754</sub> against a synthetic DNA oligonucleotide with a large 5'-3' overhang revealed exonuclease activity in the presence of Mg<sup>2+</sup> and Mn<sup>2+</sup>, although the activity of the enzyme was extremely slow showing degradation of the DNA over 19 hours (Arcus, *et al.*, 2004), suggesting this is not an optimal substrate for VapC<sub>PAE2754</sub>.



**Figure 1.13** VapC<sub>PAE0151</sub> and VapC<sub>PAE2754</sub> in their various oligomeric states. Ribbon diagram of (a) the VapC<sub>PAE2754</sub> and (b) the VapC<sub>PAE0151</sub> monomers. Central parallel  $\beta$  sheets shown in green and  $\alpha$  helices in magenta for VapC<sub>PAE2754</sub> and violet for VapC<sub>PAE0151</sub>. The four conserved acidic residues thought to constitute the active site are shown as sticks. Ribbon diagrams of (c) the VapC<sub>PAE2754</sub> and (d) the VapC<sub>PAE0151</sub> dimer, conserved acidic residues are shown as sticks as well as the conserved tyrosine residue from each monomer that sits between the active sites. (e) Surface diagram of the VapC<sub>PAE2754</sub> tetramer with electrostatic charges (red, negative and blue positive). (f) Surface diagram of the VapC<sub>PAE0151</sub> dimer with electrostatic charges (red, negative and blue positive). Note the two negatively charged active site pockets in the middle of each monomer.

### 1.3 Toxin-Antitoxin Systems in the *Mycobacteria* species

The *Mycobacterium tuberculosis* complex (MTBC) includes other bacteria that cause TB, namely *Mycobacterium bovis*, *Mycobacterium africanum*, *Mycobacterium canetti* and *Mycobacterium microti*. Ramage *et al.* hypothesised that the majority of TA systems in the MTBC complex would be present in the progenitor to the MTBC, or alternatively, TA systems were acquired independently in each bacterial species (Ramage, *et al.*, 2009). How the *M. tuberculosis* genome evolved to have so many TA systems and maintain these systems is unknown, but is a subject of much interest.

#### 1.3.1 Toxin-Antitoxin Systems Present in *Mycobacterium tuberculosis*

Recently Ramage *et al.* (2009) undertook a comprehensive analysis of TA systems from *M. tuberculosis*, and identified 88 putative TA systems on the *M. tuberculosis* chromosome. Of these putative TA systems 37% were located in HGT regions, strengthening the hypothesis that TA systems in the MTBC were acquired by HGT. Thirty gene pairs were identified to function as TA systems and of these, 20 belonged to the VapBC (PIN domain) TA family, four to the MazEF TA family, two to the RelBE TA family, one to the HigBA TA family and three belonged to a new family of TA systems.

The *Rv0909-Rv0910* operon was the only TA system found in all genomes analysed in the study and showed no homology to other TA families, indicating it is a new TA system (Ramage, *et al.*, 2009). The toxin Rv0910 inhibits growth which is relieved upon expression of Rv0909. It does not inhibit translation and displays no ribonuclease activity *in vivo* suggesting this novel TA family has a different mechanism of inhibition to that of the VapBC TA family. Thirty of the 88 putative TA systems identified are functional. Four of these systems are involved in the stress response to hypoxia and macrophage infection (Ramage, *et al.*, 2009).

#### ***The Mycobacterium tuberculosis HigBA Toxin-Antitoxin System***

A functional *higBA* locus has been demonstrated on the chromosome of *M. tuberculosis*. Expression of HigB inhibits growth in *E. coli* that is relieved upon

---

expression of HigA (Gupta, 2009). The N-terminus of HigA is predicted to contain a HTH DNA binding domain and binds to inverted repeat sequences in the *higBA* promoter region (Fivian-Hughes & Davis, 2010).

### ***The Mycobacterium tuberculosis RelBE Toxin-Antitoxin System***

Based on their homology to the *E. coli* RelBE TA system, three Rel TA modules (*relBE*, *relFG* and *relJK*) were identified on the chromosome of *M. tuberculosis* (Korch, Contreras, & Clark-Curtiss, 2009).

RelJ represses transcription but RelB and RelF act as transcriptional activators. RelE and RelG toxins themselves repress transcription in the presence of their cognate antitoxins (Korch, *et al.*, 2009). Korch *et al.* (2009) hypothesise that RelE and RelG may (1) act as transcriptional repressors when both toxin and antitoxin are present in the cell and (2) interfere with the antitoxin binding to the promoter or (3) titrate the antitoxin away from the promoter. Based on these hypotheses it is suggested that “...the alternative *M. tuberculosis rel* module regulation patterns suggest differential needs for Rel proteins during the bacillus’ unique pathogenic lifestyle.” (Korch, *et al.*, 2009)

All three Rel toxin proteins inhibit mycobacterial growth and can be reversed by expression of the cognate antitoxin but the extent of growth inhibition varied between Rel toxin proteins (Korch, *et al.*, 2009). The *relE*, *relF* and *relK* genes are expressed in *M. tuberculosis* during infection of human macrophages at 110 hours post-infection, suggesting Rel proteins are important later in the infection cascade (Korch, *et al.*, 2009). The mechanism of Rel toxin proteins in growth regulation of *M. tuberculosis* is yet to be determined, but Korch *et al.* suggest that they are required for survival within human macrophages and they function as modulators of programmed cell stasis (PCS) rather than PCD and are therefore involved in the persistence of *M. tuberculosis* (Korch, *et al.*, 2009).

### ***The Mycobacterium tuberculosis MazEF Toxin-Antitoxin System***

The *M. tuberculosis* genome contains nine *mazEF* TA systems (Pandey & Gerdes, 2005). Genes encoding MazF were identified on the *M. tuberculosis* chromosome by homology searches using *mazF* from *E. coli*. Homologues of MazF from *E. coli* were examined for endoribonuclease activity both *in vivo* and *in vitro*. The recognition sequences of these proteins are outlined in Section 1.1.8. Bioinformatic analysis revealed the target pentad sequences of MazF-mt3 and MazF-mt7 is under-represented in PE and PPE genes in *M. tuberculosis* (Zhu, *et al.*, 2008); although the functions of these genes is unknown at present, some of these proteins have been shown to exhibit an immune response by localising to the mycobacterial cell wall, playing a part of the host-pathogen interactions.

### ***The VapBC Toxin-Antitoxin System***

TA systems are unusually abundant in the genome of *M. tuberculosis*. Of the approximately 88 TA systems on the *M. tuberculosis* genome 45 of these are VapBC systems, more than any other intracellular pathogen (Pandey & Gerdes, 2005). Each VapB protein will only inhibit its cognate toxin (VapC), no cross talk was observed between VapB antitoxins and VapC toxins (Ramage, *et al.*, 2009). VapC proteins Rv0301, Rv1561, Rv2829c, Rv3408 all act to inhibit translation and this inhibition is relieved by expression of their cognate antitoxin. Rv0301 and Rv1561 proteins demonstrate ribonuclease activity against MS2 RNA, with Rv1561 exhibiting more ribonuclease activity than Rv0301 (Ramage, *et al.*, 2009). Addition of VapB to VapC inhibited the ribonuclease RNase activity.

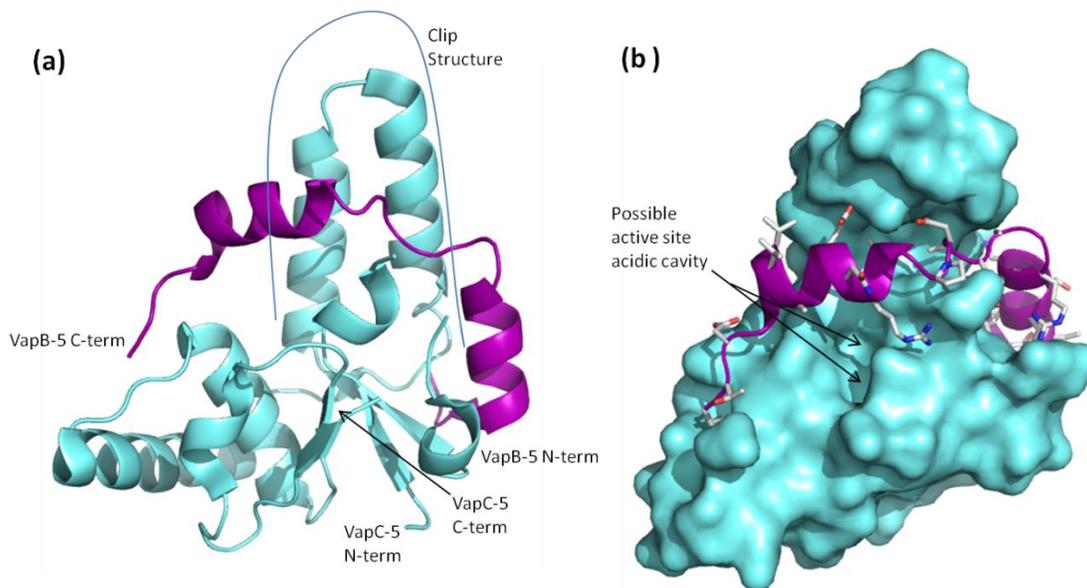
Ribonuclease activity of VapC-5 (Rv0627) was demonstrated against a general RNA substrate of unknown sequence (Miallau, *et al.*, 2009). This activity was Mg<sup>2+</sup> dependent and no ribonuclease activity of VapC-5 was observed with other divalent cations such as Mn<sup>2+</sup> and Zn<sup>2+</sup> as co-factors (Miallau, *et al.*, 2009). The ribonuclease activity of VapC-5 was weak, especially compared to that of VapC protein Rv1561 (Ramage, *et al.*, 2009), which is likely due to the majority of the protein being in complex with VapB-5. This is contrary to the fact that no ribonuclease activity is

---

observed for other VapC proteins in complex with VapB (Mattison, *et al.*, 2006; Ramage, *et al.*, 2009). VapC-5 displayed no degradation of dsDNA, consistent with VapC from *H. influenzae* (Daines, *et al.*, 2007).

The response of TA systems to hypoxic conditions both in culture and during infection of macrophages was measured indirectly. An increase in transcription of the TA operon was used as a read-out of TA system activation (Ramage, *et al.*, 2009). The HigBA homologue *Rv1955-Rv1956* and the VapBC homologue *Rv2009-Rv2010*, were shown to be induced during hypoxia and the VapBC homologues *Rv1560-1561* and *Rv0549-Rv0550* were induced during macrophage infection (Ramage, *et al.*, 2009), suggesting a role for these TA systems in the stress response.

The VapBC-5 TA structure is consistent with other PIN domain proteins. VapC-5 forms an  $\alpha/\beta/\alpha$  domain with a compact core domain of a five-stranded parallel  $\beta$  sheet surrounded by five  $\alpha$  helices (Figure 1.14). A clip structure of two  $\alpha$  helices protrudes from the core domain. The N-terminal proposed DNA binding region of VapB-5 is missing from the structure. In this case VapBC-5 was treated with a protease prior to purification and crystallisation. VapB-5 consists of two  $\alpha$  helices joined by a long, possibly flexible loop (Figure 1.14). The C-terminus of VapB-5 binds tightly to VapC-5 in a wide hydrophobic groove on its surface between the core domain and the clip structure as in Figure 1.14. VapBC-5 forms a dimer in the crystal by interactions between VapC-5 proteins.



**Figure 1.14. Structure of VapBC-5 from *M. tuberculosis*.** (a) Ribbon diagram of VapB (purple) bound to VapC (cyan). N and C termini of both proteins are labelled, as well as the clip structure of VapC. (b) Surface diagram of VapC and ribbon diagram of VapB with amino acid side chains shown as sticks, to illustrate interactions with VapC. The acidic cavity that shelters the active site residues and possibly binds  $Mg^{2+}$  is indicated by black arrows. From PDB file 3DB0.

VapC-5 has four conserved acidic residues (Asp<sup>26</sup>, Glu<sup>57</sup>, Asp<sup>115</sup> and Asp<sup>135</sup>) which are characteristic of PIN domains and are proposed to constitute the active site. Ribonuclease activity assays of VapC-5 have shown  $Mg^{2+}$  ribonuclease activity (Miallau, *et al.*, 2009), consistent with the proposed active site residues. Although there is low sequence similarity (around 24%) there is strong structural similarity between VapC-5 and VapC<sub>PAE2754</sub> from *P. aerophilum* and FitB from *N. gonorrhoeae*. However, the FitA structure differs from that of VapB-5 (Miallau, *et al.*, 2009).

VapB-5 is proposed to inhibit the ribonuclease activity of VapC-5 by causing a direct or indirect change in the conformation of the active site residues, resulting in a catalytically unfavourable conformation (Miallau, *et al.*, 2009). This is different to the direct inhibition of FitB by FitA seen for the FitAB complex (Mattison, *et al.*, 2006).

Miallau *et al.* (2009) propose a catalytic mechanism for VapC-5 similar to that of the endo- and exo-nuclease FEN-1, a member of the FLAP nuclease superfamily.

---

Superimposition of the hypothesised acidic active site residues from VapC-5 onto FEN-1 shows very similar geometry, therefore a two metal ion catalytic mechanism for VapC-5 was proposed (Miallau, *et al.*, 2009). VapC-5 requires  $Mg^{2+}$  for activity which is consistent with this mechanism but it remains to be seen if only one not two  $Mg^{2+}$  ions, are required for activity, as crystallisation of VapC<sub>PAE0151</sub> from *P. aerophilum* demonstrated the presence of only one metal ion in the active site (Bunker, *et al.*, 2008). Comparisons of substrate interactions for the PIN and NYN (Nedd4-BP1, YacP nuclease) domains with the FLAP nuclease domains suggests that PIN and NYN domains are more likely to coordinate a single metal for catalysis (Anantharaman & Aravind, 2006).

The role of VapBC systems in *M. tuberculosis* is largely unknown and is a subject of great interest to tuberculosis researchers. The *vapBC* operons are often found associated with virulence factors, transposases and repetitive elements, suggesting a role in virulence (Arcus, *et al.*, 2005). Or it is possible that different subsets of TA systems are activated in response to different stressors which would enable the organism to adapt to many different stress conditions. It is also possible that the ribonuclease activity of VapC proteins is different from protein to protein; some proteins may target a limited number of mRNAs and therefore not inhibit bulk translation. Together this would enable the organism to erase its transcriptional profile thereby reprogramming the proteome leading to a rapid change in the metabolic state of the cell during conditions of stress. These contrasting hypotheses motivated the research direction at the beginning of this PhD project.

### **1.3.2 The VapBC Toxin-Antitoxin System from *Mycobacterium smegmatis***

The large number of *vapBC* operons on the chromosome of *M. tuberculosis* raises the question as to the biological role of *vapBC* operons. *M. smegmatis* contains only three TA operons in its genome; *phd/doc*, *mazEF* and *vapBC*. *M. smegmatis* has only one *vapBC* operon, reducing the possibility of cross interaction with other chromosomal TA systems. This, along with its relatively fast growth rate makes it an ideal organism to study TA systems in mycobacteria.

---

At the beginning of this project, a bioinformatic search of the *M. smegmatis* genome yielded identification of a putative *vapBC* operon (Sevin & Barloy-Hubler, 2007). MSMEG\_1283 encodes VapB, the antitoxin and belongs to the Rv0623-like family of transcription factors, consistent with DNA binding of antitoxin proteins. MSMEG\_1284 encodes VapC, the toxin and is a member of the PIN domain family. Three highly conserved acidic residues that constitute the active site of PIN domain proteins can be found in *M. smegmatis* VapC (Asp<sup>4</sup>, Asp<sup>99</sup> and Glu<sup>188</sup>). VapC from *M. smegmatis* shows homology (81% sequence similarity) to the VapC protein Rv0624 from *M. tuberculosis*. Rv0624 is found downstream of its cognate antitoxin Rv0623 and is toxic upon induction in *M. smegmatis* and this growth inhibition is relieved by expression of Rv0623 (Ramage, *et al.*, 2009).

There is a one base pair overlap between *vapB* and *vapC* genes in the *M. smegmatis* *vapBC* operon and the two genes are transcribed as a single leaderless mRNA (Robson, *et al.*, 2009). Deletion of either *vapB* or *vapBC* had no effect on the growth of *M. smegmatis* mc<sup>2</sup>155 under normal growth conditions. VapBC binds to its promoter DNA to regulate its transcription (Robson, *et al.*, 2009), by binding to inverted repeat sequences upstream of the start codon of *vapB* that overlap the -10 and -35 sequences. Conditional expression of VapC in both wild-type and  $\Delta$ *vapBC* backgrounds inhibited growth of *M. smegmatis*; growth was inhibited to a lesser extent in the wild-type strain presumably due to the presence of endogenous VapB neutralising VapC (Robson, *et al.*, 2009). Upon removal of the inducer (tetracycline) growth was restored indicating that VapC is bacteriostatic not bactericidal, however this effect does not require the presence of the antitoxin (Robson, *et al.*, 2009). This has also been seen for RelE, which suggests cells have a mechanism to degrade the toxin in the absence of synthesis of the antitoxin. Expression of the VapBC complex has no effect on growth.

The biochemical function of VapC from *M. smegmatis* and VapC proteins from *P. Aerophilum*, or prokaryotic PIN domains in general, was unknown at the beginning of my research.

## 1.4 Objectives

These observations defined the four main objectives to my doctoral research, as outlined below:

### *Objective One*

Express and purify VapB, VapC and VapBC proteins from *Mycobacterium smegmatis* in their host *M. smegmatis*. Characterise and crystallise the VapBC complex.

### *Objective Two*

Determine the mechanism of growth inhibition of VapC in *M. smegmatis*.

### *Objective Three*

Biochemically characterise the nuclease activity for VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub>, the two VapC proteins from *P. aerophilum*, and for VapC<sub>MS1284</sub> from *M. smegmatis*.

### *Objective Four*

Reconcile the biochemical activity of VapC proteins with emerging biological data from our collaborators at the University of Otago and other laboratories.

---

## Chapter Two: Materials & Methods

*Tables of bacterial strains, plasmids, primers buffers and media used in this study can be found in Appendix A. Gene and protein information can be found in Appendix B.*

### 2.1 General Materials & Methods

#### 2.1.1 DNA Manipulations

##### 2.1.1.1 Mycobacterium smegmatis Genomic DNA

*M. smegmatis* genomic DNA was obtained from Emma Summers (Proteins & Microbes Lab, University of Waikato, Hamilton)

##### 2.1.1.2 Escherichia 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. Plasmid DNA was eluted in 50 µl elution buffer.

##### 2.1.1.3 Alkaline Lysis Plasmid Preparation Method

The alkaline lysis plasmid preparation method was used for extraction of plasmids from *M. smegmatis*. Overnight cultures were centrifuged to pellet cells. Cell pellets were resuspended in 200 µl cold GTE (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0). A 300 µl aliquot of room temperature 0.2 M NaOH/1% (w/v) SDS (made fresh) was added followed by 300 µl 3 M potassium acetate pH 4.8 and chilled on ice for 10 minutes. Samples were then centrifuged at 13000 g for 10 minutes at room temperature. A one µl aliquot of 20 mg.ml<sup>-1</sup> RNase A was added to a new 1.5 ml tube, the supernatant transferred to the new tube and incubated at 37 °C for 20 minutes. DNA was then extracted with 0.5 ml chloroform, centrifuged at 13000 g for 1 minute to separate the phases, transferred to a new tube then undergone a second chloroform extraction. The DNA was then precipitated by adding an equal volume of isopropanol and centrifuging for 10 minutes at 13000 g. The pellet was

---

washed with 500  $\mu\text{l}$  70% (v/v) ethanol and left to dry before resuspending in 64  $\mu\text{l}$  of water and transferring to a new tube. The old tube was washed with 32  $\mu\text{l}$  water and pooled with the previous. The DNA was then precipitated by addition of 24  $\mu\text{l}$  4 M NaCl and 120  $\mu\text{l}$  13% (w/v) PEG 8000, mixed thoroughly and chilled on ice for 20 minutes. Samples were then centrifuged at 13000 g for 25 minutes at 4 °C, supernatant removed and the pellet washed with 400  $\mu\text{l}$  70% ethanol. The pellet air dried before resuspending in 20  $\mu\text{l}$  nuclease free water.

#### ***2.1.1.4 Agarose Gel Electrophoresis***

DNA fragments were separated via agarose gel electrophoresis. The percent of agarose in the gel depended on the size of DNA; as a general rule samples < 200 bp were run on a 2% (w/v) gel, 200 - 400 bp 1.5% (w/v), 400 – 1000 bp 1% (w/v) and >1000 bp 0.8% (w/v) agarose in TAE buffer. Samples were mixed with 5 x DNA loading dye prior to loading onto the gel. Agarose gels were either stained with 0.5  $\mu\text{g}\cdot\text{ml}^{-1}$  ethidium bromide or 1 x SYBR Safe<sup>TM</sup> DNA gel stain (Invitrogen, USA). Gels were visualised by UV light (ethidium bromide) or on a blue light box (Invitrogen, USA) and images captured. Band sizes were determined by comparison with the 1kb-Plus DNA ladder (Invitrogen, USA).

#### ***2.1.1.5 DNA Extraction from an Agarose Gel***

DNA was separated and visualised with SYBR Safe<sup>TM</sup> DNA dye as described in Section 2.1.1.4. The desired band was cut from the gel using a scalpel blade cleaned with 70% (v/v) ethanol. DNA was eluted from the gel using the QIAquick Gel Extraction Kit (Qiagen, Netherlands) according to manufacturer's instructions. DNA was eluted in 30  $\mu\text{l}$  of elution buffer.

#### ***2.1.1.6 DNA Quantification***

DNA was quantified using a Nanodrop ND-1000 Spectrophotometer (Nanodrop Technologies, USA). This measures absorbance of DNA at 260 nm, quantifying the amount of DNA.





#### **2.1.1.10 DNA Ligation**

DNA ligations were performed according to manufacturer's recommendations using 5 U of T4 DNA Ligase (Invitrogen, USA) in 20  $\mu$ l reactions and incubated for 6 hours at room temperature or at 14 °C overnight.

#### **2.1.1.11 DNA Transformation**

##### **2.1.1.12 Preparation of Electrocompetent *Escherichia coli***

A glycerol stock of the desired strain of *E. coli* was streaked onto a Luria-Bertani (LB)-agar plate and incubated at 37 °C overnight. A single colony was then used to start a 5 ml LB culture which was grown overnight at 37 °C in a shaking incubator at 180 rpm. The overnight culture was used to seed a 500 ml LB culture, which was then grown at 37 °C in a shaking incubator at 180 rpm until an OD<sub>600</sub> of 0.5 - 0.7 was reached. The cells were then chilled on ice for at least 20 minutes before transferring to a pre-chilled sterile centrifuge bottle and centrifuged at 4000 g for 15 minutes at 4 °C. The supernatant was discarded and the cell pellet resuspended in 500 ml ice-cold sterile 10% (v/v) glycerol. The cells were again centrifuged at 4000 g for 15 minutes at 4 °C and the supernatant discarded. The cell pellet was resuspended in 250 ml ice cold sterile 10% (v/v) glycerol and centrifuged at 4000 g for 15 minutes at 4 °C. The supernatant was discarded and the cell pellet resuspended in 20 ml ice cold sterile 10% (v/v) glycerol and transferred to a pre-chilled 50 ml falcon tube and centrifuged at 4000 g for 15 minutes at 4 °C. The supernatant was discarded and the pellet resuspended in 1 to 2 ml ice cold sterile 10% (v/v) glycerol. Aliquots (50  $\mu$ l) of resuspended cell were then flash frozen in liquid nitrogen and stored at -80 °C.

##### **2.1.1.13 Electroporation of *Escherichia coli***

A 1  $\mu$ l sample of plasmid DNA or ligation reaction was added to 50  $\mu$ l freshly thawed (on ice) electrocompetent *E. coli* cells of the desired strain. The mixture was placed in a 2 mm electroporation cuvette (BioRad Laboratories, USA), tapped to settle the mixture between the electrodes and electroporated with a Bio-Rad Gene Pulser<sup>TM</sup> (Bio-Rad Laboratories, USA) at 2.5 kV, 25  $\mu$ F capacitance and 200  $\Omega$  resistance. One ml of SOC media was immediately added to the electroporated cells and

---

incubated at 37 °C shaking at 200 rpm for 30 minutes. An aliquot of the cells were plated on LB agar plates containing the appropriate antibiotic for selection and incubated at 37 °C overnight. Colonies from these agar plates were then used to start 5 ml LB or terrific broth cultures containing the appropriate antibiotic. These cultures were incubated at 37 °C shaking at 180 rpm overnight and used to isolate plasmid DNA according to the method in Section 2.1.1.2.

#### **2.1.1.14 Preparation of Electrocompetent *Mycobacterium smegmatis***

A glycerol stock of the desired strain of *M. smegmatis* was streaked onto an LB agar plate including 0.05% (v/v) tween 80 (LBT) and incubated at 37 °C overnight. A single colony from this plate was used to start a 5 ml LBT culture which was grown overnight at 37 °C in a shaking incubator at 200 rpm until an OD<sub>600</sub> of around 0.7 was reached. One ml of the overnight culture was used to inoculate 100 ml of 7H9/ADC + 0.05% (v/v) tween 80 media. The 100 ml culture was then grown at 37 °C in a shaking incubator at 200 rpm until an OD<sub>600</sub> of 0.5 - 0.7 was reached. The cells were then chilled on ice for 1.5 hours before transferring to a pre-chilled sterile centrifuge bottle and centrifuged at 4000 g for 20 minutes at 4 °C. The supernatant was discarded immediately as mycobacterial cell pellets are very flimsy and likely to break up and float away. The cell pellet was resuspended in 100 ml ice cold sterile 10% (v/v) glycerol. The cells were again centrifuged at 4000 g for 20 minutes at 4 °C and the supernatant discarded leaving a little supernatant to resuspend the cells in. The cell pellet was resuspended in the remaining supernatant, transferred to a 50 ml falcon tube and centrifuged at 4000 g for 20 minutes at 4 °C. The supernatant was discarded and the cell pellet resuspended in 0.2 ml ice cold sterile 10% (v/v) glycerol then aliquoted into eppendorf tubes (40 µl volumes) and flash frozen in liquid nitrogen and stored at -80 °C until needed.

#### **2.1.1.15 Electroporation of *Mycobacterium smegmatis***

Electroporation of *M. smegmatis* requires adjustment of transformation parameters. One µl of plasmid DNA or ligation reaction was added to 40 µl freshly thawed (on ice) electrocompetent *M. smegmatis* cells of the desired strain and 260 µl sterile

---

10% (v/v) glycerol. The mixture was placed in a 2 mm electroporation cuvette (Bio-Rad Laboratories, USA), tapped to settle the mixture between the electrodes and electroporated with a Bio-Rad Gene Pulser™ (Bio-Rad Laboratories, USA) at 2.5 kV, 25  $\mu$ F capacitance and 1000  $\Omega$  resistance. One ml of 7H9/ADC media was immediately added to the electroporated cells and incubated at 37 °C shaking at 200 rpm for 3 hours. An aliquot of the cells were plated on 7H10/ADC + 0.05% (v/v) tween 80 agar plates containing the appropriate antibiotic for selection and incubated at 37 °C for 3 days.

### **2.1.2 Glycerol Stocks**

Glycerol stocks for long term storage of transformed bacteria were made by the addition of 0.5 ml of overnight culture (LB for *E. coli* or LBT for *M. smegmatis* + appropriate antibiotic) to 0.5 ml sterile 50% (v/v) glycerol. Glycerol stocks were stored at -80 °C.

### **2.1.3 Protein Purification and Manipulation**

#### **2.1.3.1 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) Protein Analysis**

SDS-PAGE gels were cast in a Hoeffer gel casting system. SDS-PAGE gels consisted of a stacking gel (5% (v/v) acrylamide) overlaid on a resolving gel (10, 12, 15 or 18% (v/v) acrylamide depending on 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) and included 0.1% (w/v) SDS, and were polymerised by addition of ammonium persulphate (APS) (0.05% (w/v)) and TEMED (0.05% (v/v)).

Protein samples were mixed in a 3:1 ratio with 4 x SDS loading buffer and heated to 90 °C for 5 minutes before loading onto gel. Gels were run in 1 x SDS-PAGE running buffer at 15 mA until the dye front entered the resolving gel and then at 20 - 25 mA until the dye front reached the end of the gel.

---

### ***2.1.3.2 Native Polyacrylamide Gel Electrophoresis (Native-PAGE) Protein Analysis***

Native-PAGE gels were cast and run as for SDS-PAGE gels (Section 2.1.3.1) except without SDS in the gel, running buffer or loading dye.

### ***2.1.3.3 Coomassie Blue Stain for Protein Gel Electrophoresis***

Gels were stained using the quick coomassie blue staining method (Wong, Sridhara, Bardwell, & Jakob, 2000) unless otherwise stated. Gels were placed in a microwaveable box with 50 ml Fairbanks staining solution A, microwaved for 30 seconds on full power then cooled to room temperature while shaking gently. The stain was then removed and Fairbanks staining solution B added and microwaved as above. This was then repeated for Fairbanks staining solutions C & D, solution D being the destaining solution. Protein sizes were estimated by comparison with the protein sizes of the Precision Plus Protein <sup>TM</sup> Unstained Standard (Bio-Rad Laboratories, USA).

### ***2.1.3.4 Silver Staining for Protein Gel Electrophoresis***

Where stated, gels were silver stained for protein. Gels were placed in a glass bowl, covered with fixer solution, microwaved on high for 30 seconds and left shaking for 5 – 10 minutes. The gel was then covered with 20% methanol and microwaved for 30 seconds then left shaking for 5 minutes. The gel was then covered with 0.02% (w/v) sodium thiosulphate, left shaking for 1.5 – 2 minutes (2 minutes maximum) then washed with water twice. Cold silver stain was added and left shaking for 20 - 30 minutes before washing briefly with water. To rinse the gel 10 - 15 ml of developer was added (made fresh), then removed and the remaining 35 - 40 ml of developer added. Development was stopped by addition of 10% (v/v) acetic acid.

### ***2.1.3.5 Measurement of Protein Concentration***

Protein concentration was measured using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, USA). The Nanodrop measures absorbance at 280 nm and

---

the accompanying software calculates protein concentration using the Beer-Lambert equation:

$$A = \epsilon \cdot c \cdot l$$

A = absorbance at 280 nm

$\epsilon$  = theoretical molar extinction coefficient ( $M^{-1}cm^{-1}$ )

c = concentration (M)

l = pathlength (cm)

Theoretical molar extinction coefficients were calculated using the online tool ProtParam (Gasteiger, *et al.*, 2005), by submission of the amino acid sequence.

#### ***2.1.3.6 Concentration of Protein Samples***

Protein samples were concentrated using 20 ml, 2 ml or 500  $\mu$ l Vivaspin concentrators (Sartorius AG, Germany) with molecular weight cut offs of 5 or 10 kDa. Protein samples were added to the upper reservoir of the concentrator and spun at 3000 g (20 ml), 3600 g (2 ml) or 9000 g (500  $\mu$ l) at 4 °C until the desired volume or concentration was reached.

#### ***2.1.3.7 Dialysis of Protein Samples***

Spectra Por<sup>®</sup> dialysis tubing of 6 - 8 kDa molecular weight cut off (Spectrum Laboratories, USA) was used for dialysis of protein samples. The dialysis tubing was pre-wet in target buffer, the protein sample was added to the tubing and the ends of the tubing sealed with dialysis clips. The protein sample was then left to dialyse against 1 L of target buffer with gentle stirring for 2 to 3 hours at 4 °C. The target buffer was then replaced with fresh buffer and the protein dialysed with gentle stirring overnight at 4 °C.

#### ***2.1.3.8 Fast Performance Liquid Chromatography (FPLC)***

Fast performance liquid chromatography was performed using either an ÄKTA Basic<sup>™</sup> or ÄKTA Prime<sup>™</sup> FPLC system (GE Healthcare, UK). Elution of proteins was monitored by absorbance at 280 nm.

### ***2.1.3.9 Immobilised Metal Affinity Chromatography (IMAC)***

A 5 ml HiTrap<sup>TM</sup> Chelating HP column (GE Healthcare, UK) was used to purify proteins by immobilised metal affinity chromatography (IMAC). Prior to purification columns were primed with 5 ml 100 mM NiCl<sub>2</sub> and washed with 20 ml water, followed by 20 ml of lysis buffer that will be used in the purification. Columns were stripped of Ni<sup>2+</sup> ions using 10 ml 100 mM EDTA pH 8.0 and re-primed between purifications of different proteins.

### ***2.1.3.10 Purification of His-Tagged Proteins via IMAC***

The supernatant from lysed large scale expression cultures was filtered through 1.2, 0.45 and 0.2 µm Minisart filters (Sartorius AG, Germany). If the supernatant was very viscous as tends to be the case with large scale expression cultures of *M. smegmatis*, it was filtered only through 1.2 and 0.45 µm Minisart filters. The filtered supernatant was then loaded onto an IMAC column, which was pre-equilibrated with the appropriate lysis buffer at a flow rate of 1 - 2 ml.min<sup>-1</sup>. The column was then attached to either an ÄKTA Prime<sup>TM</sup> or ÄKTA Basic<sup>TM</sup> FPLC system. The column was washed with 15 - 25 ml of lysis buffer, at a flow rate of 1 ml.min<sup>-1</sup> to remove unbound proteins. Bound proteins were removed by running a gradient of 0 – 50% elution buffer (lysis buffer + 1 M imidazole) over 50 ml at a flow rate of 1 ml.min<sup>-1</sup>. Fractions containing the desired protein were then analysed by SDS-PAGE (Section 2.1.3.1).

### ***2.1.3.11 Size Exclusion Chromatography***

Size exclusion chromatography was used to further purify proteins. Size exclusion chromatography was performed using either a HiLoad<sup>TM</sup> 16/60 Superdex<sup>TM</sup> or a 10/300 Superdex<sup>TM</sup> column (analytical size exclusion column) (GE Healthcare, UK). Protein solutions were concentrated to between 1 and 5 ml and filtered with a Nanosep<sup>®</sup> MF 0.2 µm filter (Pall, USA) before loading onto a pre-equilibrated column.

### ***2.1.3.12 Calibration of Size Exclusion Column S200 10/300***

Calibration of the S200 10/300 analytical size exclusion column was carried out using the high & low molecular weight gel filtration calibration kits (GE Healthcare, UK) according to manufacturer's instructions. Briefly, all proteins were reconstituted to the recommended concentration with 50 mM phosphate buffer pH 7.4, 200 mM NaCl with the exception of carbonic anhydrase which was resuspended in water as per kit recommendations. Two protein mixes were prepared; Protein Mix A contained ferritin, conalbumin, carbonic anhydrase and ribonuclease with final concentrations of 0.3 mg.ml<sup>-1</sup> for ferritin and 3 mg.ml<sup>-1</sup> for the remaining proteins. Protein Mix B contained aldolase, ovalbumin, ribonuclease A and aprotinin with final concentrations of 4 mg.ml<sup>-1</sup> for aldolase and ovalbumin and 3 mg.ml<sup>-1</sup> ribonuclease A and aprotinin. Each protein mix was then fractionated and the elution volume for each protein determined. The void volume of the column was determined by the elution volume of Blue Dextran 2000 (1 mg.ml<sup>-1</sup>).

A calibration curve could then be determined for the column by calculation of the  $K_{av}$  (average elution volumes) values for the calibration kit proteins using the equation:

$$K_{av} = (V_e - V_o) / (V_c - V_o) \quad \text{where } V_o = \text{column void volume}$$
$$V_e = \text{elution volume}$$
$$V_c = \text{geometric column volume}$$

The calibration curve was prepared by plotting  $K_{av}$  against the log molecular weight and a linear equation determined (Appendix A.4) to calculate the molecular weight of sample proteins.

---

## 2.2 Methods relating to Chapter Three: Protein Expression, Purification & Characterisation

### 2.2.1 Cloning of *vapB*, *vapC* and *vapBC* into pYUB1049

pYUB1049 is an *E. coli* – mycobacterial shuttle vector. The open reading frames (ORFs) encoding the individual genes *vapB* (MSMEG\_1283) and *vapC* (MSMEG\_1284) and the *vapBC* operon (MSMEG\_1283 and MSMEG\_1284) were amplified as in Section 2.1.1.7 from *M. smegmatis* genomic DNA using primers engineered (Section 2.1.1.7) to contain an NcoI restriction site in the forward primer and a BamHI restriction site in the reverse primer. The amplified products were then digested with NcoI/BamHI restriction enzymes (Invitrogen, USA) (Section 2.1.1.9), purified using spin columns (Roche Applied Science, Switzerland) (Section 2.1.1.8) and ligated (Section 2.1.1.10) into the pYUB1049 shuttle vector between the NcoI and BamHI restriction sites enabling expression with a C-terminal His-tag. The constructs were then transformed into *E. coli* TOP10 cells as described in Section 2.1.1.13 and plated on low salt LB agar medium supplemented with 0.05% (w/v) tween 80 and 50  $\mu\text{g}\cdot\text{ml}^{-1}$  hygromycin B to select for positive transformants. Constructs were sequenced to ensure correct insertion. Following plasmid purification the constructs were then transformed into *M. smegmatis* mc<sup>2</sup>4517 electrocompetent cells by electroporation as described in Section 2.1.1.15. Positive colonies were selected by plating the transformants on 7H10 agar media supplemented with ADC (albumin, dextrose, catalase supplement), 0.05% (w/v) tween 80 and 50  $\mu\text{g}\cdot\text{ml}^{-1}$  kanamycin and hygromycin B.

### 2.2.2 Protein Expression and Purification using *Mycobacterium smegmatis* as a Host

#### 2.2.2.1 Small Scale Protein Expression and Purification from *Mycobacterium smegmatis*

Small scale expression tests were used as a screen for expression of VapB, VapC and VapBC in *M. smegmatis*. A single transformed colony was selected and used to inoculate a PA-0.5G/tween 80 seeder culture. This culture was grown for 48 hours at

---

37 °C and used at a 1:100 dilution to inoculate a ZYP-5052/tween 80 expression culture. The expression culture was grown at 37 °C and aliquots of culture were taken at 24, 48 and 96 hours to check for protein expression. Expression test cultures were spun at 4500 g for 20 minutes to pellet cells. Cell pellets were resuspended in 200 µl lysis buffer (50 mM phosphate buffer pH 7.4, 200 mM NaCl, 20 mM imidazole), 5 µl was removed and added to 10 µl lysis buffer and 5 µl 4 x SDS loading dye (whole-cell sample). The cells were then lysed by sonication on ice using a small probe in 5 - 10 second bursts at a low setting using a Misonix Sonicator (U.S.A) until cells were lysed. The lysate was then spun at 13000 g to separate soluble and insoluble fractions. The insoluble fraction was resuspended in 200 µl lysis buffer and 15 µl of soluble and insoluble fractions added to 5 µl of 4 x SDS loading buffer respectively.

Small scale His-tag binding tests were performed on the soluble fraction of the cell lysate as below. 50 µl of Ni Sepharose (GE Healthcare, UK) was transferred to a 1.5 ml eppendorf tube. The Ni sepharose was then washed with 1 ml of the appropriate lysis buffer and centrifuged at 2000 g for 30 seconds to collect the Ni sepharose in the bottom of the tube. The supernatant was removed and 200 µl of soluble fraction (lysate) added. The mixture was then incubated at room temperature (22 °C) for 15 minutes whilst shaking at 650 rpm. The sample was then centrifuged again at 2000 g for 30 seconds to collect the Ni sepharose in the bottom of the tube. The supernatant removed and the Ni sepharose washed by resuspending in 1 ml of lysis buffer. The sample was then centrifuged at 2000 g for 30 seconds, supernatant removed and the washing cycle repeated two more times. A 25 µl aliquot of 4 x SDS loading buffer was added to the Ni sepharose and samples analysed by SDS-PAGE.

#### ***2.2.2.2 Large Scale Protein Expression and Purification from Mycobacterium smegmatis***

VapB, VapC and VapC protein expression was performed in auto-induction media supplemented with 0.05% (w/v) tween 80 and 50 µg.ml<sup>-1</sup> kanamycin and hygromycin B. A single transformed colony was selected and used to inoculate a PA-0.5G/tween 80 seeder culture. This culture was grown for 48 hours at 37 °C and was used at a

---

1:100 dilution to inoculate a ZYP-5052/tween 80 expression culture, which was grown at 37 °C for 96 hours for maximal protein expression.

Cells from large scale expression cultures were harvested by spinning at 7000 g for 25 minutes at 4 °C. Cell pellets were resuspended in 50 mM phosphate buffer pH 7.4, 200 mM NaCl with the addition of an EDTA free protease inhibitor tablet (Roche Applied Science, Switzerland). Resuspended cells were then lysed by sonication on ice using a large probe in 30 second bursts at setting 6 - 7 using a Misonix Sonicator (U.S.A) for a total of 4 - 5 minutes. The cell lysate was centrifuged at 16000 g for 20 minutes at 4 °C to separate the soluble and insoluble fractions. The soluble fraction containing the His-tagged protein was loaded onto a HisTrap HP column (GE Healthcare Life Sciences) and purified as in Section 2.1.3.10. Fractions containing the protein of interest were concentrated as in Section 2.1.3.6 and purified further by size exclusion chromatography SEC using a Superdex 200 16/60 column (GE Healthcare, UK) as in Section 2.1.3.11 in the same buffer (50 mM phosphate buffer pH 7.4, 200 mM NaCl).

Samples for SDS-PAGE were prepared as follows, pellets were resuspended in the original volume of lysis buffer, and 3 µl of the supernatant (soluble fraction) and the resuspended pellet were transferred to a tube containing 12 µl of lysis buffer and 5 µl of 4 x SDS loading buffer.

### **2.2.3 Cloning of *vapC*<sub>MS1284</sub> into pMAL-c2x MBP-fusion vector**

The ORF encoding the individual *vapC* gene (MSMEG\_1284) was amplified as in Section 2.1.1.7 from *M. smegmatis* genomic DNA using primers engineered (Section 2.1.1.7) to contain an EcoRI restriction site in the forward primer and a BamHI restriction site in the reverse primer. The amplified products were then digested with EcoRI and BamHI restriction enzymes (Invitrogen, USA) (Section 2.1.1.9), purified using spin columns (Roche Applied Science, Switzerland) (Section 2.1.1.8) and ligated (Section 2.1.1.10) into pMAL-c2x between the EcoRI and BamHI restriction sites enabling expression of a maltose binding protein (MBP)-VapC fusion protein.

---

The constructs were then transformed into *E. coli* BL21 (DE3) electrocompetent cells as described in Section 2.1.1.13 and plated on LB agar medium supplemented with 100  $\mu\text{g}\cdot\text{ml}^{-1}$  ampicillin to select for positive transformants. Constructs were sequenced to ensure correct insertion.

## **2.2.4 Protein Expression and Purification of a MBP-VapC Fusion Protein**

### ***2.2.4.1 Small Scale MBP-VapC Protein Expression and Purification Tests***

Small scale expression tests were used as a screen for expression of an MBP-VapC fusion protein in *E. coli*. A single transformed colony was selected and used to inoculate a rich medium + glucose seeder culture. This culture was grown for 24 hours at 37 °C and used at a 1:100 dilution to inoculate a rich medium + glucose expression culture and grown at 37 °C until an OD<sub>600</sub> of around 0.5 was reached. Protein expression was then induced with 0.3 mM IPTG and aliquots of culture were taken at 1, 2 and 3 hours to check for protein expression.

Ten ml expression test cultures were harvested and lysed as for small scale protein expression from *M. smegmatis* with the sonication times reduced in 50 mM phosphate buffer pH 7.4, 200 mM NaCl, 1 mM EDTA (lysis buffer). Small scale MBP binding tests were performed on the soluble fraction of the cell lysate as below.

A 50  $\mu\text{l}$  aliquot of amylose resin (New England Biolabs, U.S.A) was transferred to a 1.5 ml eppendorf tube. The resin was washed with 1 ml of the appropriate lysis buffer (50 mM phosphate buffer pH 7.4, 200 mM NaCl, 1 mM EDTA) and centrifuged at 2000 g for 30 seconds to collect the resin in the bottom of the tube. The supernatant was removed and the wash step repeated. The resin was resuspended in 50  $\mu\text{l}$  lysis buffer and 50  $\mu\text{l}$  of soluble fraction (supernatant from lysate) was added; the mixture was then incubated on ice for 15 minutes. The sample was then centrifuged again at 2000 g for 30 seconds to collect the resin in the bottom of the tube. The supernatant was removed and the resin washed by resuspending in 1 ml of lysis buffer, the sample was then centrifuged at 2000 g for 30 seconds, the supernatant removed and the wash

---

step repeated two more times. Protein was either eluted off the resin or 25  $\mu$ l of 4 x SDS loading buffer was added to the resin. Preliminary elution experiments of the MBP-tagged protein from the amylose resin were conducted, protein was eluted from the column with a variety of maltose concentrations in elution buffer (50 mM phosphate buffer pH 7.4, 200 mM NaCl, 1 mM EDTA + maltose), ranging from 10mM - 100 mM maltose concentrations. A 50  $\mu$ l aliquot of elution buffer was added to the resin and the mixture incubated at room temperature (22 °C) for 30 minutes shaking at 600 rpm. The sample was then centrifuged again at 2000 g for 30 seconds to collect the resin in the bottom of the tube and the supernatant containing the desired protein removed. Samples were subsequently analysed by SDS-PAGE.

#### **2.2.4.2 Large Scale MBP-VapC Expression and Purification**

Expression of a MBP-VapC fusion protein was performed in rich media + glucose supplemented with 100  $\mu$ g.ml<sup>-1</sup> ampicillin. A single transformed colony was selected and used to inoculate a rich media + glucose culture. This culture was grown for 24 hours at 37 °C and was used at a 1:100 dilution to inoculate a rich media + glucose expression culture and grown at 37 °C until an OD<sub>600</sub> of around 0.5 was reached. Protein expression was then induced with 0.3 mM IPTG and cells incubated for a further 2 hours at 37 °C for maximal protein expression.

*E. coli* expression cultures were harvested and lysed as for *M. smegmatis* expression cultures. Cell pellets were resuspended in 50 mM phosphate buffer pH 7.4, 200 mM NaCl, 1 mM EDTA with the addition of an EDTA free protease inhibitor tablet (Roche Applied Science, Switzerland). Resuspended cells were then lysed by sonication on ice using a large probe in 30 second bursts at setting 6 - 7 using a Misonix Sonicator (U.S.A) for a total of 3 minutes. The cell lysate was centrifuged at 16000 g to separate the soluble and insoluble fractions. The soluble fraction containing the MBP-VapC fusion protein was loaded onto a HiTrap MBP column (GE Healthcare, UK) and purified as in Section 2.2.5.1. Fractions containing the fusion protein were concentrated as in Section 2.1.3.6 and subjected to either size

---

exclusion chromatography (Section 2.1.3.11) or Factor Xa cleavage to remove the MBP protein from VapC (Section 2.2.5.2)

## **2.2.5 Amylose Affinity Chromatography**

A 1 ml HiTrap<sup>TM</sup> MBP column (GE Healthcare, UK) was used to purify proteins by amylose affinity chromatography. Prior to purification columns were washed with 5 ml water followed by 5 ml of lysis buffer, (50 mM phosphate buffer pH 7.4, 200 mM NaCl, 1 mM EDTA). The column was regenerated by flushing with 3 ml of water, 3 ml of 1.5 M NaCl, 3 ml of water, 3 ml of 0.5 M NaOH, 3 ml of water then 3 ml of 20% (v/v) ethanol.

### ***2.2.5.1 Purification of MBP-Tagged Proteins via Amylose Affinity***

The supernatant from lysed large scale expression cultures was filtered through 1.2, 0.45 and 0.2  $\mu\text{m}$  Minisart filters (Sartorius AG, Germany). The filtered supernatant was then loaded onto a MBP column, pre-equilibrated with lysis buffer at a flow rate of 1 - 2  $\text{ml}\cdot\text{min}^{-1}$ . The column was then attached to either an ÄKTA Prime<sup>TM</sup> or ÄKTA Basic<sup>TM</sup> FPLC system. The column was washed with 12 ml of lysis buffer at a flow rate of 1  $\text{ml}\cdot\text{min}^{-1}$  to remove unbound proteins. Bound proteins were removed by running 5 ml of elution buffer (50 mM phosphate buffer pH 7.4, 200 mM NaCl, 1 mM EDTA 10 mM maltose) through the column at a flow rate of 1  $\text{ml}\cdot\text{min}^{-1}$ . Fractions containing the desired protein were then analysed by SDS-PAGE (Section 2.1.3.1).

### ***2.2.5.2 MBP-Affinity Tag Cleavage by Factor Xa Protease***

Factor Xa protease was used to cleave MBP from MBP-tagged protein. An initial evaluation of fusion protein cleavage was carried out using Factor Xa at 1, 2 and 3% (v/v) the fusion protein. The reaction mixture was incubated for 2, 4, 6 or 24 hours at room temperature (22 °C) with shaking at 300 rpm. Low concentrations of SDS (0.01% and 0.05% (w/v)) were added to improve cleavage of MBP.

### 2.2.6 Denaturing Purification of VapC from *Escherichia coli*

Cloning of the *vapC* gene from *M. smegmatis* into the pProEX expression vector was previously carried out by Emma Summers and the protein purified according to the following method. Expression of VapC was performed in LB media, protein expression was induced with 1 mM IPTG at an OD<sub>600</sub> 0.4 - 0.6 and incubated at 37 °C overnight. Cells from expression cultures were harvested and lysed as for the MBP-VapC purification above. Cell pellets were resuspended in 50 mM phosphate buffer pH 7.4, 200 mM NaCl with the addition of an EDTA free protease inhibitor tablet (Roche Applied Science, Switzerland). The insoluble fraction was resuspended in 50 mM phosphate buffer pH 7.4, 200 mM NaCl + 8 M urea and left to resuspend for 1.5 hours. The resuspended insoluble pellet was then spun at 10000 g for 10 minutes, the supernatant removed and loaded onto a HisTrap HP column (GE Healthcare Life Sciences) and the protein purified as in Section 2.1.3.10.

### 2.2.7 Protein Refolding

#### 2.2.7.1 Refolding Screen Solutions

A modified version of the Hampton refolding screen (Hampton Research, USA), consisting of 16 solutions was used in refolding screens. These screened the effect of: the presence or absence of divalent cations (Mg<sup>2+</sup>/Ca<sup>2+</sup> and EDTA); buffer composition and pH (MES/Tris-HCl); high or low concentrations of salt (NaCl/KCl); presence or absence of PEG; presence or absence of detergent (Triton X); presence or absence of polar (arginine) or non-polar (sucrose) additives; presence or absence of chaotropes (guanidine-HCl); and the type of oxidising/reducing agent (GSH/GSSG or DTT).

Refolding screen solutions can be found in Appendix A.

#### 2.2.7.2 Small Scale (20 µl) Refolding Screen

Protein purified by IMAC (Section 2.1.3.9) under denaturing conditions (Section 2.2.6) was used in refolding screens. Denatured protein was dialysed in 20 µl dialysis buttons against each of the 16 refolding screen solutions at 4 °C for 1 - 2 days.

---

Buttons were examined and scored for precipitate and then analysed by native-PAGE (Section 2.1.3.2). Successful protein refolding was determined by an absence of precipitate and the ability of the protein to run on a native polyacrylamide gel (Section 2.1.3.2).

### **2.2.7.3 Scale up (200 $\mu$ l) of Protein Refolding**

Protein purified by IMAC (Section 2.1.3.9) under denaturing conditions (Section 2.2.6) was placed in a 200  $\mu$ l dialysis button and dialysed into successful refolding conditions for 1 - 2 days at 4 °C. Samples were then run on a size exclusion S200 10/300 column and peak elution analysed. Fractions were analysed by SDS-PAGE and native-PAGE.

### **2.2.8 Disruption of the VapBC Complex from *Mycobacterium smegmatis***

The VapBC complex was purified as in Section 2.2.2.2. A variety of screens were used to attempt to disrupt the VapBC complex from *M. smegmatis* so VapC could be isolated and characterised. The VapBC complex is expressed as an operon in the mycobacterial expression vector pYUB1049, enabling expression of a C-terminal His-Tag on VapC which can be utilised in disruption of the complex. The VapBC complex was purified (Section 2.2.2.2) and subjected to a variety of screens aimed at disrupting the interaction between VapC and VapB without denaturing the proteins.

The VapBC complex was bound to Ni sepharose as in Section 2.2.2.1, centrifuged at 2000 g for 30 seconds to collect the Ni sepharose in the bottom of the tube and the supernatant removed. One ml of disruption buffer was added to the Ni sepharose and incubated at room temperature with shaking at 650 rpm for 20 minutes to disrupt the VapB-VapC interaction. Samples were then centrifuged at 2000 g for 30 seconds and the supernatant removed and kept. The Ni sepharose was washed by resuspending in 1 ml of disruption buffer (see Section 2.2.8.1 below), the sample was centrifuged at 2000 g for 30 seconds, the supernatant removed and the washing cycle repeated two more times. A 25  $\mu$ l aliquot of 4 x SDS loading buffer was added to the Ni sepharose and samples were analysed by SDS-PAGE.

### 2.2.8.1 *VapBC Disruption Screens*

The buffers in the screens below were used to screen disruption of the VapBC complex:

#### pH Screen

50 mM sodium acetate, 200 mM NaCl, 20 mM imidazole + 0.01% (v/v) triton X-100 at pH 3.5, 4.0, 4.5, 5.0 and 5.5.

50 mM phosphate buffer pH 7.4, 200 mM NaCl, 20 mM imidazole + 0.01% (v/v) triton X-100 at pH 6.0, 6.5, and 7.0.

Purified VapBC complex was added to 20 µl dialysis buttons and dialysed against 1 ml of the buffers below in a 24 well plate at 4 °C overnight. Disruption was analysed by binding to Ni sepharose as in Section 2.2.8 above.

#### Salt screen

50 mM phosphate buffer pH 7.4, 200 mM NaCl, 20 mM imidazole with NaCl concentrations ranging from 0.5 M – 2.0 M.

#### Guanidium screen

50 mM phosphate buffer pH 7.4, 200 mM NaCl, 20 mM imidazole with guanidium concentrations ranging from 0.5 M – 6.0 M.

#### Detergent Screen

50 mM phosphate buffer pH 7.4, 200 mM NaCl, 20 mM imidazole.

Plus one of the following detergents, 0.078 M tween 20, 0.01 M tween 80, 2% triton X-100 (v/v), 0.1% SDS (w/v), 0.5% SDS (w/v), 20 mM zwittergent 3-10, 1.5 mM zwittergent 2-12, 0.1% (w/v) lauryl dimethyl amine oxide (LDAO) and 1% (w/v) LDAO.

---

Each concentration of detergent was at less than its reported critical micelle concentration (CMC).

#### Urea Screen

50 mM phosphate buffer pH 7.4, 200 mM NaCl, 20 mM imidazole with urea concentrations ranging from 0.5 M – 8.0 M.

#### Urea + Detergent Screen

50 mM phosphate buffer pH 7.4, 200 mM NaCl, 20 mM imidazole, 8 M urea  
Plus one of the following detergents, 0.078 M tween 20, 0.01 M tween 80, 2% triton X-100 (v/v), 0.1% (w/v) SDS and 0.5% (w/v) SDS.

### **2.2.9 Tryptic Digest of the VapBC Complex from *Mycobacterium smegmatis***

#### **2.2.9.1 Trypsin Digest of VapBC Complex**

The VapBC complex from *M. smegmatis* was purified as in Section 2.2.2.2 and digested with 0.1 mg.ml<sup>-1</sup> trypsin (Sigma Aldrich, U.S.A) for 5, 10, 15, 20, 30, 45 and 60 minutes at room temperature. Reactions were stopped by addition of 4 x SDS loading dye and heating at 90 °C for 5 minutes. Samples were analysed by SDS-PAGE (Section 2.1.3.1). A 1 hour time point was optimal for trypsin digestion of VapBC.

#### **2.2.9.2 Trypsin Inhibitors**

Trypsin digest reactions were stopped by the addition of 0.1 mg.ml<sup>-1</sup> trypsin inhibitor from *Glycine max* (soybean) (Sigma Aldrich, U.S.A) and incubating for 15 minutes at room temperature while shaking at 500 rpm on an orbital shaker.

### **2.2.10 Purification of VapC from *Mycobacterium smegmatis***

#### **2.2.10.1 Purification of VapC by Anion Exchange Chromatography**

Trypsin digests (Section 2.2.9.1) of purified VapBC complex inhibited with trypsin inhibitor (Section 2.2.9.2) were dialysed against 50 mM phosphate buffer pH 7.4, 100 mM NaCl overnight (Section 2.1.3.7) and subjected to anion exchange

---

chromatography to separate the trypsin + inhibitor from VapC remaining after the digest as in Section 2.2.10.1 .

A 5 ml HiTrap<sup>TM</sup> Q Anion Exchange column (GE Healthcare, UK) was used to purify the VapBC trypsin digest (Section 2.2.9.1) by anion exchange chromatography. Prior to purification the column was washed with 20 ml of water, and then equilibrated in 40 ml of start buffer (50 mM phosphate buffer pH 7.4, 100 mM NaCl). Columns were regenerated by washing with 30 ml of water followed by 30 ml of 20% (v/v) ethanol. The digest was filtered through a 0.2 µm Minisart filter (Sartorius AG, Germany) and loaded onto an anion exchange column, pre-equilibrated with start buffer (50 mM phosphate buffer pH 7.4, 100 mM NaCl) at a flow rate of 1 - 2 ml.min<sup>-1</sup>. The column was then attached to either an ÄKTA Prime<sup>TM</sup> or ÄKTA Basic<sup>TM</sup> FPLC system. The column was washed with 25 ml of start buffer at a flow rate of 1 ml.min<sup>-1</sup> to remove the trypsin. Bound proteins were removed by running a gradient of 0 - 100% (v/v) elution buffer (50 mM phosphate buffer pH 7.4, 1 M NaCl) over 50 ml at a flow rate of 1 ml.min<sup>-1</sup>. Fractions containing the desired protein were then analysed by SDS-PAGE (Section 2.1.3.1).

### ***2.2.10.2 Size Exclusion Chromatography of VapC***

VapC resulting from a trypsin digest of VapBC (Section 2.2.9.1) was purified away from trypsin by anion exchange chromatography (Section 2.2.10.1) and further purified by size exclusion chromatography. The sample was analysed on an analytical S75 10/300 gel filtration column as in Section 2.1.3.11.

### **2.2.11 Dynamic Light Scattering**

The state of aggregation and polydispersity of protein samples was measured by dynamic light scattering (DLS). Protein samples were centrifuged at 16000 g for 10 minutes at 4 °C to remove any precipitated protein. A 15 µl sample was added to a clean and dry quartz cuvette and dynamic light scattering analysis performed on a DynaPro MTSC Dynamic Light Scattering instrument (Protein Solutions Inc., UK).

### 2.2.12 VapB Peptides

Two short peptides were made by IRL laboratories (N.Z.):

IRLVA3 -18 amino acids

Arg-Glu-Glu-Leu-Ala-Ala-Ile-Arg-Arg-Arg-Cys-Ala-Ala-Leu-Pro-Val-Leu-Asp

IRLVA4-18 amino acids

Asp-Asp-Arg-Thr-Ala-Glu-Ser-Ile-Leu-Gly-Tyr-Asp-Asp-Arg-Gly-Leu-Pro-Ser

Each peptide was resuspended in 50 mM phosphate buffer pH 7.4, 200 mM NaCl to a concentration of 5 mg.ml<sup>-1</sup>.

#### 2.2.12.1 Disruption of the VapBC complex

Molar ratios of peptide:VapBC complex at 1:1, 2:1, 5:1 and 10:1 for each peptide were mixed to give a final volume of 25 µl. Samples were then incubated at 37 °C, shaking for one hour. Native-PAGE loading dye was added to the sample and 20 µl loaded onto a 10% native-PAGE gel (Section 2.1.3.2) to check for disruption of the VapBC complex by one of the VapB peptides.

#### 2.2.12.2 Binding of VapB Peptides to VapC

Each peptide was combined with purified VapC and binding monitored as in Section 2.2.12.1 above. Samples were analysed by native-PAGE and by ribonuclease activity assays Sections 2.4.10.1 and 2.4.8.3.

### 2.2.13 DNA binding VapBC

#### 2.2.13.1 DNA Oligonucleotide Design

Based on Jennifer Robson's gel shift assay results with the VapBC complex from *M. smegmatis* (Robson, *et al.*, 2009), DNA oligonucleotides were designed to test DNA binding of the VapBC complex.

Two sets of oligonucleotides were designed, a positive control (MS Promoter) containing the inverted repeat to which VapBC binds and a negative control (MS Control) containing only one half of the inverted repeat and another half to which VapBC does not bind. Forward and reverse oligonucleotides were designed to anneal and form dsDNA.

Promoter sequence

```
5' GTCGTATAGATTCATGAAAGTTTGTCTATAATCA 3'
3' CAGCATATCTAAGTACTTCAAACAGATATTAGT 5'
```

Inverted repeats underlined

The negative control termed MS Control, consisted of inverted repeats for the previously annotated start site. Forward and reverse oligonucleotides were designed to anneal and form double-stranded DNA.

Control Sequence

```
5' GGCGATCTATCAAGGTCGTATAGATTCAT 3'
3' CCGCTAGATAGTTCCAGCATATCTAAGTA 5'
```

Inverted repeats underlined

### ***2.2.13.2 Annealing of DNA Oligonucleotides***

DNA oligonucleotides were annealed by one of the methods below to produce dsDNA. Equal molar ratios of forward and reverse oligonucleotides were added to either 10 x PCR buffer (Invitrogen, USA), 50 mM MgCl<sub>2</sub> to give a 1 x final concentration of PCR buffer and 8 mM MgCl<sub>2</sub> or 10 µl binding buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 150 mM NaCl). The sample was then heated to 95 °C for 5 minutes and cooled to 25 °C over 45 minutes.

### ***2.2.13.3 DNase Treatment of the VapBC Complex***

Purified VapBC complex (Section 2.2.2.2) was DNase treated by addition of 10 µg DNase (Sigma Aldrich, U.S.A) per ml of VapBC protein and incubated at 37 °C for 1, 2, 3, 4, 5 and 7 hours to establish the optimum time point for DNase digestion. A 3 hour time point was used in further DNase digestion experiments. DNase activity

---

was inhibited by addition of 25 mM EDTA. Samples were analysed by SDS-PAGE (Section 2.1.3.1) and native-PAGE (Section 2.1.3.2).

#### ***2.2.13.4 RNase A Treatment of the VapBC Complex***

Purified VapBC complex (Section 2.2.2.2) was RNase A treated by addition of 10 µg RNase A (Sigma Aldrich, U.S.A) per ml of VapBC protein and incubated at 37 °C for 3 hours. Samples were analysed by SDS-PAGE (Section 2.1.3.1) and native-PAGE (Section 2.1.3.2).

#### ***2.2.13.5 DNA Exchange***

To exchange or bind DNA to purified VapBC complex (Section 2.2.2.2) the DNA oligonucleotides above (Section 2.2.13.1) were annealed to form dsDNA (Section 2.2.13.2) and added to the VapBC complex with and without prior DNase treatment (Section 2.2.13.3) at a variety of molar ratios. Ratios of 1:10, 1:1, 4:1 and 10:1 DNA:VapBC protein were incubated at room temperature for 2.5 hours. Samples were analysed by native-PAGE (Section 2.1.3.2), DNA binding or exchange was deemed successful if there was a shift of the protein on the native polyacrylamide gel.

#### ***Biotin Labelled DNA Oligonucleotide Exchange***

The MS promoter DNA forward oligonucleotide was labelled at the 5' end with a biotin molecule (Invitrogen, USA). The oligonucleotide was then annealed to its reverse complement strand as in Section 2.2.13.2 to produce a biotin labelled dsDNA oligonucleotide. This was added to VapBC complex as in Section 2.2.13.5 above. A 90 µl sample of binding buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 150 mM NaCl) was added to 10 µl of VapBC-DNA complex which was then added to 20 µl streptavidin magnetic particles (Roche Applied Science, Switzerland) and pre-washed with binding buffer to ensure binding of the biotin labelled oligonucleotide to the particles. The sample was then incubated at room temperature for 10 - 15 minutes shaking at 550 rpm. The streptavidin magnetic particles were collected at the bottom of the tube with a magnet and the supernatant removed and kept for further analysis (unbound sample). The particles were then washed twice with 200 µl of wash buffer

---

(50 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 M NaCl). An equal volume (20  $\mu$ l) of 4 x SDS loading buffer was added to the particles. Unbound sample, washes and beads were analysed by SDS-PAGE (Section 2.1.3.1).

#### ***2.2.13.6 Visualisation of DNA Bound to VapBC***

To visualise DNA bound to the purified VapBC complex either before or after DNA exchange, a variety of methods were employed. DNA-protein interactions were disrupted by heating samples to 90 °C for 10 minutes and centrifuged at 13000 g for 10 minutes to pellet denatured protein. Both denatured and native protein were analysed by the following the methods:

##### ***Nanodrop***

To visualise the amount of DNA bound to VapBC protein, samples were analysed by a Nanodrop ND-1000 spectrophotometer according to Sections 2.1.1.6 and 2.1.3.5.

##### ***260 nm/280 nm Absorbance (Spectrophotometer)***

DNA bound to VapBC protein was also visualised using a Helios  $\gamma$  spectrophotometer (Thermospectronic, USA). A 1 ml sample was added to a quartz cuvette and absorbance measured at 260 and 280 nm to determine presence of DNA and protein.

##### ***Agarose Gel Electrophoresis***

DNA bound to VapBC protein was visualised by agarose gel electrophoresis. Protein was run on a 1% (w/v) agarose gel as in Section 2.1.1.4 and compared with the 1 kb-Plus DNA ladder (Invitrogen, U.S.A) or the 25 bp step ladder (Promega, USA).

##### ***Urea Denaturing PAGE***

DNA bound to the VapBC protein complex was visualised under urea-denaturing conditions as in Section 2.4.2.1. The denaturing conditions disrupt the interaction between the protein and DNA.

### **2.2.13.7 Exonuclease Treatment of VapBC Complex**

Purified VapBC complex was treated with exonuclease III and exonuclease VII. An aliquot containing 600 units of exonuclease III (Epicentre, USA) was added to VapBC complex along with 10 x exonuclease III reaction buffer to a final concentration of 1 x reaction buffer. The reaction was then incubated at 37 °C for 1 hour at 500 rpm. Exonuclease III reaction buffer was then removed by concentrating the reaction to one tenth its volume in a spin concentrator (Section 2.1.3.6), then bringing the volume back to the original volume with 50 mM phosphate buffer pH 7.4, 200 mM NaCl. Approximately 80 units of exonuclease VII (Epicentre, USA) was then added to the VapBC complex along with 5 x exonuclease VII reaction buffer to give a final concentration of 1 x reaction buffer. The reaction was then incubated at 37 °C for 2 hours. Exonuclease digests were analysed by SDS-PAGE (Section 2.1.3.1) and size exclusion chromatography (Section 2.1.3.11).

### **2.2.14 Initial Crystallisation Trials**

Initial crystallisation trials for VapC and VapBC proteins were set up using a medium-throughput robotic system (Moreland, *et al.*, 2005) at the School of Biological Sciences at the University of Auckland. This initial trial consisted of five in-house screens consisting of a total of 480 conditions from various screens (Moreland, *et al.*, 2005) in five 96 well Intelli-plates (Hampton Research, USA). A Multiprobe<sup>®</sup> II HTEX dispensing robot (Perkin-Elmer, USA) was used to dispense 75 µl of each condition into the reservoir wells of the Intelli-plate. A Honeybee Sitting Drop crystallisation robot (Cartesian<sup>™</sup> Dispensing Systems, USA) was used to dispense protein and reservoir solutions (100 nl each) into the crystallisation well and the plate was then sealed with Clearseal<sup>™</sup> film (Hampton Research, USA).

### **2.2.15 Optimisation of Crystallisation Conditions (Fine Screening)**

Once promising crystallisation conditions were found in the initial crystallisation trials, fine screening was used to optimise these conditions. This was carried out in a variety of ways.

### ***2.2.15.1 Sitting Drops in Crystalclear Strips***

A 50  $\mu\text{l}$  aliquot of each fine screen condition was added to the reservoir of a Crystalclear P Strip<sup>TM</sup> with no indent (Hampton Research, USA). Protein and reservoir solutions (0.5  $\mu\text{l}$  of each) were pipetted onto the stage above the reservoir well and the plate sealed with Clearseal film<sup>TM</sup> (Hampton Research, USA)

### ***2.2.15.2 Hanging Drops***

The tops of the wells of a Crystalquick 24-well plate (Greiner Bio-one, Germany) were greased with glisseal grease and 500  $\mu\text{l}$  of each fine screen condition was added to the well. One  $\mu\text{l}$  of protein and reservoir solution were pipetted onto a 22 mm siliconised glass cover slip (square) (Hampton Research, USA). The cover slip was then inverted and placed above the pre-greased well.

### **2.2.16 Microseeding**

Hanging crystallisation drops were set up according to methods above used for fine screening (Section 2.2.15.2) and allowed to equilibrate for 2 - 3 hours before seeding. A cat's whisker cleaned with ethanol and dried, was dipped into the source drop brushing existing crystals before being streaked through the new pre-equilibrated drops.

### **2.2.17 Testing of Crystals by X-Ray Diffraction**

X-ray diffraction for crystal testing was performed at the Home-Source at the School of Biological Sciences, The University of Auckland using  $\text{CuK}\alpha$  radiation ( $\lambda=1.5418 \text{ \AA}$ ). X-rays were generated by a MicroMax<sup>TM</sup> -007HF generator (Rigaku, Japan) operated at 50 kV and 100 mA with a rotating copper anode and Osmic VariMaz optics. The cryo-loop was mounted onto the goniometer using a MAR345dtb goniometer setup with an Easymount<sup>TM</sup> automatic sample mounting system (MAR Research, Germany). Crystals were kept at 110 -113 K with a cold stream of nitrogen gas produced by a Cobra cryosystem (Oxford Cryosystems, UK).

---

## **2.2.18 MALDI TOF Mass Spectrometry of Whole Protein**

### ***2.2.18.1 Sample Preparation***

Whole protein samples for analysis by MALDI-MS were prepared as follows. A 9:1 mixture of 2, 5-dihydroxybenzoic acid and 2-hydroxy-5-methoxybenzoic acid (Super-DHB) was used as a matrix as follows; 10 mg of the solid matrix was added to 30  $\mu$ l of 2:1 ACN:0.1% (v/v) TFA and mixed well. Samples were sonicated in a water-bath sonicator (Elma, Germany) for 10 minutes. The matrix was then spun at 13000 rpm for 10 minutes. Protein samples and calibrant were mixed in a 1:1 ratio (0.5  $\mu$ l:0.5  $\mu$ l) with matrix and spotted onto an AnchorChip<sup>TM</sup> MALDI-TOF target plate (Bruker Daltonics, USA) and left to air dry.

### ***2.2.18.2 MALDI-TOF Set Up***

An Autoflex<sup>TM</sup> II MALDI-TOF mass spectrometer (Bruker Daltonics, USA) was used to analyse samples. Samples were analysed in linear mode, with the mass range selector set to 'medium range', pulsed ion extraction of 450 ns, gain to 2500 V, acceleration voltage to 20 kV and a range of 5 - 20 kDa collected. Laser power was typically at around 60%. Spectra for the whole protein calibration standard II (Bruker Daltonics, USA) were first collected and the spectrometer calibrated with an automatic polynomial correction.

### ***2.2.18.3 Analysis of Mass Spectra***

Spectra were saved and exported to FlexAnalysis<sup>TM</sup> software (Bruker Daltonics, USA), and peaks identified and labelled. Peaks were confirmed to be full length protein, or variations thereof.

---

## 2.3 Methods Relating to Chapter Four: Mechanism of Growth Inhibition by VapC in *Mycobacterium smegmatis*

### 2.3.1 Strains and plasmids used

The VapC and VapBC tetracycline inducible constructs were created by Jennifer Robson according to the methods outlined in Robson *et al.* (2009) and electroporated into *M. smegmatis* mc<sup>2</sup>155 (wild-type) and *M. smegmatis* JR121 ( $\Delta vapBC$ ) strains (A table of bacterial strains and plasmids can be found in Appendix A).

### 2.3.2 Conditional expression of VapC in *Mycobacterium smegmatis*

Initial starter cultures of appropriate strains were grown in LBT to an OD<sub>600</sub> between 0.2 and 0.4 and used to inoculate a second starter culture in Hartmans de Bont (HdB) minimal medium. HdB starter cultures were then grown to an OD<sub>600</sub> between 0.1 and 0.2 and used to inoculate 100 ml of HdB minimal medium. All media were supplemented with hygromycin B (50  $\mu\text{g}\cdot\text{ml}^{-1}$ ). For all strains, growth was monitored until an OD<sub>600</sub> of 0.2 was reached and VapC expression was induced with tetracycline (20  $\text{ng}\cdot\text{ml}^{-1}$ ). Aliquots of culture were taken over time to monitor optical density and the rate of thymidine, uridine or methionine incorporation.

### 2.3.3 Pulse-Chase Methodology

*M. smegmatis* cultures were grown at 37 °C shaking at 200 rpm in 100 ml HdB minimal medium; starter cultures were prepared as in Section 2.3.2. Cells were induced with 20  $\text{ng}\cdot\text{ml}^{-1}$  tetracycline at an OD<sub>600</sub> of 0.2. Samples were taken at 3 hour intervals after induction for 24 hours, along with a pre-induction sample. At these time points 0.5 ml samples were removed and pulsed with a 20  $\mu\text{l}$  solution containing 1  $\mu\text{Ci}$  [<sup>35</sup>S]methionine, [<sup>3</sup>H]uridine or [<sup>3</sup>H]thymidine (for investigation of inhibition of translation, transcription or DNA replication, respectively) and 6.25  $\mu\text{l}$  10  $\text{ng}\cdot\text{ml}^{-1}$  unlabelled methionine, uridine or thymidine. Samples were incubated at 37 °C for 15 minutes, then chased with 50  $\mu\text{l}$  10  $\text{mg}\cdot\text{ml}^{-1}$  unlabelled methionine, uridine or thymidine for 60 minutes. Samples were then centrifuged at 13000 g for 15 minutes, the supernatant removed and the cell pellet frozen at -80 °C. Positive and negative

---

controls included addition of 50  $\mu\text{g}\cdot\text{ml}^{-1}$  chloramphenicol (global inhibitor of translation) and no addition of tetracycline respectively.

Thawed cell pellets were resuspended in 100  $\mu\text{l}$  water followed by 100  $\mu\text{l}$  40% (*w/v*) TCA then incubated on ice for 60 minutes to ensure precipitation of macromolecules (DNA, RNA and protein). Samples were then collected on a 25 mm-diameter cellulose acetate filter (pore size 0.45  $\mu\text{m}$ ; Advantec, USA) pre-wet with 20% (*w/v*) TCA. The sample tube was then washed with 0.5 ml 20% (*w/v*) TCA and added to the filter, followed by 5 ml cold 70% (*v/v*) ethanol. The filters were removed from the filtering unit and added to scintillation vials containing 5 ml Ultima Gold scintillation liquid (Packard, USA). Radioactivity was counted in a Tri Carb 2700TR Liquid Scintillation Analyser (United Technologies Packard, USA) at the Institute of Cellular & Molecular Biosciences, Newcastle University, UK.

The CPM data was processed by dividing the CPM value for each time point by the CPM value for the uninduced (pre-induction sample) and then normalising to growth (OD).

#### **2.3.4 Methionine Uptake Methodology**

*M. smegmatis* cultures were grown at 37 °C in 5 ml HdB minimal medium containing 2.5  $\mu\text{Ci}$  (0.5  $\mu\text{Ci}\cdot\text{ml}^{-1}$ ) [ $^{35}\text{S}$ ]methionine. Cells were induced with 20  $\text{ng}\cdot\mu\text{l}^{-1}$  tetracycline at  $\text{OD}_{600}$  of 0.2. Positive and negative controls were as for the post-chase methodology. Samples of 0.5 ml were taken at 30 minute intervals for 3 hours post-induction, along with a pre-induction sample. The  $\text{OD}_{600}$  of each sample was measured before adding 0.5 ml of 20% (*w/v*) ice-cold TCA to immediately precipitate macromolecules. Samples were kept on ice for the remainder of the experiment. Samples were processed as in Section 2.3.3 and radioactivity counted in a Wallac 1400 DSA ver. 2.3 Liquid Scintillation Analyser at the AgResearch Ruakura campus, Hamilton, NZ.

---

The CPM data was analysed by dividing the CPM value for each time point by the CPM value for the pre-induction sample, then normalising to growth (OD).

### 2.3.5 VapC Antibody Production

VapC protein was expressed and purified as inclusion bodies in *E. coli* as in Section 2.2.6. Antibodies to VapC proteins were produced at the AgResearch antibody facility, Ruakura, Hamilton. Antibody production was in four month old New Zealand white rabbits. A pre-immune serum sample was taken from each rabbit and then a 0.5 ml primary immunisation consisting of 0.2 ml 1 mg.ml<sup>-1</sup> VapC protein and 0.3 ml Freund's Complete adjuvant was injected subcutaneously. Every 2 weeks for 6 weeks after the primary immunisation the rabbits were injected subcutaneously with a secondary immunisation consisting of 0.1 ml 1 mg.ml<sup>-1</sup> VapC protein, 0.1 ml phosphate buffered saline (PBS) and 0.3 ml Freund's Incomplete adjuvant. Rabbits were anaesthetized with sodium pentobarbitone prior to the terminal bleed.

### 2.3.6 Total Protein Determination

Total protein was estimated via the Lowry protein estimation method (Meyers, *et al.*, 1998) as follows: One ml of culture was spun for 5 minutes at 9000 g to pellet cells, the supernatant was removed and the cell pellet washed with PBS without redissolving the pellet and spun again for 5 minutes at 9000 g. The pellet was then resuspended in 100 µl 1 M sodium hydroxide (NaOH) and heated to 99 °C for 10 minutes. 20 µl 5 M hydrochloric acid (HCl) was then added to neutralise the NaOH, followed by 880 µl PBS to make the volume up to 1 ml before spinning for 30 minutes at 9000 g to pellet cell debris. The supernatant was then used for fluorescent readings of total protein using the Quanti-iT total protein assay kit (Invitrogen, USA) for use with a Qubit Fluorometer (Invitrogen, USA). A Quanti-iT working solution was made by diluting the Quanti-iT reagent 1:200 with the Quanti-iT buffer and mixing well. To each 0.6 ml thin walled PCR tube 190 µl of working solution was added to 10 µl of standard or protein sample, vortexed and left at room temperature for 15 minutes before reading. The fluorometer was calibrated using three protein

---

standards before reading samples. Calibration of the instrument was successful if there was a greater than 30 fold difference between standards 1 and 2 and a greater than 1.4 fold difference between standards 2 and 3.

### **2.3.7 Western Blot**

Cell pellets containing 250 µg total cellular protein were subjected to SDS-PAGE along with 250 µg purified VapBC complex as a positive control and the Benchmark<sup>TM</sup> prestained protein ladder (Invitrogen, USA). Antigens and other cellular proteins were electrophoretically transferred to a nitrocellulose membrane (Whatman, UK) as follows. A transfer sandwich was set up and added to the transfer apparatus (BioRad Laboratories, USA) along with ice and a stirrer magnet, and transfer buffer added so the membrane was submerged. Transfer was conducted at 100 V for 1 hour. After transfer the membrane was removed and efficient transfer checked by staining with Ponceau for 5 minutes then rinsing with water.

The membrane was then blocked with 10% (w/v) non-fat milk powder for 1 hour at room temperature or at 4 °C overnight. A rabbit anti-VapC antibody diluted 1:750 with TBS-T + 0.5 M NaCl was added to the membrane and incubated at room temperature for 1 hour, then the membrane subjected to 5 x 5 minute washes with TBS-T + 0.5 M NaCl. After washing, a goat anti-rabbit HRP conjugate antibody diluted 1:1000 in TBS-T + 0.5 M NaCl was added to the membrane, incubated at room temperature for 1 hour and the membrane then subjected to 5 x 5 minute washes with TBS-T + 0.5 M NaCl. The membrane was treated with SuperSignal chemiluminescence substrate (Pierce, USA) at a 1:1 dilution with water. Finally the membrane was visualised using a Fujifilm intelligent dark box.

## **2.4 Methods Relating to Chapter Five: VapC Ribonuclease Activity**

### **2.4.1 VapB and VapC Protein Purification**

VapC from *M. smegmatis* was expressed, isolated and purified as in Chapter Three, Sections 2.2.2.2, 2.2.9.1 and 2.2.10. VapC<sub>P<sub>AE</sub>2754</sub> was expressed and purified as outlined in Arcus *et al.* (2004) and VapC<sub>P<sub>AE</sub>0151</sub> was expressed and purified as

---

outlined in Bunker *et al.* (2005). VapB<sub>PAE0152</sub>, the antitoxin to VapC<sub>PAE0151</sub> was expressed and purified and the PAE0151/2 VapBC complex formed as in Arcus *et al.* (2004). VapB<sub>PAE2755</sub>, the antitoxin to VapC<sub>PAE2754</sub> was expressed and purified as in (Bunker, 2005).

## 2.4.2 RNA Manipulation

### 2.4.2.1 Urea Denaturing Polyacrylamide Gel Electrophoresis Protein Analysis (Urea Denaturing-PAGE)

Urea denaturing-PAGE gels were cast in a Hoeffer gel casting system. The gels consisted solely of resolving gel (10, 15 or 20% (v/v) acrylamide depending on size of RNA to be analysed).

All urea denaturing-PAGE gels were made up with 30% (w/v) acrylamide with an acrylamide:bisacrylamide ratio of 37.5:1 (Bio-Rad Laboratories, USA) and included 6 M urea and were polymerised by addition of ammonium persulfate (APS) (0.05% (w/v)) and TEMED (0.05% (v/v)).

RNA samples were mixed in a 1:1 ratio with 2 x formamide loading dye and heated to 70 °C for 5 min before loading onto gel(s). Gels were pre-run in 1 x TBE buffer at 50 V for 30 minutes, wells were then flushed with 1 x TBE and samples loaded. Gels were then run at 150 V for one hour or until the xylene cyanol dye front reached the end of the gel.

### 2.4.2.2 Visualisation of RNA

Urea denaturing-PAGE gels were stained with ethidium bromide or using SYBR<sup>TM</sup> Green II RNA Stain (Invitrogen, U.S.A) and visualised under UV light or on a blue light box (Invitrogen, USA). Band sizes were determined by comparison with the low range ssRNA ladder (New England Biolabs, USA).

### 2.4.2.3 RNA Quantification

RNA was quantified using a Nanodrop ND-1000 Spectrophotometer (Nanodrop Technologies, USA). This measures absorbance of RNA at 260 nm, quantifying the amount of RNA.

### 2.4.3 Production of Rotavirus RNA

Purified rotavirus double layered particles (DLPs), the transcriptionally-active particles derived from infectious virions following the loss of the outer capsid layer were obtained from John Taylor, School of Biological Sciences, University of Auckland.

Rotavirus particles ( $1 \text{ mg.ml}^{-1}$ ) were dialysed (Section 2.1.3.7) in a 20  $\mu\text{l}$  dialysis button against 50 ml 100 mM Tris-HCl pH 8.0 with two buffer changes to remove EDTA and caesium chloride (CsCl). The DLPs can become quite sticky in the absence of CsCl, hence the use of a dialysis button to keep the area of dialysis membrane low.

Transcription reactions were set up in a 200  $\mu\text{l}$  volume and contained  $0.1 \text{ }\mu\text{g.}\mu\text{l}^{-1}$  rotavirus DLPs, 100 mM Tris-HCl pH 8.0, 1 mM rNTPs (rATP, rCTP, rGTP and rUTP) and 6 mM  $\text{MgCl}_2$ . A control reaction was also set up without rNTPs. Reactions were incubated at  $37 \text{ }^\circ\text{C}$  for 60 minutes and RNA recovered by addition of half the reaction volume (i.e. 100  $\mu\text{l}$ ) 7.5 M ammonium acetate and 2.5 volumes ethanol and incubated at  $-80 \text{ }^\circ\text{C}$  for one hour. RNA was pelleted by centrifugation at 13000 g,  $4 \text{ }^\circ\text{C}$  for 30 minutes, washed with 500  $\mu\text{l}$  70% (v/v) ethanol, centrifuged as above and the supernatant removed and the pellet air dried. The pellet was resuspended in 25  $\mu\text{l}$  RNase free water, nanodropped (Section 2.4.2.3) and run on a urea-denaturing PAGE gel (Section 2.4.2.1).

#### 2.4.3.1 Incorporation of $^{33}\text{P-UTP}$ into Rotavirus RNA

$^{33}\text{P}$ -labelled rotavirus RNA was made by incorporation of  $^{33}\text{P-UTP}$  into the rotavirus RNA transcription reaction. The transcription reaction was essentially carried out as

in Section 2.4.3 above with the exception of the rNTPs added to the reaction. 1 mM CTP, GTP and ATP and 0.1 mM unlabelled UTP was added to the reaction along with 2  $\mu$ l  $^{33}$ P-UTP. Reactions were incubated for the same time period and RNA recovered as in Section 2.4.3 above.

#### **2.4.4 Urea-Denaturing Polyacrylamide Sequencing Gel**

Urea-denaturing sequencing gels consisted solely of a 10% (v/v) resolving gel and were cast between large, sequencing gel glass plates. All urea denaturing-sequencing gels were made with 30% (w/v) acrylamide with an acrylamide:bisacrylamide ratio of 37.5:1 (Bio-Rad Laboratories, USA), included 6 M urea and were polymerised by addition of APS (0.05% (w/v)) and TEMED (0.05% (v/v)).

RNA samples were mixed in a 1:1 ratio with 2 x formamide loading dye and heated to 70 °C for 5 minutes before loading onto gel(s). Gels were pre-run in 1 x TBE buffer at 60 W for 30 minutes, wells were then flushed with 1 x TBE, comb inserted and samples loaded. Gels were then run at 40 W for one hour or until the xylene cyanol dye front reached the end of the gel.

Glass plates were separated and the gel transferred to pre-cut cartridge paper before drying at 80 °C for 1.5 hours on a Gel Dryer (Bio-Rad Laboratories, USA). Once dry, gels were exposed to film overnight (in the dark) and images developed in a darkroom by submersing the film in developer solution (Kodak, USA) for approximately 2 minutes then into fixer solution (Kodak, USA) for approximately 2 minutes. Films were then rinsed well with water and hung up to dry.

#### **2.4.5 Isolation of Total RNA from *Mycobacterium smegmatis***

*Method developed by Ali Ruthe, Proteins & Microbes Laboratory, University of Waikato*

*M. smegmatis* mc<sup>2</sup>155 cultures grown in LBT were incubated at 37 °C for approximately three days. Cells were immediately added to four times the volume of 5 M GITC pH 7 (typically 2 ml of cells to 8 ml GITC) and pelleted by centrifugation.

---

The supernatant was removed and cells resuspended in 0.5 ml fresh 5 M GITC and transferred to a 2 ml screw cap tube containing approximately 0.3 g of 0.1 mm and 2.5 mm zirconia beads. Beads were mixed in a Fastprep FP120 bead beater (Thermosavant, USA) for 20 seconds (setting 6.5) left to cool for one minute and the process repeated 3 times for increasing periods (i.e. 20, 25, 30 and 30 seconds). Samples were centrifuged briefly at 4 °C to remove foam. \*One tenth of the volume (50 µl) of 2 M sodium acetate pH 4.0 was added and mixed gently. An equal volume (500 µl) of water saturated phenol was added, vortexed then rotated for 10 minutes. 100 µl of 1-bromo, 3-chloro propanate was added, shaken for one minute and incubated on ice for five minutes. Samples were then centrifuged to separate the phases and the top layer removed and the method repeated from the \* above with volumes decreased accordingly. Again the top layer was removed but added to an equal volume of isopropanol and chilled on ice for 30 minutes. The precipitated RNA was then centrifuged at 13000 g, 4 °C, for 30 minutes, 1 ml 70% ethanol added, centrifuged for 15 minutes at 4 °C, supernatant removed, 1 ml 100% ethanol added, centrifuged for 15 minutes at 4 °C, supernatant removed. The pellet was then centrifuged again briefly and any residual ethanol removed. The pellet was resuspended in 25 µl 10 mM Tris-HCl pH 7.0, 0.5 mM MnCl<sub>2</sub> and DNase treated by addition of 1 µl DNase (Promega, USA) per µg RNA present and incubated at 37 °C shaking at 600 rpm for 30 minutes. The DNase reaction was stopped by addition of 1.1 µl EGTA stop solution per µl of DNase added and heated to 65 °C for 10 minutes.

#### **2.4.6 Two Base Combination RNA Oligonucleotide**

An RNA oligonucleotide was designed to include every combination of two bases, this was termed the 2 base combo RNA oligo. There are 16 possible combinations of two bases (4<sup>2</sup>). Overlapping combinations of these were determined, flanked by GAGA on each end of the oligo. The probability of the oligonucleotide forming hairpin structures was determined using Geneious Pro Version 5.02 (Biomatters, NZ). The oligo with no or very little likelihood of forming hairpin or dimer structures was chosen.

2 base combo RNA oligonucleotide:

<sup>5'</sup>GAGACUUAGCCAUGUCGGAAGAGA<sup>3'</sup>

The 2 base combo RNA oligo was end labelled using polynucleotide kinase (PK) with <sup>33</sup>P by the following reaction, 20 μM RNA oligo was added to 1 μl 10 x PK buffer, 0.5 μl PK enzyme, 1 μl <sup>33</sup>P γATP and DEPC water to a final volume of 10 μl. The reaction was incubated at 37 °C for 30 minutes and stopped by heating to 55 °C for 10 – 15 minutes. Molar ratios of 1:10, 1:100 and 1:1000 labelled to unlabelled RNA oligo were trialled in ribonuclease activity assays (Section 2.4.10).

#### **2.4.7 Fluorogenic Substrate**

*Substrate from Ashley Easter, Proteins & Microbes Laboratory, University of Waikato*

FrG was ordered from IDT (USA) at a 10 μmole synthesis scale. 6-FAM<sup>TM</sup> is a fluorophore attached via a C<sub>6</sub>N linker to the first dAMP of the oligonucleotide, BHQ-1<sup>®</sup> is a fluorescence quencher attached 3' to the final dGMP and **rG** indicates the ribo-guanidine residue in the centre of the oligonucleotide, surrounded by dNMP residues. This acts to limit possible RNA cleavage to one site.

Oligonucleotide: <sup>5'</sup>6-FAM-AAGTCrGACATCAG-BHQ-1<sup>3'</sup>

FrG was resuspended in DEPC to water to give a final concentration of 1 mM. Dilutions required for the assay reactions were made up to the final volume in the appropriate assay buffer.

##### **2.4.7.1 Fluorometer Setup**

A Hitachi F-7000 fluorescence spectrophotometer (Hitachi High Technologies, Japan) fitted with a temperature controlled Hitachi cell holder along with 5 mm path-length thick-walled quartz cuvettes was used for all assays. A sample volume of

---

300  $\mu\text{l}$  was used for all experiments, time-scans used a 0.5 second data pitch and were recorded for 3 minutes with PMT set to 450 V, excitation (ex.) 485 nm, emission (em.) 518 nm, 5 nm slits and 2 millisecond response time. Three-dimensional (3D) scans of the substrate with ex. 445 - 495 nm and em. 505 - 555 nm was performed to determine the optimal ex. and em. wavelengths for 10  $\mu\text{M}$  FrG.

#### **2.4.8 The Pentaprobe System**

*The pentaprobe system was obtained from Joel McKay, School of Molecular Bioscience, University of Sydney.*

The pentaprobe sequence covers every combination of five bases, and was made as six overlapping dsDNA plasmid molecules (Kwan, Czolij, MacKay, & Crossley, 2003). RNA can be transcribed from the plasmids, giving RNA segments covering every combination of five bases. Pentaprobe sequences can be found in Appendix B.

The pentaprobe pack consists of 12 plasmids. Each single strand of pentaprobe was cloned under a T7 promoter in pcDNA3 in Joel McKay's laboratory, University of Sydney. The complementary strands have also been cloned under the same promoter in pcDNA3. 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 with the exception of A926 and A932 were cloned into pcDNA3 using KpnI and XbaI restriction sites. Due to cloning difficulties, inserts A926 and A932 were cloned into pcDNA3 using blunt-end cloning and the EcoRV restriction site. This results in a larger gap or "linker" before pentaprobe sequences start compared with inserts cloned using KpnI and XbaI restriction sites.

##### **2.4.8.1 Pentaprobe Plasmid Preparation**

The set of 12 pentaprobe plasmids were electroporated into *E. coli* DH5 $\alpha$  electrocompetent cells (Section 2.1.1.13). Colonies from transformations were used to inoculate three 5 ml LB cultures including 100  $\mu\text{g}\cdot\text{ml}^{-1}$  ampicillin for each

---

transformation and incubated at 37 °C, 200 rpm overnight for subsequent plasmid extraction (Section 2.1.1.2).

#### **2.4.8.1.1 Restriction Enzyme Digest of Pentaprobe Plasmids**

Pentaprobe plasmids were linearised by digestion with the *ApaI* restriction enzyme according to Section 2.1.1.9 and subsequently purified using the Qiagen PCR purification kit as in Section 2.1.1.8.

To fill in the overhang left by *ApaI* digestion, 1 µl Klenow enzyme (2 U.µl<sup>-1</sup>) (Roche Applied Science, Switzerland), 2 µl 10 x klenow filling buffer, 2 µl 10 mM dNTPs and 5 µl of water was added to 10 µl of linearised pentaprobe pcDNA3 plasmid and incubated for 15 minutes at 37 °C at 300 rpm. The reaction was purified using the Qiaquick PCR purification kit as in Section 2.1.1.8.

#### **2.4.8.2 PCR Amplification of Pentaprobe Inserts**

Primers were designed as in Section 2.1.1.7. The forward primer was designed to flank the T7 promoter in pcDNA3 and the reverse to flank the end of the pentaprobe insert, ensuring amplification of the T7 promoter with the pentaprobe insert.

The pentaprobe insert was PCR amplified using the *Pfx*<sup>®</sup> DNA polymerase system (Invitrogen, USA) as in Section 2.1.1.7 and gel purified as in Section 2.1.1.5.

#### **2.4.8.3 In Vitro Transcription of Pentaprobe Inserts**

RNA was transcribed from the linearised plasmid or pentaprobe PCR product (Section 2.4.8.2 above) using the T7 MEGAscript<sup>®</sup> kit (Ambion, USA). The reaction consisted 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 to a final volume of 20 µl then incubated at 37 °C for 4 hours (time determined by time course transcription reaction). TURBO DNase (1 µl) was then added and the reaction incubated at 37 °C for 15 minutes, shaking at 600 rpm.

---

Transcribed RNA was purified by sodium acetate and ethanol precipitation as in Section 2.4.11.1. Secondary structure of RNA molecules was predicted using RNA Structure Version 4.6 (Mathews, 2006).

#### ***2.4.8.4 Producing Pentaprobe Double-Stranded RNA***

Single-stranded pentaprobe RNA was annealed with its reverse complement RNA strand (i.e. 927 with 933) to form dsRNA. Equal molar ratios of each strand were combined with nuclease free water to give a final volume of 5  $\mu$ l, incubated at 70 °C for 15 minutes and then cooled slowly to room temperature. Annealed RNA was run on an agarose gel (Section 2.1.1.4) and stained with acridine orange (Section 2.4.9) to distinguish between ds- and ssRNA.

#### ***2.4.8.5 Producing Pentaprobe Single Stranded DNA by Asymmetric PCR***

Asymmetric PCR was used to generate ssDNA from dsDNA by adjusting the ratio of forward and reverse primers to preferentially amplify one strand. A single-strand of pentaprobe DNA was amplified from the template pentaprobe PCR product (Section 2.4.8.2) using the *Pfx*<sup>®</sup> DNA polymerase system as in Section 2.1.1.7. However 1:50 and 1:100 reverse: forward primer ratios were used instead of 1:1 ratios and 35 cycles employed to increase amplification. PCR reactions were analysed by agarose gel electrophoresis (Section 2.1.1.4) by staining with acridine orange (Section 2.4.9). Single-stranded DNA products were gel purified according to the method in Section 2.1.1.5.

#### ***2.4.8.6 Design of 932 Pentaprobe Oligonucleotides***

RNA oligonucleotides were designed to cover bases 54 - 242 of pentaprobe 932 RNA. Nine overlapping oligonucleotides of 30 – 36 bases were designed to cover the 188 base RNA segment. RNA Structure Version 4.6 (Mathews, 2006) was used so secondary structure of the oligonucleotides mimicked their region on the whole 932 pentaprobe RNA molecule.

### 2.4.9 Acridine Orange Nucleic Acid Staining

Acridine orange hemi (zinc chloride) salt (Sigma Aldrich, USA) is a fluorescent pH indicator that differentially stains single-stranded and double-stranded nucleic acids. Agarose and polyacrylamide gels were stained with  $30 \mu\text{g}\cdot\text{ml}^{-1}$  acridine orange solution (acridine orange dissolved in 10 mM sodium phosphate monohydrate pH 7.0) for 15 minutes at room temperature in the dark. Gels were destained by running under hot water for 30 minutes before visualising on a UV light box. Double-stranded nucleic acids stain bright, fluorescent green whereas single-stranded nucleic acids stain a bright reddish-orange.

### 2.4.10 VapC Ribonuclease Activity Assay General Method

#### 2.4.10.1 *Mycobacterium smegmatis* VapC Ribonuclease Assay

##### Assay Reactions (Time Course)

Ribonuclease activity assays for VapC<sub>MS1284</sub> from *M. smegmatis* were set up as follows: 1  $\mu\text{g}$  purified VapC<sub>MS1284</sub> protein and 1  $\mu\text{g}$  purified RNA (concentration can vary), 6  $\mu\text{l}$  assay buffer (20 mM phosphate buffer pH 7.4, 20 mM NaCl, 10 mM MgCl<sub>2</sub>\*), was added to nuclease free water to give a final volume of 10  $\mu\text{l}$ . Individual assay reactions were set up for each time point to reduce the possibility of RNase contamination i.e. four time points, four assay reactions. Assays 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<sub>2</sub>.

##### Negative Controls

###### *RNA Only Negative Control (0 hour and 1 hour)*

Assay reactions were set up as above but no VapC<sub>MS1284</sub> was added to the reaction, to ensure there is no RNase contamination of the assay buffer. Two reactions were set up, a 0 hour reaction which was stopped immediately and a 1 hour reaction which was incubated at 37 °C for 1 hour (or the assay duration).

---

*EDTA Negative Control*

Assay reactions were set up as above but 20 mM EDTA was added to the reaction, to ensure there is no RNase contamination of the protein purification. One reaction was set up and incubated at 37 °C for 1 hour (or the assay duration).

*VapBC Negative Control*

Assay reactions were set up as above but purified VapBC complex was added to the reaction instead of purified VapC<sub>MS1284</sub>. One reaction was set up and incubated at 37 °C for 1 hour (or the assay duration).

Time course assay reactions were stopped by addition of 10 µl formamide loading dye, or for analysis by mass spectrometry assays were heat inactivated by either incubating at 70 °C for 10 minutes or inactivated by the addition of salt and ethanol for precipitation (Section 2.4.11.1).

**2.4.10.2 *Pyrobaculum aerophilum* VapC Ribonuclease Assay***Assay Reactions (Time Course)*

Ribonuclease activity assays for VapC proteins from *P. aerophilum* were carried out as follows: 1 µg purified VapC<sub>PAE2754</sub> or VapC<sub>PAE0151</sub> protein and 1 µg purified RNA (concentration can vary), 6 µl assay buffer (20 mM Tris-HCl pH 9.2, 20 mM NaCl, \*10 mM MgCl<sub>2</sub>) added to nuclease free water to a final volume of 10 µl. Individual assay reactions were set up for each time point to reduce the possibility of RNase contamination i.e. four time points, four assay reactions. Assays were incubated at 37 °C for the appropriate time (time course).

\* N.B. 10 mM or 1 mM MnSO<sub>4</sub> was used as a cofactor in early ribonuclease activity assays instead of MgCl<sub>2</sub>.

*Negative Controls*

Negative controls for the *P. aerophilum* VapC ribonuclease activity assays were the same as the *M. smegmatis* VapC activity assays except the assay buffer. The assay

---

buffer above (20 mM Tris HCl pH 9.2, 20 mM NaCl, 10 mM MgCl<sub>2</sub>) was used in the *P. aerophilum* activity assays.

## **2.4.11 MALDI TOF Mass Spectrometry of RNA Oligonucleotides**

### ***2.4.11.1 Sample Purification***

932 RNA oligonucleotides 3, 4 and 5 were ordered HPLC purified from Sigma Aldrich (USA). RNA oligonucleotides and assay reactions need to be desalted for use in MALDI-TOF MS so a variety of RNA cleanup methods were trialled.

#### ***Sodium Acetate & Ethanol Precipitation***

RNA samples were precipitated to remove contaminants by addition of one tenth the volume of RNA of 3 M sodium acetate pH 5.2 and three times the volume 100% ethanol. Samples were precipitated on ice for 30 minutes, centrifuged at 13000 g for 20 minutes at 4 °C then washed with 70% (v/v) ethanol and centrifuged as above. The supernatant was then removed and the RNA pellet resuspended in nuclease free water and stored at -80 °C.

#### ***Ammonium Acetate & Ethanol Precipitation***

Ammonium acetate (5 M) was added to each assay reaction to give a final concentration of 2 M. Three times the volume ethanol was added and RNA precipitated as in the method for sodium acetate and ethanol precipitation above.

#### ***Lithium Chloride & Ethanol Precipitation***

Lithium chloride (7.5 M) was added to each assay reaction to give a final concentration of 0.8 M. Three times the volume ethanol was added and RNA precipitated as in the method for sodium acetate and ethanol precipitation above.

#### ***ZipTip<sup>TM</sup> Cleanup of RNA***

ZipTips (Eppendorf, Germany) were used according to manufacturer's instructions to cleanup RNA oligonucleotides for MALDI-MS. RNA was eluted using the indirect elution method.

---

### ***Cation Exchange Beads***

A 100  $\mu$ l aliquot of SP sepharose fast flow cation exchange beads (GE Healthcare, UK) were washed with 1 ml DEPC water to remove the storage solution (0.2 M sodium acetate, 20% (v/v) ethanol). The beads were centrifuged briefly to collect in the bottom of the tube, supernatant removed and beads washed with 1 ml 1 M diammonium citrate, centrifuged, supernatant removed and the diammonium citrate wash repeated a further two times. Beads were washed with 1 ml 50 mM diammonium citrate in the final wash, centrifuged, supernatant removed and beads resuspended in 50  $\mu$ l 50 mM diammonium citrate.

A 4  $\mu$ l sample of the ammonium ion loaded cation exchange beads were added to heat inactivated assay reactions, incubated for 10 minutes at room temperature shaking at 850 rpm, then centrifuged to collect beads at the bottom of the tube.

#### ***2.4.11.2 Sample Preparation***

Matrix for MALDI MS of RNA oligonucleotides was prepared fresh daily. The matrix solution consisted of 5 mg of 3-hydroxypicolinic acid (3-HPA), 10  $\mu$ l 2.5 M diammonium citrate, 125  $\mu$ l of acetonitrile (ACN) and 365  $\mu$ l of DEPC water. The solution was vortexed well until all matrix dissolved then centrifuged for 5 minutes at 13000 g.

A 1  $\mu$ l aliquot of matrix solution was spotted onto either an Anchorchip<sup>TM</sup> or stainless steel target plate (Bruker Daltonics, USA) and left to air dry. A 1  $\mu$ l sample and 1  $\mu$ l of the oligonucleotide calibration standard (Bruker Daltonics, USA) were spotted onto dried matrix and left to air dry.

#### ***2.4.11.3 MALDI-TOF Set Up***

An Autoflex<sup>TM</sup> II MALDI-TOF mass spectrometer (Bruker Daltonics, USA) was used to analyse samples. Samples were analysed in linear mode, with the mass range selector set to 'low range', pulsed ion extraction of 150 ns, gain to 2500 V, acceleration voltage to 20 kV and a range of 2 - 12 kDa collected. Laser power was

---

typically at around 60%. Spectra for the oligonucleotide calibration standard (Bruker Daltonics, USA) were collected first and the spectrometer calibrated with an automatic polynomial correction.

#### ***2.4.11.4 Analysis of Mass Spectra***

Spectra were saved and exported to FlexAnalysis<sup>TM</sup> software (Bruker Daltonics, USA) and peaks identified and labelled. Spectra were also analysed in DataAnalysis (Bruker Daltonics, USA) and the baseline subtracted. The online tool Mongo Oligonucleotide Calculator Version 2.06 (Rozenski, 1999) was used to determine theoretical masses of RNA oligonucleotide fragments to compare to masses from spectra. The Mongo Oligonucleotide Calculator allowed the addition of 5' and 3' modifications to the RNA oligonucleotide.

---

---

## Chapter Three: Protein Expression, Purification & Characterisation

### 3.1 Introduction

This chapter focuses on the expression and purification of VapBC and VapC proteins; VapBC for crystallisation attempts and characterisation of the VapBC complex, and VapC for functional studies. As other groups have found (Mattison, *et al.*, 2006; Miallau, *et al.*, 2009; Ramage, *et al.*, 2009) it is difficult to express VapC as a soluble protein due to its toxicity. Therefore a variety of methods were trialled to isolate VapC for subsequent functional studies. The VapBC protein complex was characterised and used for DNA binding studies in the laboratory of collaborators (Robson, *et al.*, 2009).

#### 3.1.1 Protein Expression in *Mycobacterium smegmatis*

One of the major problems with expressing proteins in *Escherichia coli* is the formation of insoluble protein (Baneyx, 1999). Earlier work on the VapBC proteins from *M. smegmatis* showed the proteins to be insoluble when expressed in *E. coli* (Summers, 2007). These early expression trials of the VapB, VapC and VapBC proteins also revealed that *E. coli* does not recognise the one base pair overlap between *vapB* and *vapC* genes in the operon. Expression of the *vapBC* operon resulted in insoluble VapB protein only. (Summers, 2007). One strategy to overcome the formation of insoluble protein is to express the proteins in a host that is more closely related to the organism from which the protein is derived. *M. smegmatis* is a non-pathogenic mycobacterial species which is relatively fast growing with a doubling time of approximately three hours (Snapper, Melton, Mustafa, Kieser, & Jacobs, 1990) and will most likely have mycobacterial chaperones to aid protein folding, making it an ideal host for expression of mycobacterial proteins. Using *M. smegmatis* strain mc<sup>2</sup>4517 as an expression host, the mycobacterial expression vector pYUB1049 typically yields around 7 mg of soluble protein per litre of culture (Bashiri, Squire, Baker, & Moreland, 2007). *M. smegmatis* can be cultured in

---

---

autoinduction media using protocols developed for *E. coli* giving advantages of decreased handling (cultures do not require monitoring by optical density and induction) and also removes the cost of the ADC enrichment traditionally used for growth of mycobacterial cultures (Bashiri, *et al.*, 2007). Selenomethionine substituted proteins can also be expressed using this method, making multi-wavelength anomalous diffraction (MAD) an option for determining the structures of mycobacterial proteins.

### **3.1.2 Protein Expression Using an MBP-Fusion Protein in *Escherichia coli***

Expression of toxic proteins fused to another protein can promote soluble expression of the toxic protein. Expression of VapC as a fusion protein could overcome toxicity issues (due to ribonuclease activity) and prevent the formation of insoluble protein. There are many examples of fusion proteins facilitating solubility, including thioredoxin (TRX), glutathione S-transferase (GST), maltose binding protein (MBP), Protein A, Ubiquitin and DsbA. MBP has been shown to be a better solubilising agent than GST or TRX (Kapust & Waugh, 1999). MBP has a beneficial effect on the folding of some proteins in *E. coli*; it is proposed that it functions as a general molecular chaperone when in the context of a fusion protein (Kapust & Waugh, 1999). The solubility of MBP fusion proteins in *E. coli* is dependent on the order in which the two domains are synthesised. For soluble protein expression the passenger protein must be fused to the C-terminus of MBP (Sachdev & Chirgwin, 1998). Studies by Ramage *et al.* (2009) showed that solubility of mycobacterial VapC proteins can be increased by expression in *E. coli* as an MBP fusion protein (Ramage, *et al.*, 2009). Expression of VapC proteins as MBP fusions would also potentially overcome toxicity issues associated with expressing these proteins. The active VapC toxin is thought to form a dimer, therefore the MBP-VapC fusion would prevent dimerisation of the toxin thereby reducing its toxicity. The affinity of MBP for an immobilised ligand such as amylose can be used for purification of the MBP fusion protein, but binding can occur with low affinity and can be disrupted by proteins bound to MBP (Baneyx, 1999). Inclusion of a protease site between MBP and the target protein allows for deliberate cleavage of the MBP after purification. This is

---

---

essential as often the presence of MBP bound to the target protein will inhibit its activity. Cleavage of MBP from the target protein yields soluble protein for use in functional studies.

### **3.1.3 Protein Refolding**

Previous results have shown that VapC expression in *E. coli* results in insoluble aggregates. To identify the function of a protein it must be in the native and active form. Active protein can be recovered by solubilising the aggregated protein by denaturation and then refolding of the denatured protein. Careful screening and examination of refolding conditions can promote refolding to active protein. However, refolding is not straight forward and requires an extensive trial and error approach. Proteins can be denatured using chaotropic agents such as urea or guanidine HCl, thiocyanate salts or detergents like SDS (De Bernardez Clark, 1998). In most cases 6 – 8 M urea and 6 - 7 M guanidine HCl are needed to unfold and solubilise proteins (Tsumoto, Ejima, Kumagai, & Arakawa, 2003). Methods for refolding include one-step dialysis, step-wise dialysis, gel filtration, dilution, mixing and solid phase refolding (De Bernardez Clark, 1998; Tsumoto, *et al.*, 2003). One-step dialysis is where denatured, unfolded protein samples are dialysed against a refolding buffer. Over time the concentration of denaturant decreases and the concentration of refolding buffer increases, thereby promoting folding into the intermediate and native structures (Tsumoto, *et al.*, 2003). Refolding buffers must also promote correct formation of disulphide bonds. Glutathione (GSH), glutathione disulfide (GSSG) and dithiothreitol (DTT) are commonly used to assist oxidation/reduction thereby allowing disulphide bonds to break and reform.

### **3.1.4 Auto-regulation of the VapBC Family**

Many VapB proteins show homology to RHH domains. RHH domains are similar to transcription factors and their general function is to confer sequence-specific DNA binding.

---

---

***DNA Binding of FitAB (VapBC from Neisseria gonorrhoeae)***

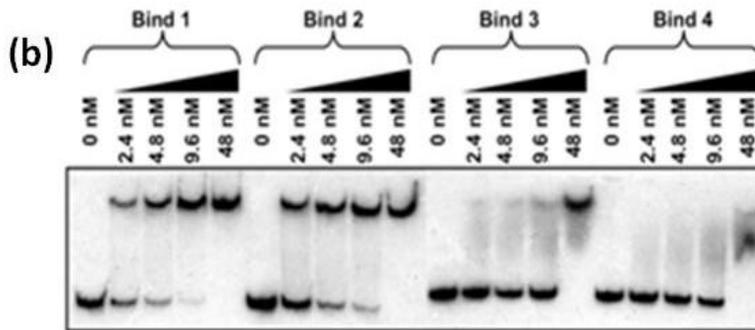
The FitAB complex from *N. gonorrhoeae* has been shown to bind to its own promoter DNA through FitA, involving the RHH domain. FitB interacts with FitA increasing its DNA binding activity and greatly enhancing its solubility (Wilbur, *et al.*, 2005). The FitAB complex binds to the inverted repeat (IR) TGCTATCA and its partner TGATAGCA separated by 13 base pairs, which is termed IR36 and is found in the promoter region of *fitAB*, thereby regulating its own transcription. The crystal structure of FitAB bound to IR36 was published in 2006 and revealed that four FitAB heterodimers form a tetramer that binds to an inverted repeat in the *fitAB* promoter region with high affinity, through the RHH motif of FitA (Mattison, *et al.*, 2006).

***DNA Binding of VapBC from Mycobacterium smegmatis***

Electrophoretic mobility shift assays (EMSAs) with purified VapBC complex were conducted by collaborators and used to determine which of the inverted repeat (IR) sequences present in the *vapBC* promoter region are involved in VapBC DNA binding (Robson, *et al.*, 2009). VapBC binds to inverted repeat 1 (IR-1), shown by a gel shift (Figure 3.1). Mutation of each half site of IR-1 resulted in a loss of the gel shift, however shifts at high VapBC concentrations were observed, reflecting weaker binding of VapBC to half the inverted repeat. The inverted repeat sequence in the *vapBC* promoter overlaps the -35 and -10 sequences (Figure 3.1) implying interference in RNA polymerase binding.



**Figure 3.1** *In vitro* DNA binding of purified VapBC complex. (a) Schematic of the *vapBC* promoter region. Bind 1 indicates the sequence used for 'Bind 1', Bind 2, 3, and 4 indicate mutation of possible binding sequences. Base pair changes overlapping potential VapBC binding sites. Divergent arrows show sequences for IR-1 and IR-2, CCCCCC above the sequence indicates nucleotide changes for each half of IR-1 and IR-2. Bold nucleotides indicate -10 and -35 promoter elements. +1 indicates the transcriptional start site and the start codon of VapB. (b) EMSA experiments with increasing concentrations of VapBC with DNA fragments 'Bind 1 - Bind 4'. Figure adapted from Robson *et al.*



Indeed, the transcriptional activity of the *vapBC* promoter of *M. smegmatis*  $\Delta$ *vapBC* has three fold more transcriptional activity compared to that of *M. smegmatis* mc<sup>2</sup>155, indicating that the *vapBC* operon is subject to auto-regulation (Robson, *et al.*, 2009).

### 3.1.5 Protein Crystallisation

The three-dimensional (3D) structure of a protein is intrinsically linked to its function. Therefore, the determination of a 3D structure of a protein can give invaluable insights into the molecular details of its function and molecular interactions with other proteins. High resolution protein structures can be determined by two main methods, nuclear magnetic resonance spectroscopy (NMR) and X-ray crystallography. NMR can solve structures of proteins in solution but is limited to smaller proteins of <30 kDa. X-ray crystallography requires the growth of protein crystals, but is not limited by the size of the protein.

---

## 3.2 Results

### 3.2.1 Cloning of *vapB*, *vapC* and *vapBC* into pYUB1049

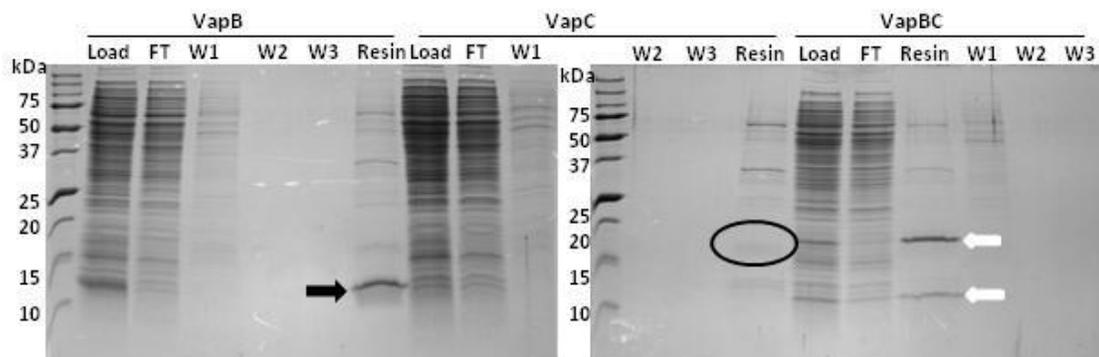
The *vapB*, *vapC* and *vapBC* genes were successfully cloned into the *E. coli* - mycobacteria shuttle vector, pYUB1049, between the NcoI and BamHI restriction sites, enabling expression of protein with a C-terminal His tag.

Ligation reactions were first transformed into electrocompetent *E. coli* TOP10 cells as this strain gives reliable, reproducible results for selection of hygromycin B resistance compared to other *E. coli* strains such as DH5 $\alpha$ . Ligations were transformed into *E. coli* cells first to facilitate plasmid purification and sequencing before transforming into *M. smegmatis* mc<sup>2</sup>4517 cells. Purification of plasmids from *M. smegmatis* is difficult due to the waxy nature of their cell walls and hence the initial transformation into *E. coli*.

Successful insertion of the *vapB*, *vapC* and *vapBC* genes into pYUB1049 was verified by extracting the plasmid from positive *E. coli* transformants and sequencing the plasmid using T7 promoter and T7 terminator primers. These primers flank the multiple cloning site and result in sequences in both the forward and reverse direction.

### 3.2.2 Small Scale Expression Tests of VapB, VapC and VapBC

Small scale protein expression tests were used to determine insoluble or soluble protein expression of the VapB, VapC and VapBC proteins in *M. smegmatis*. Small scale expression and His-tag binding tests were conducted to check binding affinity. Soluble expression and His-tag binding for VapB and the VapBC proteins was successful (Figure 3.2), but no expression of VapC, the toxin protein was seen (Figure 3.2).

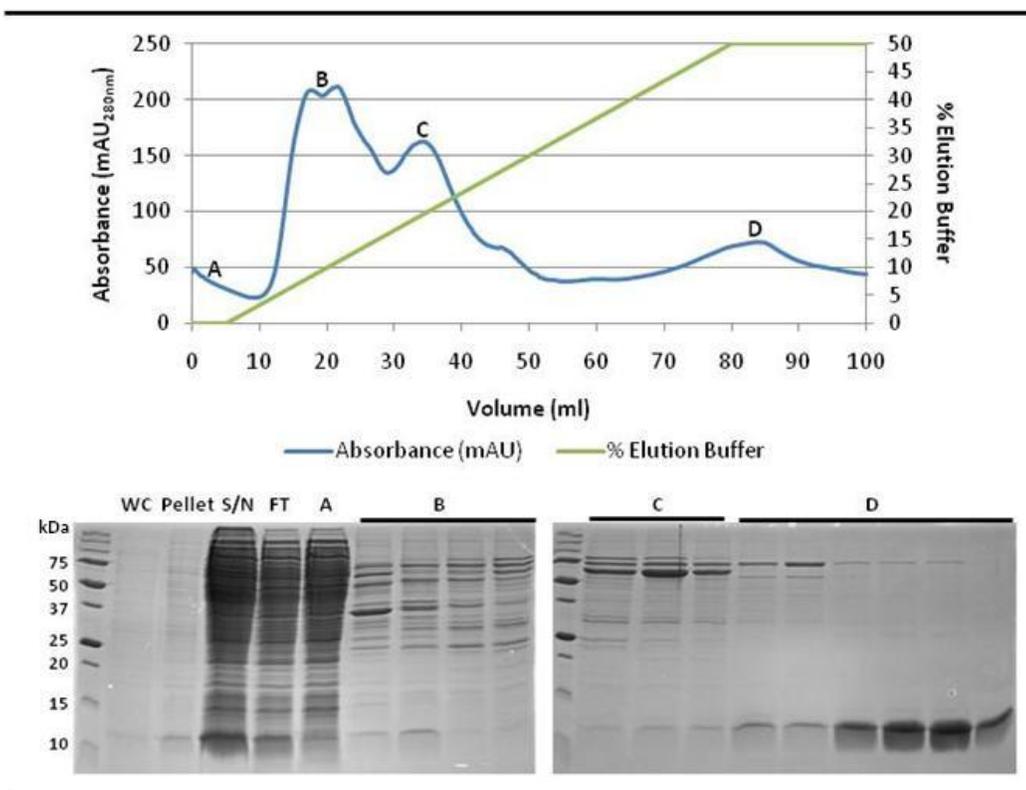


**Figure 3.2. Small scale expression screen of VapB, VapC and VapBC.** Expression and His-tag binding of VapB (black arrow) and VapBC (white arrows) but no expression of VapC observed (empty black circle). Load, soluble fraction loaded onto the Ni resin, FT, flow through, W1, wash one, W2, wash two etc. Resin, Ni resin with protein bound after wash steps.

It was assumed that there was no expression of VapC, due to its toxicity. When expressed with its inhibitor, the VapB antitoxin, expression of both the VapC and VapB proteins was observed (Figure 3.2).

### 3.2.3 VapB Expression & Purification in *Mycobacterium smegmatis*

Soluble VapB was observed in small scale expression tests (Figure 3.2). Expression cultures were subsequently scaled up (Section 2.2.2.2) and protein purified by IMAC. The chromatogram and corresponding SDS-PAGE gel in Figure 3.3 depict the purification, column load and column flow through fractions. Non-specific proteins bind to the column but are eluted early in the elution gradient, whereas VapB eluted late in the gradient at around 50% elution buffer, corresponding to ~500 mM imidazole.

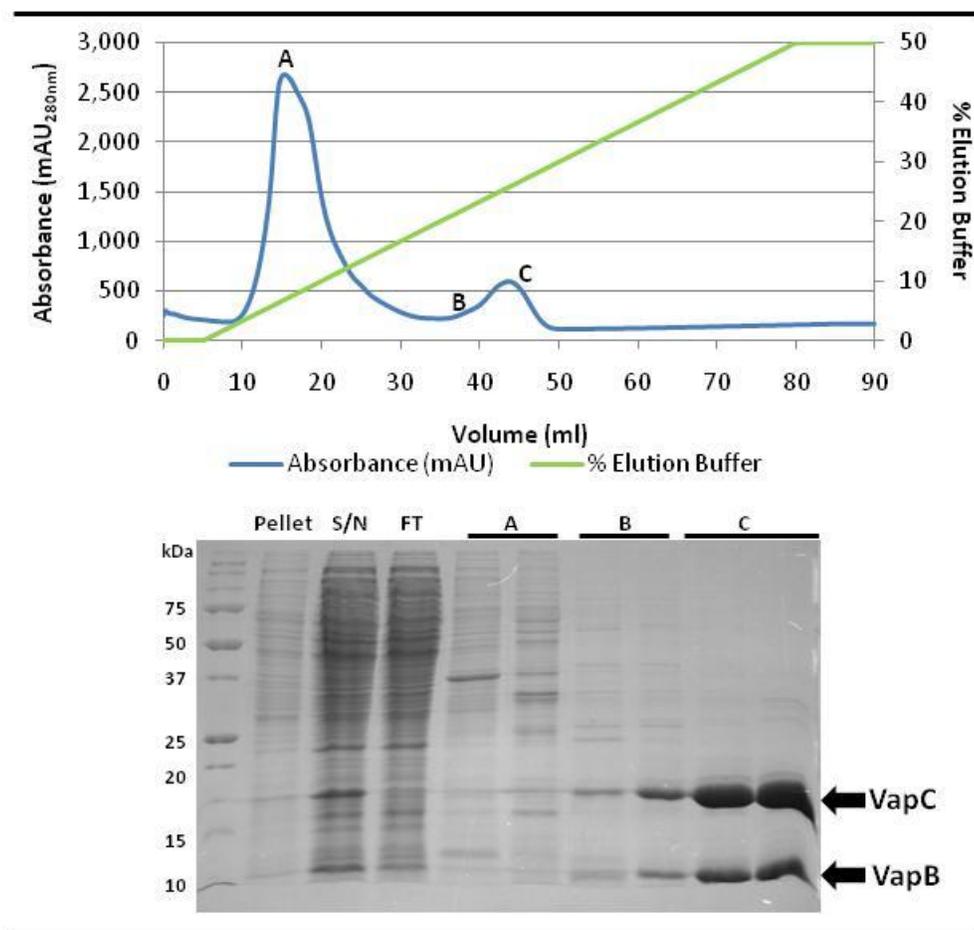


**Figure 3.3. VapB IMAC purification and corresponding 15% SDS-PAGE gel.** The chromatogram depicts the UV absorbance and elution profile of VapB. The SDS-PAGE gel denotes total cellular protein (WC), insoluble protein (pellet), soluble protein (supernatant, S/N) and column flow through (FT). Fractions and their elution position are shown by letters A-D.

VapB was further purified by size exclusion chromatography (Section 2.1.3.11). VapB eluted in the void volume of a S200 gel filtration column, indicating that the protein is aggregating during the purification process.

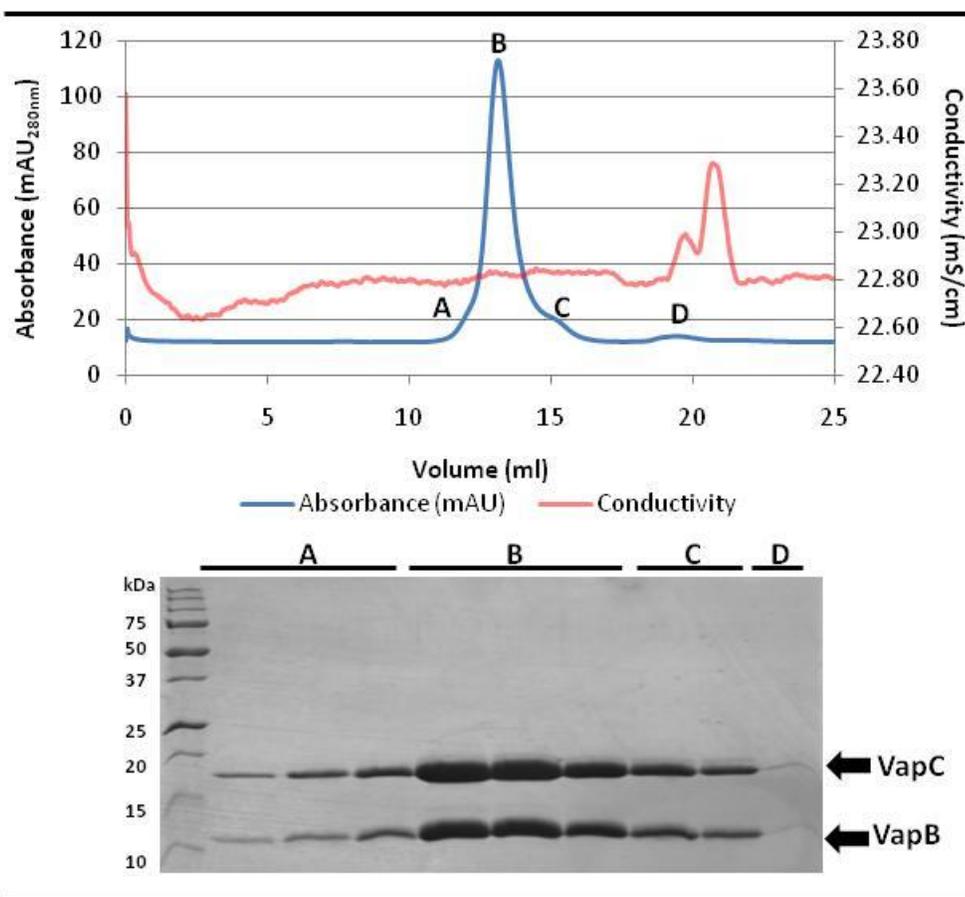
### 3.2.4 VapBC Expression & Purification in *Mycobacterium smegmatis*

Soluble expression of the VapBC protein complex was observed in small scale expression trials (Figure 3.2) so expression cultures were scaled up (Section 2.2.2.2). Soluble VapBC expression was also observed in large scale expression cultures which allowed the VapBC complex to be purified via IMAC (Section 2.1.3.10) (Figure 3.4). Non-specific proteins eluted early in the gradient, VapBC eluted later at ~250 - 300 mM imidazole.



**Figure 3.4. VapBC IMAC purification and corresponding 15% SDS-PAGE gel.** The chromatogram depicts the UV absorbance and elution profile of VapBC. The SDS-PAGE gel denotes insoluble protein (pellet), soluble protein (supernatant, S/N) and column flow through (FT). Fractions and their elution position are shown by letters A-C. VapB and VapC proteins are shown on the gel.

The VapBC protein complex was further purified by gel filtration chromatography. The complex elutes as a single peak on an S200 10/300 analytical size exclusion column (Figure 3.5).



**Figure 3.5. VapBC size exclusion purification and corresponding 15% SDS-PAGE gel.** The chromatogram depicts the UV absorbance and conductivity profile of VapBC. The SDS-PAGE gel denotes fractions and their elution position is shown by letters A-D.

### 3.2.5 Determination of the Molecular Weight of the VapBC Complex

#### *Size Exclusion Chromatography*

Calibration of the S200 10 300 size exclusion column using the high and low molecular weight Gel Filtration Calibration kits (GE Healthcare, UK) allowed the construction of a calibration curve (Figure 3.6) and the subsequent calculation of the molecular weight of the VapBC complex. The elution volume of the VapBC complex was 13.1 ml (Figure 3.5). This value was substituted into the equation below:

$$K_{av} = (V_e - V_o) / (V_c - V_o)$$

Where  $V_e$  = elution volume

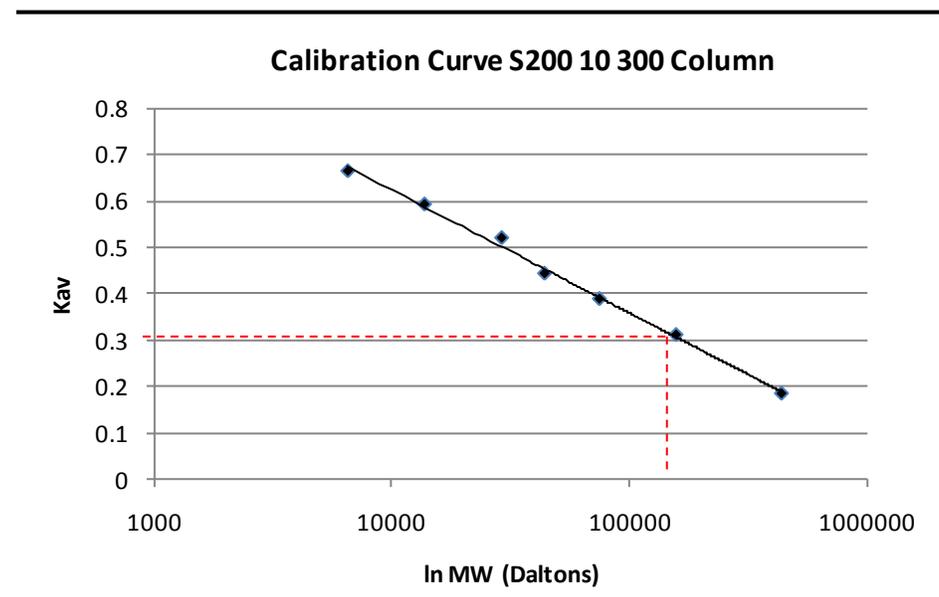
$V_o$  = column void volume

$V_c$  = geometric column volume

$$K_{av} \text{ VapBC} = (13.1 \text{ ml} - 8.11 \text{ ml}) / (24 \text{ ml} - 8.11 \text{ ml})$$

$$= 0.31$$

The molecular weight of the VapBC complex was then calculated using the formula from the calibration curve below (Figure 3.6).



**Figure 3.6. Calibration curve for the S200 10 300 size exclusion column.** Kav calculated according to formula below. Dashed red line indicates position of the VapBC complex on standard curve.

The equation from the calibration curve (Figure 3.6):

$y = -0.1151n(x) + 1.6833$  was rearranged to calculate the molecular weight (MW) of the VapBC complex:

$$MW = e^{((K_{av} - 1.6833) / -0.115)}$$

$$MW_{\text{VapBC}} = e^{((0.31 - 1.6833) / -0.115)}$$

$$= 155\,735.3 \text{ Da}$$

The calculated molecular weight of the VapBC complex is 156 kDa.

### ***Dynamic Light Scattering of the VapBC complex***

Dynamic light scattering of purified VapBC complex revealed a hydrodynamic radius of 51 Å and from this, the calculated molecular weight of the VapBC complex is

approximately 152 kDa. In addition the DLS showed a high degree of monodispersity implying that VapBC is homogeneous in solution.

DLS and size exclusion chromatography both overestimate the molecular weight of the VapBC complex due to the assumption of a spherical particle in the molecular mass calculation (Wen, Arakawa, & Philo, 1996). However the hydrodynamic radius of the VapBC complex is in accordance with the hydrodynamic radius of the crystal structure of FitAB from *N. gonorrhoeae* bound to DNA (Mattison, *et al.*, 2006). This suggests that VapBC from *M. smegmatis* is a tetramer of VapBC heterodimers bound to an approximately 36 bp segment of DNA.

### 3.2.6 Overexpression of VapB, VapC and VapBC does not affect growth of *Mycobacterium smegmatis*

To investigate the toxicity of VapC and the effect of VapBC expression in *M. smegmatis*, growth of the *M. smegmatis* mc<sup>2</sup>4517 strains harbouring pYUB1049 (empty vector), pYUBMS1283 (VapB), pYUBMS1284 (VapC) and pYUBMS1283/4 (VapBC) constructs were compared in ZYP-5052 auto-induction media. Cultures were set up in triplicate and OD<sub>600</sub> readings were taken every 6 hours over a 96 hour period.

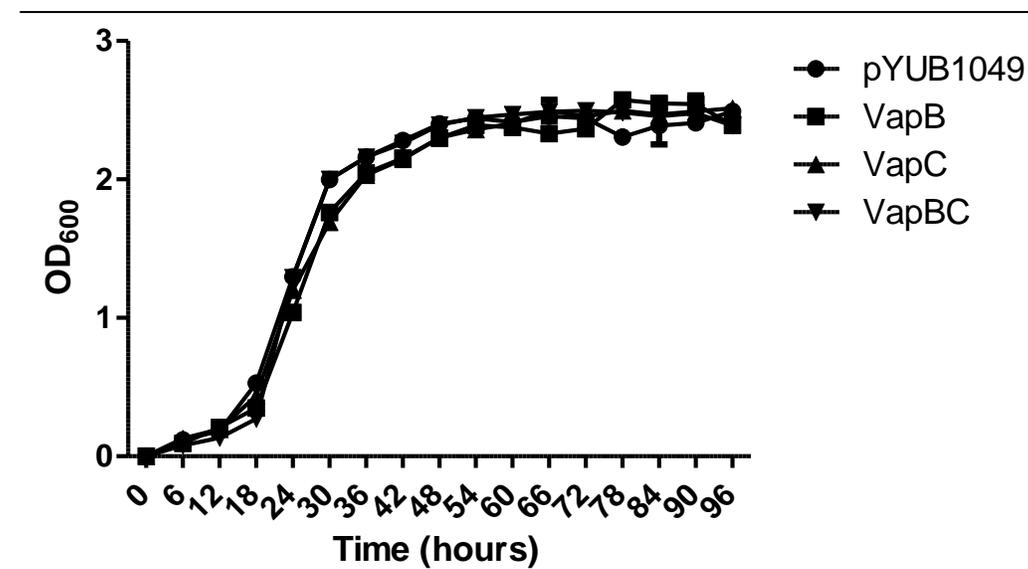


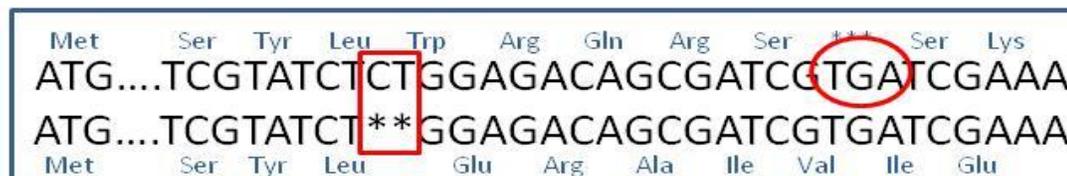
Figure 3.7. Growth of *M. smegmatis* mc<sup>2</sup>4517 expression strains over a 96 hour period. Legend denotes expression strains (VapB, VapC and VapBC), pYUB1049, vector only.

There is no difference in growth between the strains expressing VapB, VapC and VapBC and the control (empty vector). Expression of protein was confirmed by small scale His-tag binding tests (Section 2.2.2.1). Expression of VapB and the VapBC complex was observed but no expression of VapC protein was seen. The strain harbouring the pYUBMS1284 construct is not expressing VapC presumably due to its toxicity to the cell which may have resulted in a mutation in the *vapC* gene. VapBC and VapB were expressed at high levels with no effect on growth.

### 3.2.7 Mutation of the VapC Expression Construct

No VapC expression was observed in either the soluble or insoluble fractions in small scale expression tests (Figure 3.2). Scale up of expression cultures also revealed no expression of VapC in the soluble or insoluble fractions. His-tag affinity binding tests were performed on both soluble and insoluble fractions to definitively confirm the lack of expression of VapC in either fraction.

The pYUBMS1284 plasmid was extracted using the alkaline lysis plasmid preparation method (Section 2.1.1.3) from *M. smegmatis* mc<sup>2</sup>4517 expression cultures and the *vapC* insert PCR amplified using T7 forward and reverse primers and sequenced using the same primers. This revealed a two bp insertion into the *vapC* gene, giving rise to a frame shift and insertion of a stop codon further downstream (Figure 3.8). Plasmids extracted from expression cultures grown from fresh transformations of pYUBMS1284 into *M. smegmatis* mc<sup>2</sup>4517 showed the same two bp insert in three independent experiments.

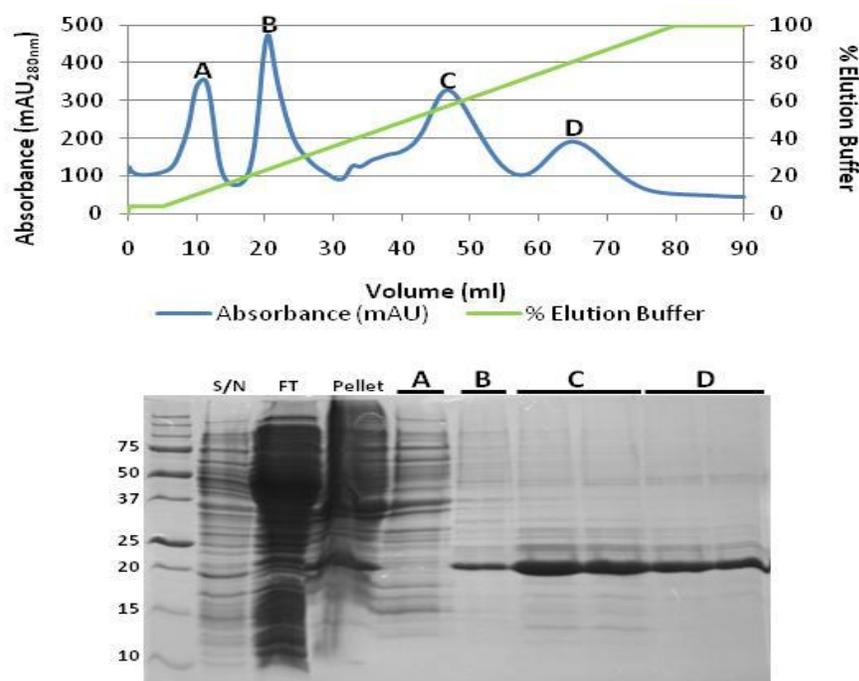


**Figure 3.8 VapC gene sequence and corresponding amino acids.** The red box shows the two bp insertion upon over-expression of VapC and the red circle shows that this insertion leads to a stop codon downstream. The top sequence is the gene sequence after expression and the bottom before expression. Amino acids are noted and \*\*\* denotes stop codons.

pYUBMS1284 was sequenced before transformation into *M. smegmatis* and the correct gene sequence in the correct orientation was observed. Due to the waxy nature of the mycobacterial cell wall, extraction of the pYUBMS1284 plasmid was difficult and multiple samples of multiple cultures were lysed in attempts to extract the plasmid. This was successful only three times and each time, the same two bp insert was observed in the same position (Figure 3.8).

### 3.2.8 Denaturing Purification of VapC

VapB and VapC expression in *E. coli* resulted in expression of insoluble protein. As no VapC expression was observed in *M. smegmatis*, expression of VapC in *E. coli* was trialled. A denaturing purification of VapC was performed (Section 2.2.6), by adding 8 M urea to lysis and elution buffers. VapC was purified by IMAC under denaturing conditions as shown in the chromatogram and SDS-PAGE gel below (Figure 3.9). VapC eluted late in the elution gradient at between 80 - 100% elution buffer, corresponding to 400 - 500 mM imidazole.



**Figure 3.9. VapC denaturing IMAC purification and corresponding 15% SDS-PAGE gel.** The chromatogram depicts the UV absorbance and elution profile of VapC. The SDS-PAGE gel denotes insoluble protein (pellet), soluble protein (supernatant, S/N) and column flow through (FT). Fractions and their elution position are shown by letters A-D.

### 3.2.9 Refolding of VapC

#### *Refolding of VapC Expressed in Escherichia coli*

Once purified, VapC was subjected to a protein refolding screen (Section 2.2.7.2) which was a modified version of the Hampton refolding screen. This consisted of 16 different conditions as outlined in Section 2.2.7.1 (refolding solutions in appendix A). Small scale refolding screens of VapC yielded promising results for refolding solutions one and two.

The successful refolding conditions above were scaled up to a 200 µl dialysis button (Section 2.2.7.3) and refolded protein analysed by native-PAGE and size exclusion chromatography. No refolded VapC was observed on a native-PAGE gel and the protein eluted in the void volume of the size exclusion column. VapC refolding was unsuccessful upon scale up from a 20 µl dialysis button to a 200 µl dialysis button and this was repeated a further two times with no protein refolding observed.

#### *Refolding of VapC Expressed in Mycobacterium smegmatis*

Purified VapBC complex (Section 2.2.2.2) was denatured by addition of 6 M guanidium and the VapC protein recovered by IMAC purification (Section 2.2.6). As *vapBC* was cloned into the pYUB1049 vector to enable expression of a C-terminal His-tag (Section 2.2.1) upon denaturation of the VapBC complex, VapC includes the His-tag and can be recovered by Ni<sup>2+</sup> affinity.

Refolding solutions six and seven yielded promising results and were scaled up to a 200 µl dialysis button (Section 2.2.7.3) and refolded protein analysed. No refolded VapC was observed on a native-PAGE gel and the protein eluted in the void volume of the size exclusion column. VapC refolding was unsuccessful upon scale up from a 20 µl dialysis button to a 200 µl dialysis button.

### 3.2.10 Cloning of *vapC* into pMAL-c2x

The *vapC* gene was successfully cloned (Section 2.2.3) into the pMAL-c2x vector, to enable expression of a MBP-VapC fusion protein. The *vapC* gene was successfully

---

amplified by PCR using primers containing restriction site EcoRI in the forward primer and BamHI in the reverse primer enabling digestion of the *vapC* insert and insertion into the pMAL-c2X vector. The expressed protein is retained in the cytoplasm of the cell whereas the pMAL-p2x vector results in the expressed protein being exported to the periplasm. The ligation was then transformed into *E. coli* BL21 (DE3) cells. Plasmid was extracted from successful transformants and sequencing confirmed correct insertion and orientation of *vapC* into pMAL-c2x and the construct was named pMAL-MS1284.

### **3.2.11 Expression and Purification of a MBP-VapC Fusion Protein**

The *E. coli* strain harbouring the pMAL-MS1284 construct was cultured in rich media + glucose. Glucose is essential in the growth media to repress the maltose genes on the chromosome of the *E. coli* host, one of which is amylase which can degrade the amylose resin used in the purification of the MBP-VapC fusion.

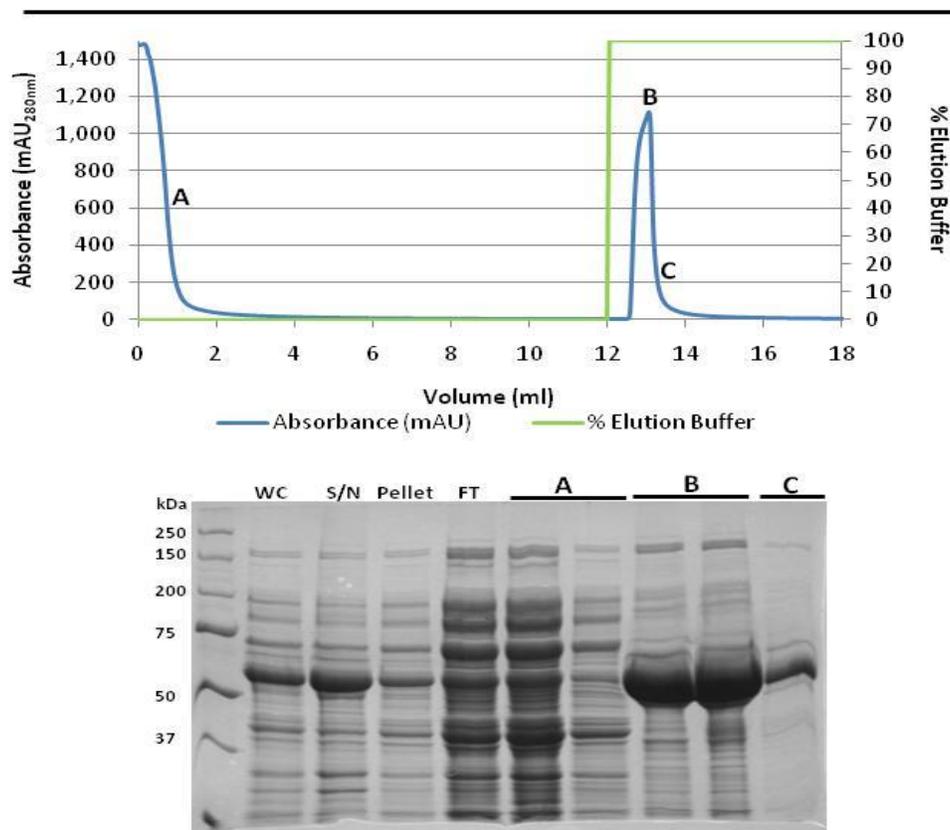
#### ***Small Scale MBP-VapC Expression***

A small scale expression trial (Section 2.2.4.1) was conducted to determine the behaviour of the MBP-VapC fusion protein. Expression cultures were sampled at one, two and three hours post-induction, the samples were lysed and run on a SDS-PAGE gel to determine if the MBP-VapC fusion was soluble and to determine which time point yields maximum protein expression. The MBP-VapC fusion was present in the soluble fraction and maximum protein expression was observed three hours post-induction (data not shown). Small scale affinity binding (Section 2.2.4.1) tests revealed that the MBP-VapC fusion bound to the amylose resin and could be eluted with 10 mM maltose. A range of maltose concentrations from 10 mM - 100 mM were employed to get complete elution of the MBP-VapC fusion protein from the amylose resin. However, there appeared to be no difference in elution of MBP-VapC from the amylose resin at any of the concentrations and some MBP-VapC still remained bound to the resin.

### ***Large Scale MBP-VapC Expression & Purification***

As small scale expressions tests were successful, protein expression was scaled up and MBP-VapC was purified using a 1 ml amylose resin column. EDTA was included in both the lysis buffer and the elution buffer to inhibit Ca<sup>2+</sup>-dependent proteases.

Non-specific proteins were removed early from the column during the wash stage; MBP-VapC was eluted with 10 mM maltose in a single peak shown in Figure 3.10. Fractions containing eluted MBP-VapC had a lot of other contaminating proteins, which are either being retained by the amylose resin or are binding to the MBP-VapC fusion itself. Further purification steps such as size exclusion chromatography were attempted to remove unspecific proteins that were purified alongside MBP-VapC (Section 2.2.10.2).

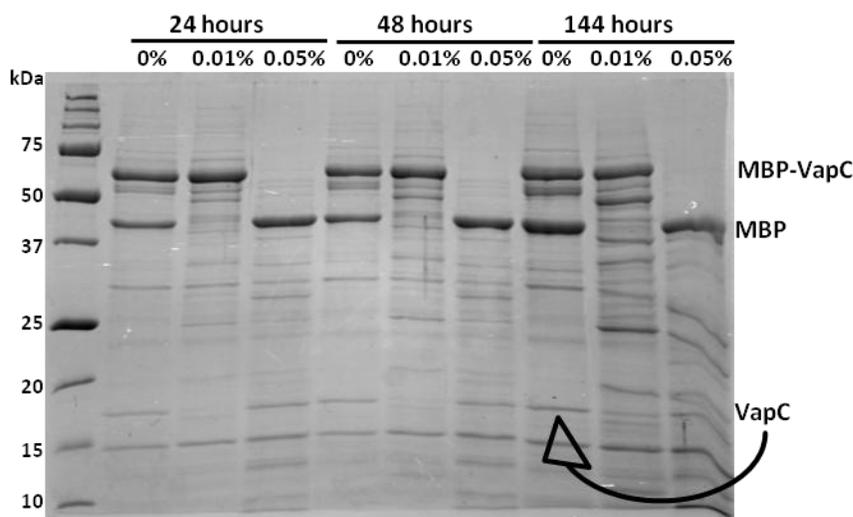


**Figure 3.10 MBP-VapC amylose purification and corresponding 10% SDS-PAGE gel.** The chromatogram depicts the UV absorbance and elution profile of MBP-VapC fusion protein. The SDS-PAGE gel denotes total cellular protein (WC) insoluble protein (pellet), soluble protein (supernatant, S/N) and column flow through (FT). Fractions and their elution positions are shown by letters A-C.

### *Factor Xa Cleavage of MBP-VapC*

The pMAL vectors contain sequences coding for the recognition of specific proteases between *malE* (gene encoding MBP) and the multiple cloning site. The pMAL-c2x vector encodes a Factor Xa cleavage site. Factor Xa cleaves after a four amino acid recognition sequence so that few or no vector-derived residues remain attached to the protein of interest.

Typically Factor Xa cleavage is carried out with Factor Xa at 1% of the fusion protein. Ratios of one, two and three percent Factor Xa: MBP-VapC fusion at time points of two, four, eight and 24 hours were trialled to determine which concentration resulted in the most efficient cleavage of MBP-VapC (Figure 3.11). No obvious difference in efficiency was seen between the various ratios at various time points (data not shown). Inclusion of low concentrations of SDS (0.01 - 0.05% (*w/v*)) has been shown to increase the cleavage efficiency of Factor Xa. SDS (0.01 and 0.05% (*w/v*)) was included in Factor Xa cleavage reactions of the MBP-VapC fusion protein at two percent (*w/v*) Factor Xa: fusion protein and reactions were incubated for 24, 48 and 144 hours (Figure 3.11).

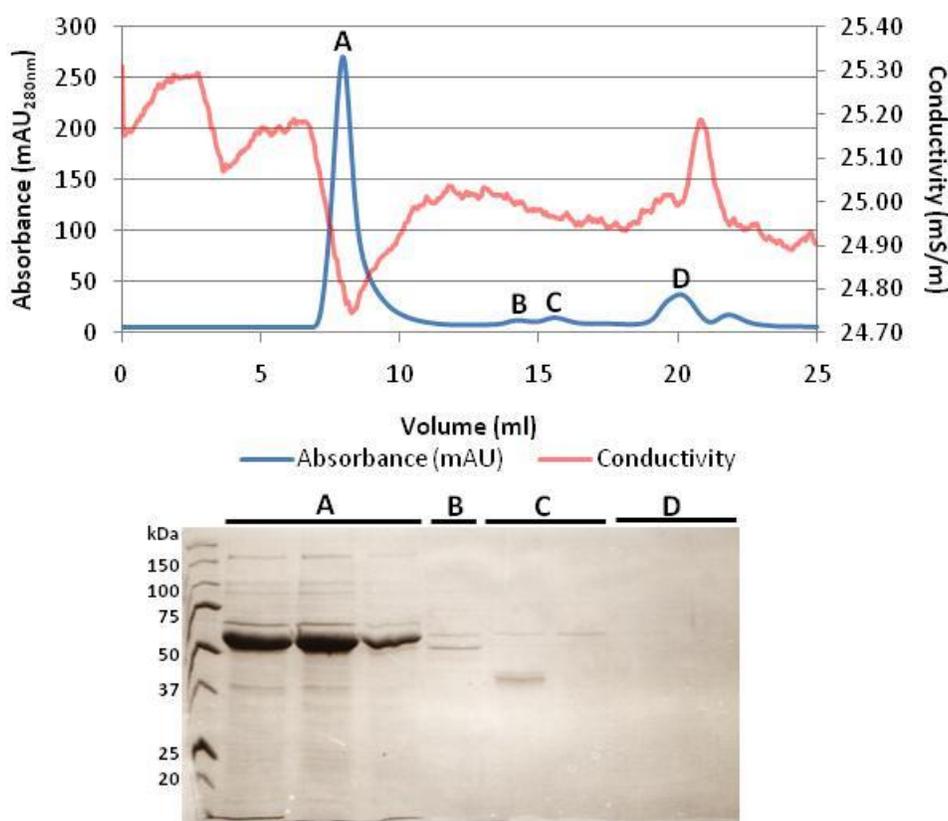


**Figure 3.11 Factor Xa cleavage of the MBP-VapC fusion.** Percentage values denote concentrations of SDS included in the cleavage reaction (2% Factor Xa). Time values along the top denote incubation time at room temperature. MBP-VapC, position of fusion protein, MBP, position of MBP alone VapC and arrow denotes position of VapC on the 15% SDS-PAGE gel.

Inclusion of 0.05% SDS in the Factor Xa cleavage reaction improved the efficiency of Factor Xa in cleaving MBP from VapC (Figure 3.11). All MBP-VapC fusion was cleaved after 24 hours incubation at room temperature. Surprisingly, the presence of 0.01% SDS is less efficient than when no SDS was present in the reaction.

### *Size Exclusion Purification of VapC*

The MBP-VapC fusion protein was subjected to size exclusion chromatography both before and after Factor Xa cleavage. The MBP-VapC fusion protein eluted in the void volume of the column, indicating that the protein is forming a soluble aggregate. Attempts to reduce aggregation by the addition of 5% (v/v) glycerol and/or 0.1% tween 80 (v/v) to the lysis buffer were unsuccessful and interfered with the binding of the fusion protein to the amylose column.



**Figure 3.12 MBP-VapC size exclusion purification and corresponding 10% SDS-PAGE gel.** The chromatogram depicts the UV absorbance and conductivity profile of MBP-VapC. The silver-stained SDS-PAGE gel denotes fractions and their elution position is shown by letters A-D. Note that the MBP-VapC fusion is eluted in the void volume of the column.

---

The MBP-VapC fusion protein forms a soluble aggregate as shown by elution in the void volume of a size exclusion column (Figure 3.12), which explains the difficulties in cleaving MBP from VapC using Factor Xa. The cleavage site between the two proteins would be obscured due to protein aggregation. Inclusion of small amounts of SDS would reduce aggregation, allowing Factor Xa to cleave MBP from VapC. However, further analysis of the cleavage reaction by size exclusion chromatography showed that VapC without MBP also forms a soluble aggregate, as it also eluted in the void volume of the size exclusion column. Therefore, the VapC-MBP purification was not pursued.

### **3.2.12 Isolation of VapC**

To determine the function of VapC from *M. smegmatis*, functional VapC protein was required. Expression of VapC using *M. smegmatis* as a host did not produce any VapC and expression using *E. coli* as a host produced insoluble protein that could then be purified under denaturing conditions. Subsequent attempts to refold VapC were ineffective. Soluble VapC can be expressed when in complex with VapB, so disruption of the interaction between VapB and VapC was pursued to yield functional VapC protein.

#### ***Disruption of the VapBC Complex***

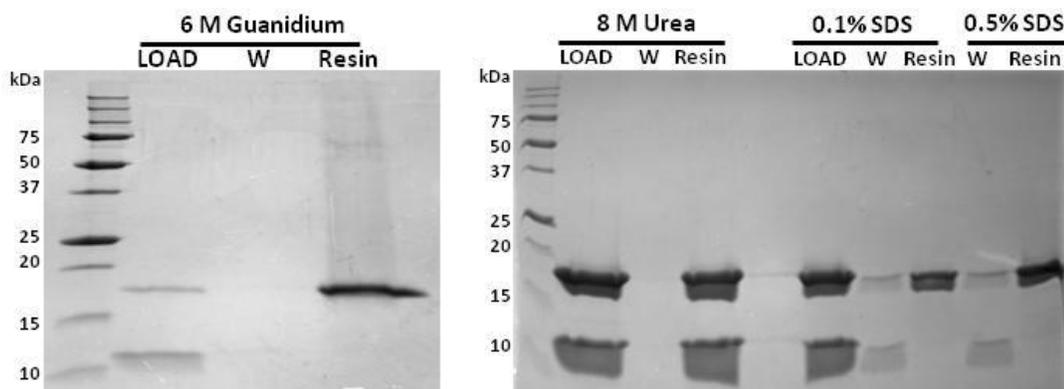
A variety of substances including urea, guanidium, detergents, salt and pH were trialled to disrupt the VapBC complex purified from *M. smegmatis* by disrupting the interaction between VapB and VapC.

A gradient ranging from 0.5 M to 8.0 M urea was ineffective at disrupting the interaction between VapB and VapC (Figure 3.13), even though at 8 M urea denaturation of the two proteins would presumably occur. Disruption of the VapBC complex was analysed by nickel affinity tests, as the His-tag is present on VapC. If the complex is disrupted only VapC will be present on the nickel resin and VapB removed during the wash steps. If the interaction between the two proteins is not disrupted both VapB and VapC will be present on the nickel resin. Detergents such as

tween 20, tween 80, triton X-100, lauryl dimethyl amine oxide (LDAO), zwittergent 3-10 and zwittergent 3-12 were trialled with and without 8 M urea in an attempt to disrupt the hydrophobic interaction between VapB and VapC. This was also unsuccessful; again both VapB and VapC proteins were present on the nickel resin.

A NaCl gradient ranging from 0.5 - 2.0 M was employed to disrupt the interaction between VapB and VapC, again this was unsuccessful with both VapB and VapC proteins present on the nickel resin.

A range of buffers varying in pH from 3.5 to 7.0 were used in a further attempt to disrupt the interaction between VapB and VapC. By altering the net charges of the proteins it was hoped that this would alter the interaction between them. However both VapB and VapC proteins were present on the nickel resin after treatment with these buffers.



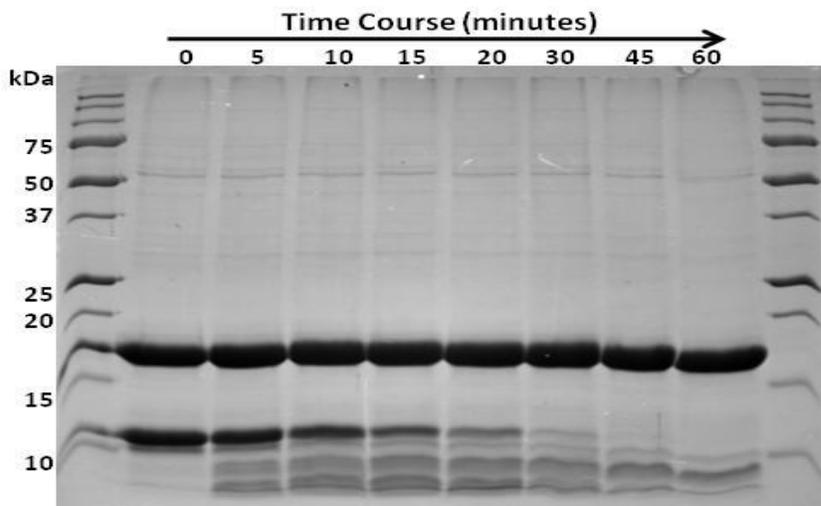
**Figure 3.13. Disruption of the VapBC complex with 6 M Guanidium, 8 M Urea, 0.1 and 0.5% SDS (15% SDS-PAGE Gel).** Load denotes VapBC complex + treatment, W, wash to remove unbound protein and Resin, protein remaining bound to the Ni resin.

Six M guanidium proved successful at disrupting the VapBC complex, only VapC was present bound to the nickel resin after treatment (Figure 3.13). Both 0.1% and 0.5% SDS were also successful in disrupting the interaction between VapB and VapC (Figure 3.13). It is likely that 6 M guanidium and 0.5% and 0.1% SDS denature VapC (Tsumoto, *et al.*, 2003; Vallejo & Rinas, 2004), which then requires a refolding step.

### *Trypsin Digest of the VapBC Complex*

The calculated stability indices of VapB (46.36) and VapC (24.32) were calculated by Protparam (Gasteiger, *et al.*, 2005) and classify VapB as less stable than VapC and therefore more susceptible to proteolytic degradation (a common feature of TA systems). These properties were used in an attempt to remove VapB from VapC, to leave functional VapC.

Treating the VapBC complex with  $0.1 \text{ mg.ml}^{-1}$  trypsin resulted in degradation of VapB after an incubation period of one hour at room temperature, depicted in Figure 3.14.

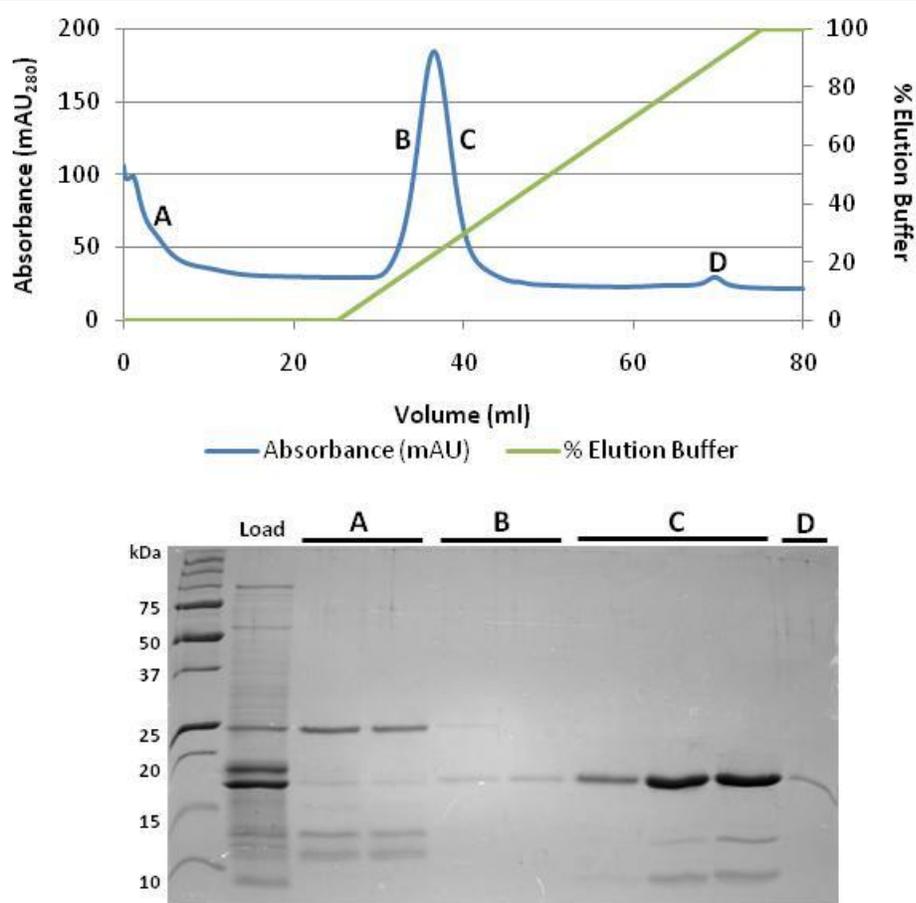


**Figure 3.14. Trypsin Digest of VapBC.** Time course trypsin digest of VapBC resulting in degradation of VapB over time (15% SDS-PAGE gel).

Reactions were stopped by a trypsin inhibitor. For use in functional studies VapC needed to be purified away from trypsin + inhibitor and any remaining VapB degradation products. Size exclusion chromatography could not be used as the molecular weights of VapC and trypsin are very similar. VapC has a His-tag, but small scale His-tag affinity binding tests revealed that the His-tag no longer remained on VapC as VapC did not bind to the nickel resin (data not shown). Therefore, trypsin must also remove the His-tag from VapC.

### 3.2.13 Purification of VapC Following Trypsin Digestion of the VapBC Complex

The isoelectric points of VapC and trypsin are quite different, therefore anion exchange chromatography could be used to purify VapC from trypsin. Trypsin has a predicted isoelectric point of 10.5 (ProtParam) whereas VapC has a predicted isoelectric point of 4.9 (ProtParam). As the VapBC complex is in a buffer at pH 7.4 VapC has a net negative charge, whereas trypsin would have a net positive charge. VapC would therefore bind to an anion exchange column, whereas trypsin will flow through.

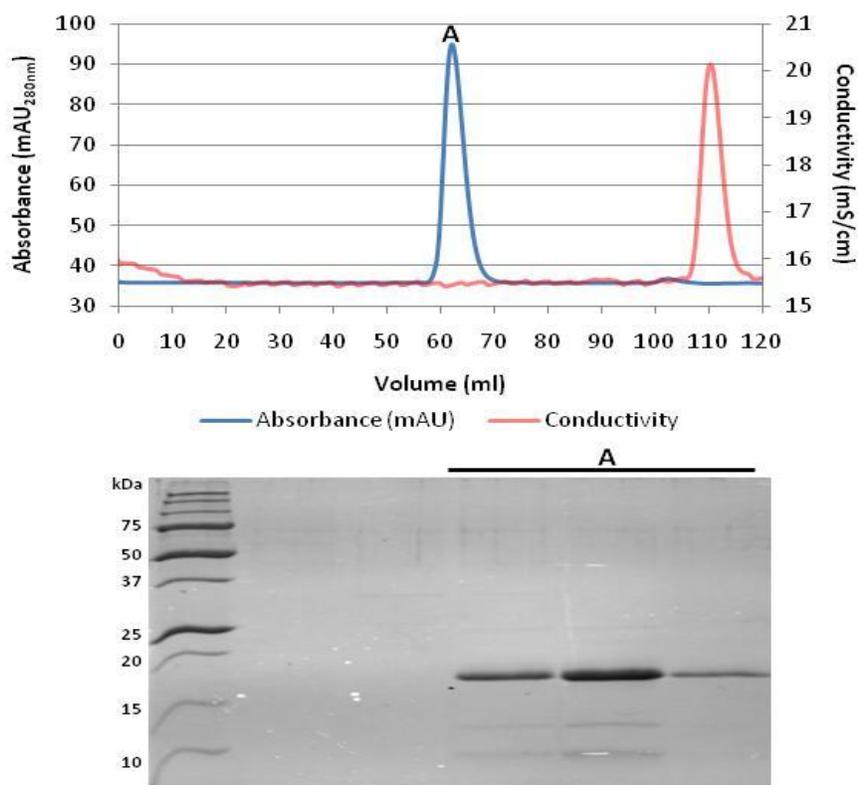


**Figure 3.15. Anion exchange purification of a trypsin digest of VapBC and corresponding 15% SDS-PAGE gel.** Trypsin is eluted early off the column (A) whereas VapC is bound by the column and eluted early in the NaCl gradient (B & C).

Purification of VapC by anion exchange chromatography was successful at isolating the toxin (Figure 3.15). VapBC was dialysed into 50 mM phosphate buffer pH 7.4,

100 mM NaCl as VapC did not bind to the anion exchange column when in 200 mM NaCl. Trypsin elutes early from the column whereas VapC remains bound and elutes at 200 - 300 mM NaCl. Two smaller molecular weight bands are present with the VapC samples on the SDS-PAGE gel (Figure 3.15) and these are different in size to VapB (11.6 kDa). They appear to be small amounts of VapC degraded by trypsin. MALDI MS revealed two different sized species of VapC at 14.94 kDa and 16.33 kDa. 16.3 kDa correlates to VapC including the C-terminal His-tag, which has a calculated molecular weight of 16.30 kDa (Gasteiger, *et al.*, 2005). The 14.94 kDa band correlates to VapC without the C-terminal His-tag which would have a molecular weight of 14.91 kDa (Gasteiger, *et al.*, 2005).

Further purification of VapC by size exclusion chromatography showed that VapC forms a dimer; VapC elutes as a single peak at 63.5 ml corresponding to a molecular weight of ~32 kDa.



**Figure 3.16. Size exclusion chromatography of VapC and corresponding 15% SDS-PAGE gel.** Chromatogram depicts the UV absorbance and elution profile of VapC and corresponding PAGE gel purified VapC.

---

### 3.2.14 VapB Peptide Inhibition of VapC

Using the FitAB complex from *N. gonorrhoeae* (Mattison, *et al.*, 2006) as a model, two peptides were made to mimic the portion of VapB that binds to VapC and inhibits its activity. One peptide (IRLVA3) was designed to mimic the VapB  $\alpha$ -helix 3 proposed to bind to a hydrophobic groove on the surface of VapC, similar to that of the FitAB complex. The other peptide (IRLVA4) was designed to mimic the VapB N-terminal peptide that inserts an arginine into the VapC active site, thereby inhibiting its activity.

Peptides IRLVA3 and IRLVA4 were added to both the VapBC complex and VapC protein alone. Analysis by native-PAGE showed no binding of either peptide to the VapBC complex or VapC protein (data not shown). There was no change in the isoelectric point of either proteins following addition of the peptides. Ribonuclease activity assays against a known RNA substrate (Section 5.2.5) of VapC for each peptide + VapC showed equivalent ribonuclease activity to VapC alone indicating that neither peptide had bound to VapC producing inhibition (data not shown).

### 3.2.15 VapBC DNA Binding

The VapBC complex from *M. smegmatis* has been shown previously to bind to inverted repeats in its promoter DNA (Figure 3.1). An aim of this study was to crystallise the VapBC complex bound to inverted repeat sequences in its cognate promoter DNA.

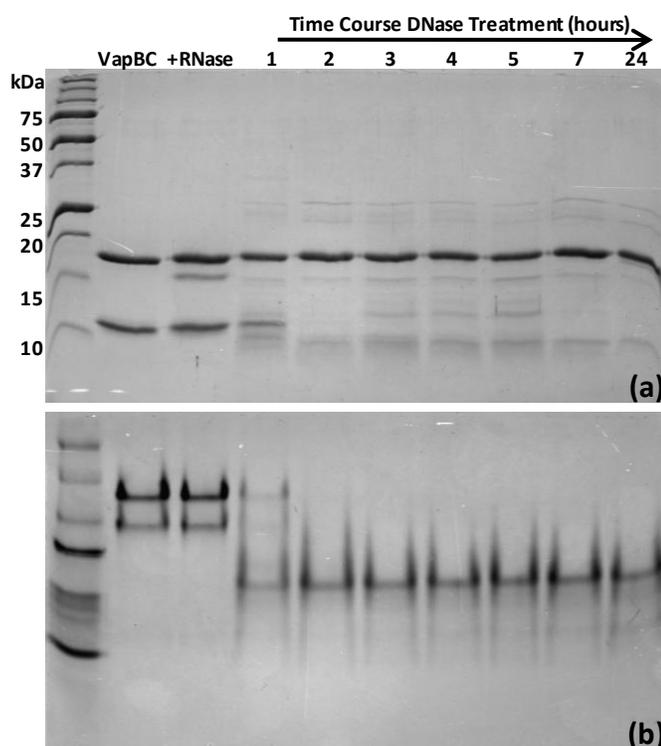
Two complementary 32 base DNA oligonucleotides that included the two inverted repeat sequences for VapBC binding were designed. Two complementary 32 base control oligonucleotides (Section 2.2.13.1) were also designed that consisted of the incorrect inverted repeat in the VapBC promoter region. The complementary oligonucleotides were annealed to form dsDNA.

The VapBC complex was incubated with the dsDNA oligonucleotides as in Section 2.2.13.5. No shift in the protein position on a non-denaturing native-PAGE gel was

observed. Therefore there was no binding of VapBC to the promoter dsDNA oligonucleotide or the control dsDNA oligonucleotide.

### ***DNase and RNase Treatment of the VapBC Complex***

As no binding to inverted repeats in the promoter DNA oligonucleotide was observed it was thought that as VapBC is purified from its host organism it may be co-purified with DNA or RNA. To investigate this, the VapBC complex was treated with DNase and RNase enzymes. Treatment with RNase had no effect on the VapBC complex as seen on native and SDS-PAGE gels (Figure 3.17).



**Figure 3.17 DNase and RNase treatment of the VapBC Complex.** (a) SDS-PAGE gel and (b) non-denaturing PAGE gel of RNase and DNase (time course) treated VapBC. Negative control, VapBC. VapB degradation shown in (a) and a protein shift seen in (b) upon DNase treatment.

Treatment of the VapBC complex with DNase resulted in degradation of VapB as in Figure 3.17, this also resulted in a shift on a native-PAGE gel implying that once DNA is removed, VapB is susceptible to degradation. Although the VapBC complex undergoes two purification steps (IMAC and size exclusion chromatography) and protease inhibitors are added to the purification, very small amounts of proteases could be present. Degradation of VapB from the VapBC complex was also observed after one month at 4 °C and after 72 hours at room temperature, indicating that VapB is particularly susceptible to proteolytic degradation. A time course of VapBC DNase

---

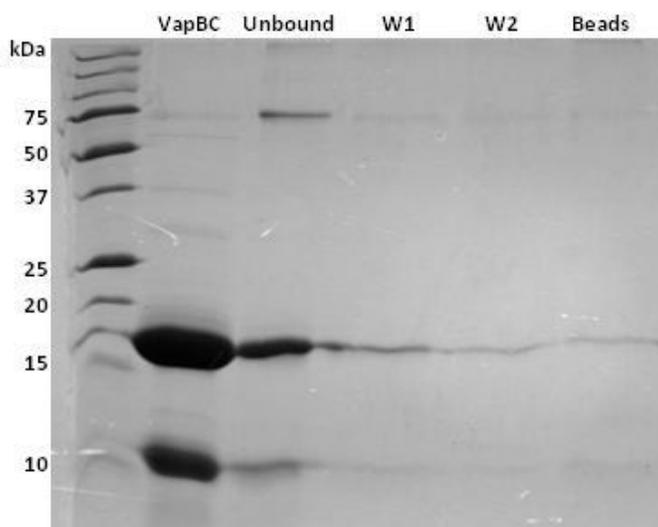
treatment was used to determine if VapB degrades instantly upon removal of DNA or if there is a window before VapB degrades so another synthetic piece of DNA could be bound to the complex. VapB is degraded after one hour of DNase treatment (Figure 3.17). A small amount of VapB remains but the majority is degraded upon removal of DNA. After 2 - 24 hours of DNase treatment all VapB has degraded leaving only VapC. Thus the VapBC complex cannot be treated with DNase to remove co-purified DNA before crystallisation or addition of a synthetic piece of DNA.

### ***Visualisation of DNA bound to VapBC***

If the VapBC complex is purified with endogenous DNA, the DNA should be able to be visualised by UV absorbance or on an agarose or acrylamide gel. However, this was not the case. When measuring the DNA by UV absorbance either by nanodrop or by a spectrophotometer only very small values were observed. For example, VapBC complex at a concentration of  $25 \mu\text{g}\cdot\text{ml}^{-1}$  (determined by absorption at 280 nm) has  $4.1 \text{ ng}\cdot\mu\text{l}^{-1}$  DNA bound (determined by absorption at 260nm). When the VapBC complex is analysed by agarose gel electrophoresis before and after denaturation of the protein complex no DNA is observed (data not shown). The only evidence for this was a small amount of DNA observed in the well of a urea-denaturing PAGE gel.

### ***Biotin Labelled DNA Oligonucleotide Exchange***

A promoter oligonucleotide (Section 2.2.13.1) was designed to include a biotin molecule at the 5' end of the DNA. The biotin labelled oligonucleotide was annealed to its reverse complement DNA oligonucleotide to form dsDNA for DNA binding studies. The VapBC complex was incubated with excess biotin labelled dsDNA oligonucleotide and binding of the VapBC complex was determined by binding the biotin labelled oligonucleotide to streptavidin beads and washing to remove unbound VapBC. A small amount of VapBC has exchanged from the DNA it is co-purified with to the biotin labelled dsDNA oligonucleotide (Figure 3.18). However, only a very small portion has exchanged, most of the VapBC complex has not bound to the beads.



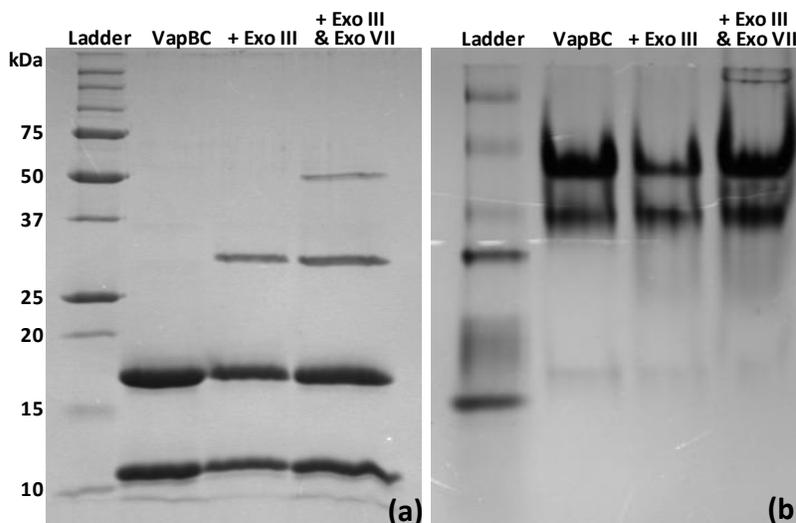
**Figure 3.18. VapBC DNA exchange with biotin labelled dsDNA oligonucleotide.** The majority of the VapBC complex has not bound to the streptavidin beads (unbound), a small amount is removed in the wash steps (W1 & W2) and a small amount of VapBC has bound to the biotin-labelled oligonucleotide (Beads). Untreated VapBC shown in Lane 1 (VapBC).

### 3.2.16 Exonuclease Treatment of VapBC

Purified VapBC complex was treated with exonuclease III and exonuclease VII to trim back the DNA that co-purifies with the complex, in the hope that the complex would be more suitable for crystallisation.

A combination of two exonucleases were used, exonuclease III from *E. coli* acts on dsDNA in a 3' to 5' direction to produce ssDNA and exonuclease VII from *E. coli* acts on ssDNA and exhibits both 5' to 3' and 3' to 5' exonuclease activities.

An initial pilot experiment determined that exonuclease treatment of purified VapBC did not result in degradation of VapB, as shown in native and SDS-PAGE gels below (Figure 3.19).



**Figure 3.19. Exonuclease treatment of the VapBC complex.** (a) 15% SDS-PAGE and (b) 10% native PAGE gels show no degradation of VapB upon treatment with exonuclease III (+ExoIII) or upon treatment with exonuclease III and exonuclease VII (+ExoIII & ExoVII) compared to the negative control (VapBC).

The exonuclease treated complex was then subjected to analytical size exclusion chromatography (data not shown). The elution volumes of untreated and exonuclease treated VapBC samples were compared to see if there was any change in the size of the complex. The elution volumes for untreated VapBC and exonuclease-treated VapBC were very similar. The calculated molecular weight of untreated VapBC was 146 kDa and the molecular weight for exonuclease treated VapBC was 141 kDa. This tentatively suggests that there is little difference between the untreated and exonuclease treated VapBC complex. This value is slightly different to the previously calculated molecular weight of the VapBC complex of 156 kDa. This could be due to slight differences between purifications or the length of the DNA which is co-purified with VapBC.

The calculated molecular weight is smaller for the untreated VapBC complex than previously determined (Section 3.2.4) although the stoichiometry of the complex is the same; a tetramer of VapBC heterodimers. The difference in the molecular weights could be due to slight variations between purifications or the size of the DNA which VapBC is copurified with.

---

### 3.2.17 Crystallisation screens

#### *VapC Crystallisation*

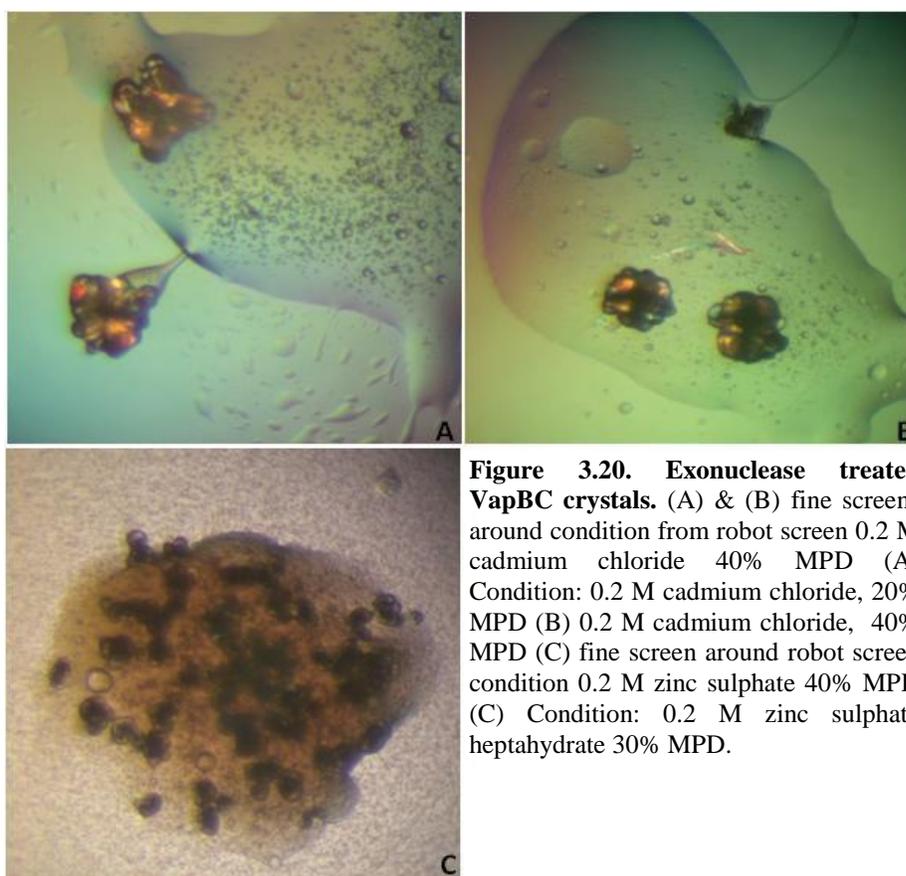
VapC protein was obtained by a tryptic digest of the VapBC complex to remove VapB (Section 2.2.9.1). VapC was then purified by anion exchange and size exclusion chromatography (Sections 2.2.10.1 and 2.2.10.2). Robot crystallisation screens (Sections 2.2.14) for VapC were laid down at 8.6 and 17.1 mg.ml<sup>-1</sup>. Crystals were observed and hanging drop fine screens around these conditions were carried out as in Section 2.2.15.2, but testing of these crystals by X-ray diffraction (Section 2.2.17) determined that they were salt crystals and not protein. Crystallisation of VapC in a condition containing 0.01 M zinc sulphate heptahydrate, 0.1 M MES pH 6.5, 25% (v/v) PEGME 550 was unreproducible in a fine screen even with microseeding (Section 2.2.16).

#### *VapBC Crystallisation*

Purified VapBC was subjected to robot crystallisation screens at concentrations of 5, 18.9 and 34 mg.ml<sup>-1</sup>. None of these screens resulted in protein crystals, many conditions grew crystals but upon fine screening (hanging drops and sitting drops) around these conditions and testing by X-ray diffraction they were shown to be salt crystals. Many of these were phosphate salts due to the protein purification buffer (50 mM sodium phosphate pH 7.4, 200 mM NaCl). To eliminate this factor from protein crystallisation VapBC was dialysed into 50 mM HEPES pH 7.4, 200 mM NaCl and subjected to robot crystallisation screens at a concentration of 8.4 mg.ml<sup>-1</sup>. Precipitation of the VapBC complex was observed in the dialysis tubing upon dialysis into a HEPES buffer and no crystals were observed in any of the robot screen conditions.

Exonuclease treated VapBC was subjected to robot crystallisation screens at a concentration of 13.67 mg.ml<sup>-1</sup>. Multiple conditions grew crystals, but upon fine screening and X-ray diffraction most of these were deemed to be salt crystals. However, fine screens around conditions 0.2 M cadmium chloride, 40% MPD and 0.2 M zinc sulfate heptahydrate, 40% MPD resulted in crystals that did not diffract.

The crystals were of low quality and not well formed, fine screening around the robot screen condition resulted in slightly improved crystal quality (Figure 3.20). The crystals in Figure 3.20 also did not diffract. No cryoprotectant was needed in testing the crystals in Figure 3.20, the MPD in the crystallisation condition was a sufficient cryoprotectant. Further fine screening and additive screening around these crystallisation conditions could not improve crystal quality for X-ray diffraction. It is possible that these are also salt crystals, the low solubility of zinc and cadmium could promote the formation of cadmium and zinc phosphate crystals.



**Figure 3.20. Exonuclease treated VapBC crystals.** (A) & (B) fine screens around condition from robot screen 0.2 M cadmium chloride 40% MPD (A) Condition: 0.2 M cadmium chloride, 20% MPD (B) 0.2 M cadmium chloride, 40% MPD (C) fine screen around robot screen condition 0.2 M zinc sulphate 40% MPD (C) Condition: 0.2 M zinc sulphate heptahydrate 30% MPD.

---

---

### 3.3 Discussion

Expression and subsequent structure-function studies of the VapBC complex from *Mycobacterium smegmatis* would enable a greater understanding of its role in *M. smegmatis*, which could then be used as model for VapBC proteins in *Mycobacterium tuberculosis*.

#### 3.3.1 Expression of VapB, VapC and VapBC and Isolation of VapC

Initial attempts to express recombinant VapB, VapC and VapBC in *Escherichia coli* were unsuccessful. The proteins were insoluble and *E. coli* does not recognise the one base pair overlap between *vapB* and *vapC* when the operon was overexpressed. Salvage strategies to aid solubility including low temperature expression and lysis buffer screens were unable to rescue VapB and VapC from inclusion bodies (Summers, 2007)

To overcome solubility problems VapB, VapC and VapBC were cloned into the *E. coli*-mycobacterial shuttle vector pYUB1049 for expression in *M. smegmatis*. *M. smegmatis* proved to be ideal for the expression of VapB and VapBC. It is a non-pathogenic fast growing mycobacterial species and provides mycobacterial chaperones and cofactors needed for correct protein folding (Bashiri, *et al.*, 2007). The one base pair overlap between VapB and VapC is also recognised enabling expression of VapC in complex with VapB. No protein expression was observed upon induction with IPTG whereas when the auto-induction medium ZYP-5052 with 0.05% tween 80 was used for expression, yields of around 5 mg.l<sup>-1</sup> of culture were observed.

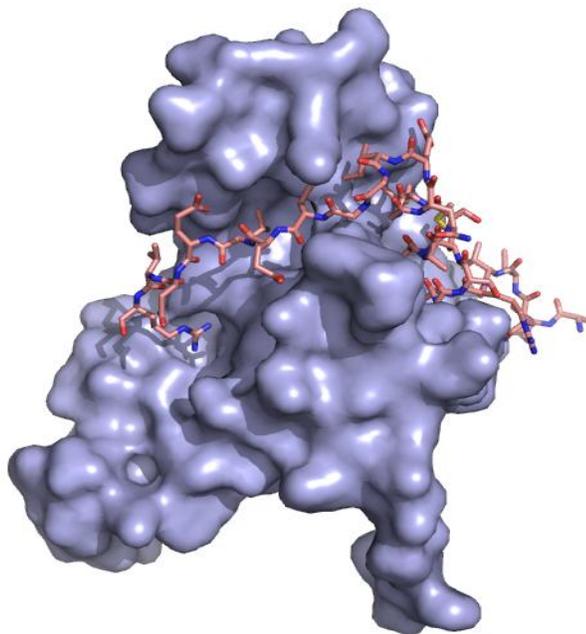
Expression of VapB and VapBC was seen but no expression of VapC. VapB expressed as a soluble aggregate and repeated attempts to express VapC were unfruitful. Sequencing of the VapC expression construct (pYUBMS1284) post-induction revealed a two base pair insertion resulting in a stop codon downstream, explaining the absence of VapC protein expression. The CT insertion was observed at the same position in a further two experiments. Basal expression of

---

VapC from the leaky T7 promoter before induction presumably results in the bacteria mutating the plasmid before expression is induced. Protein expression systems in *E. coli* have been developed which allow suppression of basal expression from leaky inducible promoters and tight control of plasmids and copy numbers, enabling expression of toxic proteins (Saida, Uzan, Odaert, & Bontems, 2006). However these systems are unavailable in mycobacteria. Expression of VapC is possible when expressed in complex with its cognate antitoxin VapB. VapB neutralises the toxicity of VapC, leading to expression of the VapBC complex.

Expression of VapC as a maltose binding protein (MBP) fusion in *E. coli* greatly increased the solubility of VapC. When in the context of a fusion protein MBP acts as a general molecular chaperone and promotes proper folding of the attached protein (Kapust & Waugh, 1999). MBP also prevents dimerisation of VapC, thereby reducing its toxicity to the cell, enabling soluble protein expression. Although expression of soluble MBP-VapC protein was seen, further investigation revealed that the MBP-VapC fusion was a soluble aggregate. The fusion protein eluted in the void volume of a size exclusion column and cleavage of MBP from VapC with the protease Factor Xa was difficult which is consistent with the MBP-VapC fusion forming a soluble aggregate. Expression of functional VapC proteins from *M. tuberculosis* as MBP fusions has been successful in other studies (Ramage, *et al.*, 2009). Incompletely or partially folded VapC could self-associate to form insoluble aggregates and the fate of the folding intermediate is dependent on its concentration inside the cell as a high concentration tends to favour aggregation (Kapust & Waugh, 1999). In the case of the MBP-VapC fusion, reduction of insoluble aggregates is seen, but the VapC protein must still be only partially folded leading to the formation of a soluble aggregate.

Refolding of VapC was unsuccessful upon scale-up and expression in *E. coli* yielded insoluble VapC protein. Expression in *M. smegmatis* resulted in no VapC expression and expression of VapC as an MBP fusion protein resulted in a soluble aggregate. However, VapC can be expressed as soluble protein in a complex with VapB.



**Figure 3.21 FitB from *N. gonorrhoeae* with FitA peptide bound.** FitB monomer shown as surface diagram and FitA shown as sticks.  $\alpha 3$  helix from Fit A binds to a hydrophobic groove on FitB. C terminal arginine of FitA sits in the active of FitB inhibiting its activity.

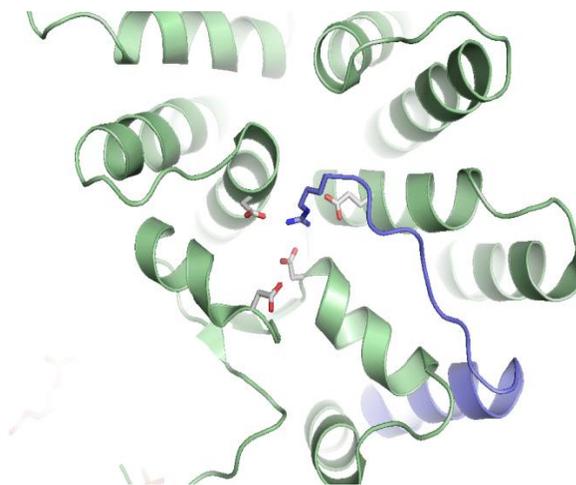
Disruption of the interaction between VapB and VapC proved difficult. The FitAB structure of a VapBC complex from *Neisseria gonorrhoeae* (Mattison, *et al.*, 2006; Wilbur, *et al.*, 2005) revealed that the FitA and FitB proteins associate tightly and the heterodimerisation interface is formed by contacts between  $\alpha 3$  and the C-terminal extended coil region of FitA (Figure 3.21) and helices  $\alpha 1$ ,  $\alpha 2$  and  $\alpha 4$  of FitB (Figure 3.21). The FitA helix fills a large exposed hydro-

phobic groove on FitB (Mattison, *et al.*, 2006). Most of the FitAB interface is hydrophobic and I hypothesise that the VapB and VapC interaction is the same in the VapBC complex from *M. smegmatis*. The binding of the  $\alpha 3$  helix of FitA to FitB is stabilised by four ionic interactions, strengthening the tight association between the two hydrophobic surfaces (Mattison, *et al.*, 2006).

A variety of substances were employed to attempt to disrupt the interaction between VapB and VapC. As with the FitAB complex the VapBC heterodimer from *M. smegmatis* associates extremely tightly. The only substances that disrupted the VapBC complex were 6 M guanidium and 0.1% and 0.5% SDS. Unfortunately, both of these presumably denature the proteins, making it impossible to obtain functional VapC. Refolding of VapC after denaturation with 6 M guanidium gave promising results, but upon scale-up, no refolded VapC was observed. Other detergents such as triton X-100, tween-20, tween-80, zwittergent and LDAO were unsuccessful at disrupting the hydrophobic interaction between VapB and VapC. Urea (8 M), was

also unsuccessful at disrupting this interaction, again reinforcing how tight the interaction is between the two proteins. A sodium chloride and pH gradient were employed to disrupt the ionic interactions that stabilise the interaction between the two hydrophobic surfaces, also with no success.

It is a well known feature of toxin-antitoxin systems that the antitoxin is more susceptible to proteolytic degradation than the toxin (Buts, *et al.*, 2005; Gerdes, *et al.*, 2005). This indeed seems to be the case with the VapBC TA loci. The instability index of VapB from *M. smegmatis* was calculated to be 46.36 (ProtParam(Gasteiger, *et al.*, 2005)), classifying the protein as unstable, whereas the instability



**Figure 3.22. Ribbon diagram of FitA (mauve) peptide bound to FitB (green).** C-terminal arginine of FitA located in the acidic pocket/active site of FitB, inhibiting its activity.

index of VapC from *M. smegmatis* was calculated to be 24.32 (ProtParam (Gasteiger, *et al.*, 2005)), classifying the protein as stable. Utilising the different properties of these two proteins the protease trypsin was used to digest away VapB, leaving VapC. The trypsin was removed by anion exchange chromatography and size exclusion chromatography indicated that VapC does in fact form a dimer. The FitAB structure revealed an arginine residue at the C-terminus of FitA located in the acidic pocket of FitB, blocking substrate access (Figure 3.22), thereby inhibiting activity (Mattison, *et al.*, 2006).

An arginine residue is also present at the C-terminus of VapB from *M. smegmatis* thereby it is assumed that VapB inhibits VapC in a similar mechanism to that of FitAB. A concern of using trypsin to remove VapB is that the VapB peptide including the C-terminal arginine that sits in the active site of VapC would remain bound, inhibiting activity of VapC. However MALDI MS results indicate that no

---

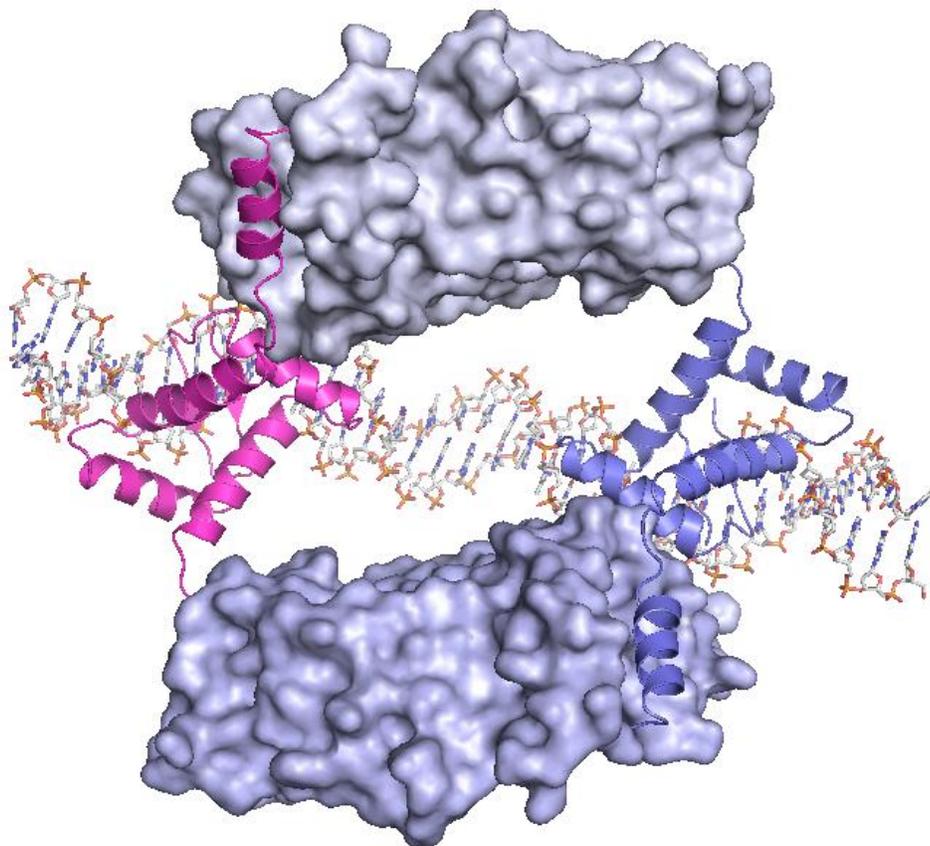
VapB peptide remains bound to VapC as the molecular weights correspond to the size of VapC with and without a C-terminal His-tag. Also, ribonuclease activity assays revealed that VapC is active after trypsin digestion of VapB indicating that no VapB peptide remains bound (see Chapter Five).

Two peptides were made to mimic the portion of VapB that binds to VapC and inhibits its activity. One, to mimic the VapB  $\alpha$ -helix 3 that is proposed to bind to a hydrophobic groove on the surface of VapC, and the other to mimic the VapB N-terminal peptide that inserts an arginine into the VapC active site, thereby inhibiting its activity. Neither peptide was successful at disrupting the VapBC complex by competing with VapB for binding to VapC nor was either peptide successful at binding to VapC and inhibiting its ribonuclease activity. Peptides mimicking antitoxins have been successful at disrupting other toxin-antitoxin systems for example the Kid/PemK toxin can be inhibited by addition of a peptide that mimics the antitoxin peptide of Kis/PemI that inhibits the ribonuclease activity of Kis/PemK (Agarwal, *et al.*, 2010). Trypsin digestion of the VapBC complex removes VapB and as VapC displays ribonuclease activity no VapB peptide remains bound. Once VapB is removed by trypsin, the peptides could exhibit weaker binding to VapC, explaining why the synthetic peptides do not bind to VapC. A longer peptide that mimics VapB could be required for binding and inhibition of VapC, perhaps both the  $\alpha$ -helix 3 and N-terminal VapB are required and should be investigated in the future.

### 3.3.2 Characterisation of the VapBC Complex from *Mycobacterium smegmatis*

The VapBC complex from *M. smegmatis* forms a complex of unknown stoichiometry. DLS revealed that the VapBC complex was monodisperse with a hydrodynamic radius of 51 Å. From this, the mass of the VapBC complex was calculated to be approximately 152 kDa. The calculated molecular mass of VapBC from size exclusion chromatography was approximately 155 kDa. This is consistent with the FitAB complex from *N. gonorrhoeae* which has a hydrodynamic radius of 51.1 Å and a molecular mass of 121 kDa (Mattison, *et al.*, 2006). The FitAB complex forms a tetramer of FitAB heterodimers when bound to an inverted repeat found in the *fitAB*

promoter. The difference in molecular masses between VapBC and FitAB is due to the assumption of a spherical particle in the DLS molecular mass calculation coupled with the unknown size of DNA purified bound to VapBC.



**Figure 3.23. Tetramer of FitAB heterodimers from *N. gonorrhoeae* bound to IR36 DNA fragment.** Pink and mauve ribbon diagrams denote FitA dimers binding to an inverted repeat sequence in the DNA. Surface diagrams depict FitB dimers, each FitB monomer is bound to a FitA monomer to form a heterodimers.

Anecdotal evidence suggests that VapBC is purified bound to DNA. Although the DNA cannot be seen on an agarose or polyacrylamide gel, upon removal of the DNA with DNase, VapB rapidly degrades. The removal of the DNA purified bound to VapBC could result in dissociation of the tetramer and as VapB is no longer bound in a rigid structure to DNA it is more susceptible to proteolytic degradation. No VapB degradation is observed upon RNase treatment, indicating that it is DNA not RNA that is co-purified with VapBC. It is unclear what degrades VapB. When expressing and purifying proteins from *M. smegmatis* it is routine to add protease inhibitors and

---

in all cases except this one protease contamination is not a problem. The incubation time and temperature while treating VapBC with DNase could not result in degradation of VapB as the equivalent temperatures and times were used for treatment with RNase and exonucleases and no degradation of VapB was observed. Therefore, I conclude that upon removal of the co-purified DNA VapB is particularly susceptible to very small amounts of contaminating proteases.

As mentioned earlier DLS data revealed the VapBC complex to be approximately 152 kDa, this is also consistent with it being purified bound to DNA. FitAB when not bound to DNA has a molecular weight of 98 kDa (Wilbur, *et al.*, 2005), whereas when bound to DNA FitAB has a molecular weight of 121 kDa. A molecular weight of 152 kDa suggests a tetramer of VapBC heterodimers bound to DNA.

The VapBC complex binds to inverted repeat sequences in its promoter DNA and the VapBC complex co-purifies with DNA but there is only one chromosome per *M. smegmatis* cell, meaning there is only one place for VapBC to bind. The inverted repeat sequence found in the promoter of *vapBC* that the VapBC complex binds (TATAGA-N<sub>13</sub>-TCTATA) is only present upstream of *vapBC* in the *M. smegmatis* mc<sup>2</sup>155 genome. If the sequence is shortened to ATAGA-N<sub>13</sub>-TCTAT by removing the two end bases it is still only present upstream of the *vapBC* operon. If the sequence is shortened to TATAG-N<sub>15</sub>-CTATA by removing the inner bases of the inverted repeat sequences it is still only present upstream of *vapBC* operon. If both the bases on the inside and the outside of the inverted repeat sequences are removed to give ATAG-N<sub>15</sub>-CTAT, the sequence is found in five places in the genome including upstream of *vapBC*. Changing the spacer region between two half sites bound by the FitAB complex did not abolish DNA binding, although FitAB bound with a diminished affinity (Wilbur, *et al.*, 2005). This led the authors to conclude that FitAB can bind independently to half sites (Wilbur, *et al.*, 2005). The two VapBC half sites (TATAGA and TCTATA) with a shorter spacer region (N<sub>12</sub>) were present in the promoter regions of four genes in the *M. smegmatis* genome. A

summary of the genes associated with variations of the inverted repeat sequences is shown in Table 3.1.

<b>TATAGA-N<sub>13</sub>-TCTATA</b>	
<b>Gene ID</b>	<b>Functional Annotation</b>
MSMEG_1283	Transcriptional regulator (VapB)
<b>ATAGA-N<sub>13</sub>-TCTAT</b>	
MSMEG_1283	Transcriptional regulator (VapB)
<b>TATAG-N<sub>13</sub>-CTATA</b>	
MSMEG_1283	Transcriptional regulator (VapB)
<b>ATAG-N<sub>13</sub>-CTAT</b>	
MSMEG_0025	Hypothetical
MSMEG_0945	Hypothetical
MSMEG_1283	Transcriptional regulator (VapB)
MSMEG_3177	Putative transcriptional regulator
MSMEG_3998	MmSAB operon regulatory protein
<b>TATAGA-N<sub>12</sub>-TCTATA</b>	
MSMEG_2154	Hypothetical
MSMEG_2991	Permease
MSMEG_3130	Hypothetical
MSMEG_4518	Hypothetical

**Table 3.1. Variations of the inverted repeat sequence found in the *vapBC* promoter region that VapBC binds.** Underneath each sequence is the places it is found in the genome of *M. smegmatis* mc<sup>2</sup>155. Gene ID is the gene associated with the sequence and its functional annotation. N is A, T C or G.

VapBC could regulate the expression of the genes in Table 3.1 but microarray analysis of conditional VapBC expression in *M. smegmatis*  $\Delta vapBC$  (Robson, 2010) revealed no repression of these genes with the exception of MSMEG\_2991 which is slightly downregulated (0.64 fold). Further work needs to be done to characterise the DNA that is co-purified with VapBC. Cloning and sequencing this DNA to determine what other sequences VapBC binds is a priority for future work.

The fact that VapBC is purified bound to DNA has made crystallisation of the VapBC complex very difficult, presumably due to the heterogeneity of the protein complex. Early crystallisation screens at a variety of concentrations and different buffers yielded no protein crystals. A sodium phosphate buffer (like that used for the purification of VapBC) has a tendency to form phosphate salt crystals with many solutions in the robot crystallisation screen. Dialysis into an organic buffer such as HEPES results in precipitation of the protein. Exchanging the DNA purified with the complex for a known DNA sequence was unsuccessful with only a small amount

---

exchanged and removal of the DNA resulted in degradation of VapB. Treatment of the VapBC complex with exonucleases III and VII was used to trim back the DNA that VapBC is bound to make the protein more amenable to crystallisation. Crystals were observed in the robot crystallisation screen, and fine screens around these conditions resulted in crystals that did not diffract. The fact that they did not diffract and that the crystals themselves were very soft is suggestive of protein crystals. Further fine screens and additive screens will be needed to confirm the presence of protein crystals.

Crystallisation attempts of VapC were unsuccessful. This is due to the purification method for VapC, the VapBC complex is digested with trypsin to remove VapB, and VapC is then purified further. A small amount of VapC degradation is observed and a mixed population of VapC with and without a C-terminal His-tag is seen with MALDI MS. Also, a small amount of VapC could have the peptide from VapB still bound, although SDS-PAGE electrophoresis, size exclusion chromatography and MALDI MS suggest this is not the case. The protein preparation is heterogeneous which does not promote protein crystallisation.

We have gained some information about the structure of the VapBC complex from *M. smegmatis*. A variety of different methods were trialled to isolate functional VapC, trypsin digestion of the VapBC complex was successful, yielding functional VapC protein. The next step was to determine the effect of VapC *in vivo* (Chapter Four) and to explore its biochemical function *in vitro* (Chapter Five).

---

## Chapter Four: Mechanism of Growth Inhibition by VapC in *Mycobacterium smegmatis*

### 4.1 Introduction

The cellular target of toxin proteins differs between different toxin-antitoxin (TA) families. CcdB and ParE inhibit DNA replication by binding to DNA gyrase however many toxin components of toxin-antitoxin systems are ribonucleases that result in inhibition of translation when released from their cognate antitoxin (Christensen-Dalsgaard, *et al.*, 2010; Christensen, *et al.*, 2001; Zhang, Zhang, Hoeflich, *et al.*, 2003). Toxin proteins RelE, MazF, Doc, HipA, VapC and HicA all inhibit translation through a variety of different mechanisms (Christensen, *et al.*, 2001; Hazan, *et al.*, 2001; Jorgensen, *et al.*, 2009; Korch & Hill, 2006; Zhang, Zhang, Hoeflich, *et al.*, 2003). When this study began, the cellular target of VapC had not been determined.

In the past year, two examples of the cellular targets of VapC have been published in the literature (Ramage, *et al.*, 2009; Winther & Gerdes, 2009). Expression of the VapC homologues Rv1031, Rv1561, Rv2829c and Rv3408 from *Mycobacterium tuberculosis* were shown to inhibit translation, preceding VapC induced growth inhibition (Ramage, *et al.*, 2009). Over expression of VapC<sub>LT2</sub> from *Salmonella* LT2 and VapC<sub>pMYSH</sub> from *Shigella* plasmid pMYSH6000 inhibited global translation but neither DNA replication or transcription, leading the authors to conclude that the target of these two VapC toxins is most likely to be found within the translation machinery (Winther & Gerdes, 2009).

Conditional expression of *vapC* from *Mycobacterium smegmatis* in both wild-type and  $\Delta vapBC$  backgrounds inhibited growth of *M. smegmatis* (Robson, *et al.*, 2009). This was greater in  $\Delta vapBC$  compared to the wild-type, presumably due to endogenous VapB neutralising VapC. Furthermore, expression of the VapBC complex had no effect on growth (Robson, *et al.*, 2009). These results from our collaborators raised the question as to the mechanism of growth inhibition by VapC

in *M. smegmatis*. To investigate this, the effect of VapC on the rate of DNA replication, transcription and translation were determined in both wild-type and  $\Delta vapBC$  *M. smegmatis* strains.

## 4.2 Results

### 4.2.1 Pulse-Chase Methionine, Uridine and Thymidine Incorporation

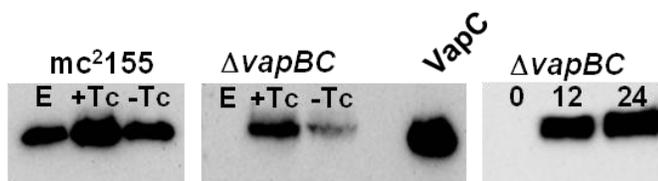
Pulse-chase experiments were conducted during exponential phase growth and DNA replication, transcription and translation were followed by the incorporation of [<sup>3</sup>H]thymidine, [<sup>3</sup>H]uridine and [<sup>35</sup>S]methionine respectively. The pulse-chase methodology involves “pulsing” with a mixture of labelled and unlabelled methionine and then chasing with unlabelled, nonradioactive (or ‘cold’) methionine to complete incorporation of [<sup>35</sup>S]methionine into protein. Radioactive/labelled methionine is mixed with unlabelled methionine so that incorporation of the isotope is linear during the pulse labelling. The rate at which the radioactivity disappears can be used to calculate the rate of turnover of a particular protein, but in this experiment the global rate of translation was measured. The pulse-chase methodology used by the Gerdes laboratory at Newcastle University (UK) for measuring DNA replication, transcription and translation in *Escherichia coli* was adapted for *M. smegmatis*. This involved a longer experimental time course as *E. coli* has a much faster doubling time compared with *M. smegmatis*. Due to this difference in doubling time longer pulse and chase periods were used; 15 minutes compared to one minute for *E. coli* and a longer chase period of 60 minutes compared to 15 minutes. A similar amount of labelled and unlabelled methionine to that used for *E. coli* was used in the *M. smegmatis* pulse-chase experiments.

During this experiment the cells were kept in exponential phase, by dilution of cultures at 12, 18 and 21 hours post-induction, to remove confounding factors of the cells entering stationary phase which decreases the rate of replication, transcription and translation. It has been reported that the greatest expression using the pMind vector occurs at 24 hours post-induction with 20 ng.ml<sup>-1</sup> tetracycline (Blokpoel, *et al.*,

2005). Therefore, a time of 24 hours post-induction with tetracycline was used for this experiment.

Polyclonal antibodies were raised against VapC purified in *E. coli* and showed excellent specificity for VapC when using *M. smegmatis* derived experiments (Figure 4.1). The presence of VapC protein in both the wild-type ( $mc^2155$ ) and  $\Delta vapBC$  strains harbouring and expressing pMind-*vapC* was determined by western blot. For each sample the same amount of total protein was loaded onto a gel for western blot analysis, allowing the expression of VapC to be compared between each sample. In the wild-type strain ( $mc^2155$ ) endogenous VapC was detected in the empty pMind control (Figure 4.1,  $mc^2155$  E) and the strain containing pMind-*vapC* without the presence of tetracycline (Figure 4.1,  $mc^2155$  -Tc). But greater amounts of VapC were detected in the wild-type strain containing pMind-*vapC* after induction with tetracycline (Figure 4.1,  $mc^2155$  +Tc).

As expected, no endogenous VapC was detected in the  $\Delta vapBC$  strain containing the empty pMind vector (Figure 4.1,  $\Delta vapBC$  E) but significant amounts of VapC were detected after induction with tetracycline in the  $\Delta vapBC$  strain harbouring pMind-*vapC* (Figure 4.1,  $\Delta vapBC$  +Tc). A small amount of VapC was detected in the knockout strain harbouring pMind-*vapC* in the absence of inducer (tetracycline) (Figure 4.1,  $\Delta vapBC$  -Tc) due to the “leakiness” of the pMind vector. Purified VapBC



**Figure 4.1.** Western blot showing VapC protein from  $mc^2155$  containing pMind (empty vector, denoted E), pMind-VapC after induction with tetracycline (denoted +Tc) and pMind-VapC without induction (denoted -Tc). For the  $\Delta vapBC$  strain, E, empty vector control, +Tc, induced and -Tc, uninduced VapC expression is shown. VapC denotes the positive control for the experiment (purified VapBC complex).  $\Delta vapBC$  0, 12 and 24 represents hours VapC detected post-induction.

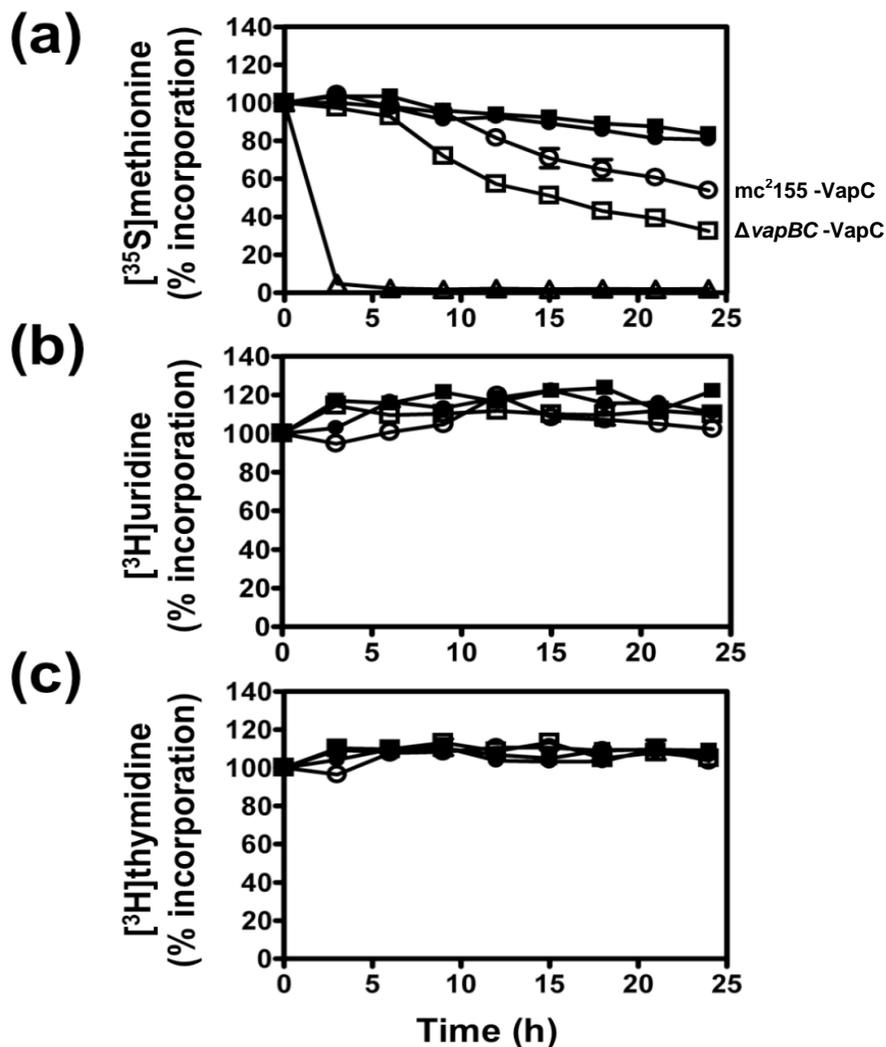
complex was included as a positive control for the western blot (Figure 4.1, VapC). In the  $\Delta vapBC$  strain VapC expression is present 12 and 24 hours post-induction, as seen in the western blot on the right in Figure 4.1, confirming protein production over the time course of the experiment.

---

Once expression of VapC was confirmed by western blot (Figure 4.1) the effect of VapC on DNA replication, transcription and translation was determined. A reduction in protein synthesis was seen in cells expressing VapC in both the wild-type and  $\Delta vapBC$  backgrounds as in Figure 4.2a. There was a more marked reduction of translation in the knockout strain compared with the wild-type. This is consistent with growth inhibition of VapC (Robson, *et al.*, 2009) and is due to endogenous VapB in the wild-type which offsets the effect of induced VapC from the pMind vector, compared to the knockout strain which lacks endogenous VapB. As a positive control cultures were treated with  $50 \mu\text{g}\cdot\text{ml}^{-1}$  chloramphenicol, a known inhibitor of translation. This reduced the rate of translation to virtually zero after three hours whereas induction of VapC slowly reduced the rate of translation over 24 hours by 50% in the  $\Delta vapBC$  strain and approximately 30% in the wild-type. These values take into account the effect of tetracycline on translation as tetracycline itself is an inhibitor of translation.

No effect on DNA replication or transcription was seen for the conditional expression of VapC in both the wild-type and  $\Delta vapBC$  strains as in Figure 4.2b and c.

Western blot results (Figure 4.1) confirm the induction of VapC protein in both wild-type and knockout strains harbouring pMind-*vapC*, confirming that the reduction of protein synthesis is due to the presence of VapC and not some spurious effect of the protein expression generally.

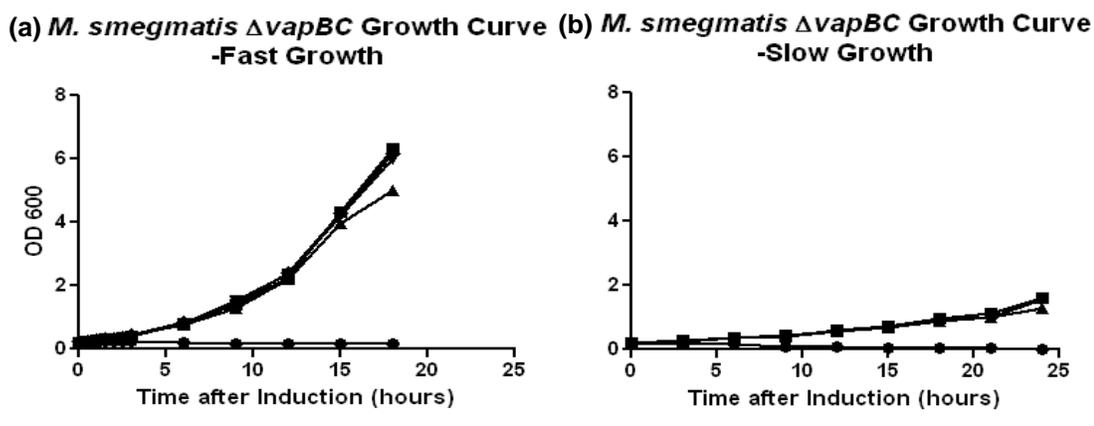


**Figure 4.2. Effect of VapC on (a) translation, (b) transcription and (c) DNA replication in *M. smegmatis*.** 100% refers to the rate of translation, transcription or DNA replication pre-induction. Time, refers to hours post-induction with tetracycline. *M. smegmatis* strain  $mc^2155$  harbouring plasmid pMIND (vector only) (closed circle), or pMIND-vapC (expressing VapC toxin) (open circle) and VapBC deletion JR121 strain harbouring the same plasmid pMIND (closed square), or pMIND-vapC (open square) were grown in HdB medium supplemented with hygromycin. Exponentially growing cultures were induced with tetracycline ( $20 \text{ ng.ml}^{-1}$ ) at time zero and the rates of (a) translation (b) transcription and (c) DNA replication were determined by the incorporation of  $[^{35}\text{S}]$  methionine,  $[^3\text{H}]$  uridine and  $[^3\text{H}]$  thymidine respectively. Chloramphenicol, an inhibitor of translation was included in (a) as a positive control (closed triangles).

#### 4.2.2 Methionine Uptake to Measure Translation Inhibition

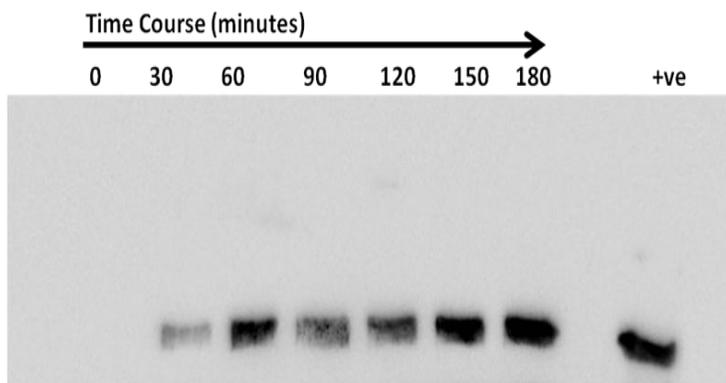
Translation inhibition experiments described above were repeated to determine the effect of the VapBC complex on translation to answer the question does the VapB antitoxin neutralise the effect of VapC, the toxin? In the above experiment (Section 4.2.1) growth of *M. smegmatis* cell cultures in HdB minimal media was very slow.

Back in the Proteins & Microbes Laboratory, University of Waikato, Hamilton, culturing *M. smegmatis* in HdB medium proved problematic. This was remedied by the use of tissue culture flasks to grow cultures. Tissue culture flasks are RNase and DNase free, completely sterile and non-pyrogenic, compared to glassware which although sterile can contain traces of detergents and other chemicals which can inhibit or slow the growth of *M. smegmatis* cultures. Although the shaking speed and the ratio of culture to head space was similar between the glass and tissue culture flasks there are differences in aeration of the cultures. Cultures in tissue culture flasks received more aeration due to the angular shape of the flask, compared to the round bottom glass flasks. These factors meant *M. smegmatis* cultures had a faster growth rate with a doubling time of three hours when grown in tissue culture flasks (Figure 4.3b) compared to the slow growth rate with a doubling time of around nine hours when grown in glass flasks (as grown at Newcastle University, UK) (Figure 4.3a).



**Figure 4.3.** Growth curves of *M. smegmatis*  $\Delta vapBC$  (square),  $\Delta vapBC$  harbouring pMind (empty vector) (inverted triangle), pMind-*vapC* (triangle) and pMind-*vapBC* (diamond) in HdB minimal media post-induction with tetracycline. Cultures were kept in exponential phase by dilution of cultures at 12, 18 and 21 hours post-induction. (a) *M. smegmatis* at a fast growth rate (tissue culture flasks) compared to (b), *M. smegmatis* at a slow growth rate (glass flasks). Chloramphenicol (positive control) (circle) inhibits growth under fast and slow growth conditions.

Due to the difference in the growth rate of *M. smegmatis* between the two laboratories (University of Waikato, Hamilton, NZ and Newcastle University, UK) it was important to determine expression of VapC at earlier time points.



**Figure 4.4. Time course of VapC expression.** Western blot showing a time course (minutes) of conditional expression of VapC from pMind-*vapC* in the  $\Delta vapBC$  background in fast growth cultures after induction with tetracycline (0 minutes represents pre-induction). The rightmost lane denotes the positive control for the experiment (purified VapBC complex).

Western blot analysis determined that VapC protein is present 30 minutes post-induction (Figure 4.4, 30 minutes) and the amount of VapC protein present increases 60 minutes post-induction, then remains at this level for the remainder of the time course (180 minutes) (Figure 4.4). The amount of VapC present three hours post-induction (Figure 4.4, 180 minutes) is comparable to that in Figure 4.1, where VapC was detected 24 hours post-induction. Based on this a time point of three hours was used in the subsequent experiment.

The greatest expression using pMind has been reported to occur 24 hours post-induction with 20 ng.ml<sup>-1</sup> tetracycline (Blokpoel, *et al.*, 2005). In previous experiments (Section 4.2.1), the greatest induction of VapC occurred 24 hours post-induction as this is when the inhibition of translation is greatest, but the cultures were at a much slower growth rate. Therefore, expression of VapC would be slow and the rate of translation would be reduced as a result of the slow growth rate. No VapC expression is detected pre-induction with tetracycline (Figure 4.4, 0 minutes) indicating no “leakiness” of the vector at this point in growth and as expected no endogenous VapC in the  $\Delta vapBC$  strain.

A new experimental design was needed to implement the translation inhibition experiments under faster growth rates. The pulse-chase methodology was variable and inconsistent when measuring methionine incorporation for faster growing *M. smegmatis* cultures. Inclusion of [<sup>35</sup>S]methionine in the culture from inoculation

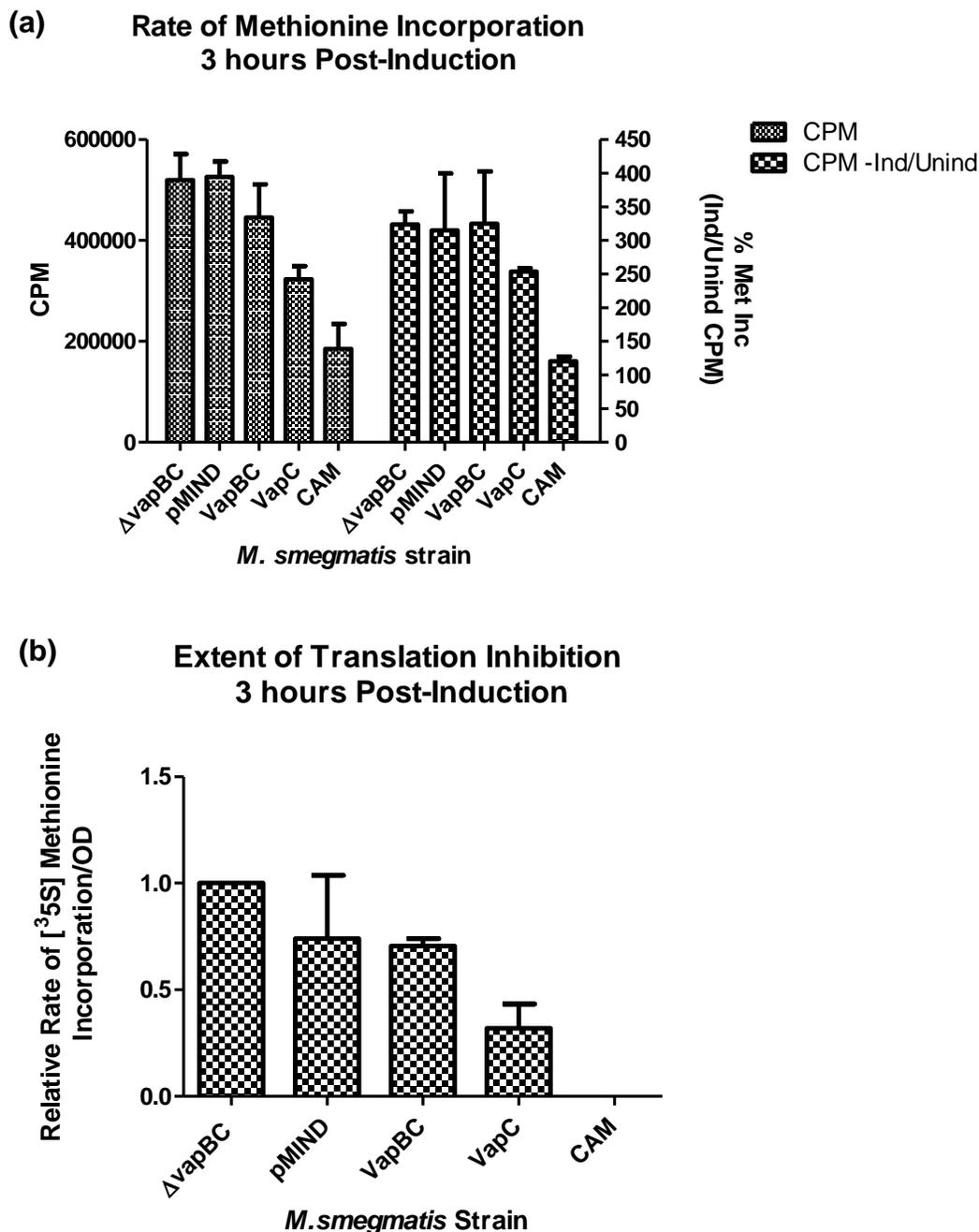
proved to be successful and gave more reliable, consistent results compared with the pulse-chase methodology.

In this new method [<sup>35</sup>S]methionine was incorporated into protein before induction. Therefore the rate of translation can be determined by comparing translation rates before and after addition of tetracycline to induce VapC expression. To ensure the amount of [<sup>35</sup>S]methionine would not limit translation in minimal media a standard curve of methionine counts was determined. From this it was determined that 0.5  $\mu\text{Ci}$  [<sup>35</sup>S]methionine per ml of culture should be added so incorporation is linear.

These experiments were conducted in the  $\Delta\text{vapBC}$  background only, as the effect of the VapBC complex on translation would be more pronounced in this strain with no interference from endogenous *vapBC* as seen in previous experiments (Figure 4.2). Induction of VapC and VapBC were compared against pMind (empty vector) and  $\Delta\text{vapBC}$  only. As a positive control, cultures were treated with 50  $\mu\text{g.ml}^{-1}$  chloramphenicol. Based on western blot analysis of VapC expression (Figure 4.4) samples were taken at 30 minute intervals up to three hours post-induction along with a pre-induction sample.

The raw data (counts per minute (CPM)), data normalised to the pre-induction sample (Figure 4.5a) and data normalised to the uninduced sample and cell number (OD) (Figure 4.5b) all show the same trend. Three hours post-induction the amount of methionine incorporated has decreased upon induction of VapC and upon treatment with chloramphenicol (Figure 4.5A). The rate of translation for cells treated with chloramphenicol drops to 0% (Figure 4.5).

Assuming that upon chloramphenicol treatment the rate of translation is zero and the rate of translation for the  $\Delta\text{vapBC}$  strain is 100%, the rate of translation three hours post-induction in the  $\Delta\text{vapBC}$  strain harbouring pMind-*vapC* drops to 65%, therefore there is a 35% drop in the rate of translation three hours post-induction of VapC.



**Figure 4.5. Effect of VapC on translation.** *M. smegmatis*  $\Delta vapBC$  strains harbouring pMind, pMind-*vapC* or pMind-*vapBC* were grown in HdB media supplemented with hygromycin. Exponentially grown cultures were induced with tetracycline and the rate of translation measured by incorporation of [<sup>35</sup>S] methionine after 3 hours. (a) Represents raw CPM data and processed CPM data normalised to the pre-induction sample. (b) The relative rate of [<sup>35</sup>S] methionine incorporation/OD represents processed CPM data that has been normalised to growth (OD). The data for pMIND, VapBC and VapC are relative to  $\Delta vapBC$  which represents 100% translation and chloramphenicol (CAM) which represents 0% translation.

### 4.3 Discussion

Many well studied TA systems like MazEF and RelBE have been shown to inhibit translation when the toxin is released from its cognate antitoxin (Christensen-Dalsgaard, *et al.*, 2010; Christensen, *et al.*, 2001; Zhang, Zhang, Hoeflich, *et al.*, 2003). Conditional expression (tetracycline regulated) of VapC in both wild-type and  $\Delta vapBC$  strains of *M. smegmatis* reduces translation (protein synthesis) and not DNA replication and transcription. This is consistent with VapC homologues from *M. tuberculosis* which have also been shown to reduce translation, although their effect on DNA replication and transcription was not determined (Ramage, *et al.*, 2009).

Previous work demonstrated that conditional expression of VapC in *M. smegmatis* resulted in inhibition of growth although no difference in OD was visible at 24 hours due to elongation of the bacterial cells a difference in colony forming units (CFUs) was observed. This effect was neutralised when co-expressed with its cognate antitoxin VapB (Robson, *et al.*, 2009). This inhibition of growth is presumably due to the effect of VapC on the rate of translation. This was one of the first studies to identify that VapC reduces the rate of translation. VapB neutralises the toxicity of VapC as no reduction of translation was observed when VapB and VapC are co-expressed. Translation inhibition studies on other TA systems also demonstrate this, for example RelB counteracts RelE mediated inhibition of translation (Pedersen, *et al.*, 2002) and translation inhibition by MazF can be relieved by expression of MazE (Zhang, Zhang, Hoeflich, *et al.*, 2003).

The reduction of translation due to the conditional expression of VapC was less pronounced in the wild-type compared to  $\Delta vapBC$  due to endogenous VapB. Western blot analysis demonstrated that endogenous VapC was present in the wild-type strain, therefore it is likely VapB will also be present, possibly in excess as VapBC had no effect on growth of *M. smegmatis* (Robson, *et al.*, 2009). The slight drop in translation seen in the negative controls over longer time periods (empty pMind

---

vector in both wild-type and  $\Delta vapBC$  background) could be due to the effect of the inducer tetracycline which is a known inhibitor of translation (Figure 4.2a).

Conditional expression of VapC reduces the rate of protein synthesis but has no effect on the rate of DNA replication or transcription. Over a longer time period (greater than 24 hours) or the same time period under faster growth conditions, it would be expected that inhibition of translation would lead to a decrease in the rate of DNA replication and transcription. A reduction in translation results in less protein being synthesised, leading to a decrease in the cellular levels of core proteins, thereby affecting cellular processes such as transcription and DNA replication.

In the  $\Delta vapBC$  *M. smegmatis* strain, under fast growth conditions a maximal reduction in translation is visible three hours post-induction with tetracycline. However under slower growth conditions maximal VapC reduction of translation is visible 24 hours post-induction with tetracycline. Under both fast and slow growth conditions the extent to which VapC reduced translation was similar, but the time points differed as a result of the different growth rates.

Two methods of methionine incorporation were used to measure the rate of translation. The pulse-chase methodology involved taking samples of an exponentially growing culture and pulsing them with a mixture of labelled and unlabelled methionine and then chasing with unlabelled methionine to complete incorporation of [<sup>35</sup>S] methionine into protein. This was suited to slow growth cultures, but when applied to fast growth cultures results were variable and inconsistent. Inclusion of [<sup>35</sup>S] methionine in the media of fast growing cultures gave reliable, reproducible results due to decreased sample handling time and by allowing a steady uptake of [<sup>35</sup>S] methionine throughout growth.

Many TA systems arrest cell growth by inhibition of translation (Christensen, *et al.*, 2001; Zhang, Zhang, Hoeflich, *et al.*, 2003) and the mechanism of translation inhibition has been well characterised for the MazEF and RelBE TA systems; RelE

---

associates with ribosomes (Christensen & Gerdes, 2003; Christensen & Gerdes, 2004) whereas MazF cleaves free mRNA in the cytoplasm independent of the ribosome (Zhang, Zhang, Hoeflich, *et al.*, 2003). The molecular mechanism of how VapC reduces translation was unknown however these results were consistent with ribonuclease activity seen for other VapC proteins from *Haemophilus influenzae* (Daines, *et al.*, 2007) and *Mycobacterium tuberculosis* (Daines, *et al.*, 2007). The potential ribonuclease activity of VapC proteins from *M. smegmatis* and *Pyrobaculum aerophilum* was investigated and is described in Chapter Five.

---

## Chapter Five: VapC Ribonuclease Activity

### 5.1 Introduction

This chapter will focus on the characterisation of the ribonuclease activity of three VapC proteins: VapC<sub>P<sub>AE</sub>2754</sub> and VapC<sub>P<sub>AE</sub>0151</sub> from *Pyrobaculum aerophilum* and VapC<sub>MS1284</sub> from *Mycobacterium smegmatis*. The ribonuclease activity of the toxic components of other TA families has been well investigated, however little is known about the proposed ribonuclease activity of VapC proteins and the catalytic mechanism of both prokaryotic and eukaryotic VapC/PIN domain proteins.

#### 5.1.1 Ribonuclease Activity of Toxin Proteins

Many toxic components of TA systems display ribonuclease activity (Christensen & Gerdes, 2003; Hurley & Woychik, 2009; Munoz-Gomez, Santos-Sierra, Berzal-Herranz, Lemonnier, & Diaz-Orejas, 2004; Zhang, Zhang, Hoeflich, *et al.*, 2003; Zhang, Zhu, *et al.*, 2005). Toxin proteins of the RelBE family generally associate with ribosomes to cleave mRNA themselves or induce ribosome-mediated cleavage of the mRNA. For example RelE from *Escherichia coli* initiates RNA cleavage in the ribosomal A site at stop codons (Pedersen, *et al.*, 2003) and HigB cleaves RNA at AAA triplet sequences and to a lesser extent AA sequences and occasionally at single A's. HigB (a member of the RelBE TA family) cleaves RNA, (unlike RelE which lacks the catalytic triad required for activity) but associates with the ribosome to do so (Hurley & Woychik, 2009). The ribonuclease activity of toxins belonging to the MazEF TA family have been well characterised (Zhang, Zhang, *et al.*, 2005; Zhang, Zhu, *et al.*, 2005). Activity of these proteins is not dependent on the ribosome nor do they induce ribosome-mediated cleavage of mRNA. Due to their endoribonuclease activity they have been termed mRNA interferases. MazF from *E. coli* targets ACA sequences, cleaving between A and C and sometimes before the first A (Zhang, Zhang, Hoeflich, *et al.*, 2003), and it cleaves only ssRNA, not RNA/RNA or RNA/DNA duplexes. Recognition sites of MazF homologues differ in different organisms; the recognition site of MazF from *Staphylococcus aureus* differs to that of its Gram negative counterparts and recognition

---

sequences also differ within the same organism. MazF toxins from *Mycobacterium tuberculosis* target different recognition sequences (Zhu, *et al.*, 2008; Zhu, *et al.*, 2006). This is unsurprising as the sequence conservation across the MazF family is low (~28%) (Mittenhuber, 1999).

The catalytic mechanism of the MazF endoribonuclease family has been well determined and is similar to that of Ribonuclease A. MazF does not require  $Mg^{2+}$  for activity and cleaves the phosphodiester bond of RNA at the 5' end to yield a free 5' OH group on the 3' cleavage product and a 2'-3' cyclic phosphate on the 3' end of the 5' cleavage product (Zhang, Zhang, *et al.*, 2005).

Of the few VapC/PIN domain proteins characterised, most demonstrate ribonuclease activity (Daines, *et al.*, 2007; Ramage, *et al.*, 2009). VapC from non-typeable *Haemophilus influenzae* degrades free RNA *in vitro* and not dsDNA or ssDNA (Daines, *et al.*, 2007) and VapC-5 from *M. tuberculosis* displays  $Mg^{2+}$ -dependent ribonuclease activity (Miallau, *et al.*, 2009), although the catalytic mechanism or sequence specificity of these enzymes has not been determined. VapC<sub>PAE2754</sub> from *P. aerophilum* displays weak exonuclease activity against a dsDNA oligonucleotide with a large 5'-3' overhang (Arcus, *et al.*, 2004).

### 5.1.2 Ribonuclease Activity of Eukaryotic PIN-Domain Proteins

PIN domain proteins from eukaryotes are associated with nonsense mediated decay (NMD) of RNA (Huntzinger, Kashima, Fauser, Sauliere, & Izaurralde, 2008) and processing of pre-18S rRNA fragments (Lamanna & Karbstein, 2009). The human PIN-domain proteins, SMG-5 and SMG-6 play essential roles in NMD pathways (Huntzinger, *et al.*, 2008). SMG-5 and SMG-6 have similar folds, however SMG-5 lacks three of the four conserved acidic residues required for catalysis and SMG-6 contains all four. The C-terminus of SMG-6 contains a PIN domain that is a  $Mn^{2+}$  dependent (and to a lesser extent  $Mg^{2+}$  dependent) endoribonuclease specific for ssRNA. It demonstrates no activity against dsDNA, ssDNA or dsRNA (Glavan, Behm-Ansmant, Izaurralde, & Conti, 2006) but ssDNA does inhibit SMG-6 RNA degradation, presumably

---

deoxyribose nucleotides are able to bind but the 2'OH of ribonucleotides is required for activity. The human protein Est1A contains a PIN domain and is involved in regulation of telomere elongation and NMD of RNA (Takeshita, Zenno, Lee, Saigo, & Tanokura, 2007). The PIN domain containing protein Nob1 is required for processing of pre-18S rRNA fragments. Its PIN domain binds to cleavage site D of the single-standed pre-18S rRNA fragment, although cleavage at this site has yet to be observed (Lamanna & Karbstein, 2009).

### 5.1.3 Choice of Substrate for Ribonuclease Activity Assays

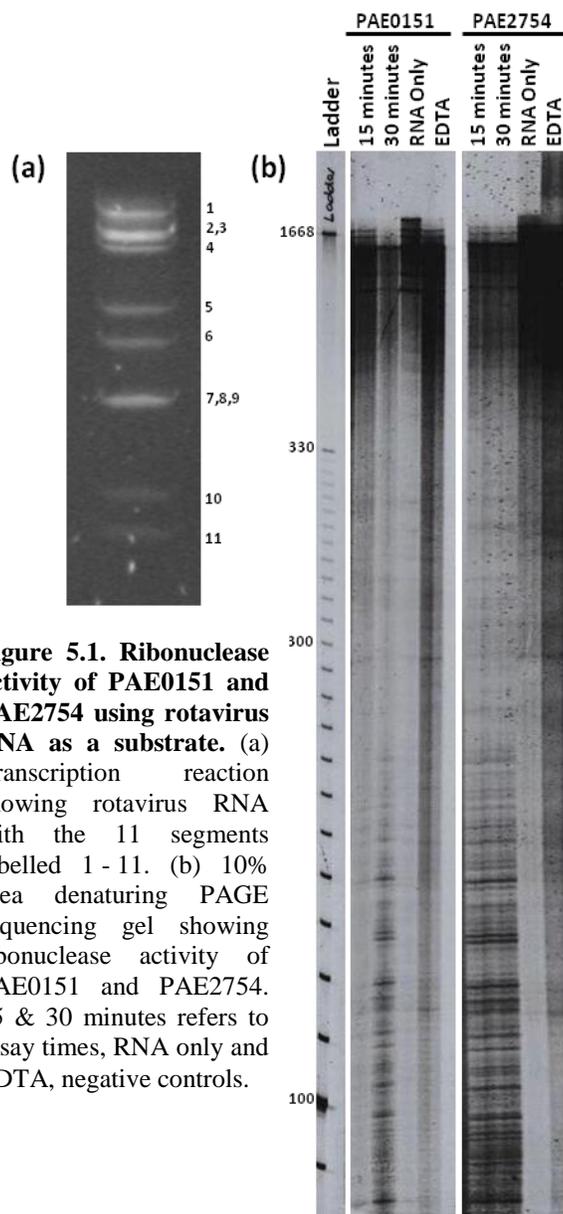
When determining the sequence specificity of ribonuclease enzymes it is important to consider the properties of the substrate. Ideally, to determine the recognition sequence the substrate should have an equal concentration of each base (A, U, C and G), be complex enough to determine recognition sequences greater than three bases and have little secondary structure. This is hard to find in a RNA substrate as the longer the RNA molecule becomes the more secondary structure it forms. A common substrate for determining recognition sequences of toxin proteins has been MS2 bacteriophage RNA (Ramage, *et al.*, 2009; Zhu, *et al.*, 2006). MS2 bacteriophage RNA is a good substrate for determining cleavage specificity of toxin proteins as it is commercially available, is 3569 bases in length so is complex enough to determine cleavage sites longer than three bases and contains an approximately equal number of each base (A, G, C & U). However, it does form extensive secondary structure, so requires the use of CspA (a major cold shock protein from *E. coli*) as an RNA chaperone.

In this chapter a range of ribonuclease activity assays were investigated and an optimal assay developed to determine cleavage sites of select VapC proteins.

## 5.2 Results

### 5.2.1 Rotavirus RNA Ribonuclease Activity Assays

Rotavirus double layered particles (DLPs) produce 11 segments of ssRNA, these segments vary in size from 3302 bases (segment 1) to 667 bases (segment 11) as in Figure 5.1a. These RNA segments can then be used as a substrate in VapC ribonuclease



**Figure 5.1. Ribonuclease activity of PAE0151 and PAE2754 using rotavirus RNA as a substrate.** (a) Transcription reaction showing rotavirus RNA with the 11 segments labelled 1 - 11. (b) 10% urea denaturing PAGE sequencing gel showing ribonuclease activity of PAE0151 and PAE2754. 15 & 30 minutes refers to assay times, RNA only and EDTA, negative controls.

activity assays. In total there is a relatively even distribution of bases. Therefore, almost every combination of up to five bases is covered. Using this substrate it would be difficult to determine the sequence specificity of VapC proteins as it would be unknown which of the 11 RNA segments is being cut. Both magnesium and manganese were used as cofactors for VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> in early ribonuclease activity assays using rotavirus RNA as a substrate. The concentration of manganese sulphate in the assay reaction was dropped to one mM from ten mM as the higher concentration resulted in precipitation of the RNA. Equivalent activity of both VapC proteins was observed with each cofactor therefore magnesium was used as a cofactor, in all subsequent experiments.

Rotavirus RNA was successfully labelled with <sup>33</sup>P-UTP. Ribonuclease activity of VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> from *P. aerophilum* was tested using the radiolabelled

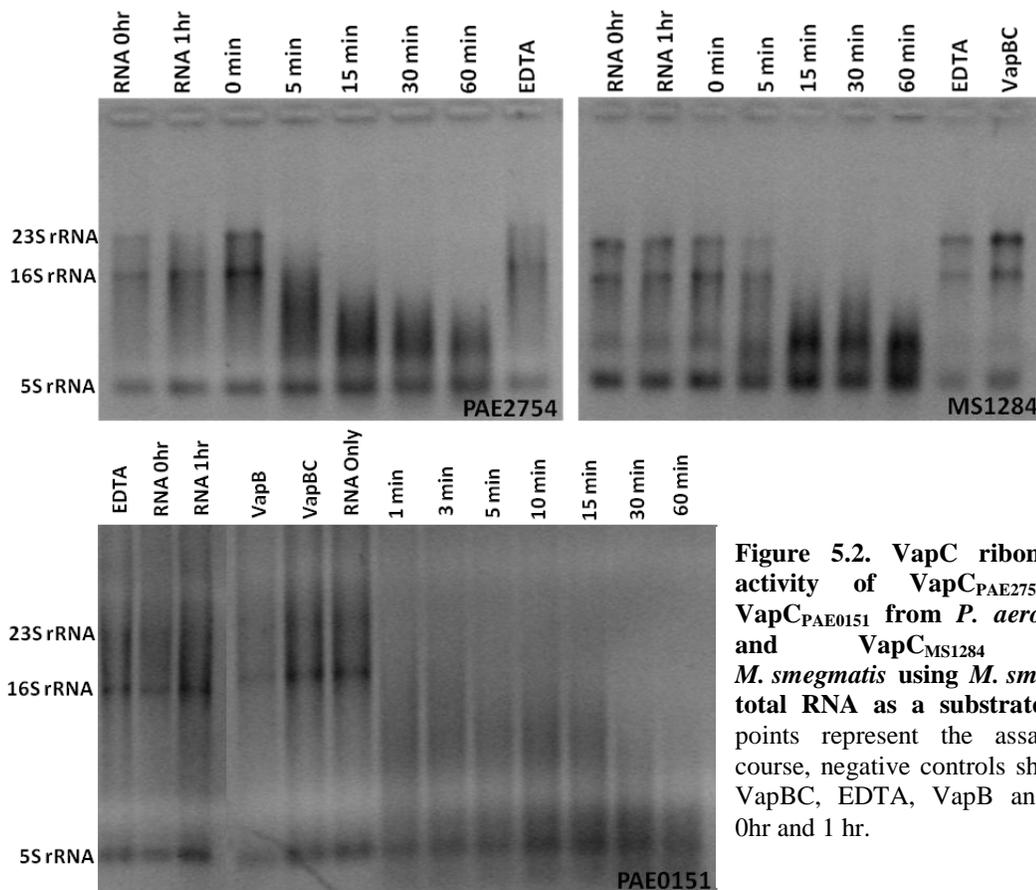
---

rotavirus RNA as a substrate. Assay reactions were run on a urea-denaturing polyacrylamide sequencing gel to obtain good resolution of lower molecular weight assay products as in Figure 5.1b. Although the gel is overexposed due to the amount of radioactivity in the sample, there is a clear difference between the negative controls (RNA only and EDTA) and the 15 and 30 minute assay samples. The negative control containing 20 mM EDTA in the assay buffer has a small amount of degradation, presumably due to trace amounts of  $Mg^{2+}$  present or trace ribonuclease contamination from the protein purification preparations. There is very little RNA degradation in the RNA only negative controls, these reactions are set up the same as the assay reactions but no VapC protein added, demonstrating there is no ribonuclease contamination in the assay buffer. Both VapC<sub>PAE0151</sub> and VapC<sub>PAE2754</sub> appear to target the same sequence as the banding pattern for the two enzymes are very similar, VapC<sub>PAE2754</sub> appears to act faster than VapC<sub>PAE0151</sub> although it is hard to tell from this experiment. When these experiments were conducted VapC<sub>MS1284</sub> from *M. smegmatis* had not been isolated. Therefore its ribonuclease activity was not tested against this substrate.

### 5.2.2 Total RNA from *Mycobacterium smegmatis* Ribonuclease Activity Assays

Many toxins cleave total RNA *in vitro* although this is very unlikely to occur *in vivo* due to tight packaging of ribosomal RNA and protection of rRNA by ribosome associated proteins. For example MazF from *E. coli* targets ACA sequences and cleaves 23S and 16S rRNA *in vitro* but not *in vivo* (Zhang, Zhang, Hoeflich, *et al.*, 2003). Total RNA could be used to screen for ribonuclease activity *in vitro*.

Total RNA was isolated from *M. smegmatis* (Section 2.4.5) was used as a substrate to detect VapC ribonuclease activity. Determining the sequence specificity of VapC proteins using total RNA is difficult, but the substrate was used for general testing of VapC proteins.



**Figure 5.2.** VapC ribonuclease activity of VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> from *P. aerophilum* and VapC<sub>MS1284</sub> from *M. smegmatis* using *M. smegmatis* total RNA as a substrate. Time points represent the assay time course, negative controls shown by VapBC, EDTA, VapB and RNA 0hr and 1 hr.

VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> from *P. aerophilum* and VapC<sub>MS1284</sub> from *M. smegmatis* all display Mg<sup>2+</sup> dependent ribonuclease activity against total RNA from *M. smegmatis*; when EDTA is included in the assay buffer ribonuclease activity is abolished. The VapC proteins degrade 23S rRNA and 16S rRNA to smaller products but not 5S rRNA as this band (lowest molecular weight band in Figure 5.2) is still present after 60 minutes with the VapC enzymes while the 23S and 16S rRNA bands are not.

Negative controls are extremely important in this assay to check there is no ribonuclease contamination of buffers, protein preparations etc. RNA 0 hr and RNA 1 hr samples show no RNA degradation indicating that there are no contaminating RNases present in the assay buffer. VapB<sub>PAE0152</sub> is the cognate antitoxin to VapC<sub>PAE0151</sub>. VapB<sub>PAE0152</sub> is purified in the same way as VapC<sub>PAE0151</sub> and shows no ribonuclease activity in Figure 5.2 above. VapBC TA complexes display no ribonuclease activity in Figure 5.2, implying that VapB inhibits the ribonuclease activity of VapC. Inclusion of EDTA to

---

inhibit VapC activity is also a negative control, as contaminating RNases like Ribonuclease A are not metal-dependent, therefore if metal-independent RNases were present in the protein preparation degradation of RNA would still be observed upon inhibition of VapC. Unfortunately, VapB<sub>PAE2755</sub>, the antitoxin (VapB) to VapC<sub>PAE2754</sub>, was insoluble. Therefore the PAE2754/5 VapBC complex could not be constructed, however it is assumed that VapB<sub>PAE2755</sub> would inhibit ribonuclease activity of VapC<sub>PAE2754</sub> like the other VapC proteins tested here.

### 5.2.3 Fluorogenic RNA Substrate

A fluorogenic substrate (termed FrG) consisting of a single RNA base, flanked by short DNA sequences and labelled with a fluorophore at one and a quencher at the opposite end was used to test the ribonuclease activity of three VapC proteins (Section 2.4.7). Upon cleavage of the substrate fluorescence is observed. No cleavage of FrG was observed by VapC<sub>PAE2754</sub>, VapC<sub>PAE0151</sub> or VapC<sub>MS1284</sub> in their appropriate assay buffers as there was no increase in fluorescence upon addition of the proteins. The sequence of the substrate is <sup>5'</sup>6-FAM-AAGTCrGACATCAG-BHQ-1<sup>3'</sup> it is therefore likely that the three VapC proteins tested do not cleave at CGA sequences. But it is possible that the overall DNA nature of the substrate might affect cleavage, not the sequence per se. A positive control, the general ribonuclease, Barnase demonstrated cleavage of this substrate (data not shown).

### 5.2.4 Two Base Combination RNA Oligonucleotide Ribonuclease Activity Assays

An RNA oligonucleotide containing every combination of two bases was designed to test the sequence specificity of the three VapC proteins (Section 2.4.6). The reasoning behind every combination of two bases was that it is unlikely the VapC proteins will have a recognition sequence of two bases (as this would make it an extremely potent toxin), but this still has to be tested and if a recognition sequence of greater than two bases is determined in other experiments the bases essential for substrate recognition could be determined using information from this experiment. The two base combination sequence was flanked by GAGA at either end as it was known from the assays on the

---

fluorometric substrate (Section 5.2.3) that VapC<sub>PAE2754</sub>, VapC<sub>PAE0151</sub> and VapC<sub>MS1284</sub> do not cut at GA sequences.

The RNA oligonucleotide was labelled at the 5' end with <sup>33</sup>P, to increase the sensitivity of the assay (Section 2.4.6). Assay reactions were set up as in Section 2.4.10, run on a 20% urea-denaturing PAGE gel and the gel exposed to film overnight. A variety of ratios of labelled to unlabelled RNA were trialled. However, all showed no degradation of the RNA oligonucleotide upon addition of VapC<sub>PAE2754</sub>, VapC<sub>PAE0151</sub> or VapC<sub>MS1284</sub> (data not shown). This indicates that the recognition sequence for these three VapC proteins is greater than two bases.

### 5.2.5 Pentaprobe RNA Ribonuclease Activity Assays

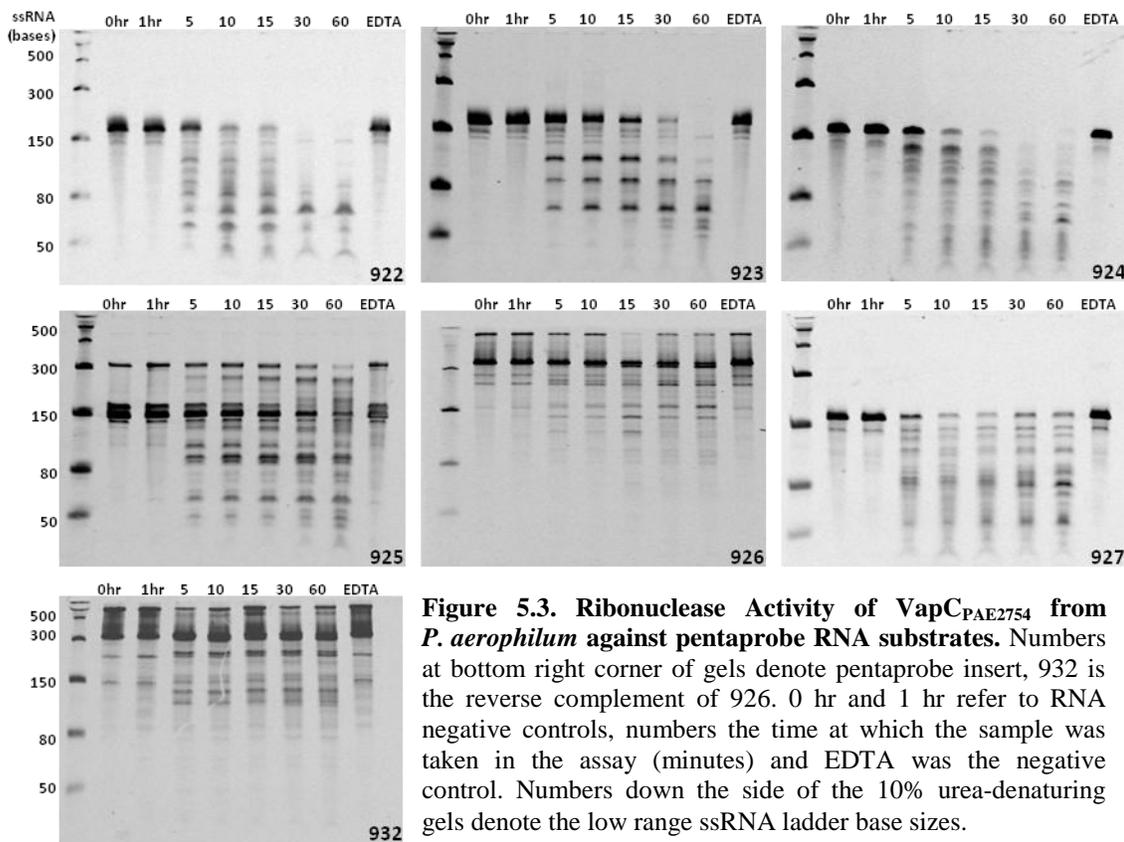
The pentaprobe pack was generated using a computer algorithm; which determined the minimum sequence to cover every combination of five bases contains 516 bases (Kwan, *et al.*, 2003). This sequence was then made as six overlapping dsDNA molecules from which RNA can be transcribed, resulting in RNA segments covering every combination of five bases. The reverse pentaprobe sequences also allow dsRNA to be synthesised. The use of pentaprobe RNA reduces the amount of secondary structure present in an RNA substrate while still allowing screening of up to pentad recognition sequences.

The pentaprobe pack contains a set of six plasmids whose overlapping inserts contain every combination of five bases. The reverse complement of these inserts is also available across the inserts of another six plasmids. From these plasmids dsDNA, ssDNA, dsRNA and ssRNA can be synthesised, therefore the activity of the VapC proteins can be tested across all these substrate types. Previous experiments with rotavirus RNA as a substrate (Section 5.2.1) have shown that VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> cleave ssRNA, therefore it was assumed to be the case for VapC<sub>MS1284</sub>. All three proteins demonstrated ribonuclease activity against total RNA from *M. smegmatis*.

Transcription of the pentaprobe RNA was first attempted using linearised plasmid that had been end-filled with klenow (Section 2.4.8.1.1). The pcDNA3 plasmid was digested

with *ApaI* which cuts the plasmid after the pentaprobe insert. Transcription using linearised pcDNA3 as a template was unsuccessful as the pcDNA3 vector does not have a T7 transcriptional terminator and therefore, very large molecular weight RNA was observed. Only a small amount of undigested, whole plasmid needs to be present for large molecular weight RNA to be formed. RNA of the correct size was also present but this would be difficult to purify away from the larger molecular weight RNA.

Construction of a PCR product including the T7 promoter was successful at producing only one, correct pentaprobe RNA molecule. To narrow down the sequence specificity of VapC proteins, ssRNA was transcribed from the PCR product of the plasmid insert. All six pentaprobe inserts were transcribed into RNA and subsequently used as substrates in VapC ribonuclease activity assays.

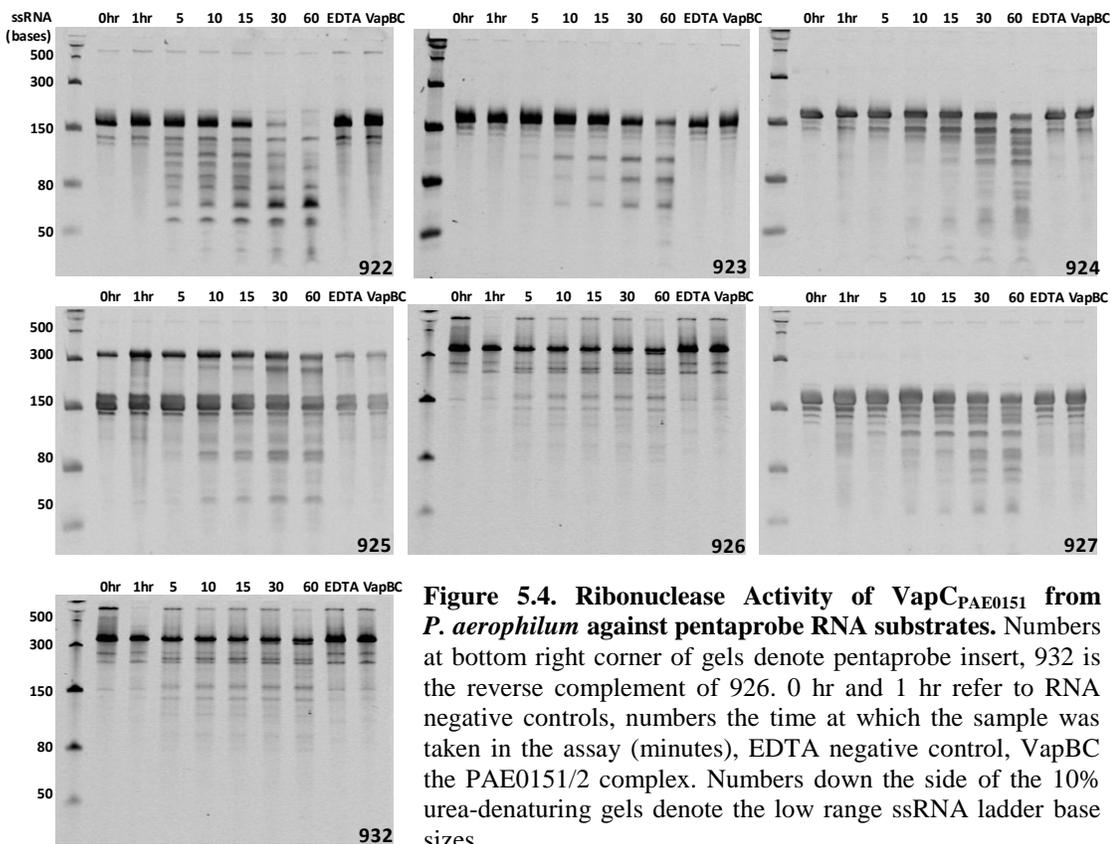


**Figure 5.3. Ribonuclease Activity of VapC<sub>PAE2754</sub> from *P. aerophilum* against pentaprobe RNA substrates.** Numbers at bottom right corner of gels denote pentaprobe insert, 932 is the reverse complement of 926. 0 hr and 1 hr refer to RNA negative controls, numbers the time at which the sample was taken in the assay (minutes) and EDTA was the negative control. Numbers down the side of the 10% urea-denaturing gels denote the low range ssRNA ladder base sizes.

VapC<sub>PAE2754</sub> displays Mg<sup>2+</sup>-dependent sequence-specific ribonuclease activity against the single-stranded pentaprobe RNA substrates (Figure 5.3). Sequence specificity of

VapC<sub>PAE2754</sub> is shown by the different fragment patterns of the cleaved RNA on the gels in Figure 5.3 of the different pentaprobe RNA molecules.

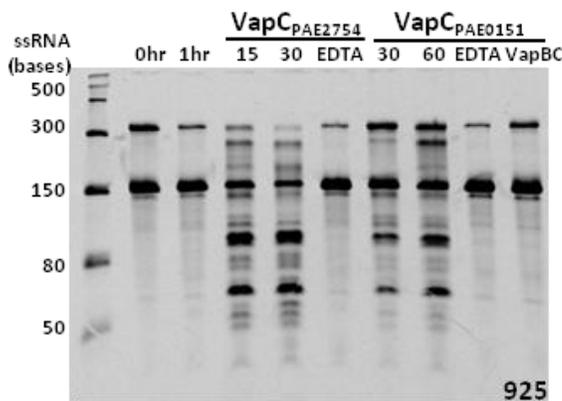
The difference in the size of RNA molecules 926 and 932 compared to the others is due to the different restriction enzymes used in cloning these fragments into pcDNA3, leading to a longer fragment when the insert is PCR amplified.



**Figure 5.4. Ribonuclease Activity of VapC<sub>PAE0151</sub> from *P. aerophilum* against pentaprobe RNA substrates.** Numbers at bottom right corner of gels denote pentaprobe insert, 932 is the reverse complement of 926. 0 hr and 1 hr refer to RNA negative controls, numbers the time at which the sample was taken in the assay (minutes), EDTA negative control, VapBC the PAE0151/2 complex. Numbers down the side of the 10% urea-denaturing gels denote the low range ssRNA ladder base sizes.

VapC<sub>PAE0151</sub> also displays Mg<sup>2+</sup>-dependent sequence-specific ribonuclease activity that against the single-stranded pentaprobe RNA substrates. Sequence specificity of VapC<sub>PAE0151</sub> is shown by the different fragment patterns of the cleaved RNA on the gels in Figure 5.4 of the different pentaprobe RNA molecules. Based on the gels in Figure 5.3 and Figure 5.4, the recognition sequences of VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> are possibly the same, as the fragment patterns of the cut RNA for each protein look similar. Under these assay conditions VapC<sub>PAE2754</sub> appears to be more active than VapC<sub>PAE0151</sub>, this is in agreement with the rotavirus ribonuclease activity assays (Section 5.2.1).

Although urea-denaturing polyacrylamide gels were used in these experiments, RNA molecules with a large amount of secondary structure (e.g. 925, 926, 927 and 932 (Table 5.1)) can run as multiple bands due to secondary structures remaining in the RNA.

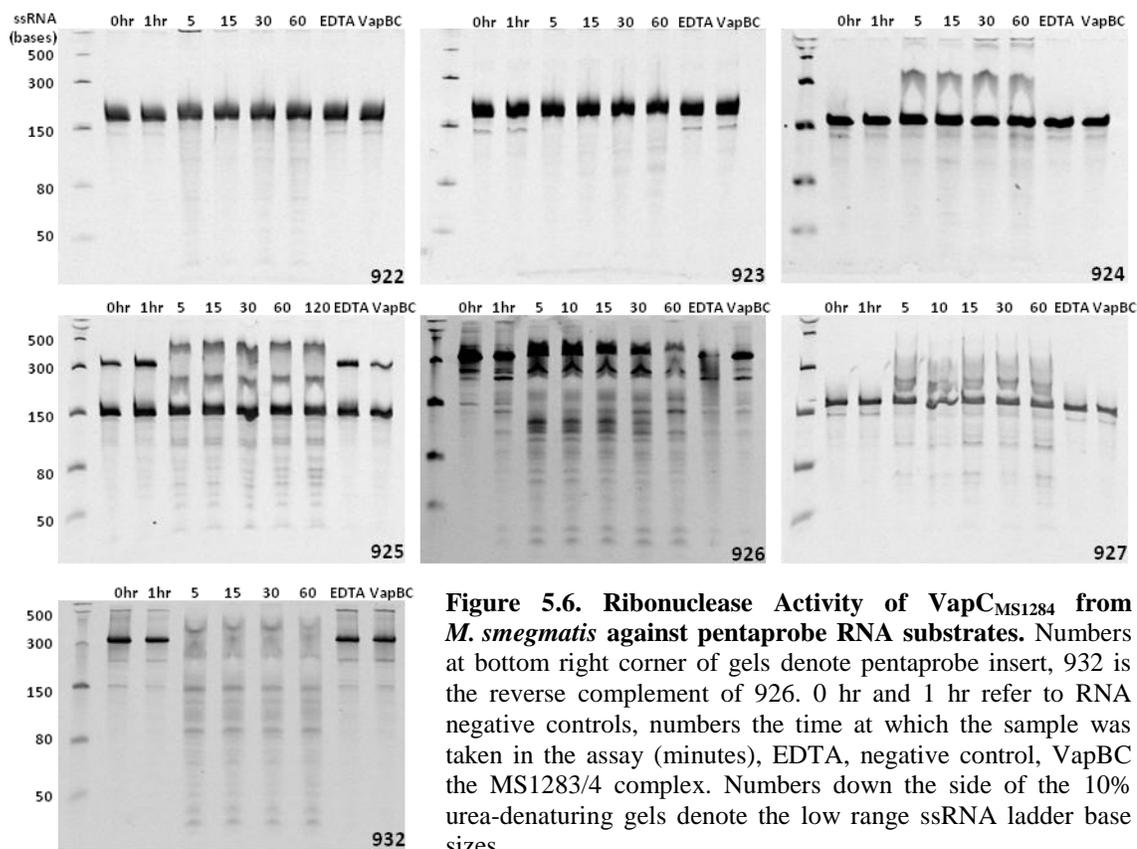


**Figure 5.5. VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> ribonuclease assay using pentaprobe 925 RNA as a substrate.** 0 hr and 1 hr refer to RNA negative controls, 15 and 30 are the assay time points (minutes), EDTA negative control and VapBC the PAE0151/2 complex. Low range ssRNA ladder shown down the side of the 10% urea-denaturing gel.

To test if the recognition sequence of VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> are the same, assays were conducted simultaneously with the two proteins. These two VapC proteins do in fact target the same sequence as the fragment pattern of the pentaprobe RNA molecule 925 in Figure 5.5 for the two proteins is identical. Of all the pentaprobe RNA molecules, 922 was cleaved the most efficiently by VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub>.

VapC<sub>MS1284</sub> from *M. smegmatis* also displays sequence-specific, Mg<sup>2+</sup>-dependent ribonuclease activity against a variety of pentaprobe RNA molecules (Figure 5.6). The sequence specificity of VapC<sub>MS1284</sub> differed from that of the VapC proteins from *P. aerophilum* which target the same sequence. VapC<sub>MS1284</sub> did not cleave pentaprobe RNAs 922, 923 or 924, but cleaved pentaprobe RNA 926 and its reverse complement 932 the most efficiently out of the six pentaprobe RNAs (Figure 5.6).

VapC<sub>MS1284</sub> also appears to bind the RNA molecules that it cleaves, as shown by the smear on the gels in the lanes where VapC<sub>MS1284</sub> is present (Figure 5.6). If the pentaprobe RNA is not cleaved by VapC<sub>MS1284</sub> the smear is not present on the gel (Figure 5.6). When EDTA is included the binding is lost and the same is seen when VapC is bound to VapB.



**Figure 5.6. Ribonuclease Activity of VapC<sub>MS1284</sub> from *M. smegmatis* against pentaprobe RNA substrates.** Numbers at bottom right corner of gels denote pentaprobe insert, 932 is the reverse complement of 926. 0 hr and 1 hr refer to RNA negative controls, numbers the time at which the sample was taken in the assay (minutes), EDTA, negative control, VapBC the MS1283/4 complex. Numbers down the side of the 10% urea-denaturing gels denote the low range ssRNA ladder base sizes.

Interestingly, the pentaprobe RNA molecules that VapC<sub>MS1284</sub> efficiently cleaves are the ones with the most predicted secondary structure as in Table 5.1 below. This suggests that VapC<sub>MS1284</sub> could be targeting both sequence and secondary structure.

Pentaprobe RNA	Free Energy kcal.mol <sup>-1</sup> (at 37 °C)
922	-25.2
923	-31.7
924	-29.9
925	-39.8
926	-96.5
927	-60
932	-88.8

**Table 5.1. Free energy values at 37 °C for each pentaprobe RNA molecule as calculated by RNA Structure Version 4.6.** This program uses 19 files of free energies for RNA-RNA interactions. The more negative the value the larger the amount of secondary structure in the RNA molecule.

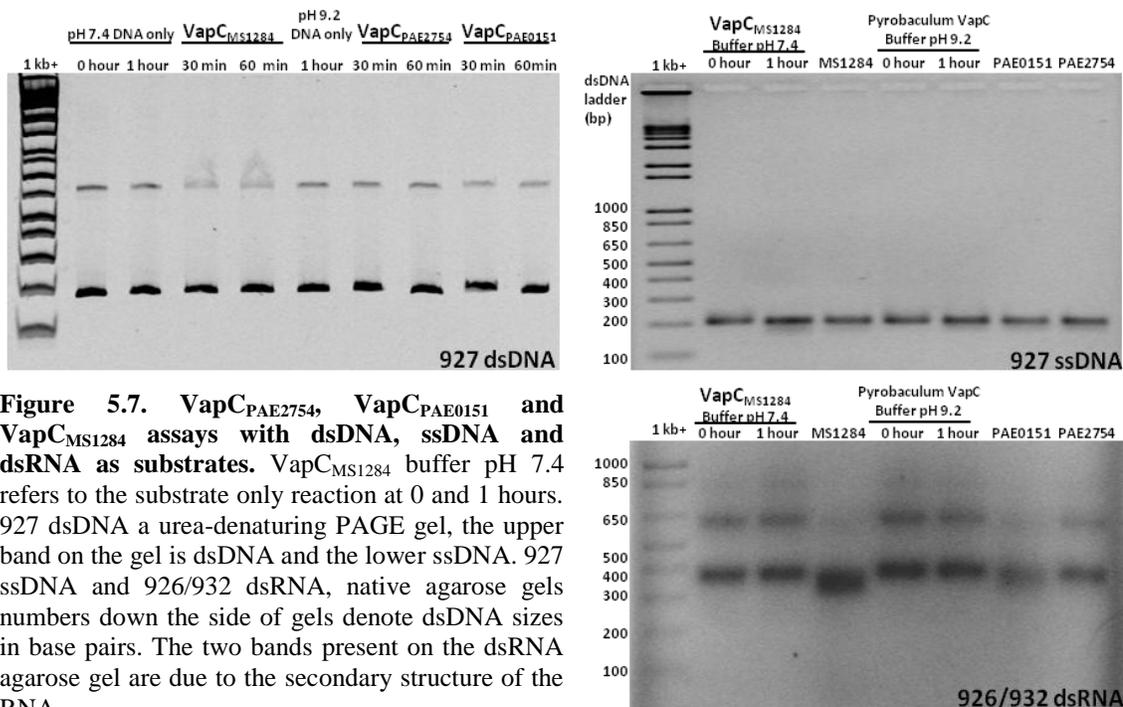
---

### 5.2.6 Other possible VapC Substrates (Double-Stranded DNA & RNA and Single-Stranded DNA)

Although VapC<sub>MS1284</sub>, VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> exhibit nuclease activity against ssRNA, it was important to test activity against other forms of nucleic acid to determine if VapC is an endoribonuclease that targets ssRNA or if it has specificity for dsRNA or even ssDNA or dsDNA.

ssDNA was synthesised by asymmetric PCR, in which one DNA strand is amplified preferentially over the other due to a difference in the ratio of forward and reverse primers (Section 2.4.8.5). The single-strand of DNA produced for this experiment was the forward strand of the 927 pentaprobe insert. dsDNA was synthesised by a traditional PCR reaction of the 927 pentaprobe insert (Section 2.4.8.2) and dsRNA was made by producing the forward (926) and reverse (932) RNA transcripts, then annealing the two transcripts together (Section 2.4.8.4).

The presence of double or single-stranded nucleic acids was tested by running the products on an agarose gel and staining with acridine orange which stains double-stranded nucleic acids fluorescent green and single-stranded nucleic acids a reddish-orange (Section 2.4.9). These substrates were gel purified prior to use in an activity assay with VapC<sub>MS1284</sub>, VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub>. Negative controls for these experiments included the substrate only in each of the reaction buffers (i.e. the VapC<sub>MS1284</sub> reaction buffer (pH 7.4) and the pyrobaculum VapC reaction buffer (pH 9.2) to ensure that the difference in pH of the assay buffers did not result in degradation of the nucleic acid at 37 °C over the assay periods.



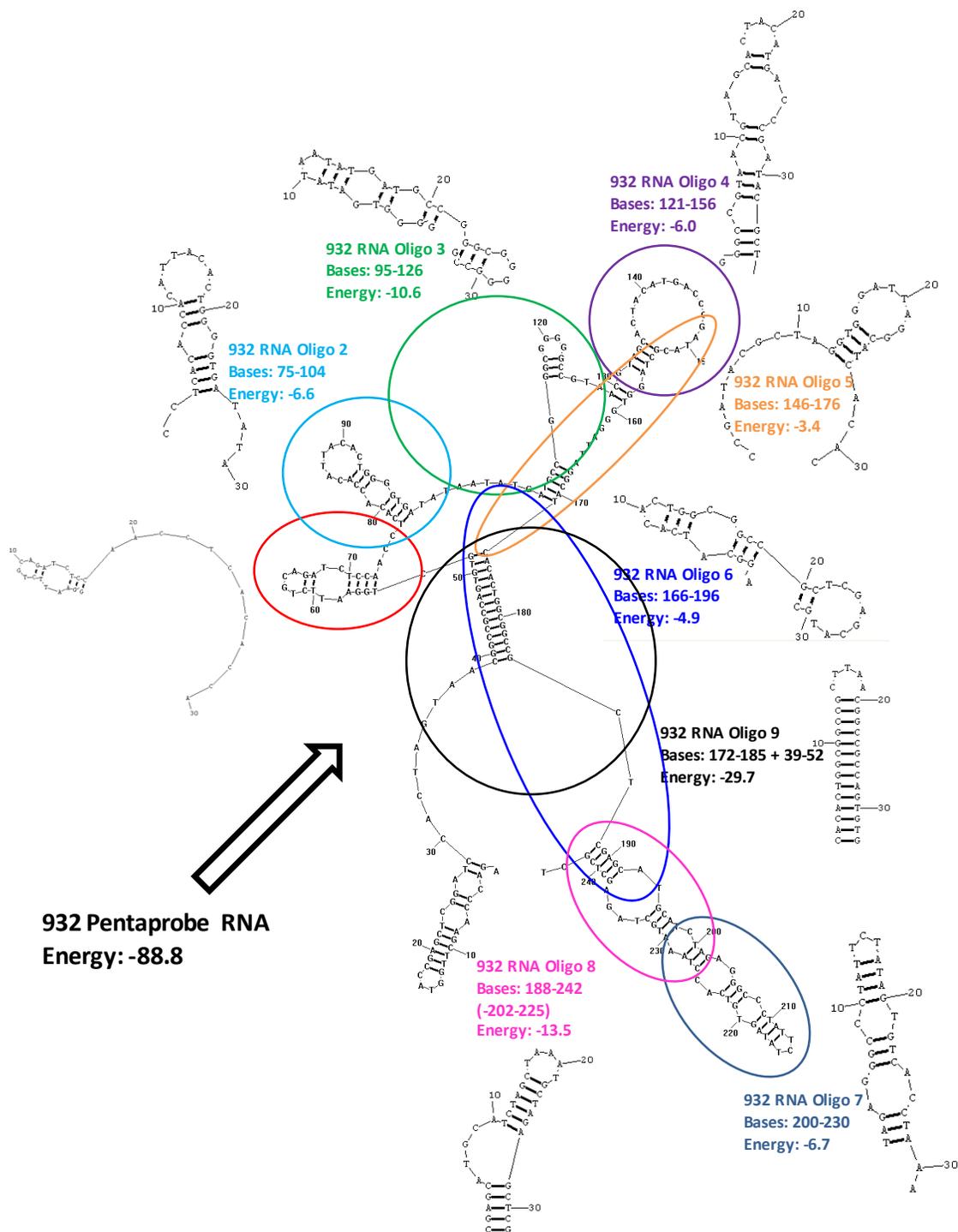
**Figure 5.7. VapC<sub>PAE2754</sub>, VapC<sub>PAE0151</sub> and VapC<sub>MS1284</sub> assays with dsDNA, ssDNA and dsRNA as substrates.** VapC<sub>MS1284</sub> buffer pH 7.4 refers to the substrate only reaction at 0 and 1 hours. 927 dsDNA a urea-denaturing PAGE gel, the upper band on the gel is dsDNA and the lower ssDNA. 927 ssDNA and 926/932 dsRNA, native agarose gels numbers down the side of gels denote dsDNA sizes in base pairs. The two bands present on the dsRNA agarose gel are due to the secondary structure of the RNA.

VapC<sub>PAE2754</sub>, VapC<sub>PAE0151</sub> and VapC<sub>MS1284</sub> do not degrade dsDNA, ssDNA or dsRNA. No degradation of any of these substrates was seen on the gels in Figure 5.7 above, although these proteins have been shown to cleave the ssRNA equivalent to the same sequence (Figure 5.3, Figure 5.4 and Figure 5.6). VapC<sub>MS1284</sub> could possibly be binding to secondary structure elements (Figure 5.6 and Figure 5.7) as there are smears present on the gel for the dsDNA bands. In addition, the upper band has disappeared upon addition of VapC<sub>MS1284</sub> to the dsRNA (Figure 5.7), again indicating binding to secondary structure elements. This binding is not seen for VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> from *P. aerophilum*.

## 5.2.7 Determination of the VapC Recognition Sequence

### 932 RNA Oligonucleotide Design

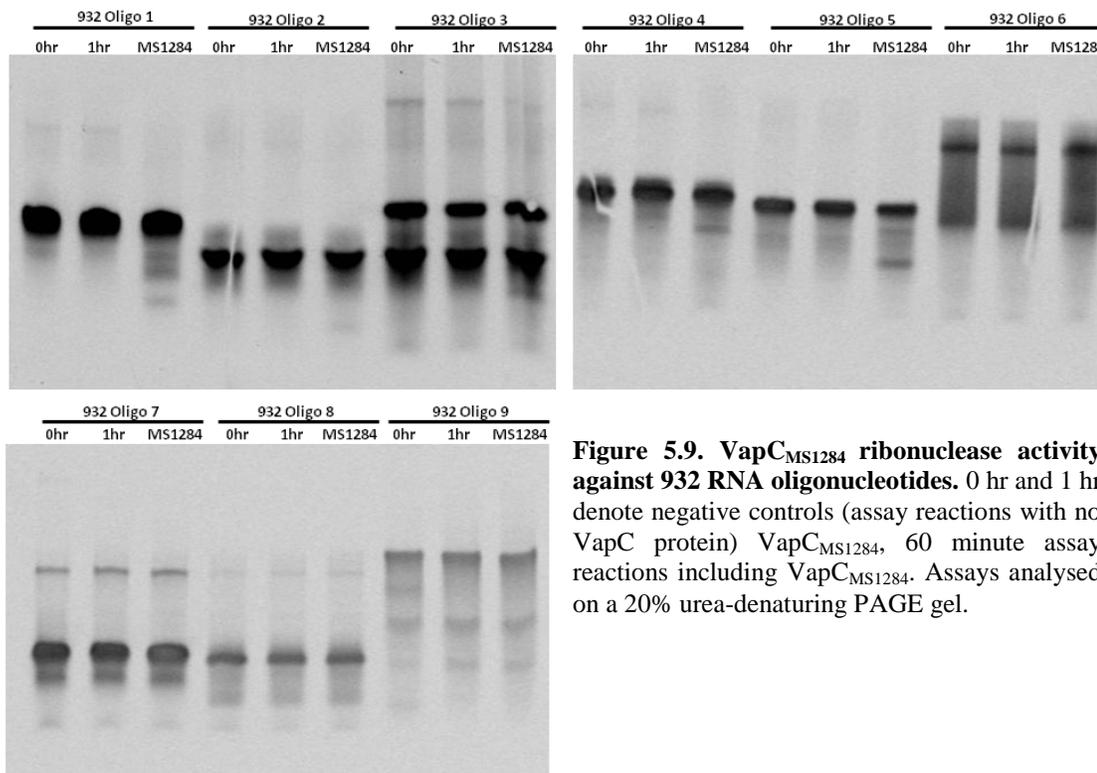
VapC<sub>MS1284</sub> cleaves the 932 pentaprobe RNA oligonucleotide most efficiently; VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> also cleave this RNA molecule. To determine the sequence specificity of these VapC proteins, overlapping RNA oligos were designed to have similar secondary structure to the original 932 RNA molecule as shown in Figure 5.8.



**Figure 5.8. Design of 932 RNA oligonucleotides 1-9.** Coloured circles represent area of 932 pentaprobe RNA covered by the corresponding oligonucleotide (coloured name). Energy:  $\text{kCal}\cdot\text{mol}^{-1}$  for the secondary structure, Bases: the bases of the 932 pentaprobe RNA covered by each oligonucleotide. Secondary structure determined by RNA Structure Ver. 4.6.

**932 RNA Oligonucleotides and VapC Ribonuclease Activity Assays**

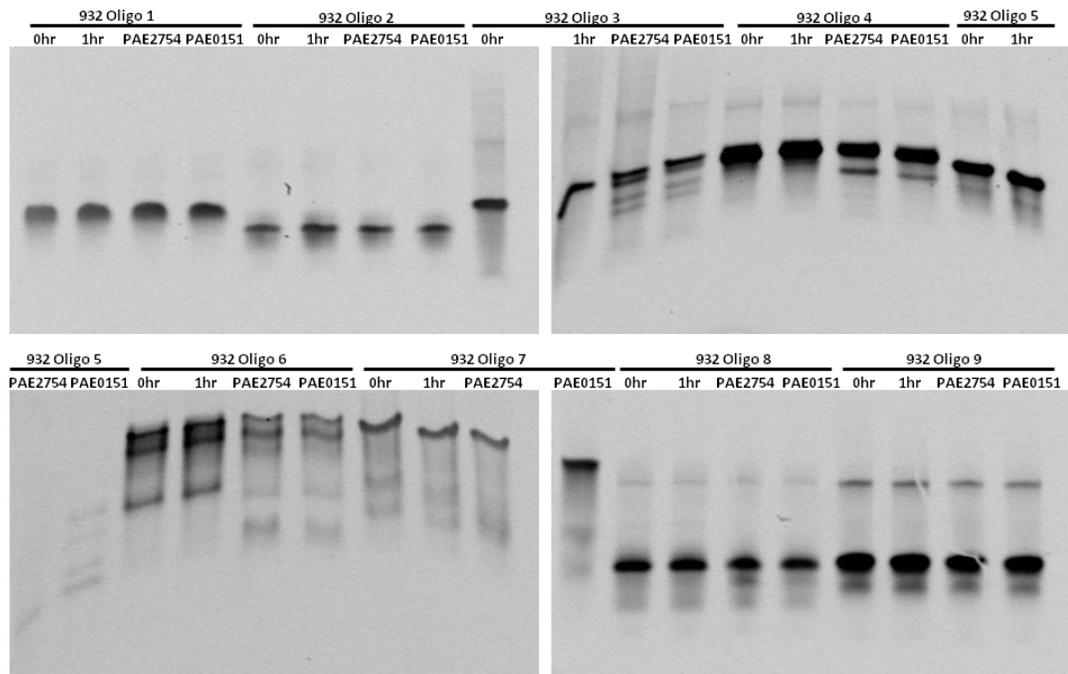
The shorter RNA oligonucleotides were used as substrates in ribonuclease activity assays. VapC<sub>MS1284</sub>, VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> were tested for those RNA oligos that were cleaved and MALDI MS used to identify the recognition site in the oligo.



**Figure 5.9. VapC<sub>MS1284</sub> ribonuclease activity against 932 RNA oligonucleotides.** 0 hr and 1 hr denote negative controls (assay reactions with no VapC protein) VapC<sub>MS1284</sub>, 60 minute assay reactions including VapC<sub>MS1284</sub>. Assays analysed on a 20% urea-denaturing PAGE gel.

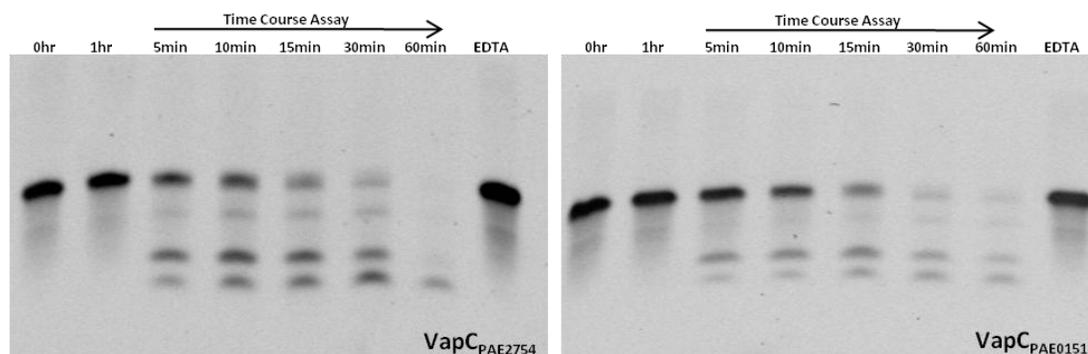
VapC<sub>MS1284</sub> cleaves 932 RNA oligos 1, 4 and 5, a very small amount of cleavage was observed for 932 RNA oligo 3 and no cleavage observed for 932 RNA oligos 2, 6, 7, 8 and 9 (Figure 5.9).

Compared to the degradation of the 932 pentaprobe RNA molecule (Figure 5.6, 932), degradation of the oligonucleotides was very slow (Figure 5.9), after 60 minutes with the enzyme very little of the RNA oligos were degraded.



**Figure 5.10. VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> ribonuclease activity against 932 RNA oligonucleotides.** 0 hr and 1 hr denote negative controls (assay reactions with no VapC protein) and VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> denote 60 minute assay reactions including protein. Assays analysed on a 20% urea-denaturing PAGE gel.

Earlier experiments demonstrated that VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> target the same recognition sequence and this was also seen in the above ribonuclease assays (Figure 5.10). VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> cleave 932 RNA oligos 3, 4, 5 and 6, cleaving 932 RNA oligo 5 most efficiently, as the whole oligo was cleaved in an hour (Figure 5.10). No degradation of 932 RNA oligos 1, 2, 7, 8 or 9 was observed. The 932 RNA oligos cleaved by VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> differs to those cleaved by VapC<sub>MS1284</sub>. This supports VapC from *M. smegmatis* having a different recognition sequence compared to the two VapC proteins from *P. aerophilum*. A time course assay was conducted with VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> using 932 RNA oligo 5 as a substrate to determine the general rate of the enzyme for use in subsequent MALDI MS experiments.



**Figure 5.11.** VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> ribonuclease activity assays using 932 RNA oligo 5 as a substrate. 0 hr and 1 hr refer to negative controls (assay reaction with no VapC protein), time course assay denotes the time course of the experiment in minutes with VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> and EDTA as the negative control. Assays analysed on a 20% urea-denaturing PAGE gel.

VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> cleave 932 RNA oligo 5 quickly. After five minutes with each of these proteins a large proportion of the RNA oligo is degraded and after 60 minutes all of the oligo is degraded and the higher molecular weight bands on the gel have also been degraded (Figure 5.11). This suggests that there is an optimal recognition sequence for VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> present in this RNA oligo and another sub-optimal sequence that VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> also recognises for cleavage.

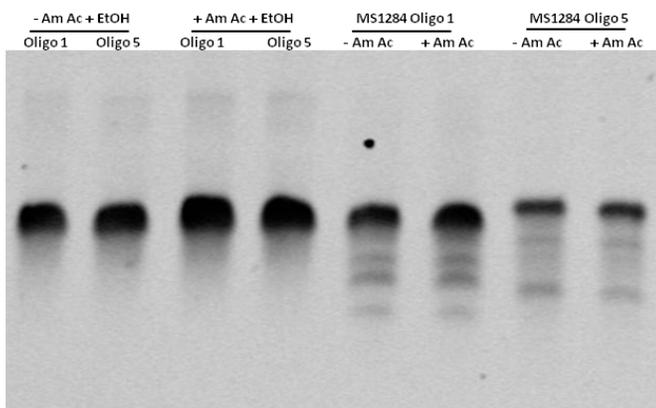
### ***MALDI-TOF MS RNA Oligonucleotide Sample Preparation***

Mass spectrometry provides a fast, sensitive and direct way of analysing ribonuclease activity assays. MALDI-TOF MS is one of the simplest and quickest mass spectrometry tools for the analysis of complex mixtures as it typically generates a singly charged oligonucleotide ion. However, sample preparation is crucial for MALDI-TOF MS as commonly used chemicals such as EDTA can interfere with ionisation of the sample and RNA tends to form adducts with sodium and potassium ions which also creates problems. It is therefore very important for the sample to be desalted before MALDI-TOF MS analysis.

A variety of desalting techniques were trialled. VapC<sub>MS1284</sub> assays were conducted in a sodium phosphate buffer and both VapC<sub>MS1284</sub> and VapC<sub>PAE2754</sub> /VapC<sub>PAE0151</sub> assays require NaCl and MgCl<sub>2</sub> for activity. Assay trials using low concentrations of MgCl<sub>2</sub> in water were unsuccessful. The use of cation exchange beads loaded with ammonium ions

to exchange the sodium ions that form adducts on the RNA was moderately successful, although the spectra collected for these samples still showed the presence of sodium adducts on the RNA. C<sub>18</sub> reverse phase pipette tips, also known as ZipTips resulted in a much cleaner RNA oligonucleotide, although the majority of the sample was lost in the process and therefore the signal upon MALDI-TOF MS was very low.

The use of lithium chloride and sodium acetate to precipitate the RNA molecules was employed, although it wasn't known if these methods would precipitate very small pieces of RNA. Lithium chloride and ethanol precipitation were relatively successful at precipitating small pieces of RNA, however some of the sample was lost in the process. Sodium acetate and ethanol precipitation of RNA oligonucleotides (from ribonuclease assays) proved successful with the majority of RNA fragments recovered, however the use of sodium acetate does not remove the problem of sodium adducts forming with the



**Figure 5.12. Ammonium acetate and ethanol precipitation of MS1284 ribonuclease activity assays using 932 RNA oligos 1 and 5 as substrates.** –Am Ac and –Am Ac + EtOH refers to untreated samples. + Am Ac and +Am Ac + EtOH refers to ammonium acetate and ethanol precipitation of samples. Oligo 1 and Oligo 5 in lanes on the left refer to negative controls and MS1284 Oligo 1 and MS1284 Oligo 5 refers to the 60 minute ribonuclease activity assay reactions (before (-Am Ac) and after (+Am Ac) precipitation).

RNA. Therefore, ammonium acetate and ethanol precipitation was trialled. The majority of RNA fragments were recovered using this technique as shown in Figure 5.12 and samples were more amenable to MALDI-TOF MS analysis. This method was successful for VapC<sub>P<sub>AE</sub>2754</sub> and VapC<sub>P<sub>AE</sub>0151</sub> assays as these are conducted in a Tris-HCl buffer. However as the VapC<sub>MS1284</sub> assays are conducted in a sodium phosphate buffer there were still

problems with sodium adducts. Therefore, an assay buffer using 20 mM ammonium phosphate at pH 7.4 instead of 20 mM sodium phosphate buffer at pH 7.4 was trialled. This proved successful; activity of VapC<sub>MS1284</sub> in the ammonium phosphate buffer was the same as that in the sodium phosphate buffer and subsequent ammonium acetate and

---

ethanol precipitation of the assay reactions resulted in cleaner spectra collected by MALDI-TOF MS.

For subsequent MALDI-TOF MS analysis VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> assays were carried out in 20 mM Tris-HCl pH 9.2, 20 mM NaCl, 10 mM MgCl<sub>2</sub> and VapC<sub>MS1284</sub> assays in 20 mM ammonium phosphate buffer pH 7.4, 20mM NaCl, 10 mM MgCl<sub>2</sub>. Assay reactions were desalted by precipitation with ammonium acetate and ethanol according to the method in Section 2.4.11.1.

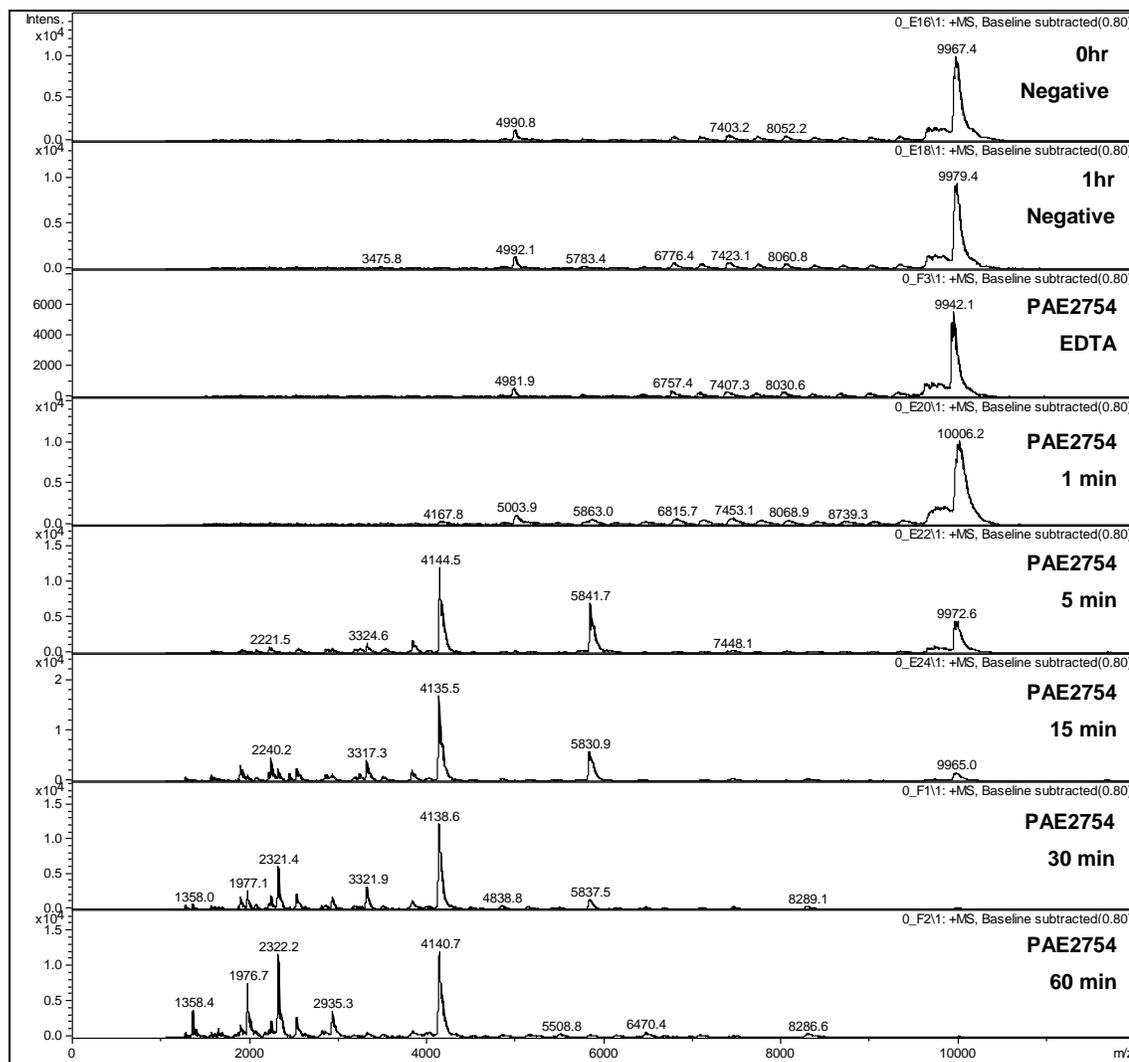
***VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> MALDI-TOF MS Analysis of 932 RNA Oligonucleotide Ribonuclease Assays***

*Raw data can be found on compact disc –Appendix C, Chapter Five.*

VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> ribonuclease activity assays of HPLC-purified 932 RNA oligos 3 and 5 were carried out as in Section 2.4.10.2 and desalted by ammonium acetate and ethanol precipitation as in Section 2.4.11.1, and prepared for MALDI-TOF analysis as in Section 2.4.11.2. Multiple spectra for each sample were collected as in Section 2.4.11.3.

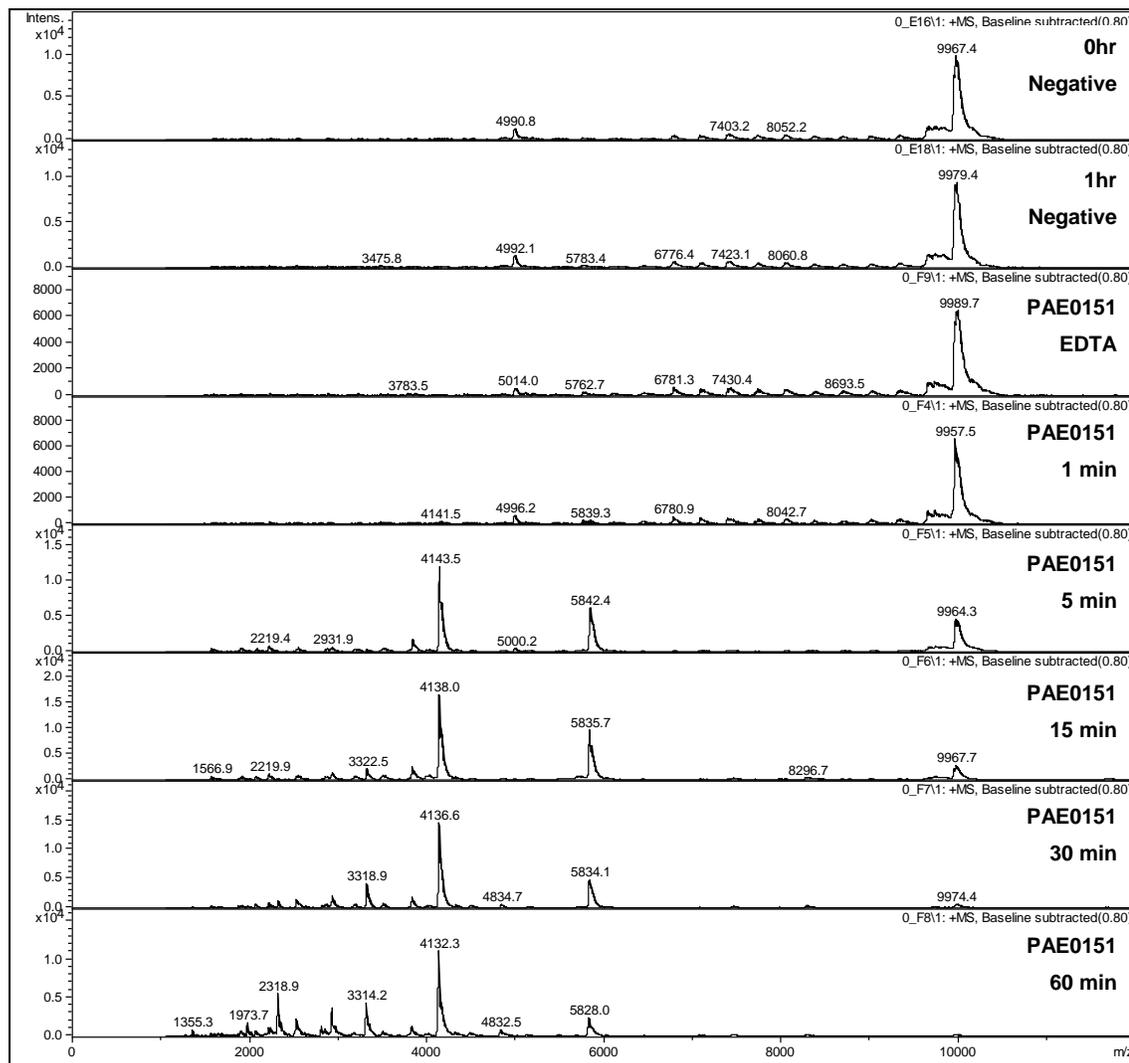
To analyse the spectra, the negative controls were overlaid with the assay reactions, and peaks present in the assay reaction but not in the negative controls were divided by the average molecular weight of an RNA nucleotide. The size (in bases) of the fragment could then be used in conjunction with the Mongo Oligonucleotide Mass Calculator (Rozenski, 1999) to determine which end of the RNA oligo the fragment was from, and if a phosphate was present on the 5' or 3' cleavage product. 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).

Earlier VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> ribonuclease activity assays revealed that these proteins cleave 932 RNA oligo 5 very efficiently. Therefore, a time course similar to that in Figure 5.11 was decided upon for MALDI-MS analysis of these assay reactions.



**Figure 5.13. 932 RNA Oligo 5 VapC<sub>PAE2754</sub> MALDI-MS results.** 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 (1 to 60 minutes) assay samples with VapC<sub>PAE2754</sub>. Data analysed by Bruker Data Analysis and baseline subtracted.

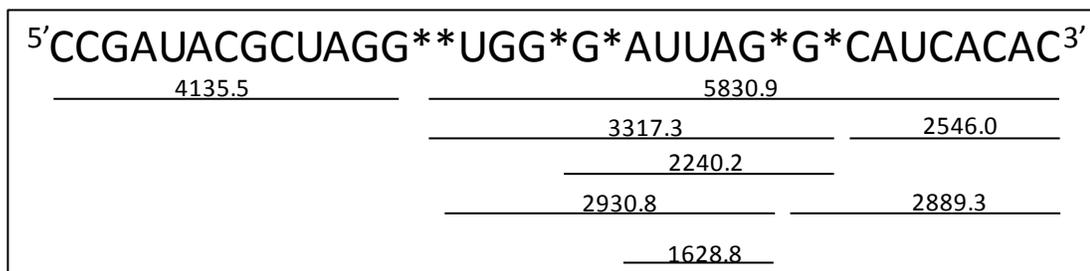
The mass spectra in Figure 5.13 and Figure 5.14 show no degradation of the RNA oligo in the negative controls. The assay reactions with RNA only (0 hr and 1 hr) showed no degradation compared with the assay reactions and addition of 20 mM EDTA to the assay reaction abolished activity of VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub>. The spectra (Figure 5.13 and Figure 5.14) obtained for both the VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> assay reaction were identical. Therefore, the two pyrobaculum VapC proteins target the same sequence.



**Figure 5.14. 932 RNA Oligo 5, VapC<sub>PAE0151</sub> MALDI-MS results.** EDTA, inclusion of 20 mM EDTA in the reaction buffer and time course (1 to 60 minutes) assay samples with VapC<sub>PAE0151</sub>. Data analysed by Bruker Data Analysis and baseline subtracted.

After one minute with either VapC<sub>PAE2754</sub> or VapC<sub>PAE0151</sub>, there are two very small peaks in Figure 5.13 and Figure 5.14 corresponding to the cut site GG\*UG, forming a 13 base 5' cleavage product ( $m/z$  4135.5) and an 18 base 3' cleavage product ( $m/z$  5830.9) including a 5' phosphate. After five minutes these peaks are very prominent, indicating that this is the preferred recognition sequence of the pyrobaculum VapC proteins. At the five minute time point a majority of the RNA oligo is degraded (Figure 5.13 and Figure 5.14). The 3' 18 base cleavage product ( $m/z$  5830.9) was degraded further by

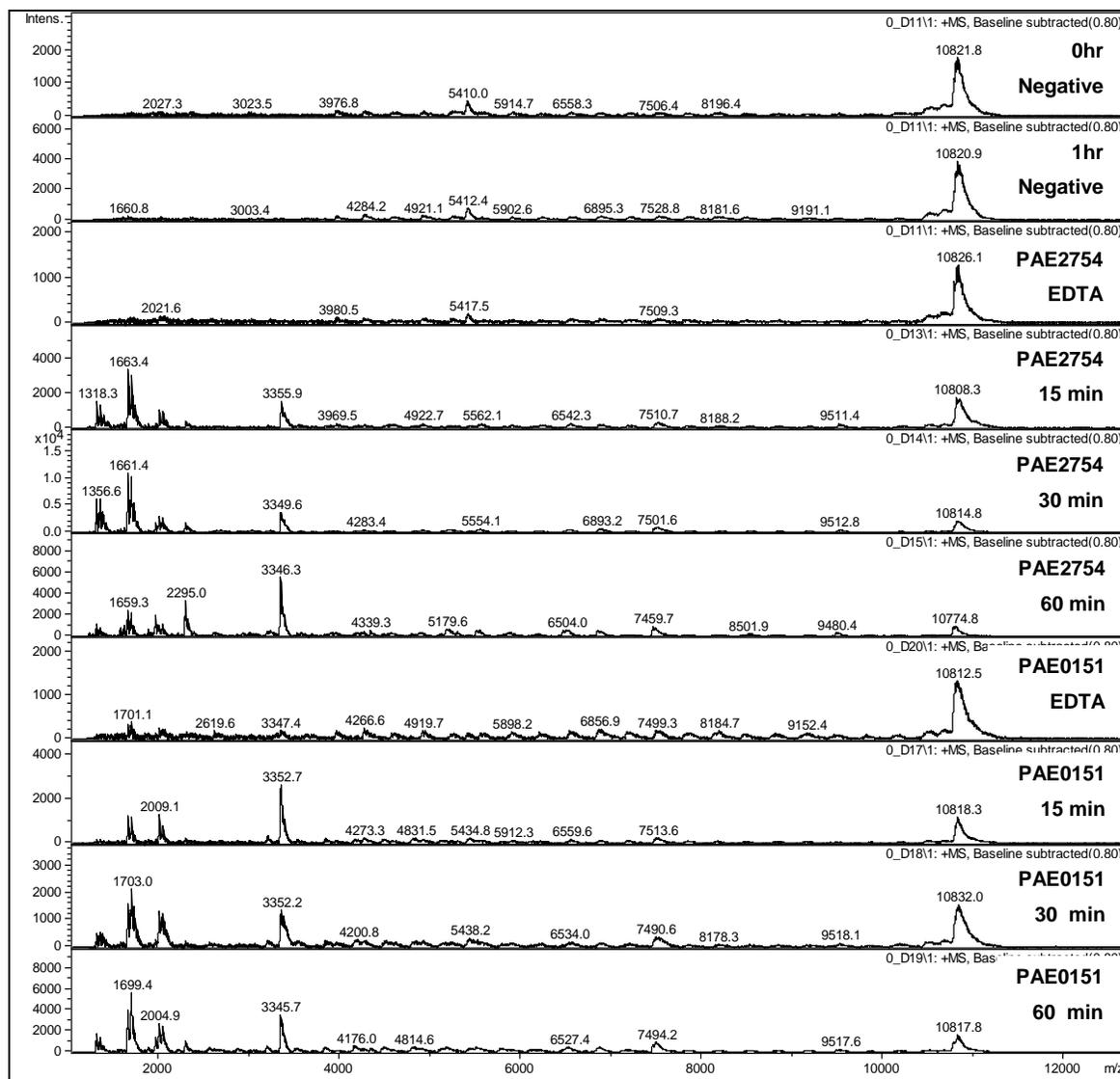
VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub>. At 15, 30 and 60 minute time points additional cut sites to those outlined above were observed at GG\*GA, GC\*AU, GG\*UG and AG\*GC again yielding a 5' phosphate on the 3' cleavage products (Figure 5.15).



**Figure 5.15. 932 RNA Oligo 5 VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> cleavage sites and corresponding masses.** Cleavage sites are indicated by \*. Lines under sequence correspond to RNA fragments and their masses (Da) determined by MADLI MS (numbers). Masses correspond to VapC<sub>PAE2754</sub> 15 minute assay reaction, similar masses were seen for other time points and VapC<sub>PAE0151</sub>.

932 RNA oligo 5 is the optimal substrate tested for VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub>. Analysis of the consensus cleavage site of VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> by taking three bases either side of the cut site revealed the optimal cut site to be GG\*UG, and a slightly less optimal cleavage site at GG\*GG. There are secondary cut sites (Figure 5.15) that VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> target suggesting there is “wobble” in the target sequence.

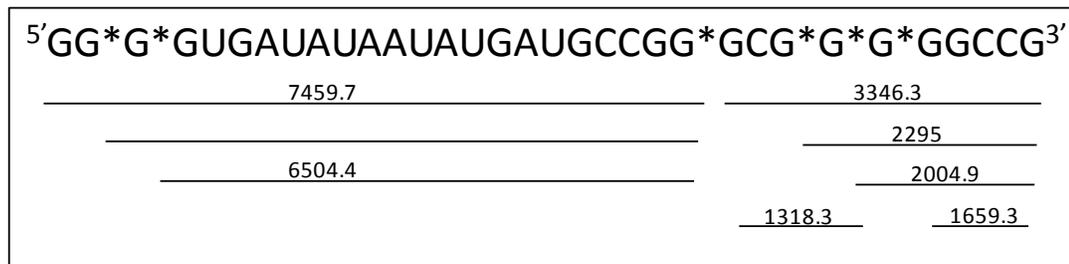
932 RNA oligo 3 was also tested as a substrate in MADLI MS experiments. The mass spectra in Figure 5.16 show no degradation of 932 RNA oligo 3 in the negative controls as also seen for 932 RNA oligo 5.



**Figure 5.16.** 932 RNA Oligo 3, VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> MALDI-MS results. 0hr and 1hr 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 VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub>. Data analysed by Bruker Data Analysis and baseline subtracted.

Analysis of the MALDI MS spectra revealed a cleavage site at GG\*GC (\* indicates cut site) after 15 minutes, leaving a 5' phosphate on the 3' cleavage product. The 3' cleavage product was then cleaved further at two places with GG\*GG sequences. These fragments ( $m/z$  7459.7 and 3346.3) were still present in the 30 and 60 minute assay samples along with further cut sites at GG\*GG and GG\*GU in the 30 minute sample and an additional cut site CG\*GG in the 60 minute sample (Figure 5.17). All 3'

cleavage products had a 5' phosphate and no phosphate was present on the 3' end of the 5' cleavage product.



**Figure 5.17. 932 RNA Oligo 3 VapC<sub>P<sub>AE</sub>2754</sub> and VapC<sub>P<sub>AE</sub>0151</sub> cleavage sites and corresponding masses.** Cleavage sites are indicated by \*. Lines under sequence correspond to RNA fragments and their masses (Da) determined by MADLI MS (numbers). Masses correspond to VapC<sub>P<sub>AE</sub>2754</sub> 60 minute assay reaction, similar masses were seen for other time points and VapC<sub>P<sub>AE</sub>0151</sub>.

The primary cut site for RNA oligo 3 is GG\*GG and GG\*GC with GG\*GU and CG\*GG being secondary cut sites (visible in later time course reactions).

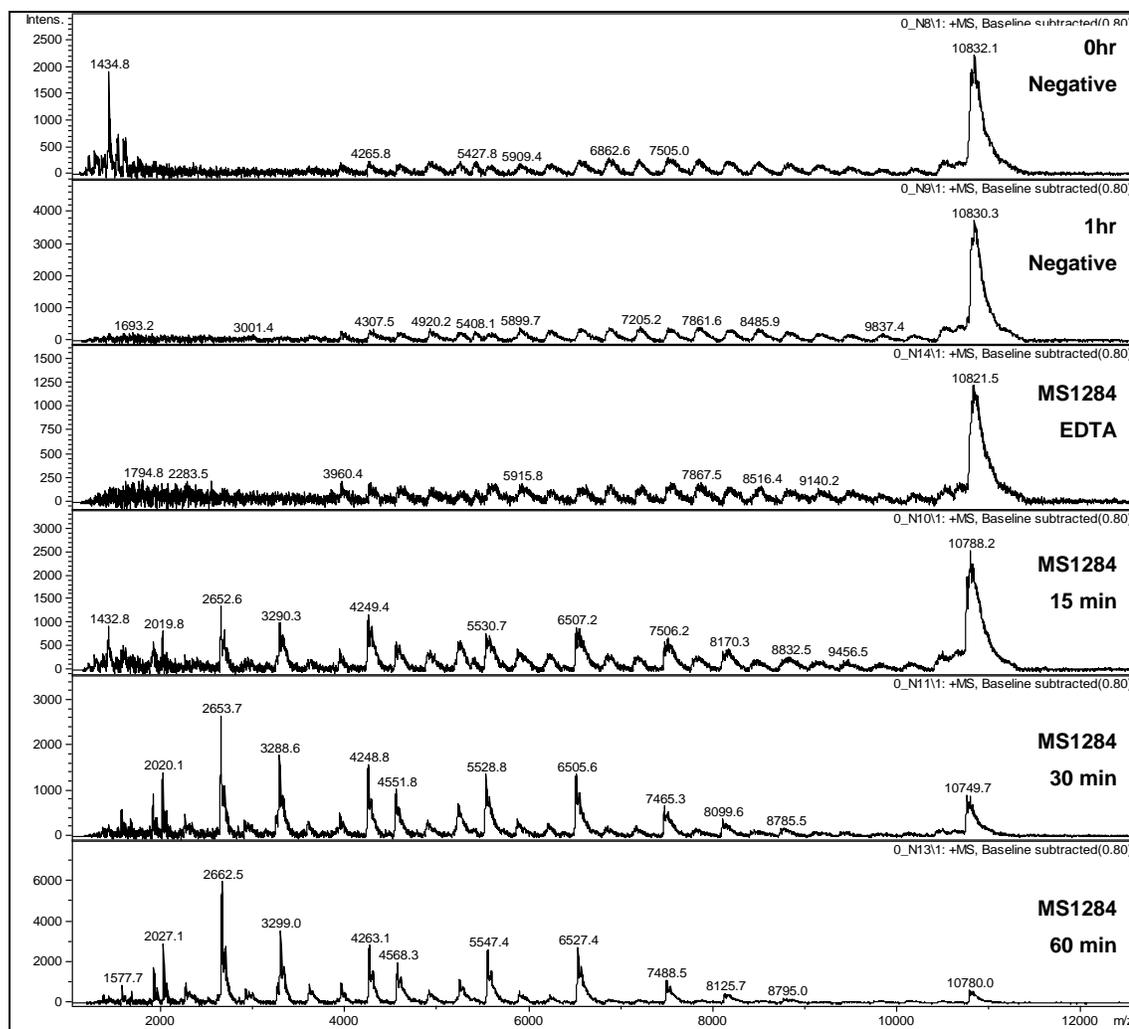
The results from both RNA oligonucleotides assayed show that VapC<sub>P<sub>AE</sub>2754</sub> and VapC<sub>P<sub>AE</sub>0151</sub> target G rich sequences. The sequence GG\*UG is the optimal sequence as cleavage at this site is demonstrated after one minute whereas all over cleavage sites were observed at later time points. GG\*GG is another strong primary cleavage site as it is present three times in RNA oligo 3 and is cut efficiently. A G residue is always positioned directly before the cleavage site, whereas there is some 'wobble' for the bases present at other positions in the target sequence.

#### ***VapC<sub>MS1284</sub> MALDI-TOF MS Analysis of 932 RNA Oligonucleotide Ribonuclease Assays***

*Raw data can be found on compact disc –Appendix C, Chapter Five.*

VapC<sub>MS1284</sub> ribonuclease activity assays of 932 RNA oligos 3 and 4 were carried out in ammonium phosphate buffer as in Section 2.4.10.1 and desalted by ammonium acetate and ethanol precipitation as in Section 2.4.11.1 and prepared for MALDI-TOF analysis as in Section 2.4.11.2. Multiple spectra for each sample were collected as in Section 2.4.11.3. Mass spectra were analysed as for the pyrobaculum VapC proteins (Section 2.4.11.4). This was more difficult for VapC<sub>MS1284</sub> compared to the pyrobaculum VapC

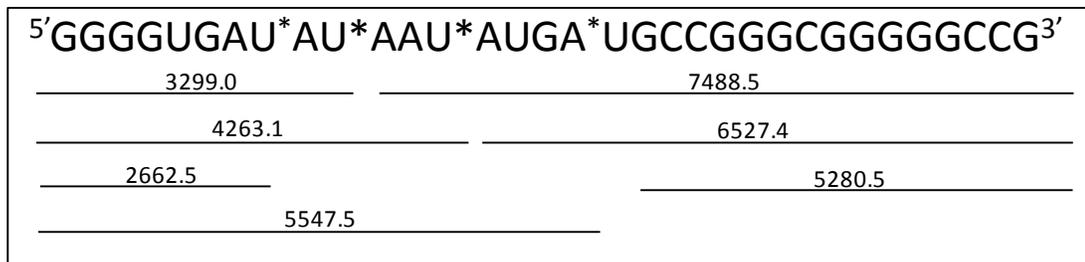
proteins, the spectra obtained had more sodium adducts. Due to the increased background of the spectra it was difficult to determine if a phosphate was present on the 5' of the 3' cleavage product or the 3' of the 5' cleavage product. It was assumed that VapC<sub>MS1284</sub> would leave a 5' phosphate on the 3' cleavage product, similar to the homologous pyrobaculum proteins. The fact that the 932 RNA oligos are not the optimal substrate for VapC<sub>MS1284</sub> also made analysis of the VapC<sub>MS1284</sub> MALDI MS results difficult. Fragments resulting from cleavage of the oligonucleotide were hard to distinguish from the background. 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).



**Figure 5.18.** 932 RNA Oligo 3, VapC<sub>MS1284</sub> MALDI-MS results. 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 minutes) assay samples with VapC<sub>MS1284</sub>. Data analysed by Bruker Data Analysis and baseline subtracted.

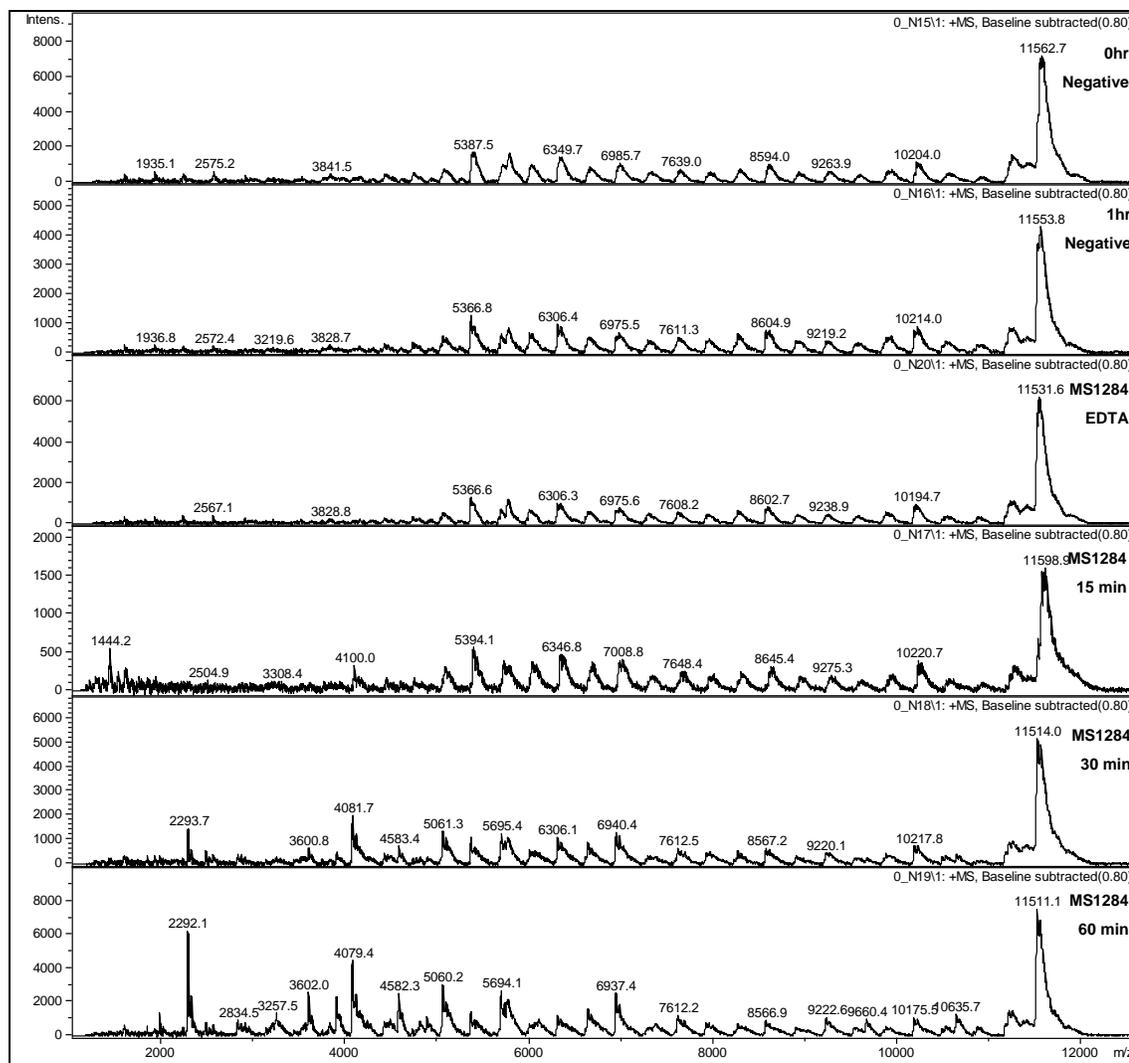
The MALDI MS spectra (Figure 5.18) shows no degradation of 932 RNA oligo 3 in the negative controls, the assay reactions with RNA only (0 hr and 1 hr) showed no degradation compared to the assay reactions and addition of 20 mM EDTA to the assay reaction abolished activity of MS1284.

Cleavage products of 932 RNA oligo 3 observed in the 15 minute assay sample were also present at 30 and 60 minutes. No additional cleavage sites were observed in the 30 and 60 minute assay samples. VapC<sub>MS1284</sub> cleaves at AU\*AA (\* denotes cleavage site) to give a 10 base 5' cleavage fragment ( $m/z$  3299.0) and a 23 base 3' cleavage product ( $m/z$  7488.5) including a 5' phosphate. It also cleaves at AU\*AU which is present twice in the 932 RNA oligo 3 sequence to yield a 13 base 5' cleavage fragment ( $m/z$  4263.1) and a 20 base 3' cleavage fragment ( $m/z$  6527.4) with a 5' phosphate and an 8 base 5' cleavage fragment ( $m/z$  2662.5) and a 16 base 3' cleavage fragment ( $m/z$  5280.5) (Figure 5.19). Cleavage at a GA\*UG sequence was also observed (Figure 5.19).



**Figure 5.19. 932 RNA Oligo 3 VapC<sub>MS1284</sub> cleavage sites and corresponding masses.** Cleavage sites are indicated by \*. Lines under sequence correspond to RNA fragments and their masses (Da) determined by MADLI MS (numbers). Masses correspond to VapC<sub>MS1284</sub> 60 minute assay reaction, similar masses were seen for other time points.

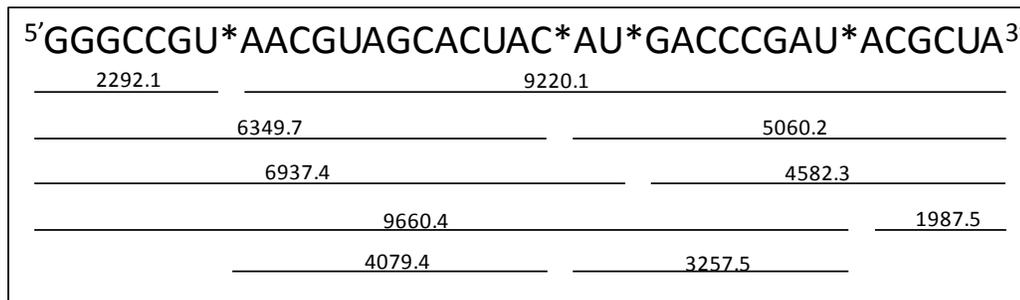
No ribonuclease contamination was present in assay preparations for 932 RNA oligo 4. The MALDI MS spectra in Figure 5.20, demonstrate that the RNA only assay (0 hr and 1 hr) and EDTA reactions exhibit no ribonuclease activity.



**Figure 5.20. 932 RNA Oligo 4 VapC<sub>MS1284</sub> MALDI-MS results.** 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 MS1284. Data analysed by Bruker Data Analysis and baseline subtracted.

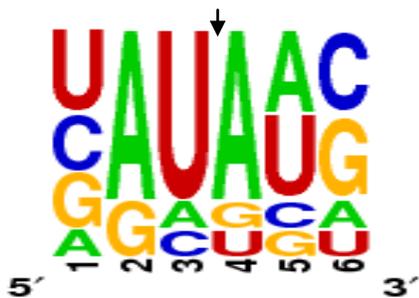
The 15 minute time point showed cleavage of 932 RNA oligo 4 at GU\*AA and AC\*AU sequences to yield fragment sizes of 29 ( $m/z$  9220.1), 20 ( $m/z$  6349.7) and 16 ( $m/z$  5060.2) bases with a 5' phosphate on the 29 and 13 base 3' fragments. After 30 and 60 minutes additional cleavage sites to those above were seen, as shown by the presence of more peaks in the spectra (Figure 5.20). AC\*AU, AU\*GA, AU\*AC recognition sites were observed leading to fragment sizes of 7 ( $m/z$  2292.1), 29 ( $m/z$  9220.1), 20 ( $m/z$  6349.7), 16 ( $m/z$  5060.2), 22 ( $m/z$  6937.4), 14 ( $m/z$  4582.3), 30 ( $m/z$  9660.4), 6 ( $m/z$

1987.5), 12 ( $m/z$  4079.4) and 10 ( $m/z$  3257.5) bases. The cut sites are outlined in Figure 5.21.



**Figure 5.21. 932 RNA Oligo 4 VapC<sub>MSI284</sub> cleavage sites and corresponding masses.** Cleavage sites are indicated by \*. Lines under sequence correspond to RNA fragments and their masses (Da) determined by MADLI MS (numbers). Masses correspond to VapC<sub>MSI284</sub> 60 minute assay reaction, similar masses were seen for other time points.

The target sequence of VapC<sub>MSI284</sub> is very different to that of its pyrobaculum homologues. VapC<sub>MSI284</sub> is targeting AT sequences. There is “wobble” in the target sequence, as many different target sequences were identified. The consensus sequence of VapC<sub>MSI284</sub> was determined to be AUAA, AUAU by analysing the residues present around the cleavage site as in Figure 5.22.



**Figure 5.22. Analysis of VapC<sub>MSI284</sub> cleavage sites.** VapC<sub>MSI284</sub> consensus cleavage sequence as calculated by WebLogo (Crooks, Hon, Chandonia, & Brenner, 2004; Schneider & Stephens, 1990). VapC<sub>MSI284</sub> cleaves after position three (as indicated by arrow). Three bases before and after the cut site were analysed. The height of the letter is proportional to the frequency of that base at that particular position. Positions labelled 1-6.

### 5.2.8 Bioinformatic Analysis of the VapC<sub>MSI284</sub> Target Sequence

Transcriptional analysis by microarray of genes affected by VapC expression revealed 106 genes were downregulated  $\geq 1.5$  fold in the  $\Delta vapBC$  *M. smegmatis* strain expressing VapC for one hour compared with the same strain expressing VapBC (Robson, 2010). Genes downregulated by VapC, and the sequences 150 bp upstream of these genes, were analysed for the prevalence of VapC target sequences. A 150 bp upstream region was included in the analysis as cleavage of the 5' untranslated region (UTR) of the mRNA

could result in degradation of the mRNA; once the stable stem loop structure formed by the 5' UTR is removed it is more susceptible to nucleolytic degradation. ATAA is the 10<sup>th</sup> rarest quartet of bases (out of 256 possible quartets) in the *M. smegmatis* genome and ATAT the 23<sup>rd</sup> rarest. Analysis of the 106 genes and their upstream regions downregulated by VapC from microarray data revealed that 39 genes downregulated by VapC contain a VapC recognition sequence (ATAT or ATAA) and the recognition sequence was present 63 times in these 39 genes and their upstream regions (Table 5.2). This is greater than the frequency of the VapC recognition sequence in the rest of the genome and those genes that are not downregulated by VapC expression. Of the genes downregulated by VapC expression in the microarray data 37% contain the VapC recognition sequence, compared to 25% for the whole genome and 24% for those not downregulated by VapC expression (Table 5.2).

	Total number of genes	Number of genes & upstream sequences with ATAT/ATAA	Number of times ATAT & ATAA present in genes & upstream sequences	Percentage of genes (+ upstream sequences) with ATAT/ATAA present
<b>Total in Genome</b>	6717	1655	3374	25
<b>No change in Microarray</b>	6512	1616	3311	25
<b>Downregulated in Microarray</b>	106	39	63	37
<b>Upregulated in Microarray</b>	99	23	31	23

**Table 5.2 Prevalence of VapC target sequences in genes and their upstream regions in the *M. smegmatis* genome. Total number of genes in the genome compared to genes and upstream regions with no change in the microarray compared to genes + upstream region both downregulated and upregulated in the microarray. Total number of genes refers to the number of genes in each subset. Genes and upstream sequences, the number of genes the VapC target sequence is present in, including 150 bp upstream of the gene. Times ATAT and ATAA present refers to the total number of times VapC target sequence is present in the genes and 150 bp upstream of a gene. Percentage of genes with ATAT/ATAA refers to the percentage of genes and upstream of the gene containing the VapC target sequence in each subset (Total number of genes/number of genes & upstream sequences with ATAT/ATTA).**

	Total number of genes	Number of genes with ATAT/ATAA	Number of times ATAT & ATAA present in genes	Percentage of genes with ATAT/ATAA present
<b>Total in Genome</b>	6717	1655	3374	25
<b>No change in Microarray</b>	6512	1459	1918	22
<b>Downregulated in Microarray</b>	106	35	43	33
<b>Upregulated in Microarray</b>	99	13	14	13

**Table 5.3. Prevalence of VapC target sequences in genes within the *M. smegmatis* genome (Total in Genome), Genes with no change in the microarray compared to genes both downregulated and upregulated in the microarray.** Total number of genes refers to the number of genes in each subset. The number of genes the VapC target sequence is present in and times ATAT and ATAA present refers to the total number of times VapC target sequence is present in the genes. Percentage of genes with ATAT/ATAA refers to the percentage of genes containing the VapC target sequence in each subset (Total number of genes with ATAT/ATTA).

Genes upregulated in the microarray (Robson, 2010) have a lower frequency (13%) of VapC cleavage sites compared to the rest of the genome (25%). When 150 bp upstream of the gene is included in the analysis there is only a slightly lower frequency of VapC cleavage sites present (23% compared to 25% for the rest of genome). Regardless if 150 bp is included in the analysis, the frequency of the VapC cleavage site is higher in the genes downregulated upon VapC expression compared to genes unaffected (no change) by VapC expression, those upregulated upon VapC expression and the total in the genome (Table 5.2 and Table 5.3).

Interestingly 70% (28 out of 39) of the genes downregulated by VapC that contain the recognition sequence in the gene or upstream region are part of an operon (Table 5.4). Although the position of the gene in the operon varies (Table 5.4) cleavage of the operon mRNA or the leader sequence (upstream of the operon) could result in degradation of the mRNA for the entire operon, affecting a subset of genes.

Gene Number	Gene		Upstream of Gene		Operons & Genes in the Operon	Gene Position in Operon	Gene Function	Pathway
	ATAT	ATAA	ATAT	ATAA				
MSMEG_0062	1	1			MSMEG_0055, 0056, 0057, 0058, 0059, 0060 and 0061	8/8	ftsK/spoIIIE family protein	Cell Division
MSMEG_0071	2		1		MSMEG_0070	2/2	Conserved hypothetical protein	
MSMEG_0131		1	1	1	N		AMP-binding enzyme, putative, Acyl CoA synthetase	Small Molecule metabolism -Degradation Fatty Acids
MSMEG_0410	1		1		MSMEG_0409, 0411 and 0412	2/3	MmpL protein	Cell Envelope Conserved membrane proteins
MSMEG_0452	1				MSMEG_0453 ( <i>aroK</i> ) and 0451	2/3	Inner membrane permease YgbN	Transport/binding proteins Carbohydrates
MSMEG_0467			3	2	N		Membrane transport protein; LysE family of lysine transporters	Transport/binding proteins efflux protein
MSMEG_0634		1			N		PAP2 superfamily protein	Family of phosphatases and haloperoxidases, phospholipid signalling
MSMEG_0635	1				N		Putative conserved exported protein	
MSMEG_0910	1	1	1		N		Hypothetical protein	
MSMEG_1202		1	1		N		Hypothetical protein	
MSMEG_1245	2				MSMEG_1242 ( <i>iscS</i> ), 1243 and 1244	4/4	phosphoadenosine phosphosulfate reductase	Energy metabolism, anaerobic respiration
MSMEG_1263	1				N		Conserved hypothetical protein	
MSMEG_1677	2				MSMEG_1682, 1681,1680 and 1679 (RD)	5/5	Aspartate ammonia-lyase	Cell Envelope Murein sacculus and peptidoglycan
MSMEG_1679	1				MSMEG_1682, 1681,1680 and 1677 (RD)	2/5	AmiB	Cell Envelope Murein sacculus and peptidoglycan
MSMEG_2121	1				MSMEG_2122, 2123 ( <i>dhaK</i> ) and 2124 (RD)	4/4	Multiphosphoryl transfer protein (MTP)	Transport/binding proteins Carbohydrates
MSMEG_2124				1	MSMEG_2121, 2123 ( <i>dhaK</i> ) and 2124 (RD)	1/4	Glycerol uptake facilitator, MIP channel	Small molecule metabolism -Degradation Carbon compounds
MSMEG_2259	1				MSMEG_2258 (RD)	1/2	Carbon starvation protein A	Adaptations and atypical conditions
MSMEG_2273	1				MSMEG_2274 ( <i>hypC</i> ), 2275 ( <i>hypD</i> ) and 2276 ( <i>hypE</i> )	1/3	[NiFe] hydrogenase maturation protein HypF	Energy metabolism, electron transport
MSMEG_2537	1				MSMEG_2533, 2534, 2535, 2536 and 2537	5/5	Transporter protein	Unknown Transporters
MSMEG_2981	1				MSMEG_2982, 2980, 2979 and 2978	2/5	Branched-chain amino acid ABC-type transport ...	Transport/ Binding proteins amino acids
MSMEG_3095				1	MSMEG_3094, 3093, 3092, 3091, 3090 and 3089 ( <i>deoC</i> )		D-ribose-binding periplasmic protein	Transport/binding proteins Carbohydrates

	Gene	Upstream of Gene						
Gene Number	ATAT	ATAA	ATAT	ATAA	Operons & Genes in the Operon	Gene Position in Operon	Gene Function	Pathway
MSMEG_3253	1				N		Conserved hypothetical protein	
MSMEG_3297	1				N		Transcriptional regulator CadC	Broad Regulatory Function repressor/activator
MSMEG_3363			1		MSMEG_3364	1/2	Regulatory protein, TetR	Broad Regulatory Function, repressor/activator
MSMEG_3375	1				MSMEG_3376	1/2	Alcohol dehydrogenase	
MSMEG_4048	3				MSMEG_4050 and 4049 (RD)	3/3	Cyclase	Other: Cyclases
MSMEG_4082	2	1	1		MSMEG_4087, 4086 ( <i>ssuD</i> ), 4085, 4084, and 4083 (RD)	6/6	Monoxygenase	Miscellaneous oxidoreductases and oxygenases
MSMEG_4978	1				MSMEG_4979 ( <i>cysD</i> ), 4977, 4976, 4975 and 4974 (RD)	5/6	Sulfate adenylyltransferase	Central intermediary metabolism, sulfur metabolism
MSMEG_5420	1				MSMEG_5417, 5418 and 5419	4/4	Tat-translocated enzyme, Dyp family of heme dependent peroxidases	Detoxification
MSMEG_5456	1				N		MK35 lipoprotein	Cell Envelope Lipoproteins
MSMEG_5571			1		MSMEG_5575, 5574, 5573, 5572 ad 5570 (RD)	5/6	ABC transporter ATP-binding protein	Transport/binding proteins Carbohydrates
MSMEG_5573	1				MSMEG_5575,5573,5572,5571 and 5570 (RD)	2/6	Sugar ABC transporter permase	Transport/binding proteins Carbohydrates
MSMEG_5575	1				MSMEG_5574, 5573, 5572, 5571 and 5570 (RD)	1/6	Repressor	Transport/binding proteins Carbohydrates
MSMEG_5586			1		MSMEG_5584 and 5587	2/3	Conserved hypothetical protein	
MSMEG_5866				1	MSMEG_5867 and 5869	1/3	Alcohol dehydrogenase B	Small molecule matabolism -Degradation Carbon compunds
MSMEG_6179		1	1		MSMEG_6175, 6176, 6177 and 6178	5/5	Acetyl -coenzyme A synthetase	Small Molecule metabolism -Degradation Fatty Acids
MSMEG_6224	1		1		N		Retinal pigment epithelial membrane protein	Miscellaneous oxidoreductases and oxygenases
MSMEG_6239		1		1	N		1,3-propanediol dehydrogenase	Small molecule matabolism -Degradation Carbon compunds
MSMEG_6550			1		MSMEG_6552, 6551, 6549 and 6548	1/5	Hypothetical protein	

**Table 5.4. Analysis of VapC<sub>MS1284</sub> target sequence in genes downregulated by VapC expression.** Gene number refers to the downregulated genes with the recognition sequence present. Gene, the presence of ATAT or ATAA in the gene, Upstream of Gene, the presence of ATAT or ATAA 150 bp upstream of the gene. Operons & Genes in the Operon indicates if the gene is in a predicted operon and what other genes are present in the operon, N, not in an operon and RD, the operon is in the reverse direction. Gene Position in Operon, indicates where the gene is in the operon.

### 5.2.9 Bioinformatic Analysis of the VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> Target Sequences

The *P. aerophilum* genome was analysed for the prevalence of the VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> target sequences. GGUG (GGUG) is present in 2169 genes and GGGG present in 2331 genes out of 3689 genes in the genome. Approximately half of the genes containing the target sequences were annotated as hypothetical and there was no pattern as to a metabolic process that could be regulated by VapC in *P. aerophilum*. Of the 2331 genes with a GGGG sequence present 2168 of these also include a GGUG sequence, the total number of genes targeted by VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> is 2332 out of 3689 genes in the genome.

## 5.3 Discussion

### 5.3.1 Identification and Analysis of the Recognition Sites for VapC<sub>MS1284</sub>, VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub>

VapC from *M. smegmatis* degraded total RNA from *M. smegmatis* and specific pentaprobe RNA molecules suggestive of a sequence-specific ribonuclease. VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> from *P. aerophilum* degraded rotavirus RNA, total RNA isolated from *M. smegmatis* and different pentaprobe RNA molecules to those degraded by VapC<sub>MS1284</sub>. Although total RNA is degraded *in vitro* it is unlikely to occur *in vivo* due to protection by ribosome associated proteins and tight packaging of ribosomal RNA. All three VapC proteins did not degrade the RNA oligonucleotide with every combination of two bases or the fluorogenic RNA substrate at CrGA (rG is the ribonucleotide base). Once the ribonuclease activity of VapC proteins had been confirmed, the sequence specificity of VapC<sub>PAE2754</sub>, VapC<sub>PAE0151</sub> and VapC<sub>MS1284</sub> could be determined.

VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> from *P. aerophilum* target the same recognition sequence (Figure 5.5), whereas VapC<sub>MS1284</sub> from *M. smegmatis* displays different sequence specificity. Ribonuclease assays using pentaprobe RNA molecules demonstrated this difference; the cleavage pattern of the RNA molecules and the number of RNA molecules cleaved, differed between VapC<sub>PAE2754</sub>/VapC<sub>PAE0151</sub> and

VapC<sub>MS1284</sub> (Figure 5.3, Figure 5.4 and Figure 5.6). Overlapping RNA oligonucleotides were designed to cover the 932 pentaprobe RNA sequence and a subsequent ribonuclease activity assay screen of these nucleotides revealed that VapC<sub>MS1284</sub> cleaved four out of the nine RNA oligonucleotides and VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> cleaved five out of the nine oligonucleotides. VapC<sub>PAE2754</sub>, VapC<sub>PAE0151</sub> and VapC<sub>MS1284</sub> all cleaved RNA oligonucleotides 3, 4 and 5.

Using MALDI-TOF analysis of 932 RNA oligonucleotides 3, 4 and 5 the recognition sites of VapC<sub>MS1284</sub>, VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> were determined. The optimal target site of VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> is GG\*UG (\* represents cut site), with the next optimal site being GG\*GG. There is flexibility for the bases present at the first, third and fourth positions but a G residue appears to be absolutely required in the second position. A U residue in the third position results in more efficient cleavage, as shown with 932 RNA oligo 5. The consensus sequence of VapC<sub>MS1284</sub> is AU\*A(A/U). There is flexibility for the base at the fourth position but there is consensus for an A residue in the first position, U in the second position and an A in the third position.

Establishing what residues in the recognition sequence for VapC<sub>PAE2754</sub>, VapC<sub>PAE0151</sub> and VapC<sub>MS1284</sub> are absolutely required for hydrolysis of the RNA is important to determine the optimal recognition sequence of VapC proteins. Future work to change the nucleotides in the first, second, third and fourth positions of the recognition sequence is required followed by ribonuclease activity assays is required to determine the optimal residue at each position.

#### ***VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> Target the Same Recognition Sequence***

VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> target a different cleavage site to VapC<sub>MS1284</sub>. It is interesting that the two VapC proteins from the same organism target the same sequence, the only difference between the two proteins is that VapC<sub>PAE2754</sub> displays faster ribonuclease activity compared to VapC<sub>PAE0151</sub>.

---

BLAST analysis (PSI-BLAST) revealed that the closest homologue of VapC<sub>PAE0151</sub> is a PIN-domain protein from *Solfolubus islandicus* and the closest homologue of VapC<sub>PAE2754</sub> is also a PIN-domain protein from *S. islandicus* but is different to the homologue of VapC<sub>PAE0151</sub>. VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> share little protein sequence similarity (39%, determined by Geneious Pro using ClustalW) suggesting that the two genes were acquired independently from *S. islandicus* by HGT. It is therefore unlikely that the two genes have arisen from a gene duplication event within *P. aerophilum*. However, as VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> target the same sequence this suggests that they arose from a gene duplication event within *S. islandicus*. So why would VapC proteins within the same organism target the same sequence? It is possible that the difference in the activities of the two proteins is the answer. Depending on the degree of regulation required, different VapC proteins could be activated under different conditions. However, there are no examples of VapC or other toxin proteins from the same organism targeting the same sequence so it is pure speculation if this differential regulation would occur.

VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> do not appear to bind the target RNA as VapC<sub>MS1284</sub> does; no shift/smear of the RNA substrate was visible in VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> activity assays. RNA binding experiments are needed to confirm this. VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> cleave the pentaprobe RNA molecules with the least amount of secondary structure most efficiently; 926 and 932 pentaprobe RNA molecules have a high degree of secondary structure and although VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> cleave these molecules, more efficient cleavage of pentaprobe RNA molecules 922, 923, 924 and 925 was observed.

VapC<sub>PAE2754</sub> has previously been shown to display nuclease activity against a dsDNA substrate with a large overhang (Arcus, *et al.*, 2004). DNase activity dsDNA was very slow especially compared with the activity using ssRNA as a substrate (Figure 5.3). No activity of VapC<sub>PAE2754</sub> was observed using ssDNA in Figure 5.7 but the time course of this experiment was a lot shorter than that used by Arcus *et al.* (2004).

---

---

***Prevalence of the VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> Target Sequences in the Genome of Pyrobaculum aerophilum***

VapC proteins VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> from *P. aerophilum* target the same recognition sequence. GGGG and GGUG are surprisingly prevalent in the *P. aerophilum* genome, with GGGG present 2331 times in genes and upstream regions, and GGUG present 2169 times out of a total of 3689 genes. Over half of the genes identified for each recognition sequence were annotated as hypothetical, making it difficult to assign a metabolic process targeted by VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub>.

Toxin proteins from other TA families have been shown to have differential toxicity between species, for example MazF from *E. coli* targets ACA sequences (Zhang, Zhang, Hoeflich, *et al.*, 2003), whereas the MazF homologue in *Staphylococcus aureus* targets UACAU sequences (Zhu, *et al.*, 2009). It is not therefore unusual that VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> target sequences present in over half the possible transcripts in the genome whereas the VapC<sub>MS1284</sub> target sequence is much more underrepresented in the *M. smegmatis* genome.

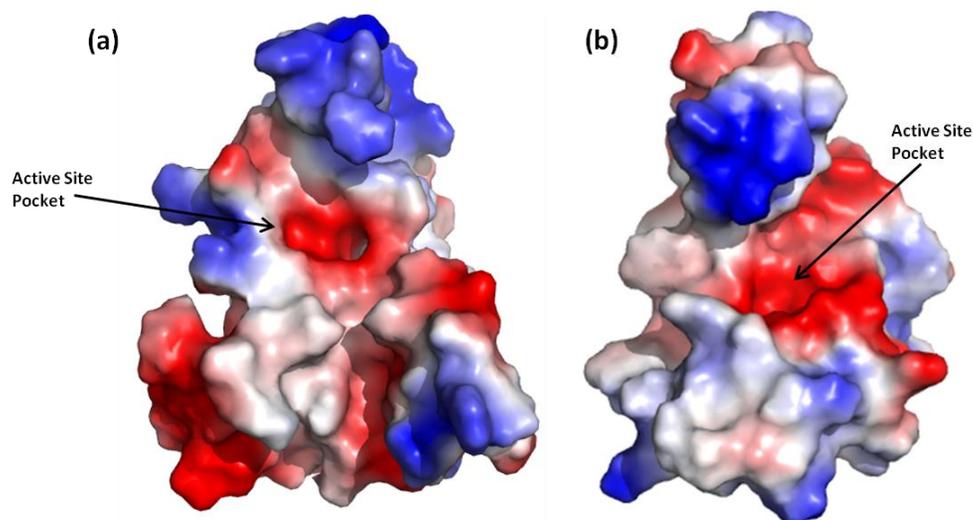
***VapC<sub>MS1284</sub> Possibly Targets Sequence and RNA Structure***

VapC<sub>MS1284</sub> cleaves the 932 pentaprobe RNA molecule more efficiently than the other pentaprobe RNA molecules. The 932 and 926 pentaprobe RNA molecules have the highest amount of secondary structure compared with the other pentaprobe RNA molecules, suggesting that VapC<sub>MS1284</sub> is targeting both sequence and secondary structure. VapC<sub>MS1284</sub> cleaves the pentaprobe RNA molecules with a high degree of secondary structure (925, 926, 927 and 932 (Table 5.1)) and a smear is visible at the top of the gel, resembling a gel shift and indicating binding of VapC<sub>MS1284</sub> to the RNA (Figure 5.6). This shift/smear is not seen when EDTA is present in the reaction and when VapB is bound to VapC, suggesting that binding of the protein to the RNA could play a part in catalysis. Binding of VapC<sub>MS1284</sub> to secondary structure was observed for assays with 927 pentaprobe dsDNA and 926/932 pentaprobe dsRNA. VapC<sub>MS1284</sub> did not cleave these substrates but the smear/shift was present upon

---

analysis by gel electrophoresis (Figure 5.7), and was not present in activity assays with ssDNA. VapC<sub>MS1284</sub> does not cleave the 932 RNA oligonucleotides of the same sequence as efficiently as the whole 932 pentaprobe RNA molecule that has a large amount of secondary structure, which strongly suggests that VapC<sub>MS1284</sub> recognises not only specific sequences but also secondary structure.

VapC<sub>MS1284</sub> has the three acidic residues that constitute the active site of PIN domains, but lacks the fourth less conserved acidic residue. The fact that VapC<sub>MS1284</sub> displays ribonuclease activity suggests that the three acidic residues Asp<sup>4</sup>, Asp<sup>99</sup> and Glu<sup>188</sup> are all that is required for activity. This might contribute to the difference in the sequence specificity of VapC<sub>MS1284</sub> to VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> or sequence recognition could occur elsewhere on the surface of the protein. If VapC<sub>MS1284</sub> is targeting secondary structure in the RNA, a binding or recognition surface could be present on the surface of the protein but as the structure of VapC<sub>MS1284</sub> has not been determined this is difficult to tell. The structure of VapC<sub>MS1284</sub> was modelled using the online structure-homology modelling server SWISS-MODEL (Arnold, Bordoli, Kopp, & Schwede, 2006; Kiefer, Arnold, Kunzli, Bordoli, & Schwede, 2009) and is shown in Figure 5.20. The surface and surface charges of VapC<sub>MS1284</sub> are quite different to that of VapC<sub>PAE0151</sub> (Figure 5.20) which could also account for the difference in sequence specificity and possibly the ability of VapC<sub>MS1284</sub> to target secondary structure and sequence. Notably the two positively charged patches on VapC<sub>MS1284</sub> could be involved in binding of the RNA.



**Figure 5.20. Surface Diagrams of (a) VapC<sub>MS1284</sub> monomer generated by SWISS-MODEL and (b) VapC<sub>PAE0151</sub> monomer.** Red represents acidic or negatively charged regions and blue positively charged regions of the protein. Active site pockets of the two monomers are labelled.

Ribonuclease assays of RNA molecules of varying stem loop structures containing the recognition sequence in stem and loop regions could determine if VapC<sub>MS1284</sub> binds and cuts the recognition sequence in loop structures and this would have important implications in mRNA degradation as stem loop structures are determinants of mRNA stability and half life (Bricker & Belasco, 1999).

### 5.3.2 The Role of VapC<sub>MS1284</sub> and Prevalence of the VapC<sub>MS1284</sub> Target Sequence in the Genome of *Mycobacterium smegmatis*

Microarray analysis of VapC expression in the  $\Delta vapBC$  *M. smegmatis* strain compared with VapBC expression in the same strain revealed a subset of downregulated genes belonging to the general glycerol metabolic pathway associated with glycerol utilisation during aerobic respiration (Robson, 2010). VapC may provide cells with a mechanism of regulating sugar metabolism by reducing the number of catabolic genes and sugar transport proteins to ultimately adjust quickly to new environments. Anabolic and catabolic reactions need to be tightly regulated to prevent the formation of toxic intermediates that can lead to cell death (Russell & Cook, 1995).

---

VapC-mediated ribonuclease activity could maintain the balance between anabolic and catabolic rates. Interestingly, VapC<sub>MS1284</sub> targets AT-rich sequences (the *M. smegmatis* genome has a GC content of 67%). The target sequences ATAT and ATAA for VapC<sub>MS1284</sub> are overrepresented in genes downregulated by VapC, 37% of the VapC downregulated genes (including 150 bp upstream) contained ATAT/ATAA, compared to 25% for the rest of the genome.

Interestingly, the VapC target site is present in three out of six genes in the operon encoding sugar transport proteins. MSMEG\_5571, MSMEG\_5573 and MSMEG\_5575 all contain the VapC recognition sequence, whereas MSMEG\_5570, MSMEG\_5572 and MSMEG\_5574 do not. However, MSMEG\_5572 and MSMEG\_5574 are downregulated by VapC expression, indicating that not only would the protein encoded by the gene containing the VapC target site be downregulated, but others in the operon as well. Cleavage of the operon mRNA or the 5' untranslated region (5'UTR) could result in degradation of the mRNA for the entire operon, affecting a subset of genes. A large number (70%) of genes that contain the recognition sequence in the gene or 5'UTR downregulated by VapC were part of an operon. As genes in an operon are grouped by function, this could have implications for specific metabolic processes in the cell (Ralston, 2008).

### 5.3.3 Comparison of VapC Ribonuclease Activity

Due to toxicity issues in expressing VapC proteins (Daines, *et al.*, 2007; Ramage, *et al.*, 2009; Wilbur, *et al.*, 2005), the ribonuclease activity of few VapC proteins have been determined. To date, the recognition sequence for any VapC protein has not been determined until now. Like VapC<sub>MS1284</sub> from *M. smegmatis* and VapC<sub>PAE0151</sub> and VapC<sub>PAE2754</sub> from *P. aerophilum*, VapC-1 from *H. influenzae* displays ribonuclease activity against total RNA but not ssDNA or dsDNA (Daines, *et al.*, 2007); the target sequence of VapC-1 has not been determined. VapC proteins Rv0301 and Rv1561 from *M. tuberculosis* cleave MS2 viral RNA (Ramage, *et al.*, 2009) but again the sequence specificity of these proteins is unknown. VapC-5 from *M. tuberculosis* also displays weak ribonuclease activity against an unknown RNA

---

substrate (Miallau, *et al.*, 2009). Due to the difficulty in isolating VapC the VapBC-5 complex is hypothesised to dissociate in the assay buffer which has a small amount of detergent, explaining the weak ribonuclease activity observed. Consistent with inhibition of FitB by FitA (Mattison, *et al.*, 2006), VapC<sub>MS1284</sub> and VapC<sub>PAE0151</sub> ribonuclease activity was inhibited by their cognate VapB antitoxins.

Like the MazF toxin, the recognition sequence of VapC differs between different organisms (Zhang, Zhang, Hoeflich, *et al.*, 2003; Zhu, *et al.*, 2006). VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> from the same organism target the same sequence, whereas MazF proteins from *M. tuberculosis* target different recognition sequences to the MazF homologue in *E. coli* (Zhu, *et al.*, 2008; Zhu, *et al.*, 2006). The target sequences of VapC<sub>PAE2754</sub>, VapC<sub>PAE0151</sub> and VapC<sub>MS1284</sub> differs from other mRNA interferases (Table 5.). Each toxin could possibly target different subsets of mRNAs to control different cellular processes within the same organism.

It is possible that VapC proteins from the same organism target different sequences, possibly targeting mRNAs specific to different metabolic pathways. *M. tuberculosis* has 45 *vapBC* operons, raising the possibility that each VapC protein targets a different metabolic pathway to strictly regulate catabolism and anabolism, and slow the growth of the organism. It would be of great interest to determine if the 45 VapC proteins from *M. tuberculosis* target different recognition sequences and if Rv0624, the homologue of VapC<sub>MS1284</sub> in *M. tuberculosis* targets the same sequence as VapC<sub>MS1284</sub>.

Strain	mRNA Interferase	Target Site
<i>Escherichia coli</i>	MazF	ACA
	PemK	UAH (H is C, A or U)
	ChpBK	ACY (Y is G, A or U)
	MqsR	GCU
<i>Bacillus anthracis</i>	PemK	C/U (pyrimidines)
<i>Mycobacterium tuberculosis</i>	MazF-mt1	CUACC and UUACA
	MazF-mt3	CUCCU and UCCU
	MazF-mt6	(U/C)U(A/U)C(U/C)
	MazF-mt7	UCGCU
<i>Myxococcus xanthus</i>	MazF-mx	GUUGC
<i>Staphylococcus aureus</i>	MazFsa	UACAU
<i>Bacillus subtilis</i>	EndoA	UAC
<i>Pyrobaculum aerophilum</i>	VapC <sub>PAE2754</sub>	GGUG or GGGG
	VapC <sub>PAE0151</sub>	GGUG or GGGG
<i>Mycobacterium smegmatis</i>	VapC	AUA(A/U)

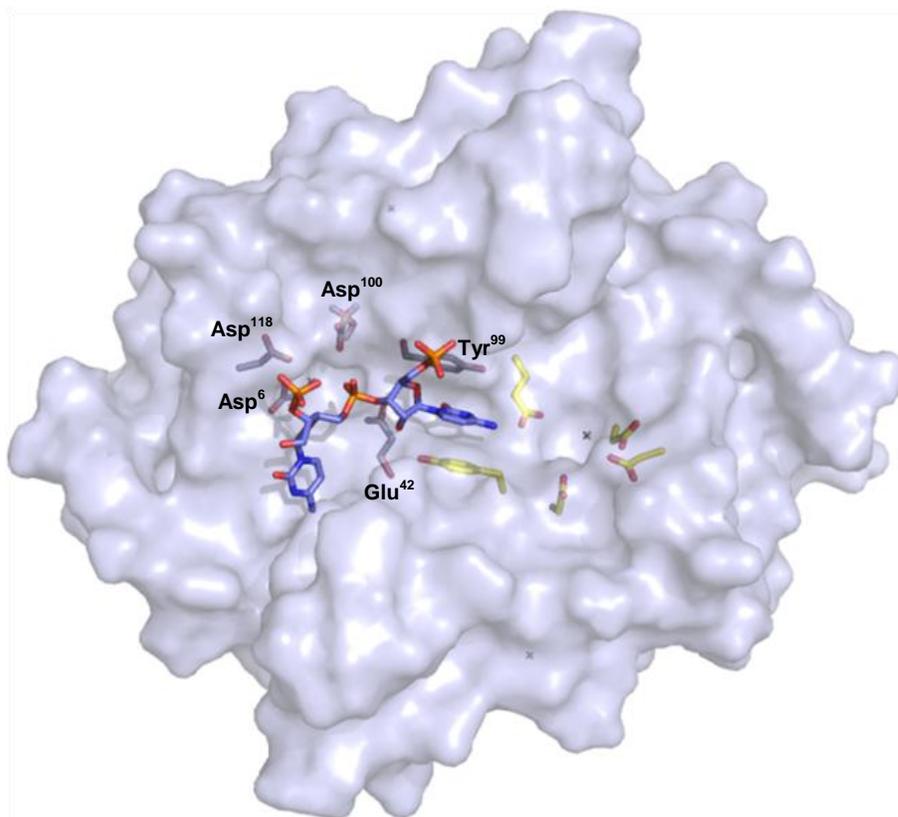
**Table 5.6. mRNA interferases in bacteria and their target sequences.** mRNA interferases are toxin proteins that act on free RNA in the cytoplasm, i.e. the RNA does not need to be translated.

VapC proteins from enterobacteria activate YoeB to induce ribosome dependent mRNA cleavage at stop codons (Winther & Gerdes, 2009). There are no other known examples of VapC activation of other TA systems to induce mRNA cleavage. VapC proteins demonstrate ribonuclease activity against a variety of RNA substrates (Daines, *et al.*, 2007; Miallau, *et al.*, 2009; Ramage, *et al.*, 2009), consistent with the structural similarity of PIN domains to Flap nucleases (Arcus, *et al.*, 2004) and the presence of three conserved acidic, catalytic residues in the active site. The catalytic mechanism of VapC/PIN domains is currently unknown but is discussed below.

#### 5.3.4 Predicted Catalytic Mechanism of VapC

VapC<sub>MS1284</sub>, VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> all form dimers, and weak tetramer interactions have been shown in the crystal structure of VapC<sub>PAE2754</sub> (Arcus, *et al.*, 2004). VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> target the same recognition sequence and cleave this recognition sequence in the same position, despite the fact that VapC<sub>PAE2754</sub> forms

a tetramer and VapC<sub>PAE0151</sub> forms a dimer. From analysis of the recognition sequences of the three VapC proteins it is assumed that only one active site in the dimer is catalytically active at a time, i.e. the substrate is positioned in the active site for cleavage of the RNA molecule by one active site only. If both sites in the dimer were active you would expect cleavage sites on the RNA to be very close i.e. N\*N\*N or N\*NN\*N (N= A, U, C or G, \* = cleavage site) and the recognition sequence to be an inverted repeat so the same sequence is present in each active site pocket. Only a single cleavage site is present in the recognition sequences of VapC<sub>MS1284</sub>, VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub>, suggesting that only one active site pocket is catalytically active at a time. The VapC<sub>PAE0151</sub> dimer is symmetric, however RNA is not symmetric under a two-fold rotation. In order for VapC<sub>PAE0151</sub> to cleave RNA and leave a 5' phosphate on the 3' cleavage product it is virtually impossible that each site is simultaneously active as the RNA substrate would not be positioned correctly in each active site for simultaneous cleavage to occur.



**Figure 5.20. Surface diagram of the VapC<sub>PAE0151</sub> dimer, with a two base RNA nucleotide docked into the active site using GOLD.** The four conserved acidic, active site residues and the tyrosine residues involved in base stacking and are shown in purple for one monomer and yellow for the other. Amino acids hypothesised to be involved in catalysis are labelled in one monomer.

A two base ribonucleotide substrate was modelled into the active site of VapC<sub>PAE0151</sub> dimer using the GOLD protein-ligand docking program (Cambridge Crystallographic Data Centre, UK). Figure 5.20 shows that this positions the phosphate backbone close to the Glu<sup>42</sup> residue in the bottom of the active site pocket. This glutamate could potentially initiate nucleophilic attack on the phosphate backbone either directly or indirectly by activation of a water molecule. The hydroxyl group present on the ribose base is also within hydrogen bonding distance to the glutamate residue, explaining the requirement for an RNA and not a DNA substrate. The Mn<sup>2+</sup>/Mg<sup>2+</sup> ion required for catalysis is co-ordinated by Asp<sup>100</sup> and Asp<sup>118</sup> residues, which could act in positioning the phosphate group on the second nucleotide and/or the phosphate backbone for cleavage. Two tyrosine residues (Tyr<sup>99</sup> from each monomer) between the two active site pockets (coloured in yellow and purple in Figure 5.20) are required for stacking of the base and subsequent positioning of the RNA molecule for cleavage. The tyrosine residue involved in base stacking is also present in VapC<sub>PAE2754</sub> (Tyr<sup>91</sup>) and VapC<sub>MS1284</sub> (Tyr<sup>96</sup>). Crystallisation of a VapC protein with a non-hydrolysable DNA substrate would show the regions on the surface of VapC which determine sequence specificity and allow the development of a predictive method of sequence specificity for other VapC proteins.

MALDI-TOF analysis of the cleavage products of VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> revealed a 5' phosphate present on the 3' cleavage product. The MALDI-TOF spectra obtained for VapC<sub>MS1284</sub> were more ambiguous but supported the presence of a 5' phosphate on the 3' cleavage product. This is in contrast to the mechanism of MazF and PemK (Kid) which cleave the phosphodiester bond at the 5' end to yield a free 5' OH group on the 3' cleavage product and a 2'-3' cyclic phosphate on the 3' end of the 5' cleavage product (Munoz-Gomez, *et al.*, 2005; Zhang, Zhang, *et al.*, 2005). In PemK only one active site in the protein dimer is occupied at a time.

By superimposition of the active site of the exo- and endo-nuclease FEN-1 onto VapC-5 from *M. tuberculosis*, Miallau *et al.* hypothesised a two metal ion catalytic mechanism (Miallau, *et al.*, 2009). Comparisons by Anantharaman & Aravind of the

---

PIN and NYN domains with the FLAP exonuclease domains suggest that PIN domains are more likely to co-ordinate a single metal ion (Anantharaman & Aravind, 2006), which is consistent with the structure of the VapC<sub>PAE0151</sub> from *P. aerophilum* which has only a single metal ion in the active site (Bunker, *et al.*, 2008).

The catalytic mechanism of PIN domains is unknown but VapC proteins cleave ssRNA at the 3' side of the phosphate to leave a 5' phosphate on the 3' cleavage product. This is consistent with other metal-dependent ribonuclease enzymes like RNase H, which activate a water molecule to initiate nucleophilic attack and cleave the 3'-O-P bond of RNA in DNA/RNA duplexes to produce 3' hydroxyl and 5' phosphate cleavage products (Tadokoro & Kanaya, 2009). Future work to establish the exact catalytic mechanism of VapC/PIN domain proteins might allow prediction of the mRNA sequences targeted by each member of this large family of proteins.

---

## Chapter Six: Discussion

The function of toxin-antitoxin (TA) systems is a subject of much debate. The biological roles of the well characterised RelBE and MazEF TA systems is not well determined despite decades of research. MazF from *Escherichia coli* is an mRNA interferase that cleaves at ACA sequences to inhibit translation (Zhang, Zhang, Hoeflich, *et al.*, 2003) and is proposed to lead to cell stasis under various stress conditions (Gerdes, *et al.*, 2005). However, others propose that induction of MazF mediates programmed cell death (PCD) after a point of no return (Amitai, *et al.*, 2004), and this *mazEF*-mediated cell death is a population phenomenon that requires a linear pentapeptide called the extracellular death factor (EDF) (Kolodkin-Gal & Engelberg-Kulka, 2008; Kolodkin-Gal, *et al.*, 2007). The only unequivocal example of MazF PCD in the literature is the role of MazF in obligatory cell lysis in *Myxococcus xanthus* (Nariya & Inouye, 2008). RelE associates with ribosomes to induce ribosome-mediated RNA cleavage at stop codons (Christensen & Gerdes, 2003) and reduces translation. Reduction of translation reduces the level of charged tRNAs which activates RelA leading to stringent control of gene expression through ppGpp (Christensen & Gerdes, 2004; Gerdes, *et al.*, 2005).

Due to the relatively recent identification of VapBC TA systems, little is known about their contemporary biological function, despite the fact that VapBC systems are more widely spread than MazEF and RelBE systems in bacteria and archaea (Arcus, *et al.*, 2005; Gerdes, *et al.*, 2005). Given the large number of *vapBC* loci in *Mycobacterium tuberculosis* the role of these TA systems is of huge interest. To investigate the function of *vapBC* operons in the mycobacteria, I have characterised the biochemistry and sequence specificity of VapC from *Mycobacterium smegmatis* and two VapC proteins from *Pyrobaculum aerophilum* to show the target(s) of these proteins, I have also structurally characterised VapBC protein complexes and investigated the possible mechanism of growth inhibition by VapC. Results from these experiments resulted in

two publications (Bunker *et al.*, 2008 & Robson *et al.*, 2009) that can be found in Appendix D.

The *vapBC* operon from *M. smegmatis* exhibits all the characteristics of a TA system. The two genes are transcribed as an operon and are subject to auto-regulation by the VapBC complex binding to inverted repeat sequences in the promoter of *vapBC* (Robson, *et al.*, 2009). The *vapBC* operon is not essential for growth or the stress response (Robson, 2010), therefore it is hypothesised that this TA system is not related to the stress response in *M. smegmatis* as seen with TA systems in other organisms (Christensen, *et al.*, 2001; Christensen, *et al.*, 2003; Hazan, *et al.*, 2004). Expression of VapC induced cell stasis in *M. smegmatis* (Robson, *et al.*, 2009) and reduced protein synthesis (and not DNA replication or transcription) which is consistent with the post-transcriptional function of VapC, via its ribonuclease activity (Section 5.3.1).

### **6.1 The Role of *Mycobacterium smegmatis* VapBC Proteins as Metabolic Managers**

Microarray analysis of transcripts downregulated in response to VapC expression (by collaborators) identified genes specifically involved in sugar transport and subsequent carbon metabolism (Robson, 2010). In particular, transcripts involved in the general glycerol metabolic pathway associated with glycerol utilisation during aerobic respiration were downregulated. VapC could therefore provide the cells with a mechanism to regulate sugar metabolism by reducing transporters and associated catabolic genes, thereby allowing the cell to balance catabolic and anabolic reactions.

An imbalance between catabolism and anabolism can produce toxic intermediates that can lead to cell death (Russell & Cook, 1995). Therefore cells must have a mechanism to adjust quickly under changing environments. This cannot be conferred by transcription factors alone as the mRNA and the encoded protein would still be present. VapC could be involved in post-transcriptional regulation of metabolic genes to fine tune the cell's immediate response to changing environments (Robson, 2010). Obligate host-associated organisms like *Mycobacterium leprae* have lost their TA systems by reductive evolution and inhabit a stable environment, TA systems may be beneficial to

---

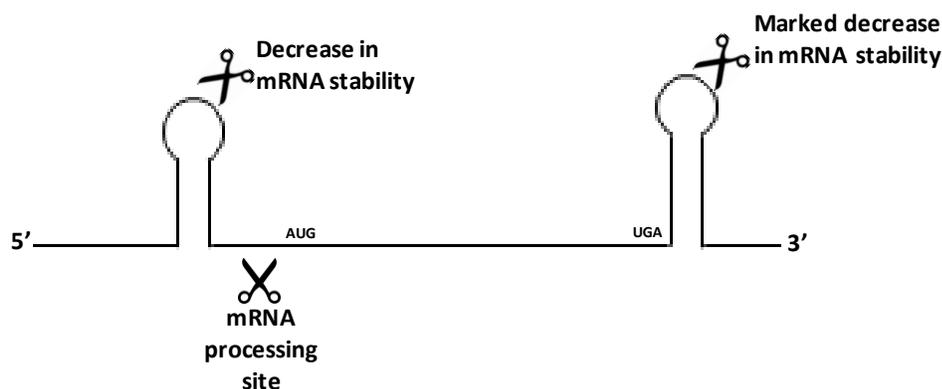
free-living organisms that face environmental changes as a mechanism of regulating their metabolism to adjust to these changes (Arcus, McKenzie, Robson, & Cook, 2011). VapC could therefore maintain the balance of anabolic and catabolic rates in *M. smegmatis* by VapC-mediated ribonuclease activity. VapC from *M. smegmatis* targets AUAA and AUAU sequences in mRNA, yet the genomes of mycobacteria species are GC rich. The *M. smegmatis* genome has a GC content of 67%, which makes it extremely interesting that VapC<sub>MS1284</sub> targets AT rich sequences. ATAA is the 10<sup>th</sup> rarest quartet in the *M. smegmatis* genome and ATAT the 23<sup>rd</sup> rarest (out of a possible 256 quartets).

The determination of the VapC target sequences *in vitro* (AUAU and AUAA) is consistent with *in vivo* microarray studies as the VapC recognition sequences are overrepresented (37%) in the genes and their upstream regions downregulated by VapC expression compared to the rest of the genome (25%). Analysis that did not include 150 bp upstream of the gene demonstrated the same trend. Interestingly, when the upstream region was not included in the analysis only 13% of genes upregulated in the microarray study contained the VapC recognition sequence.

The majority (70%) of the genes downregulated by VapC that contain the recognition sequences in the gene or 5'UTR were part of an operon. Cleavage of the operon mRNA or 5'UTR could result in degradation of the mRNA for the entire operon, affecting a subset of genes. For example, the VapC target site is present in the three genes of the operon encoding sugar transport proteins and five out of the six genes in the operon are downregulated in response to VapC. This is consistent with the theory that VapC is targeting specific metabolic processes. Genes in operons are often grouped by function (Ralston, 2008), this is also consistent with VapC targeting transcripts associated with specific metabolic processes

The general system for mRNA degradation in the mycobacteria has not been well studied, however it has been well characterised in *E. coli* (Kushner, 2002) and *Bacillus subtilis* (Condon, 2003). A major determinant of mRNA stability in both organisms is

the 5' end of the transcripts (Bricker & Belasco, 1999; Condon, 2003) termed the 5' UTR. Secondary structures increase mRNA stability and provide a barrier against nuclease degradation (Figure 6.1) (Steege, 2000). Some genes downregulated by VapC include the target sequence in the upstream region as outlined in Table 5.4. If the gene has a 5' UTR, it is feasible that VapC could cleave this, removing the secondary structure and making the mRNA more susceptible to degradation by general RNA degradation pathways within the cell (Figure 6.1). Additional evidence for this is that VapC from *M. smegmatis* appears to target a combination of secondary structure and sequence, and *in vitro* can cleave mRNA transcripts with a large amount of secondary structure.

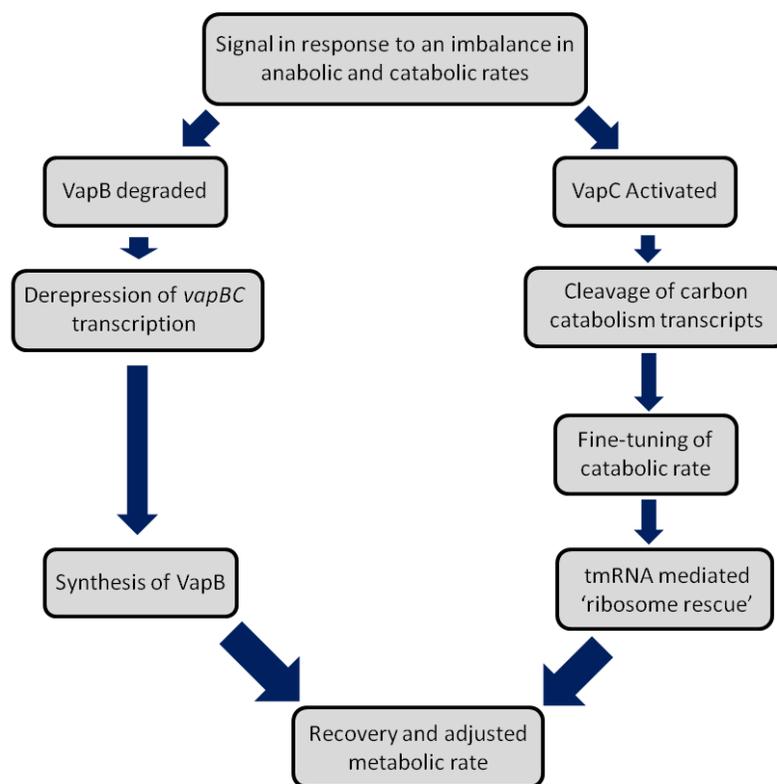


**Figure 6.1. Model of mRNA stability.** Stem loop structures in the 5'UTR and the 3' end of the RNA provides protection from general ribonucleases within the cell, thereby increasing the stability of the mRNA. If the 5' stem-loop structure is cleaved by VapC there is a decrease in mRNA stability, if the 3' stem-loop structure is cleaved there is a marked decrease in mRNA stability. The 5' stem-loop may be removed for translation to occur (mRNA processing site). Scissor icons represent possible nuclease cleavage sites, AUG, start codon and UGA stop codon.

A model similar to that of the stress response hypothesis for chromosomal TA systems has been proposed for the role of VapBC proteins as metabolic managers (Figure 6.2) (Robson, 2010). According to this model an imbalance in the catabolic and anabolic rates results in the degradation of VapB by cellular proteases causing a shift in the stoichiometry of the VapBC complex leading to derepression of the *vapBC* operon. This results in an increased number of VapBC transcripts. In a parallel pathway VapC cleaves specific mRNAs involved in glycerol metabolism (carbon catabolism) to fine tune the metabolic rate. Stalled ribosomes are rescued by tmRNA (a tRNA-mRNA

hybrid) by tagging the incomplete, stalled polypeptide chain for degradation, thereby allowing translation to terminate normally and ‘rescues’ the ribosome. This enables translation of *vapBC* mRNA restoring the ratio of VapB to VapC and repressing the *vapBC* operon.

This model is based on the findings from microarray analysis that expression of VapC results in (1) an upregulation of phosphate transporters and proteins implying that changes in intracellular phosphate levels may occur and (2) a downregulation in genes involved in glycerol metabolism which therefore reduces catabolic proteins within the cell (Robson, 2010). This prevents the formation of toxic intermediate metabolites formed by the imbalance between anabolism and catabolism. In the proposed model cells are able to recover and resume growth at a new metabolic rate (Robson, 2010).



**Figure 6.2 Model for the role of VapC proteins as metabolic managers in *M. smegmatis*.** An imbalance in catabolism and anabolism results in degradation of VapB and derepression of the *vapBC* operon, resulting in an increased number of VapBC transcripts. In a parallel pathway VapC cleaves mRNAs involved in glycerol metabolism to fine tune the metabolic rate. Stalled ribosomes are rescued by tmRNA and translation resumes. Figure adapted from Robson (2010).

---

In accordance with the model in Figure 6.2, VapB is more susceptible to proteolytic degradation compared to VapC *in vitro* (Section 3.2.12, trypsin digest of the VapBC complex). The stoichiometry of the purified VapBC complex is a tetramer of VapBC heterodimers. This is the same as that for FitAB from *Neisseria gonorrhoeae*, where the tetramer binds to inverted repeat sequences in the *fitAB* promoter. *In vitro* VapC targets AUAU and AUAA sequences and these sequences are overrepresented in genes downregulated in response to VapC, providing evidence for the mechanism of downregulation. The *vapBC* operon from *M. smegmatis* does not contain the VapC target sequence, therefore VapC cannot cleave the *vapBC* mRNA. Thus, it can be translated upon ‘ribosome rescue’ by tmRNA and the ratio of VapB to VapC for repressor complex formation restored.

Further work is needed to validate this model and experimental strategies to do this include the following: Demonstration of the cleavage specificity of VapC on selected transcripts downregulated in response to VapC using the ribonuclease activity assay in Chapter Five. Now the target sequence of VapC is known, the *in vivo* activation of VapC under specific environmental conditions could be determined by using a reporter such as green fluorescent protein (GFP) for a direct readout of toxin activity. The use of deep RNA sequencing would allow quantification as to the extent to which VapC regulates specific genes under certain conditions, and potentially allow identification of the sites at which VapC is cleaving these targeted transcripts. If VapC is targeting a combination of sequence and secondary structure this could have important implications for the mode of action and the targets of VapC; the hypothesised targets identified in this study could be underestimated if secondary structure is involved in cleavage site recognition.

## 6.2 VapC from *Pyrobaculum aerophilum*

Two VapC proteins from the thermophilic crenarchaeon *P. aerophilum* display sequence-specific ribonuclease activity. Both VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> target the

same sequence. The optimal sequence GGTG (GGUG) is present in 2169 genes in the *P. aerophilum* genome and the slightly less optimal sequence, GGGG is present in 2331 genes out of 3689 genes in the genome, 2168 genes contain both GGUG and GGGG sequences. It could be possible that the differences in the activity of the two proteins is important; VapC<sub>PAE0151</sub> is less active than VapC<sub>PAE2754</sub> under equivalent conditions so a different protein could be activated depending on the level of regulation required. No subsets of genes were identified as belonging to a metabolic process containing the target sequence that could be regulated by VapC in *P. aerophilum*.

VapC<sub>PAE0151</sub> and VapC<sub>PAE2754</sub> cleave over half the possible transcripts in the *P. aerophilum* genome, suggesting potent toxin activity. More transcripts cleaved by VapC<sub>PAE0151</sub> and VapC<sub>PAE2754</sub> are present in the *P. aerophilum* genome compared to transcripts cleaved by VapC<sub>MS1284</sub> in the *M. smegmatis* genome. This is not uncommon in TA systems; MazF from *E. coli* cleaves ACA sequences (Zhang, Zhang, Hoeflich, *et al.*, 2003) making it an extremely potent toxin, whereas MazF homologues in *M. tuberculosis* and *Staphylococcus aureus* cleave at pentad recognition sequences (Zhu, *et al.*, 2009; Zhu, *et al.*, 2008; Zhu, *et al.*, 2006) making them far less potent toxins.

The genes encoded by *vapBC* operons from *P. aerophilum* have not been studied in detail compared to the *vapBC* operon from *M. smegmatis*. The sequence specificity of VapC<sub>PAE2754</sub> and VapC<sub>PAE0151</sub> proteins have been well characterised in this study, but there has been little investigation as to the role of these proteins in *P. aerophilum*.

### 6.3 Conclusions

VapC proteins are present in 490 bacteria and archaea and this is half of all sequenced prokaryotes. These include the pathogens *N. gonorrhoeae* (Mattison, *et al.*, 2006; Wilbur, *et al.*, 2005), *M. tuberculosis* (Miallau, *et al.*, 2009; Ramage, *et al.*, 2009), pathogenic strains of *E. coli* (Iguchi, *et al.*, 2009) and *Haemophilus influenzae* (Daines, *et al.*, 2007). VapC proteins are characterised by three conserved acidic residues that constitute the active site and bind Mg<sup>2+</sup>/Mn<sup>2+</sup> which is essential for activity (Arcus, *et*

*al.*, 2005). The strict conservation of active site residues suggests all VapC proteins are ribonucleases. Three VapC proteins from *M. tuberculosis* and VapC-1 from *H. influenzae* along with two VapC proteins from *P. aerophilum* and VapC from *M. smegmatis* display ribonuclease activity but to date the only sequence specificity determined for these proteins has been that of VapC from *M. smegmatis* and VapC<sub>P<sub>AE</sub>2754</sub> and VapC<sub>P<sub>AE</sub>0151</sub> from *P. aerophilum*.

VapC proteins from different organisms target different sequences, but do VapC proteins from the same organism target the same sequence? VapC<sub>P<sub>AE</sub>2754</sub> and VapC<sub>P<sub>AE</sub>0151</sub> from *P. aerophilum* target the same sequence but it is possible that different VapC proteins within the same organism can target different sequences. If VapC from *P. aerophilum* targets a different recognition sequence to VapC from *M. smegmatis* yet they have the same conserved acidic residues it is possible that VapC proteins from the same organism also target different sequences, possibly targeting mRNAs specific to different metabolic pathways. *M. tuberculosis* has 45 *vapBC* operons, raising the possibility that each VapC protein targets a different metabolic pathway to strictly regulate catabolism and anabolism and slow the growth of the organism. It would be of great interest to determine the sequence specificity of VapC proteins from *M. tuberculosis* and other organisms by expressing and purifying VapC proteins using the methods outlined in Chapter Three and determining the target sequence using the pentaprobe system and MALDI-TOF MS as for VapC from *M. smegmatis* (Chapter Five). Although *M. tuberculosis* is a well studied organism due to its pathogenicity little is known about how the mycobacteria adapt and respond to environmental changes and adapt their metabolic rate accordingly. VapBC proteins could mediate this adaptation and thus play a role in the persistent state of tuberculosis which is one of the primary reasons for treatment difficulties, and the success of this pathogen.

## Appendices

### Appendix A: Reagents

#### A.1 Primers, Plasmids & Bacterial Strains Used in This Study

Primer	Sequence
MS 1283F	TCAGCTCCATGGCTCTAAGCATCAAACA
MS 1284R New	TATTTAGGATCCGCGTGGACCGCAGCG
MS 1283R New	TATAACGGATCCGAGGACGGCAGACCG
T7 Fwd	TAATACGACTCACTATAGGG
T7 Rev	TAGTTATTGCTCAGCGGTGG
MS Promoter Fwd	GTCGTATAGATTCATGAAGTTTGTCTATAATCA
MS Terminator Rev	CTGCATATCTAAGTACTTCAAACAGATATTAGT
MS Control Fwd	GGCGATCTATCAAGGTCGTATAGATTCAT
MS Control Rev	CCGCTAGATAGTTCCAGCATATCTAAGTA
MBP MS 1284 Fwd	CTAGAATTCATGGTTATCGACACTTCTG
MBP MS 1284 Rev	TATGGATCCTCAGTGGACCGCAGC
malE 5'	GGTCGTCAGACTGTCGATGAAGCC
M13/pUC 3'	CGCCAGGGTTTTCCCAGTCACGAC
pcDNA3 PP Fwd	AGAGAACCCACTGCTTACTGGCT
pcDNA3 PP Rev	AGCGAGCTCTAGCATTTAGGTGACA

**Table A.1. List of Primers used in the study**

Strain	Description
<i>Escherichia coli</i>	
DH10B (TOP10)	F- <i>mcrA</i> $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ) $\Phi$ 80 <i>dlacZ</i> $\Delta$ M15 $\Delta$ <i>lacX74</i> <i>deoR</i> <i>recA1</i> <i>araD139</i> $\Delta$ ( <i>ara leu</i> )7697 <i>galU</i> <i>galK</i> <i>rpsL</i> <i>endA1</i> <i>nupG</i>
DH5 $\alpha$	<i>fhuA2</i> $\Delta$ ( <i>argF-lacZ</i> )U169 <i>phoA</i> <i>glnV44</i> $\Phi$ 80 $\Delta$ ( <i>lacZ</i> )M15 <i>gyrA96</i> <i>recA1</i> <i>relA1</i> <i>endA1</i> <i>thi-1</i> <i>hsdR17</i>
BL21 (DE3)	F- <i>ompT</i> <i>hsdS<sub>B</sub></i> ( <i>r<sub>B</sub>-m<sub>B</sub>-</i> ) <i>gal</i> <i>dcm</i> (DE3)
<i>Mycobacterium smegmatis</i>	
mc <sup>2</sup> 155	Electrocompetent wild-type strain of <i>M. smegmatis</i>
mc <sup>2</sup> 4517	<i>M. smegmatis</i> expression strain with T7 RNA polymerase, Km <sup>r</sup>
JR121	mc <sup>2</sup> 155 $\Delta$ <i>vapBC</i> :: <i>aphA-3</i> ; Km <sup>r</sup>
<i>Plasmids</i>	
pYUB1049	<i>E. coli</i> mycobacterium shuttle vector with T7 promoter and encoding both C and N-terminal His-tags; Hyg <sup>r</sup>
pMind	Tetracycline inducible expression vector; Km <sup>r</sup> Hyg <sup>r</sup>
pMAL-c2x	<i>E. coli</i> expression vector with <i>malE</i> for expression of a MBP fusion protein; Amp <sup>r</sup>
pProEX-Hta/c	<i>E. coli</i> expression vector encoding a N-terminal His-tag including rTEV site; Amp <sup>r</sup>
pcDNA3	Mammalian Expression Vector; Amp <sup>r</sup>

**Table A.2. List of bacterial strains and plasmids used in the study**

## A.2 Buffers and Solutions

BLOTTO	10% (w/v) non-fat milk powder in 1 x TBS-T
5 x DNA Loading Dye	0.05% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 30% (v/v) glycerol
Developer (Silver Stain)	3 g Na <sub>2</sub> CO <sub>3</sub> , 1 ml 0.02% (v/v) sodium thiosulphate, 25 µl formalin to 50 ml with H <sub>2</sub> O
Fairbanks Staining Solution A	0.05% (w/v) coomassie blue R-250, 25% (v/v) isopropanol, 10% (v/v) acetic acid
Fairbanks Staining Solution B	0.005% (w/v) coomassie blue R-250 10% (v/v) isopropanol, 10% (v/v) acetic acid
Fairbanks Staining Solution C	0.002% (w/v) coomassie blue, 10% (v/v) acetic acid
Fairbanks Staining Solution D	10% (v/v) acetic acid
Fixer Solution (Silver Stain)	50% (v/v) methanol, 12% (v/v) acetic acid, 0.05% (v/v) formalin
Formamide Loading Buffer	0% (v/v) formamide, 5 mM EDTA, 0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyanol FF
GITC	295.4 g guanidine thiocyanate, 2.5 g N-lauroyl sarcosine, 3.9 g tri-sodium citrate, 3.6 ml 2-mercaptoethanol, 280 ml DEPC H <sub>2</sub> O (total volume 500 ml)
Native Gel Loading Buffer	0.3 M Tris-HCl, 1% (w/v) bromophenol blue pH 6.8
10 x PBS	0.2 M phosphate (0.038 M NaH <sub>2</sub> PO <sub>4</sub> , 0.162 M Na <sub>2</sub> HPO <sub>4</sub> , 1.5 M NaCl, pH 7.4
1 x PBS	100 ml 10 x PBS + 900 ml H <sub>2</sub> O
Ponceau Stain	0.2% (w/v) ponceau Red, 1% (v/v) acetic acid
Resolving Buffer	1.5 M Tris-HCl pH 8.8
4 x SDS loading buffer	200 mM Tris-HCl pH 6.8, 8% (w/v) SDS, 40% (v/v) glycerol, 0.4% (w/v) bromophenol blue, 400 mM β-mercaptoethanol
SDS running buffer	25 mM Tris-HCl pH 6.8, 0.1% (w/v) SDS 190 mM glycine
Silver Stain	100 mg AgNO <sub>3</sub> , 38 µl formalin to 50 ml with H <sub>2</sub> O
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 <sub>2</sub> O
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 <sub>2</sub> O
10 x TBS	200 mM Tris-HC pH 7.6, 1.5 M NaCl
1 x TBS – T	100 ml 10 x TBS + 899 ml H <sub>2</sub> O, 0.05% (v/v) tween 20
TE	10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0
Transfer Buffer	25 mM Tris-HCl, 192 mM glycine, 20% methanol, 0.01% (w/v) SDS
Tris-Glycine Running Buffer	25 mM Tris-HCl 250 mM glycine pH 8.5

## Refolding Screen Solutions

	Buffer	Salt	Cation 2+	PEG (%w/v) 3350	Detergent Dodecyl maltoside	Chaotrope-Guanidine HCL	Polar (P) Nonpolar (NP) Additives	Oxidation/Reduction Potential
1	50 mM Tris pH 8.2	250 mM NaCl	1 mM EDTA	0.05%	0.0 mM	0 mM	0 mM	1 mM DTT
2	50 mM MES pH 6.5	10 mM NaCl	2mM MgCl <sub>2</sub> 2mM CaCl <sub>2</sub>	0.00%	0.3 mM	500 mM	0 mM	1 mM GSH 0.1 mM GSSG
3	50mM MES pH 6.5	10mM NaCl	1mM EDTA	0.05%	0.0mM	500mM	200mM Sucrose (NP) 200mM L-Arginine (P)	0,1mM GSH 0.1mM GSSG
4	50 mM Tris pH 8.2	250 mM NaCl	2mM MgCl <sub>2</sub> 2mM CaCl <sub>2</sub>	0.00%	0.3 mM	0 mM	200 mM Sucrose (NP) 200 mM L-Arginine (P)	1 mM DTT
6	50 mM Tris pH 8.2	10 mM NaCl	1 mM EDTA	0.05%	0.3 mM	500 mM	200 mM Sucrose (NP)	1 mM DTT
7	50 mM Tris pH 8.2	10 mM NaCl	2mM MgCl <sub>2</sub> 2mM CaCl <sub>2</sub>	0.00%	0.0 mM	500 mM	200 mM L-Arginine (P)	1 mM DTT
8	50 mM MES pH 6.5	250 mM NaCl	1 mM EDTA	0.05%	0.3 mM	0 mM	200 mM L-Arginine (P)	1 mM GSH 0.1 mM GSSG
9	50 mM MES pH 6.5	250 mM NaCl	2mM MgCl <sub>2</sub> 2mM CaCl <sub>2</sub>	0.05%	0.0 mM	500 mM	200 mM Sucrose (NP)	1 mM DTT
10	50 mM Tris pH 8.2	10 mM NaCl	1 mM EDTA	0.00%	0.3 mM	0 mM	200 mM Sucrose (NP)	1 mM GSH 0.1 mM GSSG
11	50 mM Tris pH 8.2	10 mM NaCl	2mM MgCl <sub>2</sub> 2mM CaCl <sub>2</sub>	0.05%	0.0 mM	0 mM	200 mM L-Arginine (P)	1 mM GSH 0.1 mM GSSG
12	50 mM MES pH 6.5	250 mM NaCl	1 mM EDTA	0.00%	0.3 mM	500 mM	200 mM L-Arginine (P)	1 mM DTT
13	50 mM Tris pH 8.2	250 mM NaCl	1 mM EDTA	0.00%	0.0 mM	500 mM	0 mM	1 mM GSH 0.1 mM GSSG
14	50 mM MES pH 6.5	10 mM NaCl	2mM MgCl <sub>2</sub> 2mM CaCl <sub>2</sub>	0.05%	0.3 mM	0 mM	0 mM	1 mM DTT
15	50 mM MES pH 6.5	10 mM NaCl	1 mM EDTA	0.00%	0.0 mM	0 mM	200 mM Sucrose (NP) 200 mM L-Arginine (P)	1 mM DTT
16	50 mM Tris pH 8.2	250 mM NaCl	2mM MgCl <sub>2</sub> 2mM CaCl <sub>2</sub>	0.05%	0.3 mM	500 mM	200 mM Sucrose (NP) 200 mM L-Arginine (P)	1 mM GSH 0.1 mM GSSG

---

---

### A.3 Growth Media

All dehydrated media were supplied by Difco.

#### A.3.1 Solid Media

7H10-agar	1.9 g 7H10 powder, 0.5% (w/v) glycerol in 90 ml H <sub>2</sub> O, autoclave then add 10 ml ADC enrichment and 0.05% (v/v) tween 80 at 50-50 °C.
LB-agar	1% (w/v) bactotryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, 15 g.L <sup>-1</sup> agar pH 8.0
LBT-agar	1% (w/v) bactotryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, 15g.L <sup>-1</sup> agar, 0.05% (v/v) tween 80 pH 8.0
Low salt LB-agar	1% (w/v) bactotryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 15g.L <sup>-1</sup> agar pH 8.0

#### A.3.2 Liquid Media

7H9	0.47 g 7H9 powder, 0.2% glycerol in 90 ml H <sub>2</sub> O, autoclave then add 10 ml ADC enrichment and 0.05% (v/v) tween 80 at 45 °C.
Hartmans de Bont	Components added in order given: 1 x trace metal stock*, 15 mM (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 0.05% (v/v) tween 80, 0.2% (v/v) glycerol, 50 mM MOPS. Autoclave then add 8.9 M K <sub>2</sub> HPO <sub>4</sub> and 7.08 M NaH <sub>2</sub> PO <sub>4</sub>
LB	1% (w/v) bactotryptone, 0.5%(w/v) yeast extract, 1% (w/v) NaCl pH 8.0
LBT	1% (w/v) bactotryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl 0.05% (v/v) tween 80 pH 8.0
PA-0.5G	50 mM Na <sub>2</sub> HPO <sub>4</sub> , 50 mM KH <sub>2</sub> PO <sub>4</sub> , 25 mM (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 1 mM MgSO <sub>4</sub> , 0.5% (w/v) glucose, 0.1 x metals mix**, 200 µg.ml <sup>-1</sup> each of 17 amino acids (no Cys, Tyr or Met). Individual components autoclaved to sterile filtered before adding to sterile H <sub>2</sub> O.
Rich Media + Glucose	1% (w/v) bactotryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 0.2% (w/v) glucose
SOC	2% (w/v) bactotryptone or bactopectone, 0.55% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl 10 mM MgCl <sub>2</sub> , 10 mM MgSO <sub>4</sub> , 20 mM glucose.

---

---

TB (Terrific Broth)	12% (w/v) bacto-tryptone, 24 % (w/v) yeast extract, 4% (w/v) glycerol autoclave. After autoclaving combine with
ZYP-5052	1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract, 50 mM Na <sub>2</sub> HPO <sub>4</sub> , 25 mM KH <sub>2</sub> PO <sub>4</sub> , 25 mM (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 1 mM MgSO <sub>4</sub> , 0.5% (w/v) glycerol, 0.05% (w/v) glucose, 0.2% (w/v) α-lactose, 1 x metals mix

\*100 x trace metals stock for HdB media: EDTA [0.1 g], MgCl<sub>2</sub>·6H<sub>2</sub>O [1 g], CaCl<sub>2</sub>·2H<sub>2</sub>O [10 mg], NaMoO<sub>4</sub>·2H<sub>2</sub>O [2 mg], CoCl<sub>2</sub>·6H<sub>2</sub>O [4 mg], MnCl<sub>2</sub>·2H<sub>2</sub>O [10 mg], ZnSO<sub>4</sub>·7H<sub>2</sub>O [20 mg], FeSO<sub>4</sub>·7H<sub>2</sub>O [50 mg], CuSO<sub>4</sub>·5H<sub>2</sub>O to 100 ml.

\*\* 1000 x metals mix made up from the sterile stock solutions of each component to give the following concentrations: 50 μM FeCl<sub>3</sub> in 0.12 M HCl (filter sterile), 20 μM CaCl<sub>2</sub>, 10 μM MnCl<sub>2</sub>, 10 μM ZnSO<sub>4</sub>, 2 μM CoCl<sub>2</sub>, 2 μM CuCl<sub>2</sub>, 2 μM NiCl<sub>2</sub>, 2 μM Na<sub>2</sub>MoO<sub>4</sub>, 2 μM Na<sub>2</sub>SeO<sub>3</sub>, 2 μM H<sub>3</sub>BO<sub>3</sub>

---

---

**Appendix B: Gene & Protein Information****B.1 *Mycobacterium smegmatis* VapB and VapC Genomic & Protein Information**MS 1283 VapB (255bp)

ATGGCTCTAAGCATCAAACACCCGGAAGCCGACCCGGCTGGCCAGAGAGCTAGCGGCACGTAC  
GGGGGAGACACTGACCGAGGCGGTGGTGTATGGCACTGCGGGAGCGACTGGCGCGCACTGTTCG  
GCCGCACTCAGGTTCGTTCCGCTTCGTGAGGAACTCGCCGCGATCCGTCGTCGCTGCGCAGCC  
CTACCGGTGTTGGACGACCCGGACAGCCGAATCGATCCTGGGCTACGACGACCCGGTCTGCC  
GTCCTGA

MS 1283 Protein Sequence

MALSIKHPEADRLARELAARTGETLTEAVVMALRERLARTVGRTQVVPLREE  
LAAIRRRCAALPVLDDRTAESILGYDDRGLPS

84 amino acids

Molecular Weight: 9279.7

Theoretical pI: 7.98

MS 1283 VapB with pYUB1049 fusion tag (318bp)

ATGGCTCTAAGCATCAAACACCCGGAAGCCGACCCGGCTGGCCAGAGAGCTAGCGGCACGTAC  
GGGGGAGACACTGACCGAGGCGGTGGTGTATGGCACTGCGGGAGCGACTGGCGCGCACTGTTCG  
GCCGCACTCAGGTTCGTTCCGCTTCGTGAGGAACTCGCCGCGATCCGTCGTCGCTGCGCAGCC  
CTACCGGTGTTGGACGACCCGGACAGCCGAATCGATCCTGGGCTACGACGACCCGGTCTGCC  
GTCCCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCA  
CCACCACTGA

MS 1283 VapB with pYUB1049 fusion tag

MALSIKHPEADRLARELAARTGETLTEAVVMALRERLARTVGRTQVVPLREE  
LAAIRRRCAALPVLDDRTAESILGYDDRGLPSSDPNSSSVDKLAALAEHHHHHHH\*

106 amino acids

Molecular Weight: 11688.2

Theoretical pI: 6.65

MS 1284 VapC (390bp)

ATGGTTATCGACACTTCTGCCCTCGTTGCCATCTTGACCGACGAACCCGACGCCGAGTTGTT  
GGAGGGGGCGGTGGCTGACGATCCTGTTCCGACTATGTCCACCGCGTCGTATCTGGAGACAG  
CGATCGTGATCGAAAGTCGCTTCGGCGAGCCAGGTGGTCGCGAACTCGATCTGTGGTTGCAT  
CGCGCATCAGTGGCACTGGTCGCTGTGGATGCCGATCAGGCCGACGCCCGCCCGGTTGGCTTA  
CCGGAGATACGGCAAGGGGCGCCATCGTGCAGGCCTGAATTACGGCGACTGTTTTTCTATG  
CGCTCGCCAAGGTCAGCGGTCAGCCTTTGTTGTTCAAGGGCGAGGCTTTTCGGCTCACCGAC  
GTCGCTGCGGTCCACTGA

MS 1284 VapC Protein Sequence

MVIDTSALVAILTDEPDAELLEGA VADDPVRTMSTASYLETAIVIESRFGEPPGRELDLWLH  
RASVALVAVDADQADAARLAYRRY GKRHRAGLNYGDCFSYALAKVSGQPLLFKGEAFRLTD  
VAAVH

129 amino acids

Molecular Weight: 13909.7

Theoretical pI: 4.94

MS 1284 VapC with pYUB1049 fusion tag (456bp)

ATGGTTATCGACACTTCTGCCCTCGTTGCCATCTTGACCGACGAACCCGACGCCGAGTTGTT  
GGAGGGGGCGGTGGCTGACGATCCTGTTCCGACTATGTCCACCGCGTCGTATCTGGAGACAG  
CGATCGTGATCGAAAGTCGCTTCGGCGAGCCAGGTGGTCGCGAACTCGATCTGTGGTTGCAT  
CGCGCATCAGTGGCACTGGTCGCTGTGGATGCCGATCAGGCCGACGCCGCCCGGTTGGCTTA  
CCGGAGATACGGCAAGGGGGCGCCATCGTGCAGGCCTGAATTACGGCGACTGT  
TTTTCTATGCGCTCGCCAAGGTCAGCGGTTCAGCCTTTGTTGTTCAAGGGCGAGGCTTTTCG  
GCTCACCGACGTGCGTTCACGCGGATCCGAATTCGAGCTCCGTTCGACAAGCTTGCGG  
CCGCACTCGAGCACCACCACCACCACCCTGA

MS 1284 VapC with pYUB1049 fusion tag

MVIDTSALVAILTDEPDAELLEGA VADDPVRTMSTASYLETAIVIESRFGEPPGGRELDLWLH  
RASVALVAVDADQADAARLAYRRYGKGRHRAGLNYGDCFSYALAKVSGQPLLFKGEAFRLTD  
VAAVHADPNSSSVDKLAAALEHHHHHH\*

151 amino acids

Molecular Weight: 16302.2

Theoretical pI: 5.49

MS 1283/4 VapBC Operon (668bp)

ATGGCTCTAAGCATCAAACACCCGGAAGCCGACCCGGCTGGCCAGAGAGCTAGCGGCACGTAC  
GGGGGAGACACTGACCGAGGCGGTGGTGTATGGCACTGCGGGAGCGACTGGCGCGCACTGTCC  
GCCGCACTCAGGTGCTTCGCTTCGTGAGGAACTCGCCGCGATCCGTGCTCGCTGCGCAGCC  
CTACCGGTGTTGGACGACCCGACAGCCGAATCGATCCTGGGCTACGACGACCCGCGTCTGCC  
GTCCTGATGGTTATCGACACTTCTGCCCTCGTTGCCATCTTGACCGACGAACCCGACGCCGA  
GTTGTTGGAGGGGGCGGTGGCTGACGATCCTGTTCCGACTATGTCCACCGCGTCGTATCTGG  
AGACAGCGATCGTGATCGAAAGTCGCTTCGGCGAGCCAGGTGGTCGCGAACTCGATCTGTGG  
TTGCATCGCGCATCAGTGGCACTGGTCGCTGTGGATGCCGATCAGGCCGACGCCGCCCGGTT  
GGCTTACCGGAGATACGGCAAGGGGGCGCCATCGTGCAGGCCTGAATTACGGCGACTGTTTTT  
CCTATGCGCTCGCCAAGGTCAGCGGTTCAGCCTTTGTTGTTTCAGGGCGAGGCTTTTCGGCTCA  
CCGACGTGCTGCGGTCCACTGA

MS 1283/4 VapBC Complex Protein Sequence

MALSIKHPEADRLARELAARTGETLTEAVVMALRERLARTVGRTQVVPLREELAAIRRRCAA  
LPVLDDRTAESILGYDDRGLPSMVIDTSALVAILTDEPDAELLEGA VADDPVRTMSTASYLE  
TAIVIESRFGEPPGGRELDLWLHRASVALVAVDADQADAARLAYRRYGKGRHRAGLNYGDCFS  
YALAKVSGQPLLFKGEAFRLTDVAAVH\*

213 amino acids

Molecular Weight: 23171.4

Theoretical pI: 5.47

---

---

MS 1283/4 VapBC operon with pYUB1049 fusion tag (710bp)

ATGGCTCTAAGCATCAAACACCCGGAAGCCGACCGGCTGGCCAGAGAGCTAGCGGCACGTAC  
GGGGGAGACACTGACCGAGGCGGTGGTGTATGGCACTGCGGGAGCGACTGGCGCGCACTGTCC  
GCCGCACTCAGGTTCGTTCCGCTTCGTGAGGAACTCGCCGCGATCCGTCGTCGCTGCGCAGCC  
CTACCGGTGTTGGACGACCGGACAGCCGAATCGATCCTGGGCTACGACGACCGCGGTCTGCC  
GTCCTGATGGTTATCGACACTTCTGCCCTCGTTGCCATCTTGACCGACGAACCCGACGCCGA  
GTTGTTGGAGGGGGCGGTGGCTGACGATCCTGTTCCGACTATGTCCACCGCGTTCGTATCTGG  
AGACAGCGATCGTGATCGAAAGTCGCTTCGGCGAGCCAGGTGGTCGCGAACTCGATCTGTGG  
TTGCATCGCGCATCAGTGGCACTGGTCGCTGTGGATGCCGATCAGGCCGACGCCGCCCGGTT  
GGCTTACCGGAGATACGGCAAGGGGGCGCCATCGTGCAGGCCTGAATTACGGCGACTGTTTTT  
CCTATGCGCTCGCCAAGGTCAGCGGTCAGCCTTTGTTGTTCAAGGGCGAGGCTTTTCGGCTC  
ACCGACGTCGCTGCGGTCCACCGCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCG  
CACTCGAGCACCACCACCACCACCCTGA

MS 1283/4 VapBC Complex with pYUB1049 fusion tag (His Tag on MS 1284)

MALSIKHPEADRLARELAARTGETLTEAVVMALRERLARTVGRTOVVPLREELAAIRRRCAA  
LPVLDDRTAESILGYDDRGLPSMVIDTSALVAILTDEPDAELLEGA VADDPVRTMSTASYLE  
TAIVIESRFGEPPGGRELDLWLHRASVALVAVDADQADAARLAYRRYGKGRHRAGLNYGDCFS  
YALAKVSGQPLLFKGEAFRLTDVAAVHADPNSSSVDKLAAALEHHHHHHH\*

235 amino acids

Molecular Weight: 25563.9

Theoretical pI: 5.80

---

---

**B.2. Pyrobaculum Protein Information**PAE0151 Protein Sequence With pProEX fusion tag

MSYYHHHHHDYDIPTTENLYFQGAMKLVVDASAIAALYVPEERSEQAERAVSQAQELHTLD  
LAAYEVANDLWKHARRGLLREDEASNMLEELWEFFKALKVHSYAEVLKDAFALALKHGVTVY  
DAAYVALAEKIGGKLLTLDRQLAEKFPALVTP

156 amino acids

Molecular weight: 17721.09

Theoretical pI: 5.50

PAE0152 Sequence with pET151 fusion tag

MHHHHHGGKPIPNPLGLDSTENLYQGIDPFTMSEVISIRVRRGLKKELEELGINYAEAVRKF  
LEELVARERRRALERARALREELRKKGAFPPSAELIREDRDEASR

108 amino acids

Molecular Weight: 12642.42

Theoretical pI: 9.51

PAE2754 Sequence with pProEX fusion tag

MSYYHHHHHDYDIPTTENLYFQGAMAVEYLVDASALYALAAHYDKWIKHREKLAILHLTIY  
EAGNALWKEARLGRVDWAAASRHLKVKVLSFFKVLDEPPLDEVLRVAVERGLTFYDASYAYVA  
ESSGLVLVTQDRELLAKTKGAIDVETLLVRLAAQ

158 amino acids

Molecular weight 17979.52

Theoretical pI: 6.18

PAE2755 Sequence with pET151 fusion tag

MHHHHHGGKPIPNPLGLDSTENLYQGIDPFTMSVVIISVRVRELKEKAKRLGIDIRRVVE  
RALEEEIKRREEEELAKSLEELRRALSGISEREWVEAVREARNAR

105 amino acids

Molecular Weight: 12349.10

Theoretical pI: 8.17

### 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

AGACCCAAGCTTGGTACCGGAATTCTACGAATTTTCTTTTGTATTTCCTTTCGCTTTGC  
TTCTCTTCCCTTCGGTTCGTTCGGTTTACCTTGTCTTGCCTTATCTTACTTTATCTAGAG  
GGCCCTATTCTATAGTGTACCTAAATGCTAGAGCTCGCT

#### 923 + Flanking Sequences

AGACCCAAGCTTGGTACTATCTTACTTTAGTTTCATTTAATTGTGTTGTACTCTCCTCTGCG  
TTCACTTAGCTTAACTTGGTTTGGCTTGATTTGACTTCAGTTGCGCTCTATTCTATCTAGAG  
GGCCCTATTCTATAGTGTACCTAAATGCTAGAGCTCGCT

#### 924 + Flanking Sequences

AGACCCAAGCTTGGTACCGCTCTATTCTACTGTCTGTGCATTCAATCGTTGAGTTCGATCT  
AGTCTCGTCTAACCCCTCCCCTGCTCCGCTGGTCTGGCCTCGCCTATCCTACCCATCTAGAG  
GGCCCTATTCTATAGTGTACCTAAATGCTAGAGCTCGCT

#### 925 + Flanking Sequences

AGACCCAAGCTTGGTACTATCCTACCCATTGGGCTCATCTGATCCATCCGGTCCCCTCCACT  
CGGCTATGTTATGCTGTATTGAGTCGTGTCGCGTCGAGCTGCCCTAATCCCACCTCTAGAG  
GGCCCTATTCTATAGTGTACCTAAATGCTAGAGCTCGCT

#### 926 + Flanking Sequences

AGACCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCT  
GCAGATCCTAATCCCACCTAGCGTATCGGGTCATGTAGTGCTACGTTACGGCCCCCGCCGG  
CATCATATTATATCACCCAGTGTAATGTGGTGTGAGGTTGGAGCATCACACTGGCGGCCGC  
TCGAGCATGCATCTAGAGGGCCCTATTCTATAGTGTACCTAAATGCTAGAGCTCGCT

#### 927 + Flanking Sequences

AGACCCAAGCTTGGTACGTGAGGTTGGAGTCCGACCTGGAATCTCAGCCTGACGTGCCATGC  
GGTGCGATGTCACGCCGCGCCACGGTATAGTATGGTACGGGATCCCGTCTAGAGGGCCCTAT  
TCTATAGTGTACCTAAATGCTAGAGCTCGCT

#### 932 + Flanking Sequences (932 = reverse complement to 926)

AGACCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCT  
GCAGATCTCCAACCTCACACCACATTACACTGGGGTGTATAAATATGATGCCGGGCGGGGGC  
CGTAACGTAGCACTACATGACCCGATACGCTAGGTGGGATTAGGCATCACACTGGCGGCCGC  
TCGAGCATGCATCTAGAGGGCCCTATTCTATAGTGTACCTAAATGCTAGAGCTCGCT

---

---

**Appendix C: Raw Data**

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

Contents include:

Chapter Four –VapC Inhibition of Translation

- DNA replication, transcription and translation pulse-chase data Oct 08
- Methionine Incorporation Experiment August 2009

Chapter Five – MALDI MS RNA Oligos Data

- MS1284 MALDI results summary 20-08-10
- PAE2754 – PAE0151 MALDI results summary 11-08-10

---

---

**Appendix D: Publications**

---

## References

- AFIF, H., ALLALI, N., COUTURIER, M., & VAN MELDERN, L. (2001). The Ratio Between CcdA and CcdB Modulates the Transcriptional Repression of the *ccd* Poison-Antidote System. *Molecular Microbiology*, *41*(1), 73-82.
- AGARWAL, S., MISHRA, N. K., BHATNAGAR, S., & BHATNAGAR, R. (2010). PemK Toxin of *Bacillus anthracis* is a Ribonuclease. *Journal of Biological Chemistry*, *285*(10), 7254-7270.
- AIZENMAN, E., ENGELBERG-KULKA, H., & GLASER, G. (1996). An *Escherichia coli* Chromosomal "Addiction Molecule" Regulated by Guanosine 3',5'-Bispyrophosphate: A Model for Programmed Bacterial Cell Death. *Proceedings of the National Academy of Sciences*, *93*, 6059-6063.
- AMITAI, S., YASSIN, Y., & ENGELBERG-KULKA, H. (2004). MazF-Mediated Cell Death in *Escherichia coli*: a Point of No Return. *Journal of Bacteriology*, *186*(24), 8295-8300.
- ANANTHARAMAN, V., & ARAVIND, L. (2003). New Connections in the Prokaryotic Toxin-Antitoxin Network: Relationship with the Eukaryotic Nonsense-Mediated RNA Decay System. *Genome Biology*, *4*, R81.
- ANANTHARAMAN, V., & ARAVIND, L. (2006). The NYN Domains: Novel Predicted RNases with a PIN Domain-Like Fold. *RNA Biology*, *3*(1), 18-27.
- ARCUS, V. L., BACKBRO, K., ROOS, A., DANIEL, E. L., & BAKER, E. N. (2004). Distant Structural Homology Leads to the Functional Characterization of an Archaeal PIN Domain as an Exonuclease. *The Journal of Biological Chemistry*, *279*(16), 16471-14478.
- ARCUS, V. L., MCKENZIE, J. L., ROBSON, J., & COOK, G. M. (2011). The PIN-Domain Ribonucleases and the Prokaryotic VapBC Toxin-Antitoxin Array. *Protein Engineering, Design and Selection*, *24*(1-2), 33-40.
- ARCUS, V. L., RAINEY, P. B., & TURNER, S. J. (2005). The PIN-Domain Toxin-Antitoxin Array in Mycobacteria. *TRENds in Microbiology*, *13*(8), 360-365.
- ARNOLD, K., BORDOLI, L., KOPP, J., & SCHWEDE, T. (2006). The SWISS-MODEL Workspace: A Web-Based Environment for Protein Structure and Homology Modelling. *Bioinformatics*, *22*, 195-201.
- BANEYX, F. (1999). Recombinant Protein Expression in *Escherichia coli*. *Current Opinion in Biotechnology* *10*(5), 411-421.
- BARRIOS, A. F., ZUO, R., REN, D., & WOOD, T. K. (2006). Hha, YbaJ and OmpA Regulate *Escherichia coli* K12 Biofilm Formation and Conjugation Plasmids Abolish Motility. *Biotechnology and Bioengineering*, *93*, 188-200.
- BASHIRI, G., SQUIRE, C. J., BAKER, E. N., & MORELAND, N. (2007). Expression, Purification and Crystallisation of Native and Selenomethionine Labeled *Mycobacterium tuberculosis* FGC1 (Rv0407) using a *Mycobacterium smegmatis* Expression System. *Protein Expression & Purification*, *54*, 38-44.
- BERNARD, P., KEZDY, K. E., VAN MELDEREN, L., STEYAERT, J., WYNS, L., & PATO, M. L. (1993). The F Plasmid CcdB Protein Induces Efficient ATP-

- Dependent DNA Cleavage by Gyrase. *Journal of Molecular Biology*, 234, 534-541.
- BLACK, D., KELLY, A. J., MARDIS, M. J., & MOYED, H. S. (1991). Structure and Organisation of *hip*, an Operon That Affects Lethality Due to Inhibition of Peptidoglycan or DNA Synthesis. *Journal of Bacteriology*, 173(18), 5732-5739.
- BLOKPOEL, M. C., MURPHY, H. N., O'TOOLE, R., WILES, S., RUNN, E. S. C., STEWART, G. R., YOUNG, D. B., & ROBERTSON, B. D. (2005). Tetracycline-Inducible Gene Regulation in Mycobacteria. *Nucleic Acids Research*, 33(2), 1-7.
- BODOGAI, M., FERENCZI, S., BASHTOVYY, D., MICLEA, P., PAPP, P., & DUSHA, I. (2006). The *ntrPR* Operon of *Sinorhizobium meliloti* is Organised and Functions as a Toxin-Antitoxin Module. *Molecular Plant-Microbe Interactions*, 19(7), 811-822.
- BOE, L., GERDES, K., & MOLIN, S. (1987). Effects of Genes Exerting Growth Inhibition and Plasmid Stability on Plasmid Maintenance. *Journal of Bacteriology*, 169, 4646-4650.
- BRICKER, A. L., & BELASCO, J. G. (1999). Importance of a 5'-Stem-Loop for Longevity of *papA* mRNA in *Escherichia coli*. *Journal of Bacteriology*, 181(11), 3587-3590.
- BROWN, B. L., GRIGORIU, S., KIM, Y., ARRUDA, J. M., DAVENPORT, A., WOOD, T. K., PETI, W., & PAGE, R. (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 Pathogens*, 5(12), 1-15.
- BUDDE, P. P., DAVIS, B. M., YUAN, J., & WALDOR, M. K. (2007). Characterisation of a *higBA* Toxin-Antitoxin Locus in *Vibrio cholerae*. *Journal of Bacteriology*, 189(2), 491-500.
- BUNKER, R. D. (2005). *Structure-Function Studies of Two Archaeal PIN Domain Proteins*. University of Auckland, Auckland.
- BUNKER, R. D., MCKENZIE, J. L., BAKER, E. N., & ARCUS, V. L. (2008). Crystal Structure of PAE0151 from *Pyrobaculum aerophilum*, a PIN-domain (VapC) Protein from a Toxin-Antitoxin Operon. *Proteins*, 72(1), 510-518.
- BUTS, L., LAH, J., DAO-THI, M.-H., WYNS, L., & LORIS, R. (2005). Toxin-Antitoxin Modules as Bacterial Metabolic Stress Managers. *Trends in Biochemical Sciences*, 30(12), 672-679.
- CAMACHO, A. G., MISSELWITZ, R., BEHLKE, J., AYORA, S., WELFLE, K., MEINHART, A., LARA, B., SAENGER, W., WELFLE, H., & ALONSO, J. C. (2002). *In Vitro* and *In Vivo* Stability of the Epsilon2 Zeta2 Protein Complex of the Broad-Host Range *Streptococcus pyogenes* pSM1905 Addiction System. *Journal of Biological Chemistry*, 383, 1701-1713.
- CHERNY, I., ROCKAH, L., & GAZIT, E. (2005). The YoeB Toxin is a Folded Protein That Forms a Physical Complex with the Unfolded YefM Antitoxin. *The Journal of Biochemistry*, 280(34), 30063-30072.
- CHRISTENSEN-DALSGAARD, M., & GERDES, K. (2006). Two *higBA* loci in the *Vibrio cholerae* Superintegron Encode mRNA Cleaving Enzymes and can Stabilize Plasmids. *Molecular Microbiology*, 62(2), 397-411.

- CHRISTENSEN-DALSGAARD, M., JORGENSEN, M. G., & GERDES, K. (2010). Three New RelE-Homologous mRNA interferases of *Escherichia coli* Differentially Induced by Environmental Stresses. *Molecular Microbiology*, 75(2), 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*, 48(5), 1389-1400.
- CHRISTENSEN, S. K., & GERDES, K. (2004). Delayed-relaxed Response Explained by Hyperactivation of RelE. *Molecular Microbiology*, 53(2), 587-597.
- CHRISTENSEN, S. K., MAENHAUT-MICHEL, G., MINE, N., GOTTESMAN, S., GERDES, K., & VAN MELDERN, L. (2004). Overproduction of the Lon protease triggers inhibition of translation in *Escherichia coli*: involvement of the *yefM-yoeB* toxin-antitoxin system. *Molecular Microbiology*, 51(6), 1705-1717.
- CHRISTENSEN, S. K., MIKKELSEN, M., PEDERSEN, K., & GERDES, K. (2001). RelE, a global inhibitor of translation is activated during nutritional stress. *Proceedings of the National Academy of Sciences*, 98(25), 14328-14333.
- CHRISTENSEN, S. K., PEDERSEN, K., HANSEN, F. G., & GERDES, K. (2003). Toxin-Antitoxin Loci as Stress-Response-Elements: ChpAK/MazF and ChpBK Cleave Translated RNAs and are Counteracted by tmRNA. *Journal of Molecular Biology*, 332(4), 809-819.
- CLISSOLD, P. M., & PONTING, C. P. (2000). PIN Domains in Nonsense-Mediated mRNA Decay and RNAi. *Current Biology*, 10(24), R888-R890.
- CONDON, C. (2003). RNA Processing and Degradation in *Bacillus subtilis*. *Microbiology and Molecular Biology Reviews*, 67(2), 157-174.
- CONNOLLY, L. E., EDELSTEIN, P. H., & RAMAKRISHNAN, L. (2007). Why Is Long-Term Therapy Required to Cure Tuberculosis? *PLoS Medicine*, 4(3), 435-442.
- COOPER, C. R., DAUGHERTY, A. J., TACHDIJAN, S., BLUM, P. H., & KELLY, R. M. (2009). Role of *vapBC* Toxin-Antitoxin Loci in the Thermal Stress Response of *Sulfolobus solfataricus*. *Biochemical Society Transactions*, 37, 123-126.
- COOPER, T. F., & HEINEMANN, J. A. (2000). Postsegregational Killing Does Not Increase Plasmid Stability but Acts to Mediate the Exclusion of Competing Plasmids. *Proceedings of the National Academy of Sciences*, 97, 12643-12648.
- CORREIA, F. F., D'ONOFRIO, A., REJTAR, T., LI, L., KARGER, B. L., MAKAROVA, K., KOONIN, E. V., & LEWIS, K. (2006). Kinase Activity of Overexpressed HipA Is Required for Growth Arrest and Multidrug Tolerance in *Escherichia coli*. *J. Bacteriol.*, 188(24), 8360-8367.
- COUTURIER, M., BAHASSI, E. M., & VAN MELDEREN, L. (1998). Bacterial Death by DNA Gyrase Poisoning. *TRENds in Microbiology*, 6, 269-275.
- CROOKS, G. E., HON, G., CHANDONIA, J. M., & BRENNER, S. E. (2004). WebLogo: A sequence logo generator. *Genome Research*, 14, 1188-1190.

- DAINES, D. A., WU, M. H., & YUAN, S. Y. (2007). VapC-1 of Nontypeable *Haemophilus influenzae* is a Ribonuclease. *Journal of Bacteriology*, 189(14), 5041-5048.
- DAVIS, T. L., HELINISKI, D. R., & ROBERTS, R. C. (1992). Transcription and autoregulation of the stabilizing functions of broad-host-range plasmid RK2 in *Escherichia coli*, *Agrobacterium tumefaciens* and *Pseudomonas aeruginosa*. *Molecular Microbiology*, 6, 1984-1994.
- DE BERNARDEZ CLARK, E. (1998). Refolding of recombinant proteins. *Current Opinion in Biotechnology*, 9(2), 157-163.
- DOA-THI, M.-H., VAN MELDERN, L., DE GENST, E., AFIF, H., BUTS, L., WYNS, L., & LORIS, R. (2005). Molecular Basis of Gyrase Poisoning by the Addiction Toxin CcdB. *Journal of Molecular Biology*, 348, 1091-1102.
- EBERL, L., GIVSKOV, M., & SCHWAB, H. (1992). The Divergent Promoters Mediating Transcription of the par Locus of Plasmid RP4 are Subject to Autoregulation. *Molecular Microbiology*, 6, 1969-1979.
- ENGELBERG-KULKA, H., AMITAI, S., KOLODKIN-GAL, I., & HAZAN, R. (2006). Bacterial Programmed Cell Death and Multicellular Behaviour in Bacteria. *PLoS Genetics*, 2(10), e135.
- FATICA, A., TOLLERVEY, D., & DLAKIC, M. (2004). PIN Domain of Nob1p is Required for D-site Cleavage in 20s Pre-rRNA. *RNA*, 10, 1698-1701.
- FINERAN, P. C., BLOWER, T. R., FOULDS, I. J., HUMPHREYS, D. P., LILLEY, K. S., & SALMOND, G. P. C. (2009). The Phage Abortive Infection System, ToxIN, Functions as a Protein-RNA Toxin-Antitoxin Pair. *Proceedings of the National Academy of Sciences*, 106(3), 894-899.
- FINN, R. D., MISTRY, J., TATE, J., COGGILL, P., HEGER, A., POLLINGTON, J. E., GAVIN, O. L., GUNASEKARAN, P., CERIC, G., FORSLUND, K., HOLM, L., SONNHAMMER, E. L. L., EDDY, S. R., & BATEMAN, A. (2010). The Pfam Protein Families Database. *Nucleic Acids Research*, 38, D211-D222.
- FIVIAN-HUGHES, A. S., & DAVIS, E. O. (2010). Analysing the Regulatory Role of the HigA Antitoxin within *Mycobacterium tuberculosis*. *Journal of Bacteriology*, 192(17), 4348-4356.
- FOZO, E. M., HEMM, M. R., & STORZ, G. (2008). Small Toxic Proteins and the Antisense RNAs That Repress Them. *Microbiology and Molecular Biology Reviews*, 72(4), 579-589.
- FRANCUSKI, D., & SAENGER, W. (2009). Crystal Structure of the Antitoxin-Toxin Protein Complex RelB-RelE from *Methanococcus jannaschii*. *Journal of Molecular Biology*, 393(4), 898-908.
- FU, Z., DONEGAN, N. P., MEMMI, G., & CHEUNG, A. L. (2007). Characterisation of MazFSa, an Endoribonuclease from *Staphylococcus aureus*. *Journal of Bacteriology*, 189(24), 8871-8879.
- GARCIA-CONTRERAS, R., ZHANG, X.-S., KIM, Y., & WOOD, T. K. (2008). Protein Translation and Cell Death: The Role of Rare tRNAs in Biofilm Formation and in Activating Dormant Phage Killer Genes. *PLoS One*, 3(6), 1-15.
- GARCIA-PINO, A., BALASUBRAMANIAN, S., WYNS, L., GAZIT, E., DE GREVE, H., MAGNUSON, R. D., CHARLIER, D., VAN NULAND, N. A.

- J., & LORIS, R. (2010). Allosteric and Intrinsic Disorder Mediate Transcription Regulation by Conditional Cooperativity. *Cell*, 142, 101-111.
- GARCIA-PINO, A., CHRISTENSEN-DALSGAARD, M., WYNS, L., YARMOLINSKY, M., MAGNUSON, R. D., GERDES, K., & LORIS, R. (2008). Doc of Prophage P1 is Inhibited by its Antitoxin Partner Phd through Fold Complementation. *Journal of Biological Chemistry*, 283(45), 30821-30827.
- GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAND, S., WILKINS, M. R., APPEL, R. D., & BAIROCH, A. (Eds.). (2005). *Protein Identification and Analysis on the ExPASy Server*. Totowa: Humana Press.
- GAZIT, E., & SAUER, R. T. (1999). The Doc Toxin and Phd Antidote Proteins of the Bacteriophage P1 Plasmid Addiction System Form a Heterotrimeric Complex. *Journal of Biological Chemistry*, 274, 16813-16818.
- GERDES, K. (2000). Toxin-Antitoxin Modules May Regulate Synthesis of Macromolecules During Nutritional Stress. *Journal of Bacteriology*, 182(3), 561-572.
- GERDES, K. (2007). Toxin – Antitoxin (TA) Genes from Prokaryotic Cells.
- GERDES, K., CHRISTENSEN, S. K., & LOBNER-OLSEN, A. (2005). Prokaryotic Toxin-Antitoxin Stress Response Loci. *Nature Reviews Microbiology*, 3, 371-382.
- GERDES, K., GULTYAEV, A. P., FRANCH, T., PEDERSEN, K., & MIKKELSEN, N., D. (1997). Antisense RNA-Regulated Programmed Cell Death. *Annual Reviews in Genetics*, 31, 1-31.
- GERDES, K., RASMUSSEN, P. B., & MOLIN, S. (1986). Unique Type of Plasmid Maintenance Function: Postsegregational Killing of Plasmid Free Cells. *Proceedings of the National Academy of Sciences*, 83(10), 3116-3120.
- GLAVAN, F., BEHM-ANSMANT, I., IZAURRALDE, E., & CONTI, E. (2006). Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. *The EMBO Journal*, 1-9.
- GOTFREDSSEN, M., & GERDES, K. (1998). The *Escherichia coli relBE* Genes belong to a New Toxin-Antitoxin Gene Family. *Molecular Microbiology*, 29(4), 1065-1076.
- GUPTA, A. (2009). Killing Activity and Rescue Function of Genome-Wide Toxin-Antitoxin Loci of *Mycobacterium tuberculosis*. *FEMS Microbiology Letters*, 290, 45-53.
- HAYES, F. (2003). Toxins-Antitoxins: Plasmid Maintenance, Programmed Cell Death, and Cell Cycle Arrest. *Science*, 301, 1496-1499.
- HAZAN, R., & ENGELBERG-KULKA, H. (2004). *Escherichia coli mazEF*-Mediated Cell Death as a Defense Mechanism that Inhibits the Spread of Phage P1. *Molecular Genetics & Genomics*, 272, 227-234.
- HAZAN, R., SAT, B., & ENGELBERG-KULKA, H. (2004). *Escherichia coli mazEF*-Mediated Cell Death is Triggered by Various Stressful Conditions. *Journal of Bacteriology*, 186(11), 3663-3669.
- HAZAN, R., SAT, B., RECHES, M., & ENGELBERG-KULKA, H. (2001). Postsegregational Killing Mediated by the P1 Phage "Addiction Molecule" *phd-doc* Requires the *Escherichia coli* Programmed Cell Death System *mazEF*. *Journal of Bacteriology*, 183(2046-2050).

- HOWARD, B., & KEISSER, J. (1994). *Clinical and Pathogenic Microbiology*. St Louis: Mosby.
- HUANG, F., & HE, Z.-G. (2010). Characterisation of an interplay between a *Mycobacterium tuberculosis* MazF homolog, Rv1495 and its sole DNA topoisomerase I. *Nucleic Acids Research, Advance Access* 1-12.
- HUNTZINGER, E., KASHIMA, I., FAUSER, M., SAULIERE, J., & IZAURRALDE, E. (2008). SMG6 is the Catalytic Endonuclease that Cleaves mRNAs Containing Nonsense Codons in Metazoa. *RNA*, *14*, 2609-2617.
- 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*, *284*(28), 18605-18613.
- IGUCHI, A., THOMSON, N. R., OGURA, Y., SAUNDERS, D., OOKA, T., HENDERSON, I. R., HARRIS, D., ASADULGHANI, M., KUROKAWA, K., DEAN, P., KENNY, B., QUAIL, M. A., THURSTON, S., DOUGAN, G., HAYASHI, T., PARKHILL, J., & FRANKEL, G. (2009). Complete Genome Sequence and Comparative Genome Analysis of Enteropathogenic *Escherichia coli* O127:H6 Strain E2348/69. *Journal of Bacteriology*, *191*(1), 347-354.
- JIANG, Y., POGLIANO, J., HELINISKI, D. R., & KONIECZNY, I. (2002). ParE Toxin Encoded by the Broad-Host-Range Plasmid RK2 is an Inhibitor of *Escherichia coli* gyrase. *Molecular Microbiology*, *44*(4), 971-979.
- JOHNSON, E. P., STROM, A. R., & HELINISKI, D. R. (1996). Plasmid RK2 Toxin Protein ParE: Purification and Interaction with the ParD Antitoxin Protein. *Journal of Bacteriology*, *178*(5), 1420-1429.
- JORGENSEN, M. G., PANDEY, D. P., JASKOLSKA, M., & GERDES, K. (2009). HicA of *Escherichia coli* Defines a Novel Family of Translation-Independent mRNA Interferases in Bacteria and Archaea. *Journal of Bacteriology*, *191*(4), 1191-1199.
- KAMADA, K., & HANAOKA, F. (2005). Conformational Change in the Catalytic Site of the Ribonuclease YoeB Toxin by YefM Antitoxin. *Molecular Cell*, *19*(4), 497-509.
- KAMADA, K., HANAOKA, F., & BURLEY, S. K. (2003). Crystal Structure of the MazE/MazF Complex: Molecular Basis of Antidote-Toxin Recognition. *Molecular Cell*, *11*(4), 875-884.
- KAMPHUIS, M. B., BONVIN, A. M. J., MONTI, M. C., LEMONNIER, M., MUNOZ-GOMEZ, A. J., VAN DEN HEUVEL, R. H. H., DIAZ-OREJAS, R., & BOELEN, R. (2006). Model for RNA Binding and the Catalytic Site of the RNase Kid of the Bacterial *parD* Toxin-Antitoxin System. *Journal of Molecular Biology*, *357*, 115-126.
- KAPUST, R. B., & WAUGH, D. S. (1999). *Escherichia coli* Maltose-Binding Protein is Uncommonly Effective at Promoting the Solubility of Polypeptides to which it is Fused. *Protein Science*, *8*(8), 1668-1674.
- KEDZIERSKA, B., LIAN, L.-Y., & HAYES, F. (2007). Toxin-Antitoxin Regulation: Bimodal Interaction of YefM-YoeB with Paired DNA Palindromes Exerts Transcriptional Autorepression. *Nucleic Acids Research*, *35*(1), 325-339.

- KIEFER, F., ARNOLD, K., KUNZLI, M., BORDOLI, L., & SCHWEDE, T. (2009). The SWISS-MODEL Repository and Associated Resources. *Nucleic Acids Research*, 37, D387-D392.
- KIM, Y., & WOOD, T. K. (2009). 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*, 391(1), 209-213.
- KOLODKIN-GAL, I., & ENGELBERG-KULKA, H. (2006). Induction of *Escherichia coli* Chromosomal *mazEF* by Stressful Conditions Causes an Irreversible Loss of Viability. *Journal of Bacteriology*, 188(9), 3420-3423.
- KOLODKIN-GAL, I., & ENGELBERG-KULKA, H. (2008). The Extra-Cellular Death Factor (EDF): Physiological and Genetic Factors Influencing its Production and Response in *Escherichia coli*. *Journal of Bacteriology*, 190(9), 169-175.
- KOLODKIN-GAL, I., & ENGELBERG-KULKA, H. (2009). The Stationary Sigma Factor is Responsible for the Resistance of *Escherichia coli* Stationary Cells to *mazEF*-mediated Cell Death. *Journal of Bacteriology*, 191(9), 3177-3182.
- KOLODKIN-GAL, I., HAZAN, R., GAATHON, A., CARMELI, S., & ENGELBERG-KULKA, H. (2007). A Linear Pentapeptide Is a Quorum-Sensing Factor Required for *mazEF*-Mediated Cell Death in *Escherichia coli*. *Science*, 318(5850), 652-655.
- KORCH, S. B., CONTRERAS, H., & CLARK-CURTISS, J. E. (2009). Three *Mycobacterium tuberculosis* Rel Toxin-Antitoxin Modules Inhibit Mycobacterial Growth and Are Expressed in Infected Human Macrophages. *Journal of Bacteriology*, 191(5), 1618-1630.
- 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*, 188, 3826-3836.
- KUSHNER, S. R. (2002). mRNA Decay in *Escherichia coli* Comes of Age. *Journal of Bacteriology*, 184(17), 4658-4665.
- KWAN, A. H., CZOLIJ, R., MACKAY, J. P., & CROSSLEY, M. (2003). Pentaprobe: A comprehensive Sequence for the One-Step Detection of DNA-Binding Activities. *Nucleic Acids Research*, 31(20), 1-11.
- LAMANNA, A. C., & KARBSTEIN, K. (2009). Nob1 Binds the Single-Stranded Cleavage Site D at the 3'End of 18s rRNA with its PIN Domain. *Proceedings of the National Academy of Sciences*, 106(34), 14259-14264.
- LEHNHERR, H., & YARMOLINSKY, M. B. (1995). Addiction Protein PhD of Plasmid Prophage P1 is a Substrate of the ClpXP Serine Protease of *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 92, 3274-3277.
- LEWIS, K. (2000). Programmed Death in Bacteria. *Microbiology & Molecular Biology Reviews*, 64(503-514).
- LEWIS, K. (2007). Persister Cells, Dormancy and Infectious Disease. [Review]. *Nature Reviews Microbiology*, 5, 48-56.
- LI, G.-Y., ZHANG, Y., INOUE, M., & IKURA, M. (2009). Inhibitory Mechanism of *Escherichia coli* RelE-RelB Toxin-Antitoxin Module Involves a Helix Displacement Near an mRNA Interferase Active Site. *Journal of Biological Chemistry*, 284, 14628-14636.

- LIU, M., ZHANG, Y., INOUE, M., & WOYCHIK, N. A. (2008). Bacterial Addiction Molecule Toxin Doc Inhibits Translation Elongation Through its Association with the 30S Ribosomal Subunit. *Proceedings of the National Academy of Sciences*, 105(15), 5885-5890.
- MAGNUSON, R. D. (2007). Hypothetical Functions of Toxin-Antitoxin Systems. [Guest Commentary]. *Journal of Bacteriology*, 189(17), 6089-6092.
- MAKAROVA, K. S., GRISHIN, N. V., & KOONIN, E. V. (2006). The HicAB Cassette, A Putative Novel, RNA-Targeting Toxin-Antitoxin System in Archaea and Bacteria. *Bioinformatics*, 22(21), 2581-2584.
- MAKAROVA, K. S., WOLF, Y. I., & KOONIN, E. V. (2009). Comprehensive Analysis of Type 2 Toxin-Antitoxin Systems and Related Mobile Stress Response Systems in Prokaryotes. *Biology Direct*, 4(19).
- MAKI, S., TAKIGUCHI, S., MIKI, T., & HORIUCHI, T. (1992). Modulation of DNA Supercoiling activity of *Escherichia coli* DNA Gyrase by F Plasmid Proteins: Antagonistic Actions of LetA (CcdA) and LetD (CcdB) Proteins. *Journal of Biological Chemistry*, 267, 12244-12251.
- MATHEWS, D. (2006). RNA Secondary Structure Analysis Using RNAstructure. *Current Protocols in Bioinformatics*, 13, 12.16.11 - 12.16.14.
- MATTISON, K., WILBUR, J. S., SO, M., & BRENNAN, R. G. (2006). Structure of FitAB from *Neisseria gonorrhoeae* Bound to DNA Reveals a Tetramer of Toxin-Antitoxin Heterodimers Containing PIN Domains and Ribbon-Helix-Helix motifs. *Journal of Biological Chemistry*, 281(49), 37942-37951.
- MEINHART, A., ALONSO, J. C., STRATER, N., & SAENGER, W. (2003). Crystal Structure of the Plasmid Maintenance System Epsilon/Zeta: Functional Mechanism of Toxin Zeta and Inactivation of Epsilon 2 Zeta 2 Complex Formation. *Proceedings of the National Academy of Sciences*, 100, 1661-1666.
- MEYERS, P. R., BOURN, W. R., STEYN, L. M., VAN HELDEN, P. D., BEYERS, A. D., & BROWN, G. D. (1998). Novel Method for Rapid Measurement of Growth of Mycobacteria in Detergent-Free Media. *Journal of Clinical Microbiology*, 36(9), 2752-2754.
- MIALLAU, L., FALLER, M., CHIANG, J., ARBING, M., GUO, F., CASCIO, D., & EISENBERG, D. (2009). Structure and Proposed Activity of a Member of the VapBC Family of Toxin-Antitoxin Systems. VapBC-5 from *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, 284(1), 276-283.
- MINE, N., GUGLIELMINI, J., WILBAUX, M., & VAN MELDEREN, L. (2009). The Decay of the Chromosomally-Encoded *ccd0157* Toxin-Antitoxin System in the *Escherichia coli* Species. *Genetics*, 181, 1557-1566.
- MITTENHUBER, G. (1999). Occurrence of MazEF-like Antitoxin/Toxin Systems in Bacteria. *Journal of Molecular Microbiology and Biotechnology*, 1(2), 295-302.
- MONTI, M. C., HERNANDEZ-ARRIAGA, A. M., KAMPHUIS, M. B., LOPEZ-VILLAREJO, J., HECK, A. J. R., BOELENS, R., DIAZ-OREJAS, R., & VAN DEN HEUVEL, R. H. H. (2007). Interactions of the Kid-Kis Toxin-Antitoxin Complexes With the *parD* Operator-Promoter Region of Plasmid R1 are Piloted by the Kis Antitoxin and Tuned by the Stoichiometry of the Kid-Kis Oligomers. *Nucleic Acids Research*, 35(5), 1737-1749.

- MORELAND, N., ASHTON, R., BAKER, H. M., IVANOVIC, I., ARCUS, V. L., BAKER, E. N., & LOTT, J. S. (2005). A Flexible and Economical Medium-Throughput Strategy for Protein Production and Crystallisation. *Acta Crystallographica, D61*, 1378-1385.
- MOTIEJUNAITE, R., ARMALYTE, J., MARKUCKAS, A., & SUZIEDELIENE, E. (2007). *Eshcherichia coli* *dinJ-yafQ* Genes Act as a Toxin-Antitoxin Module. *FEMS Microbiology Letters*, 268, 112-119.
- MUNOZ-GOMEZ, A. J., LEMONNIER, M., SANTOS-SIERRA, S., BERZAL-HERRANZ, A., & DIAZ-OREJAS, R. (2005). RNase/Anti-RNase Activities of the Bacterial *parD* Toxin-Antitoxin System. *Journal of Bacteriology*, 187(9), 3151-3157.
- MUNOZ-GOMEZ, A. J., SANTOS-SIERRA, S., BERZAL-HERRANZ, A., LEMONNIER, M., & DIAZ-OREJAS, R. (2004). Insights into the Specificity of RNA Cleavage by the *Escherichia coli* MazF Toxin. *FEBS Letters*, 567(2-3), 316-320.
- MURRAY, P. R., & ROSENTHAL, K. S. (1998). *Medical Microbiology*. St Louis: Mosby.
- NARIYA, H., & INOUE, M. (2008). MazF, an mRNA Interferase Mediates Programmed Cell Death during Multicellular *Myxococcus* Development. *Cell*, 132, 55-66.
- NEUBAUER, C., GAO, Y.-G., ANDERSEN, K. R., DUNHAM, C. M., KELLEY, A. C., HENTSCHEL, J., GERDES, K., RAMAKRISHNAN, V., & BRODERSEN, D. R. (2009). The Structural Basis for mRNA Recognition and Cleavage by the Ribosome-Dependent Endonuclease RelE. *Cell*, 139, 1084-1095.
- OLAH, B., KISS, E., GYORGYPAL, Z., BORZI, J., CINEGE, G., CASNADI, G., BATUT, J., KONDOROSI, A., & DUSHA, I. (2001). Mutation in the *ntnR* Gene, a Member of the *vap* Gene Family, Increases the Symbiotic Efficiency of *Sinorhizobium meliloti*. *Molecular Plant-Microbe Interactions*, 14(7), 887-894.
- OVERGAARD, M., BORCH, J., & GERDES, K. (2009). RelB and RelE of *Escherichia coli* Form a Tight Complex That Represses Transcription via the Ribbon-Helix-Helix Motif in RelB. *Journal of Molecular Biology*, 73, 790-800.
- OVERGAARD, M., BORCH, J., JORGENSEN, M. G., & GERDES, K. (2008). Messenger RNA Interferase RelE Controls *relBE* Transcription by Conditional Cooperativity. *Molecular Microbiology*, 69(4), 841-857.
- PANDEY, D. P., & GERDES, K. (2005). Toxin-Antitoxin Loci are Highly Abundant in Free-Living but Lost from Host-Associated Parasites. *Nucleic Acids Research*, 33(3), 966-976.
- PEDERSEN, K., CHRISTENSEN, S. K., & GERDES, K. (2002). Rapid Induction and Reversal of a Bacteriostatic Condition by Controlled Expression of Toxins and Antitoxins. *Molecular Microbiology*, 45(2), 501-510.
- PEDERSEN, K., ZAVIALOV, A. V., PAVLOV, M. Y., ELF, J., GERDES, K., & EHRENBERG, M. (2003). The Bacterial Toxin RelE Displays Codon-Specific Cleavage of mRNAs in the Ribosomal A Site. *Cell*, 112(1), 131-140.

- PELLEGRINI, O., MATHY, N., GOGOS, A., SHAPIRO, L., & CONDON, C. (2005). The *Bacillus subtilis* *ydcDE* Operon Encodes an Endoribonuclease of the MazF/PemK Family and its Inhibitor. *Molecular Microbiology*, *56*, 1139-1148.
- PIMENTEL, B., MADINE, M. A., & DE LA CUEVA-MENDEZ, G. (2005). Kid cleaves specific mRNAs at UUACU sites to rescue the copy number of plasmid R1. *EMBO J*, *24*(19), 3459-3469.
- PRYSKAK, M. H., MOZDZIERZ, C., J., COOK, A. M., ZHU, L., ZHANG, Y., INOUE, M., & WOYCHIK, N. A. (2009). Bacterial Toxin YafQ is an Endoribonuclease that Associates with the Ribosome and Blocks Translation Elongation Through Sequence-Specific and Frame-Dependent mRNA Cleavage. *Molecular Microbiology*, *71*(5), 1071-1087.
- PUSKAS, L. G., NAGY, Z. B., KELEMAN, J. Z., RUBERG, S., BODOGAI, M., BECKER, A., & DUSHA, I. (2004). Wide-range Transcriptional Modulating Effect Of *ntrR* Under Microaerobiosis in *Sinorhizobium meliloti*. *Molecular Genetics & Genomics*, *272*, 275-289.
- RALSTON, A. (2008). Operons and Prokaryotic Gene Regulation. *Nature Education*, *1*(1).
- RAMAGE, H. R., CONNOLLY, L. E., & COX, J. S. (2009). Comprehensive Functional Analysis of *Mycobacterium tuberculosis* Toxin-Antitoxin Systems: Implications for Pathogenesis, Stress Responses, and Evolution. *PLoS Genetics*, *5*(12), 1-14.
- ROBSON, J. (2010). *Characterisation of the vapBC Toxin-Antitoxin Operon of Mycobacterium smegmatis*. University of Otago, Dunedin.
- ROBSON, J., MCKENZIE, J. L., CURSONS, R., COOK, G. M., & ARCUS, V. L. (2009). The *vapBC* Operon from *Mycobacterium smegmatis* is an Autoregulated Toxin-Antitoxin Module that Controls Growth via Inhibition of Translation. *Journal of Molecular Biology*, *390*, 353-367.
- ROTEM, E., LOINGER, A., RONIN, I., LEVIN-REISMAN, I., GABAY, C., SHORESH, N., BIHAM, O., & BALABAN, N. Q. (2010). Regulation of Phenotypic Variability by a Threshold-Based Mechanism Underlies Bacterial Persistence. *Proceedings of the National Academy of Sciences*, *107*(28), 12541-12546.
- ROWE-MAGNUS, D. A., GUEROUT, A. M., BISKRI, L., BOUIGE, P., & MAZEL, D. (2003). Comparative Analysis of Superintegrons: Engineering Extensive Genetic Diversity in the Vibrionaceae. *Genome Research*, *13*, 428-442.
- ROZENSKI, J. (1999). Mongo Oligonucleotide Calculator. 2.06.
- RUIZ-ECHEVARRIA, M. J., GIMENEZ-GALLEGO, G., SABARIEGOS-JARENO, R., & DIAZ-OREJAS, R. (1991). The *kis* and *kid* genes of the ParD Maintenance System of Plasmid R1 Form an Operon that is Autoregulated at the Level of Transcription by the Co-Ordinated Action of the Kis and Kid Proteins. *Molecular Microbiology*, *5*, 2685-2693.
- RUIZ-ECHEVARRIA, M. J., GIMENEZ-GALLEGO, G., SABARIEGOS-JARENO, R., & DIAZ-OREJAS, R. (1995). Kid, a Small Protein of the *parD* Stability System of Plasmid R1, is an Inhibitor of DNA Replication Acting at the Initiation of DNA Synthesis. *Journal of Molecular Biology*, *247*, 568-577.

- RUSSELL, J., & COOK, G. (1995). Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol. Rev.*, 59(1), 48-62.
- SAAVEDRA DE BAST, M., MINE, N., & VAN MELDEREN, L. (2008). Chromosomal Toxin-Antitoxin Systems May Act as Antiaddiction Modules. *J. Bacteriol.*, 190(13), 4603-4609.
- SACHDEV, D., & CHIRGWIN, J. M. (1998). Order of Fusions Between Bacterial and Mammalian Proteins can Determine Solubility in *Escherichia coli*. *Biophysical and Biochemical Research Communications*, 244, 933-937.
- SAIDA, F., UZAN, M., ODAERT, B., & BONTEMS, F. (2006). Expression of Highly Toxic Genes in *Escherichia coli*: Special Strategies and Genetic Tools *Current Protein & Peptide Science*, 7(1), 47-56.
- SAT, B., HAZAN, R., FISHER, T., KHANER, H., GLASER, G., & KULKA-ENGELBERG, H. (2001). Programmed Cell Death in *Escherichia coli*: Some Antibiotics and Trigger *mazEF* Lethality. *Journal of Bacteriology*, 183(6), 2041-2045.
- SCHERRER, R., & MOYED, H. S. (1988). Conditional impairment of cell division and altered lethality in *hipA* mutants of *Escherichia coli* K-12. *Journal of Bacteriology*, 170(8), 3321-3326.
- SCHNEIDER, T. D., & STEPHENS, R. M. (1990). Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Research*, 18, 6097-6100.
- SCHUMACHER, M. A., PIRO, K. M., XU, W., HANSEN, S., LEWIS, K., & BRENNAN, R. G. (2009). Molecular Mechanisms of HipA-Mediated Multidrug Tolerance and Its Neutralization by HipB. *Science*, 323, 396-401.
- SEVIN, E. W., & BARLOY-HUBLER, F. (2007). RASTA-Bacteria: A Web-Based Tool for Identifying Toxin-Antitoxin Loci in Prokaryotes. *Genome Biology*, 8(8), R155.
- SHINOHARA, M., GUO, J. X., MORI, M., NAKASHIMA, T., TAKAGI, H., NISHIMOTO, E., YAMASHITA, S., TSUMOTO, K., KAKUTA, Y., & KIMURA, M. (2010). The Structural Mechanism of the Inhibition of Archaeal RelE Toxin by its Cognate RelB Antitoxin. *Biochemical and Biophysical Research Communications*, 400(3), 346-351.
- SIMIC, M., DE JONGE, N., LORIS, R., VESNAVER, G., & LAH, J. (2009). Driving Forces of Gyrase Recognition by the Addiction Toxin CcdB. *Journal of Biological Chemistry*, 284(30), 20002-20010.
- SNAPPER, S. B., MELTON, R. E., MUSTAFA, S., KIESER, T., & JACOBS, W. R. J. (1990). Isolation and Characterisation of Efficient Plasmids Transformation Mutants of *Mycobacterium smegmatis*. *Molecular Microbiology*, 4(11), 1911-1919.
- STEEGE, D. A. (2000). Emerging Features of mRNA Decay in Bacteria. *RNA*, 6, 1078-1090.
- SUMMERS, E. (2007). Expression of VapBC from *Mycobacterium smegmatis* in *Escherichia coli*. University of Waikato.
- SUZUKI, M., ZHANG, M., LIU, M., WOYCHIK, N. A., & INOUE, M. (2005). Single Protein Production in Living Cells Facilitated by an mRNA Interferase. *Molecular Cell*, 18, 253-261.

- TADOKORO, T., & KANAYA, S. (2009). Ribonuclease H: Molecular Diversities, Substrate Binding Domains, and Catalytic Mechanism of the Prokaryotic Enzymes. *FEBS Journals*, 276(6), 1482-1493.
- TAKESHITA, D., ZENNO, S., LEE, W. C., SAIGO, K., & TANOKURA, M. (2007). Crystal Structure of the PIN Domain of Human Telomerase-Associated Protein EST1A. *Proteins*, 68, 980-989.
- TSILIBARIS, V., MAENHAUT-MICHEL, G., MINE, N., & VAN MELDERN, L. (2007). What Is the Benefit to *Escherichia coli* of Having Multiple Toxin-Antitoxin Systems in Its Genome? *Journal of Bacteriology*, 189(17), 6101-6108.
- TSUMOTO, K., EJIMA, D., KUMAGAI, I., & ARAKAWA, T. (2003). Practical Considerations in Refolding Proteins from Inclusion Bodies. *Protein Expression and Purification*, 28(1), 1-8.
- VALLEJO, L. F., & RINAS, U. (2004). Strategies for the Recovery of Active Proteins Through Refolding of Bacterial Inclusion Body Proteins. *Microbial Cell Factories*, 3(11), 1-12.
- VAN MELDEREN, L., THI, M. H. D., LEECHI, P., GOTTESMAN, S., COUTURIER, M., & MAURIZI, M. R. (1996). ATP-Dependent Degradation of CcdA by Lon Protease. Effects of Secondary Structure and Heterologous Subunit Interactions *Journal of Biological Chemistry*, 11, 1151-1157.
- WEN, J., ARAKAWA, T., & PHILO, J. S. (1996). Size-Exclusion Chromatography with On-Line Light-Scattering, Absorbance and Refractive Index Detectors for Studying Proteins and Their Interactions. *Analytical Biochemistry*, 240, 155-166.
- WHO. (2007). Tuberculosis Fact Sheet.
- WILBUR, J. S., CHIVERS, P. T., MATTISON, K., POTTER, L., BRENNAN, R. G., & SO, M. (2005). *Neisseria gonorrhoeae* FitA Interacts with FitB To Bind DNA through Its Ribbon-Helix-Helix Motif. *Biochemistry*, 44(37), 12515-12524.
- WINTHER, K. S., & GERDES, K. (2009). Ectopic Production of VapCs from *Enterobacteria* inhibits translation and *trans*-activates YoeB mRNA interferase. *Molecular Microbiology*, 72(4), 918-930.
- WONG, C., SRIDHARA, S., BARDWELL, J. C. A., & JAKOB, U. (2000). Heating Greatly Speeds Coomassie Blue Staining and Destaining. *Biotechniques*, 28, 426-432.
- WOZNIAK, R. A., & WALDOR, M. K. (2009). A Toxin-Antitoxin System Promotes the Maintenance of an Integrative Conjugative Element. *PLoS Genetics*, 5(3), 1-11.
- YAMAGUCHI, Y., & INOUE, M. (Eds.). (2009). *mRNA Interferases, Sequence-Specific Endoribonucleases from the Toxin-Antitoxin Systems* (Vol. 85). Paris: Academic Press.
- YAMAGUCHI, Y., PARKHILL, M., & INOUE, M. (2009). MqsR, A Crucial Regulator for Quorum Sensing and Biofilm Formation, is a GCU-Specific mRNA Interferase in *Escherichia coli*. *Journal of Biological Chemistry*, 284, 28746-28753.

- ZHANG, J., ZHANG, Y., & INOUE, M. (2003). Characterization of the Interactions Within the *mazEF* Addiction Module of *Escherichia coli*. *Journal of Biological Chemistry*, 278(34), 32300-32306.
- ZHANG, Y., & INOUE, M. (2009). The Inhibitory Mechanism of Protein Synthesis by YoeB, an *Escherichia coli* Toxin. *Journal of Biological Chemistry*, 284(11), 6627-6638.
- ZHANG, Y., YAMAGUCHI, Y., & INOUE, M. (2009). Characterisation of YafO, an *Escherichia coli* Toxin. *Journal of Biological Chemistry*, 285(38), 25522-25531.
- ZHANG, Y., ZHANG, J., HARA, H., KATO, I., & INOUE, M. (2005). Insights into the mRNA Cleavage Mechanism by MazF, an mRNA Interferase. *Journal of Biological Chemistry*, 280(5), 3143-3150.
- ZHANG, Y., ZHANG, J., HOEFLICH, K. P., IKURA, M., QING, G., & INOUE, M. (2003). MazF Cleaves Cellular mRNAs Specifically at ACA to Block Protein Synthesis in *Escherichia coli*. *Molecular Cell*, 12(4), 913-923.
- ZHANG, Y., ZHU, L., ZHANG, J., & INOUE, M. (2005). Characterisation of ChpBK, an mRNA Interferase from *Escherichia coli*. *Journal of Biological Chemistry*, 280(28), 26080-26088.
- ZHANG, Y. X., GUO, X. K., WU, C., BI, B., REN, S. X., WU, C. F., & ZHAO, G. P. (2004). Characterisation of a Novel Toxin-Antitoxin Module, VapBC, Encoded by *Leptospira interrogans* Chromosome. *Cell Research*, 14(3), 208-216.
- ZHU, L., INOUE, K., YOSHIZUMI, S., KOBAYASHI, H., ZHANG, Y., OUYANG, M., KATO, F., SUGAI, M., & INOUE, M. (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*, 191(18), 3248-3255.
- ZHU, L., PHADTARE, S., NARIYA, H., OUYANG, M., HUSSON, R. N., & INOUE, M. (2008). The mRNA Interferases, MazF-mt3 and MazF-mt7 from *Mycobacterium tuberculosis* Target Unique Pentad Sequences in Single-Stranded RNA. *Molecular Microbiology*, 69(3), 559-569.
- ZHU, L., ZHANG, Y., TEH, J.-S., ZHANG, J., CONNELL, N., RUBIN, H., & INOUE, M. (2006). Characterisation of mRNA Interferases from *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, 281(27), 18638-18643.
- ZIELENKIEWICZ, U., & CEGLOWSKI, P. (2005). The Toxin-Antitoxin System of the Streptococcal Plasmid pSM19035. *Journal of Bacteriology*, 187(17), 6094-6105.