

<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.

Regulation of Metabolism in Mycobacteria and Cyanobacteria

A thesis submitted in fulfilment
of the requirements for the degree
of
Doctor of Philosophy in Biological Sciences
at
The University of Waikato
By
Chelsea Joy Vickers

The University of Waikato

2015



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Abstract

Bacteria encounter changes in their environment and must adapt to these changes in order to survive. Their ability to adapt is determined by their capacity to efficiently regulate their cellular processes. The mechanisms of bacterial regulation at the transcriptional level have been investigated by structural and functional characterisation of the transcription factor AmtR from *Mycobacterium smegmatis*. *M. smegmatis* is a soil bacteria capable of utilising alternative nitrogen sources during nitrogen limitation. AmtR is an important transcription factor that regulates the cellular machinery involved in alternative nitrogen metabolism via a novel co-repressor induced mode of regulation. A second pathway that involves post-transcriptional regulation was investigated by preliminary characterisation of a previously unexplored VapBC family of Type II Toxin Antitoxin (TA) systems found in the cyanobacteria *Microcystis aeruginosa* and *Synechocystis* sp. PCC6803 (*Synechocystis*). Investigation of the biochemistry of *M. aeruginosa* VapBC systems was limited by poor expression, however a single TA system, VapBC_{MAE43230/20}, was successfully purified by incorporating the SUMO fusion protein into the expression construct. VapB_{MAE43230} and VapC_{MAE43220} are co-expressed and form a tetrameric complex that appears to be copurified with DNA. Similar to *M. aeruginosa*, only one system, VapBC_{slr1209/1210}, in *Synechocystis* was successfully purified and investigated *in vitro*. Preliminary EMSA assays indicate VapBC_{slr1209/1210} is autoregulatory and binds to the promoter region of its operon by recognising palindromic sequences. The RNA pentaprobe system demonstrates that VapC_{slr1210} has metal dependent endoribonuclease activity that can be inhibited by VapB_{slr1209}. Markerless deletion strains of five *Synechocystis* *vapBC* operons were made and growth experiments of three of the deletion strains were conducted. Growth experiments identified a reduced growth rate in two of the deletion strains (when compared to the WT strains): $\Delta vapBC_{slr1209/1210}$ and $\Delta vapBC_{ssl2138/sll1092}$. Transcriptomic analyses of both deletion strains were conducted. Preliminary data show that both VapBC systems appear to target genes involved in carbon assimilation and metabolism via both photosynthetic and heterotrophic processes. Interestingly, VapBC_{slr1209/1210} targets an operon involved in glucose transport, which is also the cellular process targeted by the only other well characterised VapBC system from *M. smegmatis* (VapBC_{1283/4}). These

Abstract

preliminary results point toward a possible similarity in biological function for VapBC systems across the prokaryotic tree.

Acknowledgements

I would firstly like to acknowledge my chief supervisor Vic Arcus. Thank you for your support and encouragement. I am very grateful to be a student of yours. I admire your passion and intellect for research and I hope we can work together again in the future. I would also like to thank my other supervisors Ian McDonald and Ray Cursons. You have both helped me immensely both with my research and support and it is really appreciated. To Judith and Emma Summers, you run a very smooth operation and it makes for an amazing research environment. Thank you for being so wonderful and great at what you do. To all the other students at staff in the Arcus Lab: thank you all for your shared enthusiasm and encouragement. In particular, I would like to thank Ali, Emma A, Abby, and Jo M for taking the time to introduce me to the elusive world of VapBCs. Jo M thank you for putting me up in Cambridge and taking on an unofficial mentoring role with me. I am grateful to have had your support and insight. Similarly, I would like to thank Jo Hobbs (Abell) for your optimism, encouragement and mentoring. You are both researchers that I admire and am grateful to have worked with.

I would also like to thank my dear friends and family. I would like to thank Vicky for being unnaturally proud of me and supporting my strange life decisions. To Eric, thank you for your unwavering understanding, love and support. I look forward to being on the same continent as you soon. To Sheree, thank you for getting it and I love you. To Daniel, I want to thank you for your joke ridden love and support and now that I will finally get a job you are going to have to find a new gag. To mum, I can not express my gratitude to you, you are a wonderful person and your unwavering supply of wine and love has seen me through. In honour of dad: ‘What if I fall? Oh, but my darling, what if you fly?’ – E.H.

Table of Contents

Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Figures	viii
List of Tables.....	xi
List of Abbreviations.....	xii
1 Introduction	1
1.1 Prokaryotic Regulation.....	1
1.1.1 Regulation of Carbon Metabolism	3
1.1.2 Regulation of Nitrogen Metabolism	6
1.2 Transcriptional Regulation.....	10
1.2.1 GlnR Transcription Factor Protein Family	11
1.2.2 The TetR Transcription Factor Protein Family	13
1.3 Post-Transcriptional Regulation.....	18
1.3.1 Toxin-Antitoxin (TA) Protein Systems.....	19
1.3.2 Type II TA Systems	20
1.3.3 VapBC TA Systems	21
1.3.4 Cyanobacteria VapBC TA Systems	28
1.4 Objectives of this Thesis	31
2 Characterisation of <i>Mycobacterium smegmatis</i> Transcription Factor AmtR	32
2.1 Introduction	32
2.2 Methods.....	36
2.2.1 DNA Manipulation and Cloning.....	36
2.2.2 Protein Expression and Purification.....	43
2.2.3 Dialysis of Protein Samples	48
2.2.4 Crystallography	48

Table of Contents

2.2.5	AmtR DNA Electrophoretic Mobility Shift Assays (EMSA).....	51
2.2.6	AmtR Signalling Molecule Identification.....	52
2.3	Results	57
2.3.1	Expression and Protein Purification of AmtR in <i>M. smegmatis</i>	57
2.3.2	AmtR X-ray Crystallographic Structures.....	58
2.3.3	Identification of AmtR Regulated Genes and Recognition Sequence	75
2.3.4	AmtR Signalling Molecules.....	81
2.4	Discussion	88
3	Preliminary Characterisation of Cyanobacteria VapBC Systems	92
3.1	Introduction.....	92
3.2	Methods.....	97
3.2.1	DNA Manipulation and Cloning.....	97
3.2.2	Protein Expression and Purification.....	104
3.2.3	Crystallisation Trials for X-ray Crystallography of <i>Synechocystis</i> <i>VapBC</i> _{slr1209/1210}	106
3.2.4	<i>VapC</i> _{slr1210} Ribonuclease Activity	107
3.2.5	<i>Synechocystis VapBC</i> _{slr1209/1210} DNA Electrophoretic Mobility Shift Assays (EMSA).....	109
3.2.6	<i>Synechocystis ΔvapBC</i> Phenotypic Characterisation	110
3.3	Results	115
3.3.1	<i>M. aeruginosa</i> NIES-843 <i>vapBC</i> Loci Identification	115
3.3.2	Identification and Cloning of Representative <i>M. aeruginosa</i> <i>VapBC</i> Systems for Expression in <i>E. coli</i>	118
3.3.3	Large Scale Expression and Protein Purification of <i>VapBC</i> _{slr1209/1210}	126
3.3.4	Preliminary <i>VapBC</i> _{slr1209/1210} DNA Electrophoretic Mobility Shift Assays (EMSA).....	127
3.3.5	Preliminary <i>VapC</i> _{slr1210} Ribonuclease (RNase) Activity	130
3.3.6	<i>Synechocystis ΔvapBC</i> Deletion Strain Construction and Characterisation	132

Table of Contents

3.4 Discussion	146
3.4.1 <i>In vitro</i> Investigation of Cyanobacteria VapBC Systems in this Study	147
3.4.2 <i>In vivo</i> Investigation of VapBC Systems	148
4 Summary Chapter.....	152
4.1 Transcriptional Regulation by <i>M. smegmatis</i> AmtR.....	152
4.2 Post-Transcriptional Regulation.....	154
4.3 Further work.....	156
4.3.1 <i>M. smegmatis</i> AmtR Transcription Factor	156
4.3.2 Cyanobacteria VapBC systems	156
References	158
Appendices.....	172
Appendix A: Primers and DNA oligonucleotides.....	172
Appendix B: Plasmids and Bacterial Strains	174
Appendix C: Buffers, Media and Solutions	176
Appendix D: Gene and Protein Information	178
Appendix E: Chapter 3 Additional Tables and Figures	183

List of Figures

Figure 1.1. Mechanism underlying inducer exclusion in enteric bacteria.	4
Figure 1.2. Carbon catabolite repression (CCR) in <i>E. coli</i>	5
Figure 1.3. Regulation of glutamine synthetase (GS) by P _{II} mediated post-translational adenylation.	8
Figure 1.4. The P _{II} mediated two component NtrBC system of enteric bacteria. ...	9
Figure 1.5. X-ray crystal structure ribbon diagram of <i>Amycolatopsis mediterranei</i> GlnR.....	11
Figure 1.6. GlnR consensus binding motif derived from the GlnR binding regions.	12
Figure 1.7. X-ray crystal structures of TetR.	13
Figure 1.8. Analysis of AmtR-GlnK interaction by native gel electrophoresis. ...	15
Figure 1.9. The AmtR box of <i>C. glutamicum</i>	16
Figure 1.10. Genomic map of AmtR regulon in <i>M. smegmatis</i>	17
Figure 1.11. Generalised schematic of the characteristics and genetic organisation of bacterial TA pairs.	21
Figure 1.12. Schematic diagram of a generalized VapBC system.	23
Figure 1.13. <i>fitAB</i> autoregulation.	24
Figure 1.14. Generalized PIN-domain structure.	25
Figure 1.15. X-ray structures showing FitA inactivation of FitB.	26
Figure 1.16. Schematic representation of the genetic structure of the <i>Synechocystis</i> VapBC10 operon.....	31
Figure 2.1. IMAC purification and corresponding SDS-PAGE analysis of AmtR.	57
Figure 2.2. SEC of AmtR and corresponding SDS-PAGE gel.	58
Figure 2.3. Crystal of <i>M. smegmatis</i> trypsin treated mercury soaked AmtR.	59
Figure 2.4. A cartoon diagram of the X-ray crystal structure of the C-terminal ligand binding domain of <i>M. smegmatis</i> AmtR.	61
Figure 2.5. Cartoon diagram of AmtR monomer showing the central triangle of the LBD made up of helices $\alpha 5$, $\alpha 6$ and $\alpha 7$	62
Figure 2.6. The <i>M. smegmatis</i> LBD.....	63

List of Figures

Figure 2.7. <i>M. smegmatis</i> AmtR non-trypsin treated crystal.	64
Figure 2.8. Cartoon diagram of the full X-ray crystallographic structure of <i>M. smegmatis</i> AmtR.	65
Figure 2.9. Alignment of the SAD and full structure of AmtR.....	66
Figure 2.10. Cartoon diagram of the AmtR dimer DBD.....	67
Figure 2.11. Electrostatic image of the AmtR dimer.	67
Figure 2.12. Alignment of the AmtR monomer with the PfmR monomer using PDBeFOLD.....	69
Figure 2.13. AmtR DBD superimposed over QacR DBD bound to DNA.	70
Figure 2.14. Comparison of (A) PfmR (Agari <i>et al</i> , 2012) and (B) AmtR binding pockets.	71
Figure 2.15. Conserved features of AmtR HTH motif.....	72
Figure 2.16. AmtR DBD domain alignment with PfmR, TetR and QacR illustrating the conserved residues Leu ₅₈ , Tyr ₅₉ , His ₆₁ and Lys ₆₅	73
Figure 2.17. Cartoon diagram of AmtR LBD showing 59 ° angle of helices $\alpha 7$ and $\alpha 8$	74
Figure 2.18. The <i>C. glutamicum</i> AmtR recognition motif identified in the upstream promoter region of the <i>M. smegmatis</i> <i>vapBC</i> operon <i>msmeg_1283/4</i>	76
Figure 2.19. Genomic and chemical schematic of the putative metabolic pathway encoded by the urea carboxylase and allophanate hydrolase (UC/AH) operon in the <i>M smegmatis</i> AmtR regulon.....	76
Figure 2.20. <i>In vitro</i> DNA binding of <i>M. smegmatis</i> AmtR to <i>msmeg_2184</i> promoter DNA.....	78
Figure 2.21. Alignment of all <i>M. smegmatis</i> putative AmtR recognition motifs used in EMSA.	79
Figure 2.22. MEME-derived consensus sequence from the upstream regions of all AmtR regulated genes.	79
Figure 2.23. Actinomycetes consensus AmtR recognition motif.	80
Figure 2.24. Protein purification and 16% SDS-PAGE analysis of AmtR, GlnK and GlnK-AMP.	81
Figure 2.25. AmtR and GlnK SEC pull down assays.	83
Figure 2.26. PAGE analysis of AmtR and GlnK SEC pull down assays.	84
Figure 2.27. Thermal stability of the AmtR protein.....	86

List of Figures

Figure 2.28. Xanthine modeled into the AmtR binding site using the ParDOCK software.	87
Figure 3.1. A visual depiction of the hidden Markov model that defines the conserved amino acids of PIN-domain proteins.	118
Figure 3.2. pOPINS+ <i>vapBC</i> _{43230/20} small scale expression trial.	122
Figure 3.3. IMAC purification and corresponding SDS-PAGE analysis of His-SUMO-VapBC _{MAE43230/20}	123
Figure 3.4. SEC of VapBC _{MAE43230/20} and corresponding SDS-PAGE gel.	124
Figure 3.5. SDS-PAGE analysis of <i>E. coli</i> _pET28b-pstI_VapBC _{slr1209/1210} small scale expression trial.	126
Figure 3.6. IMAC purification and corresponding SDS-PAGE analysis of VapBC _{slr1209/1210}	127
Figure 3.7. SEC of VapBC _{slr1209/1210} and corresponding SDS-PAGE gel.	127
Figure 3.8. Preliminary VapBC _{slr1209/1210} autoregulation analysis by EMSA.	129
Figure 3.9. Acid SEC of VapBC _{slr1209/1210} and corresponding SDS-PAGE gel. .	130
Figure 3.10. Preliminary time course assays showing ribonuclease activity of VapC _{slr1210} using the pentaprobe system.	131
Figure 3.11. Schematic of <i>Synechocystis</i> markerless Δ <i>vapBC</i> deletion strains and PCR diagnostic gels confirming deletion.	134
Figure 3.12. Growth of <i>Synechocystis</i> WT vs Δ <i>vapBC</i> strains over a 9 day period in either full light or 12 hr light-dark cycling.	136
Figure 3.13. RT-qPCR $2^{-\Delta\Delta Ct}$ fold change expression profiles of <i>glcP</i> , <i>momP</i> , and <i>slr0350</i> in Δ <i>vapBC</i> _{slr1209/1210} vs WT.....	145
Figure A.1. SDS-PAGE analysis of <i>E. coli</i> _pET28b-pstI+ <i>MICvapBC</i> small-scale expression.....	183
Figure A.2. SDS-PAGE analysis of <i>E. coli</i> _pET28b-pstI+ <i>vapC</i> small scale expression trials.....	184
Figure A.3. SDS-PAGE analysis of <i>E. coli</i> _pPROEX+ <i>vapBC</i> _{49640/50} small scale expression trial	184
Figure A.4. SDS-PAGE analysis of <i>E. coli</i> _pET28b-pstI+ <i>vapBC</i> (<i>Synechocystis</i>) small scale expression trials	185

List of Tables

Table 1.1. Summary of Type II TA families and their general mode of action....	20
Table 2.1. AmtR ligand library containing a list of molecules tested as putative AmtR ligands.....	54
Table 2.2. Data collection and refinement statistics.	60
Table 3.1. List of the 46 predicted VapBC pairs in <i>M. aeruginosa</i> (MIC_NIES_843) with NCBI nucleotide BLAST alignments.	116
Table 3.2. <i>M. aeruginosa</i> representative <i>vapBC</i> operons <i>vapBC</i> _{MAE43230/20} and <i>vapBC</i> _{MAE49640/50} for cloning, expression and purification in <i>E. coli</i> .	119
Table 3.3. Summary of <i>M. aeruginosa</i> VapBC and VapC small scale expression trials and His tag binding.	120
Table 3.4. <i>Synechocystis</i> <i>vapBC</i> operons successfully cloned into pET28b-pstI for protein expression with a C-terminal His tag and predicted protein molecular weight (MW).....	125
Table 3.5. Summary of <i>Synechocystis</i> VapBC small scale expression tests and His tag binding.	125
Table 3.6. List of genes that are differentially expressed (>2.0 x change, P<0.05) in <i>Synechocystis</i> Δ <i>vapBC</i> _{slr1209/1210} compared to WT under full light.	139
Table 3.7. List of genes that are differentially expressed in <i>Synechocystis</i> Δ <i>vapBC</i> _{slr1209/1210} compared to WT in light-dark cycling.....	140
Table 3.8. List of genes that are differentially expressed in <i>Synechocystis</i> Δ <i>vapBC</i> _{ssl2138/sll1092} compared to WT in light-dark cycling..	143
Table 3.9. List of RNA-seq determined DEG in <i>Synechocystis</i> Δ <i>vapBC</i> _{slr1209/1210} analysed by RT-qPCR.	144
Table A.1. Primers and DNA oligonucleotides used in this study.....	172
Table A.2. List of primers used for RT-qPCR.....	173
Table A.3. Plasmids and bacterial strains used in this study	174
Table A.4. General buffers and solutions used in this study.....	176
Table A.5. Media used in this study.....	177
Table A.6. <i>M. aeruginosa</i> <i>vapBC</i> operons and genes successfully cloned for protein expression and predicted protein molecular weight (MW)..	183

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
ADC	albumin dextrose catalase
Ala	alanine
Arg	Arginine
AMP	adenosine monophosphate
Amp ^r	ampicillin resistance
AmtR	ammonium transport regulator
APS	ammonium persulphate
Asp	aspartate
ATP	adenosine triphosphate
bp	base pair(s)
BLAST	basic local alignment search tool
BSA	bovine serum albumin
C	cytosine
Cas9	CRISPR associated protein 9
cAMP	cyclic adenosine monophosphate
CCR	carbon catabolite repression
CCP4	collaborative computational project
Cra	catabolite repressor-activator
CRISPR	clustered regularly interspaced short palindromic repeats
CRP	cAMP receptor protein
crRNA(s)	CRISPR RNA(s)
CsrA	carbon storage regulator protein
C-terminal	carboxyl terminus
Cyt	Cystine
DBD	DNA binding domain
DEG	Differential gene expression
DEPEC	diethyl pyruvate carbonate
DIG	digoxigenin-11-ddUTP
DMSO	dimethyl sulfoxide

List of Abbreviations

DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
doc	death on curing
EDTA	ethylene diamine tetraacetic acid (disodium salt)
EMSA	electrophoretic mobility shift assay
fit	fast intracellular trafficking
FPLC	fast protein liquid chromatography
g	times the force of gravity
G	guanine
GGA	glucosylglycerate
GITC	guanidinium isothiocyanate
Glu	glutamate
Gln	glutamine
Gly	glycine
GOGAT	glutamine:2-oxoglutarate aminotransferase
GOI	gene(s) of interest
GS	glutamine synthetase
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
HGT	horizontal gene transfer
His	histidine
His tag	poly-histidine tag
HPLC	high-performance liquid chromatography
HTH	helix-turn-helix
Hyg ^r	hygromycin resistance
IMAC	immobilised metal affinity chromatography
IPTG	isopropylthio- β -D-galactosidase
IR	inverted repeat
Kan ^r	kanamycin resistance
kb	kilobase
kDa	kilodalton
LB	Luria Bertani
LBD	ligand binding domain
Leu	leucine
Lys	lysine

List of Abbreviations

MALDI	matrix assisted laser desorption ionisation
mAu	milli-absorbance units
Met	methionine
MGS	Massey genome service
ML	mother liquor
MR	molecular replacement
mRNA	messenger RNA
MS	mass spectrometry
MQ	milli Q water – ion exchanged purified water
Native-PAGE	non-denaturing PAGE
N-terminal	amino terminus
NtrBC	nitrogen regulatory system
OD	optical density
2OG	2-oxoglutarate
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	protein databank
Phen	phenylalanine
pI	isoelectric point
PIN	PilT N-terminal domain
Pro	proline
PTS	phosphoenolpyruvate:carbohydrate phosphotransferase system
qPCR	quantitative (real time) polymerase chain reaction
RBS	ribosome binding site
RMSD	root mean standard deviation
RNA-seq	RNA sequencing
RNase	ribonuclease
rNTP	ribonucleotide
rRNA	ribosomal RNA
RRM	RNA recognition motifs
rrs	16S ribosomal ribonucleic acid (rRNA) gene
RT	reverse transcriptase/transcription

List of Abbreviations

RT-qPCR	reverse transcription quantitative (real time) polymerase chain reaction
SAD	single anomalous diffraction
SDS	sodium dodecyl sulphate
SEC	size exclusion chromatography
Ser	serine
SOC	super optimal broth
sRNA	small non-coding RNA
ss	single stranded
T	thymine
TA	toxin-antitoxin
TADB	toxin-antitoxin database
TAE	tris-acetate-EDTA
TB	tuberculosis
TBE	tris-borate-EDTA
TE	tris EDTA buffer
TEMED	N, N, N, N,-tetramethylenediamine
TetR	tetracycline repressor
Tyr	tyrosine
tmRNA	transfer messenger RNA
tRNA	transfer RNA
Trp	tryptophan
U	uracil
UV	ultra violet
Vap	virulence associated protein
v/v	volume/volume
wt	wild type
w/v	weight per volume

1 Introduction

1.1 Prokaryotic Regulation

Regulation of prokaryotic metabolism in response to the environment is essential for bacterial survival. Regulation is primarily driven by a signal transduction pathway that translates changes in environmental conditions via signalling molecules to the regulatory systems of the cell to induce a response. The regulatory response occurs at three levels: 1) transcription, 2) post-transcription and 3) post-translation for efficient energy expenditure in response to extra- and intra-cellular nutrient status. Regulation at each level is incredibly complex and varies greatly depending on both the status and type of nutrient source available.

Archetypal regulation of metabolism at the transcription level is caused by transcription factors that bind to promoter regions of genes involved in metabolism to change their transcription rate in response to nutrient status signals in the cell. The signal is typically in the form of a carbon or nitrogen containing substrate, or its metabolite, which binds directly to the transcription factor (or another signalling protein that transfers the signal to the transcription factor) to induce a transcriptional response (Babu & Teichmann, 2003; Babu, Luscombe, Aravind, Gerstein, & Teichmann, 2004; Commichau, Forchhammer, & Stülke, 2006; Huffman & Brennan, 2002). Archetypal regulation of metabolism at the post-translational level is caused by both enzymatic modification and allosteric regulation of a protein to induce activation or inactivation (Arcondéguy, Jack, & Merrick, 2001; Goodey & Benkovic, 2008; Kamberov, Atkinson, & Ninfa, 1995). Enzymes responsible for post-translational modification by enzymatic modification are known as signalling proteins such as the P_{II} family, and proteins or other molecules that function by allosteric regulation are known as effectors and activate or inactivate metabolically relevant proteins in response to changes in nutrient status (Arcondéguy et al., 2001; Leigh & Dodsworth, 2007).

Metabolism is also fine tuned at the post-transcription level by a number of processes that regulate the translation of messenger RNA (mRNA) into proteins in response to changes in the environment. The importance of post-transcriptional

Chapter One

regulation has become more evident over recent years and there is now overwhelming evidence that it plays a much larger role in cellular adaption than previously expected (Hammarlöf, Bergman, Garmendia, & Hughes, 2015). The process of post-transcriptional regulation is also much more complex than previously thought. Small non-coding RNA (sRNA), RNA binding and RNA degrading enzymes, collectively referred to as mRNA interferases, are the major players of post-transcriptional regulation in prokaryotes (Gottesman, 2005; Hammarlöf et al., 2015). A sRNA typically work by binding to its target mRNA transcript and modifying its translation by either stabilizing or destabilizing it. The sRNA transcript production can be induced or repressed in response to environmental changes (Gottesman, 2005; Vanderpool & Gottesman, 2004). sRNA can also affect gene expression and even genomic recombination by interacting with the relevant enzymatic machinery for these processes, instead of the mRNA itself. An example of this is the CRISPR system found in prokaryotes that functions as an adaptable immune mechanism to protect cells from foreign nucleic acids, such as viruses (Barrangou et al., 2007; Fineran & Charpentier, 2012).

mRNA interferases play a critical role in post-transcriptional regulation by inhibiting translation of specific mRNA transcripts involved in certain carbon and nitrogen metabolic processes in response to changes in the environment. Examples include: RNA binding proteins that bind to and change the rate of degradation of specific mRNA targets by recruiting other degrading enzymes (Liu, Yang, & Romeo, 1995), and RNA endonucleases that bind to and directly degrade specific mRNA targets (Bodogai et al., 2006; Hammarlöf et al., 2015; Inouye, 2006; McKenzie, et al., 2012b; Romeo, 1996).

Integrated regulation of carbon and nitrogen metabolism is essential in order for organisms to function efficiently and balance their nutrient status and energy utilisation in the cell. For regulation to be effective, bacteria must be able to fine tune metabolism by modulating processes at the multiple levels of gene expression in response to rapid changes in the environment. Prokaryotic carbon and nitrogen metabolism both have unique modes of regulation. Below are

examples of the most well understood regulatory mechanisms present in prokaryotes.

1.1.1 Regulation of Carbon Metabolism

For efficient utilisation of carbon to occur the processes of carbon uptake, assimilation and metabolism must be tightly controlled. The best example of this is carbon catabolite repression (CCR), a phenomenon found in all prokaryotes and considered the paradigm of cellular regulation (Deutscher, 2008). CCR is the best studied example of a signal transduction pathway that occurs when bacteria are exposed to multiple carbon sources and prioritise the uptake and metabolism of only one (Deutscher, 2008; Saier, 1998). CCR is conserved in all prokaryotes and several different mechanisms have evolved to carry out the process; however, each mechanism involves components of the phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS) and protein phosphorylation (Deutscher, 2008; Postma, Lengeler, & Jacobson, 1993; Saier, 1998).

The PTS is responsible for the regulation of importing and phosphorylating sugars. Glucose is the most energy efficient carbon source for prokaryotes and glucose levels in the cell determine the phosphorylated state of the PTS. Each phosphorylated state of the PTS interacts with a different metabolic component of the cell to affect carbon assimilation and catabolism. High intracellular glucose levels induce an un-phosphorylated state of the PTS. This process is also called inducer exclusion, and is considered to be a major CCR mechanism, where the uptake of the favourable carbon substrate, generally glucose, directly alters the state of the PTS system, which results in the inhibition of other carbon transporters (Deutscher, 2008) (Figure 1.1). This results in the bacteria investing energy only into the transport of the most suitable carbon source for metabolism.

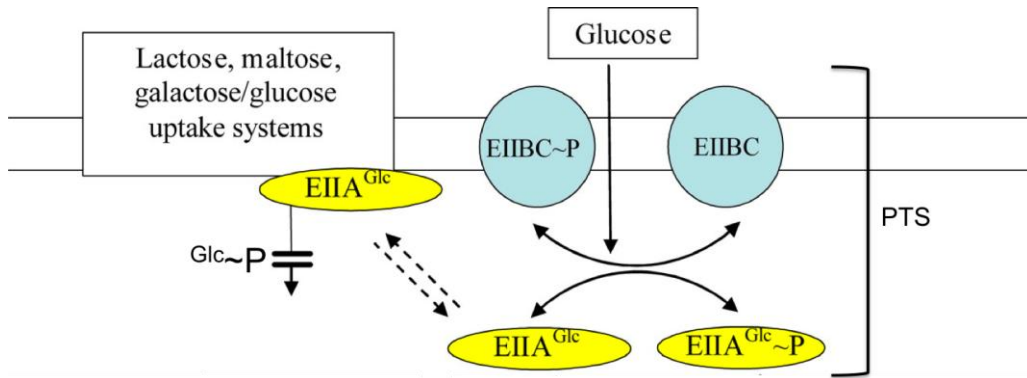


Figure 1.1. Mechanism underlying inducer exclusion in enteric bacteria. The transport of glucose into the cell results in the phosphorylation of glucose and net dephosphorylation of the PTS proteins (EIIBC and EIICA) and inducer exclusion. The dephosphorylated EIICAGlc blocks the import of other sugars (e.g. lactose, maltose and galactose) by binding to the corresponding transporter or kinase to inactivate it. Figure from Bowden et al. (2009).

The PTS also directly interacts with global transcription factors to induce a change in expression of multiple cohorts of genes involved in carbon metabolism. A well-known example is the CRP/cAMP mediated regulation of gene expression. The un-phosphorylated state of the PTS system, which is a result of increased intracellular glucose, represses the expression of cAMP synthesising enzymes resulting in reduced levels of both cAMP and the cAMP receptor protein (CRP) (Figure 1.2). CRP is a global transcription factor that controls the transcription of many catabolic genes involved in glucose uptake and alternative carbon metabolism, and cAMP is its co-factor. De-phosphorylation of the PTS results in the repression of these processes to reduce wastage of cellular energy when sufficient glucose is present in the cell (Deutscher, 2008; Ishizuka, Hanamura, Inada, & Aiba, 1994). This is a common regulatory mechanism in prokaryotes that allows for a fast global physiological response to changes in nutrient status.

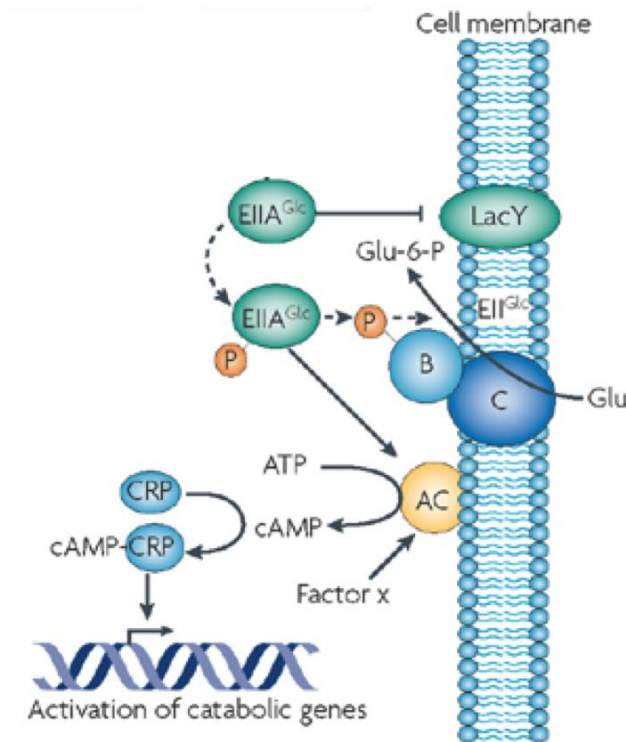


Figure 1.2. Carbon catabolite repression (CCR) in *E. coli*. When phosphorylated, the PTS (EIIAGlc-p) binds and activates adenylate cyclase (AC), which leads to cyclic AMP (cAMP) synthesis. High cAMP concentrations trigger the formation of cAMP-CRP complexes, which bind and activate the promoters of catabolic genes. In its de-phosphorylated form the PTS (EIIAGlc) cannot activate AC and the catabolic genes are repressed. Figure from Gorke & Stulke (2008).

Another mechanism of CCR involves the global carbon transcription factor catabolite repressor-activator (Cra). Cra regulates genes in the PTS system, as well as genes involved in carbon uptake, assimilation and metabolism, that function independent of the PTS system (Ramseier, 1995). Cra functions by both inhibiting and activating the expression of different cohorts of its regulon via binding different carbon metabolites as signalling molecules. It can simultaneously inhibit a cohort of genes, while inducing another in response to a range of different intracellular carbon source levels (Ramseier, 1995; Saier, 1998).

Recent work has revealed that several key enzymes in central carbohydrate metabolic pathways are controlled at the post-transcriptional level (Hammarlöf et al., 2015; Liu et al., 1995; McKenzie, et al., 2012b; Romeo, 1996). For example, the carbon storage regulator protein (CsrA), responsible for the regulation of glycogen biosynthesis, negatively regulates the expression of the genes responsible for this process during exponential growth. CsrA acts as an mRNA interferase and represses the expression of its regulon by binding to mRNA transcripts, which accelerates degradation and inhibits translation (Liu et al., 1995). CsrA has also been shown to simultaneously activate glycolysis during exponential phase, and, as a result, is suggested to be an important player in appropriately directing the intracellular carbon flux (Liu et al., 1995; Romeo,

1996). Additionally, research into an mRNA interferase Virulence associated protein (VapC₁₂₈₄) from *Mycobacterium smegmatis* belonging to the Type II toxin-antitoxin (TA) protein family was characterised as a post-transcriptional regulator of carbon metabolism. VapC₁₂₈₄ selectively binds and degrades mRNA transcripts involved in carbon uptake and metabolism in response to environmental changes (McKenzie, et al., 2012b).

For efficient energy expenditure and rapid response to environmental conditions prokaryotic regulation of carbon metabolism occurs at both the transcription and post transcription stages of gene expression. Transcriptional regulation in the form of CCR is present in all bacteria and is an integral component of regulating carbon metabolism, allowing bacteria to selectively transport and metabolise optimal carbon sources available in their environment (Deutscher, 2008). An additional layer of regulation is observed by the mRNA interferases that allow rapid re-generation of the mRNA transcript pool in response to changing nutrient signals. These examples illustrate only a subset of processes present in bacteria that regulate carbon metabolism; however, they do provide a good example of the diversity of mechanisms available and the sophistication involved at each level of regulation.

1.1.2 Regulation of Nitrogen Metabolism

Nitrogen regulatory mechanisms vary significantly across the phylogenetic spectrum of prokaryotes. However, in most prokaryotes, nitrogen metabolism is regulated using a similar signal transduction pathway as with CCR in carbon metabolism. This process is mediated by signal transduction proteins that belong to one of the most conserved protein families in prokaryotes: the P_{II} superfamily (Arcondéguy et al., 2001; Leigh & Dodsworth, 2007). Examples of P_{II} protein include GlnB, GlnK, NifI₁ and NifI₂. A principle characteristic of P_{II} proteins is their ability to be switched between two forms by covalent modification when a conserved tyrosine residue in the T-loop of the structure is uridylylated or adenylylated (Arcondéguy et al., 2001; Williams, Bennett, Barton, Jenkins, & Robertson, 2013). The covalent state of the protein is determined by the intracellular nitrogen status, and results in inactivation or activation of the P_{II} protein (Arcondéguy et al., 2001; Kamberov et al., 1995). Activated P_{II} proteins

Chapter One

function to post-translationally modify other proteins involved in nitrogen metabolism, including 1) enzymatic proteins, resulting in inactivation or activation of specific nitrogen metabolic processes, and 2) transcription factors to induce changes in the expression of the genes encoding proteins involved in nitrogen metabolism (Arcondéguy et al., 2001; Leigh & Dodsworth, 2007).

One well-studied example is the role of P_{II} proteins in the regulation of ammonia assimilation. Ammonia is the least energy expensive nitrogen source for prokaryotes, and all other nitrogen sources have to be converted to this to be metabolically useable, at a cost to the organism (Arcondéguy et al., 2001; Leigh & Dodsworth, 2007). As a consequence, prokaryotic regulation of nitrogen metabolism is driven by ammonia availability and assimilation. In most prokaryotes, the major mechanism of ammonia assimilation is the glutamine synthetase (GS)/glutamate synthase (glutamine:2-oxoglutarate aminotransferase or GOGAT) pathway (Arcondéguy et al., 2001; Nolden, Farwick, Kramer, & Burkovski, 2001a). The GS/GOGAT pathway is ubiquitous in bacteria and is responsible for converting glutamate to glutamine. GOGAT then transfers the amido group of glutamine to 2-oxoglutarate (2OG) (the precursor of ammonia assimilation).

GS is highly regulated at both the transcriptional and post-transcriptional level. It can be reversibly modified by adenylation at a conserved tyrosine residue, which alters its enzymatic activity in response to nitrogen availability (Arcondéguy et al., 2001) (Figure 1.3). GS is adenylylated under an increase of nitrogen, which results in its inactivation. This modification is controlled by the P_{II} proteins GlnB and GlnK (Adler, Purich, & Stadtman, 1975; Arcondéguy et al., 2001). It is proposed that the physiological significance of this in-activation during nitrogen surplus is to prevent ATP and glutamate depletion, allowing rapid growth during favourable growth conditions (Kustu, Hirschman, Burton, Jelesko, & Meeks, 1984).

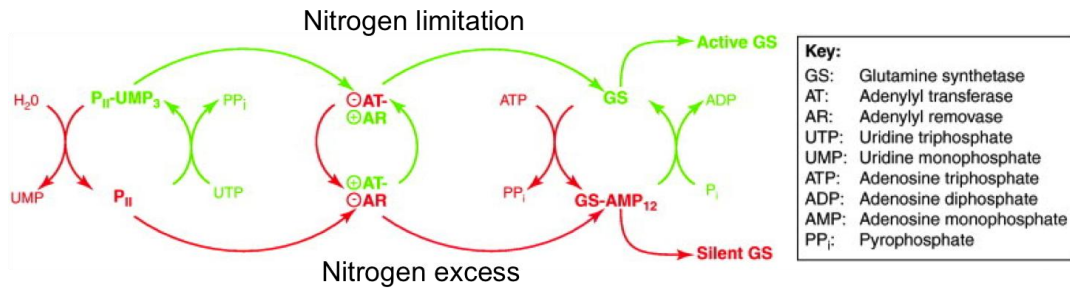


Figure 1.3. Regulation of glutamine synthetase (GS) by P_{II} mediated post-translational adenylation. The activity of GS is regulated by the bifunctional enzyme GS-AT, which in turn is controlled by the regulatory protein P_{II}. P_{II} is subject to uridylylation (P_{II}-UMP₃) and deuridylylation (P_{II}), and binds to GS-AT in both the modified and unmodified states with opposing effects on the activities of the enzyme. P_{II}-UMP₃ activates the AR activity of GS-AT but inhibits its adenylyl transferase activity, thus leading to activation of GS via deadenylylation. Conversely, in the de-uridylylated state, P_{II} inhibits the deadenylation or removease activity of GS-AT and stimulates the adenylylation activity towards GS, thereby inhibiting GS. Green and red arrows indicate stimulatory and inhibitory steps in the regulation of GS, respectively. Figure from Itzen, Blankenfeldt, & Goody (2011).

Similar to carbon metabolism, activation and repression of nitrogen-regulated genes in bacteria is coordinated by the action of global transcription factors. A well studied example is the global nitrogen regulatory (NtrBC) system present in enteric bacteria (Keener & Kustu, 1988). The transcription factor NtrC is phosphorylated in response to low nitrogen availability, which results in its activation and subsequent upregulation of its regulon (Figure 1.4) (Austin & Dixon, 1992; Kern et al., 1999). The phosphorylation of NtrC is dependent on the P_{II} proteins GlnB and GlnK (Arcondéguy et al., 2001; Atkinson & Ninfa, 1998; Kamberov et al., 1995). GlnB and GlnK are post-translationally modified in response to nitrogen starvation, which provides an intracellular switch that in turn activates a signal transduction cascade that results in the phosphorylation of NtrC (Arcondéguy et al., 2001; Kamberov et al., 1995). The result is the induced expression of the genes in the NtrC regulon, which encode for ammonium transport and assimilation, alternative nitrogen uptake and metabolism, and nitrogen fixation (Arcondéguy et al., 2001; Atkinson & Ninfa, 1998; Claverie-Martin & Magasanik, 1991; Schneider, Kiupakis, & Reitzer, 1998). In bacteria that do not have an NtrC system, such as the firmicutes and some actinobacteria, the global transcription factors GlnR, TnR and AmtR function as the major P_{II} mediated nitrogen-limiting response players (Arcondéguy et al., 2001; Jakoby,

Nolden, Meier-Wagner, Kramer, & Burkovski, 2000; Schreier, Brown, Hirschi, Nomellini, & Sonenshein, 1989; Tiffert et al., 2008).

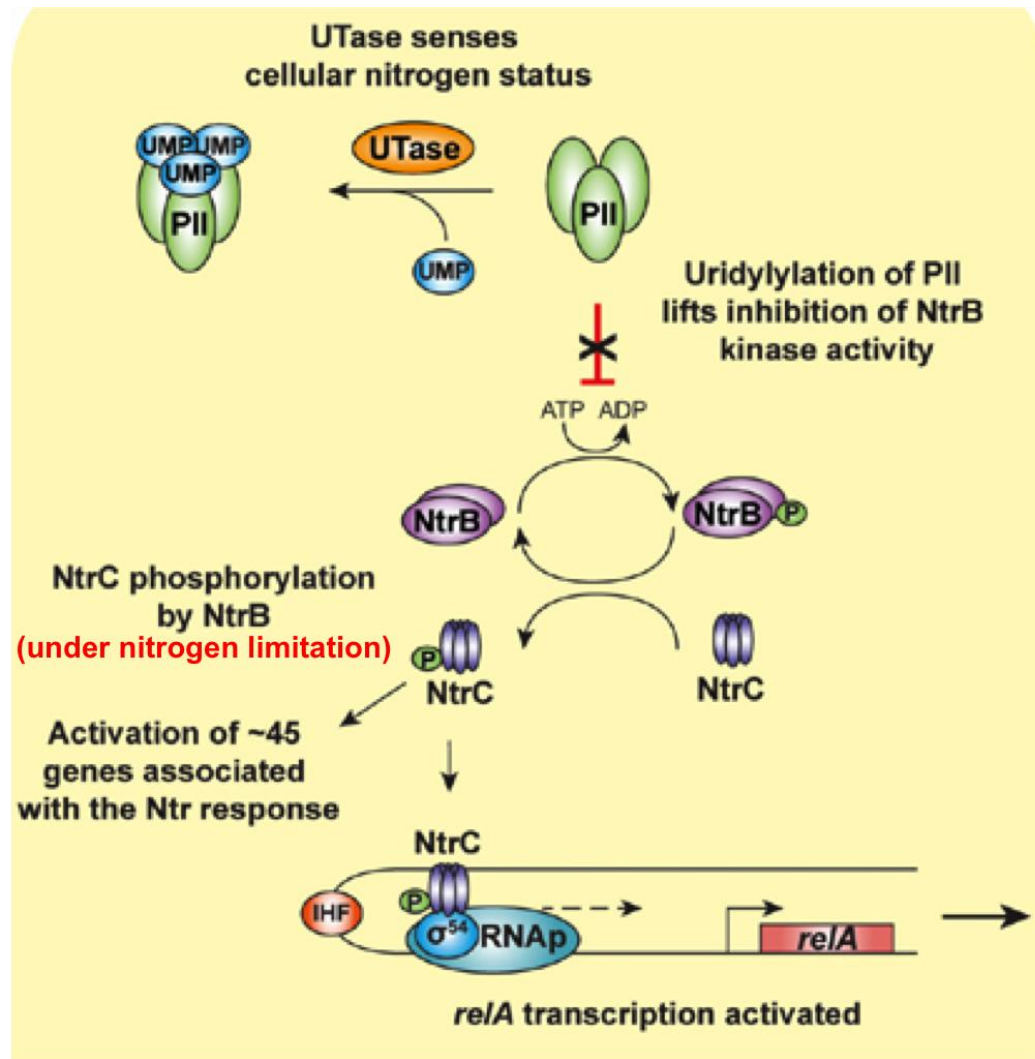


Figure 1.4. The P_{II} mediated two component NtrBC system of enteric bacteria. Under nitrogen limiting conditions the P_{II} proteins phosphorylate the NtrB (NtrB-P) kinase, which subsequently phosphorylates NtrC (NtrC-P). NtrC-P is activated and binds to promoter regions upstream of the genes in its regulon to induce expression. The induced expression is referred to as the Ntr response and involves genes that encode proteins involved in alternative nitrogen transport and metabolism. Figure from Brown et al. (2014).

P_{II} signalling proteins play a pivotal role in nitrogen metabolism by acting as an intra cellular switch that allows a global response within the cell at both the transcriptional and post-transcriptional level. The P_{II} proteins post-translationally modify global transcription factors, like NtrC, to induce a transcriptional response of genes encoding nitrogen related processes. They also post-translationally modify critical nitrogen metabolic enzymes, like GS, to activate or inactivate them in response to changes in nitrogen availability.

1.2 Transcriptional Regulation

Transcription factors are ultimately responsible for inducing the transcriptional response to changes in nutrient status provided by signal transduction pathways. Transcription factors regulate the transcription of genes into mRNA transcripts that are then translated into proteins. They are generally homodimeric proteins that contain two domains in each monomer: a DNA binding motif and a regulatory domain. There is an extraordinary diversity and abundance of transcription factors that span the prokaryotic kingdom, with over 30 different superfamilies as defined by their type of DNA binding motif (Huffman & Brennan, 2002; Wilson, Charoensawan, Kummerfeld, & Teichmann, 2007).

Transcription factors function by binding to specific DNA sequences in the promoter regions of genomic DNA that encode proteins (Huffman & Brennan, 2002). Binding of the transcription factor to DNA is generally dictated by the presence or absence of a signalling molecule or protein. Transcription factors can either activate or repress transcription, or both. Previously, it was assumed that the DNA binding motif family and its structural location (i.e. C-terminal or N-terminal) was indicative of a transcription factor's regulatory function. However, it is now suggested that the regulatory function is dictated by the position of its binding site on DNA (Babu & Teichmann, 2003).

Generally, activator transcription factors bind upstream of the transcription start site and repressors can bind both upstream and downstream of the transcription start site. Dual regulators have equal ratios of binding sites upstream and downstream of the transcription start site (Babu & Teichmann, 2003). Transcription factors binding at upstream sites function to stabilise RNA polymerase, the enzyme responsible for transcribing DNA into RNA. Repressor transcription factors that bind upstream generally function by recruiting other repressor proteins to downstream sites. Transcription factors that bind at downstream sites inhibit RNA polymerase activity by steric hindrance. However, there is a diverse range of activation and inhibition mechanisms including DNA structure alteration and unwinding, protein-molecule interaction and protein-protein interactions making transcription factor regulation a very complex process to generalise (Babu & Teichmann, 2003; Huffman & Brennan, 2002). This section

will focus on transcription factors that are major players of nitrogen regulation in actinomycetes, as this is a large focal point of the research presented in this thesis.

1.2.1 GlnR Transcription Factor Protein Family

GlnR is a global response regulator that belongs to the OmpR transcription factor family. Information from X-ray crystallography structures and experimental work has shown that GlnR functions as a homodimer, where each monomer is made of a regulatory domain and a DNA binding domain (Figure 1.5) (Ling et al., 2015; Tiffert et al., 2008).

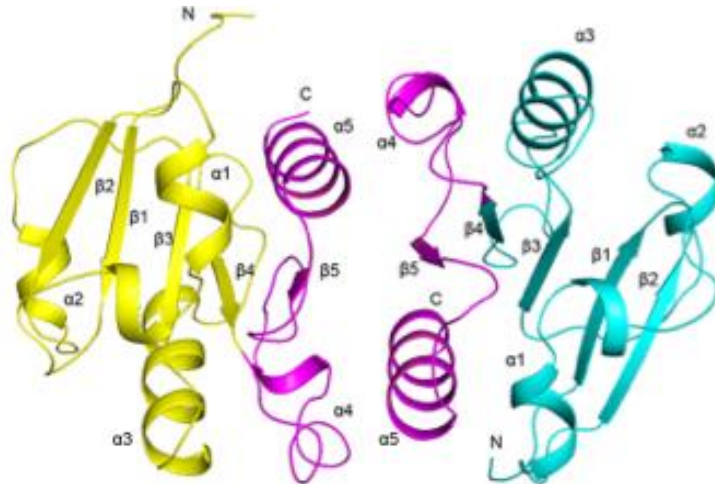


Figure 1.5. X-ray crystal structure ribbon diagram of *Amycolatopsis mediterranei* GlnR. The two molecules constituting the homodimer are coloured in yellow and cyan. The dimer interface is shown in magenta. Figure from Lin et al. (2014).

GlnR is present in several members of the actinomycetes phylum that do not contain a NtrBC system including *Streptomyces coelicolor*, *S. avermitilis*, *M. smegmatis* and many pathogenic mycobacteria. GlnR controls over 100 genes, with over 25 % of genes encoding ammonium uptake and assimilation, alternative nitrogen metabolism, and P_{II} proteins (Amon et al., 2008; Fink, Weisschuh, Reuther, Wohlleben, & Engels, 2002; Jenkins, Barton, Robertson, & Williams, 2013; Jenkins, Robertson, & Williams, 2012; Jessberger et al., 2012). The regulon also contains a number of genes encoding enzymes that are predicted to be involved in several processes that generate ammonium from urea, suggesting that urea is an important alternative nitrogen source for bacteria in this phylum (Jenkins et al., 2012). The GlnR DNA binding motif has been identified (AC/T-

n9-AC), which is present in every known GlnR binding site (Figure 1.6) (Jenkins et al., 2013).

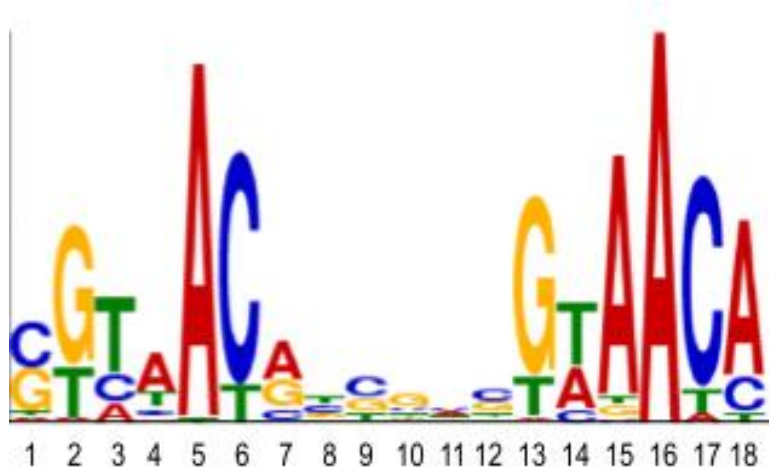


Figure 1.6. GlnR consensus binding motif derived from the GlnR binding regions. The height of letters represents the frequency of corresponding nucleotides in the GlnR binding site. Figure from Jenkins et al. (2013).

Activation of GlnR is induced upon nitrogen limitation, however the signal that stimulates GlnR to activate the transcription of its regulon is still unknown (Amon et al., 2008; Jenkins et al., 2013; Jessberger et al., 2012; Tiffert et al., 2008). OmpR proteins are characterised as two-component response regulators that are activated by post-translational phosphorylation of a conserved amino acid residue, similar to the NtrBC system (Amon et al., 2008; Kenney, 2002). Conservation of the putative GlnR phosphorylation site is essential in the transcriptomic response to nitrogen limitation (Jenkins et al., 2012), however, no corresponding kinase has yet been identified (Amon et al., 2008). Metabolic footprinting of *M. smegmatis* has revealed an increased intracellular ratio of 2-OG:glutamine/glutamate, as well as intracellular levels of a sugar derived compound called glucosylglycerate (GGA) during nitrogen limitation. It is suggested that 2-OG and GGA could be the nitrogen limitation signal that stimulates GlnR to activate transcription of its regulon (Behrends, Williams, Jenkins, Robertson, & Bundy, 2012).

GlnR is an example of a transcription factor that functions as a global regulator of nitrogen metabolism in several members of the actinomycetes phylum, however not all of the genes involved in nitrogen metabolism in these organisms are under the control of GlnR. One explanation for this is that there is a secondary nitrogen regulator. In fact, there is strong evidence that the TetR-type

transcription regulator AmtR plays a secondary role in mycobacteria nitrogen regulation (Amon et al., 2008).

1.2.2 The TetR Transcription Factor Protein Family

The tetracycline repressors (TetR) are one of the most abundant, widely distributed and understood transcription factor families. The TetR protein family is classified by the original TetR protein, which regulates the *tet* genes using tetracycline and magnesium as the signalling molecule. The *tet* genes encode proteins that impart tetracycline resistance. TetR forms a homodimer with each monomer composed of 10 α -helices that form a N-terminal helix turn helix (HTH) motif DNA binding domain (DBD) and a large C-terminal ligand binding domain (LBD) (Ramos et al., 2005) (Figure 1.7). The LBD is also responsible for dimerization of the protein.

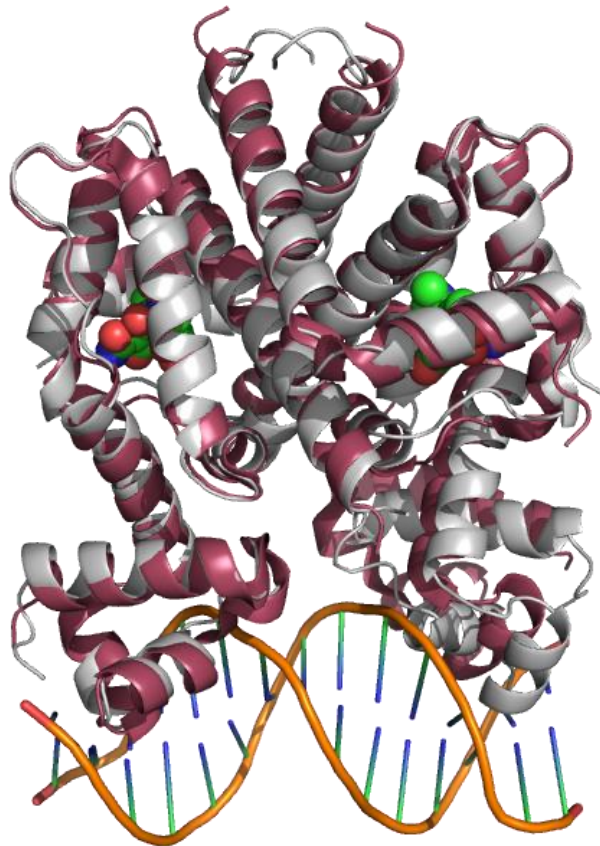


Figure 1.7. X-ray crystal structures of TetR. Overlay of DNA-bound (pink) and magnesium-tetracycline-bound (grey). Magnesium tetracycline (Mg^{2+} -Tet) atoms are shown as spheres and DNA is represented as a ball-and-stick model. PDB coordinates for DNA-bound structure from Orth et al. (2000) (PDB ID 1QPI) and Mg^{2+} -Tet structure from Kisker et al. (1995) (PDB ID 2TRT). Figure was drawn using PyMol (Delano, 2002).

Members of the TetR family show conservation in the DNA binding motif and little to no conservation in the regulatory domain. This is likely to reflect the diversity of different signalling molecules recognised in the regulatory domains and hence variations in functionality (Ramos et al., 2005). The X-ray crystal structures of TetR unbound, bound to tetracycline and bound to DNA have been solved (Hinrichs et al., 1994; Kisker, Hinrichs, Tovar, Hillen, & Saenger, 1995; Orth et al., 1998; Orth, Schnappinger, Hillen, Saenger, & Hinrichs, 2000). Comparative analyses of the structures have revealed insight into the specificity and function of TetR. The structures show that the DNA binding motif exclusively binds to the major groove of a palindromic DNA sequence, however there is no major groove interaction at the central base pair (Figure 1.7). This suggests that the spacing, rather than the sequence, within the two half sites of recognition is critical for recognition and binding (Huffman & Brennan, 2002). Binding of TetR to DNA is disrupted by binding of tetracycline into the LBD. This induces a conformational change that results in the DBD rotating via a pendulum motion so that the two recognition helices are separated, disrupting the contact with DNA (Huffman & Brennan, 2002).

The regulatory functions of members of the TetR transcription factor superfamily includes multidrug resistance, global regulation, catabolism, nitrogen metabolism, antibiotic biosynthesis, osmotic stress and pathogenicity (Ramos et al., 2005). All known members of the family function as repressors and most have the same mechanism of protein-DNA interaction disruption by a signalling molecule binding to the regulatory region (Ramos et al., 2005). TetR transcription factors are particularly abundant in organisms living in fluctuating environments, pathogenic bacteria and extremophiles, for example *Corynebacterium*, *Mycobacterium*, and *Deinococcus* (Ramos et al., 2005). Considering the diversity in function and high representation in organisms subjected to either fluctuating or extreme environments, it has been suggested that TetR proteins play a critical role in metabolic adaptation.

1.2.2.1 AmtR Transcription Factor Protein Family

AmtR belongs to the TetR protein family and, like other TetR transcription factors, it is found in a range of bacteria that are exposed to fluctuating

environments and is recognised as a global nitrogen regulator (Williams et al., 2013). The role of AmtR in this diverse range of bacteria is not well understood, however it has been extensively investigated in the soil actinomycetes *Corynebacteria glutamicum*. *C. glutamicum* is utilized industrially for its ability to synthesise high yields of extra cellular amino acids. Amino acid production is entirely dependent on nitrogen metabolism, and, as a result, nitrogen metabolism of *C. glutamicum* is now well understood.

C. glutamicum does not have an NtrBC system and instead the AmtR transcription factor functions as the global response regulator to nitrogen limitation (Burkovski, 2003; Jakoby et al., 2000; Nolden et al., 2001a). The AmtR regulon includes 35 genes that are involved in nitrogen metabolism which are repressed under nitrogen sufficient conditions. The repression is lifted by direct interaction with the adenylylated form of the P_{II} protein GlnK, a nitrogen assimilation regulatory protein, which is only present under nitrogen limiting conditions (Figure 1.8) (Beckers et al., 2005; Buchinger et al., 2009; Muhl et al., 2009; Nolden, Ngouoto-Nkili, Bendt, Kramer, & Burkovski, 2001b; Strosser, Ludke, Schaffer, Kramer, & Burkovski, 2004).

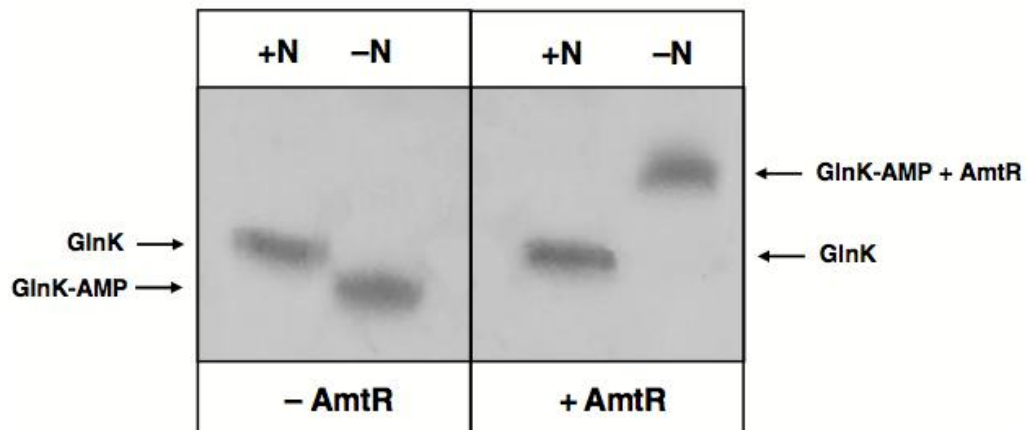


Figure 1.8. Analysis of AmtR-GlnK interaction by native gel electrophoresis. *C. glutamicum* GlnK overexpressing cell cultures were grown in either nitrogen surplus (+N) or nitrogen limiting (-N) conditions. As a result, GlnK was present in either its unmodified (GlnK) or adenylylated form (GlnK-AMP), respectively. Fractions of these cells were subjected to native polyacrylamide gel electrophoresis, either in the absence (left panel) or presence (right panel) of purified AmtR. Cell extracts containing GlnK-AMP resulted in a shift of AmtR motility. Figure from Beckers et al. (2005).

The AmtR-DNA binding motif is defined as a semi palindromic sequence of tttCTATN₆AtAGat/aA (with bases represented by capital letters being highly conserved) found on both the sense or antisense strand of the promoter region (Figure 1.9) (Beckers et al., 2005; Muhl et al., 2009). AmtR-DNA binding has been demonstrated both *in vitro*, and, in some cases, *in vivo* with a co-factor independent binding mechanism (Beckers et al., 2005; Jakoby et al., 2000; Muhl et al., 2009). It is assumed that the number of upstream binding sites and conservation of the binding motif determines the strength of repression.

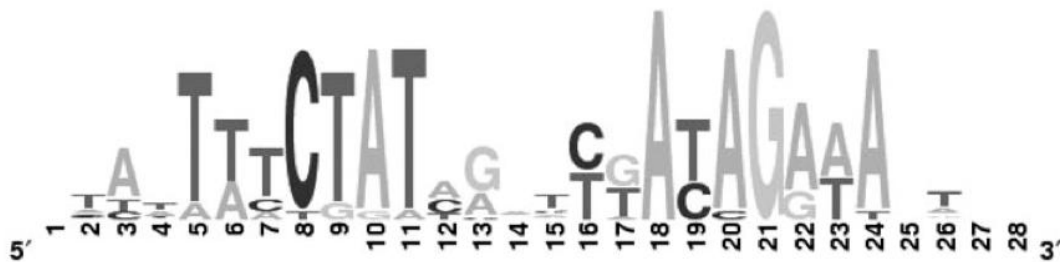


Figure 1.9. The AmtR box of *C. glutamicum*. The AmtR binding motif was established based on the binding sites predicted by bioinformatics analyses, which were verified by RNA hybridization analyses, real-time RT qPCR and gel retardation tests. The height of letters represents the frequency of corresponding nucleotides in the AmtR box. Figure from Beckers et al. (2005).

The AmtR regulon is composed of genes that encode proteins that respond to nitrogen limitation including ammonia uptake and assimilation, alternative nitrogen uptake and metabolism, and the master nitrogen signalling P_{II} protein GlnK, showing indirect autoregulation (Beckers, Bendt, Kramer, & Burkovski, 2004; Beckers, Nolden, & Burkovski, 2001; Bendt, Beckers, Silberbach, Wittmann, & Burkovski, 2004; Jakoby et al., 2000; Muhl et al., 2009; Nolden et al., 2001a). This system of nitrogen regulation has only been observed in *C. glutamicum* and introduces a new player, at the level of transcription, to the intricate network of prokaryotic nitrogen regulation.

S. avermitilis and *M. smegmatis* are both members of the actinomycetes that do not have a NtrBC system and instead contain both GlnR and AmtR. In *S. avermitilis* the GlnR is suggested to work as a crucial nitrogen regulator, although functional studies have not confirmed this (Chen, Zhu, Zheng, Jiang, & Lu, 2013). The AmtR has been identified as a regulator involved in alternative urea metabolism and amino acid transport, suggesting that it plays an important role in

Chapter One

nitrogen metabolism (Chen et al., 2013). In *M. smegmatis*, GlnR has been identified as an important nitrogen regulator with its regulon containing the genes responsible for ammonia uptake, assimilation and signal transduction P_{II} proteins, as mentioned earlier. Although GlnR has been identified as an important nitrogen regulator in mycobacteria, it has also been shown that a number of genes that encode for critical proteins required in response to nitrogen limitation are not under the control of GlnR. The regulation of these genes is poorly understood and it has been suggested that these genes are under control of AmtR (Amon et al., 2008).

Using the *C. glutamicum* AmtR DNA binding motif, a bioinformatics analysis of the *M. smegmatis* genome was conducted to identify putative genes under the control of AmtR, however, none were found (Amon et al., 2008). Further bioinformatic analyses were conducted to characterise the AmtR regulon using a different approach (Jessberger et al., 2012). The search criteria for a gene under putative AmtR control were based on the degree of (I) co-occurrence and co-localization with the *amtR* gene and (II) involvement in nitrogen metabolism. Interestingly, the genes identified were homologous to the genes in the AmtR regulon of *S. avermitilis* (Chen et al., 2013; Jessberger et al., 2012). The genes appear to form an operon based on the co-localization in the genome and simultaneous up regulation in response to nitrogen levels. The operon encodes for an amino acid permease, a urea carboxylase and associated proteins and an amidases (Figure 1.10).

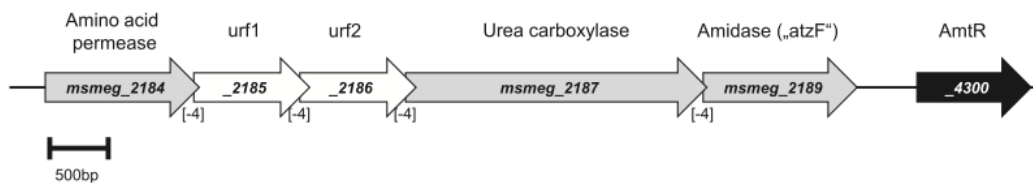


Figure 1.10. Genomic map of AmtR regulon in *M. smegmatis*. The arrows indicate length and transcriptional orientation of annotated genes. Number in square brackets show the lengths of intergenic regions in bp. The AmtR gene (*msmeg_4300*) is found elsewhere in the genome and does not form part of the operon. Figure from Jessberger et al. (2012).

To validate this work, *M. smegmatis* *amtR* deletion strains were grown under nitrogen limitation and demonstrated the de-regulation of this operon in the

absence of AmtR (Jessberger et al., 2012). This work confirmed that the *M. smegmatis* AmtR controls a regulon of genes encoding for alternative nitrogen uptake and urea metabolism, which is upregulated in nitrogen limitation. As mentioned above, the high prevalence of urea uptake and metabolic genes in *M. smegmatis* suggests that urea is an important alternative nitrogen source (Amon et al., 2008). This study shows that AmtR plays a role in regulating this alternative metabolism. Co-factor independent binding of AmtR to the promoter region of the operon was also confirmed and no tested putative signalling molecules were observed to influence binding. The signal to lift repression, like GlnR in mycobacteria, is currently unknown. Elucidating the signal for these transcription factors is critical to improve our understanding of nitrogen metabolic regulation. These initial findings suggest *M. smegmatis* AmtR has a unique role in nitrogen regulation and further investigation into the structure and function of AmtR will provide greater understanding of the regulatory mechanisms functioning in mycobacteria.

1.3 Post-Transcriptional Regulation

The process of transcription transcribes genetic information from DNA into mRNA, which can then be translated into a functional protein. The transcription of genes into mRNA is regulated by transcription factors in response to environmental signals. Post-transcriptional regulation is the next level of regulation, where the translation of mRNA is modulated in response to environmental signals by a range of different regulators.

Prokaryotic mRNA is very unstable with half lives on the order of minutes (Picard, Dressaire, Girbal, & Coccagn-Bousquet, 2009). The instability of mRNA is attributed to the proteins and small non-coding RNA that interact with it and determine its fate. Small non-coding RNAs (sRNA) mediate translation by a diversity of regulatory mechanisms. In brief, sRNA binds to RNA recognition motifs (RRM) in targeted mRNA transcripts to induce a conformational change to the secondary structure, which can inhibit translation by either increasing the degradation rate of the transcript or inhibiting the translational machinery from initiating translation (Picard et al., 2009). They can also activate translation by stabilising the transcript and recruiting the translational machinery to the ribosome

binding site (RBS) (Picard et al., 2009; Vierke, Engelmann, Hebbeln, & Thomm, 2003). RNA binding and degradation proteins, collectively named mRNA interferases, also play a critical role in post-transcriptional regulation.

An interesting protein based regulatory system is the toxin-antitoxin (TA) protein system. A number of toxins in the TA systems, such as the Type-II TA VapBC protein systems, function as selective mRNA interferases that inhibit translation (Inouye, 2006). These systems are well documented as stress-response elements and have been shown to specifically target mRNA transcripts involved in both carbon and nitrogen metabolism (Bodogai et al., 2006; Inouye, 2006; McKenzie, et al., 2012b). The remainder of this section will focus on TA systems, as this is the other main focus of the research presented in this thesis.

1.3.1 Toxin-Antitoxin (TA) Protein Systems

The TA loci were first discovered in 1986 on prokaryotic plasmids and were identified as selfish elements that play a role in the prevention of post segregational plasmid loss (Gerdes, Rasmussen, & Molin, 1986). TA systems have since been found to be highly abundant in the chromosomes of a range of prokaryotes, but have very different proposed functions, including fine tuning the physiology of the cell in response to environmental cues (Arcus, McKenzie, Robson & Cook, 2011; Buts, Lah, Dao-Thi, Wyns, & Loris, 2005; McKenzie, et al., 2012b). TA systems are arranged as a bicistronic operon with the gene encoding the antitoxin found upstream of gene encoding the toxin, which typically overlap by 2-12 bp. Both the toxin and the antitoxin are co-expressed and form a complex that is inactive and binds to the promoter region upstream of its own operon to allow autoregulation.

There are five types of TA systems classified by the nature and mode of the antitoxin, where the toxin is always a protein. In Type I systems, the antitoxin is an un-translated, antisense RNA, which is complementary to the toxin mRNA (Fozo, Hemm, & Storz, 2008; Gerdes, 2012). It regulates toxin expression by inhibiting the toxin's translation. In Type II systems, the antitoxin is a protein and inhibits toxin activity by binding to the toxin, forming an inactive complex (Gerdes, Christensen, & Løbner-Olesen, 2005). In Type III systems, the antitoxin

is a small, non-antisense RNA molecule that binds to the toxin to inhibit activity (Fineran et al., 2009). In Type IV systems, the antitoxin is a protein that competes with the toxin for the same binding target (Dy, Przybilski, Semeijn, Salmond, & Fineran, 2014; Masuda, Tan, Awano, Wu, & Inouye, 2012). In Type V systems, the antitoxin protein cleaves the toxin encoding mRNA (Fineran et al., 2009; Short et al., 2013). Type II TA systems were the first discovered and are the most prevalent within the prokaryotic kingdom, and will be the focus of this thesis.

1.3.2 Type II TA Systems

The Type II TA systems are composed of diverse families, each with different modes of action, but ultimately function to induce bacteriostasis in the cell in response to specific signals (Arcus et al., 2011; Gerdes et al., 2005; Mutschler, Gebhardt, Shoeman, & Meinhart, 2011; Ning, Jiang, Liu, & Xu, 2013a; Yamaguchi, Park, & Inouye, 2011) (Table 1.1).

Table 1.1. Summary of Type II TA families and their general mode of action.

Type II TA Family	Mode of Action
MazEF	Induces bacteriostasis via translation inhibition by mRNA cleavage in response to nutrient stress (Christensen, Pedersen, Hansen, & Gerdes, 2003)
RelBE	Induces bacteriostasis via translation inhibition by mRNA cleavage in response to nutrient stress (Pedersen et al., 2003)
ParDE	Inhibits DNA replication and transcription by poisoning topoisomerase II (Jiang, Pogliano, Helinski, & Konieczny, 2002)
CcdAB	Inhibits DNA replication and transcription by poisoning topoisomerase II (Afif, Allali, Couturier, & Van Melderren, 2001; Van Melderren, 2002)
Phd-Doc	Induces bacteriostasis via translation inhibition by ribosomal inactivation (Liu, Zhang, Inouye, & Woychik, 2008)
HicAB	Induces bacteriostasis via translation inhibition by mRNA and tmRNA cleavage in response to nutrient stress (Jorgensen, Pandey, Jaskolska, & Gerdes, 2009)
HipBA	Induces bacteriostasis via translation inhibition by inactivation of tRNA by phosphorylation (Schumacher et al., 2009)
MosAT	Promotes maintenance of an integrative conjugative element that confers antibiotic resistance (Wozniak & Waldor, 2009)
VapBC	Induces bacteriostasis via translation inhibition by mRNA cleavage (Arcus et al., 2011)

The majority of Type II toxins function by inhibiting translation which leads to growth arrest (Brown & Shaw, 2003; Yamaguchi et al., 2011) (Figure 1.11). The toxin is inactivated by forming a complex with the antitoxin (Gerdes et al., 2005). The high prevalence of Type II systems on chromosomes of bacteria led to the proposal that chromosomal Type II TA systems are stress-response elements contributing to prokaryote adaptation to stressful environments (Gerdes et al., 2005). This proposal has been supported by a series of recent experimental findings. For example, in *E. coli*, activation of Type II TA systems is triggered by various stressors (Christensen-Dalsgaard, Jørgensen, & Gerdes, 2010; Hazan, Sat, & Engelberg-Kulka, 2004). It is suggested that the general Type II TA system mode of action is to induce reversible growth arrest under stressed conditions via inhibition of translation. This allows stressed cells to remain in a dormant or non-growing stress-tolerant state until more favourable environmental conditions return.

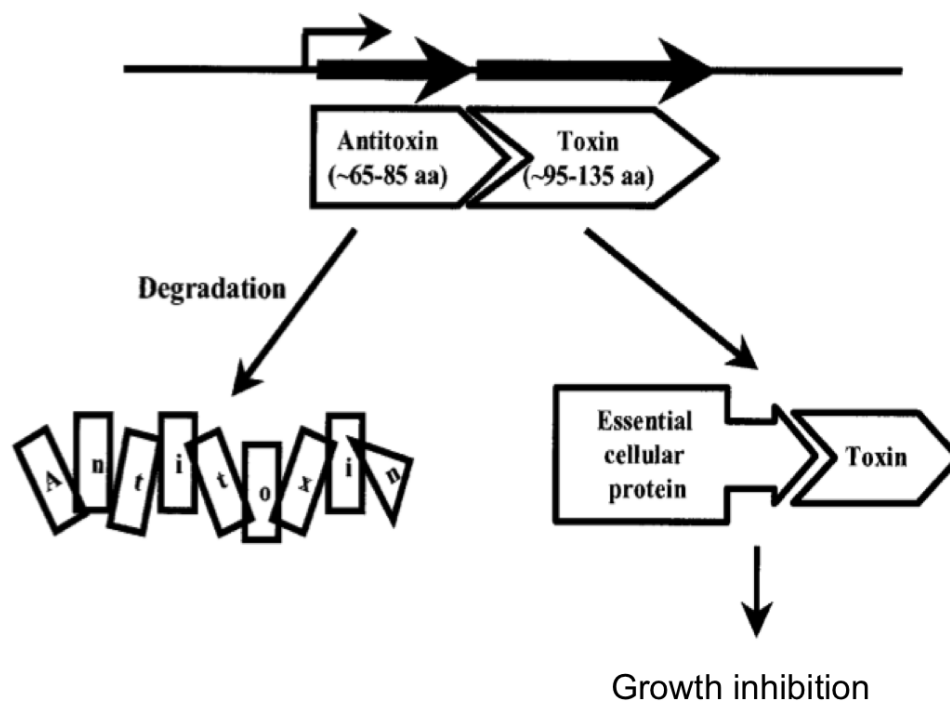


Figure 1.11. Generalised schematic of the characteristics and genetic organisation of bacterial TA pairs. Figure modified from Brown & Shaw (2003).

1.3.3 VapBC TA Systems

The VapBC family is defined by their toxic Virulence associated protein (VapC), which belongs to the PilT N-terminal domain (PIN domain) family of proteins. The VapBC TA family is the largest of the Type II TA systems but they

remain the least well characterised, as a consequence of their relatively recent identification and diverse functions (Arcus, Bäckbro, Roos, Daniel, & Baker, 2004; Arcus, Rainey, & Turner, 2005; Clissold & Ponting, 2000; Gerdes et al., 2005). VapBC TA systems have become a point of interest because of their overrepresentation in the genomes of pathogenic and environmentally hazardous bacteria, providing a new potential target for controlling these organisms.

1.3.3.1 Expansion of VapBC Systems in Pathogenic and Environmentally Hazardous Bacteria

VapBC TA systems are overrepresented in the genomes of pathogenic and environmentally hazardous bacteria such as the human pathogen *M. tuberculosis* and the fresh water toxic bloom-forming cyanobacteria *Microcystis aeruginosa* (Arcus et al., 2011). *M. tuberculosis* has 47 *vapBC* operons encoded in its genome, where the related non-pathogenic environmental *M. smegmatis* has just one (Arcus et al., 2011; McKenzie, et al., 2012b). *M. aeruginosa* has 34 *vapBC* operons encoded in their genome, where the related non-toxic *Synechocystis* sp. *PCC6803* (*Synechocystis*) has only 12 (Kaneko et al., 2001; Leplae et al., 2011; Makarova, Wolf, & Koonin, 2009). This overexpansion suggests a functional link between toxicity and VapBC systems.

1.3.3.2 VapBC Autoregulation

VapBC pairs are arranged as a bicistronic operon with both the antitoxin (VapB) and the toxin (VapC) co-expressed to form an inactive complex (VapBC) that binds to the promoter region upstream of its own operon to allow autoregulation. The VapB is responsible for binding to inverted repeat (IR) units in the promoter region, which is typically enhanced by the interaction of VapB and VapC (Figure 1.12) (Arcus et al., 2011).

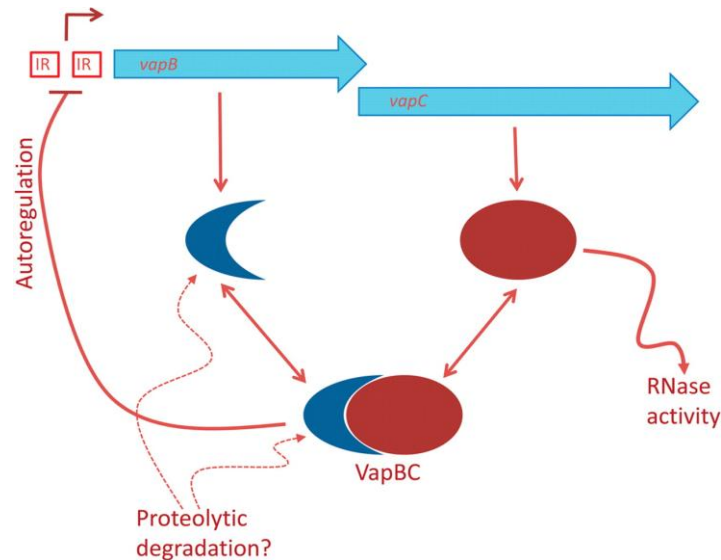


Figure 1.12. Schematic diagram of a generalized VapBC system. The antitoxin (VapB) binds to the promoter region of the TA operon. This binding is enhanced by the toxin (VapC). The toxin gene encodes a toxic protein that often leads to growth arrest via ribonuclease-mediated inhibition of translation. When the toxin is bound with the antitoxin, it becomes inactive and forms a benign protein complex. The antitoxin is more susceptible to proteolytic degradation. Figure from Arcus et al. (2011).

The fast intracellular trafficking locus (*fitAB*) from the sexually transmitted pathogen *N. gonorrhoeae* is the most well characterised VapBC system with a 1 bp overlap between the two genes with *fitA* upstream of *fitB*. FitA has a DNA binding domain that binds to an IR sequence (TGCTATCA-N12-TGATAGCA) within the promoter region of the operon to allow autoregulation (Figure 1.13) (Mattison, Wilbur, So, & Brennan, 2006; Wilbur et al., 2005). The X-ray crystal structure of FitAB bound to DNA shows four FitAB heterodimers bound to DNA (Figure 1.13) (Mattison et al., 2006).

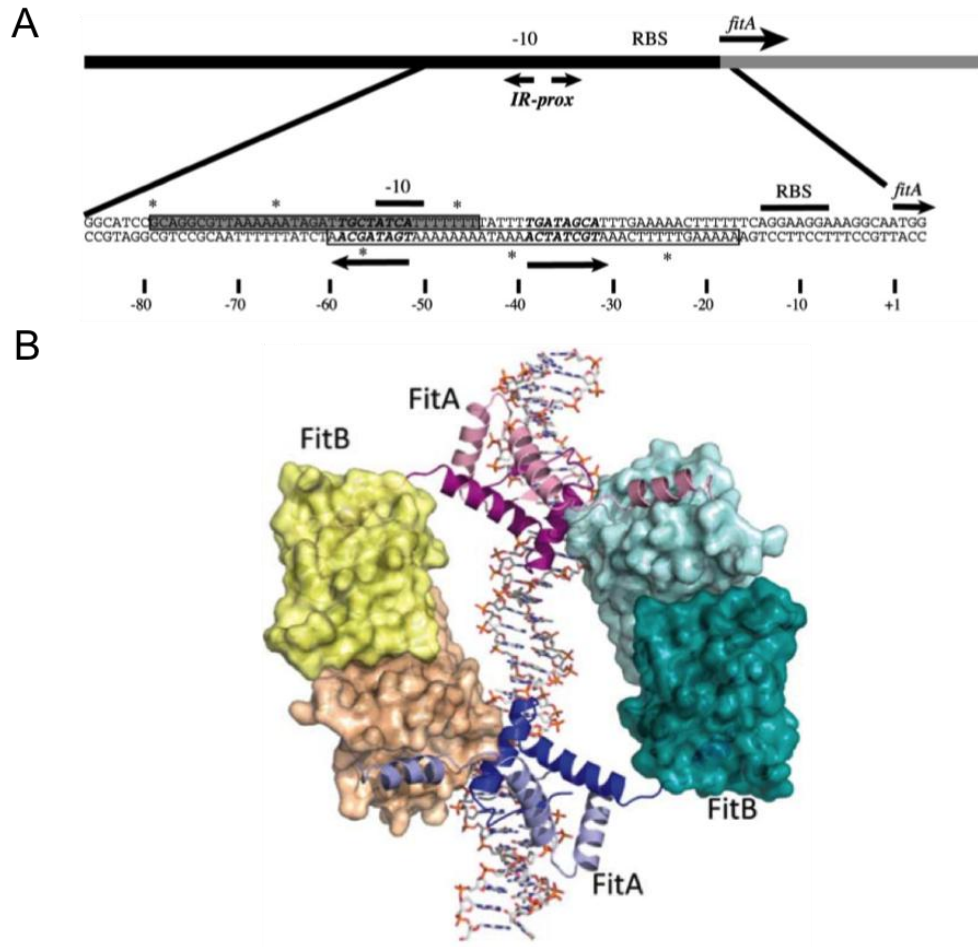


Figure 1.13. *fitAB* autoregulation. (A) Schematic of the *fitAB* promoter region that is autoregulated by FitAB. The IR sequence that FitA binds to is shown upstream of the FitA start codon and is highlights in bold. Figure from Wilbur et al. (2005). (B) The structure of FitAB from *N. gonorrhoeae* bound to IR from its promoter DNA. At the top of the figure FitA forms a dimer (two chains coloured pink and purple). The dimer binds to one half of the DNA IR. The other FitA dimer is seen at the bottom of the figure (two chains coloured blue and light blue) binding to the other half of the DNA IR. Each FitA monomer binds to a FitB monomer-the two FitB dimers lie to the left and right of the DNA (FitB chains are coloured clockwise from top-right - light green, dark green, tan and yellow). Double stranded DNA lies behind the hetero-octomeric FitAB structure. Figure from Arcus et al. (2011).

FitA binds to this sequence with weak affinity, which is improved when FitA forms a complex with FitB (Wilbur et al., 2005). FitAB binds to its promoter region when *N. gonorrhoeae* is extracellular to represses transcription of the *fitAB* operon. Upon invasion into epithelial cells, FitAB is released from the DNA and the complex dissociates to release FitB. Active FitB then slows replication by inhibiting translation via specific mRNA degradation (Mattison et al., 2006).

1.3.3.3 VapC Endonuclease Function to Inhibit Translation

The VapC mode of action is to inhibit translation by specific mRNA degradation, resulting in reversible growth arrest. VapC proteins belong to the PilT N-terminal domain (PIN domain) family of proteins. PIN domain containing proteins are associated with RNA degradation by virtue of their ribonuclease activity (Anantharaman & Aravind, 2003; Clissold & Ponting, 2000; Fatica, Tollervey, & Dlakić, 2004; Lamanna & Karbstein, 2009). The PIN domain contains four conserved acidic residues (Asp₅, Glu₄₂, Asp₁₀₄ and Asp₁₃₃) within an RNase-H-like fold, which form a negatively charged active site to bind a divalent cation (Mg²⁺ or Mn²⁺) and facilitate metal-dependent nuclease activity (Figure 1.14) (Arcus et al., 2004; Arcus et al., 2011; Mattison et al., 2006; Miallau et al., 2009).

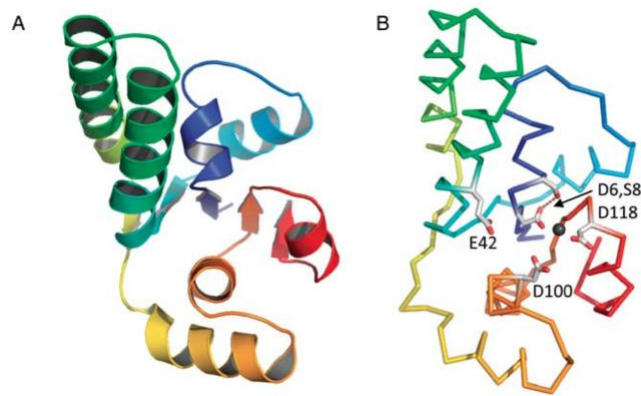


Figure 1.14. Generalized PIN-domain structure. (A) A cartoon representation of a PIN-domain structure coloured blue-red from N-terminus to C-terminus with the central 5-stranded parallel β sheet typical of a PIN-domain. (B) The same structure as a C- α trace showing the four conserved acidic residues and Mn²⁺ (black) in the active site. Figure from Arcus et al. (2011).

VapC is inactivated by direct interaction with VapB as shown by the structure of the FitAB complex (Figure 1.13). Nucleic acid cleavage by FitB is inhibited by binding of the FitA C-terminal Arg₆₈ into the active site of FitB where it interacts with three of the acidic residues (Asp₅, Glu₄₂, and Asp₁₀₄), forming strong electrostatic interactions that are difficult to disrupt (Figure 1.15). The complex is further stabilised by the hydrophobic interaction between the C-terminal FitA helix α 3 and the FitB helices α 1, α 2 and α 4 (Figure 1.15).

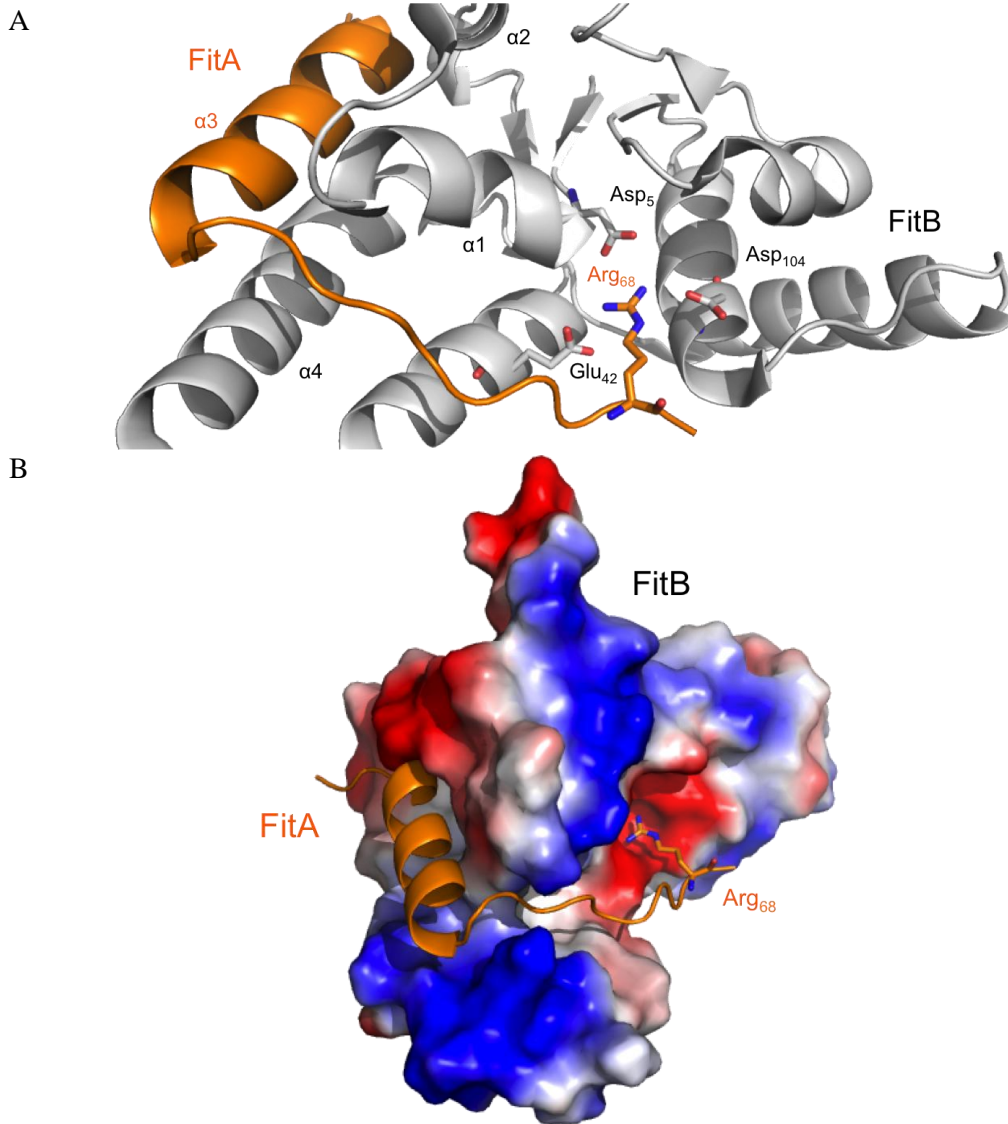


Figure 1.15. X-ray structures showing FitA inactivation of FitB. (A) The C-terminal Arg₆₈ from FitA (monomer peptide shown in orange with Arg₆₈ depicted as a stick and coloured by atom) binds into the active site of the FitB monomer (shown in grey with acidic active site residues depicted as a stick and coloured by atom) (B) FitB depicted as an electrostatic surface to show the acidic properties of the active site that Arg₆₈ interacts with and the hydrophobic region formed within FitB helices $\alpha 1$, $\alpha 2$ and $\alpha 4$ that the FitA helix $\alpha 3$ interacts with to stabilise the complex. PDB co-ordinates from Mattison et al. (2006) (PDB ID 2H10). Figures were drawn using PyMol (Delano, 2002).

VapB is susceptible to degradation by cellular proteases, which is induced under stress. This causes the cellular levels of antitoxin to drop, subsequently inducing activation of the toxin. The mode of action of VapBC indicates that they play an important role in growth regulation in response to changing environments (Arcus et al., 2005; Arcus et al., 2011; Daines, Wu, & Yuan, 2007; Mattison et al., 2006; Puskás et al., 2004; Robson, McKenzie, Cursons, Cook, & Arcus, 2009).

1.3.3.4 Functions of Characterised VapBC Systems

Several VapBC systems from a range of bacteria have been shown to inhibit translation, and, in some cases, induce growth arrest under stress conditions or when overexpressed *in vivo*. However, cellular targets and specific biological roles of these systems are yet to be elucidated (Arcus et al., 2011; Bodogai et al., 2006; Cooper, Daugherty, Tachdjian, Blum, & Kelly, 2009; Daines et al., 2007; Ning, Jiang, Liu, & Xu, 2013a; Sharp et al., 2012; Winther & Gerdes, 2009; Y. X. Zhang et al., 2004).

As mentioned earlier, FitAB in *N. gonorrhoea* is hypothesised to slow intracellular trafficking and replication during host infection (Mattison et al., 2006). The mechanism involves FitAB binding to its promoter region when *N. gonorrhoeae* is extracellular to repress transcription of its operon. Upon invasion into epithelial cells, FitAB is released from the DNA and the complex dissociates to release FitB. Active FitB then slows replication by inhibiting translation via specific mRNA degradation (Mattison et al., 2006).

The nitrogen-fixing symbiont of legumes *Sinorhizobium meliloti* carries a large number of TA modules in its genome. One of these VapBC systems, NtrPR, has been implicated in the regulation of a range of metabolic transcripts, and, more specifically, is hypothesised to adjust metabolic processes during symbiosis and other stress conditions (Puskás et al., 2004). Another VapBC system (VapBC-5) was shown to affect efficiency of symbiosis and nitrogen fixation capacity (Cinege, Bodogai, Oláh, & Kiers, 2014).

M. smegmatis is a relative of the pathogenic bacteria *M. tuberculosis*. Because of its fast growing and non-pathogenic physiology, it is a valuable tool for investigating mycobacterial metabolism. As a result, the sole VapBC system (VapBC_{1283/4}) has been well characterised. VapC₁₂₈₄ functions as a metal dependent endoribonuclease that specifically binds and cleaves single stranded RNA (ssRNA) at the cut sites AUA(U/A) (McKenzie et al., 2012b). This cut site was found to be abundant in *M. smegmatis* mRNA transcripts that were downregulated in the response to VapC overexpression in microarray experiments. A large proportion of these downregulated genes encode proteins

Chapter One

responsible for carbon transport and metabolism indicating that VapBC_{1283/4} functions as a post-transcriptional regulator of carbon metabolism (McKenzie et al., 2012b).

M. tuberculosis has 47 *vapBC* operons encoded in their genome. Of the systems that have been characterised, all VapC proteins inhibit translation by endoribonuclease activity. In some cases, translation inhibition resulted in reversible growth arrest (Arcus et al., 2011; Ramage, Connolly, & Cox, 2009; Sharp et al., 2012). All functional VapBC systems monitored *in vivo* under stress conditions, in both culture and during infection of macrophages, showed that each condition induced different VapBC systems (Ramage et al., 2009). Also, no cross talk between VapBC systems was identified (Ramage et al., 2009). Although the role of these VapBC systems remains largely unknown, these findings suggest that each system may play an individual role in response to different environmental cues, suggesting they function as an intricate response system during adaptation. For organisms that host a large number of VapBC systems, this would provide the capacity for efficient adaption to fluctuating and/or inhospitable conditions.

1.3.4 Cyanobacteria VapBC TA Systems

Cyanobacteria are capable of autotrophy via oxygenic photosynthesis and are the most important photosynthetic bacterial phylum. They are primary producers by nature of their autotrophy, which is essential for the existence of all higher organisms (Herrero, 2008). They produce organic carbon from CO₂, water, and light energy with gaseous oxygen being generated as the by-product of water splitting (Herrero, 2008). The biological production of oxygen is hypothesised to have converted the atmosphere of early Earth from an anaerobic, inhospitable environment to an atmosphere that supported aerobic respiration (Dismukes et al., 2001; Herrero, 2008). Aerobic respiration is the most efficient biological process for generating energy from organic carbon and is dependent on oxygen. Aerobic respiration supports higher metabolic demands and subsequently more complex and larger organisms (Herrero, 2008). Cyanobacteria have been utilized for industrial purposes due to their ability to generate renewable energy, cyanobacterial based fuels, ethanol, medicinal extracts and health supplements

Chapter One

(Deng & Coleman, 1999; Deng & Chow, 2010; Ostensvik, Skulberg, Underdal, & Hormazabal, 1998; Quintana, Van der Kooy, Van de Rhee, Voshol, & Verpoorte, 2011; Rittmann, 2008). It is obvious why understanding the physiology of cyanobacteria has become one of the largest fields in biology.

Members of the cyanobacteria phylum can be found in almost every terrestrial and aquatic habitat. Aquatic cyanobacteria can be found in both marine and freshwater systems, and terrestrial cyanobacteria form biofilms that can be found anywhere from rich temperate soils to rocks in the deserts of Antarctica (Thajuddin & Subramanian, 2010). Aquatic cyanobacteria can reproduce explosively in certain conditions and are known for their extensive blooms. These blooms are detrimental to other species in close proximity and can often be toxic to humans and animals because of the cyanotoxins produced (Briand, Jacquet, Bernard, & Humbert, 2003; Falconer & Humpage, 2005; Paerl, Fulton, Moisander, & Dyble, 2001). Blooms often result in closure of recreational waters and the treatments used to eliminate the blooms are expensive and can be fatal to livestock and other aquatic organisms (Frangeul et al., 2008; Steffensen, 2008).

The toxic bloom forming cyanobacteria *M. aeruginosa* has an overrepresentation of TA systems in its genome, compared to its non-toxic relatives. For example the *M. aeruginosa* genome encodes 113 TA systems, where its non-toxic relative *Synechocystis Sp. PCC6803* (*PCC8603*) has only 19 (Toxin-Antitoxin Database). Of the 113 TA systems encoded in the *M. aeruginosa* genome, a large proportion of the TA loci (34) are annotated as *vapBC* operons (Arcus et al., 2011). In comparison, *PCC6803* has only 12 *vapBC* loci (Kaneko et al., 2001; Leplae et al., 2011; Makarova et al., 2009). This expansion suggests a functional link between toxicity and VapBC systems, however the function of VapBC systems in cyanobacteria is not known. In fact, only one of the 12 *PCC6803* VapBC systems has been partially characterised, and none from any toxic-forming species (Ning, Liu, Xu, Zhuang, Wen, et al., 2013b).

1.3.4.1 *Synechocystis* VapBC Systems

The *Synechocystis* genome was the third prokaryote and first photosynthetic genome to be sequenced (Kaneko et al., 1995) and is used as the

Chapter One

model organism to investigate the physiology of cyanobacteria. *Synechocystis* has 12 annotated *vapBC* loci and until recently none had been characterised. Ning et al., (2013) investigated the VapBC10 (*ssr2962/slr1767*) system of *Synechocystis* using *E. coli* as the model organism and were able to confirm that the operon functions as a typical VapBC system. They showed *vapB10* and *vapC10* form a bicistronic operon and both genes are co-transcribed under normal growth conditions. Overexpression of VapBC10 had no effect on growth but overexpression of VapC10 caused growth arrest in *E. coli*, which could be reversed by introducing overexpressing VapB10. Because the VapC10 shows the characteristic bacteriostatic phenotype of a VapC toxin it is proposed that VapC10 exerts its growth arrest via translation inhibition, like other characterised VapC toxins, however this has not been confirmed experimentally.

VapBC systems are normally characterised by their repressive autoregulation where VapB binds to the promoter region of the operon and represses expression. The efficiency of binding is enhanced when VapC is bound to VapB and it is proposed that under normal growth conditions the toxin and antitoxin are co-expressed to form a non-toxic TA complex which represses TA expression. Under specific stress conditions this repression is lifted by the degradation of VapB, which results in activation of VapC (Gerdes et al., 2005). This is not the case for *Synechocystis* VapBC10, which shows an alternative mechanism of autoregulation.

Ning et al. (2013) showed that VapB10 binds to the promoter region of its own operon and positively autoregulates transcription. They identified the specific DNA sequence that VapB10 recognises: the imperfect IR sequence TTTCCCT-2N-AGGGTAA present in the promoter region upstream of the *ssr2962/slr1767* operon (Figure 1.16). They also showed VapC10 is unable to bind to the promoter region and actually abolishes binding of VapB10 to DNA. These results reveal that VapB10 can positively autoregulate the operon, which can be inhibited by VapBC10.



Figure 1.16. Schematic representation of the genetic structure of the *Synechocystis* VapBC10 operon. The imperfect IR sequence that VapB binds to is shown upstream of the VapBC10 operon within the promoter region. Figure from Ning et al. (2013).

Because this is the only VapBC system to be characterised in cyanobacteria, and the functional mechanism is still unknown, it is difficult to draw any conclusions about the role that these VapBC systems play in growth regulation. However, because of the link between the number of VapBC systems and the toxicity of some organisms it is a very promising avenue to investigate. Obtaining a better understanding of the role that VapBC systems play may also provide a novel strategy for bloom control.

1.4 Objectives of this Thesis

This thesis will investigate bacterial regulation at both (I) the transcription level by the structural and functional characterisation of the transcription factor AmtR in the actinomycetes *M. smegmatis* and (II) the post-transcription level with preliminary characterisation of three VapBC systems present in the cyanobacteria *Synechocystis* and one VapBC system from the toxic bloom forming *M. aeruginosa*. The *M. smegmatis* AmtR was initially investigated as a putative regulator of the *M. smegmatis* VapBC_{1283/4} system, with initial characterisation by X-ray crystallography. Functional characterisation of AmtR revealed that it has no regulatory control over VapBC_{1283/4}, however it does play an instrumental role in regulation of alternative nitrogen metabolism in response to nitrogen limitation. The cyanobacteria VapBC systems investigated in this thesis were previously unexplored and through a combination of *in vitro* and *in vivo* analyses I present the preliminary characterisation and putative function of these poorly understood systems.

2 Characterisation of *Mycobacterium smegmatis*

Transcription Factor AmtR

2.1 Introduction

Nitrogen and carbon are essential for bacterial growth and survival, and, as a consequence, bacteria have evolved processes for efficient uptake and metabolism of these nutrients in response to their environment. Soil bacteria are subjected to environments with consistent nitrogen limitation and carbon surplus. This has implications for the necessity of regulation and diversification of nitrogen metabolism in these organisms. Transcription factors play an essential role in these processes.

A well studied example of transcriptional regulation of nitrogen metabolism in bacteria is the global nitrogen regulatory system (NtrBC), which is responsible for response to nitrogen limitation (see Figure 1.4 in Chapter 1) (Arcondéguy et al., 2001; Chubukov et al., 2014 Morett & Segovia, 1993). NtrC is activated by phosphorylation in response to nitrogen limitation, which results in the upregulation of its regulon (Austin & Dixon, 1992; Kern et al., 1999). The phosphorylation of NtrC is dependent on the P_{II} proteins GlnB and GlnK (Arcondéguy et al., 2001; Atkinson & Ninfa, 1998; Kamberov et al., 1995). The NtrC regulon is composed of genes that encode for ammonium transport and assimilation, alternative nitrogen uptake and metabolism, and nitrogen fixation (Arcondéguy et al., 2001; Atkinson & Ninfa, 1998; Claverie-Martin & Magasanik, 1991; Schneider et al., 1998).

In bacteria that do not have an NtrBC system, including a number of actinobacteria, the global transcription factors GlnR, TnR and AmtR function as the major players in nitrogen-limiting response (Arcondéguy et al., 2001; Jakoby et al., 2000; Schreier et al., 1989; Tiffert et al., 2008). The regulation of nitrogen metabolism is very diverse in actinomycetes and has been intensively studied in the organisms *Corynebacterium glutamicum*, *Streptomyces coelicolor* and *Streptomyces avermitilis* (Chen et al., 2013; Jakoby et al., 2000; Tiffert et al., 2008).

In *C. glutamicum* the response to nitrogen limitation is regulated via the global nitrogen transcription regulator AmtR, which acts as a repressor of at least 35 genes under nitrogen surplus (Beckers et al., 2005; Jakoby et al., 2000). AmtR directly interacts with the P_{II} protein GlnK, under nitrogen limiting conditions, leading to the release of AmtR from its target DNA and upregulation of its regulon (Beckers et al., 2005). The *C. glutamicum* AmtR regulon is composed of genes that encode proteins responsible for responding to nitrogen limitation, including ammonia uptake and assimilation, and alternative nitrogen uptake and metabolism (Beckers et al., 2001; 2004; Bendt et al., 2004; Jakoby et al., 2000; Muhl et al., 2009; Nolden et al., 2001a). This system of nitrogen regulation has been observed only in *C. glutamicum* and introduces a new player to the intricate network of prokaryotic nitrogen regulation. In *S. coelicolor*, the transcription of genes involved in response to nitrogen limitation is mediated by the transcriptional regulator GlnR (Tiffert et al., 2008). *S. avermitilis* contains both GlnR and AmtR, and GlnR is suggested to work as a crucial nitrogen regulator, although functional studies have not confirmed this (Chen et al., 2013). AmtR has been identified as a regulator involved in alternative urea metabolism and amino acid transport, playing an important role in nitrogen metabolism (Chen et al., 2013).

Although these studies have provided insight into nitrogen regulation of members of the actinobacteria phylum, there is limited understanding of nitrogen regulation in the genus mycobacteria. *Mycobacterium smegmatis* is a soil bacteria that belongs to the fast-growing mycobacterial group and is commonly used as a model organism to study pathogenic species like *M. tuberculosis*. The preferred nitrogen source of *M. smegmatis* is ammonium but it is capable of using different nitrogenous compounds under changing environmental conditions. This scavenging ability allows *M. smegmatis* to utilize a broader range of nitrogenous sources compared to its pathogenic relatives (Amon, Titgemeyer, & Burkovski, 2009).

M. smegmatis lacks a NtrBC system for global nitrogen response and, like *S. avermitilis*, encodes both GlnR and AmtR proteins. GlnR has been identified as an important nitrogen regulator, as it controls the genes that encode ammonia

uptake and assimilation, alternative nitrogen uptake and metabolism, and the signal transduction P_{II} proteins (Amon et al., 2008; Fink et al., 2002; Jenkins et al., 2012; 2013; Jessberger et al., 2012). A number of genes that encode proteins involved in critical nitrogen metabolic processes are not under the control of GlnR and it has been suggested that these genes are under control of AmtR (Amon et al., 2008).

Bioinformatic analyses of the *M. smegmatis* genome revealed a number of genes homologous to the *S. avermitilis* regulon that are putatively under AmtR control (Jessberger et al., 2012). This was confirmed by transcriptome analysis of the *M. smegmatis amtR* deletion strain, which also showed that AmtR repression is lifted only under nitrogen limitation. Co-factor independent binding of AmtR to the upstream promoter region of the genes was also confirmed and no putative signalling molecules were identified (Jessberger et al., 2012). This study indicates that the *M. smegmatis* AmtR controls a regulon of genes involved in alternative nitrogen uptake and metabolism, and the expression of this regulon is induced under nitrogen limiting conditions when AmtR repression is lifted. Like GlnR, the signal to lift repression in mycobacteria is not known. Elucidating the signal for these transcription factors to lift repression is critical to improve our understanding of nitrogen metabolic regulation.

These initial findings suggest *M. smegmatis* AmtR may have a unique role in metabolic adaptation to nitrogen limitation, and further investigation into the structure and function of AmtR will provide greater understanding of the regulatory mechanisms functioning in mycobacteria. In this chapter, I present my work on the structural and functional characterisation of the *M. smegmatis* AmtR. Firstly, I have solved the X-ray crystallographic structure of both the truncated C-terminal regulatory domain of AmtR and the full structure including both the C-terminal regulatory domain and the N-terminal DNA binding domain. This is the first AmtR structure to be solved: it reveals the regulatory domain contains a deep binding pocket and the DNA binding domain is defined by a helix-turn-helix (HTH) motif. Secondly, in collaboration with the Cook Laboratory at Otago University, we have confirmed the *M. smegmatis* AmtR regulon and determined the DNA recognition sequence. Subsequently, I have determined a modified actinobacteria AmtR binding motif using representative actinobacteria AmtR

Chapter Two

DNA recognition sequences and mutation work done in this thesis. Finally, we have identified two potential signalling molecules specific to AmtR that appear to function via a co-repressor mechanism of regulation that has not been observed in the TetR protein family until now.

2.2 Methods

2.2.1 DNA Manipulation and Cloning

2.2.1.1 Gene, Primer and Oligonucleotide Design

2.2.1.1.1 *M. smegmatis* *amtR* Gene Design

The *M. smegmatis* *amtR* gene (msmeg_4300) was ordered from GeneArt (Life Technologies, Carlsbad CA) with a 5' NcoI restriction site and a 3' HindIII restriction site with the stop codon removed to allow for ligation into pYUB28b with a C-terminal His tag. The *amtR* gene was extracted from the GeneArt plasmid by restriction digest as per Section 2.2.1.6 using NcoI and HindIII and inserted into the pYUB28b expression vector (Bashiri, Rehan, Greenwood, Dickson, & Baker, 2010). The vector was transformed into *E. coli* TOP10 cells prior to transformation and protein expression in the *M. smegmatis* 4517 expression strain.

2.2.1.1.2 *M. smegmatis* Primer Design

Primers for amplification of the *glnK* gene (msmeg_2426) from the *M. smegmatis* mc²155 genomic DNA for cloning were initially designed to include the 5' restriction site NdeI and the 3' restriction site HindIII necessary for ligation into pYUB28b with a N-terminal His tag. An extra 3-6 bases were included to allow for efficient cleavage, depending on the restriction enzyme used. These primers failed to amplify *glnK* so a new set of primers were made without the restriction enzyme sites, which successfully amplified *glnK*. The PCR product of the second set of primers was used as the template for PCR with the first set of primers to add the restriction sites to the *glnK* gene. The gene was then digested as per Section 2.2.1.6 with NdeI and HindIII and inserted into the pYUB28b expression vector. The vector was then transformed into *E. coli* TOP10 cells prior to transformation and protein expression in the *M. smegmatis* 4517 expression strain. All putative sequences were cloned in silico into the pYUB28b vector using Geneious R6 to ensure the protein sequence was correct and the His tags were in frame.

Primers for generation of DNA fragments for electrophoretic mobility shift assays (EMSA) were designed to amplify specific regions from the promoter sites of putative AmtR regulated genes in the *M. smegmatis* genome. Primers were also designed to amplify a non-specific region of the genome that contained no putative binding site, which were used as the non-specific control. All primers were supplied by IDT (Coralville IA) and resuspended in 1 x TE.

2.2.1.1.3 *M. smegmatis* Oligonucleotide Design

DNA oligonucleotides required for AmtR DNA EMSA assays and AmtR:DNA co-crystallisation trials were designed as complementary single stranded sequences of recognition regions, or variations of recognition regions, from the promoter sites of putative AmtR regulated genes. All oligonucleotides were supplied by IDT and resuspended in 1 x TE. The oligonucleotides were annealed by mixing with annealing buffer (1 x TE + 150 mM NaCl) in a 1:4 annealing buffer to DNA ratio. The mixture was incubated at 95 °C for 10 minutes and brought down to room temperature to incubate overnight. DNA quantification was conducted as per Section 2.2.1.5.

For recognition regions that were too large to order as oligonucleotides, primers were designed (Section 2.2.1.1.2). The recognition regions were PCR amplified from the *M. smegmatis* mc²155 genome as per Section 2.2.1.3.1 with subsequent agarose gel electrophoresis analysis (Section 2.2.1.4.1), and purification and quantification as per Section 2.2.1.5.

A list of all genes, primers and oligonucleotides ordered are presented in Table A.1 in the Appendices.

2.2.1.2 DNA Digoxigenin-II-ddUTP (DIG) Labelling for EMSA Assays

DNA was 3'-end labelled with digoxigenin-11-ddUTP (DIG) using the DIG Gel Shift kit, 2nd Generation (Roche, Germany) as per manufactures instructions. The labelling reaction was as follows:

Chapter Two

100ng	DNA
1 x	labelling buffer
1.25 mM	CoCl ₂
1 mM	DIG ddUTPs
400 U	Terminal transferase

The reaction mixture was incubated at 37 °C for 15 min and enzyme activity was stopped by the addition of 0.11 mM EDTA (pH 8). DNA was diluted to 0.4 ng/μl for all gel shift assays.

2.2.1.3 Polymerase Chain Reactions (PCR)

2.2.1.3.1 PCR Amplification from *M. smegmatis* mc²155 Genomic DNA

PCR reactions were carried out with the proofreading enzyme KAPA HiFi DNA polymerase (Kapa Biosystems, Wilmington MA). The annealing temperature was 60 °C unless otherwise stated. PCR reactions were carried out in 25 μl volumes using the following concentrations: 1 x High GC amplification buffer, 0.3 mM deoxynucleotide mix (dATP, dCTP, dTTP & dGTP), 2 mM MgSO₄, 5% DMSO, 1 U 2G DNA polymerase, 0.3 μM of each primer and 10 ng template DNA (*M. smegmatis* mc²155 genomic DNA). The following cycling conditions were used for amplification:

95 °C	2.00	(min:sec)
98 °C	0.20	x 29
55 °C	0.15	
72 °C	0.45	
72 °C	5.00	

2.2.1.3.2 PCR Screening of pYUB28b + *amtR* Transformants

pYUB28b plasmids purified from *E. coli* (Section 2.2.1.11.), were checked for the presence of the *amtR* gene insert prior to sequencing by performing a PCR using T7 primers. PCR reactions were carried out with *Taq* DNA polymerase (Invitrogen, Carlsbad CA) in 25 μl volumes using the following concentrations: 1 x PCR buffer, 0.2 mM dNTP mix (dATP, dCTP, dTTP & dGTP), 1.5 mM

Chapter Two

MgSO₄, 2.5 U *Taq* DNA polymerase, 1 µM of each primer and 30-60 ng template DNA. The following cycling conditions were used for amplification:

94 °C	2.00	(min:sec)
94 °C	0.30	x 29
55 °C	0.30	
72 °C	0.45	
72 °C	5.00	

PCR reactions were visualised as in Section 2.2.1.4 below.

2.2.1.4 DNA Gel Electrophoresis

2.2.1.4.1 Agarose Gel Electrophoresis

DNA fragments were separated via agarose gel electrophoresis. Samples were mixed with 10 x DNA loading dye prior to loading, and run on a 1% TAE gel. Agarose gels were made with 1x SYBR safeTM DNA gel stain (Invitrogen). Gels were visualised by blue light and band sizes were determined by visual comparison against a 1kb-Plus DNA ladder (Invitrogen).

2.2.1.4.2 Native Acrylamide TBE Gel Electrophoresis

Digoxigenin-11-ddUTP (DIG) labelled DNA from EMSA was run on 6 % native acrylamide (19:1) TBE (Tris-borate-EDTA) gels. Gels were pre-run at 70 V for 30 min in 1 x TBE before loading DNA. Gels were then run at 150 V for 45 min-1 hr until the dye front reached the bottom of the gel.

2.2.1.5 DNA Purification and Quantification

DNA from agarose gels was separated and visualised with SYBR safeTM DNA dye as described in Section 2.2.1.4.1. The desired bands were extracted using a sterile scalpel blade and DNA purified using the QIAquick Gel Extraction Kit (Qiagen, Netherlands) according to the manufacturer's instructions. DNA was eluted in 30 µl of elution buffer. PCR products from solution were purified using a QIAquick PCR Product Purification Kit (Qiagen) according to the manufacturer's instructions. DNA purity and concentration was estimated by measuring A₂₆₀/A₂₃₀

Chapter Two

nm and A_{260}/A_{280} nm ratios using a NanoDrop®ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham MA).

2.2.1.6 Restriction Enzyme Digestion

Restriction enzymes were purchased from Invitrogen, and digests were performed on purified PCR products or plasmid DNA either singly, due to buffer incompatibilities, or as double digests if buffers were compatible. 1 μ l of restriction enzyme was added to 28-38 μ l of DNA along with the appropriate amount of buffer, BSA and water for a 40-100 μ l reaction according to the manufacturer's instructions. Digests were incubated at 37 °C for 3 hrs, with subsequent purification as per Section 2.2.1.5.

2.2.1.7 DNA Ligation

DNA ligations were set up using 1 U of T4 DNA ligase (Invitrogen), and a vector to insert molar ratio of 1:3, where possible, in a total volume of 10-50 μ l. Ligations were incubated overnight at 14 °C.

2.2.1.8 Preparation of Electrocompetent *E. coli* TOP10 Cells

A glycerol stock of *E. coli* TOP10 cells was streaked onto a low salt LB agar plate and incubated at 37 °C overnight. A 10 ml starter culture of low salt LB was inoculated with a single colony and grown overnight at 37 °C, shaking at 200 rpm. The overnight culture was used to inoculate 1 L of low salt LB, which was grown in the same conditions as the starter culture until it reached OD_{600} 0.5-0.7. All subsequent steps were performed with ice-cold 10 % glycerol, in pre-chilled, sterile containers. The cells were chilled on ice for at least 20 min, centrifuged (4,000 g, 15 min, 4 °C), the supernatant removed and the pellet resuspended in 500 ml 10 % glycerol. The spin and wash step was repeated three further times, resuspending in 250 ml, 20 ml then 2 ml respectively. 50 μ l aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

2.2.1.9 Electroporation of *E. coli*

Ligation reactions (10-50 μ l) or plasmid DNA (1 μ l) was added to 50 μ l freshly thawed electrocompetent *E. coli* TOP10 cells. The mixture was placed in a 2 mm electroporation cuvette (BioRad Laboratories, Hercules CA) and electroporated with a BioRad Gene PulserTM (BioRad Laboratories) at 2.5 kV, 25 μ F capacitance and 200 Ω resistance. Transformed cells were recovered immediately in 0.9 ml of SOC media and incubated at 37 °C, shaking at 200 rpm for 30 min. Aliquots of the cells (50 & 100 μ l) were plated into low salt LB agar plates containing hygromycin (50 μ g/ml) for selection and incubated at 37 °C overnight. Single colonies from these plates were used to inoculate 5 ml low salt LB cultures containing hygromycin and incubated at 37 °C, shaking at 200 rpm overnight. These cultures were used to isolate plasmid DNA according to the method in Section 2.2.1.11 below and to make glycerol stocks as per Section 2.2.1.15.

2.2.1.10 Cloning of *amtR* and *glnK* into pYUB28b

pYUB28b is an *E. coli*-mycobacterial shuttle vector with hygromycin resistance. The *M. smegmatis amtR* and *glnK* genes were digested (Section 2.2.1.6), purified (Section 2.2.1.5.) and ligated (Section 2.2.1.7) into pYUB28b between the selected restriction sites within the multiple cloning site (MCS), allowing for the appropriate N- or C-terminal His tag. Vectors containing the inserted gene were transformed into electrocompetent *E. coli* TOP10 cells (Section 2.2.1.9) and plated onto low salt LB agar plates containing hygromycin for selection. The plasmid constructs were purified then sequenced (Section 2.2.1.11 and Section 2.2.1.12) to confirm correct insertion before being transformed into electrocompetent *M. smegmatis* 4517 (Section 2.2.1.14).

2.2.1.11 *E. coli* Plasmid DNA Extraction

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

2.2.1.12 DNA Sequencing

DNA was sequenced by Massey Genome Service (MGS) using Sanger sequencing on an ABI3730 DNA Analyzer (Applied Biosystems, Foster City CA) at Massey University.

2.2.1.13 Preparation of Electrocompetent *M. smegmatis* 4517 Cells

A glycerol stock of *M. smegmatis* mc² 4517 cells was streaked onto a low salt LB agar plate supplemented with 0.05 % (w/v) tween 80 (LBT) + kanamycin (50 µg/ml) and incubated in a humidity chamber at 37 °C for three days. A culture of 5 ml low salt LBT + kanamycin was inoculated with a single colony and grown for 48 hrs at 37 °C at 200 rpm. The 5 ml culture was used to inoculate 500 ml of 7H9 media + ADC + Tween 80 + Kanamycin, which was grown at 37 °C at 200 rpm until it reached an OD₆₀₀ of 0.6-0.7. All subsequent steps were performed with ice-cold 10 % glycerol, in pre-chilled, sterile 50 ml falcon tubes. The cells were chilled on ice for 1.5 hrs, centrifuged (4,000 g, 20 min, 4 °C), the supernatant removed and the pellet resuspended in 50 ml 10 % glycerol. The spin and wash step was repeated two more times, resuspending in 50 ml then 5 ml respectively. The cells were pooled and transferred to a single 50 ml falcon tube and centrifuged (as above) one final time before resuspending in 500 µl 10 % glycerol, flash freezing 40 µl aliquots in liquid nitrogen and storing at -80 °C.

2.2.1.14 Electroporation of *M. smegmatis*

Plasmid DNA (1 µl) was added to 40 µl freshly thawed electrocompetent *M. smegmatis* mc² 4517 cells and 40 µl of sterile 10 % glycerol. The mixture was placed in a 2 mm electroporation cuvette (BioRad Laboratories) and electroporated with a BioRad Gene PulserTM (BioRad Laboratories) at 2.5 kV, 25 µF capacitance and 1000 Ω resistance. Transformed cells were recovered immediately in 0.9 ml of 7H9 media and incubated at 37 °C, shaking at 200 rpm for 3 hrs. Aliquots of the cells (50 µl and 100 µl) were plated onto 7H10 + ADC + Tween 80 agar plates containing hygromycin (50 µg/ml) and kanamycin (50 µg/ml) for selection and incubated at 37 °C for three days in a humidity chamber. These cultures were used to make glycerol stocks as per Section 2.2.1.15 and to conduct small scale expression trials (Section 2.2.2.6).

2.2.1.15 Glycerol Stocks

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

2.2.2 Protein Expression and Purification

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

SDS-PAGE gels were cast in a Hoefer mini gel casting system. SDS-PAGE gels consisted of a 5 % acrylamide stacking gel overlaid on a 16.5 % acrylamide resolving gel. All SDS-PAGE gels were made up with 30 % acrylamide with an acrylamide:bisacrylamide ratio of 37.5:1 (v/v) (BioRad Laboratories), including 0.1 % filtered SDS, and were polymerised by the addition of 0.05 % ammonium persulphate (APS) and TEMED.

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

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

12 % Native-PAGE gels were cast and run as for SDS-PAGE gels (Section 2.2.2.1), with the exception of no SDS in the gel, running buffer or loading dye.

2.2.2.3 Coomassie Blue Stain for Protein Gel Electrophoresis

Gels were stained by colloidal coomassie staining using the quick stain method (Wong, Sridhara, Bardwell, & Jakob, 2000) with Fairbanks staining

Chapter Two

solutions A-D. Gels were microwaved for 30 s in 50 ml Fairbanks staining solution A, cooled to room temperature while shaking gently, the stain removed, and then repeated as above with stains B, C and lastly D (de-stain).

2.2.2.4 Quantification of Protein Concentration

A NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific) was used to measure absorbance at 280 nm and the Beer-Lambert equation ($A = \epsilon \cdot c \cdot l$) applied to correlate absorbance with concentration. Here A = absorbance at 280 nm, ϵ = the theoretical extinction coefficient of the protein, c = concentration, and l = pathlength. Theoretical extinction coefficients were calculated based on the amino acid composition using Geneious Pro (Version 6.1.7 Biomatters Ltd, NZ) and the ExPASy ProtParam tool (Wilkins et al., 1998).

2.2.2.5 Concentration of Protein Samples

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

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

All *M. smegmatis* expression cultures included kanamycin and hygromycin at 50 μ g/ml for selection. Expression of *M. smegmatis* AmtR and GlnK were initially confirmed with small scale His tag binding tests using Ni Sepharose beads. A single transformed colony either from the freshly transformed plate or streaked from a glycerol stock was used to inoculate a 5 ml defined media PA-0.5G + 0.05 % Tween 80 seeder culture. This culture was grown for 48 hr at 37 °C, shaking at 200 rpm, then 100 μ l was used to seed a 10 ml ZYP-5052 + 0.05 % Tween 80 expression culture, which was grown for 90 hr at 37 °C, shaking at 200 rpm. Cultures were centrifuged (4,600 g, 20 min, 4 °C) and cell pellets were resuspended in 1 ml lysis buffer (50 mM phosphate buffer, 200 mM NaCl, 20 mM imidazole, pH 7.4) before 5 μ l was removed and added to 10 μ l lysis buffer and 5 μ l 4 x SDS loading dye (named whole-cell sample).

Chapter Two

The remainder of the cells were sonicated on ice using a fine tip probe for 15 s bursts at setting 3.5 with a Misonix Sonicator (Misonix Inc., Farmingdale NY) until cells were lysed. The lysate was centrifuged at 13,000 g for 20 min at 4 °C. 10 µl of the supernatant, containing the soluble protein, was added to 5 µl of lysis buffer and 5 µl of 4 x SDS loading dye (named soluble sample). The remainder of the supernatant was transferred to a 1.5 ml eppendorf tube containing 50 µl of lysis buffer pre-equilibrated Ni Sepharose beads. The pellet, containing the insoluble protein, was resuspended in 250 µl of lysis buffer and 5 µl was added to 10 µl of lysis buffer and 5 µl of 4 x SDS loading dye (named insoluble sample). The supernatant and beads were incubated at room temperature for 15 min at 1,000 rpm and then left to settle for 5 min. The beads were then centrifuged at 3,000 g for 1 min and 10 µl of the supernatant was added to 5 µl of lysis buffer and 5 µl of 4 x SDS loading dye (named flow-through sample). The remainder of the supernatant was removed. The beads were then washed with 1 ml of lysis buffer, left to settle for 5 min and then centrifuged (3,000 g for 1 min). 15 µl of the supernatant was added to 5 µl of 4 x SDS loading dye (named wash sample). The remainder of the supernatant was removed and the wash step was repeated with no aliquots taken for sampling. 50 µl of 4 x SDS loading buffer was added to the beads, then 20 µl from all six samples were analysed by SDS-PAGE.

2.2.2.7 Large Scale Protein Expression and Sonication from *M. smegmatis*

2.2.2.7.1 Large Scale Autoinduction Protein Expression

Large scale protein expression cultures were grown as for small scale protein expression cultures in Section 2.2.2.6, with the exception that 10 ml seeder cultures were used to inoculate 1 L expression cultures to maintain the 1:100 ratio. Cultures were centrifuged (4,600 rpm, 20 min, 4 °C) and cell pellets either frozen at -80 °C or resuspended in 25 ml lysis buffer (50 mM Phosphate buffer, 200 mM NaCl, 20 mM imidazole, pH 7.4) for sonication.

2.2.2.7.2 *In vivo* Large Scale Nitrogen Limiting Induced GlnK Adenylation

For *in vivo* nitrogen limiting induced adenylation of GlnK, *M. smegmatis* containing the vector pYUB28b+*glnK* was streaked out on a low salt LBT agar plate + kanamycin and hygromycin (50 µg/ml) and grown at 37 °C in a humidity

chamber for 3-5 days. A single colony from the plate was used to inoculate a 10 ml seeder culture of LBT broth + kanamycin and hygromycin (50 µg/ml). The culture was incubated at 37 °C, shaking at 200 rpm for 24 hrs. The seeder (0.2 ml) was used to inoculate a 10 ml seeder culture of modified Sauton's minimal media (Williams et al., 2013) + 30 mM NH₄Cl₂ which was incubated at 37 °C, shaking at 200 rpm for 24 hrs. The Sauton's seeder (5 ml) was used to inoculate 1 L of modified Sauton's media + 30 mM NH₄Cl₂ which was incubated at 37 °C, shaking at 200 rpm until they reached OD₆₀₀ 0.5-0.7 and were induced with 0.1 M IPTG. Cells were further incubated until they reached late log phase and harvested by centrifugation at 4,600 g for 20 min, 4 °C. The supernatant was removed and cells were washed by resuspension in modified Sauton's media without NH₄Cl₂. The wash step was repeated and cells were transferred to 1 L of modified Sauton's + 1 mM NH₄Cl₂. The cells were incubated for another 12 hours to allow all available nitrogen to be metabolised and induce GlnK adenylation as per Williams et al. (2013). Cultures were centrifuged (4,600 g, 20 min, 4 °C) and cell pellets either frozen at -80 °C or resuspended in 25 ml lysis buffer (50 mM Phosphate buffer, 200 mM NaCl, 20 mM imidazole, pH 7.4) for sonication.

All media recipes are presented in the Appendices.

2.2.2.8 *M. smegmatis* Large Scale Protein Purification

Cells were sonicated on ice using a large probe for 30 s bursts at setting 7-8.5 with 1 min rests in between using a Misonix Sonicator (Misonix Inc.) until cells were lysed (3-8 min). The lysate was centrifuged at 13,000 rpm for 20 min at 4 °C. The supernatant was filtered through 1.2 µm and 0.45 µm Minisart filters (Sartorius AG) and 5 µl was added to 10 µl lysis buffer and 5 µl 4 x SDS loading buffer for later SDS-PAGE analysis. The pellet was resuspended in approximately 10 ml of lysis buffer and 5 µl of this was added to 10 µl lysis buffer and 5 µl 4 x SDS loading buffer for later SDS-PAGE analysis. The protein in the remainder of the supernatant was initially run through a HisTrap HP column (GE Healthcare, UK) and purified via Immobilised Metal Affinity Chromatography (IMAC) on an ÄKTA FPLC system (Section 2.2.2.8.1). Fractions containing the protein of interest were purified further by size exclusion chromatography (SEC) using a Superdex column (GE Healthcare) (Section 2.2.2.8.2). Aliquots of fractions (15

Chapter Two

μl) containing the protein of interest from both IMAC and SEC were added to 5 μl of 4 x SDS loading buffer for analysis with SDS-PAGE.

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

A 5 ml HiTrap™ chelating HP column (GE Healthcare) was pre-equilibrated with lysis buffer (50 mM phosphate buffer, 200 mM NaCl, 20 mM imidazole, pH 7.4) at a flow rate of 1 ml/min. The supernatant from the post-sonicated cell lysate was loaded onto the IMAC column which was then attached to an ÄKTA Basic™ or ÄKTA Purifier™ FPLC system (GE Healthcare). Unbound proteins were removed by washing the column with 15-25 ml of lysis buffer at a flow rate of 1 ml/min, before bound protein was removed using a gradient of 0-50% elution buffer (lysis buffer + 1 M imidazole) over 50 ml at a flow rate of 1 ml/min. The IMAC column was stripped of Ni²⁺ ions using 10 ml 100 mM EDTA pH 8.0, washed with 20 ml water and re-primed with 5 ml 100 mM NiCl₂ between purifications of different proteins.

2.2.2.8.2 Size Exclusion Chromatography (SEC)

Proteins were purified by size exclusion chromatography (SEC) using either a preparative HiLoad™ 16/60 Superdex™ 200 pg (S200 16/60) or analytical Superdex™ 200 10/300 (S200 10/300) column (GE Healthcare), which have a separation range of 10-600 kDa. Proteins were concentrated to 5 ml or 0.5 ml respectively, and filtered with a Nanosep® MF 0.2 μm filter (Pall Corporation, New York NY) before loading and running on a pre-equilibrated (50 mM phosphate buffer, 200 mM NaCl, pH 7.4) column.

SEC of proteins bound to co-purified genomic DNA was performed using the S200 10/300 column as above but in high salt (2 M NaCl) buffer to dissociate the co-purified genomic DNA. The protein was combined at 1:1 volume ratio with high salt PBS buffer (4 M NaCl), concentrated down to 0.5 ml and filtered with a Nanosep® MF 0.2 μm filter (Pall Corporation) before loading and running on a pre-equilibrated (50 mM phosphate buffer, 2M NaCl, pH 7.4) column.

2.2.3 Dialysis of Protein Samples

To remove the high salt concentration and any co-purified ligands, the protein was dialysed into a low salt buffer (20 mM phosphate, 200 mM NaCl, pH 7.4) using 6-8 kDa molecular cut off Spectra Por® dialysis tubing (Spectrum Laboratories, Rancho Dominguez CA). Samples were loaded into pre-wet tubing and the ends were sealed using dialysis clips before dialysing into 1 L of gently stirring buffer. The buffer was replaced four times over a period of 15 hrs.

2.2.4 Crystallography

2.2.4.1 Initial Crystallisation Trials of AmtR

Initial crystallisation trials for AmtR employed sitting-drop vapour diffusion and were set up using the high throughput Mosquito® crystal robotic system (TTP Labtech Ltd, UK), using PEGRx HT, Crystal Screen HT, Index HT and SaltRx Ht crystallisation screens (Hampton Research, Aliso Viejo CA). Aliquots of each screen condition (100 µl) were manually pipetted into 96 well Intelli-plate reservoirs (Hampton Research) before 100 nl of protein and reservoir solutions were robotically added into the crystallisation well. All plates were sealed with ClearSeal™ Film (Hampton Research) and stored at 18 °C.

Initial trials used AmtR purified protein and subsequent trials used AmtR co-purified with up to 5 mM of a co-repressor xanthine or allantoin and 400-500 µg of a 26 bp DNA oligonucleotide containing the recognition sequence.

2.2.4.2 Optimisation of Crystallisation Conditions (Fine Screening)

Once promising crystallisation conditions had been identified, a variety of optimisation steps were employed to produce crystals of the highest possible quality suitable for X-ray analysis as follows.

2.2.4.2.1 Hanging Drops

Crystallization fine-screens employed hanging-drop vapour diffusion in 24-well plates (Hampton Research) at 18 °C. Plates were pre-greased with Glisseal N then 0.5 ml of fine screen ‘mother liquor’ (ML) was transferred into

Chapter Two

the well. In general, the fine screen ML trialled contained a variation of the original concentration of the macromolecule, precipitant, or pH. 1 μ l of protein and ML was pipetted onto a 22 mm square siliconised coverslip (Hampton Research), which was then inverted over the reservoir of ML and sealed by the grease.

2.2.4.2.2 Protein Concentration and Age Trials

A combination of AmtR concentration and age was trialled to determine the optimal conditions for quality crystal formation. Protein concentrations ranged from 5 mg/ml to 15 mg/ml and protein age ranged from 1 hr to 1 week post purification.

2.2.4.2.3 Tryptic Digest of Protein

Purified AmtR was subjected to tryptic digestion to preferentially digest away the N-terminal DNA binding domain, in order to crystallise only the C-terminal regulatory domain. Trypsin from bovine pancreas (Sigma Aldrich, St. Louis MO) was added to AmtR at ratios of 1 mg:100 mg, 1 mg:500 mg, and 1 mg:1,000 mg respectively. All further trypsin crystallization fine-screens used trypsin at a ratio of 1 mg:100 mg. Trypsin was added to the protein, which was immediately used to lay down crystal fine screens to limit the activity of the trypsin on the protein prior to beginning the crystallisation process.

2.2.4.2.4 DNase Treatment of Protein

Purified AmtR was subjected to DNA removal by DNase I (from bovine pancreas) (Sigma Aldrich) to remove co-purified DNA. DNase I was added to protein to give 1 mg:800 mg ratio respectively. DNase I was added to AmtR and incubated at 37 °C overnight, before being used to lay down crystallisation fine screens.

2.2.4.2.5 Heavy Metal Soaks

A post-soak with mercury was performed on crystals to incorporate the heavy metal into the crystal at cysteine positions. ML (250 μ l) containing 5 mM

Chapter Two

mercury(II) chloride was transferred into a pre-greased well. 2 μ l of the ML/mercury(II) chloride was pipetted onto a coverslip and a crystal was added to this drop. The coverslip was inverted above the well and sealed overnight before looping the crystal and freezing it for testing.

2.2.4.2.6 Ligand-Additive Screens

For co-crystallisation crystals were grown in hanging drops (Section 2.2.4.2.1) with the ML containing 0.1 mM – 100 mM of an additive. Additives were selected from the AmtR ligand library (Table 2.1. in Section 2.2.6.2.). For post soaking, AmtR crystals grown in non-additive ML conditions were transferred to a new hanging drop containing 0.1 mM-100 mM of an additive and left overnight.

2.2.4.2.7 DNA Co-Crystallisation

Crystals were grown in hanging drops (Section 2.2.4.2.1) with AmtR protein that was co-purified with a co-repressor and a 26 bp DNA oligonucleotide containing the AmtR recognition sequence. The ML contained up to 10 mM of co-repressor.

2.2.4.3 Testing Crystal X-Ray Diffraction

Protein crystals were tested for X-ray diffraction on an Agilent Supernova X-ray Diffractometer (Agilent Technologies, Liberty Lake WA) at the University of Waikato using a Titan S2 CCD detector, or at the Australian Synchrotron using an ADSC Quantum 210r detector.

2.2.4.4 X-Ray Data Collection

AmtR protein crystals were flash-frozen in liquid nitrogen with up to 30% glycerol prior to data collection. For data collection, crystals were mounted in a stream of cold N₂ gas at 100 K. Data was collected at the macromolecular crystallography beamline (MX1) at the Australian Synchrotron, Melbourne, Australia using an ADSC Quantum 210r detector.

2.2.5 AmtR DNA Electrophoretic Mobility Shift Assays (EMSA)

Purified AmtR that had been treated for co-purified DNA removal by high salt SEC (Section 2.2.2.8.2) and subsequent dialysis (Section 2.2.3) was used in electrophoretic mobility shift assays (EMSA) using the DIG Gel Shift kit, 2nd Generation (Roche). The target DNA fragments were labelled as described in Section 2.2.1.2.

2.2.5.1 EMSA Reactions

The protein:DNA ratio was based on the molarity of AmtR (24.330 g/mol) calculated by ProtParam (web.expasy.org/protparam/) and DNA which was calculated using the Promega BioMath Calculator (<http://www.promega.com/a/apps/biomath/index.html?calc=ugpmols>). Assay protein:DNA ratios began at equal molar and subsequent higher ratios of protein to DNA. Assays were conducted as per manufacturer's instructions with each reaction containing 1.5 x Binding Buffer, 0.5 µg of Poly dI-dC, 0.4 ng of DIG labelled DNA and various concentrations of AmtR. To confirm positive shifts were a result of protein binding to DNA and not to the DIG label, assays were set up with unlabelled target DNA added in 120 X excess at the highest labelled DNA:protein ratio. This assay was labelled specific control. To determine the specificity of protein binding, assays were set up with unlabelled non-specific DNA (the promoter region of the *M. smegmatis phoH2* gene which contains no putative AmtR recognition sequence) added in 103 X excess (to account for size difference of the DNA fragments). This assay was labelled the non-specific control.

Assays were prepared on ice and incubated at room temperature for 20 min. Loading dye was added to each assay and then run on a 6 % native acrylamide TBE gel (Section 2.2.1.4.2)

2.2.5.2 DNA Southern Blotting and Cross Linking

DNA was transferred from the 6 % native acrylamide TBE gel to a Biotransfer[®] nylon transfer membrane (Pall Corporation) by southern blotting. The membrane was placed over the gel and several pieces of blotting paper were

Chapter Two

placed on top of the membrane and weighted by a 150 ml schott bottle filled with water to increased capillary movement. Blotting was done at room temperature for 30 min. DNA was fixed to the membrane by UV crosslinking using a Gibco BRL UV Crosslinker (Life Technologies).

2.2.5.3 Membrane Washing and Blocking

The membrane was rinsed in 50 ml washing buffer (0.1 M maleic acid, 0.15 M NaCl, pH 7.5, 0.3 % Tween 20) for 5 min and blocked for 30 min in 1 X blocking solution (Roche). The membrane was then incubated in 1 X blocking solution + antibody (Roche) for 30 min followed by two 15 min wash steps with washing buffer. The membrane was equilibrated in detection buffer (0.1 M Tris, 0.1 M NaCl, pH 9.5) for 5 min before doing the chemiluminescent immunoassay.

2.2.5.4 Chemiluminescence Immunoassay

A chemiluminescence reaction using the alkaline phosphatase-conjugated antibody bound to the DIG labelled DNA and Disodium 3-(4-methoxyspiro {1, 2-dioxetane-3, 2'-(5'-chloro) tricyclodecan}-4-yl) phenyl phosphate (CSPD) was done as per the manufacturer's instructions. The reaction was incubated at room temperature for 5 min, followed by a 10 min incubation at 37 °C to enhance the reaction. The alkaline phosphatase dephosphorylates the CSPD resulting in light emission. The light emission was captured by exposing the membrane to Amersham Hyperfilm X-ray film (GE Healthcare) in a light protective cassette for a minimum of 1 hr. The film was then developed in Kodak developer (Sigma Aldrich) for 2 min and fixed using Kodak fixer (Sigma Aldrich) for 30 sec. The film was visualised and captured using white light.

2.2.6 AmtR Signalling Molecule Identification

To determine the signalling molecule required for the regulation of the AmtR regulon a variety of techniques were implemented including protein pull down assays, fluorescence binding assays, AmtR + *msmeg_2184* promoter region gel shift assays, crystallization by additive screens and high throughput thermal melt assays.

2.2.6.1 AmtR-GlnK Pull Down Experiments

To test for protein interaction between AmtR and GlnK, both proteins were overexpressed into *M. smegmatis* overexpression culture conditions (Section 2.2.2.8) and purified by IMAC (Section 2.2.2.8.1.). To test for protein interaction between AmtR and adenylyated GlnK (GlnK-AMP), AmtR was expressed as above and *M. smegmatis* pYUB28b+*glnK* was grown in modified Sautons media + 1 mM NH₄Cl₂ (Section 2.2.2.7.2.) To confirm adenylylation of GlnK (GlnK-AMP) the protein was purified by IMAC and SDS-PAGE analysed (Section 2.2.2.1.).

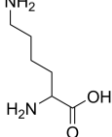
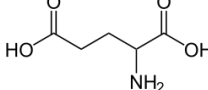
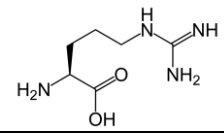
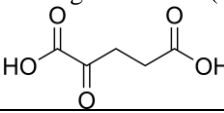
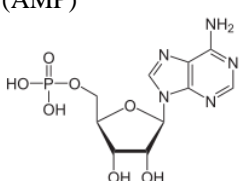
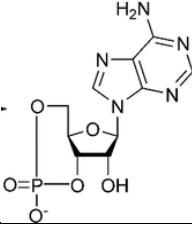
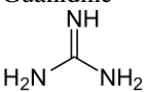
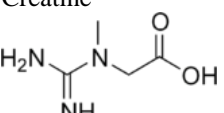
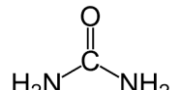
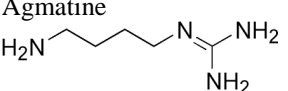
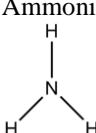
2.2.6.1.1 Incubation of AmtR and GlnK

IMAC purified AmtR and GlnK, and AmtR and GlnK-AMP were incubated together at equal monomeric molar ratios for 20 min at room temperature then moved to 4 °C overnight. The solution was then run through a pre-equilibrated (50 mM phosphate buffer, 200 mM NaCl, pH 7.4) analytical Superdex™ S200 10/300 SEC column (GE Healthcare). A SEC pull down was implemented over a traditional IMAC pull down as both proteins included a terminal His tag. Protein peaks from the elution traces were analysed for interactions by a change in size of the protein complex coming off the column, which was verified by different band profiles in both SDS-PAGE and native-PAGE (Section 2.2.2.1 and Section 2.2.2.2.).

2.2.6.2 Ligand Library Generation

The *M. smegmatis* AmtR regulon is primarily composed of genes involved in alternative nitrogen metabolism. As a result the ligand library was generated by selecting biological molecules that were (I) substrates or metabolites in nitrogen metabolism (II) general cellular signalling molecules and (III) compatible with the chemistry of the AmtR binding pocket in the regulatory domain (based on the structure we solved) (Table 2.1).

Table 2.1. AmtR ligand library containing a list of molecules tested as putative AmtR ligands.

Ligand group	Example	Reasoning
Amino acids (a.a) $\begin{array}{c} \text{R} \\ \\ \text{H}_2\text{N}-\text{C}-\text{COOH} \\ \\ \text{H} \end{array}$	Lysine 	<ul style="list-style-type: none"> – Nitrogen substrate and/or metabolite – AmtR binding site chemistry
	Glutamate 	
	Arginine 	
Metabolic signaling molecules	2-Oxoglutaric acid (2-OG) 	<ul style="list-style-type: none"> – Key intermediate in Krebs cycle – Metabolite of a.a metabolism and synthesis
	Adenosine monophosphate (AMP) 	
	Cyclic AMP 	
Ammonia metabolism	Guanidine 	<ul style="list-style-type: none"> – AmtR binding chemistry – Nitrogen metabolism
	Creatine 	
	Urea 	<ul style="list-style-type: none"> – AmtR regulon
	Agmatine 	
	Ammonia 	

The molecules were screened as putative AmtR ligands with purified and extensively dialysed AmtR by a range of methods including (1) fluorescence binding assays, (2) EMSA of AmtR bound to the *msmeg_2184* promoter region to test if putative ligands disrupted the protein-DNA interaction, (3) co-crystallization and post-soaking of AmtR crystals with putative ligands for X-ray crystallography, (4) high throughput differential scanning fluorimetry and (5) modelling of putative ligands into the binding pocket of the AmtR regulatory domain.

2.2.6.3 Fluorescence Binding Assays

The intrinsic fluorescence excitation and emission of AmtR was determined by a 3D wavelength scan using a F-7000 fluorescence spectrophotometer (Hitachi HTA, USA). Binding assays were conducted using a 280 nm excitation wavelength scan with peak emission of AmtR at approximately 330 nm. Molecules from the ligand library were screened for binding to AmtR with concentrations of ligand from 0 mM-100mM. Binding of ligand was represented by a change in AmtR emission wavelength or intensity proportional to the amount of ligand added to the assay.

2.2.6.4 Ligand Screening via AmtR + DNA Complex Disruption by EMSA

The binding of AmtR to its recognition sequence ATCTGTCATCTGACAG was used as a screen for ligand binding. The TetR protein family that AmtR belongs to is characterised by signalling molecule induced disruption of protein:DNA complexes. To test if ligands from the library caused disruption of the AmtR:DNA, complex EMSA shifts were conducted (Section 2.2.5.1) with the addition of ligands (0 mM-100 mM).

2.2.6.5 AmtR Co-Crystallisation and Post Soaking Additive Screens

AmtR co-crystallisation fine screens and post soaking were set up (Section 2.2.4.2) with the addition of 0 mM-100 mM ligand from the library to the ML.

2.2.6.6 High Throughput Differential Scanning Fluorimetry

A differential scanning fluorimetry assay was performed at the University of Waikato by collaborating PhD student Michael Petridu to detect ligand interactions that promote protein stability (Niesen, Berglund, & Vedadi, 2007). Experiments were performed with final concentrations of 5 μ M purified AmtR protein and 5 x SYPRO® Orange Dye (Thermo Fisher Scientific) in single wells of an opaque 96-well plate. When started, DNA was added at a molar ratio of 30:1 (AmtR:DNA). To detect ligand interactions, compounds were obtained from the nitrogen substrate PM3 Biolog Plate (Biolog Inc., Hayward CA), resuspended in 50 μ L PBS buffer to make final ligand concentrations of 2 mM. The melting temperature of AmtR was determined using a LightCycler®480 (Roche). Samples were equilibrated for 5 minutes at 25 °C and then the temperature was increased to 90 °C at a rate of 1 °C/min, taking a fluorescence reading every 0.07 °C. The melting point of the protein was calculated by the LightCycler®480 software v1.5.1.62.

2.2.6.7 ParDOCK Modelling of Xanthine into AmtR Binding Pocket

Xanthine was manually modeled into the AmtR binding site in a position where there was unresolved density that could not be satisfied by the model during molecular replacement. The structure was then run through a series of refmac5 refinements to optimize ligand positioning. To further optimize ligand positioning the structure was run through the ParDOCK program (Gupta, Gandhimathi, Sharma, & Jayaram, 2007). ParDOCK is an all-atom energy based Monte Carlo, rigid protein ligand docking program, which predicts the binding mode of the ligand in receptor target sites. The ParDOCK output file was visualized in PyMOL (Delano, 2002) to identify the predicted hydrogen bonding interactions between the residues in the binding site and xanthine.

2.3 Results

2.3.1 Expression and Protein Purification of AmtR in *M. smegmatis*

Cloning and small scale expression trials were previously conducted and optimized by Jo McKenzie, a Post Doctoral fellow in the Arcus Lab. Protein expression cultures were scaled up (Section 2.2.2.7) and protein was initially purified by IMAC (Section 2.2.8.1) and analysed by SDS-PAGE (Section 2.2.2.1). The chromatogram and corresponding SDS-PAGE gel (Figure 2.1) depict the column load, flow through and purification fractions. The AmtR monomer is 24.3 kDa, as shown by the SDS-PAGE gel. Non-specifically bound protein eluted early off the column at a low concentration of imidazole (100-250 mM), whereas AmtR-His tagged protein eluted later in the gradient with a greater concentration of imidazole (400-650 mM).

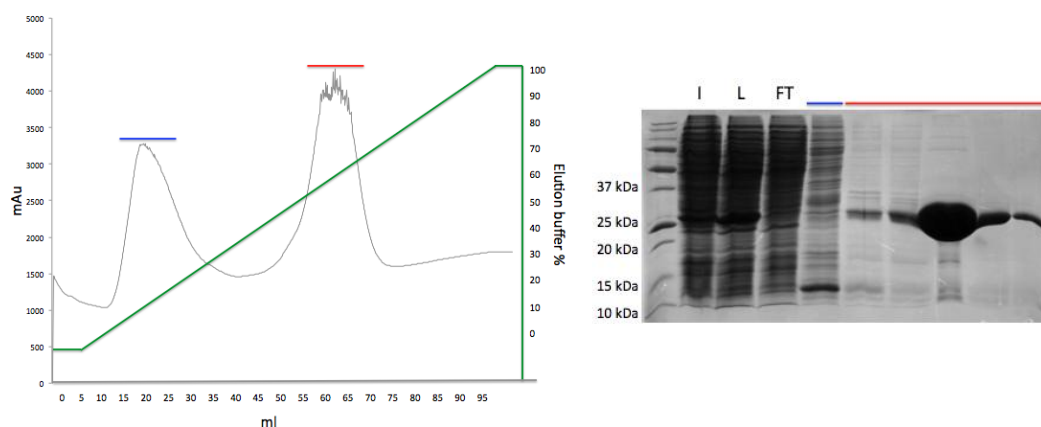


Figure 2.1. IMAC purification and corresponding SDS-PAGE analysis of AmtR. The chromatogram depicts the UV absorbance and elution profile and the corresponding SDS-PAGE gels depict the insoluble fraction after cell lysis (I), soluble fraction loaded onto the column (L) and the flow through fraction containing unbound non-specific protein (FT). Fractions and their elution position are shown by the coloured bars. The blue line represents the fractions containing the non His tagged protein and the red represents the fractions containing AmtR-His tagged.

AmtR was further purified by SEC (Section 2.2.2.8.2) (Figure 2.2). Proteins belonging to the TetR family generally function as dimers and AmtR eluted off the column as a single peak within the size range of 35-67 kDa, suggesting it is in a dimeric state (48.6 kDa).

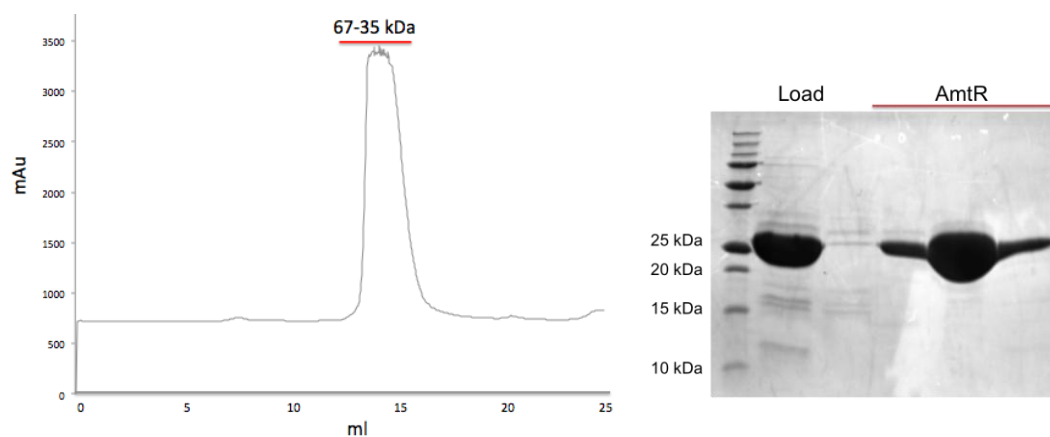


Figure 2.2. SEC of AmtR and corresponding SDS-PAGE gel. The chromatogram depicts the UV absorbance profile of protein elution and the gel depicts the load and elution fractions. The position of the AmtR elution peak and approximate size is shown by the red bar.

SEC purified protein was subjected to subsequent high salt (2 M NaCl) induced dissociation of any AmtR:DNA complexes that were present due to co-purified genomic DNA, which may inhibit downstream *in vitro* analyses. The protein was then purified again by SEC in high salt buffer to ensure complete removal of contaminating DNA. To remove the high salt concentration and any co-purified ligands the protein was exhaustively dialysed into a low salt buffer (200 mM NaCl). The purified and dialysed AmtR was used for crystallisation trials, electrophoretic mobility DNA shift assays (EMSA) and ligand binding experiments.

2.3.2 AmtR X-ray Crystallographic Structures

2.3.2.1 AmtR C-terminal LBD Domain Structure

2.3.2.1.1 Trypsin Treatment of Protein and Crystallisation

Initial attempts to produce high quality crystals of the non truncated AmtR protein proved unsuccessful. Purified AmtR was then subjected to tryptic digestion to preferentially digest away the N-terminal DNA binding domain in order to crystallise only the C-terminal regulatory domain. AmtR trypsin treated crystals grew after 14 days in the crystallisation condition 100 mM Na acetate pH 5.0, 2.2 M Na formate with protein at 10 mg/ml. Crystals diffracted to 3 Å but the structure could not be solved by molecular replacement.

Chapter Two

2.3.2.1.2 Heavy Metal Post Soaking of Trypsin Treated AmtR Crystals

Trypsin treated crystals (Figure 2.3) were soaked overnight in the same conditions as Section 2.3.2.1.1 with the addition of mercury (5 mM HgCl₂). These crystals showed significantly better diffraction (1.84 Å), and data was collected on these crystals to solve the AmtR structure.

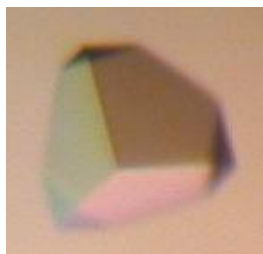


Figure 2.3. Crystal of *M. smegmatis* trypsin treated mercury soaked AmtR. Crystals grew after 14 days in 100 mM Na acetate pH 5.0, 2.2 M Na formate with protein at 10 mg/ml. Crystal diffracted to 1.84 Å and a dataset was collected with 100 % completeness in space group I 2 3.

2.3.2.1.3 Data Collection

Data was collected on the above trypsin treated crystal soaked in mercury to 1.84 Å with 100% completeness at the wavelength 12290.0 eV (6 eV above the L_{III} edge of mercury) (Table 2.2). The data set was indexed and integrated using iMosflm 7.0.9 (Leslie & Powell, 2007) then scaled and truncated using Aimless (Evans & Murshudov, 2013) in the CCP4 platform suite 6.4.0 (Collaborative Computational Project, Number 4, 1994). The resolution limit of 1.84 Å for the C-terminal domain dataset gives a very high R-merge value 0.201 (1.662) (Table 2.2). This is possibly the result of the very high redundancy. This resolution limit of 1.84 Å can be justified by the CC½ of 0.859 in the outer shell and the mean I/σ(I) 2.9 in outer shell (Table 2.2).

2.3.2.1.4 Structure Solution and Refinement

Phasing was achieved using the Single-wavelength Anomalous Diffraction (SAD) technique to solve the C-terminal AmtR structure. Initial mercury sites were identified and refined using SHELX (Sheldrick, 2007) and the solution was autosolved in space group I 2 3 using autorickshaw (Panjikar, Parthasarathy, Lamzin, Weiss, & Tucker, 2005) (Table 2.2). The structure was refined by initial restrained refinement in refmac5 (Murshudov et al., 2011; Murshudov, Vagin, & Dodson, 1997) in the CCP4 suite 6.4.0. Subsequent cycles of manual refinement using COOT 0.7 (Emsley, Lohkamp, Scott, & Cowtan, 2010) followed by refinement in Refmac5 on the CCP4 suite 6.4.0 were done (Table 2.2).

Table 2.2. Data Collection and Refinement Statistics.

	SAD/Truncated structure	MR/Full Structure
<i>Data collection</i>		
Wavelength	1.0072 Å	0.95370 Å
Space group	I 2 3	C 2
Unit cell dimensions (Å)	111.1, 111.1, 111.1	181.8, 105.8, 39.57
Unit cell dimensions (°)	90, 90, 90	90, 94.5, 90
Resolution (Å)	32.07-1.84, (1.88-1.84)	37.25-1.98 (2.03-1.98)
R-merge	0.201 (1.662)	0.100 (0.694)
CC (1/2)	0.999 (0.859)	0.997 (0.769)
Completeness (%)	100 (100)	97.1 (90.5)
Redundancy	42.7 (39.7)	6.7 (5.7)
No. of observations	856794 (48790)	336413 (18392)
No. of unique reflections	20068 (1230)	50329 (3241)
Mean I/σ(I)	69.1 (2.9)	9.6 (2.1)
<i>Phasing Statistics</i>		
Anomalous completeness	99.9	-
Anomalous multiplicity	22.1	-
DelAnom correlation between half sets	0.764 (0.035)	-
<i>Refinement</i>		
R-factor	0.183	0.212
R-free	0.210	0.252
Protein atoms	1097	4549
Metal atoms	9 Hg	-
Solvent atoms	147	126
Average B value (Å ²)	25.7	28.8
<i>RMSD</i>		
Bond angles (°)	2.670	1.851
Bond lengths (Å)	0.028	0.018

2.3.2.1.5 Structural Analysis

The crystal structure showed trypsin had cleaved off the N-terminal 3-helical DNA binding domain (DBD), leaving the larger 7 α -helical C-terminal ligand binding domain (LBD) (Figure 2.4). The structure crystallised as a homodimer with one molecule in the asymmetric unit.

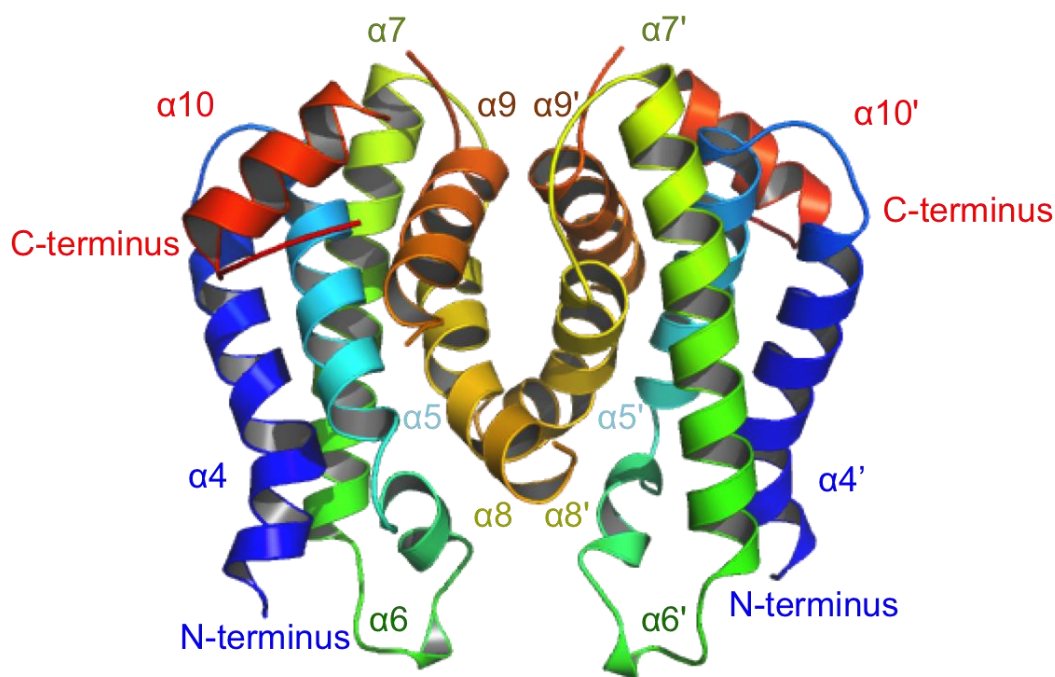


Figure 2.4. A cartoon diagram of the X-ray crystal structure of the C-terminal ligand binding domain of *M. smegmatis* AmtR. The structure was solved using SAD on a mercury soaked crystal. The structure resolved to 1.84 Å in 1 2 3 symmetry. The structure is solved as a homodimer with each monomer composed of a 7- α helical domain. The N-terminus is helix $\alpha 4$ and is coloured dark blue. The C-terminus is helix $\alpha 10$ and is coloured red. This figure and all other structural figures have been drawn using PyMOL (Delano, 2002).

The structure of the LBD begins at the N-terminal helix $\alpha 4$ which is a long helix composed of 24 residues (Asp₆₆-Gly₈₉) that runs parallel with helix $\alpha 7$ (Figure 2.4). Helix $\alpha 4$ is responsible for connecting the DBD to the LBD. Helix $\alpha 4$ is followed by a short loop region before the start of helix $\alpha 5$. The LBD domain is built around a central triangle composed of helices $\alpha 5$, $\alpha 6$ and $\alpha 7$, with helix $\alpha 5$ and helix $\alpha 7$ running antiparallel with each other so that the start of helix $\alpha 5$ and the end of helix $\alpha 7$ are close to each other (Figure 2.5). Helix $\alpha 6$ curves around helix $\alpha 4$ (Figure 2.5) and is composed of 9 residues but has only one full turn of a helix. Helix $\alpha 8$ runs at an angle that is nearly parallel with helix $\alpha 5$. Helix $\alpha 9$ is positioned behind the central triangle and runs antiparallel with helix $\alpha 8$. (Figure 2.5) Helix $\alpha 8$ and helix $\alpha 9$ constitute the dimerization interface

forming an orthogonally packed four-helix bundle with the same helices (helix $\alpha 8'$ and helix $\alpha 9'$) from the other monomer (Figure 2.4).

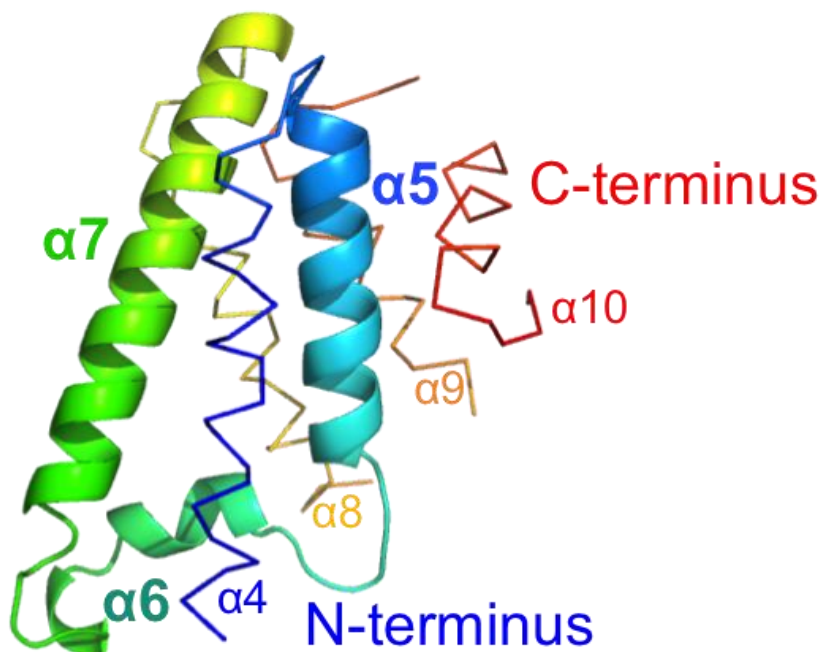


Figure 2.5. Cartoon diagram of AmtR monomer showing the central triangle of the LBD made up of helices $\alpha 5$, $\alpha 6$ and $\alpha 7$. Helices $\alpha 5$ and $\alpha 7$ run antiparallel to each other with the N-terminus of helix $\alpha 5$ interacting with the C-terminus of helix $\alpha 7$. Helix $\alpha 6$ forms the base of the triangle by wrapping around helix $\alpha 4$ and is the most disordered region of the structure. Helices $\alpha 8$ and $\alpha 9$ sit at the back of the triangle and run antiparallel to each other. Helices $\alpha 4$, $\alpha 8$, $\alpha 9$ and $\alpha 10$ have been shown in ribbon form to highlight the conserved triangle of helices $\alpha 5$, $\alpha 6$ and $\alpha 7$.

The LBD forms a large binding pocket in the centre of each monomer, which is accessed by a long entrance tunnel (Figure 2.6). The AmtR binding pocket sits in the lower section of the triangle in the space between helices $\alpha 4$ and $\alpha 5$ with a volume of approximately 125.70 \AA^3 (Figure 2.6). The pocket consists of 11 residues from helices $\alpha 4$, $\alpha 5$, $\alpha 6$, $\alpha 7$ and $\alpha 8$: 7 polar residues (Thr₇₆, Asp₁₀₅, Tyr₁₀₄, Glu₁₀₈, Tyr₁₄₄, Arg₁₆₇ and Glu₁₇₀) and 5 non-polar residues (Pro₈₀, Leu₁₀₁, Leu₁₀₉, Leu₁₄₀ and Phen₁₆₆) (Figure 2.6). Glu₁₇₀ on helix $\alpha 8$ shows density for alternative conformations and sits at the entrance of the binding pocket. It is speculated that this residue may act as a gate to the binding pocket that can change position based on the conformation of AmtR and help stabilise ligand binding.

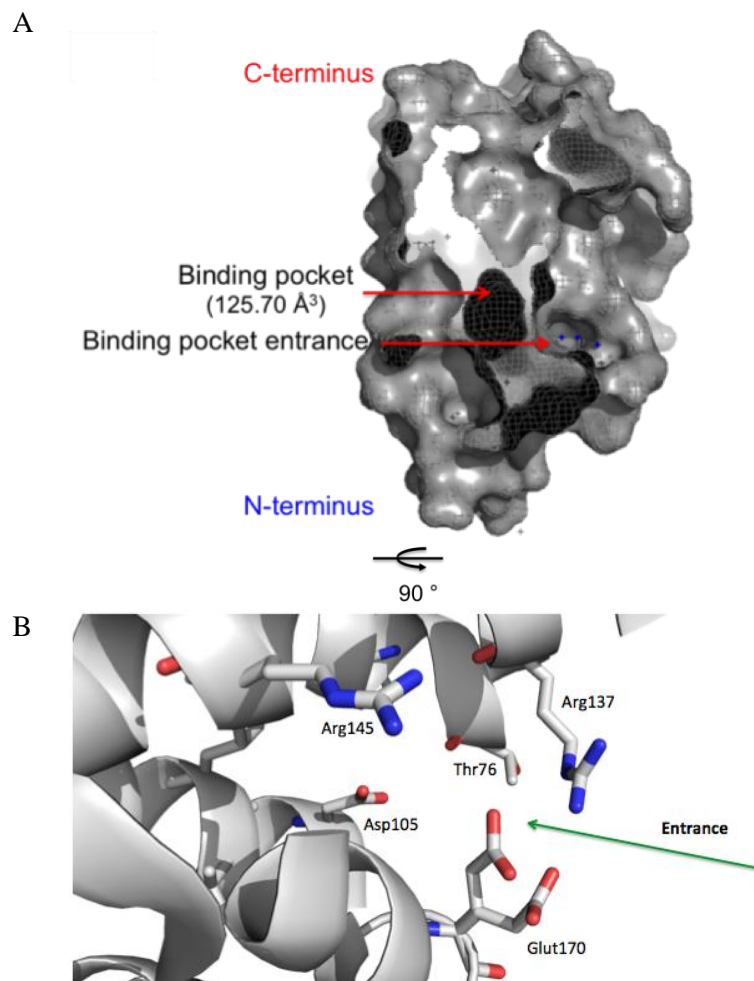


Figure 2.6. The *M. smegmatis* LBD (A) A mesh lined space filling model of the AmtR LBD monomer structure showing the long entrance tunnel along the grey surface with three water molecules (in blue) sitting in it. The binding site is shown in black with a mesh surface. (B) The AmtR LBD flipped 90 ° to illustrate the putative interacting residues in binding pocket as sticks from helices $\alpha 5$, $\alpha 6$, $\alpha 7$ and $\alpha 8$. Glu₁₇₀ is shown with its alternate conformations illustrating its potential role as a gate to the binding pocket.

2.3.2.2 AmtR Full Structure

2.3.2.2.1 DNase Treatment of Protein and Crystallisation

A good quality non-trypsin treated AmtR crystal (Figure 2.7) was obtained only from fresh (1 hr post-purification) AmtR that had been DNase treated in the same crystallisation condition as trypsin treated crystals (100 mM Na acetate pH 5.0, 2.2 M Na formate with protein at 10 mg/ml). The crystal diffracted to 1.98 Å and data was collected on this crystal for structure solution by molecular replacement (MR).



Figure 2.7. *M. smegmatis* AmtR non-trypsin treated crystal. The crystal grew only from fresh (1 hr post-purification) AmtR that had been DNase treated in 100 mM Na acetate pH 5.0, 2.2 M Na formate with protein at 10 mg/ml. The crystal diffracted to 1.98 Å and a dataset was collected with 97.1 % completeness in space group C2.

2.3.2.2.2 Data Collection

Data was collected on the above crystal to 1.98 Å and a dataset was collected with 97.1 % completeness in space group C2 at the wavelength 12669.5 eV (Table 2.2). The data set was indexed and integrated using iMosflm 7.0.9 (Leslie & Powell, 2007) then scaled and truncated using Aimless (Evans & Murshudov, 2013) in the CCP4 platform suite 6.4.0 (Collaborative Computational Project, Number 4, 1994).

2.3.2.2.3 Structure Solution and Refinement

The structure solution was determined by molecular replacement (MR) using the truncated AmtR structure solved by SAD as a model. MR was done in Phaser (McCoy et al., 2007) in the CCP4 suite 6.4.0. The N-terminal domain was then built in using Buccaneer (Cowtan, 2006; Cowtan, 2007). The structure was refined by initial restrained refinement in refmac5 (Murshudov et al., 1997; Murshudov et al., 2011) refinement in the CCP4 suite 6.4.0. Subsequent cycles of manual refinement using COOT 0.7 (Emsley et al., 2010) followed by refinement in Refmac5 on the CCP4 suite 6.4.0 were done (Table 2.2). The Coordinate files and structure factors for the full structure has been deposited in the Protein Data Bank with PDB code 5E57.

2.3.2.2.4 Structural Analysis

The structure crystallized as a homodimer with each monomer composed of 10 α -helices with connecting turns and loops that constitute two domains (Figure 2.8). The N-terminal domain is a Helix-Turn-Helix (HTH) DNA binding motif and is composed of helices α_1 , α_2 , and α_3 . The C-terminal domain is the LBD as described in Section 2.3.1.1.

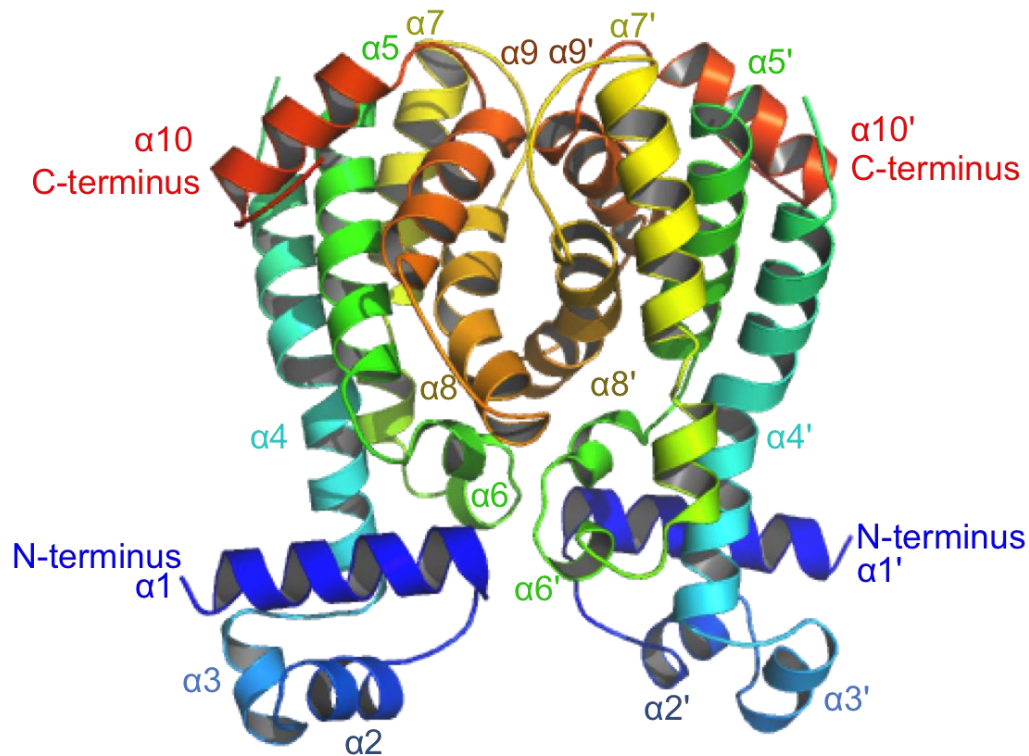


Figure 2.8. Cartoon diagram of the full X-ray crystallographic structure of *M. smegmatis* AmtR. The structure was solved by molecular replacement to 1.98 Å with C2 symmetry. The structure solved as a homodimer with each monomer composed of 10 α helices constituting a N-terminal HTH DNA binding domain and a C-terminal LBD domain.

Alignment of the SAD LBD structure with the full structure of AmtR showed a Root Mean Standard Deviation (RMSD) of 0.36 Å. The main area of disparity is at the N-terminus where the full structure has the DBD present and the SAD structure does not. There is a difference of 5.8 Å at the N-terminus of the helix $\alpha 4$ between the structures, as a result of the DBD changing the orientation of this region in the full structure (Figure 2.9). Helices $\alpha 6$ of both structures are variable in both secondary structure and position and also align poorly. There is a difference of 12.2 Å between the last turn of the helix in both structures (Figure 2.9). This observed variability is important for the signal transduction to be transferred from the ligand binding site to the DBD. The N-terminal region of helix $\alpha 7$ has a difference of 3.2 Å, as a result of the full structure having a more significant kink in the helical structure (Figure 2.9). Helices $\alpha 5$, $\alpha 8$, $\alpha 9$ and $\alpha 10$ align well. Helix $\alpha 10$ in the full structure has an extension containing a disordered region followed by a small helix that is composed of the C-terminal 6-His tag.

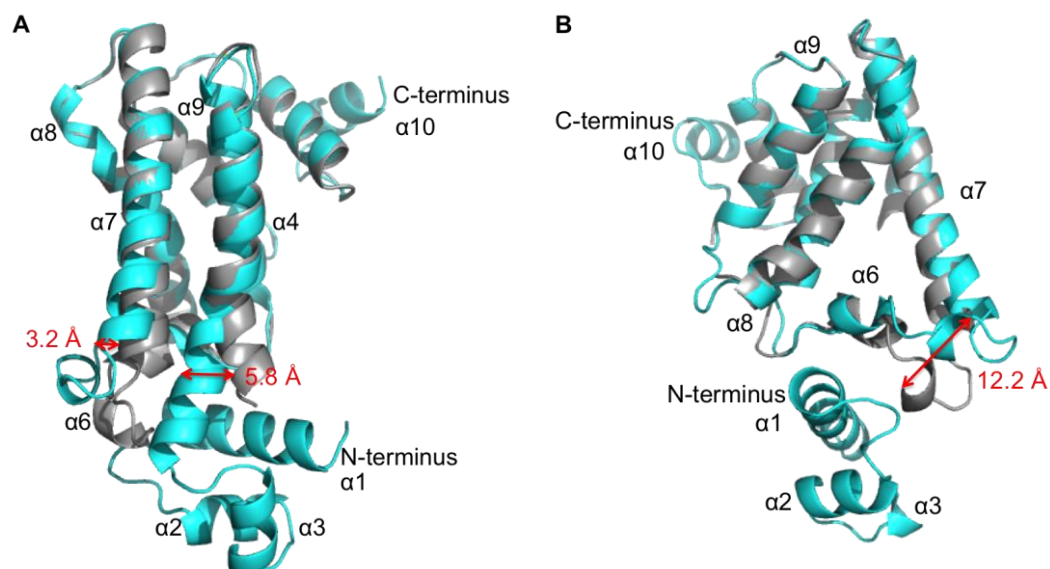


Figure 2.9. Alignment of the SAD and full structure of AmtR. The SAD structure is coloured grey and the full structure is coloured cyan. The SAD structure is composed of the C-terminus LBD and the full structure is composed of both the N-terminus DBD and the C-terminus LBD with an extension to the helix $\alpha 10$ containing a 6-His tag. (A) Good alignment for the LBD with the exceptions in the N-terminus of helices $\alpha 4$ and $\alpha 7$. (B) Rotation of structures by 90° to show the poor alignment in variable region of helix $\alpha 6$.

2.3.2.2.5 The N-Terminal DNA Binding Domain (DBD) of AmtR

The AmtR DBD domain is 46 residues and is composed of a long helix $\alpha 1$ (14 residues (Arg₂₃-Thr₃₆)), followed by a HTH motif composed of helices $\alpha 2$ and $\alpha 3$ which are orientated to insert into the major groove of DNA (Figure 2.10). Helix $\alpha 2$ is composed of 6 residues (Arg₅₄-His₆₀) and forms a small helix. Helices $\alpha 2$ and $\alpha 3$ are at an 80.5° angle and joined by a 5 residue turn. Helix $\alpha 3$ is composed of 5 residues (Ala₅₆-His₆₀) and forms a short helix that constitutes the recognition helix of the HTH motif (Figure 2.10).

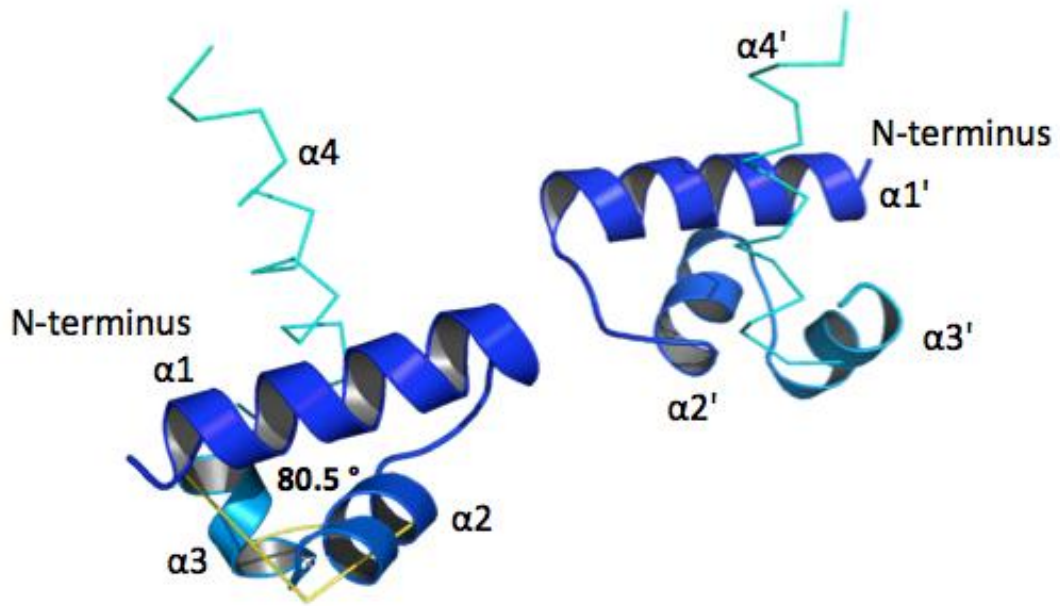


Figure 2.10. Cartoon diagram of the AmtR dimer DBD. Helix $\alpha 1$ is the longest helix shown in dark blue followed by helix $\alpha 2$ shown in lighter blue and then the recognition helix $\alpha 3$ shown in aqua blue. Helix $\alpha 4$ is shown in cyan and is represented in ribbon form to help illustrate the helices that compose only in the DBD. The 80.5° angle between the helices that compose the HTH motif, $\alpha 2$ and $\alpha 3$, is shown in yellow.

The DBD has a positively charged surface (Figure 2.11), with the residues responsible for this charge from helices $\alpha 1$ (Arg₂₃), $\alpha 2$ (Arg₄₅ & Arg₄₆) and $\alpha 3$ (His₆₀) and the loops between helices $\alpha 1$ - $\alpha 2$ (His₃₇), $\alpha 2$ - $\alpha 3$ (Arg₅₄) and $\alpha 3$ - $\alpha 4$ (His₆₁ & Lys₆₅).

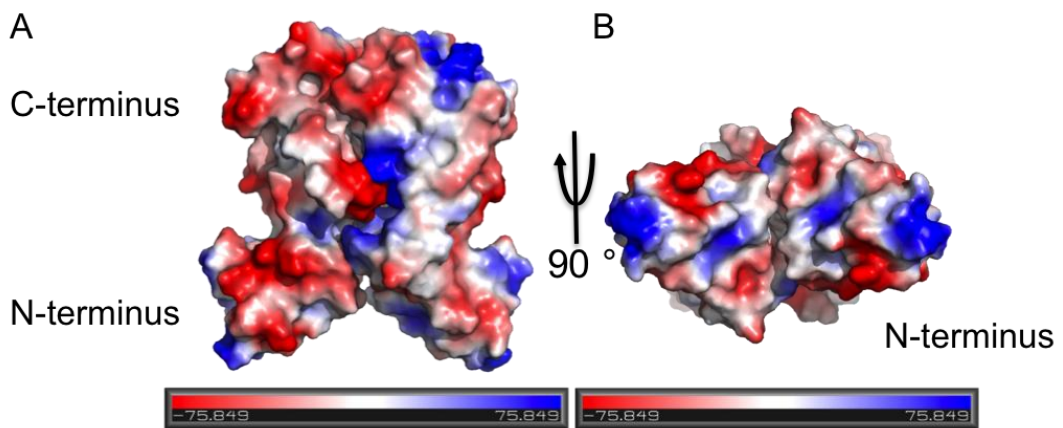


Figure 2.11. Electrostatic image of the AmtR dimer. (A) Full structure showing electrostatic properties of the AmtR dimer surface. (B) AmtR dimer rotated 90° to show the positive charge of the DBD.

2.3.2.2.6 Comparison of AmtR to Characterised TetR Structures

Although there are over 100 TetR structures in the PDB, no pair of any solved TetR shares more than 36 % sequence identity and less than 20 have been functionally characterized (Yu, Reichheld, Savchenko, Parkinson, & Davidson, 2010). The lack of conservation is seen in the ligand binding domain with the average pairwise identity at only 9 % (Yu et al., 2010). The structures range from 177 to 237 residues in length and all possess nine conserved helices where helix 1 to helix 3 comprises the DBD domain and helix 4 to helix 9 comprises the LBD domain (Yu et al., 2010).

The *M. smegmatis* AmtR structure was compared with structures in the PDB using the PDBeFOLD server. The closest structure was a putative TetR from *Rhodococcus jostii* RHA1, a close relative of mycobacteria (McLeod et al., 2006) (PDB code 3HIM; Z = 6.64; RMSD, 2.40 Å; number of matched residues, 168; sequence identity, 23 %). The structure was solved with C2 symmetry to 2.2 Å. No functional work has been done on this protein so no further comparisons were made. The second alignment was with the TetR protein, phenylacetic acid and fatty acid metabolism regulator (PfmR), from the extremely thermophilic bacterium, *Thermus thermophilus* HB8 (Agari, Sakamoto, Kuramitsu, & Shinkai, 2012) (PDB code 3VPR; Z = 7.0; RMSD, 2.31 Å; number of matched residues, 159; sequence identity, 18 %) (Figure 2.12). Other alignments included another putative transcriptional regulatory protein also from *R. jostii* RHA1 (PDB code 2IBD; Z = 6.71; RMSD, 2.49 Å; number of matched residues, 163; sequence identity, 21 %), the nucleoid occlusion factor SlnA protein from *E. coli* (PDB code 4GCL; Z = 6.30; RMSD, 2.87 Å; number of matched residues, 162; sequence identity, 19 %), the nucleoid occlusion factor SlnA protein from *Klebsiella pneumonia* (PDB code 4GCK; Z = 5.91; RMSD, 2.93 Å; number of matched residues, 161; sequence identity, 18%) and the transcriptional regulator HlyIIIR from *Bacillus cereus* (PDB code 2JK3; Z = 5.45; RMSD, 2.46 Å; number of matched residues, 145; sequence identity, 15 %).

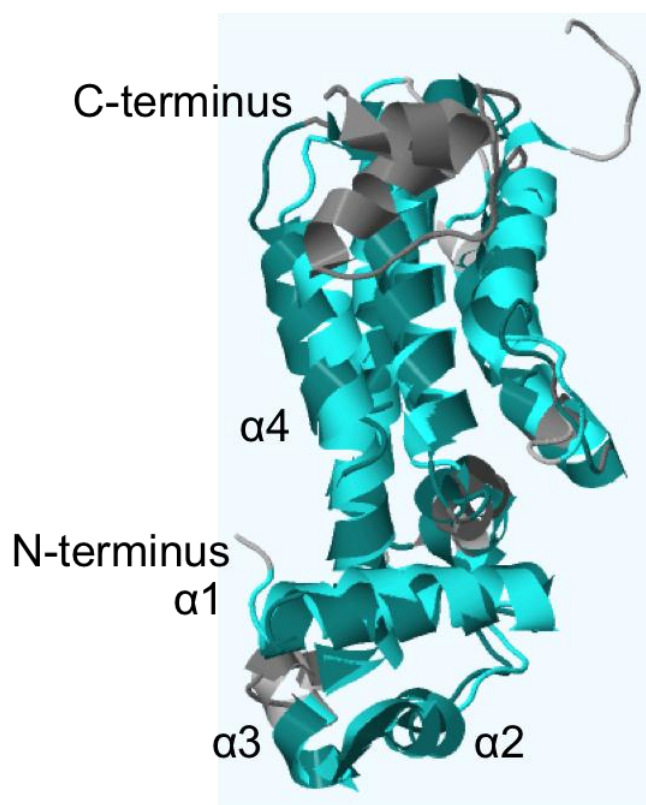


Figure 2.12. Alignment of the AmtR monomer with the PfmR monomer using PDBeFOLD. AmtR is coloured dark cyan and PfmR is coloured light cyan. Areas of no alignment are coloured grey. Areas of the poorest alignment are observed in the C-terminal LBD.

PfmR is characterised as a transcriptional repressor of the operons involved in phenylacetic acid and fatty acid metabolism. The N-terminal domain of PfmR resembles that of the DNA-binding domain from the QacR family of TetR transcription factors (Agari et al., 2012). QacR transcription factors are unique in their DNA binding mechanisms, as they generally recognise a longer sequence of DNA and involve the binding of a pair of QacR dimers (Schumacher et al., 2002). TetR DNA binding is generally by one dimer to a 15 bp DNA sequence upstream of the gene or promoter that it regulates (Orth et al., 2000).

The AmtR HTH motif aligns well with the QacR HTH motif (RMSD, 0.818 Å; number of matched residues, 49) (Figure 2.13). It also contains 4 (Thr₄₄, Tyr₅₉, His₆₁ & Lys₆₅) of the 7 QacR DBD residues (Thr₂₅, Ser₃₄, Ser₃₅, Tyr₄₀, Tyr₄₁, His₄₂, & Lys₄₆), which form hydrogen bonds with the backbone of the major groove of DNA in the QacR-DNA complex. The alignment shows that like QacR, the AmtR helix $\alpha 2$ is orientated to insert into the major groove of DNA and the N-terminus of the $\alpha 3$ points towards the centre of the DNA major groove, while the helix axis runs parallel to the major groove. This alignment suggests

that AmtR binds DNA in the same mechanism as QacR. Interestingly the residues responsible for binding to specific nucleotides in QacR are absent in AmtR, suggesting that binding to the major groove of DNA by AmtR is achieved by the QacR mechanism but the specificity in DNA recognition is not shared.

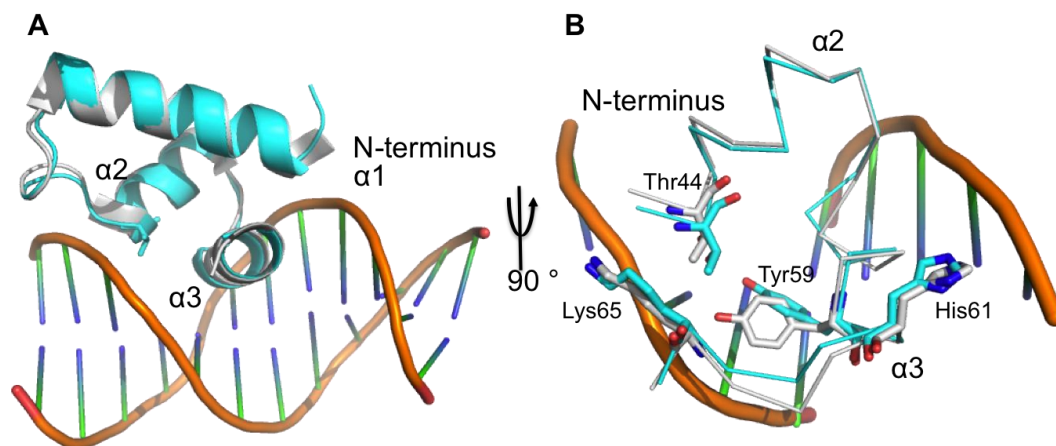


Figure 2.13. AmtR DBD superimposed over QacR DBD bound to DNA. AmtR is in cyan and QacR is in grey. (A) Cartoon diagram of interaction of QacR DBD bound to DNA with AmtR DBD superimposed over QacR. **(B)** Ribbon diagram of alignment rotated 90 ° to show conserved DNA binding residues Thr₄₄, Tyr₅₉, His₆₁ and Lys₆₅.

As mentioned earlier, it is common for TetR proteins to be disparate in the LBD, as a result of the difference in ligand binding and regulatory control. This is the case for AmtR and PfmR. Although the DBD regions align relatively well, the LBD regions show poor alignment and diverge significantly in their binding pocket volume and composition. Interestingly, PfmR appears to have a glutamate gate (Glu₁₂₃) at the entrance site to the binding pocket, similar to the Glu₁₇₀ observed in AmtR (Figure 2.14).

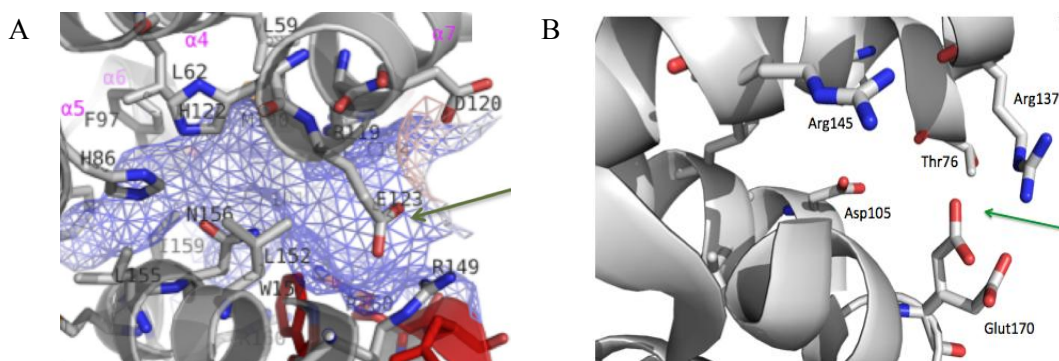
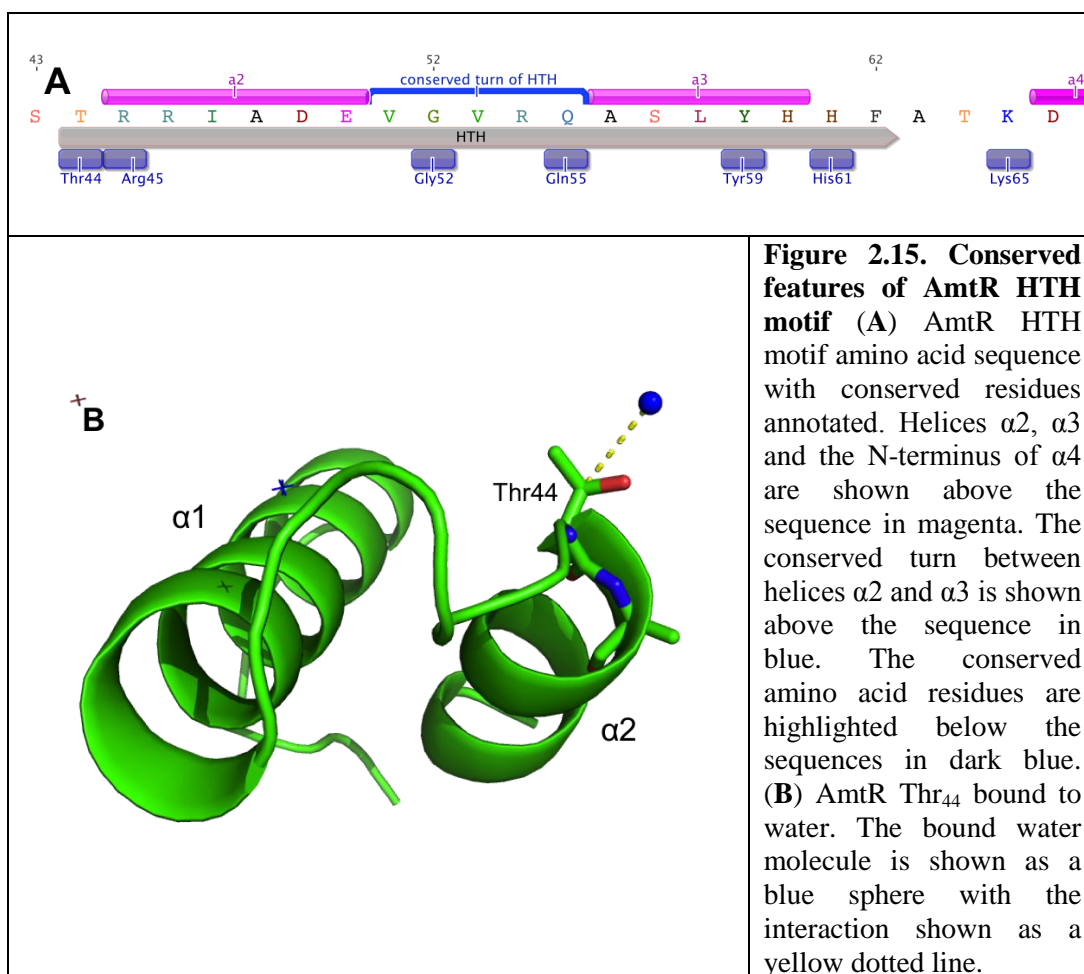


Figure 2.14. Comparison of (A) PfmR (Agari *et al*, 2012) and (B) AmtR binding pockets. Both pockets are composed of aromatic residues, hydrophobic residues and polar residues characteristic of TetR binding pockets. Both structures show a glutamate residue at the entrance of the binding pocket that may act as a gate for ligand binding. In PfmR, the glutamate bridge is labelled E123 and sits at the entrance of the cavity. In AmtR, the glutamate bridge is labelled Glu₁₇₀ and also sits at the entrance of the cavity. Direction of ligand movement into binding pocket is indicated by green arrow in both figures.

2.3.2.2.7 Conserved Structural Features of AmtR

The DBD domain is the most structurally conserved region of TetR structures (Yu *et al.*, 2010), which is composed of a typical HTH motif (Yu *et al.*, 2010). The AmtR DBD domain shares this conservation with the two short helices, $\alpha 2$ and $\alpha 3$ composing the HTH motif. Generally, there is a water molecule bound to a threonine that sits in the turn between helices $\alpha 1$ and $\alpha 2$, two residues beside the start of helix $\alpha 2$, and to the arginine in the first position of helix $\alpha 2$. The binding of these water molecules help orientate the residues in the HTH motif for binding to DNA (Orth *et al.*, 2000). In the AmtR structure the threonine is substituted to a serine (Ser₄₃) and the arginine (Arg₄₅) is conserved as the first position in helix $\alpha 2$ (Figure 2.15). Interestingly, the residue in between Ser₄₃ and Arg₄₅ is a threonine (Thr₄₄), and, from the structure, appears to be the only residue in this region that has a water molecule bound to it (Figure 2.15). Helices $\alpha 2$ and $\alpha 3$ are normally at a 120 ° angle and joined by a tight 4 residue turn with glycine found in the second position. The AmtR structure varies in that the angle of the turn between helices $\alpha 2$ and $\alpha 3$ is 80.5 ° and the turn is 5 residues long instead of 4 (Figure 2.15).



Helix α_3 of TetR structures is the most conserved helix, with each subfamily having high conservation, indicating that closely related TetR proteins within subfamilies recognise similar DNA sequences (Yu et al., 2010). This helix is termed the recognition helix and is responsible for inserting into the major groove of the DNA and interacting with the nucleotides of the palindromic sequence, with the exception of the 3 central base pairs. Upon binding, the major groove is widened at recognition half sites and narrowed at the central base pairs of the palindromic sequence due to partial unwinding (Orth et al., 2000).

The core residues responsible for binding DNA are highly conserved in TetR proteins and are located either in helix α_3 or in the loop between helices α_3 and α_4 (Yu et al., 2010). Figure 2.16 shows the sequence alignment of the AmtR DBD domain with representative DBD domains of the TetR family (PfmR, TetR & QacR) illustrating the conserved residues Leu₅₈, Tyr₅₉, His₆₁ and Lys₆₅. Residues Tyr₅₉, His₆₁ and Lys₆₅ are responsible for hydrogen bonding to the phosphate groups in the major groove containing the recognition sites of the

palindromic DNA. The side chain of Lys₆₅ also forms hydrogen bonds with oxygen atoms in residues in the loop between helices $\alpha 1$ and $\alpha 2$, which are thought to stabilise the region (Yu et al., 2010). Lys₆₅ also forms salt bridges with the DNA phosphate backbone, showing that it is a critical residue for DNA binding and functions to both stabilize the DBD domain and interact with DNA.

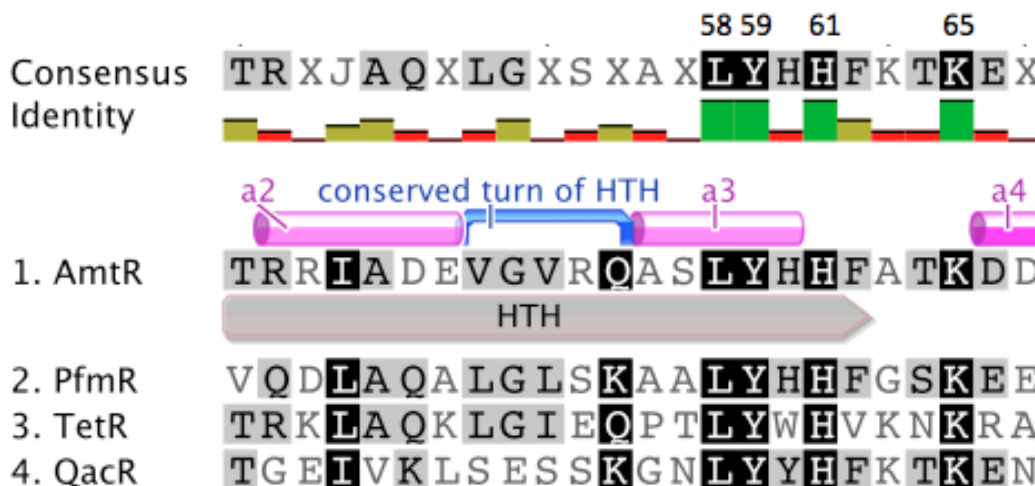


Figure 2.16. AmtR DBD domain alignment with PfmR, TetR and QacR illustrating the conserved residues Leu₅₈, Tyr₅₉, His₆₁ and Lys₆₅. Residues Tyr₅₉, His₆₁ and Lys₆₅ are responsible for interacting with nucleotides in the major groove of DNA containing the recognition sites of palindromic sequence. Lys₆₅ plays an instrumental role in both DBD stabilisation and DNA binding.

The LBD domains of TetR structures are inherently disparate because of the diversity of regulatory functions played by the TetR family, however there are a small number of conserved structural features. The LBD is built around a central triangle composed of helices $\alpha 5$, $\alpha 6$ and $\alpha 7$. Helices $\alpha 5$ and $\alpha 7$ generally run antiparallel with each other so that the start of helix $\alpha 5$ and the end of helix $\alpha 7$ are close to each other. This structure is also observed in AmtR as mentioned earlier. The variable region of the helix $\alpha 6$ seen in AmtR is conserved in all TetR structures. This is a result of the requirement of helix $\alpha 6$ a to be displaced when the ligand is bound into the binding site (Orth et al., 2000). The movement of helix $\alpha 6$ displaces helix $\alpha 4$ causing the HTH motif to swing in a pendulum-like motion by about 5 ° (Orth et al., 2000). This movement shifts the recognition helix $\alpha 3$ along the major groove of the DNA to disrupt the protein-DNA complex (Orth et al., 2000).

There is only one position in the LBD of TetR structures where there is a strongly conserved residue and it is a glycine that is found in the loop between

helices $\alpha 7$ and $\alpha 8$, and the angle between these helices is generally conserved at 58° with a standard deviation of 6° (Yu et al., 2010). The conserved glycine has been shown to play a critical role in maintaining the angle and stabilization of the LBD (Yu et al., 2010). This is also observed in AmtR, which actually has two glycines (Gly₁₅₆ & Gly₁₅₇) present in the loop and the angle between helices $\alpha 7$ and $\alpha 8$ is 59° (Figure 2.17). Helices $\alpha 8$ and $\alpha 9$ constitute the dimerization interface forming an orthogonally packed four-helix bundle with the same helices (helix $\alpha 8'$ and helix $\alpha 9'$) from the other monomer.

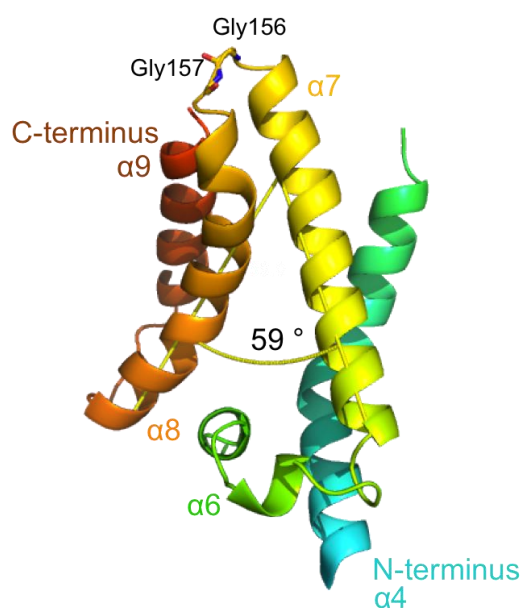


Figure 2.17. Cartoon diagram of AmtR LBD showing 59° angle of helices $\alpha 7$ and $\alpha 8$. Helix $\alpha 7$ is coloured yellow and helix $\alpha 8$ is coloured rust. The two conserved glycine residues Gly₁₅₆ and Gly₁₅₇ are also shown in the loop between helices $\alpha 7$ and $\alpha 8$.

The size of TetR ligand binding cavities range from 25 to 5278 \AA^3 . Each monomer possesses one cavity so that each dimer can bind two identical ligands simultaneously (Yu et al., 2010). The binding cavities are generally characterised as hydrophobic with an enrichment of aromatic residues (Phe, Trp and Tyr). The hydrophobic binding cavities have been shown to bind to small hydrophobic molecules or ring-containing molecules, which allow favourable stacking and van der Waals interactions with the aromatic residues (Yu et al., 2010). The AmtR binding cavity is approximately 125.7 \AA^3 and consists of 11 residues containing 7 polar residues and 5 non-polar residues including 4 aromatic residues. Glu₁₇₀ on $\alpha 9$ forms a potential gate of the pocket that extends into a well-defined entrance site that is formed by helices $\alpha 6$ - $\alpha 9$. The similarity of the AmtR binding pocket to

that of other TetR structures indicates that AmtR likely binds a small hydrophobic molecule or a ring containing ligand.

In summary, the AmtR structure is a homodimer, with each monomer containing a N-terminal DBD and a C-terminal LBD. The DBD shares conserved residues and structural characteristics with the transcription factor QacR, suggesting that it binds DNA as a pair of dimers in the same mechanism as QacR. Like all TetR structures, the AmtR LBD is unique and forms a large binding pocket in the lower centre of each monomer, which is accessed by a long entrance site. The pocket has an enrichment of aromatic residues and both hydrophobic and hydrophilic residues, suggesting that it binds a ring containing signalling molecule. The residue Glu₁₇₀ on helix $\alpha 8$ sits at the entrance of the pocket and shows density for alternative conformations indicating that it may act as a gate that can change position based on the conformation of AmtR and help stabilise ligand binding.

2.3.3 Identification of AmtR Regulated Genes and Recognition Sequence

2.3.3.1 AmtR Regulon

To determine the genes under the control of *M. smegmatis* AmtR an initial search of the *M. smegmatis* genome using variations of the *C. glutamicum* AmtR recognition motif was conducted and one putative semi-palindromic sequence in the promoter region upstream of the *vapBC*_{msmeg_1283/4} operon was identified. The *C. glutamicum* AmtR binding motif is ATCTATAGn₁₋₄ATAG (Figure 2.18), which is present in the promoter region of the *vapBC*_{msmeg_1283/4} operon at -29 to -47 upstream of the operon start site; however, the interspacing region between the two half sites is 9 bp compared to the 1-4 bp seen in *C. glutamicum*.

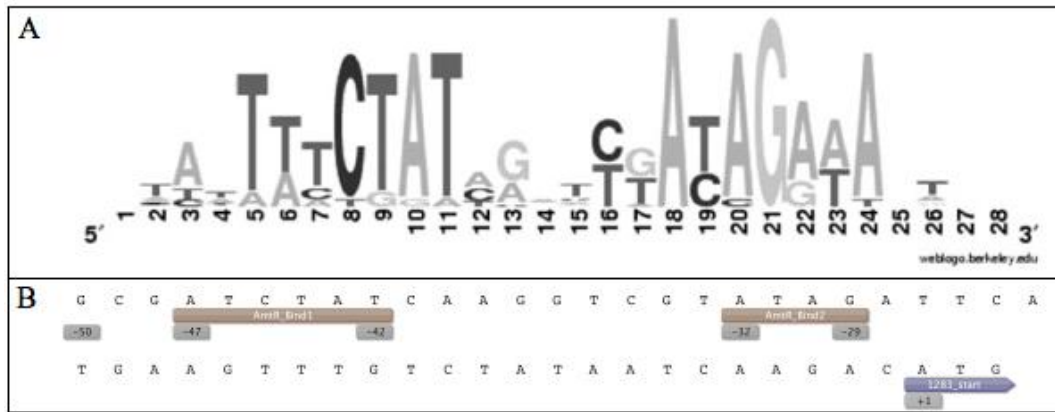


Figure 2.18. The *C. glutamicum* AmtR recognition motif identified in the upstream promoter region of the *M. smegmatis* *vapBC* operon *msmeg_1283/4* (A) The AmtR recognition motif of *C. glutamicum*. The height of letters represents the frequency of corresponding nucleotides in the motif. Figure from Beckers et al. (2005). (B) The *M. smegmatis* *vapBC* operon (*msmeg_1283/4*) promoter region showing *C. glutamicum* AmtR bind motif upstream of the start codon for the *msmeg_1283* gene.

Additionally, microarray analysis of the *M. smegmatis* AmtR deletion (*M. smegmatis* $\Delta amtr$) strain by our collaborators at Otago University identified a number of genes found to be differentially expressed in the absence of AmtR. The main genes found to be regulated by AmtR were *msmeg_2184*, *msmeg_2185*, *msmeg_2186*, *msmeg_2187* and *msmeg_2189*. These appear to form an operon and are annotated to encode an amino acid permease, two hypothetical proteins, a urea carboxylase (UC), and an allophanate hydrolase (AH) respectively (Figure 2.19).

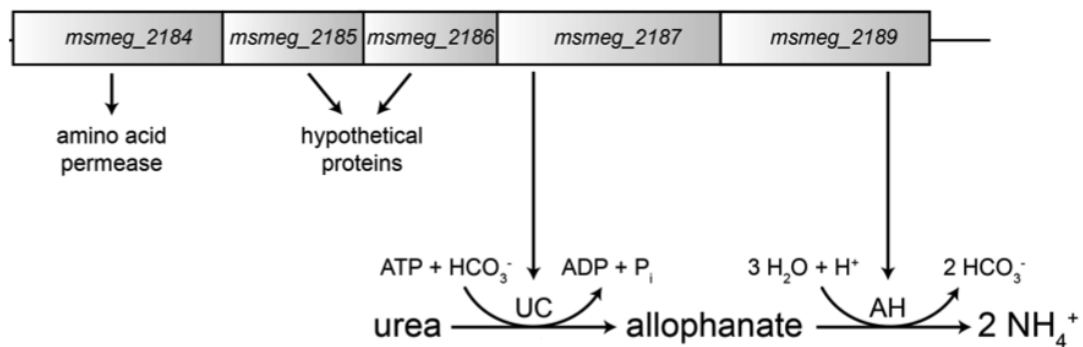


Figure 2.19. Genomic and chemical schematic of the putative metabolic pathway encoded by the urea carboxylase and allophanate hydrolase (UC/AH) operon in the *M. smegmatis* AmtR regulon. The AmtR protein functions as transcriptional repressor of the UC/AH encoding operon under nitrogen surplus. When nitrogen levels drop, AmtR is released from the *msmeg_2184* promoter region, which activates transcription of the genes encoding UC/AH, an energy-dependent urea degradation pathway (Petridis, 2015).

None of the genes identified in the above operon have been functionally characterised in *M. smegmatis* and their function remains hypothetical; however, in other organisms, the UC and AH proteins are known to form an enzyme complex that converts urea to ammonium via an ATP-dependent process (Figure 2.19) (Whitney, Cooper & Magasanik, 1973). This indicates that in *M. smegmatis*, AmtR controls a regulon of genes involved in alternative nitrogen uptake (amino acid permease) and urea metabolism (UC and AH), confirming that AmtR functions as a nitrogen regulator in mycobacteria. The results of the microarray analysis and subsequent confirmation of the *M. smegmatis* AmtR regulon are elaborated in Michael Petridu's PhD thesis (Petridis, 2015).

2.3.3.2 AmtR Recognition DNA Sequence

The upstream regions of all differentially expressed genes in the *M. smegmatis* $\Delta amtr$ strain were analysed and a 16 bp consensus sequence of TCTGTn₄₋₆ACAG was identified as a putative AmtR recognition motif. Promoter regions from the differentially expressed genes *msmeg_2184* (P_{msmeg_2184}) and *msmeg_2185* (P_{msmeg_2185}) were selected as representative binding sites and tested for AmtR binding by electrophoretic mobility shift assays (EMSA) using purified and exhaustively dialysed AmtR protein from *M. smegmatis* (Sections 2.2.2.8 and 2.2.3).

AmtR caused a mobility shift only when incubated with P_{msmeg_2184} at a DNA:protein molar ratio of 1:50 (Figure 2.20). To determine specificity of binding, unlabelled specific DNA was added in 120 x excess resulting in complete abolition of the shift (Figure 2.20). Non-specific DNA (promoter region of *msmeg_phoH2* gene) was added in 103 x excess (accounting for size difference between specific and non-specific DNA) and did not affect the mobility shift (Figure 2.20). This demonstrates that AmtR binds specifically to the upstream promoter region of *msmeg_2184*, albeit with low affinity.

To identify the *M. smegmatis* AmtR binding motif, P_{msmeg_2184} was mutated and the EMSA was repeated. Using the consensus sequence TCTGTn₄₋₆ACAG determined by the microarray experiment, we mutated six nucleotides to adenine in the 40 bp DNA fragment (Figure 2.20). The shift was not observed when the

Chapter Two

EMSA DNA fragment was replaced by the mutated sequence (Figure 2.20). This result demonstrates that the specificity of the nucleotides present in the two half sites of the recognition sequence are critical for binding. To test for ligand-dependent binding, we repeated the above assays in the presence of various concentrations of putative ligands, however no conclusive results were obtained.

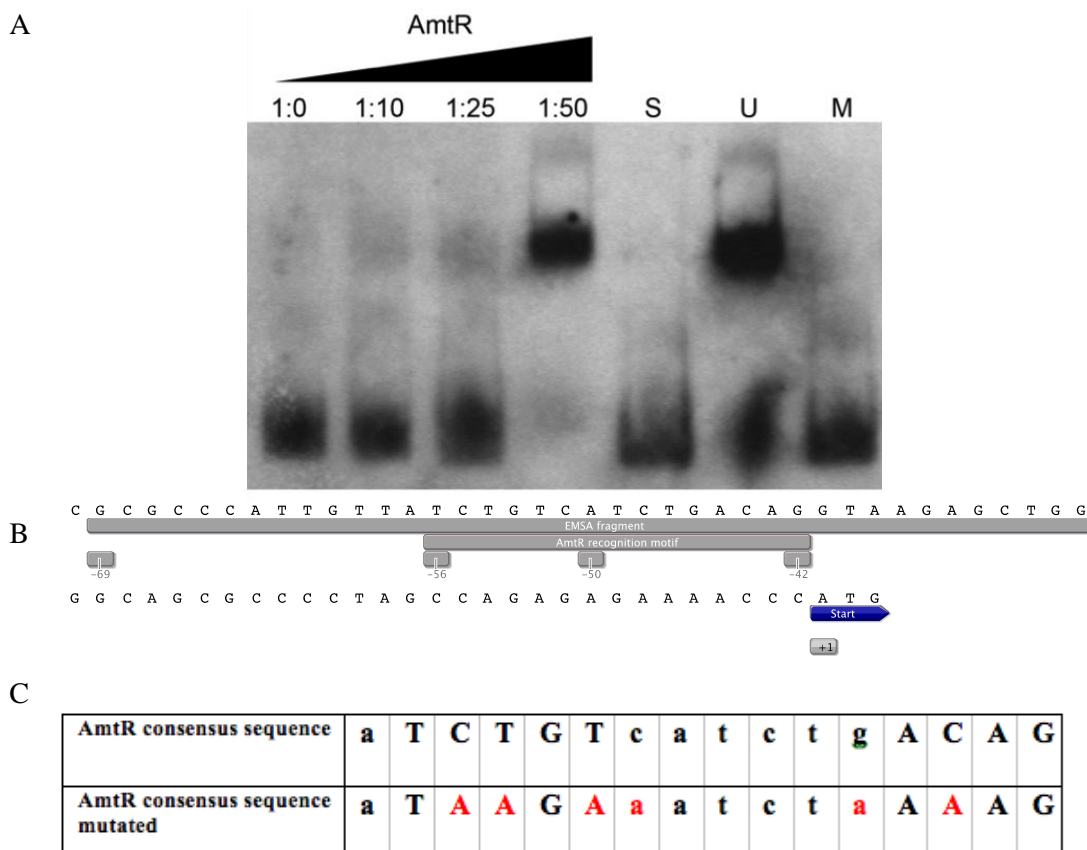


Figure 2.20. *In vitro* DNA binding of *M. smegmatis* AmtR to *msmeg_2184* promoter DNA. (A) EMSA with AmtR protein and unmutated vs mutated DIG-labelled 40-bp DNA fragment including the putative AmtR-binding site in *P_{msmeg_2184}*. The molar ratios of DNA to protein are shown above. Controls for binding specificity (lane S) using 103-fold excess of unlabelled *P_{msmeg_2184}* (lane S) and using 103-fold excess of unlabelled promoter region of the unspecific *phoH2* gene (lane U) are shown. EMSA using a mutated promoter region of *P_{msmeg_2184}* is shown in lane M. **(B)** *P_{msmeg_2184}* containing the AmtR binding recognition sequence at -42 bp to -56 bp. **(C)** Unmutated vs mutated AmtR-binding region in *P_{msmeg_2184}*. Nucleotides in the inverted repeats were mutated (shown in red) using site-directed mutagenesis.

The alignment of all the putative *M. smegmatis* AmtR recognition sites used in this study (Figure 2.21) shows that the *P_{msmeg_2184}* and *P_{msmeg_2185}* sequences are very similar, sharing 11 nucleotides, including 9 nucleotides in the two half sites. The main difference between the two sequences is that the spacing region between the two half sites is 6 bp in *P_{msmeg_2184}* and 4 bp in *P_{msmeg_2185}*. Although both sequences are very similar, *M. smegmatis* AmtR only recognised

and bound the P_{msmeg_2184} sequence. The specificity of AmtR for P_{msmeg_2184} and not P_{msmeg_2185} , in combination with the microarray analysis and the location of each gene to each other in the genome suggests that *msmeg_2184*, *msmeg_2185*, *msmeg_2186*, *msmeg_2187* and *msmeg_2189* form an operon, which is controlled by AmtR at the promoter region of *msmeg_2184*. The *vapBC_{msmeg_1283/4}* promoter region sequence, which AmtR does not recognise, also shares 11 nucleotides with the *msmeg_2184* recognition motif (Figure 2.21). Similar to *msmeg_2185*, the main difference between *msmeg_2184* and *vapBC_{msmeg_1283/4}* is the spacing between the two half sites, where it is 6 bp in *msmeg_2184* and 9 bp in *vapBC_{msmeg_1283/4}*. This indicates that the spacing region between each half site of the recognition sequence is critical for binding.

<i>msmeg_2184</i>	A T C T G T C A T C T G _ _ _ A C A G
<i>msmeg_2185</i>	G T C T G T G A C C _ _ _ _ _ A C A G
<i>msmeg_1283/4</i>	A T C T A T C A A G G T C G T A T A G

Figure 2.21. Alignment of all *M. smegmatis* putative AmtR recognition motifs used in EMSA. AmtR bound to *msmeg_2184* but not *msmeg_2185* or *vapBC_{msmeg_1283/4}*. Nucleotides in red represent shared conservation with the *msmeg_2184* AmtR recognition motif.

Recent characterisation of AmtR from the actinomycetes family member, *Streptomyces avermitilis*, has revealed another organism specific AmtR recognition motif and a putative actinomycetes consensus semi-palindromic 22 bp AmtR recognition motif: $n_4CTnTCn_4GACnGn_4$ (Figure 2.22) (Chen et al., 2013).

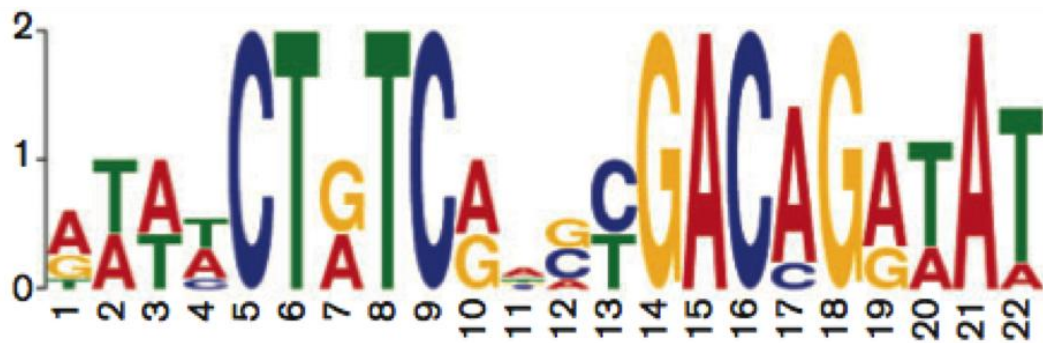


Figure 2.22. MEME-derived consensus sequence from the upstream regions of all AmtR regulated genes. The height of each letter in the logo is proportional to the frequency of the nucleotide appearance. Figure from Chen et al., (2013).

To determine the new consensus sequence for the actinomycetes AmtR recognition motif, given that we determined the *M. smegmatis* AmtR binding sequence as ATCTGTCATCTGACAG, the actinomycetes representative AmtR recognition sequences were aligned to look for conserved nucleotides (Figure

2.23). The first half site of the *M. smegmatis* recognition motif is similar to the first half site of the *C. glutamicum* AmtR recognition motif, with the exception of the 5th bp being guanine instead of adenine, respectively. The spacing region between the two half sites is 6 bp in both the *M. smegmatis* and *C. glutamicum* AmtR recognition motifs. The second half site of the AmtR motif in both organisms varies only at the 16th bp, where it is a cytosine in *M. smegmatis* and a thymine in *C. glutamicum*. In *S. avermitilis*, there are conserved nucleotides in both of the half sites, with the second half site being identical to the *M. smegmatis* second half site. The main difference between *S. avermitilis* and the two other motifs is the spacing region, which is extended by 2 bp. Based on the alignments of these representative AmtR recognition motifs the new actinomycetes AmtR consensus sequence for binding is the semi-palindromic 16-18 bp sequence ATnTnTn₆₋₈AnAG (Figure 2.23).

Organism	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
<i>M. smegmatis</i>	A	T	C	T	G	T	C	A	T	C	T	G	_	_	A	C	A	G
<i>C. glutamicum</i>	A	T	C	T	A	T	A	G	N	N	N	N	_	_	A	T	A	G
<i>S. avermitilis</i>	A	T	T	T	C	T	G	T	C	G	C	C	C	G	A	C	A	G
Actinomycetes consensus	A	T	N	T	N	T	N	N	N	N	N	N	N	N	A	N	A	G

Figure 2.23. Actinomycetes consensus AmtR recognition motif. Alignment of the actinomycetes representative AmtR recognition sequences from *M. smegmatis*, *C. glutamicum* and *S. avermitilis* show the conserved regions of the motif. Red nucleotides represent nucleotides shared between the motifs (positions 1, 2, 4, 6, 15, 17 and 18).

Using a combination of microarray analysis and EMSA, this study has revealed the *M. smegmatis* AmtR regulon and the recognition sequence required for AmtR binding. The AmtR regulon is composed of genes that form an operon, which encode for proteins involved in alternative nitrogen uptake and urea metabolism. By testing native and mutated putative AmtR recognition sequences by EMSA, it was possible to determine the *M. smegmatis* AmtR recognition sequence, which is very similar to the *C. glutamicum* AmtR recognition motif. This study also revealed that both the specificity of nucleotides present in the two half sites and the spacing between these two half sites of the recognition sequence are crucial for AmtR binding. By comparing the *M. smegmatis* AmtR recognition sequence and all other representative AmtR DNA binding motifs, it was possible to generate a modified actinomycetes AmtR consensus recognition motif.

2.3.4 AmtR Signalling Molecules

AmtR, like all TetR transcription factors, is known to act as a repressor of gene expression when bound to DNA in the promoter region of the genes under its regulation. In *C. glutamicum*, AmtR repression is lifted as a result of direct interaction with the adenylylated P_{II} signalling protein GlnK, which only occurs under nitrogen limiting conditions (Beckers et al., 2005). To investigate GlnK as a signalling protein for *M. smegmatis* AmtR, both proteins (AmtR and GlnK in both the native (GlnK) and adenylylated (GlnK-AMP) form) (Figure 2.24) were purified and subjected to Size Exclusion Chromatography (SEC) pull down assays. To produce GlnK-AMP *M. smegmatis* overexpression cultures were grown in nitrogen limiting conditions to induce the post-translational adenylylation of GlnK. Adenylylation was confirmed by the enrichment of a band at 15 kDa which only appeared under nitrogen limiting growth conditions (Figure 2.24). The results were validated with both denaturing and native protein gel electrophoresis (PAGE) analysis, however mass spectrometry is required to confirm accurate protein composition.

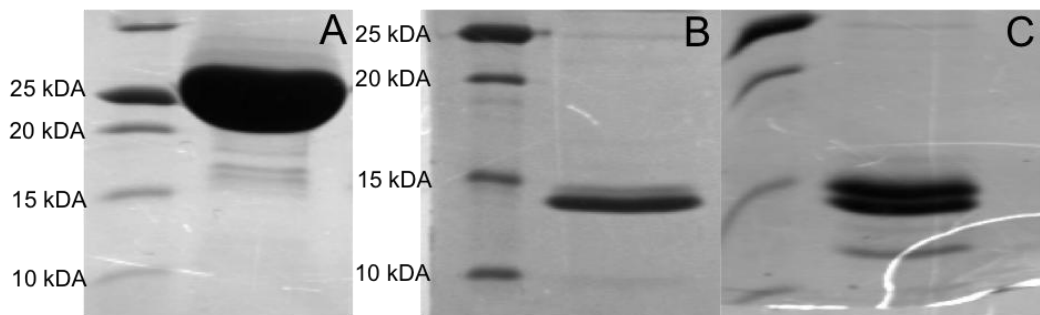


Figure 2.24. Protein purification and 16% SDS-PAGE analysis of AmtR, GlnK and GlnK-AMP. (A) A band at approximately 25 kDa representing AmtR was purified from an *M. smegmatis* overexpression culture. (B) A band at approximately 14 kDa shows GlnK purified from an *M. smegmatis* overexpression culture grown in nitrogen sufficient conditions. (C) Two bands showing GlnK and GlnK-AMP purified from *M. smegmatis* overexpression cultures grown in nitrogen limiting conditions. The band at 15 kDa represents GlnK-AMP and the band at approximately 14 kDa represents GlnK.

For the SEC pull down assays, initially the purified GlnK, GlnK-AMP and AmtR were run individually on a SEC column to determine the size and elution trace of each protein unbound (Figure 2.25). GlnK proteins are generally trimeric with each monomer approximately 14 kDa. The *M. smegmatis* GlnK monomer is 14.4 kDa and eluted off the SEC column as a trimer of approximately 43 kDa (Figure 2.25). GlnK-AMP is also generally a trimer with each monomer able to be

Chapter Two

adenylylated. The *M. smegmatis* GlnK-AMP also eluted off the SEC column as a trimer of approximately 44 kDa (Figure 2.25). AmtR is generally dimeric with each monomer approximately 25 kDa in size. The *M. smegmatis* AmtR monomer is 24.3 kDa and the purified protein elutes off the column as a dimer of approximately 49 kDa (Figure 2.25).

Purified AmtR was then incubated with either purified GlnK or GlnK-AMP and then run on a SEC column. If an AmtR-GlnK interaction did occur, with the possibility of each GlnK monomer interacting with one AmtR dimer, the SEC elution trace should have a peak from approximately 91 kDa (1 AmtR dimer : 1 GlnK trimer) to 135 kDa (1 AmtR dimer : 2 GlnK trimers). No new peaks in this size range were observed and an increased peak in the size range 43-48 kDa was observed (Figure 2.25), which suggested that no interaction was observed between AmtR and GlnK or AmtR and GlnK-AMP in this pull down experiment.

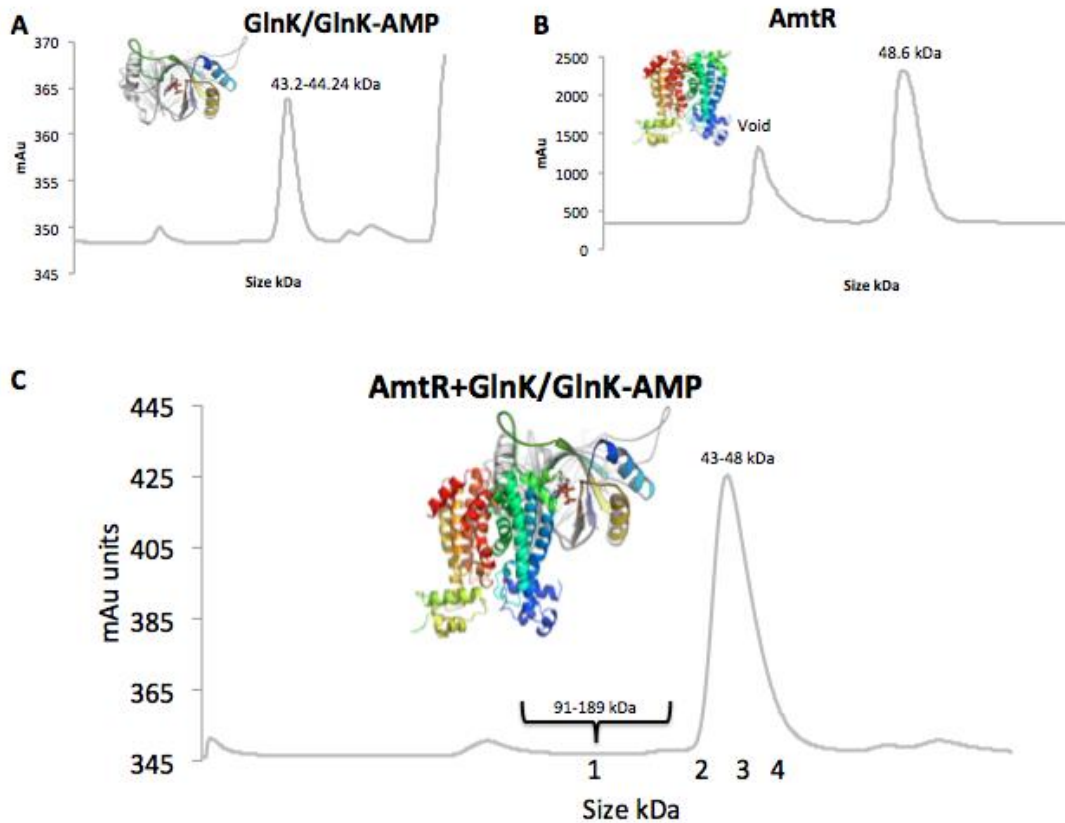


Figure 2.25. AmtR and GlnK SEC pull down assays. (A) SEC elution trace of purified GlnK and GlnK-AMP showing both proteins eluted off the column as trimers of approximately 43-48 kDa. (B) SEC elution trace of purified AmtR showing the protein eluted off the column as a dimer of approximately 49 kDa. (C) AmtR + GlnK (native & adenylylated) pull down assay SEC elution trace. The trace showed that both proteins come off in the same size range as the control elutions and there is no interaction between the proteins, which would have been represented by a peak within the size range 91-189 kDa. The numbering on the trace represents where samples of elution fractions were used for PAGE verification.

To validate this observation, elution fractions from each peak were analysed by both 16 % SDS and native-PAGE to determine the protein composition of each peak (Figure 2.26). The native-PAGE gel analysis shows that in the load containing AmtR and GlnK-AMP the proteins are unbound. Fraction 1, which represents the size range from 91-135 kDa where the bound proteins were expected to elute, contained no protein. Fractions 2, 3 and 4 of the single peak in the elution trace had both AmtR and GlnK-AMP present, with AmtR in higher abundance in the left side and GlnK-AMP in higher abundance on the right side of the peak, as expected due to their size differences. The native-PAGE analysis of the same fractions in the peak showed there was no interaction between the proteins (Figure 2.26). The results of the pull down assay demonstrate that AmtR

and GlnK (in its native or modified form) do not interact. This suggests that GlnK is not a signalling protein for AmtR to lift repression, and, therefore, the *M. smegmatis* AmtR regulation is not controlled by the same signalling mechanism as the *C. glutamicum* AmtR.

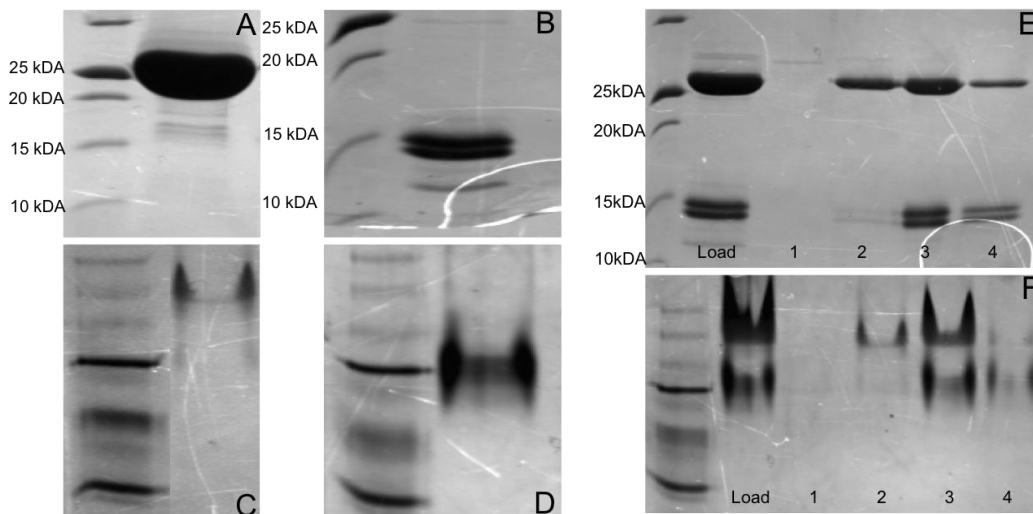


Figure 2.26. PAGE analysis of AmtR and GlnK SEC pull down assays. (A) 16 % SDS PAGE of AmtR shown in the monomeric state at ~24 kDa. (B) 16 % SDS PAGE of GlnK-AMP, represented as the upper band in its monomeric state at ~15 kDa (C) Native PAGE of AmtR in its dimeric state. Ladder is provided for comparative value only as sizes are inaccurate in non-denaturing conditions (D) Native gel electrophoresis of GlnK-AMP in its trimeric state. (E) 16 % PAGE of AmtR + GlnK-AMP SEC pull down assay elution peaks. The load is AmtR + GlnK-AMP before being run through the SEC. Fraction 1 represents the fractions within the size range 91-135 kDa. Fraction 2 represents the left side, fraction 3 represents the centre and fraction 4 represents the right side of the only peak observed. (F) Native gel electrophoresis of AmtR and GlnK-AMP SEC pull down assay elution peaks with the same lanes as (E).

AmtR belongs to the TetR protein family, which is characterised by their regulatory control being mediated by the binding of small signalling molecules. The signalling molecule induces a conformational change in the TetR structure, which disrupts the protein-DNA interaction and results in expression of the genes. To test if *M. smegmatis* AmtR controlled its regulon in this manner, we generated a ligand library of potential molecules that could act as signalling molecules for AmtR. Because the microarray analysis showed that the AmtR regulon is primarily composed of genes involved in alternative nitrogen metabolism, the ligand library (Table 2.1 in Methods Section 2.2.6.2) was generated by selecting biological molecules that were (I) substrates or metabolites in nitrogen metabolism (II) general cellular signalling molecules and (III) had compatible chemistry with the AmtR binding pocket.

All putative signalling molecules were screened against purified and extensively dialysed AmtR by a range of methods, including (1) Fluorescence binding assays, (2) electrophoretic mobility shift assays (EMSA) of AmtR bound to the *msmeg_2184* promoter region to test if putative ligands disrupted the protein-DNA interaction, (3) co-crystallization and post-soaking of AmtR crystals with putative ligands for X-ray crystallography and (4) modelling of putative ligands into the binding pocket of the AmtR regulatory domain. Using this multi-technique approach, no conclusive results were obtained to confidently identify any of the above molecules as a ligand for AmtR.

Next, high throughput differential scanning fluorimetry utilizing Biolog plates containing 96 different nitrogenous sources or metabolites per plate were used to screen for signalling molecules (Petridis, 2015). Purified and extensively dialysed AmtR in both the presence and absence of $P_{msmeg_1283/4}$ were used for this screen. An increase in the melting temperature (T_m) of AmtR was observed in presence of DNA showing that AmtR is stabilised when it is bound to DNA (Figure 2.27). The effect of each of the 96 molecules on the AmtR and AmtR + DNA T_m was then tested. Only two of the 96 molecules, xanthine and allantoin, showed a change in the T_m of the AmtR protein. Interestingly, these molecules both caused an increase in the AmtR T_m and the shift was only observed in the presence of DNA (Figure 2.27). This suggests that AmtR stability is further increased when bound to both DNA and either of these molecules.

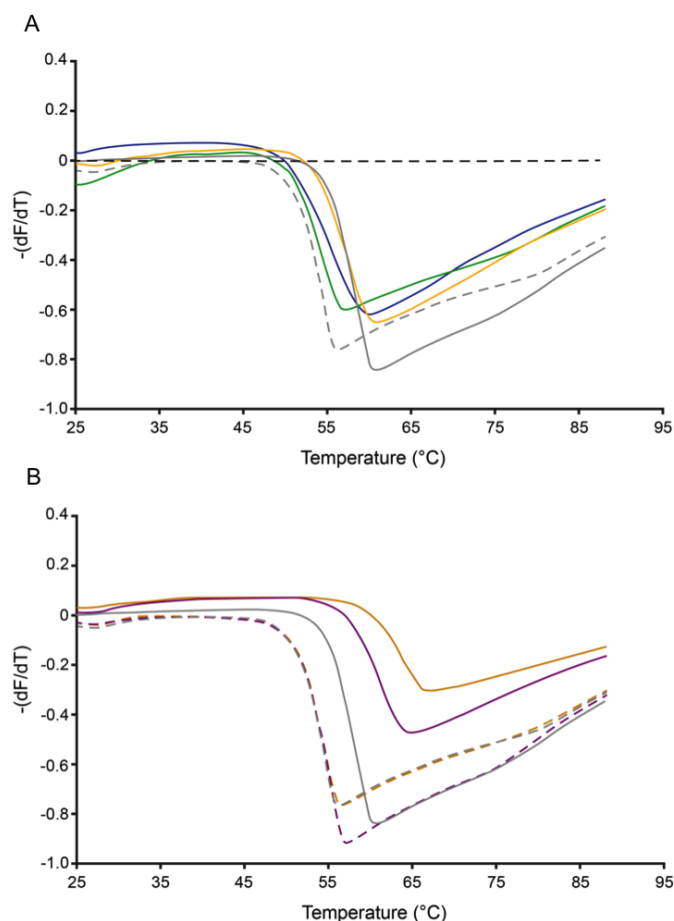


Figure 2.27. Thermal stability of the AmtR protein. (A) Increase in melting temperature (T_m) of the AmtR protein in the presence of DNA (solid grey: AmtR + DNA; dashed grey: AmtR – DNA). The effect of various exemplary compounds is shown (solid yellow: AmtR + DNA + nitrate; solid green: AmtR + DNA + urea; solid blue: AmtR + DNA + glutamate). DNA only (dashed black) has no effect. (B) Increase in the T_m of the AmtR protein after addition of xanthine and allantoin in presence, but not in absence of DNA (solid grey: AmtR + DNA; dashed grey: AmtR – DNA; solid purple: AmtR + DNA + xanthine; dashed purple: AmtR – DNA + xanthine; solid orange: AmtR + DNA + allantoin; dashed orange: AmtR – DNA + allantoin) (Petridis, 2015).

To validate this unexpected result, xanthine was modelled into the AmtR binding pocket (Figure 2.28) using ParDOCK software. Xanthine was manually modeled into the AmtR binding site in a position where there was unresolved density that could not be satisfied by the model during molecular replacement and ligand positioning was optimized using ParDOCK. The modelling shows that xanthine hydrogen bonds with four of the residues that line the binding pocket (Thr₇₆, Asp₁₀₅, Arg₁₃₇ and Phen₁₆₆). Xanthine interacts with the carbonyl group on Phen₁₆₆ and the side chains of Thr₇₆, Asp₁₀₅, and Arg₁₃₇. The strongest interacting residue in the binding pocket is Asp₁₀₅ which interacts with both the nitrogen in the amide group and the nitrogen in the adjacent amine group within xanthine.

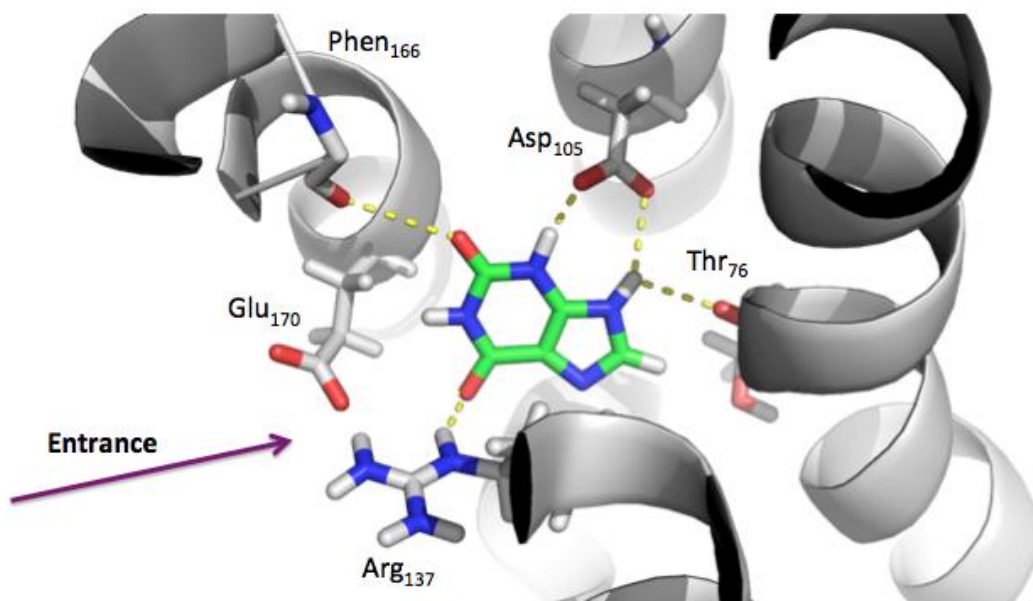


Figure 2.28. Xanthine modeled into the AmtR binding site using the ParDOCK software. The AmtR structure is coloured grey with interacting residues represented as sticks and coloured by atom to show hydrogen bonding. Xanthine is coloured by atom with carbon represented in green. Hydrogen bonding is represented by yellow dotted lines.

In summary, screening for signalling molecules of AmtR has revealed that the *M. smegmatis* AmtR does not have the same P_{II} protein-protein signalling mechanism as its close relative *C. glutamicum*. Small molecule screening has revealed that AmtR may use the nitrogenous molecules xanthine and allantoin as signalling molecules, however not in the same manner as other TetR structures. In fact, these molecules appear to stabilise the AmtR:DNA complex. Using this high throughput screening technique, we have identified two putative ligands that appear to be functioning via a co-repressor mode of regulation to control AmtR regulation.

2.4 Discussion

M. smegmatis is a soil bacteria that is exposed to fluctuating nutrient levels and in particular, limited nitrogen availability. It is characterised by its scavenging ability to transport and utilise a range of nitrogenous compounds in the absence of ammonia (Amon et al., 2009). The genome has a high prevalence of genes involved in urea transport and metabolism, suggesting that urea is an important alternative nitrogen source during nitrogen limitation (Amon et al., 2008). *M. smegmatis* is also one of the members of the actinomycetes that lacks a NtrBC system, which is utilised in most organisms for global nitrogen response. Instead, it has both transcription factors GlnR and AmtR, which are responsible for global nitrogen regulation. GlnR has been identified as an important regulator of nitrogen metabolism, however a number of the genes involved in urea transport and metabolism are not under the control of GlnR. The role of AmtR remained largely unknown but has been indicated as a regulator of these alternative urea metabolic pathways under nitrogen limitation.

Based on the work presented in this thesis and through collaborative efforts with the Cook Laboratory at Otago University, we have structurally and functionally characterised the *M. smegmatis* AmtR. The X-ray crystallographic structure of *M. smegmatis* AmtR is the first AmtR structure of the TetR family to be solved and forms a homodimer with each monomer composed of 10 α -helices that constitute two domains. The C-terminal domain is the ligand binding domain (LBD) defined by helices $\alpha 5$ - $\alpha 10$ that has a deep central binding pocket. The N-terminal domain is the DNA binding domain (DBD) composed of helices $\alpha 1$ - $\alpha 3$ that is defined by the Helix-Turn-Helix (HTH). The AmtR regulon has been identified as a cohort of genes including an operon involved in alternative nitrogen uptake and urea metabolism via a two-step energy dependent process. Subsequently, the AmtR DNA recognition sequence was identified and used to generate a modified actinobacteria AmtR binding motif. The signalling mechanism used to lift *M. smegmatis* AmtR repression has been identified with two co-repressor signalling molecules that appear to function via a mechanism of regulation that has not been observed in the TetR protein family until now.

The N-terminal domain of the AmtR structure is defined by the HTH motif and contains conserved residues and a positively charged surface that allows interaction with DNA. The AmtR DBD aligns best with PfmR, a transcriptional repressor from *Thermus thermophilus* HB8 that also belongs to the TetR protein family. The DBD of PfmR resembles the QacR family of TetR transcription factors. QacR transcription factors are unique in their DNA binding mechanisms mainly because they bind DNA as pairs of dimers, where TetR DNA binding is generally by one dimer. The similarities of AmtR and QacR suggest that AmtR may bind DNA as a pair of dimers in the same mechanism as QacR.

The *M. smegmatis* AmtR recognition sequence has been identified as the 16 bp semi-palindromic ATCTGTCATCTGACAG, which shares similarities to its close relative, the *C. glutamicum* AmtR recognition motif. Both the nucleotide specificity and the size of the spacing regions between the two half sites of the recognition sequence region are critical for binding. The specificity of AmtR for *msmeg_2184* over *msmeg_2185* supports the microarray analysis that *msmeg_2184*, *msmeg_2185*, *msmeg_2186*, *msmeg_2187* and *msmeg_2189* form an operon which is controlled by AmtR at the promoter region of *msmeg_2184*. Using the *M. smegmatis* AmtR binding sequence, a new actinomycetes AmtR consensus sequence was generated: the semi-palindromic 16-18 bp sequence ATnTnTn₆₋₈AnAG.

The *M. smegmatis* AmtR LBD forms a large binding pocket in the centre of each monomer, which is accessed by a long entrance site. It is built around a central triangle composed of helices $\alpha 5$, $\alpha 6$ and $\alpha 7$ with a disordered helix $\alpha 6$. The disorder is a result of the requirement of helix $\alpha 6$ to be displaced when the ligand is bound into the binding pocket. The movement of helix $\alpha 6$ induces a conformational change that results in the disruption of the protein-DNA complex. The binding cavity is approximately 125.70 \AA^3 and consists of both hydrophobic and hydrophilic residues, with an enrichment of aromatic residues, similar to other TetR binding pockets, suggesting AmtR binds a ring containing signalling molecule. The residue Glu₁₇₀ on $\alpha 8$ forms a potential gate at the entrance site of the pocket that may influence the stability or efficiency of ligand binding.

AmtR, like all TetR transcription factors, acts as a repressor of gene expression when bound to DNA in the promoter region of the genes that it regulates. Conservation of the general structural features of the LBD and the presence of a large binding pocket suggested that ligand binding is a likely scenario to lift AmtR repression of its regulon. However, in *C. glutamicum* the AmtR is known to lift repression as a result of direct interaction with the adenylylated form of the P_{II} signalling protein GlnK, which only occurs under nitrogen limiting conditions. SEC pull down assays showed that AmtR and GlnK (in its native or adenylylated form) do not interact, suggesting that GlnK is not a signalling protein for AmtR. We conclude that the *M. smegmatis* AmtR regulation is not controlled by the same signalling mechanism as the *C. glutamicum* AmtR.

To test if small molecule binding is a likely scenario to lift AmtR repression we screened a library of potential ligands and identified two molecules, xanthine and allantoin, as potential ligands for AmtR. Interestingly, the interaction was only observed when AmtR was bound to the *msmeg_2184* promoter region. This unexpected result suggests that the AmtR-DNA interaction is actually stabilised by the presence of either of these molecules. All characterised TetR regulators have shown the opposite effect in the presence of their ligands i.e. when the ligand is bound to the protein it induces a conformational change that results in the disruption of the protein-DNA interaction. This is the first example where the repression of a TetR protein is potentially enhanced by the presence of a ligand, which appears to be functioning as a co-repressor. Although this has not been observed in the TetR protein family, it is a common regulatory mechanism present in other transcription factor protein families. An interesting example is the transcriptional repressor PurR, a member of the LacI family, which regulates purine biosynthesis (Schumacher, Choi, Zalkin, & Brennan, 1994). The PurR-DNA complex is dependant on a co-repressor, which can be either hypoxanthine or guanine. Based on these findings we suggest that AmtR functions in the same regulatory manner as PurR, where the absence of the co-repressor releases AmtR from the DNA and lifts the repression of its regulon. One can speculate that the dissipation of xanthine and allantoin is directly related to a decrease in intracellular nitrogen levels. The result would be the release of AmtR repression and subsequent expression of the proteins involved in alternative nitrogen uptake and urea metabolism to respond the limiting nitrogen levels.

Chapter Two

By structurally and functionally characterising the *M. smegmatis* AmtR transcription factor we have begun to elucidate a novel co-repressor mechanism of nutrient signalling that allows the AmtR regulon to provide alternative nitrogen metabolism in the form of urea metabolism. This is of significance as *M. smegmatis* inhabits an environment of constant nitrogen limitation, so the ability to respond with alternative ammonia production in the form of urea metabolism provides a strong adaptive advantage. These results support findings that AmtR functions as an important nitrogen regulator in mycobacteria.

3 Preliminary Characterisation of Cyanobacteria VapBC Systems

3.1 Introduction

Regulation in bacteria is essential for adaptation. For regulation to be effective bacteria must be able to fine tune metabolism by modulating processes at multiple levels of gene expression for efficient energy expenditure in response to nutrient status. The regulatory response occurs at three levels: transcription, post-transcription and post-translation. The process of transcription transposes genetic information from DNA to mRNA, which can then be translated to a functional protein. Transcription of genes into mRNA is regulated by transcription factors (as mentioned in Chapter 2). Post-translational regulation is where the synthesised protein is modified by a range of post-translational regulators to alter the proteins active state. Post-transcriptional regulation is where the population of mRNA is modulated in response to environmental signals by a range of regulators. The importance of post-transcriptional regulation has become increasingly evident over recent years and there is now overwhelming evidence that it plays a much larger role in cellular adaption than previously thought (Hammarlöf et al., 2015).

mRNA is relatively unstable with half lives in the order of minutes, allowing prokaryotic cells to rapidly alter their transcript pool in response to environmental changes (Picard et al., 2009). The instability of mRNA is attributed to the post-transcriptional regulators with which it interacts to determine its fate. These regulators function by recognising specific characteristics of the target RNA, including sequence and/or secondary structure to alter subsequent translation efficiency. An interesting example of this form of regulation is the toxin-antitoxin (TA) protein system. A number of toxin families in the TA systems function as selective mRNA interferases which result in the inhibition of translation (Inouye, 2006). These systems are well documented as stress-response elements and have also been shown to target mRNA transcripts involved in metabolism (Bodogai et al., 2006; Inouye, 2006; McKenzie et al., 2012b).

Chapter Three

There are five types of TA systems, however the Type II TA systems are the most prevalent within the prokaryotic kingdom. They are classified by having a protein antitoxin and toxin. The antitoxin inhibits the toxin activity by binding to the toxin, forming an inactive complex (Gerdes et al., 2005). The majority of Type II toxins are mRNA endoribonucleases that target specific mRNA transcripts to inhibit translation (Yamaguchi et al., 2011). The VapBC family is the largest family by number of the Type II systems. They are defined by their toxic virulence associated protein (VapC), which belongs to the PilT N-terminal domain (PIN domain) family of proteins.

The VapBC TA system is arranged as a bicistronic operon (Figure 1.12 in Chapter 1) with the gene encoding the antitoxin found upstream of the gene encoding the toxin, which typically overlaps by 2-12 bp (Mattison et al., 2006; Ning, Jiang, Liu, & Xu, 2013a; Robson et al., 2009). Both proteins are co-expressed and form an inactive complex that binds to the promoter region upstream of its own operon to allow autoregulation (Figure 1.12). The VapB contains a DNA binding domain that is responsible for binding to inverted repeat (IR) units in the promoter region, which is typically enhanced by the interaction of VapC (Arcus et al., 2011). VapC inhibits translation by specific mRNA degradation via metal dependant ribonuclease activity (Anantharaman & Aravind, 2003; Arcus et al., 2004; Arcus et al., 2011; Clissold & Ponting, 2000; Fatica et al., 2004; Lamanna & Karbstein, 2009; Mattison et al., 2006; Miallau et al., 2009). Under normal growth conditions VapC and VapB are co-expressed forming a non-toxic complex which autorepresses VapBC expression. Under specific stress conditions this repression is lifted by the degradation of VapB and results in activation of VapC and subsequent growth arrest via translation inhibition (Gerdes et al., 2005). This allows stressed cells to remain in a dormant or in a non-growing stress-tolerant state until more favourable environmental conditions return. The mode of action of VapBC indicates that they play an important role in growth regulation in response to adaptation (Arcus et al., 2005; Arcus et al., 2011; Daines et al., 2007; Mattison et al., 2006; Puskás et al., 2004; Robson et al., 2009).

Several VapBC systems from a range of bacteria have been shown to inhibit translation, and, in some cases, induce growth arrest under stress

Chapter Three

conditions or when overexpressed *in vivo*. However, cellular targets and specific biological roles of these systems are yet to be elucidated (Arcus et al., 2011; Bodogai et al., 2006; Cooper et al., 2009; Daines et al., 2007; Ning, Jiang, Liu, & Xu, 2013a; Sharp et al., 2012; Winther & Gerdes, 2009; Zhang et al., 2004). For example, FitAB in *N. gonorrhoea* is hypothesised to slow intracellular trafficking and replication during host infection (Mattison et al., 2006). NtrPR in the nitrogen-fixing symbiont of legumes *Sinorhizobium meliloti* is hypothesised to adjust metabolic processes during symbiosis and other stress conditions (Puskás et al., 2004). VapBC-5 (also in *S. meliloti*) affects efficiency of symbiosis and nitrogen fixation capacity (Cinege et al., 2014). Several VapBC systems in the pathogenic *Mycobacterium tuberculosis* have been implicated in the stress response to different environmental cues, suggesting they function as an intricate response system during adaptation. The sole VapBC system in *M. smegmatis*, the non-pathogenic relative of *M. tuberculosis*, has been characterised extensively with the cellular targets and biological role identified: VapBC_{1283/4} acts as a post-transcriptional regulator of carbon metabolism (McKenzie et al., 2012b).

VapBC TA systems have become a point of interest because of their overrepresentation in the genomes of pathogenic and environmentally hazardous bacteria. For example, *M. tuberculosis* has 47 *vapBC* operons encoded in its genome, where the non-pathogenic environmental *M. smegmatis* has just one (Arcus et al., 2011; McKenzie et al., 2012b). The toxic bloom-forming cyanobacteria *M. aeruginosa* has 34 *vapBC* operons encoded in the genome, where the non-toxic *Synechocystis* sp. *PCC6803* (*Synechocystis*) has only 12 (Kaneko et al., 2001; Leplae et al., 2011; Makarova et al., 2009). This expansion suggests a functional link between toxicity and VapBC systems. As mentioned earlier, several VapBC systems in *M. tuberculosis* have been implicated in stress response to different environmental cues, suggesting they function as an intricate response system during adaptation. For organisms that host a large number of VapBC systems this would provide the capacity for efficient adaption to fluctuating and/or inhospitable conditions. The function of VapBC systems in cyanobacteria is not known. In fact, only one of the 12 *Synechocystis* VapBC systems have been characterised, and none from any toxin-forming species (Ning, Liu, Xu, Zhuang, Wen, et al., 2013b).

Chapter Three

Cyanobacteria are primary producers by nature of their autotrophy, which is essential for the existence of all higher organisms (Herrero, 2008). Members of the cyanobacteria phylum can be found in almost every terrestrial and aquatic habitat. Aquatic cyanobacteria can reproduce explosively in certain conditions and are known for their extensive blooms. These blooms are detrimental to other species in close proximity and can often be toxic to humans and animals because of the cyanotoxins produced (Briand et al., 2003; Falconer & Humpage, 2005; Paerl et al., 2001). Blooms often result in closure of recreational waters and the treatments used to eliminate the blooms are expensive and can be fatal to livestock and other aquatic organisms (Frangeul et al., 2008; Steffensen, 2008).

Synechocystis is used as the model organism to investigate the physiology of cyanobacteria. It has 12 annotated *vapBC* loci, and, until recently, none were characterised. Ning et al. (2013) investigated VapBC10 (*ssr2962/slr1767* where the first letter “s” stands for *Synechocystis*, the second letters indicate gene length with “s” indicating less than 100 codons and “l” indicating longer than 100 codons, and the last letter “r” indicating direction of transcription on the circular genome is in the right direction) of *Synechocystis*. They were able to confirm that the operon functions as a typical VapBC system, with the exception of its mode of autoregulation. *vapB10* and *vapC10* form a bicistronic operon and both genes are co-transcribed under normal growth conditions. Expression of VapBC10 has no effect on growth but expression of VapC10 caused growth arrest in *E. coli*, which could be reversed by expressing VapB10. Because VapC10 shows the characteristic bacteriostatic phenotype of a VapC toxin, it is proposed that VapC10 exerts its growth arrest via translation inhibition, like other characterised VapC toxins, however this has not been confirmed experimentally.

VapBC systems are normally characterised by their repressive autoregulation, where VapB binds to the promoter region of the operon and represses expression. The efficiency of binding is enhanced when VapC is bound to VapB. Ning et al. (2013) showed, firstly, that VapB10 binds to the promoter region of its own operon and positively autoregulates transcription. They identified the specific DNA sequence that VapB10 recognises: the imperfect IR sequence TTTCCCT-2N-AGGGTAA present in the promoter region upstream of the *ssr2962/slr1767* operon (Figure 1.16 in Chapter 1). Secondly, they showed

Chapter Three

that VapC10 abolishes binding of VapB10 to DNA. These results reveal that VapB10 can positively autoregulate the operon, which can be inhibited by VapBC10.

Because this is the only characterised VapBC system in cyanobacteria and the functional mechanism is still unknown, it is difficult to draw any conclusions about the biological role of these VapBC systems. However, because of the link between the number of VapBC systems and the toxicity of the organism it is a very promising avenue to investigate. Obtaining a better understanding of the role that VapBC systems play may also provide a novel strategy for bloom control. This chapter presents research on the preliminary characterisation of both *M. aeruginosa* and *Synechocystis* VapBC systems.

3.2 Methods

3.2.1 DNA Manipulation and Cloning

3.2.1.1 Gene, Primer and Oligonucleotide Design

3.2.1.1.1 *M. aeruginosa* NIES-843 Gene Design

The *M. aeruginosa* NIES-843 *vapBC* operons *MAE49640/50* and *MAE43230/20* were ordered from GeneArt (Life Technologies) with a 5' PstI restriction site and a 3' HindIII restriction site with the stop codon removed to allow for ligation into pET28b-PstI with a C-terminal His tag. The operons were extracted from the GeneArt plasmid by restriction digest as per (Section 2.2.1.6) using PstI and HindIII and inserted into the pET28b-PstI expression vector and the vector was transformed into *E. coli* DH5 α cells prior to transformation and protein expression in *E. coli* BL21 (DE3) cells.

3.2.1.1.2 *M. aeruginosa* NIES – 843 Primer Design

Primers were designed to amplify *vapC* genes from GeneArt *vapBC* plasmids with a 5' PstI restriction site and a 3' HindIII restriction site with the stop codon removed to allow for ligation into pET28b-PstI with a C-terminal His tag. An extra 3-6 bases were included to allow for efficient cleavage, depending on the restriction enzyme used. The PCR products were then digested as per Section 2.2.1.6 with the corresponding restriction enzymes, and inserted into the expression vector. The vector was transformed into *E. coli* DH5 α cells prior to transformation and protein expression in *E. coli* BL21 cells.

Primers were designed for cloning *M. aeruginosa* NIES-843 *vapBC* operons *MAE49640/50* and *MAE43230/20* into alternative *E. coli* expression vectors when pET28b-pstI cloning failed. Primers were designed to amplify the *vapBC* operons from GeneArt plasmids and include the appropriate 5' and 3' restriction sites for ligation with either an N-terminal or C terminal His tag, depending on the plasmid. An extra 3-6 bases were included to allow for efficient cleavage, depending on the restriction enzyme used. Where only a N-terminal His tag was possible a stop codon was re-inserted into *vapC*. The operons were then digested as per Section 2.2.1.6 with the corresponding restriction enzymes and

Chapter Three

inserted into the expression vector, and the vector was transformed into *E. coli* DH5 α cells prior to transformation and protein expression in *E. coli* BL21 (DE3) cells. Table A.3 in the Appendices shows the plasmids used for expression attempts.

Primers were also designed for cloning *M. aeruginosa* NIES-843 *vapBC* operons *MAE49640/50* and *MAE43230/20* using the In-Fusion Ready pOPIN vector Suite (Berrow et al., 2007) when pET28b-pstI and alternative plasmid cloning failed. Primers were designed to amplify the *vapBC* operons from GeneArt plasmids and include the 5' and 3' 15 bp extensions homologous with the sites of linearization of the pOPIN vectors to enable successful In-Fusion cloning. The operons were inserted for expression with an N-terminal His tag (a stop codon was re-inserted into *vapC*) and, depending on the plasmid, an N-terminal fusion protein to reduce VapC toxicity. The operons were inserted into the expression vector and the vector was transformed into *E. coli* DH5 α cells prior to transformation and protein expression in *E. coli* BL21 (DE3) cells. Table A.3 in the Appendices shows all plasmids used for expression trials. All putative sequences were cloned in silico into the vector using Geneious R6 to ensure the protein sequence was correct and the His tags were in frame. Successful transformant plasmids containing the insert were confirmed by sequencing.

3.2.1.1.3 *Synechocystis* Primer Design

Primers for amplification of *Synechocystis* *vapBC* operons from the *Synechocystis* genomic DNA for cloning were initially designed to include the 5' restriction site PstI and the 3' restriction site HindIII with the stop codon removed for ligation into pET28b-PstI with a C-terminal His tag. An extra 3-6 bases were included to allow for efficient cleavage, depending on the restriction enzyme used. The operons were then digested as per Section 2.2.1.6 with PstI and HindIII and inserted into the pET28b-PstI expression vector, and the vector was transformed into *E. coli* DH5 α cells prior to transformation and protein expression in *E. coli* BL21 cells. All putative sequences were cloned insilico into the pET28b-PstI vector using Geneious R6 to ensure the protein sequence was correct and the His tags were in frame.

Chapter Three

Primers for generation of ‘left’ and ‘right’ DNA fragments were designed to be cloned into pUC19 plasmids for generation of unmarked *Synechocystis vapBC* deletion strains (*Synechocystis* $\Delta vapBC$) via homologous chromosomal DNA exchange. The primers to construct the ‘left’ fragment were designed to amplify up to 200 bp of the 5’ end of *vapB* with 0.7-1.0 kb of 5’ upstream non-protein coding DNA with a 5’ EcoRI restriction site and a 3’ SacI restriction site. The primers to construct the ‘right’ fragment were designed to amplify up to 200 bp of the 3’ end of *vapC* with 0.7-1.0 kb of 3’ downstream non-protein coding DNA with a 5’ BamHI restriction site and a 3’ XbaI restriction site. An extra 1-2 bases were included to allow for efficient cleavage, depending on the restriction enzyme used. The fragments were then digested (Section 2.2.1.6) with the respective enzymes and inserted into the pUC19 vector, which was transformed into *E. coli* DH5 α cells prior to transformation into *Synechocystis* for the generation of the knock out strains. All putative sequences were cloned in silico into the pUC19 vector using Geneious R6 to ensure the sequence was correct.

Primers for generation of DNA fragments for electrophoretic mobility shift assays (EMSA) were designed to amplify specific regions from the promoter site of the *Synechocystis vapBC*_{slr1209/1210} operon in the *Synechocystis* genome. The PCR products were 3’-end labelled with digoxigenin-11-ddUTP (DIG) (Section 2.2.1.2).

All primers were supplied by IDT and resuspended in 1 x TE. A list of all genes, primers and oligonucleotides ordered are present in Table A.1 in the Appendices.

3.2.1.2 Polymerase Chain Reactions (PCR)

3.2.1.2.1 PCR Amplification from *Synechocystis* Genomic DNA and GeneArt

Plasmids

PCR reactions were carried out with the proofreading enzyme KAPA HiFi DNA polymerase (Kapa Biosystems). The annealing temperature was 55 °C unless otherwise stated. PCR reactions were carried out in 25 μ l volumes using the following concentrations: 1 x High GC amplification buffer, 0.3 mM

Chapter Three

deoxynucleotide mix (dATP, dCTP, dTTP & dGTP), 2 mM MgSO₄, 1 x Enhancer buffer, 1 U 2G DNA polymerase, 0.3 µM of each primer and 10 ng template DNA. The following cycling conditions were used for amplification:

95 °C	2.00	(min:sec)
98 °C	0.20	x 29
55 °C	0.15	
72 °C	0.45	
72 °C	5.00	

Ligation, preparation of *E. coli* DH5α and BL21 cells and electroporation of *E. coli* cells were done as per Section 2.2.1.7 and 2.2.1.9 in Chapter 2.

3.2.1.2.2 PCR Screening of Transformants.

Plasmids purified from *E. coli* (Section 2.2.1.11), were checked for the presence of the target DNA insert prior to sequencing by performing a PCR using either target specific or T7 primers. PCR reactions were carried out with *Taq* DNA polymerase as per Section 2.2.13. PCR reactions were visualised as in Section 2.2.1.4.1. Alternatively *E. coli* transformants were confirmed by colony PCR where a colony was taken from the transformant plate and used to inoculate a *Taq* PCR reaction.

3.2.1.3 DNA Purification and Quantification

DNA from agarose gels were separated and visualised with SYBR safe™ DNA dye as described in Section 2.2.1.5. The desired bands were extracted using a sterile scalpel blade and DNA purified using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions. DNA was eluted in 30 µl of elution buffer. PCR products from solution were purified using a Qiaquick PCR Product Purification Kit (Qiagen) according to the manufacturer's instructions. DNA purity and concentration were estimated by measuring A₂₆₀/A₂₃₀ nm and A₂₆₀/A₂₈₀ nm ratios using a NanoDrop®ND-1000 Spectrophotometer (Thermo Fisher Scientific).

3.2.1.4 Cloning of *vapBC* Operons into *E. coli* Expression Plasmids

3.2.1.4.1 pET28b-pstI Cloning

pET28b-pstI is an *E. coli* expression vector with kanamycin resistance. The *vapBC* operons and *vapC* genes were digested as per Section 2.2.1.6, purified (Section 2.2.15) and ligated (Section 2.2.1.7) into pET28b-pstI between the selected restriction sites within the multiple cloning site (MCS) allowing for the C-terminal His tag. Vectors containing the inserted gene or operon were transformed into electrocompetent *E. coli* DH5 α cells (Section 2.2.1.9) and plated onto LB agar plates containing kanamycin for selection. The plasmid constructs were purified, then sequenced (Section 2.2.1.12) to confirm correct insertion before being transformed into electrocompetent *E. coli* BL21 (DE3) (Section 2.2.1.9).

3.2.1.4.2 pETDUET and pPROEX – Htb

pETDUET is an *E. coli* expression vector with ampicillin resistance that allows co-expression of two proteins from different cloning sites with a variation of terminal His tags. pPROEX HTb is an *E. coli* expression vector with ampicillin resistance that allows expression of proteins with a N-terminal His tag. Vectors containing the inserted gene or operon were transformed into electrocompetent *E. coli* DH5 α cells (Section 2.2.1.9) and plated onto LB agar plates containing the appropriate antibiotic for selection. The plasmid constructs were purified then sequenced (Section 2.2.1.12) to confirm correct insertion before being transformed into electrocompetent *E. coli* BL21 (Section 2.2.1.9).

3.2.1.4.3 pOPIN Cloning

The pOPIN vector suite in combination with the In-Fusion enzyme (Clontech Laboratories, Mountain View CA) is a one step, high-throughput cloning system for problematic protein expression. Primer extensions are used to flank the 5' and 3' of each insert to allow a single PCR product to be inserted into multiple vectors containing different fusion partners and ampicillin resistance. *vapBC* operons were amplified with primers that incorporated the homologous extensions and then ligated as per the manufacturers instructions into pOPIN plasmids. Vectors containing the inserted operon were transformed into

Chapter Three

electrocompetent *E. coli* DH5 α cells (Section 2.2.1.9) and plated onto LB agar plates + ampicillin for selection. The plasmid constructs were purified, then sequenced (Section 2.2.1.12) to confirm correct insertion before being transformed into electrocompetent *E. coli* Bl21 (Section 2.2.1.9).

3.2.1.5 Preparation of pUC19 Mutant Constructs

pUC19 is an *E. coli* vector with ampicillin resistance. The fragments were digested as per Section 2.2.1.6, purified (Section 2.2.15) and ligated (Section 2.2.1.7) into pUC19 between the selected restriction sites within the MCS. Vectors containing the inserted fragments were transformed into electrocompetent *E. coli* DH5 α cells (Section 2.2.1.9) and plated onto LB agar plates containing kanamycin for selection. The plasmid constructs were purified then sequenced (Section 2.2.1.12) to confirm correct insertion before subsequent cloning of the next fragment into the MCS. The end result was a pUC19 plasmid with both the 5' left and 3' right fragments cloned into the MCS. When both fragments were successfully cloned into pUC19 (pUC19 + L&R construct), a sucrose sensitive and kanamycin resistance cassette (sacRB/Kan(R)) was excised from the pUM24 plasmid by BamH1 digestion and cloned into the BamH1 site in the pUC19 L&R between the left and right fragments (pUC19 + L&R + sacRB/Kan(R) construct).

Plasmid DNA extraction was done as per Section 2.2.1.11 and DNA sequencing was done as per Section 2.2.1.12 in Chapter 2.

3.2.1.6 Preparation of *Synechocystis* Unmarked $\Delta vapBC$ Strains

3.2.1.6.1 Generation of Marked $\Delta vapBC$ Strains

A *Synechocystis* wild type (WT) glycerol stock was streaked onto BG-11 agar plates and grown under low light (approximately 5-20 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$) at 30 °C for 4-5 days. Colonies from the plate were used to inoculate a 50 ml BG-11 culture which was incubated for 3-4 days under normal light (40 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$) at 30 °C, shaking at 150 rpm. Multiple colonies were required for inoculation to prevent photo-inhibition of cells, which occurs when initial biomass is too low. 1-2 ml of culture was centrifuged at 5000 g for 5 min at room temperature. The supernatant was discarded and the cell pellet was washed by

Chapter Three

resuspending in 1 ml of BG-11 media and centrifugation at 5000 g for 5 min at room temperature. The supernatant was removed and fresh BG-11 was added to a final volume of 100-150 μ l. The cells were transferred to a 15 ml falcon tube and 2 μ g of pUC19 + L&R + sacRB/kan(R) construct was added. The solution was mixed and the cells were incubated, with the tube horizontal, for 4-5 hrs under low light at 30 °C and mixed every 1-2 hrs.

Aliquots of both 20 μ l and the remaining cells were plated out onto BG-11 agar plates and incubated at 30 °C under low light for 24 hrs. Post-incubation, 2.5-3 ml of 0.6% agar solution containing kanamycin at 5 μ g/ml, cooled to ~ 42 °C, was added on top of the plates to form an even surface layer being careful not to disrupt colonies. When the agar surface set, the plates were further incubated for 5-7 days under normal light at 30 °C or until colonies appeared. Up to 6 colonies were streaked out on a BG-11 + kanamycin agar plate (30 μ g/ml) and incubated for 7 days under normal light at 30 °C. Confirmation of deletions was done by PCR diagnostic tests (Section 3.2.1.6.3).

3.2.1.6.2 Generation of Unmarked $\Delta vapBC$ Strains

Colonies from the deletion strain agar plate were used to inoculate a 30-50 ml BG-11 media, which was incubated for 3-4 days under low light (5-20 μ mol photons $m^{-2} s^{-1}$) to induce DNA uptake at 30 °C, shaking at 150 rpm. 2-10 ml of the culture was centrifuged at 4000 g for 5 min at room temperature. The supernatant was removed and the cells were washed as in Section 3.2.1.6.1. Fresh BG-11 was added to get a final volume of 100-200 μ l, depending on cell density and 2 μ g of pUC19 + L&R construct was added. The cells were mixed gently and transferred to a 10-20 ml flask and incubated for 4-5 hrs under low light at 30 °C, shaking at 150 rpm and mixed every 1-2 hrs. 900 μ l of BG-11 media was added and the culture was incubated for a further 4 days at 30 °C, 150 rpm under normal light.

Aliquots of 20 μ l, 50 μ l, 100 μ l and the remaining cells were plated on BG-11 + 5 % (w/v) sucrose plates and incubated under low light at 30 °C for 7 days or until colonies appeared. Colonies were then patched out onto both BG-11 + kanamycin and BG-11 + sucrose agar plates to detect unmarked deletion strains.

Chapter Three

Colonies that grew only on BG-11 + sucrose agar plates were then tested by colony PCR (Section 3.2.1.6.3) to confirm deletion and cassette removal. Successful deletion strains were streaked out onto BG-11 + sucrose agar plates to ensure all chromosomes had the unmarked deletion present.

3.2.1.6.3 Colony PCR Confirmation of $\Delta vapBC$ Strains

To confirm marked deletion strains a small proportion of cells from each initial colony of marked $\Delta vapBC$, transformants and WT were subjected to cell lysis via bead beating: cells were mixed with 50 μ l sterile water containing 0.3 g of 0.1 mm and 2.5 mm zirconia beads and shaken in a bead beater for 2 min. Cells were centrifuged at 13000 g for 5 min and 5 μ l of the supernatant was used as the DNA template for a Taq PCR (Section 2.2.13) using the forward left fragment primer and the reverse right fragment primer for each operon. Transformants produced a band >5 kb containing the flanking regions (2-2.5 kb) of the operon and the 3.7 kb cassette that replaced the operon, which is too large to be visualised by gel electrophoresis. The absence of a band represented a successful marked deletion. The WT band was approximately 2.5 kb (operon + flanking regions). Successful deletion colonies were streaked onto a new BG-11 + kanamycin agar plate to ensure all chromosomes had the deletion present.

To confirm unmarked deletion strains, the above procedure was done with cells from each initial colony of unmarked $\Delta vapBC$ transformants and WT. The PCR was done with primers designed to amplify only the operon. Successful deletion strains showed a region of the operon removed, which was represented by a smaller band on the gel compared to the WT PCR product.

3.2.1.7 Glycerol Stocks

Glycerol stocks for long term storage of transformed or mutated bacteria at -80 °C were made by adding 0.5 ml of cells to 0.5 ml of 50 % sterile glycerol.

3.2.2 Protein Expression and Purification

SDS-Polyacrylamide gel electrophoresis (SDS-PAGE) protein analysis and quantification was done as per Section 2.2.2 in Chapter 2.

3.2.2.1 Concentration of Protein Samples

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

3.2.2.2 Small Scale Protein Expression and Purification from *E. coli*

All *E. coli* expression cultures included kanamycin at 50 μ g/ml for selection. Expression of VapBC complexes were initially confirmed with small scale His tag binding tests using Ni Sepharose beads. A single transformed colony either from the freshly transformed plate or streaked from a glycerol stock was used to inoculate a 5 ml LB seeder culture. This culture was grown overnight at 37 °C, shaking at 200 rpm then 100 μ l was used to seed a 10 ml LB expression culture which was grown for approximately 2 hrs until the cells reached OD₆₀₀ 0.5-0.7. Cells were then induced with 1 mM IPTG and incubated overnight at 37 °C (unless otherwise stated), shaking at 200 rpm. Cultures were then centrifuged (4,600 g, 20 min, 4 °C). The rest of the expression trial was done as per Section 2.2.2.6 in Chapter 2.

3.2.2.3 Large Scale Protein Expression and Sonication from *E. coli*

3.2.2.3.1 Large Scale Protein Expression

Large scale protein expression cultures were grown as for small scale protein expression cultures in Section 3.2.2.2 with the exception that 10 ml seeder cultures were used to inoculate 1 L expression cultures to maintain the 1:100 ratio. Cultures were centrifuged (4,600 g, 20 min, 4 °C) and cell pellets either frozen at -80 °C or resuspended in 25 ml lysis buffer (50 mM sodium phosphate buffer, 200 mM NaCl, 20 mM imidazole, pH 7.4) for sonication.

All media recipes are presented in the Appendices.

3.2.2.4 *E. coli* Large Scale Purification

Cells were sonicated on ice using a small probe for 20 s bursts at setting 3-5 with 20 sec rests in between using a Misonix Sonicator (Misonix Inc.) until cells were lysed (1-2 min). The lysate was centrifuged, filtered and SDS-PAGE analysed as per Section 2.2.2.7. The supernatant containing the protein was initially run through a HisTrap HP column (GE Healthcare) and purified via Immobilised Metal Affinity Chromatography (IMAC) on an ÄKTA FPLC system (GE Healthcare) as in Section 2.2.2.8.1. Fractions containing the protein of interest were purified further by size exclusion chromatography (SEC) using a Superdex column (GE Healthcare) as in Section 2.2.2.8.2. Aliquots of fractions (15 µl) containing the protein of interest from both IMAC and SEC were added to 5 µl of 4 x SDS loading buffer for analysis with SDS-PAGE.

3.2.2.4.1 Acid SEC for VapB_{slr1209} Removal and VapC_{slr1210} Purification

VapBC_{slr1209/1210} protein from IMAC purification was concentrated to 0.5 ml using vivaspin concentrators (10 kDa molecular weight cut off) (Sartorius AG) and then diluted to 10 ml with SEC buffer (50 mM phosphate buffer, 200 mM NaCl, pH 3.0) to bring protein solution to pH 3. Protein was then concentrated using vivaspin concentrators (10 kDa molecular weight cut off) to 0.5 ml and filtered with a Nanosep® MF 0.2 µm filter (Pall Corporation) before loading and running on a pre-equilibrated (50 mM phosphate buffer, 200 mM NaCl, pH 3.0) SEC column as per Section 2.2.2.8.2. Fractions containing purified VapC were pooled, brought to pH 7.4, and concentrated before being tested for ribonuclease activity.

3.2.3 Crystallisation Trials for X-ray Crystallography of *Synechocystis*

VapBC_{slr1209/1210}

Initial crystallisation trials for VapBC_{slr1209/1210} employed sitting-drop vapour diffusion as per Section 2.2.4.1 using protein concentrated to 10 mg/ml. Once promising crystallisation conditions had been identified, hanging drop vapour diffusion fine-screens were employed as per Section 2.2.4.2.1 to produce crystals suitable for X-ray analysis. Crystals were tested for X-ray diffraction on an Agilent Supernova X-ray Diffractometer (Agilent Technologies, Liberty Lake

WA) at the University of Waikato using a Titan S2 CCD detector or at the Australian Synchrotron using an ADSC Quantum 210r detector. All Crystals tested for xray diffraction were identified as non protein salt crystals.

3.2.4 VapC_{slr1210} Ribonuclease Activity

An assay to test for ribonuclease activity of VapC proteins was used, which was originally developed by Jo McKenzie from the Proteins and Microbes Laboratory at the University of Waikato (McKenzie et al., 2012a).

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

Urea denaturing-PAGE gels were cast in a Hoefer mini gel casting system and consisted of a 10 % acrylamide resolving gel. Gels were made with 30 % acrylamide with an acrylamide:bisacrylamide ratio of 37.5:1 (Bio-Rad Laboratories), including 6 M urea, and polymerised with 0.05 % ammonium persulphate (APS) and TEMED. RNA samples were mixed in a 1:1 ratio with 2 x formamide loading dye and denatured for 5 min at 70 °C prior to loading onto the gel. Gels were pre-run in 1 x TBE buffer at 70 V for 30 min and wells flushed to remove any leached urea before samples were loaded. Gels were run at 150 V until the xylene dye front reached the end of the gel. For RNA band size estimation, 10 µl of prepared Low Range ssRNA ladder (New England Biolabs, Ipswich MA) was run on each gel. Gels were post-stained with 1x SYBR safeTM DNA gel stain (Invitrogen), and visualised by blue light.

3.2.4.2 The Pentaprobe System

The pentaprobe system is made of six overlapping dsDNA plasmid molecules and covers every combination of five bases (Kwan, Czolij, Mackay, & Crossley, 2003). Each single strand of pentaprobe was cloned under a T7 promoter in pcDNA3 in Joel McKay's laboratory, School of Molecular Bioscience, University of Sydney. The plasmids containing pentaprobe strands in the forward direction are termed A922, A923, A924, A925, A926 and A927 and in the reverse direction A928, A929, A930, A931, A932 and A933. A922 is complementary to A928, A923 to A929 etc. RNA fragments covering every

Chapter Three

combination of five bases were transcribed from this set of 12 plasmids (sequences are presented in Appendices).

The set of 12 pentaprobe inserts were produced by Jo McKenzie (Proteins and Microbes Laboratory, University of Waikato) from plasmids obtained from Joel MckKay, School of Molecular Bioscience, University of Sydney. Pentaprobe inserts were then amplified using the Pfx® DNApolymerase system (Invitrogen) with the forward primer flanking the T7 promoter in pcDNA3 and reverse flanking the end of the pentaprobe insert, and gel purified.

3.2.4.2.1 *In vitro* Transcription of Pentaprobe Inserts

RNA was transcribed from Jo McKenzie's pentaprobe PCR products using the Ambion T7 MEGAscript® kit (Thermo Fisher Scientific). Reactions of 1 µg DNA template, 2 µl each rNTP (75 mM stock), 2 µl T7 polymerase enzyme mix, 2 µl 10 x reaction buffer and nuclease free water in a total volume of 20 µl were incubated at 37 °C for 4 hrs, then DNase treated with 1 µl TURBO™ DNase for 15 min at 37 °C, shaking at 600 rpm. Transcribed RNA was purified by ammonium acetate and ethanol precipitation. RNA was precipitated on ice for 1 hr or at -20 °C overnight, with 8 µl 5 M ammonium acetate and three times the volume (54 µl) of 100 % ethanol. Samples were centrifuged (13,000 g, 20 min, 4 °C) and supernatant was removed. The pellet was washed with 50 µl 70 % ethanol and centrifuged as above. The supernatant was removed and the pellet was air dried and then resuspended in 5 µl nuclease free water, and stored at -80 °C. Secondary structure of RNA molecules were predicted using RNAfold (Gruber, Lorenz, Bernhart, Neuböck, & Hofacker, 2008).

3.2.4.3 VapC Ribonuclease Activity Assay

Ribonuclease activity of VapC against pentaprobe RNA was measured over 1 hr with 3 different negative controls: 1) RNA only, to ensure there is no RNase contamination of the assay buffer; 2) with EDTA included in the buffer, to rule out the possibility of contaminating metal-independent RNase that co-purified with VapC, and 3) with the VapBC complex, to show lack of VapC ribonuclease activity when in complex with VapB.

Chapter Three

3.2.4.3.1 Assay Reactions

10 µl reactions containing 1 µg purified VapC protein, 1 µg RNA, 6 µl assay buffer (20 mM phosphate buffer pH 7.4, 20 mM NaCl, 10 mM MgCl₂) and nuclease free water were incubated at 37 °C for 1 hr. For the RNA only negative controls, assay reactions were set up as above but no VapC was added to the reaction. Two reactions were set up, 0 hr (which was stopped immediately) and 1 hr. For the EDTA negative control, an assay was set up as per the normal assays but using assay buffer containing 20 mM EDTA. For the VapBC control, the assay was set as per the normal assay but purified VapBC was added to the reaction instead of VapC. All assay reactions were stopped by the addition of 10 µl 2 x formamide loading dye.

3.2.5 *Synechocystis* VapBC_{slr1209/1210} DNA Electrophoretic Mobility Shift Assays (EMSA)

Purified VapBC was used in electrophoretic mobility shift assays (EMSA) using the DIG Gel Shift kit, 2nd Generation (Roche). The target DNA fragments were labelled as described in Section 2.2.1.2.

3.2.5.1 EMSA Reactions

The protein:DNA ratio was based on the molarity of the VapBC complex (tetrameric 115.24 g/mol) calculated by ProtParam (web.expasy.org/protparam/) and DNA, which was calculated using the Promega BioMath Calculator (<http://www.promega.com/a/apps/biomath/index.html?calc=ugpmols>). Assay Protein:DNA ratios began at equal molar and subsequent higher ratios of protein to DNA. Assays were conducted as per manufacturer's instructions with each reaction containing 1.5 x Binding Buffer, 0.5 µg of Poly dI-dC, 0.4 ng of DIG labelled DNA and various concentrations of protein. To confirm positive shifts were a result of protein binding to DNA and not to the DIG label, assays were set up with unlabelled target DNA added in 120 x excess at the highest labelled DNA: protein ratio. This assay was labelled specific control. To determine specificity of the protein binding assays were set up with unlabelled non-specific DNA (pET28bpstI_*vapBC*_{slr1209/1210} with no promoter region) added in 120 x excess. This assay was labelled the non-specific control.

Chapter Three

Assays were prepared on ice and incubated at room temperature for 20 min. Loading dye was added to each assay and then run on a 6 % native acrylamide TBE gel as per Section 2.2.1.4.2. DNA southern blotting and cross linking was done as per Section 2.2.5.2. Membrane washing and blocking was done as per Section 2.2.5.3. The chemiluminescence immunoassay and visualisation was done as per Section 2.2.5.4.

3.2.6 *Synechocystis* $\Delta vapBC$ Phenotypic Characterisation

3.2.6.1 *Synechocystis* WT vs PCC6803 $\Delta vapBC$ Unmarked Growth Experiments

Synechocystis WT and $\Delta vapBC$ glycerol stocks were streaked onto BG-11 agar plates and grown under low light (approximately $5\text{--}20\ \mu\text{mol photons m}^{-2}\ \text{s}^{-1}$) at $30\ ^\circ\text{C}$ for 4-5 days. Colonies from the plate were used to inoculate 3 x 200 ml BG-11 starter cultures per strain. Starter cultures were incubated under low to moderate light ($20\text{--}40\ \mu\text{mol photons m}^{-2}\ \text{s}^{-1}$) at $30\ ^\circ\text{C}$, shaking at 150 rpm for 5 days or until they reached OD_{750} 0.2-0.7. Multiple colonies were required for inoculation to prevent photo-inhibition of cells, which occurs when initial biomass is too low. Each starter culture was used to inoculate 1 x 100 ml growth culture at OD_{750} 0.1, for the full light condition, and 1 x 100 ml growth culture OD_{750} 0.1, for the 12 hrs light-dark cycling condition. Each growth condition was run simultaneously. Each culture was sampled every 12 hrs for 9 days with OD_{600} (contamination), OD_{680} (chlorophyll) and OD_{750} (cell density) recorded. Cultures were visualised under microscope to check for culture health, morphology and contamination.

3.2.6.2 Isolation of Total RNA from *Synechocystis*

Total RNA was isolated based on the Acidic Guanidinium Isothiocyanate/Phenol/Chloroform (AP-GITC) procedure (Chomczynski & Sacchi, 1987) with a few modifications, namely, 1-bromo-3-chloropropane (BCP) was used instead of chloroform and two successive phenol-BCP extractions were performed instead of one. A 50 ml sample of *Synechocystis* WT and *Synechocystis* $\Delta vapBC$ cultures were taken at 84 hrs (late-exponential phase for full light experiment and early exponential phase for light-dark experiment).

Chapter Three

Samples were restricted to this late time point as a result of low biomass at earlier stages of growth. Cells were centrifuged at 4,000 g for 15 min at 4 °C and the supernatant was removed. The pellet was resuspended in 5 ml of 5 M GITC, pH 7, and centrifuged as above. The pellet was then resuspended in 0.5 ml GITC, pH 7 and transferred to a sterile tube containing 0.3 g of 0.1 mm and 2.5 mm zirconia beads and stored at -80 °C until use. Cells were disrupted by bead beating four times for 20, 25, 30 and 35 s respectively with 1 min rests in between, then centrifuged to remove the resulting foam. 50 µl of 2 M, pH 4 Sodium acetate (NaOAc) and 0.5 ml water saturated phenol were added, vortexed for 15 s, then rotated for 10 min at room temperature. After addition of 100 µl BCP, the sample was shaken vigorously for 1 min, incubated for 5 min on ice and centrifuged (13,000 g, 15 min, 4 °C). The aqueous phase was re-extracted by repeating the two previous steps in a new tube, scaling up the volumes accordingly to account for any increase in aqueous phase volume. The RNA contained in the aqueous phase was transferred to a new tube and precipitated with an equal volume of isopropanol before being incubated at -40 °C overnight. The RNA was pelleted by centrifugation (13,000 g, 30 min, 4 °C), washed in 1 ml 70 % ethanol and centrifuged again (13,000 g, 15 min, 4 °C). Finally the RNA was washed in 1 ml 100 % ethanol, centrifuged (13,000 g, 15 min, 4 °C), briefly air-dried and then solubilised in 20 µl nuclease free water. Isolated RNA was stored at -80 °C until use.

3.2.6.3 RNA-seq Processing and Data Analysis

Total RNA purity and concentration was estimated by measuring A_{260}/A_{230} and A_{260}/A_{280} ratios using a NanoDrop® ND-1000 Spectrophotometer (Thermo Fisher Scientific). 23s and 16s band representation and further quality control tests were done at BGI, China (<http://www.genomics.cn/en/index>). DNA removal, cDNA library construction, RNAseq and bioinformatics analysis (Quantification) was also done at BGI. Briefly, the RNA-seq raw reads were mapped to the reference genome *Synechocystis* sp. *PCC6803* chromosomal DNA (RefSeq NC_020286.1) to summarise the distribution of reads between samples. RNA reads from the bacteria *E. coli* CFT073 and *Bacillus subtilis subsp natto* 195 were identified in each sample during the mapping procedure and removed for downstream analysis. Bioinformatic analysis involved pairwise comparisons

Chapter Three

between control (wild type strain (WT)) and treatment (Δ strain) in both full light and light-dark cycling (for WT vs $\Delta vapBC_{slr1209/1210}$) and light-dark cycling (for WT vs $\Delta vapBC_{ssl2138/sll1092}$). Differentially expressed genes were identified by determining the Log2 expression ratio ($\Delta strain/WT strain$) where differential expression was termed significant if the expression ratio was >2.0 Fold with a P-value <0.05 .

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

3.2.6.4.1 Genomic DNA Removal and cDNA Library Construction

DNA removal from total RNA was achieved using DNase (Promega, Madison WI) treatment. 1 μ g of total RNA was treated with 1 μ l of RNase-free DNase and incubated at 37 °C for 3 hrs, shaking at 600 rpm. DNase was inactivated by the addition of 1 μ l 25 mM EDTA (pH 8) and incubation at 65 °C for 10 min. The cDNA library was constructed from the DNase treated total RNA by reverse transcription using Bioline Tetro Reverse Transcriptase (RT) enzyme (Bioline, UK) with random hexamer primers. Total RNA was incubated with 100 ng primers, 200 x U RT, 0.1 μ M dNTPS and 1 X RT buffer. The following cycling conditions were used for cDNA amplification:

25 °C	10.00 (min:sec)
45 °C	30.00
85 °C	5.00

To confirm there was no contaminating genomic DNA, a negative control was set up for each sample as above, but with no reverse transcriptase. An end point PCR test on both samples was done to confirm no amplification from the negative control. cDNA was frozen at -20 °C until required.

3.2.6.4.2 Primer Design

Primers for each gene were designed using Primer 3 v.0.4.0 (<http://bioinfo.ut.ee/primer3-0.4.0/>). Where possible, primers were designed to have similar properties so that PCR for different genes could be performed in the same run. All primers were supplied by IDT and resuspended in 1 x TE. A list of all primers is presented in Table A.2 in the Appendices.

Chapter Three

3.2.6.4.3 Reaction Set Up

All RT-qPCR reactions were carried out in triplicate 20 µl volumes using the 5 x HOT FIREPol® EvaGreen® qPCR Supermix (DNature, NZ). A master mix was prepared for each primer set consisting of 1 x Supermix buffer (containing Optimized buffer, GC-enhancer, 2.5 mM MgCl₂, dNTP blend containing dUTP/dTTP, EvaGreen® dye, and HOT FIREPol® DNA Polymerase), 100 nM primers, 10 µl of diluted cDNA and DNase free water to bring the volume to 20 µl. cDNA was diluted 1:50 to make a final concentration of 1:5 for qPCR analysis.

3.2.6.4.4 qPCR Run

RT PCR was performed on the Rotor-Gene 6000 (Corbett Research, Australia). The following cycling conditions were used for amplification:

95 °C	12.00	(min:sec)
95 °C	0.15	x 35
60 °C	0.30	
72 °C	0.30	
80 °C	0.10	

Data was acquired on the 80 °C step in the green channel for SYBR Green. Immediately following amplification, a melt curve was performed consisting of 90 sec at 55 °C, followed by 5 sec steps with 0.5 °C increase in temperature per step rising to 95 °C. The identity of the PCR product was verified by the melt curve profile. Each RT-qPCR run included a no template control (NTC) per amplicon.

3.2.6.4.5 Analysis

Firstly, the geometric mean of the two housekeeping genes (*rrn 16s* and *rnpB*) was calculated ($Ct\ HKG = \sqrt{(rrn\ 16s\ Ct \times rnpB\ Ct)}$) for both the WT and the *ΔvapBC* sample qPCR using their average Ct values generated from the Rotor-Gene 6000 software. The genes of interest (GOI) were then normalised against the housekeeping genes by calculating the ΔCt in both the WT sample ($Ct\ GOI_{wt} - Ct\ HKG_{wt}$) and the Δ sample ($Ct\ GOI_{\Delta} - Ct\ HKG_{\Delta}$). The $\Delta\Delta Ct$ for each GOI was

Chapter Three

then calculated ($\Delta Ct_{\Delta} - \Delta Ct_{wt}$) and the fold difference ($2^{(-\Delta\Delta Ct)}$) where differential expression was termed significant if the expression ratio was >2.0 Fold.

3.3 Results

3.3.1 *M. aeruginosa* NIES-843 *vapBC* Loci Identification

M. aeruginosa NIES-843 (MIC_NIES_843) was the strain chosen for this investigation because it is the only fully assembled and annotated *M. aeruginosa* genome available. Further information regarding the choice of strains chosen for this research is in the Appendices. To identify all putative *vapBC* loci in *M. aeruginosa* NIES 843, the Toxin-Antitoxin Database (TADB) (Shao et al., 2011) was consulted. The TADB is a web-based tool to identify Type II TA loci in bacterial genomes based on sequence alignment and conserved domains. *M. aeruginosa* NIES-843 has one chromosome and no plasmids so all *vapBC* loci are chromosomal. 46 VapBC systems are predicted in the genome (Table 3.1).

Alignments from NCBI nucleotide BLAST for all predicted 46 VapBC systems show that I) 27 have high conservation with VapBC systems in the closely related organism *M. aeruginosa* PCC 7806 (Table 3.1) (Frangeul et al., 2008), II) 18 have very low or no conservation with VapBC systems in *M. aeruginosa* PCC 7806 and are homologous to other cyanobacteria, including *Cyanothece* and *Calothrix sp.*, *Nodularia spumigena*, *Anabaena cylindrica* PCC 7122, *Nostoc sp.* PCC 7524, *Synechococcus sp.* PCC 7502, *Oscillatoria acuminata* PCC 6304, *Anthrospira platensis*, *Synechocystis* PCC 6803 and *Cyanobacterium aponinum* PCC 10605, III) 1 VapBC system has highest conservation with a VapBC system in the proteobacteria *Burkholderia rhizoxinica*, and IV) none of the 46 are closely related to each other, ruling out gene duplication as a recent mechanism for their expansion in the genome.

Table 3.1. List of the 46 predicted VapBC pairs in *M. aeruginosa* (MIC_NIES_843) with NCBI nucleotide BLAST alignments. Highlighted pairs indicate VapBC systems chosen for *in vitro* characterization

TA Pair (VapB-VapC)	VapC/B (MAE)	NCBI BLAST Alignment (Nucleotide)
COG2442-PIN (16) ^a COG2442 proteins contain a DNA/RNA-binding 3-helical bundle and are among the most abundant PIN-associated predicted antitoxins. Present in variety of bacterial and few archaeal genomes (abundant in cyanobacteria/chloroflexi).	05270/05280	78% conservation with <i>Nodularia spumigena</i> (96% coverage)
	10430/10420	99% conservation with <i>M. aeruginosa</i> PCC 7806 (PCC 7806) (50% coverage)
	18190/18180	99% conservation to PCC 7806 (100% coverage)
	21350/21340	95% conservation to PCC 7806 (100% coverage)
	24740/24730	70% conservation with <i>Anabaena cylindrica</i> PCC 7122 (PCC7122) (43% coverage)
	30350/30360	79% conservation with <i>Nostoc sp.</i> PCC7524 (98% coverage)
	32540/32550	87% conservation with PCC 7806 (98% coverage)
	32560/32550	87% conservation with PCC 7806 (49% coverage)
	34480/34490	84% conservation with PCC 7806 (96% coverage)
	34790/34800	99% conservation to PCC 7806 (100% coverage)
	34850/34840	99% conservation with PCC 7806 (100% coverage)
	35570/35580	68% conservation with <i>Synechococcus sp.</i> PCC 7502 (88% coverage)
	42140/42130	98% conservation with PCC 7806 (100% coverage)
	45530/45540	84% conservation with <i>Cyanothece sp.</i> (98% coverage)
	47160/47150	99% conservation with PCC 7806 (100% coverage)
	49650/49640	99% conservation with PCC 7806 (100% coverage)
COG2886-PIN (6) ^a COG2886 are an uncharacterised family of proteins associated with PIN genes which are abundant in Archaea and cyanobacteria.	01050/01060	95% conservation with PCC 7806 (100% coverage)
	04250/04260	97% conservation with PCC 7806 (100% coverage)
	04270/04280	97% conservation with PCC7806 (60% coverage)
	04320/04330	97% conservation with PCC7806 (96% coverage)
	43220/43230	100% conservation with PCC 7806 (37% coverage)
	46970/46960	76% conservation with PCC 7113 (31% coverage)
Xre-PIN (6) ^a Xre proteins have HTH DNA binding domains and have diverse regulatory roles not always related to TA systems.	03580/03590	72% conservation with <i>Oscillatoria acuminata</i> PCC 6304 (94 % coverage)
	08700/08690	80% conservation with <i>Anthrospira platensis</i> (71% coverage)
	15740/15750	98% conservation with PCC 7806 (92% coverage)
	20360/20370	99 conservation with PCC 7806 (92% coverage)
	34530/34520	99% conservation with PCC 7806 (100% coverage)
	47340/47330	99% conservation with PCC 7806 PCC 7806 (100% coverage)

<p>PHD-PIN (5^a)</p> <p>PHD has a DNA binding domain and are associated with a range of Type II toxins</p>	06990/07000	75% conservation with Proteobacteria <i>Burkholderia rhizoxinica</i> (19% coverage)
	23820/23810	83% conservation with <i>Cyanothece sp.</i> ATCC 51142 (66% coverage)
	33180/33190	88% conservation with <i>Cyanothece sp.</i> PCC 7822 (99% coverage)
	56140/56150	97% conservation with PCC 7806 (67% coverage)
	58010/58020	81% conservation with <i>Calothrix sp.</i> PCC 7507 (53% coverage)
<p>RHH-PIN (4^a)</p> <p>Antitoxin contains ribbon-helix- helix (RHH) DNA binding domain.</p>	07680/07670	93% conservation with PCC 7806 (33% coverage)
	13550/13540	76% conservation with PCC 6803 (92% coverage)
	17540/17530	75% conservation with <i>Cyanobacterium aponinum</i> PCC 10605 (79% coverage)
	37970/37980	88% conservation with ATCC 51142 (76% coverage) 75% conservation with <i>Nostoc sp.</i> PCC 7524 (78% coverage)
<p>AbrB-PIN (2^a)</p> <p>AbrB proteins have a DNA binding domains and are common in thermophilic prokaryotes</p>	31220/31210	100% conservation with PCC 7806 (39% coverage) 89% conservation with <i>Oscillatoria acuminata</i> PCC 6304 (95% coverage)
	54730/54740	67% conservation with PCC 7524 (39% coverage)
<p>VapC* (2^a)</p>	18130/18120	99% conservation with PCC 7806 (100% coverage)
	62610/62620	98% conservation with PCC 7806 (99% coverage)
<p>PIN* (5^a)</p>	03490/03480	99% conservation with PCC 7806 (100% coverage)
	07520/07510	66% conservation with PCC 7502 (65% coverage)
	08780/08770	80% conservation with <i>Calothrix sp.</i> PCC 6303 (100% coverage)
	20360/20350	99% conservation with PCC 7806 (85% coverage)
	62210/62200	79% conservation with <i>Cyanothece sp.</i> PCC 7424 (91% coverage)

^a Number of loci in genome

* TADB prediction of toxin only

3.3.2 Identification and Cloning of Representative *M. aeruginosa* VapBC Systems for Expression in *E. coli*

Two of the 46 *vapBC* operons from *M. aeruginosa* NIES-843 were chosen as representative systems and codon optimised for *E. coli* expression. The two operons *vapBC*_{MAE49640/50} and *vapBC*_{MAE43230/20} (Table 3.2) were chosen based on a selection criteria that included I) VapC was >112 aa and included all conserved acidic residues for catalytic activity (Figure 3.1), II) there must be at least 1 bp overlap between *vapB* and *vapC*, and III) protein sequence comparisons shows either a) the protein shares high similarity to other bloom forming strains or is a ubiquitously conserved VapBC system, or b) The system appears unique to allow investigation of potentially novel VapBC functions in cyanobacteria. The operons were designed for insertion into the pET28b-pstI vector for *E. coli* expression with a C-terminal His tag. To increase the chances of expression, the overlapping region between *vapB* and *vapC* were reduced to 1-4 bp.

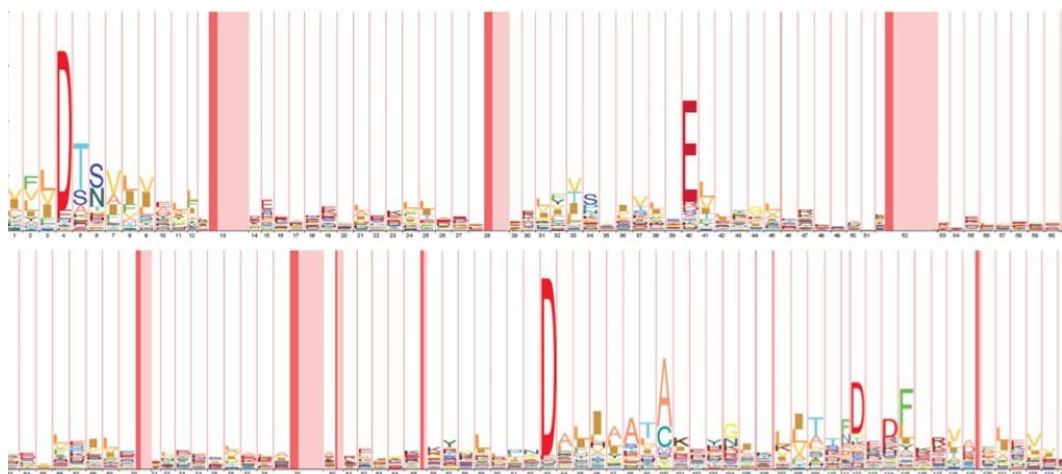


Figure 3.1. A visual depiction of the hidden Markov model that defines the conserved amino acids of PIN-domain proteins. The height of each letter is proportional to the amount of ‘information’ it provides about the respective position in the PIN-domain family. The width of each column is also an indication of the importance of this position in defining the family. Sequence regions that contain insertions are shown as dark and light pink columns. Figure from Arcus et al. (2011).

Table 3.2. *M. aeruginosa* representative *vapBC* operons *vapBC*_{MAE43230/20} and *vapBC*_{MAE49640/50} for cloning, expression and purification in *E. coli*.

VapBC Description: VapBC _{MAE43230/20}		Reason for Selection
Locus ID	4985	Novel VapB: COG2886 is an uncharacterised family of small proteins associated with PIN genes. Found mostly in archaea and cyanobacteria
Size	730 bp with 8 bp overlap	
VapC	VapC _{MAE43220} : PIN domain with conserved residues D7, E39, D92 and D99	All COG2886-PIN TA pairs share ≥95% sequence similarity to PCC 7806 (with exception of TA ID 4992) (Makarova et al., 2009)
VapB	VapB _{MAE43230} : COG2886 domain contains a HTH DNA binding motif	There are only 6 copies of COG2886-PIN TA pairs in the MIC NIES-843 genome
Nucleotide Sequence Analysis	100% sequences conservation to PCC 7806 over conserved regions and poor conservation in non conserved regions	VapC has all conserved acidic residues. The last D residue is at 99 instead of 112 High sequence conservation similarity with PCC 7806
VapBC Description: VapBC _{MAE49640/50}		Reason for Selection
Locus ID	4999	Ubiquitous VapB: COG2442 is a family among the most abundant new PIN-associated predicted antitoxins. Present in a variety of bacterial and a few archaeal genomes but are most abundant in cyanobacteria and chloroflexi (Makarova et al. 2009).
Size	605 bp with 4 bp overlap	
VapC	VapC _{MAE49650} : PIN domain with conserved residues D9, E42, D94, D113	VapC has all conserved acidic residues. The last D residue is at 108 instead of 112.
VapB	VapB _{MAE49640} : COG2442 domain contains a DNA/RNA-binding 3-helical bundle	The most abundant TA pair in the MIC NIES-843 genome is COG2442-PIN with 16 identified.
Nucleotide Sequence Analysis	99% sequences similarity to PCC 7806 with 100% coverage over entire operon.	High sequence conservation with PCC 7806.

Attempts to clone each of the *M. aeruginosa* *vapBC*, *vapB* or *vapC* into the expression vectors pET28b-pstI, pETDUET, pPROEX or the pOPIN vector suite were conducted. Each *vapBC* operon or *vapB* and *vapC* gene that was successfully cloned into its respective vector for protein expression was confirmed by sequencing vector DNA from positive transformants with T7 or M13 forward and reverse primers that flanked the multiple cloning site (MCS). Successfully cloned operons and genes were tested for protein expression in *E. coli* with small scale expression trials.

3.3.2.1 Small Scale Expression Trials

Small scale expression trials of VapBC and VapC were carried out to evaluate protein expression and Ni²⁺ binding affinity. A summary of the expression trials (Table 3.3) show that most expression vectors were unsuccessful, with only one operon being successfully expressed (*vapBC*_{MAE43230/20}) in 1 of the 5 vectors tested (pOPINS). To determine if the overlap between *vapB* and *vapC* was a translational hindrance for *E. coli* and the reason for unsuccessful expression, the *vapC* genes from each operon were extracted from the GeneArt_*vapBC* plasmids and inserted into pET28b-pstI for expression. Expression trials show there was no soluble overexpression of either VapC (Figure 3.3). Attempts to improve expression by repeating at temperatures from 18-37 °C were not successful.

Table 3.3. Summary of *M. aeruginosa* VapBC and VapC small scale expression trials and His tag binding.

Vector	His tag	Gene	Protein	Expression?
pET28b-pstI	C	<i>vapBC</i> _{MAE49640/50}	VapBC _{MAE49640/50}	N
		<i>vapBC</i> _{MAE43230/20}	VapBC _{MAE43230/20}	N
		<i>vapC</i> _{MAE49650}	VapC _{MAE49650}	N
		<i>vapC</i> _{MAE43220}	VapC _{MAE43220}	N
pPROEX	N	<i>vapBC</i> _{MAE49640/50}	VapBC _{MAE49640/50}	N
pOPINF	N	<i>vapBC</i> _{MAE43230/20}	VapBC _{MAE43230/20}	N
pOPINM	N	<i>vapBC</i> _{MAE43230/20}	VapBC _{MAE43230/20}	N
pOPINS	N	<i>vapBC</i> _{MAE43230/20}	VapBC _{MAE43230/20}	S

Abbreviations: no protein expression (N), and soluble protein expression (S).

To allow co-expression of VapB and VapC without the overlap, both genes were extracted from the GeneArt_*vapBC* plasmid and attempted to be inserted into two different cloning sites in the multiple cloning site (MCS) of the pETDUET vector. Transformation of the *vapB* into the vector for both operons proved unsuccessful and expression in this system could not be trialled. Alternatively, an attempt was made to clone the *vapBC* operons into the expression vector pPROEX, an alternative to pET28b-pstI, with an N-terminal His tag. VapBC_{MAE43230/20} cloning was unsuccessful but VapBC_{MAE49640/50} was

Chapter Three

successful and expression was tested with the empty plasmid used as the negative control. No expression was observed (Figure A.3 in the Appendices).

Because the standard expression systems mentioned above proved unsuccessful, a high throughput pOPIN vector suite cloning system ((Berrow et al., 2007) commonly used for problematic protein expression was adopted. The *vapBC* operons were extracted from the GeneArt plasmids and had homologous extensions incorporated to allow insertion into pOPIN plasmids with various fusion proteins and tags. Table A.3 in the Appendices lists all plasmids used for expression trials. Cloning of *vapBC*_{MAE49640/50} was unsuccessful, however cloning of *vapBC*_{MAE43230/20} was successful and expression trials were conducted.

From the expression trials only the pOPINS vector showed promising VapBC expression (Figure 3.2). pOPINS contains an N-terminal His tag followed by a SUMO fusion protein. SUMO proteins are commonly used as fusion proteins that bind to and increase solubility and expression of otherwise toxic or insoluble proteins (Marblestone et al., 2006). The SUMO protein with an N-terminal His tag (His-SUMO) is 13.21 kDa, which can be seen in the SDS-PAGE gel of the empty vector expression trial (Figure 3.2). SUMO-His bound to VapB (His-SUMO-VapB) is 22.17 kDa, which can be seen on the gel of the pOPINS_43230/20 expression trial. Both proteins appear larger on the gels, which is discussed later. VapC is 17.57 kDa, which is not obviously expressed in this screen. To confirm if both VapB and VapC were being expressed, this system was taken into the large scale expression and purification system.

Chapter Three

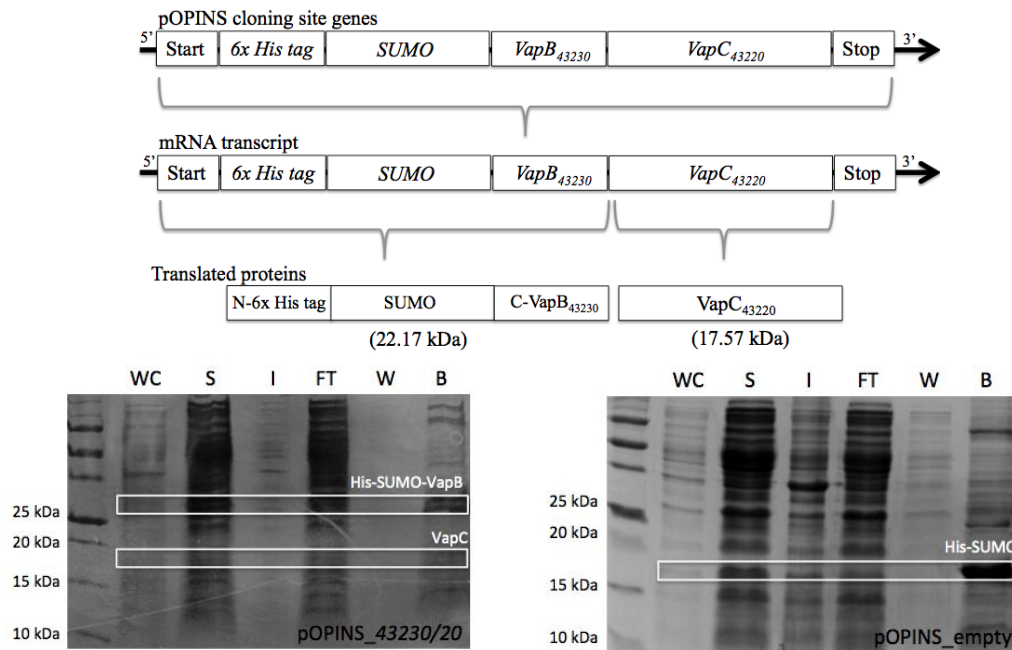


Figure 3.2. pOPINS+*vapBC*_{43230/20} small scale expression trial. Schematic of the pOPINS system indicating the N-terminal His tag-SUMO-VapB fusion protein and VapC expression. SDS-PAGE analysis of small scale expression trial with white boxes indicating expected band size. Labels: whole cell fraction (WC), soluble fraction (S), insoluble fraction (I), flow through (FT), wash (W) and nickel beads with protein bound after wash steps (B).

3.3.2.2 Large Scale Expression and Protein Purification of VapBC_{MAE43230/20}

Following evidence of soluble protein expression of VapBC_{MAE43230/20} in small scale screens (Figure 3.2), protein expression cultures were scaled up as in Section 3.2.2.3.1 and protein was purified by IMAC (Figure 3.3). Non-specific protein eluted early from the column at a low concentration of imidazole, whereas His tag associated protein eluted later in the gradient with a greater concentration of imidazole. His-SUMO protein elutes earlier in the gradient (~100-150 mM imidazole), His-SUMO-VapB protein elutes slightly later (~150 mM imidazole) and His-SUMO-VapBC protein elutes as the last peak (~200-325 mM imidazole) (Figure 3.3).

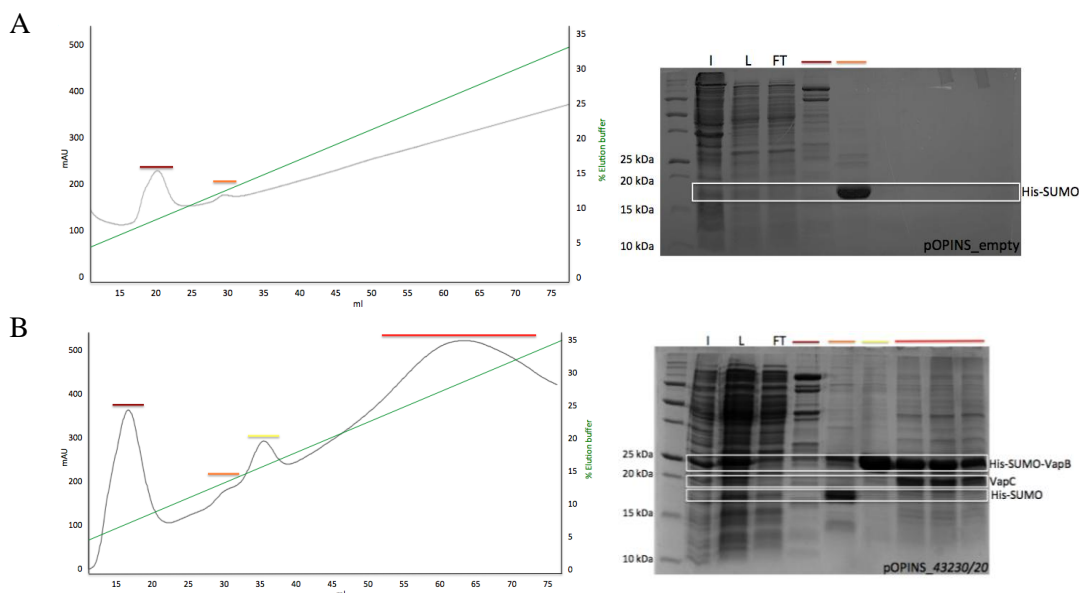


Figure 3.3. IMAC purification and corresponding SDS-PAGE analysis of His-SUMO-VapBC_{MAE43230/20}. The chromatograms depict the UV absorbance and elution profile of each purification. The corresponding SDS-PAGE gels depict the insoluble fraction (I), soluble fraction loaded onto the column (L) and the flow through (FT). Fractions and their elution position are shown by the coloured bars (A) IMAC purification of the empty pOPINS vector showing elution peak of His-SUMO at ~100-150 mM imidazole concentration. (B) IMAC purification of pOPIN_43230/20 showing elution of His-SUMO as in (A), His-SUMO-VapB_{MAE43230} at ~150 mM imidazole, and His-SUMO-VapBC_{MAE43230/20} later in the gradient as the largest peak at ~200-325 mM imidazole concentration.

The protein was further purified by size exclusion chromatography (SEC) using a Superdex column (GE Healthcare) (Figure 3.4). The majority of VapBC elutes in the first peak which corresponds to proteins ≥ 669 kDa or soluble aggregated protein. A small proportion of the complex elutes off the column in the second peak, which corresponds to approximately 440 kDa. This is much larger than the expected tetramer (159 kDa); however, the corresponding gel fraction shows no sign of aggregation, suggesting that the protein is stable in this fraction. The last peak in the trace contains the free VapB, shown by the gel fraction, which elutes off the column at approximately 67 kDa. This is also much larger than the expected dimer of VapB (44 kDa). The position of the peak suggests a tetrameric state of VapB (89 kDa). The unexpected increased size of the proteins observed in both small scale and large scale protein expression and purification could be explained by co-purification with genomic DNA. VapB binds DNA and the tetrameric state of VapB or VapBC is only possible in the presence of DNA, suggesting that VapBC_{MAE43230/20} is being co-purified with DNA. The chromatogram also shows significantly more His-SUMO-VapB expressed compared to His-SUMO-VapBC. VapBC protein co-expression is thought to

Chapter Three

occur in a ratio of up to 10:1 VapB:VapC. This is a result of the overlap present within the operon. The translational machinery has to slip at the end of the VapB transcript and re-start translation at the overlapping start site of the VapC transcript. This slippage occurs at a lower rate than the standard initiation of transcription for VapB, resulting in more VapB being translated than VapC. This result indicates that the *vapBC*_{MAE43230/20} is being transcribed as a normal *vapBC* operon.

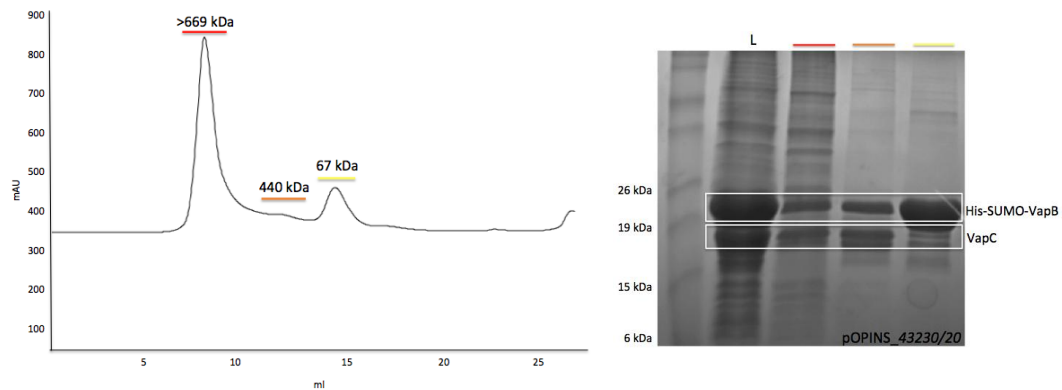


Figure 3.4. SEC of VapBC_{MAE43230/20} and corresponding SDS-PAGE gel. The chromatogram depicts the UV absorbance profile of protein elution and the gel depicts the fractions and their elution position (and approximate size) is shown by the coloured bars.

In summary, expression of VapBC_{MAE49640/50} was unsuccessful and expression of VapBC_{MAE43230/20} was successful only when expressed with a SUMO fusion protein attached to VapB_{MAE43230}. VapB_{MAE43230} and VapC_{MAE43220} are co-expressed with an excess of VapB being expressed relative to VapC. This expression ratio indicates that *vapBC*_{MAE43230/20} is expressed as a typical *vapBC* operon, where expression is characterised by a ratio of up to 10:1 VapB:VapC. VapB_{MAE43230} and VapC_{MAE43220} may form a tetrameric VapBC_{MAE43230/20} protein complex, which may be bound to DNA and unstable, with the majority of the complex forming a large species (possible soluble aggregate). VapB_{MAE43230} also appears to be purified bound to genomic DNA and forms a tetramer. Further optimisation is required to reduce the aggregation of the complex and removal of co-purified DNA before *in vitro* characterisation can be conducted.

3.3.2.3 Cloning of *Synechocystis* *vapBC* into pET28b-pstI

Synechocystis VapBC systems where deletion strains had been made for characterisation by growth experiments (Section 3.2.1.6) were used for cloning

Chapter Three

and expression trials. *vapBC* operons successfully cloned into pET28b-pstI with a C-terminal His tag for protein expression (Table 3.4) were confirmed by sequencing and tested for protein expression using small scale expression trials in *E. coli*.

Table 3.4. *Synechocystis vapBC* operons successfully cloned into pET28b-pstI for protein expression with a C-terminal His tag and predicted protein molecular weight (MW).

<i>vapBC</i> Operon	Expected MW (kDa)		
	VapB	VapC	VapBC
<i>vapBC_{slr1209/1210}</i>	12.37	16.46	28.83
<i>vapBC_{slr1327/ssr2201}</i>	8.40	17.98	26.38
<i>vapBC_{ssl2138/sll1092}</i>	7.70	16.33	24.03
<i>vapBC_{ssr2962/slr1767}</i>	8.23	13.98	22.21

3.3.2.4 Small Scale Expression Trials

Small scale expression trials were carried out to evaluate protein expression and His tag binding affinity (Table 3.5). Only one of the four VapBC systems (*vapBC_{slr1209/1210}*) tested shows soluble expression of both VapB and VapC (Table 3.5).

Table 3.5. Summary of *Synechocystis* VapBC small scale expression tests and His tag binding.

Operon	Protein	Expression?	
		VapB	VapC
<i>vapBC_{slr1209/1210}</i>	VapBC _{slr1209/1210}	S	S
<i>vapBC_{slr1327/ssr2201}</i>	VapBC _{slr1327/ssr2201}	N	N
<i>vapBC_{ssl2138/sll1092}</i>	VapBC _{ssl2138/sll1092}	N	I
<i>vapBC_{ssr2962/slr1767}</i>	VapBC _{ssr2962/slr1767}	S	N

Abbreviations: no protein expression (N), Insoluble protein expression (I) and Soluble protein expression (S).

All VapBC systems expressed poorly in the pET28b-pstI expression system, with the exception of VapBC_{slr1209/1210} (Figure 3.5). VapBC_{slr1327/ssr2201} shows no protein expression (Figure A.4, Appendices). VapBC_{ssl2138/sll1092} expressed only VapC, which was insoluble (Figure A.4, Appendices).

Chapter Three

VapBC_{ssr2962/slr1767} expressed only VapB, with the majority of the protein in the insoluble fraction (Figure A.4, Appendices). VapBC_{slr1209/1210} expressed very well and showed high affinity for the nickel beads (Figure 3.5), and was taken into the large scale expression and purification system.

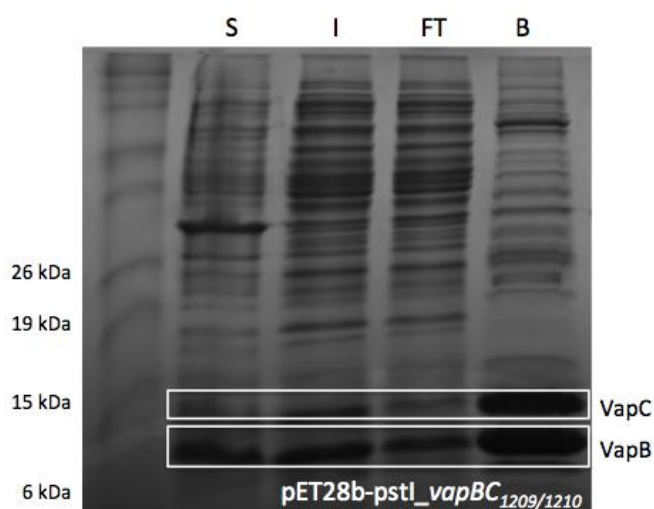


Figure 3.5. SDS-PAGE analysis of *E. coli* pET28b-pstI_VapBC_{slr1209/1210} small scale expression trial. White boxes indicate expected band sizes. Labels: soluble fraction (S), insoluble fraction (I), flow through (FT) and nickel beads with protein bound after wash steps (B).

3.3.3 Large Scale Expression and Protein Purification of VapBC_{slr1209/1210}

Following evidence of soluble protein expression of VapBC_{slr1209/1210} in small scale screens, expression cultures were scaled up (Section 3.2.2.3.1); however, scaling the cultures up at 37 °C resulted in insolubility of VapBC_{slr1209/1210}. The large scale protein expression cultures were subsequently incubated at 30 °C post induction, which increased the solubility of the VapBC complex. Protein was purified by IMAC and the chromatograms and corresponding SDS-PAGE gels (Figure 3.6) depict the column load, flow through and purification fractions. Non-specific bound protein eluted early off the column at a low concentration of imidazole (~90-120 mM), whereas His tag associated protein eluted later in the gradient with a greater concentration of imidazole (~130-220 mM).

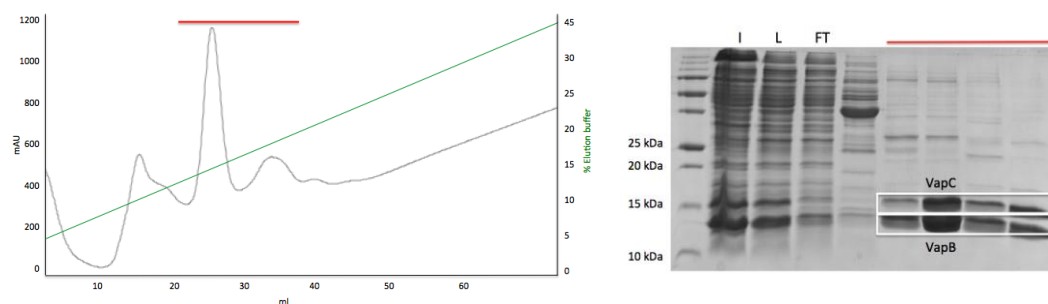


Figure 3.6. IMAC purification and corresponding SDS-PAGE analysis of VapBC_{slr1209/1210}. The chromatogram depicts the UV absorbance and elution profile. The corresponding SDS-PAGE gels depict the insoluble fraction (I), soluble fraction loaded onto the column (L) and the flow through (FT). Fractions and their elution position are shown by the red bar.

The protein was further purified by SEC (Figure 3.7). The SEC chromatogram shows that VapBC_{slr1209/1210} is being purified as a complex. VapBC_{slr1209/1210} elutes as a single peak at approximately 67 kDa based on elution volume. The expected size of VapBC_{slr1209/1210} is either a dimer of 57.66 kDa, or tetramer of 115.32 kDa when bound to DNA. The elution profile suggests that the protein is not bound to DNA and is in a dimeric state.

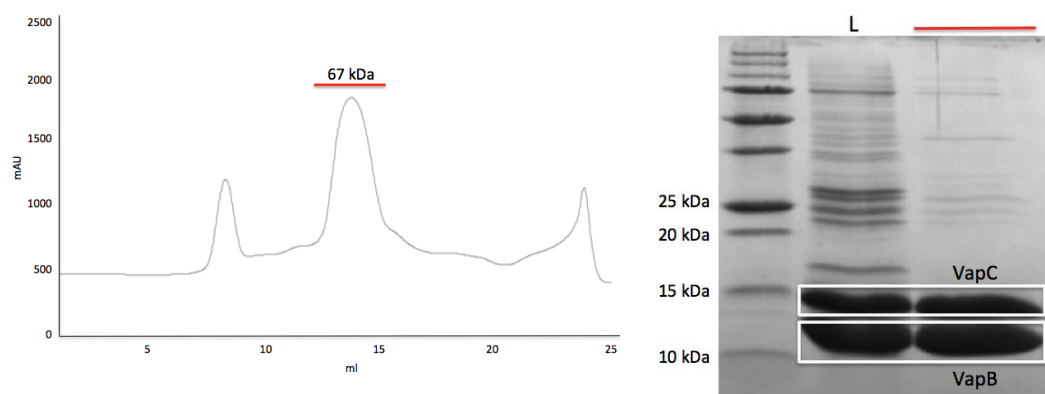


Figure 3.7. SEC of VapBC_{slr1209/1210} and corresponding SDS-PAGE gel. The chromatogram depicts the UV absorbance profile of protein elution and the gel depicts the fractions. Their elution positions are shown by the red bar.

SEC purified protein was used for crystallisation trials and electrophoretic mobility DNA shift assays (EMSA), or further treated for disruption of the complex to analyse VapB_{slr1209} or VapC_{slr1210} individually.

3.3.4 Preliminary VapBC_{slr1209/1210} DNA Electrophoretic Mobility Shift Assays (EMSA)

EMSAs were conducted with purified VapBC_{slr1209/1210} to investigate if it is capable of autoregulation. As an initial attempt to identify the VapBC_{slr1209/1210}

Chapter Three

recognition sequence, a 150 bp DNA fragment ($P_{slr1209A}$) of the *vapB*_{slr1209} promoter region was designed and tested by EMSA, and no shift was observed (Figure 3.8). The promoter region was subsequently investigated for palindromic sequences indicative of VapB recognition motifs and two overlapping palindromic sequences were identified (Figure 3.8). One palindromic sequence occurs between the bases 39 to 150 bp upstream of the start codon with each inverted repeat (IR) containing 10 bp and a 92 bp spacing region (Long IR1&2) (Figure 3.8). The other palindromic sequence occurs 147 to 168 upstream of the start codon, with each IR containing 4 bp and a 14 bp spacing region (Short IR1&2) (Figure 3.8). A new DNA fragment was designed to include these regions by extending $P_{slr1209A}$ by 50 bp at the 5' end ($P_{slr1209B}$) (Figure 3.8) and the EMSA was repeated (Figure 3.8). A shift in mobility was observed at 3.0-4.5 pM VapBC when incubated with $P_{slr1209B}$ (DNA:protein molar ratio 1:1000-1500) (Figure 3.8). To determine the specificity of binding unlabelled specific DNA and non-specific DNA, pET28bpstI_*vapBC*_{slr1209/1210} was added in 120 x excess, which did not affect the shift (Figure 3.8). These results suggest that VapBC_{slr1209/1210} does bind to the promoter region of its own operon, however the assay conditions need further optimisation to confirm specificity.

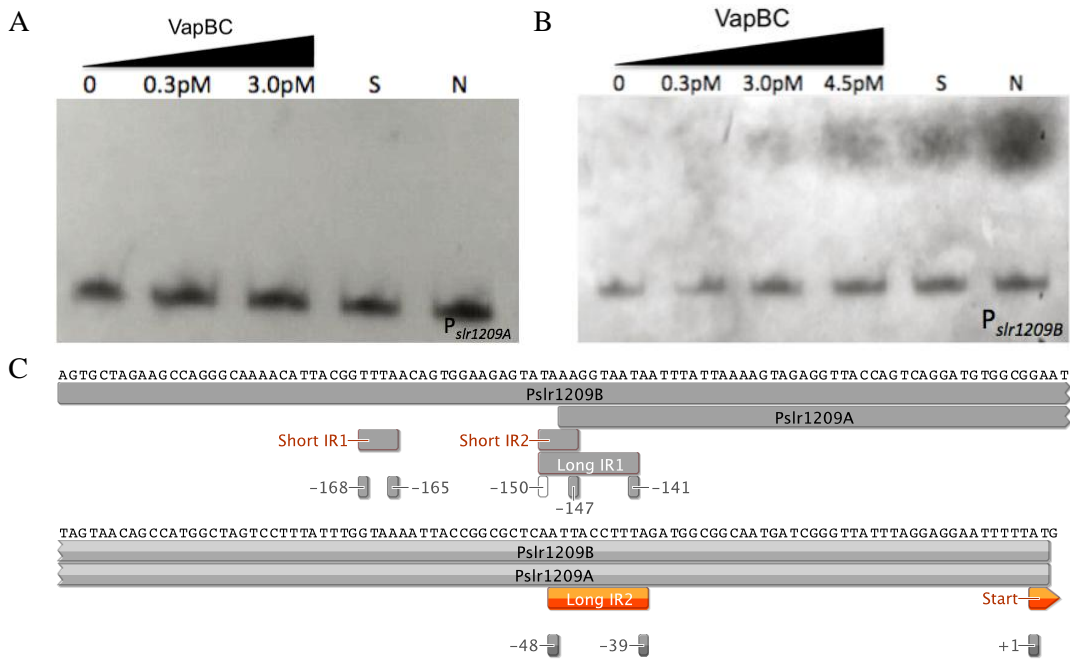


Figure 3.8. Preliminary VapBC_{slr1209/1210} autoregulation analysis by EMSA. Labels: Negative control with 3.0 fM DIG labelled DNA and no protein (0), 3.0 fM DIG labelled DNA and 0.3 pM protein to give a 1:100 DNA:protein ratio (0.3pM), 3.0 fM DIG labelled DNA and 3.0 pM protein to give a 1:1000 DNA:protein ratio (3.0pM), 3.0 fM DIG labelled DNA and 4.5 pM protein to give a 1:1500 DNA:protein ratio (4.5pM), 3.0 fM DIG labelled DNA and 4.5 pM protein to give a 1:1500 DNA:protein ratio (4.5pM) + 120 X unlabelled specific DNA (S), 3.0 fM DIG labelled DNA and 4.5 pM protein to give a 1:1500 DNA:protein ratio (4.5pM) + 120 X unlabelled non-specific DNA (pET28b-pstI_vapBC_{slr1209/1210}) (N). (A) VapBC_{slr1209/1210} EMSA with P_{slr1209A}. (B) VapBC_{slr1209/1210} EMSA with P_{slr1209B}. (C) Schematic of P_{slr1209} containing the potential palindromic binding recognition sequences at -39 to -150 bp (Long IR1&2) and -147 to -168 bp (Short IR1&2) upstream of the operon start codon.

In summary, these preliminary results indicate the VapBC_{slr1209/1210} does bind to the promoter region of its own operon, suggesting that it is autoregulatory. Binding is observed only when the full palindromic sequences are present from both putative recognition motifs. This suggests that these sequences are important for protein:DNA interaction, in particular the region containing the overlapping IR half sites at 141 to 150 upstream of the start site. Further investigation into these sequences is important to determine the recognition motif of VapBC_{slr1209/1210}, however, given the spacing region between the two putative motifs, the short palindromic sequence is a likely candidate. It is important to note these results are preliminary and optimisation of the assay conditions as well as further investigation into alternative regulatory mechanisms are important to elucidate the mechanism of regulation in this system.

3.3.5 Preliminary VapC_{slr1210} Ribonuclease (RNase) Activity

VapC proteins, in general, have been tested for nucleotidase activity on double stranded (ds) DNA (dsDNA), single stranded (ss) DNA (ssDNA), dsRNA and ssRNA, and show exclusive activity on ssRNA (Cook et al., 2013). To test VapC_{slr1210} RNase activity, the VapBC_{slr1209/1210} complex was disrupted by acid SEC (Section 3.2.2.4.1) to release VapC_{slr1210} (Figure 3.9). The SEC trace and SDS-PAGE analysis show that acidification (pH 3) of the purification buffer is successful in liberating VapC_{slr1210} from VapB_{slr1209}. The fractions that were enriched in VapC were pooled and concentrated for screening of RNase activity using the pentaprobe system (Section 3.2.4.2).

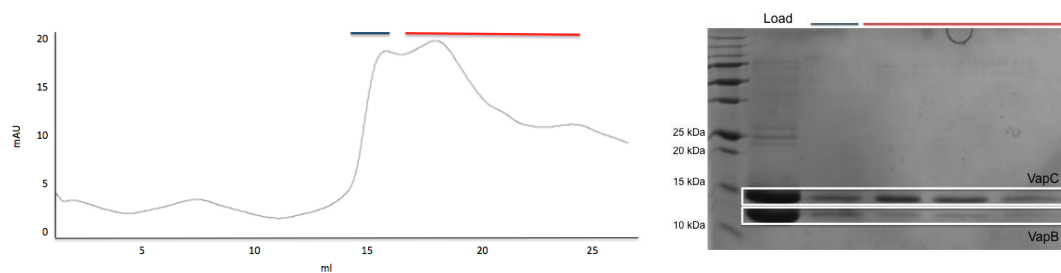


Figure 3.9. Acid SEC of VapBC_{slr1209/1210} and corresponding SDS-PAGE gel. The chromatogram depicts the UV absorbance profile of protein elution and the SDS-PAGE analysis shows the protein loaded onto the column and the subsequent purification fractions. Labels: VapBC_{slr1209/1210} in purification buffer at pH 3 (Load), blue line represented VapBC_{slr1209/1210} still in complex that elutes off the column as the first peak. The red line represents free VapC_{slr1210} protein elution.

3.3.5.1 *In vitro* RNase Pentaprobe Assays

To test purified VapC_{slr1210} for general ssRNA RNase activity, pentaprobe screens were conducted. Pentaprobe RNase activity screens, collectively, contain every possible combination of five bases, and have been previously described (Section 3.2.4.2) (McKenzie et al., 2012a). All six pentaprobe RNAs (from the forward direction) were tested as substrates for VapC_{slr1210} RNase activity (Figure 3.10). Comparison of assays after one hour in the presence and absence of VapC_{slr1210} indicated that there was slight degradation of pentaprobe 927 exclusively. To confirm RNase activity of VapC_{slr1210} on pentaprobe 927 the assay was repeated with the following controls: assay conducted with no VapC to check for any contaminating RNases in the assay buffers (RNA), VapBC_{slr1209/1210} was added instead of VapC_{slr1210} to see if the presence of VapB_{slr1209} inhibited VapC_{slr1210} activity (VapBC), and EDTA was added to the assay buffer

Chapter Three

to test for metal dependant endonuclease activity and to check for any metal independent RNase contamination (EDTA). All assays were incubated for 1 hr (Figure 3.10). VapC_{slr1210} displayed metal dependant RNase activity against the 927 pentaprobe RNA, which was abolished in the presence of EDTA. The banding pattern of the degraded 927 pentaprobe show two prominent bands between 50-80 bp, suggesting that VapC_{slr1210} selectively cleaves in the centre of the RNA molecule in an endonuclease mode, to produce two RNA fragments ~70 bp and 80 bp (Figure 3.10). VapBC_{slr1209/1210} showed no RNase activity, indicating that VapB_{slr1209} inhibits VapC_{slr1210} activity. No RNA degradation was observed in the RNA or EDTA controls indicating that there was no RNase contamination. Sequence specificity is shown by the different banding patterns produced by the cleaved fragments compared to general smearing that would be observed with non-specific RNase activity (Figure 3.10).

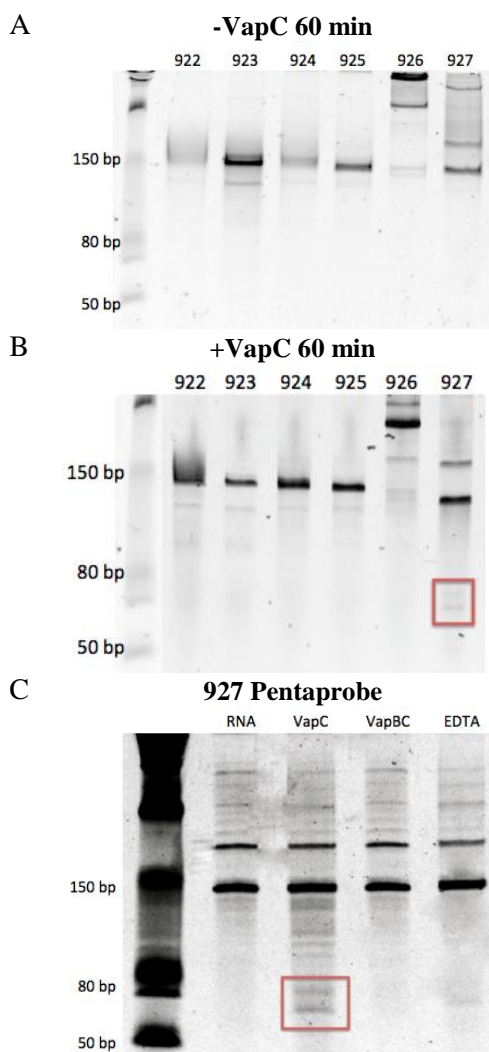


Figure 3.10. Preliminary time course assays showing ribonuclease activity of VapC_{slr1210} using the pentaprobe system. Each pentaprobe assay is labelled at the top of each lane. Numbers down the side of the urea-denaturing gels denote the low range ssRNA ladder base sizes. The multiple bands observed are due to the secondary structures in the RNA molecules. The red boxes indicate prominent degradation fragments (A) Each pentaprobe assay after 1 hr incubation in the absence of VapC. (B) Each pentaprobe assay after 1 hr in the presence of VapC. (C) The 927 pentaprobe assay with appropriate controls. Labels: assay run in the absence of VapC (RNA), Assay run in the presence of VapC (VapC), assay run in the presence of VapBC (VapBC), assay run in the presence of VapC with EDTA added to assay buffer (EDTA).

In previous studies using the pentaprobe system, binding of specifically degraded pentaprobes was observed by an EMSA-type shift, indicating a dual

Chapter Three

recognition function of the VapC by binding and cleaving its RNA substrate. The binding was abolished in the presence of VapB or EDTA (McKenzie et al., 2012a; McKenzie et al., 2012b). VapC_{slr1210} does not appear to bind to the 927 pentaprobe, suggesting it does not have the dual function.

To determine sequence specificity of VapC_{slr1210} future work will include oligomers of the 927 pentaprobe around the putative central cut site, which then will be used as substrates for VapC_{slr1210}. Cleaved oligonucleotides will then be subjected to MALDI-TOF MS (McKenzie et al., 2012a) to identify the cleaved fragments and target recognition sites.

3.3.6 *Synechocystis* Δ vapBC Deletion Strain Construction and Characterisation

3.3.6.1 *Synechocystis* Δ vapBC Deletion Strain Generation

In order to investigate the role of cyanobacterial VapBC systems, *in vivo* attempts were made to generate markerless Δ vapBC deletion strains of each of the 12 chromosomal *Synechocystis* VapBC systems. Of the 12 VapBC systems present in *Synechocystis*, a total of five markerless *Synechocystis* Δ vapBC deletion strains were successfully constructed (Figure 3.11). Generation of other deletion strains were unsuccessful due to mutations arising in coding regions or incomplete segregation of all 12 chromosomes to ensure mutation is present in every chromosome.

The generation of each markerless deletion strain required, firstly, the generation of a marked deletion strain, containing a cassette with selection markers (kanamycin resistance and sucrose sensitivity) to identify successful homologous recombination. Briefly, the marked deletion strain was generated by transforming the plasmid pUC19 + L&R + sacRB/Kan(R) construct (Section 3.2.1.5) into the wild type (WT). Kanamycin was introduced as the selection for homologous recombination and cells were passaged for 14 days. Successful deletions were confirmed by PCR diagnostics (Section 3.2.1.6.3) using the forward left fragment primer and the reverse right fragment primer for each operon. Transformants produced a band >5 kb containing the flanking regions (2-2.5 kb) of the operon and the 3.7 kb cassette that replaced the operon, which is too

Chapter Three

large to be visualised by gel electrophoresis. The absence of a band represented a successful marked deletion. The WT band was approximately 2.5 kb (operon + flanking regions). Successful deletion colonies were streaked onto new kanamycin agar plate to ensure all chromosomes had the deletion present.

Colonies from the marked deletion strain agar plate were then transformed with plasmid pUC19 + L&R construct (Section 3.2.1.5) and incubated for 4 days. Post- transformation cells were plated on agar plates with 5% (w/v) sucrose and incubated for 7 days or until colonies appeared. Colonies were then patched onto both agar plates + kanamycin and agar plates + sucrose to detect unmarked deletion strains. Colonies that grew only on sucrose agar plates were streaked onto sucrose agar plates again to ensure all chromosomes had the unmarked deletion present. Transformants were screened by colony PCR (Section 3.2.1.6.3) to confirm deletion and cassette removal using primers designed to amplify only the *vapBC* operon. Successful deletion strains showed a region of the operon removed, which was represented by a smaller band on the gel compared to the WT PCR product (Figure 3.11).

Chapter Three

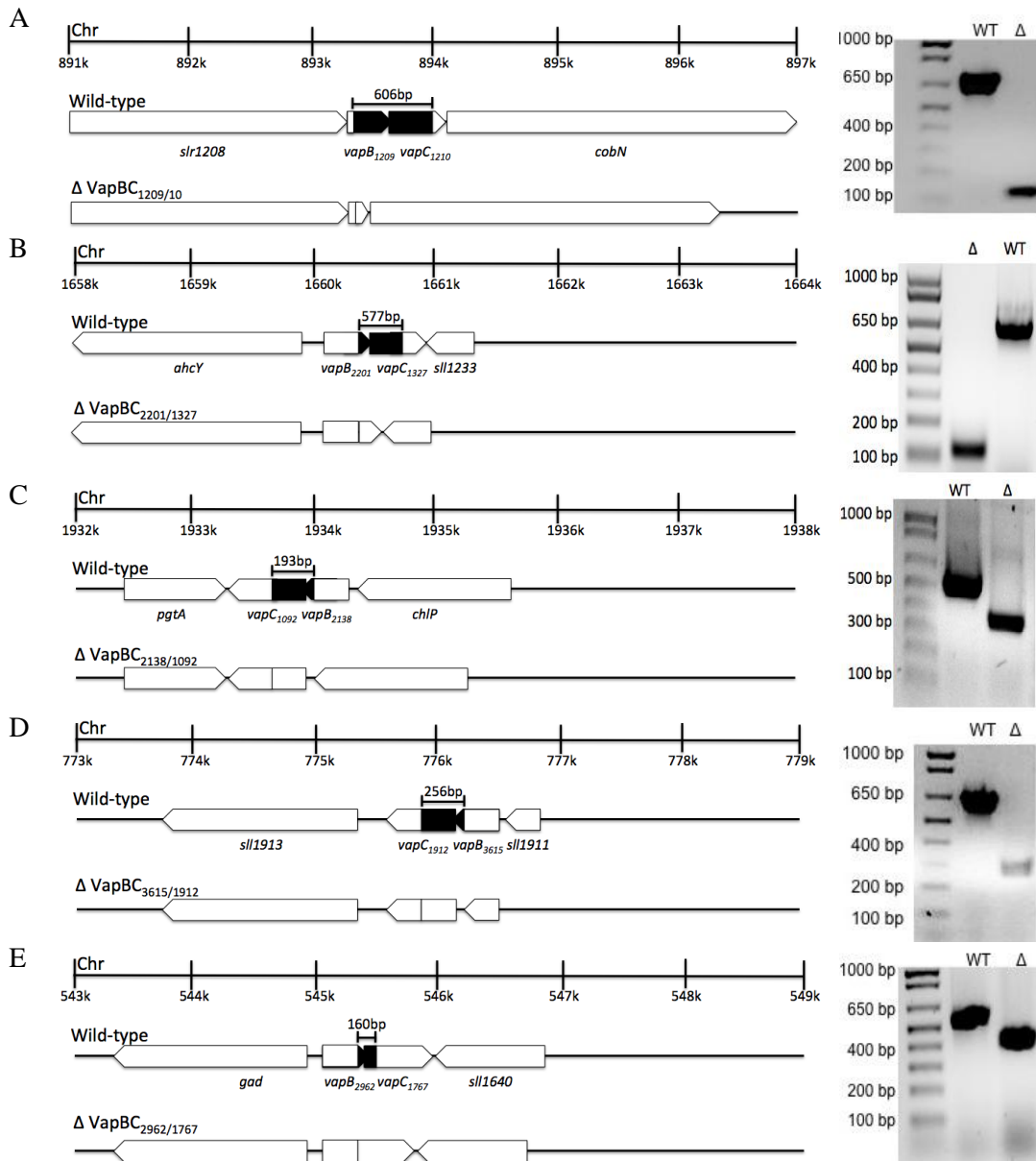


Figure 3.11. Schematic of *Synechocystis* markerless $\Delta vapBC$ deletion strains and PCR diagnostic gels confirming deletion. The top scale indicates the position of the operon on the chromosome. Each gene within the operon and adjacent to the operon are represented by arrowed boxes. The region of DNA within the operon that was removed, including regions of both *vapB* and *vapC*, are shown in black. The size of the deleted region is shown in the WT strain and the gel represents this by change in size of the operon PCR product band of WT vs Δ .

3.3.6.2 *Synechocystis* $\Delta vapBC$ Deletion Strain Growth Experiments

To investigate if deletion of the VapBC systems had an affect on growth, three deletion strains were grown alongside the WT in BG-11 media under both full light ($40 \mu\text{mol photons m}^{-2} \text{s}^{-1}$) and 12 hr light-dark cycling conditions. Cultures were set up in triplicate and OD₇₅₀ readings were taken every 12 hrs over a 9 day period (Figure 3.12). OD₆₀₀ (contamination) and OD₆₈₀ (chlorophyll) measurements

Chapter Three

showed identical growth patterns to the OD₇₅₀ cell density measurements (Figure 3.12) indicating an absence of fast growing bacteria and no uncoupling of photosynthesis (as measured by chlorophyll) and cell growth. The *Synechocystis* $\Delta vapBC_{ssr2201/slr1327}$ strain showed no difference in growth compared to the WT strain in both full light and light-dark cycling (Figure 3.12). The *Synechocystis* $\Delta vapBC_{slr1209/1210}$ strain showed no difference in growth compared to the WT in light-dark cycles but grew slower than the WT under full light, especially during mid-late exponential growth (Figure 3.12). The *Synechocystis* $\Delta vapBC_{ssl2138/sll1092}$ strain grew slower than the WT in both conditions. In full light the deletion strain grew slower only for the first 12 hrs and in light-dark cycling the deletion strain grew slower for the full 9 days until the last time point (216 hrs) (Figure 3.12).

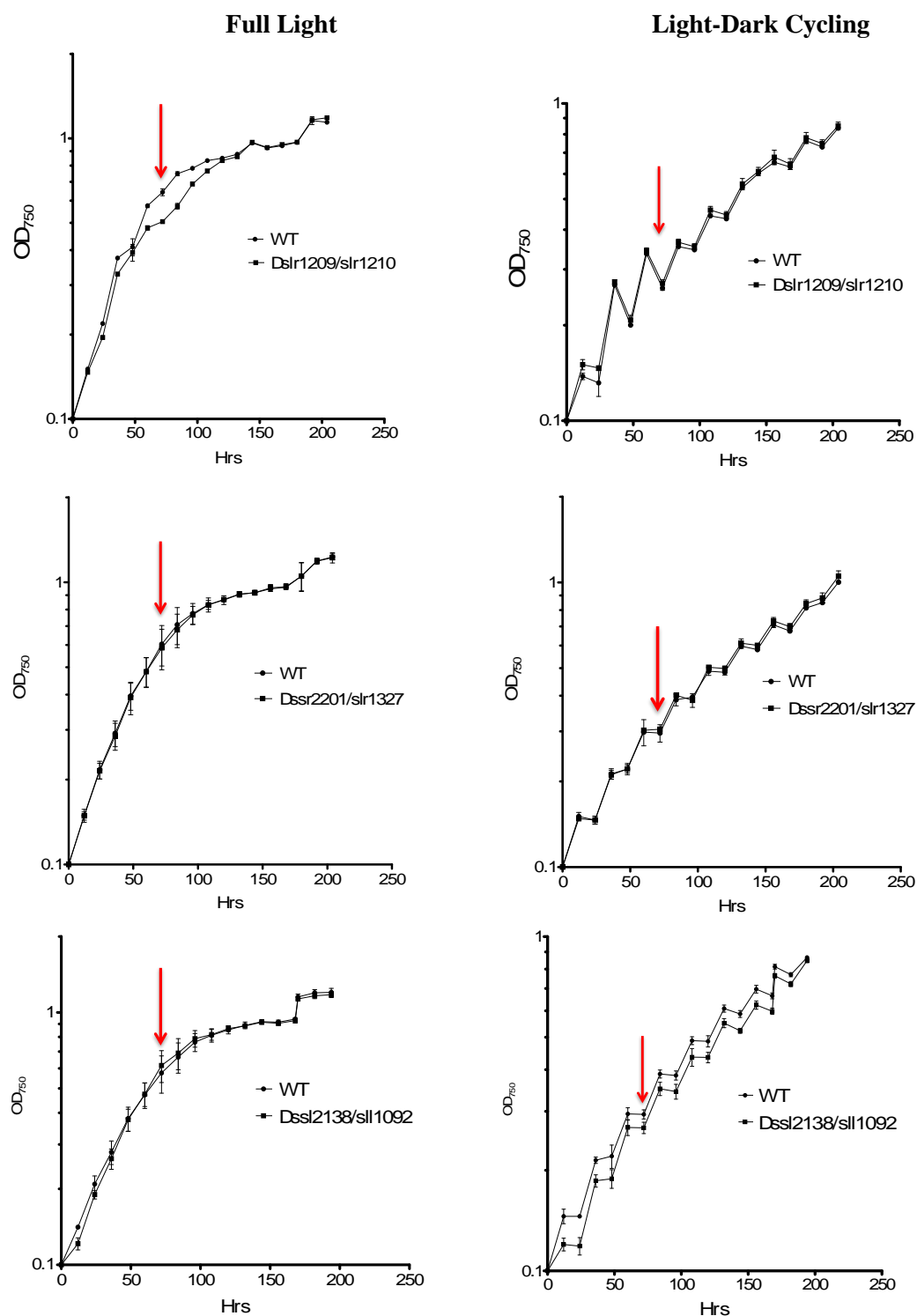


Figure 3.12. Growth of *Synechocystis* WT vs $\Delta vapBC$ strains over a 9 day period in either full light or 12 hr light-dark cycling. The graph is representative of one independent experiment with three biological replicates taken at each time point. The error bars represent SEM. Red arrows indicate time point (84 hrs) where samples were harvested for RNA-seq and RT-qPCR analyses.

No obvious changes in culture appearance or cell morphology were observed between strains during the duration of the growth experiments. For each strain where a difference in growth was observed compared to the WT the growth,

experiments were repeated and RNA was extracted at mid to late exponential growth (Figure 3.12) for subsequent RNAseq and RT-qPCR analysis.

3.3.6.3 RNA-seq Analysis of *Synechocystis* WT vs $\Delta vapBC$ Strains

The *Synechocystis* $\Delta vapBC_{slr1209/1210}$ and $\Delta vapBC_{ssl2138/sll1092}$ strains both showed a reduced growth rate compared to the WT. RNA-seq was used to investigate the transcriptional profiles of each strain and how the deletion of the VapBC systems resulted in slower growth for each of these strains. Cells were harvested at mid to late exponential phase (84 hours) to obtain a high enough yield of RNA for subsequent processing and analysis. Purified RNA was sent to BGI genome sequencing facility in Hong Kong for quality assessment, sample processing, cDNA library generation and sequencing, and analysis.

3.3.6.3.1 RNA-seq Analysis of *Synechocystis* WT vs $\Delta vapBC_{slr1209/1210}$

To confirm deletion of the *vapBC_{slr1209/1210}* operon the expression profile of each gene was assessed in the WT and deletion strain dataset using fragments per kilobase of transcript per million mapped reads (FPKM). In full light the WT expression profile for *vapB_{slr1209}* is 70.03 and *vapC_{slr1210}* is 5.29 showing expression of both of genes is low. In full light the deletion strain showed no expression of *vapB_{slr1209}*, however *vapC_{slr1210}* is indicated as present (2.89). The raw transcript reads show that there are only 3 raw reads for this gene, suggesting that this could be an artefact of the RNA-seq analysis, rather than true expression. In the light dark cycling conditions the WT expression profile for *vapB_{slr1209}* is 42.98 and *vapC_{slr1210}* is 5.43, again showing expression is low. In the deletion strain there is no expression of either *vapB_{slr1209}* or *vapC_{slr1210}*, confirming that this operon has been deleted from the genome.

Comparison of the transcriptional response of the WT compared to $\Delta vapBC_{slr1209/1210}$ grown under full light showed 16 differentially expressed genes (DEG) ($>2.0 \times$ change, $P < 0.05$), with 5 genes upregulated and 11 genes downregulated in the deletion strain (Table 3.6). The most upregulated gene in the deletion strain is a putative transposase (*slr0352*), which has a 7.62 Fold increase in expression. Other upregulated genes include those encoding two hypothetical proteins (*ssr3402* and *sll5075*), the Photosystem II accessory pigment CpcG2

Chapter Three

(*cpcG2*) and a putative oxidoreductase (*slr2008*). The range of upregulation for these genes was between 2.22 Fold to 2.71 Fold. The most downregulated gene in the deletion strain is a putative transposase (*sll0650*), which is downregulated 11.03 Fold. Other downregulated genes include genes encoding another putative transposases (*slr0800*), five hypothetical proteins (*ssr6032*, *slr6075*, *slr1056* and *ssr1768*), the glucose transport protein (GlcP, *glcP*), the sulphate transport system binding protein (SbpA, *sbpA*), an assembly protein for Photosystem II (*slr0286*), a putative cation transporter (*sll0671*) and a putative endonuclease (*sll0709*). The range of downregulation for these genes was between 2.09 Fold to 10.82 Fold (Table 3.6). Although each DEG is statistically supported, all but 3 show very low expression (FPKM <74) (Table 3.6). As a result, the observed differential expression is likely artificially inflated and an artefact of the RNA seq analysis of low copy number comparisons, rather than true differential expression. Therefore, the genes *cpcG2* and *glcP*, where the difference in FPKM is significant ($\Delta\text{FPKM} > 100$), are considered the only significant DEG from this analysis (Table 3.6).

Table 3.6. List of genes that are differentially expressed (>2.0 x change, P<0.05) in *Synechocystis* $\Delta vapBC_{slr1209/1210}$ compared to WT under full light. Rows coloured in orange indicate DEG present in both full light and light-dark cycling. Rows coloured green indicate DEG with significant differences in FPKM (Δ FPKM>100).

Locus	Expression ratio ^a	P-value	WT expression (FPKM)	Δ expression (FPKM)	Encoded protein
<i>slr0352</i>	7.62	3.03E-06	0.09	17.66	Putative transposase
<i>ssr3402</i>	2.71	4.96E-05	11.01	72.21	Hypothetical protein
<i>cpcG2</i>	2.38	2.94E-33	45.87	239.25	Phycocyanin linker protein (CpcG2)
<i>slr2008</i>	2.31	2.50E-05	14.85	73.61	Putative oxidoreductase (MrpC)
<i>sll5075</i>	2.22	6.63E-05	12.19	56.83	Hypothetical protein
<i>sll0650</i>	-11.03	5.21E-09	41.87	0.02	Putative transposase
<i>ssr6032</i>	-10.82	5.28E-05	36.23	0.02	Hypothetical protein
<i>glcP</i>	-3.00	6.60E-77	177.47	22.13	Glucose transport protein (GlcP)
<i>sbpA</i>	-2.56	4.52E-05	13.72	2.33	Sulphate transport system substrate binding protein (SbpA)
<i>slr0286</i>	-2.37	5.38E-09	25.56	4.93	Assembly protein for Photosystem II
<i>sll0671</i>	-2.32	9.77E-06	16.91	3.39	Putative cation transporter
<i>sll0709</i>	-2.16	1.36E-10	32.51	7.31	Putative endonuclease
<i>slr6075</i>	-2.10	2.13E-10	19.17	4.47	Hypothetical protein
<i>slr1056</i>	-2.09	1.57E-13	30.67	7.22	Hypothetical protein
<i>ssr1768</i>	-2.08	1.41E-4	98.2	23.15	Hypothetical protein
<i>slr0800</i>	-2.08	6.18E-11	122.52	28.9	Putative transposase

^a Log2 expression ratio ($\Delta vapBC_{slr1209/1210}$ /WT).

Comparison of the transcriptional response of the WT compared to $\Delta vapBC_{slr1209/1210}$ grown under light-dark cycling showed 7 DEG (>2.0 x change, P<0.05) with 3 genes upregulated and 4 genes downregulated in the deletion strain (Table 3.7). The most upregulated gene in the deletion strain is a putative transposase (*sll1997*), which is upregulated 10.92 Fold. The 2 other upregulated

Chapter Three

genes also encode putative transposases (*sll1255* and *slr0350*) (5.96 Fold and 10.24 Fold upregulated, respectively). The most downregulated gene in the deletion strain is a putative transposase (*sll0651*), which is downregulated 8.47 Fold. Other downregulated genes include genes that encode another putative transposase (*sll0315*), the glucose transport protein (GlcP, *glcP*), and a putative major outer membrane protein (MomP, *sll0772/momP*). These genes are downregulated from 2.74 Fold to 5.97 Fold (Table 3.7). Although each DEG is statistically supported, all but 2 show very low expression (FPKM <70) (Table 3.7), and, as mentioned earlier, the observed differential expression of these genes is likely an artefact of RNA seq analysis. As a result the genes *glcP*, and *momP*, where difference in FPKM is significant ($\Delta\text{FPKM} > 100$), are considered the only significant DEG from this analysis (Table 3.7).

Table 3.7. List of genes that are differentially expressed in *Synechocystis* $\Delta\text{vapBC}_{slr1209/1210}$ compared to WT in light-dark cycling. Rows coloured in orange represent DEG present in both full light and light-dark cycling. Rows coloured green indicate DEG with significant differences in FPKM ($\Delta\text{FPKM} > 100$).

Locus	Expression ratio ^a	P-value	WT expression (FPKM)	Δ expression (FPKM)	Encoded protein
<i>sll1997</i>	10.92	1.66E-06	0.01	19.37	Putative transposase
<i>sll1255</i>	10.24	1.52E-09	0.01	12.08	Putative transposase
<i>slr0350</i>	5.96	1.64E-15	0.47	29.33	Putative transposase
<i>sll0651</i>	-8.46	1.37E-05	24.57	0.07	Putative transposase
<i>sll0315</i>	-5.97	1.81E-20	68.75	1.1	Putative transposase
<i>glcP</i>	-3.30	3.62E-135	337.03	34.26	Glucose transport protein (GlcP)
<i>sll0772</i> (<i>momP</i>)	-2.74	4.75E-104	262.05	39.19	Probable porin, major outer membrane protein (MomP)

^a Log2 expression ratio ($\Delta\text{vapBC}_{slr1209/1210}/\text{WT}$).

In summary, the RNA-seq analysis identified two genes, *cpcG2* and *glcP*, grown under full light and two genes, *glcP* and *momP*, grown under light-dark cycling that are differentially expressed in the $\Delta\text{vapBC}_{slr1209/1210}$ deletion strain. *cpcG2* is the only upregulated gene in the deletion strain and is only upregulated

Chapter Three

in full light conditions. CpcG2 is a structural protein that composes part of the light harvesting antennae (phycobilisome) in Photosystem II of the photosynthetic pathway. Based on the observed differential expression, under full light conditions it appears VapBC_{slr1209/1210} normally downregulates CpcG2 expression. The genes *glcP* and *momP* are downregulated in both conditions, where *glcP* is significantly downregulated in both conditions ($>2.0 \times$ change, $P < 0.05$, $\Delta\text{FPKM} > 100$) and *momP* is significantly downregulated in light-dark and 1.66 Fold ($P = 1.33 \times 10^{-35}$, $\Delta\text{FPKM} = 103.59$) in full light. GlcP belongs to the sugar transport protein family, and is responsible for glucose and fructose transportation during heterotrophic growth in *Synechocystis*. MomP is a hypothetical protein annotated as a putative outer membrane protein and is found 91 bp downstream of *glcP*. Because these genes are downregulated in the deletion strain, it appears that VapBC_{slr1209/1210}, when present, may play a role in inducing the expression of these genes. Collectively, it appears that VapBC_{slr1209/1210} plays a role in repressing the expression of CpcG2, which is involved in photosynthesis, and inducing the expression of GlcP and MomP, which sit adjacent to each other in the genome, forming an operon responsible for glucose uptake.

3.3.6.3.2 RNA-seq Analysis of *Synechocystis* WT vs $\Delta\text{vapBC}_{ssl2138/sll1092}$

To confirm deletion of the *vapBC*_{ssl2138/sll1092} operon the expression profile of each gene in the operon was assessed in the WT and deletion strain using FPKM. The WT expression profile for *vapB*_{ssl2138} is 3.21 and *vapC*_{sll1092} is 39.13 showing expression for both genes is low. The deletion strain also showed expression of both *vapB*_{ssl2138} (4.75) and *vapC*_{sll1092} (37). This result was unexpected, considering the previous confirmation of the deletion from the genome (Figure 3.11). To investigate this further, the raw transcript read numbers of each gene were compared between the WT and deletion strain datasets. The WT dataset has 2 transcript reads of *vapB*_{ssl2138}, where the deletion strain has 1, and 50 transcript reads for *vapC*_{sll1092} where the deletion strain has 16. There are two possible explanations for this. The first is that expression of both genes is so low that the transcript reads are unable to be quantified accurately. The second explanation is that the operon has not been successfully deleted from the genome and further segregation is required to ensure the deletion in all chromosomes; however, this should have been detected in the initial diagnostic. The

Chapter Three

transcriptional response of the WT compared to $\Delta vapBC_{ssl2138/sll1092}$ was subsequently analysed however, the results of this comparison are provided as a preliminary assessment only.

Comparison of the transcriptional response of the WT compared to $\Delta vapBC_{ssl2138/sll1092}$ grown under light-dark cycling showed 19 DEG (>2.0 x change, $P < 0.05$), with 5 genes upregulated and 14 genes downregulated in the deletion strain (Table 3.8). The most upregulated gene in the deletion strain is a putative transposase (*sll0650*), which is upregulated 11.76 Fold. All other upregulated genes encode for manganese transport (*mntA*, *mntB* and *mntC*) and a putative cell invasion protein (*sipA*), with upregulation from 2.02 Fold to 2.88 Fold. The most downregulated gene in the deletion strain is a hypothetical protein (*ssr3517*), which is downregulated 11.96 Fold. Other downregulated genes encode for 3 putative transposases (*slr0462*, *sll0651* and *sll1397*), 5 hypothetical proteins (*slr1851*, *ssr1251*, *slr1611*, *sll0253* and *ssr6048*), the Chlorophyll *a* ring formation protein (*ChlA_{II}*, *chlA_{II}*), the Phosphoenolpyruvate synthase (*Pps*, *pps*) and a putative glycosyltransferase (*slr1065*). The range of downregulation of these genes are from 2.02 Fold to 10.87 Fold (Table 3.8). Although each DEG is statistically supported, 12 show very low expression (FPKM < 78) (Table 3.8) indicating that the differential expression observed may be an artefact of the RNA seq analysis, as mentioned earlier, rather than a true differential expression. However, the genes *mntA*, *mntB* and *mntC* encode proteins involved in manganese transport and appear to form an operon. Although these genes have poor expression (FPKM < 75), each one is upregulated in the deletion strain, suggesting that this may be a true dysregulation of an entire operon. As a result, this operon and the 7 downregulated genes *slr1851*, *ssr1251*, *sll0253*, *ssr6048* and *sll1863* (hypothetical proteins), *chlA_{II}* and *pps* are considered significant DEG from this analysis (Table 3.8).

Table 3.8. List of genes that are differentially expressed in *Synechocystis* $\Delta vapBC_{ssl2138/ssl1092}$ compared to WT in light-dark cycling. Rows coloured green indicate DEG with significant differences in FPKM ($\Delta FPKM > 100$). Rows coloured in blue indicate low expression DEG that form the manganese transport machinery operon and are all upregulated indicating true differential expression irrelevant of low FPKM.

Locus	Expression ratio ^a	P-value	WT expression (FPKM)	Δ expression (FPKM)	Encoded protein
<i>sll0650</i>	11.76	8.37E-09	0.01	34.76	Putative transposase
<i>mntC</i>	2.88	1.22E-26	10.05	73.78	Manganese Transporter (MntC)
<i>mntB</i>	2.58	4.62E-22	12.42	74.28	Manganese transport system membrane protein (MntB)
<i>sipA</i>	2.46	1.50E-4	9.16	50.24	Putative cell invasion protein (SipA)
<i>mntA</i>	2.02	2.51E-10	12.87	52.15	Manganese Transport ATP-binding protein (MntA)
<i>ssr3571</i>	-11.96	1.03E-4	39.76	0.01	Hypothetical protein
<i>slr0462</i>	-10.87	1.03E-4	18.78	0.01	Putative transposase
<i>sll0651</i>	-5.18	1.01E-05	24.57	0.68	Putative transposase
<i>slr1851</i>	-4.71	3.84E-25	140.09	5.37	Hypothetical protein
<i>sll1397</i>	-4.63	4.34E-13	37.71	1.52	Putative transposase
<i>ssr1251</i>	-3.91	2.52E-09	166.33	11.04	Hypothetical protein
<i>slr1611</i>	-3.29	3.41E-08	57.32	5.84	Hypothetical protein
<i>chlA_{II}</i>	-3.15	3.84E-34	110.57	12.47	Chlorophyll <i>a</i> ring formation protein (ChlA _{II})
<i>sll0665</i>	-2.56	1.91E-05	55.64	9.41	Putative transposase
<i>sll0253</i>	-2.17	8.77E-92	928.26	206.13	Hypothetical protein
<i>ssr6048</i>	-2.12	3.11E-09	220.35	50.73	Hypothetical protein
<i>pps</i>	-2.05	1.50E-289	660.86	159.24	Phosphoenolpyruvate synthase (Pps)
<i>sll1863</i>	-2.03	1.04E-28	530.36	129.63	Hypothetical protein
<i>slr1065</i>	-2.02	3.81E-14	77.69	19.2	Putative glycosyltransferase

^a Log2 expression ratio ($\Delta vapBC_{ssl2138/ssl1092}$ /WT).

Aside from the manganese transport operon, all significant DEG are downregulated in the deletion strain, with the majority encoding hypothetical

Chapter Three

proteins (5/7). The two characterised genes are *pps*, which encodes Pps, a protein involved in pyruvate and acetyl-CoA metabolism, and *chlA_{II}*, which encodes ChlA_{II}, an assembly protein of Chlorophyll *a*. Both of these genes are downregulated in the deletion strain indicating that VapBC_{ssl2138/sll1092} plays a role in inducing the expression of these genes. Although these results are preliminary, they do suggest that VapBC_{ssl2138/sll1092} plays a role in repressing the expression of an entire operon responsible for manganese transport, and inducing the expression of a cohort of hypothetical proteins and genes involved in both photosynthesis and the glucose metabolism.

3.3.6.4 RT-qPCR Analysis

Because of the success in protein expression of VapBC_{slr1209/1210} and subsequent preliminary *in vitro* characterisation of this system, the DEG identified in the Δ vapBC_{slr1209/1210} RNA-seq analysis were chosen for further RT-qPCR analysis (Table 3.9). To test whether transposases had true differential expression, four were selected for RT-qPCR validation. Additionally, the genes *glcP* and *momP* were chosen for validation as they were consistently downregulated in the deletion strain for both conditions tested, making them the most promising candidate of VapBC characterisation.

Table 3.9. List of RNA-seq determined DEG in *Synechocystis* Δ vapBC_{slr1209/1210} analysed by RT-qPCR. Rows in green indicate DEG that were successfully validated by RT-qPCR (>2.0 x change).

Gene	Protein	RNA-seq expression ratio ^a	WT strain expression profile (FPKM)	Δ strain expression profile (FPKM)	RT-qPCR relative expression in (KO vs WT) ^b
<i>sll1997</i>	Transposase	↑10.92	0.01	19.37	Not detected
<i>sll1255</i>	Transposase	↑10.24	0.01	12.08	Not detected
<i>slr0350</i>	Transposase	↑5.96	0.47	29.33	↑1.24
<i>sll0315</i>	Transposase	↓5.97	68.75	1.1	Not detected
<i>glcP</i>	GlcP	↓3.30	337.03	34.26	↓83.33
<i>momP</i>	Momp	↓2.74	262.05	39.19	↓3.15

^a Log2 expression ratio (Δ vapBC_{ssl2138/sll1092}/WT).

^b $2^{(-\Delta\Delta Ct)}$ for upregulated genes and its reciprocal ($1/(2^{(-\Delta\Delta Ct)})$) for downregulated genes to represent ratio relevant to WT.

RNA-seq determined differential expression of all transposases tested could not be validated by RT-qPCR given the required parameters of this test (>2.0 x change) (Table 3.9). This result supports the previous assumption that the enrichment of transposases, and other genes with low expression, in the total DEG of the RNA-seq analyses are likely an artefact of the analysis, rather than actual differential expression. Both *glcP* and *momP* show significant differential expression, with both genes being downregulated in the deletion strain, validating the RNA-seq results (Table 3.9) (Figure 3.13).

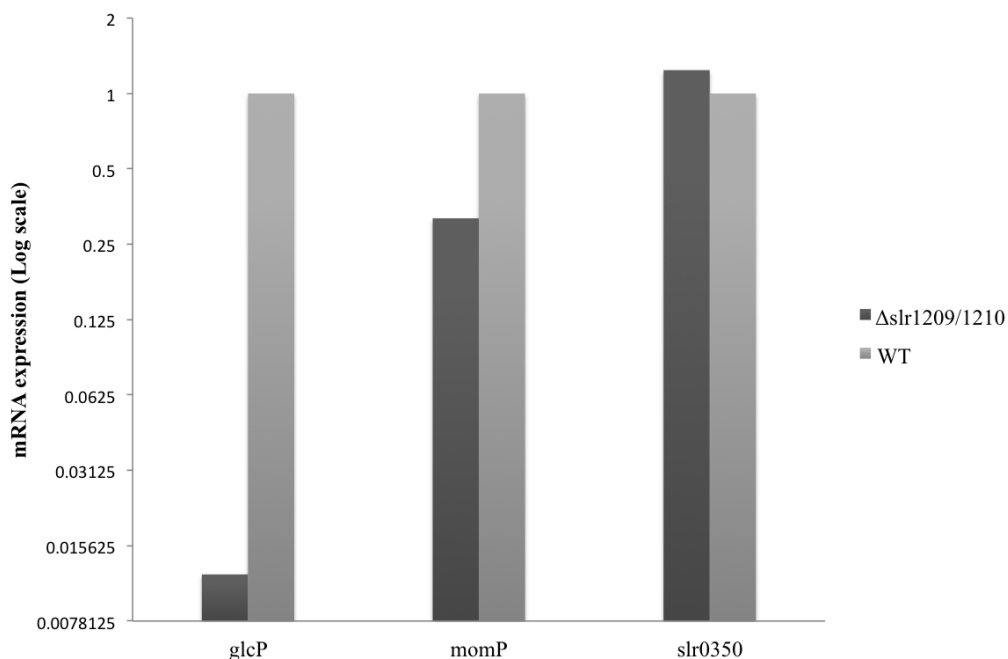


Figure 3.13. RT-qPCR $2^{-\Delta\Delta C_t}$ fold change expression profiles of *glcP*, *momP*, and *slr0350* in $\Delta vapBC_{slr1209/1210}$ vs WT. RT-qPCR analysis sets the WT gene mRNA expression to an arbitrary expression level of 1 for comparison to the $\Delta vapBC_{slr1209/1210}$ gene expression.

In summary, these results provide an indication, in combination with the RNA-seq results, that the $VapBC_{slr1209/1210}$ system affects the expression of these genes. However, these results are preliminary and require further optimisation, with increased sample numbers to allow for the appropriate statistics to be done. Further investigation of the relationship between $VapBC_{slr1209/1210}$ and the genes *glcP* and *momP*, as putative cellular targets provide an exciting avenue to elucidate the role of this $VapBC$ system in cyanobacteria.

3.4 Discussion

VapBC systems are the largest family of the Type II TA systems and have become a point of interest because of their overrepresentation in the genomes of pathogenic and environmentally hazardous bacteria. Several VapBC systems from a range of bacteria have been implicated in persistence and stress response via inducing growth arrest by inhibiting translation; however, the cellular targets and specific biological roles of most of these systems are yet to be elucidated.

The sole VapBC_{1283/4} system in *M. smegmatis*, the non-pathogenic relative of *M. tuberculosis*, is one of the best characterised systems and has provided significant insight into the biological functioning of VapBC systems. VapBC_{1283/4} targets and cleaves mRNA transcripts that encode the glycerol transport machinery. In the $\Delta vapBC_{1283/4}$ strain glycerol uptake was greater than that of the WT, however the increased rate of consumption was uncoupled to biomass generation. From this, a model was proposed that VapBC_{1283/4} regulates the rate of glycerol utilisation to match the anabolic demands of the cell, allowing for fine-tuning of the catabolic rate at a posttranscriptional level.

M. tuberculosis has an expansion of VapBC systems (47) encoded in its genome. Although preliminary characterisations have been done for a number of these systems, the actual biological targets remain to be elucidated. However, studies have implicated these VapBC systems as an intricate stress response array, where specific cohorts of VapBC systems are activated in response to different environmental triggers to induce bacteriostasis via translation inhibition. The dormancy, or persistence of *M. tuberculosis*, which is a result of bacteriostasis, is one of the most challenging characteristics of this pathogen in terms of human infection and treatment, as current treatments are in the form of antibiotics that target active cells. As a result, VapBC systems are a valuable biological target for investigating the capacity to prevent persistence and increasing susceptibility to antibiotic treatment.

The environmentally hazardous toxic bloom forming *M. aeruginosa* shares physiological characteristics with *M. tuberculosis*, in particular, its ability to persist in a dormant state while nutrients are limiting. *M. aeruginosa* also has an

Chapter Three

expansion of VapBC systems (46 predicted by TADB) encoded in its genome, however, none of these systems have been investigated. In fact, there is only one VapBC system in cyanobacteria that has been characterised: VapBC10 (*ssr2962/slr1767*) from *Synechocystis*, the model organism of cyanobacteria. VapC10 was shown to induce the characteristic bacteriostatic phenotype of a VapC toxin, which is reversed upon the introduction of VapB. It is proposed that VapC10 exerts its growth arrest via translation inhibition, like many other characterised VapC toxins, however this has not been confirmed experimentally. Because this is the only VapBC system to be characterised in cyanobacteria and the functional mechanism is still unknown, it is difficult to make any generalisations about the biological role that these VapBC systems play. In this study, two *M. aeruginosa* VapBC systems and three *Synechocystis* VapBC systems have been investigated.

3.4.1 *In vitro* Investigation of Cyanobacteria VapBC Systems in this Study

To investigate the VapBC systems in *M. aeruginosa*, two representative VapBC systems were selected: VapBC_{MAE49640/50} and VapBC_{MAE43230/20}. Expression in *E. coli* of VapBC_{MAE49640/50} was unsuccessful and expression of VapBC_{MAE43230/20} was successful only when expressed with a SUMO fusion protein attached to VapB_{MAE43230}. VapB_{MAE43230} and VapC_{MAE43220} are co-expressed with an excess of VapB being expressed relative to VapC. This expression ratio indicates that *vapBC*_{MAE43230/20} is expressed as a typical *vapBC* operon, where expression is characterised by a ratio of up to 10:1 VapB:VapC as a result of the overlap present within the operon. VapB_{MAE43230} and VapC_{MAE43220} form a tetrameric VapBC_{MAE43230/20} protein complex bound to genomic DNA, which appears to be unstable, with the majority of the complex forming a soluble aggregate. VapB_{MAE43230} also appears to be purified bound to genomic DNA and forms a tetramer. Further optimisation is required to reduce the aggregation of the complex and removal of co-purified DNA before *in vitro* characterisation can be conducted.

Because of the difficulty faced with expression of the *M. aeruginosa* VapBC systems, *Synechocystis* VapBC systems were trialled for expression. All VapBC systems expressed poorly, with the exception of VapBC_{slr1209/1210}. VapB_{slr1209} and VapC_{slr1210} are co-expressed and form a stable dimeric complex

Chapter Three

that is not bound to DNA. The ratio of expression could not be determined for this systems as the His tag was bound to VapC, resulting in excess VapB being lost during the initial IMAC purification step. Disruption of the complex was achieved by subjecting the complex to acidic buffering conditions, which resulted in a change in net charge of each protein and subsequent disruption the ionic interactions that stabilize the complex.

Preliminary EMSA assays indicate autoregulation by VapBC_{slr1209/10} by binding to the promoter region of the *vapBC_{slr1209/1210}* operon. Analysis of the promoter region identified two overlapping palindromic sequences as putative VapB_{slr1209} recognition motifs, and EMSA assays show that the presence of these sites are necessary for binding to occur. To validate these results, additional mutation work is required to confirm the recognition sites, and assays should be optimised to confirm the specificity of binding.

Using the RNA pentaprobe system VapC_{slr1210} was confirmed to have metal specific endoribonuclease activity. VapC_{slr1210} ribonuclease activity was observed only on pentaprobe 927, which was inhibited when EDTA or VapB_{slr1209} was added to the assays. The degradation pattern shows two prominent bands between 50-80 bp, suggesting that VapC_{slr1210} selectively cleaves the RNA at the centre of the 150 bp molecule in an endonuclease mode to produce two RNA fragments ~70 bp and 80 bp. The site of cleavage is certainly not optimal due to the slow rate of cleavage. Binding of VapC_{slr1210} to the 927 pentaprobe was not observed, suggesting there is no dual recognition function by binding and cleaving its RNA substrate. This lack of dual function has also been observed, using the pentaprobe system, in characterised VapC proteins from *M. tuberculosis* and *Pyrobaculum aerophilum*. The only exception to this is the sole *M. smegmatis* VapC₁₂₈₄, where binding of the pentaprobe was observed. The binding was abolished in the presence of VapB or EDTA.

3.4.2 *In vivo* Investigation of VapBC Systems

Markerless deletion strains of five different *Synechocystis* *vapBC* operons were made, and growth experiments of three of the deletion strains were conducted to determine the affect of the deletion of these VapBC systems on fitness of the organism. Growth experiments of *Synechocystis* WT vs Δ *vapBC*

Chapter Three

strains identified a reduced growth rate in two of the deletion strains: $\Delta vapBC_{slr1209/1210}$ and $\Delta vapBC_{ssl2138/sll1092}$. Transcriptomic analyses of both deletions strains were subsequently carried out to investigate the cause of the observed phenotypes.

RNA-seq analysis of the WT in both full light and light-dark cycling show that both $VapBC_{slr1209/1210}$ $VapBC_{ssl2138/sll1092}$ systems are expressed in relatively low abundance. RNA-seq analysis of the $\Delta vapBC_{ssl2138/sll1092}$ transcriptome (light-dark cycling) show transcripts present for both $VapB_{ssl2138}$ (1) and $VapC_{sll1092}$ (16), which are very low and, as a result, difficult to acknowledge as true indication of expression. RNA-seq analysis of the $\Delta vapBC_{slr1209/1210}$ transcriptome in light-dark cycling confirmed the deletion of the $vapBC_{slr1209/1210}$ operon, however in the full light analysis there are 3 raw transcript reads for $VapC_{slr1210}$, which is also so minimal that it is difficult to acknowledge as true expression. Given that deletions were confirmed prior to RNA-seq analysis, for the purposes of this study, it is assumed that all expression profile comparisons of poorly expressed genes cannot be accurately quantified and are unreliable. This was supported by RT-qPCR analysis, where a number of RNA-seq identified DEG that were poorly expressed could not be validated. As a result, a significant number of DEG identified by BGI analysis have been removed from this study, including all transposases.

Comparison of the transcriptional response of the WT compared to $\Delta vapBC_{ssl2138/sll1092}$ grown under light-dark cycling showed 10 DEG with 3 genes, *mntA*, *mntB* and *mntC*, upregulated and 7 genes, *slr185I*, *ssr125I*, *sll0253*, *ssr6048*, *sll1863* (hypothetical proteins), *chlA_{II}* and *pps*, downregulated in the deletion strain. All 3 upregulated genes form an operon that encodes manganese transport machinery (MntA, MntB and MntC). Accumulation of manganese is essential for the assembly of the Photosystem II manganese cluster in *Synechocystis* and subsequent oxygenic photosynthesis (Bartsevich & Pakrasi, 1996). Because this manganese transport operon is upregulated in the deletion strain, it is possible that they are targets of $VapBC_{ssl2138/sll1092}$, which acts to inhibit their translation; however, further validation of these genes as targets is required. The two characterised downregulated genes encode Pps, a protein involved in pyruvate and acetyl-CoA metabolism and ChlA_{II}, an assembly protein of

Chapter Three

Chlorophyll *a*. As both of these genes are downregulated in the deletion strain, it is possible that VapBC_{ssl2138/sll1092} plays a role in inducing the expression of these genes, however further validation is required.

Comparison of the transcriptional response of the WT compared to *ΔvapBC_{slr1209/1210}* grown under both full light and light-dark cycling showed differential expression of genes exclusively involved in photosynthesis and glucose uptake. In full light, the two differentially expressed genes are *cpcG2*, which is upregulated, and *glcP*, which is downregulated. Under light-dark cycling *glcP* is also downregulated, as well as the gene *momP*, which is found 91 bp downstream of *glcP*. Both *glcP* and *momP* downregulation was validated by preliminary RT-qPCR analysis and *cpcG2* was not tested.

cpcG2 encodes CpcG2, a structural protein that composes part of the light harvesting antennae (phycobilisome) in Photosystem II of the photosynthetic pathway. Based on the observed differential expression, under full light conditions, it appears VapC_{slr1210} normally downregulates *cpcG2* expression. Given that this protein is overexpressed in the deletion strain, which is likely due to increase light harvesting ability and subsequent carbon fixation and glucose production, it is interesting that the effect is a reduced growth phenotype compared to the WT. As mentioned earlier, the only other well characterised biological targets of VapBC systems are that of *M. smegmatis* VapBC_{1283/4}, which also include the cellular machinery responsible for carbon utilisation and metabolism. Deletion of VapBC_{1283/4} had an uncoupling affect of carbon uptake and biomass production, which identified the role of VapBC_{1283/4} in regulating the rate of glycerol utilisation to match the anabolic demands. It also appears that the deletion of VapBC_{slr1209/1210} in *Synechocystis* results in a similar dysregulation between autotrophic processes of carbon fixation, via CpcG2, and biomass production, as shown by the reduced growth phenotype.

glcP encodes GlcP, which is responsible for glucose transportation during heterotrophic growth in *Synechocystis* and sits upstream of *momP*. Previous studies have investigated both *glcP* and *momP* for their role in sugar transport in *Synechocystis* and if they function as a polycistronic operon (Schmetterer, 1990; Zhang, Durand, Jeanjean, & Joset, 1989; Zhang, Jeanjean, & Joset, 1998). All

Chapter Three

studies identified GlcP as an essential glucose importer, however the function of MomP still remains unknown. Zhang et al. (1989) state that it is highly possible that both genes are co-transcribed as a single mRNA as the distance between the two genes is quite short. Pfam (Finn et al., 2014) analysis of the MomP protein sequence shows the closest alignment is to an OprB family member, a carbohydrate selective-porin protein from *Pseudomonas aeruginosa*, indicating a putative role in carbohydrate transport. Because these genes are downregulated in the deletion strain, it appears that VapBC_{slr1209/10} may play a role in inducing the expression of these genes.

With the exception of VapBC_{ssl2138/sll1092} also affecting the expression of an operon encoding magnesium transport and a cohort of hypothetical proteins, the general observation is that both VapBC systems appear to target genes involved in carbon assimilation and metabolism via both photosynthetic and heterotrophic processes. Both VapBC systems appears to target genes involved in photosynthesis, in particular the light harvesting apparatuses, which are essential for phototrophic carbon fixation. In terms of heterotrophic growth, VapBC_{ssl2138/sll1092} targets genes involved in the initial states of glycolysis, and VapBC_{slr1209/1210} targets an operon involved in glucose transport. Given the similarity in cellular targets between *M. smegmatis* VapBC_{1283/4} and both *Synechocystis* VapBC systems, in particular VapBC_{slr1209/1210}, it is tempting to speculate that these protein systems share similar biological functions; that is, they fine-tune the regulation of carbon metabolism by controlling the rate of glucose utilisation (via both carbon fixation and glucose transport) to match the anabolic demands of the cell.

Although these results are preliminary, they have provided insight to allow us to begin elucidating the cellular targets of these uncharacterised VapBC systems and their potential role in cyanobacteria. Future work to test the hypotheses developed from these preliminary findings is discussed in Chapter 4.

4 Summary Chapter

In this thesis I have focused on the mechanisms of transcription and post transcriptional regulation in the context of the environmental organisms *M. smegmatis* and cyanobacteria (*M. aeruginosa* and *Synechocystis*). These organisms are characterised as slow growing opportunists that are exposed to constantly fluctuating environments. The tight regulation of their cellular processes in response to their changing environments is crucial for their adaptation, making them suitable candidates for this research. The role of the poorly understood transcription factor AmtR in *M. smegmatis* was investigated and characterised for its role in adaptation to nitrogen limitation. The mRNA interferases, VapBC systems, were investigated and identified primarily as putative post-transcription regulators of carbon metabolism in *M. aeruginosa* and *Synechocystis*.

4.1 Transcriptional Regulation by *M. smegmatis* AmtR

There is an extraordinary diversity and abundance of transcription factors that span the prokaryotic kingdom, with over 30 different superfamilies. Transcription factors are responsible for regulation of the expression of genes involved in specific cellular processes in response to the environment. One of the most well understood transcription factors is the nitrogen regulatory system (NtrBC), present in most bacteria, which is responsible for response to nitrogen limitation. In bacteria that do not have an NtrBC system, including a number of actinobacteria, the global transcription factors GlnR and AmtR function as the major nitrogen-limiting response players. In the actinobacteria *C. glutamicum*, the response to nitrogen limitation is regulated via AmtR, which acts as a repressor of over 35 genes under nitrogen surplus. AmtR interacts with the P_{II} protein GlnK, under nitrogen limitation, leading to the release of AmtR from its target DNA and subsequent upregulation of its regulon. The *C. glutamicum* AmtR regulon is composed of genes that encode proteins responsible for ammonia uptake and assimilation, and alternative nitrogen uptake and metabolism.

The actinobacteria *M. smegmatis* also lacks a NtrBC system but encodes both GlnR and AmtR proteins. GlnR has been identified as an important nitrogen

Chapter Four

regulator that controls the genes encoding ammonia uptake and assimilation, alternative nitrogen uptake and metabolism, and the signal transduction P_{II} proteins. The preferred nitrogen source of *M. smegmatis* is ammonium but it is capable of scavenging different nitrogenous compounds under changing environmental conditions. The genome has a high prevalence of genes involved in urea transport and metabolism, suggesting that urea is an important alternative nitrogen source during nitrogen limitation. AmtR has been identified as a repressor of a number of genes involved in alternative urea metabolism and amino acid transport during non-nitrogen limiting conditions.

The first AmtR structure to be solved (described here) reveals the regulatory domain that contains a deep binding pocket and the DNA binding domain is defined by a helix-turn-helix (HTH) motif. AmtR forms a homodimer with each monomer composed of 10 α -helices that constitute two domains. The DBD aligns with the QacR family of TetR transcription factors, which are unique in their DNA binding mechanisms as they bind DNA as pair of dimers where TetR DNA binding is generally by one dimer. The similarities of AmtR and QacR suggest that AmtR binds DNA as a pair of dimers in the same mechanism as QacR. The large binding pocket in the centre of the LBD is accessed by a long entrance site and contains a residue Glu₁₇₀ on helix α 8 which forms a potential gate at the entrance site of the pocket that may influence the stability or efficiency of ligand binding.

Our work validated previous work that identified the *M. smegmatis* AmtR regulon, and subsequently we determined the DNA recognition sequence. The AmtR regulon has previously been identified as a cohort of genes involved in alternative nitrogen uptake and urea metabolism (Jessberger et al., 2012). We confirmed this with microarray analysis of the *M. smegmatis* AmtR deletion strain, which showed deregulation of an operon that encodes proteins responsible for alternative nitrogen uptake and urea metabolism: an amino acid permease, two hypothetical proteins, a urea carboxylase (UC), and an allophanate hydrolase (AH) respectively. The UC and AH proteins are known to form an enzyme complex that converts urea to ammonium via an ATP-dependent process. Subsequently, the AmtR DNA recognition sequence in the promoter region of this operon was identified by EMSA mutation work. The recognition motif is the 16

bp semi-palindromic ATCTGTCATCTGACAG, which shares similarities to its close relative, the *C. glutamicum* AmtR recognition motif. Both the nucleotide specificity and the size of the spacing regions between the two half sites of the recognition sequence region are critical for binding. Using the *M. smegmatis* AmtR binding sequence, a new actinomycetes AmtR consensus sequence was generated: the semi-palindromic 16-18 bp sequence ATnTnTn₆₋₈AnAG.

Finally, we identified two potential signalling molecules specific to AmtR that appear to function via a co-repressor mechanism of regulation. Previously, no information was known regarding how the signal for nitrogen limitation was transferred to AmtR in *M. smegmatis*. AmtR, like all TetR transcription factors, acts as a repressor of gene expression when bound to DNA in the promoter region of the genes that it regulates. *M. smegmatis* AmtR was tested for GlnK binding but no interaction was observed with either the native or adenylylated GlnK, suggesting that GlnK is not a signalling protein for AmtR. Screening of small molecules as potential ligands identified two molecules, xanthine and allantoin, as potential ligands for AmtR. Interestingly, the interaction was only observed when AmtR was bound to DNA, suggesting that the AmtR-DNA interaction is stabilised by the presence of these molecules. This is the first example where the repression of a TetR protein is potentially enhanced by the presence of a co-repressor, and it is the absence of the co-repressor that releases AmtR from the DNA to lift repression. Xanthine and allantoin are both metabolites of nitrogen metabolism and decreases in their intracellular levels are valid signals for intracellular nitrogen limitation. This dissipation signal would directly result in AmtR repression being lifted, and subsequent induction of the proteins involved in alternative nitrogen uptake and urea metabolism would occur as a response to nitrogen limitation.

4.2 Post-Transcriptional Regulation

Post-transcriptional regulation occurs by a diverse range of mechanisms. This thesis has focused on one particular mechanism: the VapBC family of the Type II Toxin Antitoxin (TA) systems found in the cyanobacteria *M. aeruginosa* and *Synechocystis*. VapBC systems are a point of interest because of their

Chapter Four

overrepresentation in the genomes of pathogenic and environmentally hazardous bacteria, including the toxic bloom formers *M. aeruginosa*.

Investigation of *M. aeruginosa* VapBC systems was limited by poor cloning and expression. However, one system, VapBC_{MAE43230/20}, had increased expression after incorporating the SUMO fusion protein into the expression system. Preliminary *in vitro* investigation of this system showed that VapB_{MAE43230} and VapC_{MAE43220} are co-expressed and formed a tetrameric complex that appears to be bound to DNA. Significantly more VapB is expressed relative to VapC, suggesting that the *vapBC*_{MAE43230/20} operon is expressed as a typical *vapBC* operon. Future characterisation of this operon will provide great insight into the role of VapBC systems in *M. aeruginosa*.

Due to the difficulties involved with *M. aeruginosa* VapBC investigations, the cyanobacteria model organism *Synechocystis* VapBC systems were subsequently investigated. Similar to the *M. aeruginosa*, all but one VapBC system proved too difficult to express and purify. The sole soluble VapBC system, VapBC_{slr1209/1210}, was investigated *in vitro*. VapB_{slr1209} and VapC_{slr1210} are co-expressed and form a dimeric complex that does not appear to be bound to DNA and could be disrupted by acid dissociation. Preliminary EMSA assays indicate VapBC_{slr1209/1210} is autoregulatory and binds to the promoter region of its own operon by recognising palindromic sequences upstream of the start codon. Using the RNA pentaprobe system, VapC_{slr1210} shows metal specific endoribonuclease activity, which could be inhibited by VapB_{slr1209}.

Markerless deletion strains of five different *Synechocystis* *vapBC* operons were made, and growth experiments of three of the deletion strains were conducted to determine the effect of the deletion of these VapBC systems on fitness of the organism. Growth experiments identified a reduced growth rate in two of the deletion strains: Δ *vapBC*_{slr1209/1210} and Δ *vapBC*_{ssl2138/sll1092}. Transcriptomic analyses of both deletions strains were subsequently carried out to investigate the cause of the observed phenotypes.

With the exception of VapBC_{ssl2138/sll1092} also affecting the expression of a cohort of hypothetical proteins, the general observation is that both VapBC

Chapter Four

systems appear to target genes involved in carbon assimilation and metabolism via both photosynthetic and heterotrophic processes. Interestingly, VapBC_{slr1209/1210} targets an operon involved in glucose transport, which is also the cellular process targeted by the only other well characterised VapBC system, *M. smegmatis* VapBC_{1283/4}, which targets the cellular machinery responsible for carbon transport and metabolism and functions to regulate the rate of glycerol utilisation to match the anabolic demands. Given the similarity in cellular targets between *M. smegmatis* VapBC_{1283/4} and both *Synechocystis* VapBC systems, in particular VapBC_{slr1209/1210}, it is tempting to speculate that these protein systems share similar biological functions. As a result of this and initial characterisation of VapBC_{slr1209/1210}, it is proposed that VapBC_{slr1209/1210} may also fine-tune the regulation of carbon metabolism by controlling the rate of glucose utilisation (via both carbon fixation and glucose transport) to match the anabolic demands of the cell.

4.3 Further work

4.3.1 *M. smegmatis* AmtR Transcription Factor

Further investigation of the identified co-repressor ligands xanthine and allantoin should be carried out to determine their roles in the regulatory process. Co-crystallisation of AmtR and each of these ligands will provide significant information regarding ligand binding and AmtR:DNA interactions. Subsequently, the X-ray crystallographic structure of the AmtR:DNA complex (most with ligand bound) would be the most informative protein structure for this investigation. Binding affinities of AmtR to each ligand can be used to determine the strength of signalling provided by each ligand and the relevant concentrations that resemble ligand depletion and subsequent AmtR repression to be lifted. Finally, characterisation of the urea metabolic pathway that AmtR regulates and its role in alternative nitrogen regulation should be investigated.

4.3.2 Cyanobacteria VapBC systems

Structural analysis of all possible VapBC structures should be completed to elucidate more information regarding both the VapBC complex properties and the individual protein structural and functional characteristics. Continued protein purification and functional characterisation of the VapBC_{MAE43230/20} system would also be valuable, and any functional characterisation of this system should be

Chapter Four

compared to *Synechocystis* VapBC systems to determine conservation of functions between the organisms.

To overcome current protein expression and solubility problems associated with expression of cyanobacteria derived proteins in the *E. coli* expression system, a *Synechocystis* expression system could be developed. The system could be designed similar to the method adopted to generate markerless knock out strains described in Chapter 3. Homologous recombination via pUC19 constructs could introduce a stronger promoter and a terminal 6 x histidine tag to the *vapBC* operons. The result would allow for *in vivo* over expression of the VapBC proteins that could subsequently be purified by IMAC with the terminal 6 x histidine tag.

Further RT-qPCR should be completed in order to validate more DEG identified in RNA-seq work. Continued characterisation of the VapBC_{slr1209/1210} system should be expanded to include investigations of substrate specificity by *in vitro* RNase activity assays in combination with MALDI_TOF MS of oligonucleotides from the 927 pentaprobe containing the putative central cut sites. Once cut sites have been determined, they can be searched for in the cellular targets identified in the transcriptomic analyses. EMSA should be additionally optimized to identify both the mode of autoregulation and the VapB_{slr1209} recognition motif in the promoter region. Overexpression studies would be valuable to identify if VapC_{slr1210} can cause the characteristic bacteriostatic effect, which can be relieved by addition of VapB_{slr1209}. Finally, characterisation of remaining *Synechocystis* *vapBC* deletion strains would provide a more comprehensive understanding of these systems in cyanobacteria.

References

- Adler, S. P., Purich, D., & Stadtman, E. R. (1975). Cascade control of *Escherichia coli* glutamine synthetase. Properties of the PII regulatory protein and the uridylyltransferase-uridylyl-removing enzyme. *The Journal of Biological Chemistry*, 250(16), 6264–6272.
- Afif, H., Allali, N., Couturier, M., & Van Melderren, L. (2001). The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison-antidote system. *Molecular Microbiology*, 41(1), 73–82.
- Agari, Y., Sakamoto, K., Kuramitsu, S., & Shinkai, A. (2012). Transcriptional repression mediated by a TetR family protein, PfmR, from *Thermus thermophilus* HB8. *Journal of Bacteriology*, 194(17), 4630–4641.
- Amon, J., Bräu, T., Grimrath, A., Hänssler, E., Hasselt, K., Höller, M., et al. (2008). Nitrogen control in *Mycobacterium smegmatis*: nitrogen-dependent expression of ammonium transport and assimilation proteins depends on the OmpR-type regulator GlnR. *Journal of Bacteriology*, 190(21), 7108–7116.
- Amon, J., Titgemeyer, F., & Burkovski, A. (2009). A genomic view on nitrogen metabolism and nitrogen control in mycobacteria. *Journal of Molecular Microbiology and Biotechnology*, 17(1), 20–29.
- Anantharaman, V., & Aravind, L. (2003). New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. *Genome Biology*, 4(12), R81. doi:10.1186/gb-2003-4-12-r81.
- Arcondéguy, T., Jack, R., & Merrick, M. (2001). P(II) signal transduction proteins, pivotal players in microbial nitrogen control. *Microbiology and Molecular Biology Reviews*, 65(1), 80–105.
- Arcus, V. L., Bäckbro, 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), 16471–16478.
- 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 & Selection : PEDS*, 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.
- Atkinson, M. R., & Ninfa, A. J. (1998). Role of the GlnK signal transduction protein in the regulation of nitrogen assimilation in *Escherichia coli*. *Molecular Microbiology*, 29(2), 431–447.
- Austin, S., & Dixon, R. (1992). The prokaryotic enhancer binding protein NTRC has an ATPase activity which is phosphorylation and DNA dependent. *The EMBO Journal*, 11(6), 2219–2228.

References

- Babu, M. M., & Teichmann, S. A. (2003). Functional determinants of transcription factors in *Escherichia coli*: protein families and binding sites. *Trends in Genetics*, 19(2), 75–79.
- Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M., & Teichmann, S. A. (2004). Structure and evolution of transcriptional regulatory networks. *Current Opinion in Structural Biology*, 14(3), 283–291.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819), 1709–1712.
- Bartsevich, V. V., & Pakrasi, H. B. (1996). Manganese transport in the cyanobacterium *Synechocystis* sp. PCC 6803. *The Journal of Biological Chemistry*, 271(42), 26057–26061.
- Bashiri, G., Rehan, A. M., Greenwood, D. R., Dickson, J. M. J., & Baker, E. N. (2010). Metabolic engineering of cofactor F420 production in *Mycobacterium smegmatis*. *PLoS ONE*, 5(12), e15803. doi:10.1371/journal.pone.0015803.
- Beckers, G., Bendt, A. K., Kramer, R., & Burkovski, A. (2004). Molecular identification of the urea uptake system and transcriptional analysis of urea transporter- and urease-encoding genes in *Corynebacterium glutamicum*. *Journal of Bacteriology*, 186(22), 7645–7652.
- Beckers, G., Nolden, L., & Burkovski, A. (2001). Glutamate synthase of *Corynebacterium glutamicum* is not essential for glutamate synthesis and is regulated by the nitrogen status. *Microbiology*, 147(Pt 11), 2961–2970.
- Beckers, G., Strosser, J., Hildebrandt, U., Kalinowski, J., Farwick, M., Kramer, R., & Burkovski, A. (2005). Regulation of AmtR-controlled gene expression in *Corynebacterium glutamicum*: mechanism and characterization of the AmtR regulon. *Molecular Microbiology*, 58(2), 580–595.
- Behrends, V., Williams, K. J., Jenkins, V. A., Robertson, B. D., & Bundy, J. G. (2012). Free glucosylglycerate is a novel marker of nitrogen stress in *Mycobacterium smegmatis*. *Journal of Proteome Research*, 11(7), 3888–3896.
- Bendt, A. K., Beckers, G., Silberbach, M., Wittmann, A., & Burkovski, A. (2004). Utilization of creatinine as an alternative nitrogen source in *Corynebacterium glutamicum*. *Archives of Microbiology*, 181(6), 443–450.
- Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., et al. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research*, 35(6), e45. doi:10.1093/nar/gkm047.
- Bodogai, M., Ferenczi, S., Bashtovyy, D., Miclea, P., Papp, P., & Dusha, I. (2006). The ntrPR operon of *Sinorhizobium meliloti* is organized and functions as a toxin-antitoxin module. *Molecular Plant-Microbe Interactions*, 19(7), 811–822.

References

- Bowden, S. D., Rowley, G., Hinton, J. C. D., Thompson, A. (2009). Glucose and glycolysis are required for successful infection of macrophages and mice by *Salmonella enterica* Serovar Typhimurium. *Infection and Immunity*, 77(7), 3117-3126.
- Briand, J. F., Jacquet, S., Bernard, C., & Humbert, J. F. (2003). Health hazards for terrestrial vertebrates from toxic cyanobacteria in surface water ecosystems. *Veterinary Research*, 34(4), 361–377.
- Brown, J. M., & Shaw, K. J. (2003). A novel family of *Escherichia coli* toxin-antitoxin gene pairs. *Journal of Bacteriology*, 185(22), 6600–6608.
- Brown, D. R., Barton, G., Pan, Z., Buck, M., Wigneshweraraj, S. (2014). Combinational stress response: direct coupling of two major stress responses in *Escherichia coli*. *Microbial Cell*, 1(9): 315-317.
- Buchinger, S., Strösser, J., Rehm, N., Hänsler, E., Hans, S., Bathe, B., et al. (2009). A combination of metabolome and transcriptome analyses reveals new targets of the *Corynebacterium glutamicum* nitrogen regulator AmtR. *Journal of Biotechnology*, 140(1-2), 68–74.
- 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.
- Burkovski, A. (2003). Ammonium assimilation and nitrogen control in *Corynebacterium glutamicum* and its relatives: an example for new regulatory mechanisms in actinomycetes. *FEMS Microbiology Reviews*, 27(5), 617–628.
- 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.
- Chen, Y., Zhu, H., Zheng, G., Jiang, W., & Lu, Y. (2013). Functional analysis of TetR-family regulator AmtRsav in *Streptomyces avermitilis*. *Microbiology*, 159(Pt 12), 2571–2583.
- Chomczynski, P., & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, 162(1), 156–159.
- 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.
- Christensen-Dalsgaard, M., Jørgensen, 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.
- Chubukov, V., Gerosa, L., Kochanowski, K., & Sauer, U. (2014). Coordination of microbial metabolism. *Nature Reviews Microbiology*, 12(5), 327-340.
- Cinege, G., Bodogai, M., Oláh, B., & Kiers, A. (2014). A vapBC-type toxin-

References

- antitoxin module of *Sinorhizobium meliloti* influences symbiotic efficiency and nodule senescence of *Medicago sativa*. *Environmental Microbiology*, 16(12), 3714–3729.
- Claverie-Martin, F., & Magasanik, B. (1991). Role of integration host factor in the regulation of the glnHp2 promoter of *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 88(5), 1631–1635.
- Clissold, P. M., & Ponting, C. P. (2000). PIN domains in nonsense-mediated mRNA decay and RNAi. *Current Biology*, 10(24), R888–890.
- Commichau, F. M., Forchhammer, K., & Stülke, J. (2006). Regulatory links between carbon and nitrogen metabolism. *Current Opinion in Microbiology*, 9(2), 167–172.
- Cook, G. M., Robson, J. R., Frampton, R. A., McKenzie, J., Przybilski, R., Fineran, P. C., & Arcus, V. L. (2013). Ribonucleases in bacterial toxin-antitoxin systems. *Biochimica Et Biophysica Acta*, 1829(6-7), 523–531.
- Cooper, C. R., Daugherty, A. J., Tachdjian, 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(Pt 1), 123–126.
- Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D Biological Crystallography*, 62(Pt 9), 1002–1011.
- Cowtan, K. (2007). Fitting molecular fragments into electron density. *Acta Crystallographica Section D Biological Crystallography*, 64(Pt 1), 83–89.
- 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.
- Das, U., Pogenberg, V., Subhramanyam, U. K. T., Wilmanns, M., Gourinath, S., & Srinivasan, A. (2014). Crystal structure of the VapBC-15 complex from *Mycobacterium tuberculosis* reveals a two-metal ion dependent PIN-domain ribonuclease and a variable mode of toxin-antitoxin assembly. *Journal of Structural Biology*, 188(3), 249–258.
- Delano, W. L. (2002). The PyMOL Molecular Graphics System. <http://www.citeulike.org/group/340/article/1186294>
- Deng, M. D., & Coleman, J. R. (1999). Ethanol synthesis by genetic engineering in cyanobacteria. *Applied and Environmental Microbiology*, 65(2), 523–528.
- Deng, R., & Chow, T.-J. (2010). Hypolipidemic, antioxidant, and antiinflammatory activities of microalgae *Spirulina*. *Cardiovascular Therapeutics*, 28(4), e33–e45. doi:10.1111/j.1755-5922.2010.00200.x.
- Deutscher, J. (2008). The mechanisms of carbon catabolite repression in bacteria. *Current Opinion in Microbiology*, 11(2), 87–93.
- Dismukes, G. C., Klimov, V. V., Baranov, S. V., Kozlov, Y. N., DasGupta, J., &

References

- Tyryshkin, A. (2001). The origin of atmospheric oxygen on Earth: the innovation of oxygenic photosynthesis. *Proceedings of the National Academy of Sciences*, 98(5), 2170–2175.
- Dy, R. L., Przybilski, R., Semeijn, K., Salmond, G. P. C., & Fineran, P. C. (2014). A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. *Nucleic Acids Research*, 42(7), 4590–4605.
- Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography*, 66(Pt 4), 486–501.
- Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D Biological Crystallography*, 69(Pt 7), 1204–1214.
- Falconer, I. R., & Humpage, A. R. (2005). Health risk assessment of cyanobacterial (blue-green algal) toxins in drinking water. *International Journal of Environmental Research and Public Health*, 2(1), 43–50.
- Fatica, A., Tollervey, D., & Dlakić, M. (2004). PIN domain of Nob1p is required for D-site cleavage in 20S pre-rRNA. *RNA*, 10(11), 1698–1701.
- Fineran, P. C., & Charpentier, E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: Acquisition of new information. *Virology*, 434 (2012) 202–209
- 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.
- Fink, D., Weisschuh, N., Reuther, J., Wohlleben, W., & Engels, A. (2002). Two transcriptional regulators GlnR and GlnRII are involved in regulation of nitrogen metabolism in *Streptomyces coelicolor* A3(2). *Molecular Microbiology*, 46(2), 331–347.
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Research*, 42(D1), D222–D230.
- 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.
- Frangoul, L., Quillardet, P., Castets, A.-M., Humbert, J.-F., Matthijs, H. C. P., Cortez, D., et al. (2008). Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium. *BMC Genomics*, 9, 274. doi:10.1186/1471-2164-9-274.
- Gerdes, K. (2012). *Prokaryotic Toxin-Antitoxins*. Springer Verlag: Heidelberg.
- Gerdes, K., Christensen, S. K., & Løbner-Olesen, A. (2005). Prokaryotic toxin–

References

- antitoxin stress response loci. *Nature Reviews Microbiology*, 3(5), 371–382.
- 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.
- Goodey, N. M., & Benkovic, S. J. (2008). Allosteric regulation and catalysis emerge via a common route. *Nature Chemical Biology* 4(8) 474–482.
- Gorke, B. & Stulke J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature Reviews*, 6, 613–624.
- Gottesman, S. (2005). Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends in Genetics*, 21(7), 399–404.
- Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., & Hofacker, I. L. (2008). The Vienna RNA websuite. *Nucleic Acids Research*, 36(Web Server issue), W70–W74.
- Gupta, A., Gandhimathi, A., Sharma, P., & Jayaram, B. (2007). ParDOCK: an all atom energy based Monte Carlo docking protocol for protein-ligand complexes. *Protein and Peptide Letters*, 14(7), 632–646.
- Hammarlöf, D. L., Bergman, J. M., Garmendia, E., & Hughes, D. (2015). Turnover of mRNAs is one of the essential functions of RNase E. *Molecular Microbiology*, 98(1), 34–45.
- 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.
- Herrero, A. (2008). *The Cyanobacteria: Molecular Biology, Genomics and Evolution*. Caister Academic Press: Norfolk, UK.
- Hinrichs, W., Kisker, C., Duvel, M., Muller, A., Tovar, K., Hillen, W., & Saenger, W. (1994). Structure of the Tet Repressor Tetracycline Complex and Regulation of Antibiotic-Resistance. *Science*, 264(5157), 418–420.
- Huffman, J. L., & Brennan, R. G. (2002). Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Current Opinion in Structural Biology*, 12(1) 98–106.
- Humbert, J.F., Barbe, V., Latifi, A., Gugger, M., Calteau, A., Coursin, T., et al. (2013). A tribute to disorder in the genome of the bloom-forming freshwater cyanobacterium *Microcystis aeruginosa*. *PLoS ONE*, 8(8), e70747. doi:10.1371/journal.pone.0070747.
- Inouye, M. (2006). The discovery of mRNA interferases: Implication in bacterial physiology and application to biotechnology. *Journal of Cellular Physiology*, 209(3), 670–676.
- Ishizuka, H., Hanamura, A., Inada, T., & Aiba, H. (1994). Mechanism of the downregulation of cAMP receptor protein by glucose in *Escherichia coli*: role of autoregulation of the *crp* gene. *The EMBO Journal*, 13(13), 3077–3082.

References

- Itzen, A., Blankenfeldt, W., & Goody, R. S. (2011). Adenylylation: renaissance of a forgotten post-translational modification. *Trends in Biochemical Sciences*, 36(4), 221–228.
- Jakoby, M., Nolden, L., Meier-Wagner, J., Kramer, R., & Burkovski, A. (2000). AmtR, a global repressor in the nitrogen regulation system of *Corynebacterium glutamicum*. *Molecular Microbiology*, 37(4), 964–977.
- Jenkins, V. A., Barton, G. R., Robertson, B. D., & Williams, K. J. (2013). Genome wide analysis of the complete GlnR nitrogen-response regulon in *Mycobacterium smegmatis*. *BMC Genomics*, 14, 301. doi:10.1186/1471-2164-14-301
- Jenkins, V. A., Robertson, B. D., & Williams, K. J. (2012). Aspartate D48 is essential for the GlnR-mediated transcriptional response to nitrogen limitation in *Mycobacterium smegmatis*. *FEMS Microbiology Letters*, 330(1), 38–45.
- Jessberger, N., Lu, Y., Amon, J., Titgemeyer, F., Sonnewald, S., Reid, S., & Burkovski, A. (2012). Nitrogen starvation-induced transcriptome alterations and influence of transcription regulator mutants in *Mycobacterium smegmatis*. *BMC Research Notes*, 6(1), 482–482.
- Jiang, Y., Pogliano, J., Helinski, 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.
- 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.
- Kamberov, E. S., Atkinson, M. R., & Ninfa, A. J. (1995). The *Escherichia coli* PII signal transduction protein is activated upon binding 2-ketoglutarate and ATP. *The Journal of Biological Chemistry*, 270(30), 17797–17807.
- Kaneko, T., Nakajima, N., Okamoto, S., Suzuki, I., Tanabe, Y., Tamaoki, M., et al. (2008). Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. *DNA Research*, 14(6), 247–256.
- Kaneko, T., Nakamura, Y., Wolk, C. P., Kuritz, T., Sasamoto, S., Watanabe, A., et al. (2001). Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Research*, 8(5), 205–227–53.
- Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., et al. (1995). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. *DNA Research*, 2(4), 153–166, 191–198.
- Keener, J., & Kustu, S. (1988). Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins NTRB and NTRC of enteric bacteria: roles of the conserved amino-terminal domain of NTRC. *Proceedings of the*

References

- National Academy of Sciences*, 85(14), 4976-4980
- Kenney, L. J. (2002). Structure/function relationships in OmpR and other winged-helix transcription factors. *Current Opinion in Microbiology*, 5(2), 135–141.
- Kern, D., Volkman, B. F., Luginbühl, P., Nohaile, M. J., Kustu, S., & Wemmer, D. E. (1999). Structure of a transiently phosphorylated switch in bacterial signal transduction. *Nature*, 402(6764), 894–898.
- Kisker, C., Hinrichs, W., Tovar, K., Hillen, W., & Saenger, W. (1995). The Complex Formed Between Tet Repressor and Tetracycline-Mg²⁺ Reveals Mechanism of Antibiotic-Resistance. *Journal of Molecular Biology*, 247(2), 260–280.
- Kustu, S., Hirschman, J., Burton, D., Jelesko, J., & Meeks, J. C. (1984). Covalent modification of bacterial glutamine synthetase: physiological significance. *Molecular & General Genetics*, 197(2), 309–317.
- Kwan, A. H. Y., 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), e124–e124.
- 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.
- Leigh, J. A., & Dodsworth, J. A. (2007). Nitrogen regulation in bacteria and archaea. *Annual Review of Microbiology*, 61, 349–377.
- Leplae, R., Geeraerts, D., Hallez, R., Guglielmini, J., Drèze, P., & Van Melderen, L. (2011). Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic Acids Research*, 39(13), 5513–5525.
- Leslie, A. G. W., Powell, H. R. (2007). Processing diffraction data with Mosflm. *Evolving Methods for Macromolecular Crystallography*, 245, 41-51.
- Ling, L. L., Schneider, T., Peoples, A. J., Spoering, A. L., Engels, I., Conlon, B. P., et al. (2015). A new antibiotic kills pathogens without detectable resistance. *Nature*, 517(7535), 454-459.
- Lin, W., et al. (2014). Atypical OmpR/PhoB subfamily response regulator GlnR of Actinomycetes functions as a homodimer, stabilised by the unphosphorylated conserved asp-focused charge interactions. *The Journal of Biological Chemistry*, 289(22), 15413-15425.
- Liu, M. Y., Yang, H., & Romeo, T. (1995). The product of the pleiotropic *Escherichia coli* gene *csrA* modulates glycogen biosynthesis via effects on mRNA stability. *Journal of Bacteriology*, 177(10), 2663–2672.
- Liu, M., Zhang, Y., Inouye, M., & Woychik, N. A. (2008). Bacterial addiction module toxin Doc inhibits translation elongation through its association with the 30S ribosomal subunit. *Proceedings of the National Academy of Sciences*, 105(15), 5885–5890.

References

- Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2009). Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biology Direct*, 4, 19. doi:10.1186/1745-6150-4-19.
- Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X., & Butt, T. R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Science*, 15(1), 182–189.
- Masuda, H., Tan, Q., Awano, N., Wu, K.-P., & Inouye, M. (2012). YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in *Escherichia coli*. *Molecular Microbiology*, 84(5), 979–989.
- 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. *The Journal of Biological Chemistry*, 281(49), 37942–37951.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, 40, 658–674.
- McKenzie, J. L., Duyvestyn, J. M., Smith, T., Bendak, K., MacKay, J., Cursons, R., et al. (2012a). Determination of ribonuclease sequence-specificity using Pentaprobates and mass spectrometry. *RNA*, 18(6), 1267–1278.
- McKenzie, J. L., Robson, J., Berney, M., Smith, T. C., Ruthe, A., Gardner, P. P., et al. (2012b). A VapBC toxin-antitoxin module is a posttranscriptional regulator of metabolic flux in mycobacteria. *Journal of Bacteriology*, 194(9), 2189–2204.
- McLeod, M. P., Warren, R. L., Hsiao, W. W. L., Araki, N., Myhre, M., Fernandes, C., et al. (2006). The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proceedings of the National Academy of Sciences*, 103(42), 15582–15587.
- 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*. *The Journal of Biological Chemistry*, 284(1), 276–283.
- Morett, E., & Segovia, L. (1993). The sigma 54 bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. *Journal of Bacteriology*, 175(19), 6067–6074.
- Muhl, D., Jessberger, N., Hasselt, K., Jardin, C., Sticht, H., & Burkovski, A. (2009). DNA binding by *Corynebacterium glutamicum* TetR-type transcription regulator AmtR. *BMC Molecular Biology*, 10(1), 73. doi:10.1186/1471-2199-10-73.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., et al. (2011). REFMAC5 for the refinement of

References

- macromolecular crystal structures. *Acta Crystallographica Section D Biological Crystallography*, 67(Pt 4), 355–367.
- Murshudov, G. N., Vagin, A. A., & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D Biological Crystallography*, 53(Pt 3), 240–255.
- Mutschler, H., Gebhardt, M., Shoeman, R. L., & Meinhart, A. (2011). A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis. *PLoS Biology*, 9(3), e1001033–e1001033. doi:10.1371/journal.pbio.1001033.
- Niesen, F. H., Berglund, H., & Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*, 2(9), 2212–2221.
- Ning, D., Jiang, Y., Liu, Z., & Xu, Q. (2013a). Characterization of a chromosomal Type II toxin–antitoxin system mazEaFa in the cyanobacterium *Anabaena* sp. PCC 7120. *PLoS ONE*, 8(2), e56035. doi:10.1371/journal.pone.0056035.
- Ning, D., Liu, S., Xu, W., Zhuang, Q., Wen, C., & Tang, X. (2013b). Transcriptional and proteolytic regulation of the toxin-antitoxin locus vapBC10 (ssr2962/slr1767) on the chromosome of *Synechocystis* sp. PCC 6803. *PLoS ONE*, 8(11), e80716. doi.org:10.1371/journal.pone.0080716.
- Nolden, L., Farwick, M., KrÄ mer, R., & Burkovski, A. (2001a). Glutamine synthetases of *Corynebacterium glutamicum*: transcriptional control and regulation of activity. *FEMS Microbiology Letters*, 201(1), 91–98.
- Nolden, L., Ngouoto-Nkili, C. E., Bendt, A. K., Kramer, R., & Burkovski, A. (2001b). Sensing nitrogen limitation in *Corynebacterium glutamicum*: the role of glnK and glnD. *Molecular Microbiology*, 42(5), 1281–1295.
- Orth, P., Cordes, F., Schnappinger, D., Hillen, W., Saenger, W., & Hinrichs, W. (1998). Conformational changes of the Tet repressor induced by tetracycline trapping. *Journal of Molecular Biology*, 279(2), 439–447.
- Orth, P., Schnappinger, D., Hillen, W., Saenger, W., & Hinrichs, W. (2000). Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nature Structural Biology*, 7(3), 215–219.
- Ostensvik, O., Skulberg, O. M., Underdal, B., & Hormazabal, V. (1998). Antibacterial properties of extracts from selected planktonic freshwater cyanobacteria--a comparative study of bacterial bioassays. *Journal of Applied Microbiology*, 84(6), 1117–1124.
- Paerl, H. W., Fulton, R. S., Moisander, P. H., & Dyble, J. (2001). Harmful freshwater algal blooms, with an emphasis on cyanobacteria. *The Scientific World Journal*, 1, 76–113.
- Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S., & Tucker, P. A. (2005). Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallographica Section D Biological Crystallography*, 61(Pt 4), 449–457.

References

- 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.
- Petridis, M. (2015). *Nitrogen metabolism in Mycobacterium smegmatis* (submitted Doctoral thesis, University of Otago, Dunedin, New Zealand).
- Picard, F., Dressaire, C., Girbal, L., & Coccagn-Bousquet, M. (2009). Examination of post-transcriptional regulations in prokaryotes by integrative biology. *Comptes Rendus Biologies*, 332(11), 958–973.
- Postma, P. W., Lengeler, J. W., & Jacobson, G. R. (1993). Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiological Reviews*, 57(3), 543–594.
- Puskás, L. G., Nagy, Z. B., Kelemen, J. Z., Rüberg, S., Bodogai, M., Becker, A., & Dusha, I. (2004). Wide-range transcriptional modulating effect of ntrR under microaerobiosis in Sinorhizobium meliloti. *Molecular Genetics and Genomics*, 272(3), 275–289.
- Quintana, N., Van der Kooy, F., Van de Rhee, M. D., Voshol, G. P., & Verpoorte, R. (2011). Renewable energy from Cyanobacteria: energy production optimization by metabolic pathway engineering. *Applied Microbiology and Biotechnology*, 91(3), 471–490.
- 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), e1000767. doi:10.1371/journal.pgen.1000767.
- Ramos, J. L., Martinez-Bueno, M., Molina-Henares, A. J., Teran, W., Watanabe, K., Zhang, X., et al. (2005). The TetR family of transcriptional repressors. *Microbiology and Molecular Biology Reviews*, 69(2), 326–356.
- Ramseier, T. M. (1995). Cra and the control of carbon flux via metabolic pathways. *Research in Microbiology*, 147(6), 489–493.
- Rittmann, B. E. (2008). Opportunities for renewable bioenergy using microorganisms. *Biotechnology and Bioengineering*, 100(2), 203–212.
- 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(3), 353–367.
- Romeo, T. (1996). Post-transcriptional regulation of bacterial carbohydrate metabolism: Evidence that the gene product CsrA is a global mRNA decay factor. *Research in Microbiology*, 147(6-7), 505–512.
- Saier, M. H. (1998). Multiple mechanisms controlling carbon metabolism in bacteria. *Biotechnology and Bioengineering*, 58(2-3), 170–174.
- Schmetterer, G. R. (1990). Sequence conservation among the glucose transporter from the cyanobacterium Synechocystis sp. PCC 6803 and mammalian

References

- glucose transporters. *Plant Molecular Biology*, 14(5), 697–706.
- Schneider, B. L., Kiupakis, A. K., & Reitzer, L. J. (1998). Arginine catabolism and the arginine succinyltransferase pathway in *Escherichia coli*. *Journal of Bacteriology*, 180(16), 4278–4286.
- Schreier, H. J., Brown, S. W., Hirschi, K. D., Nomellini, J. F., & Sonenshein, A. L. (1989). Regulation of *Bacillus subtilis* glutamine synthetase gene expression by the product of the *glnR* gene. *Journal of Molecular Biology*, 210(1), 51–63.
- Schumacher, M. A., Choi, K. Y., Zalkin, H., & Brennan, R. G. (1994). Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science*, 266(5186), 763–770.
- Schumacher, M. A., Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. A., & Brennan, R. G. (2002). Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. *The EMBO Journal*, 21(5), 1210–1218.
- 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(5912), 396–401.
- Shao, Y., Harrison, E. M., Bi, D., Tai, C., He, X., Ou, H.-Y., et al. (2011). TADB: a web-based resource for Type 2 toxin-antitoxin loci in bacteria and archaea. *Nucleic Acids Research*, 39(Database issue), D606–11. doi:10.1093/nar/gkq908.
- Sharp, J. D., Cruz, J. W., Raman, S., Inouye, M., Husson, R. N., & Woychik, N. A. (2012). Growth and translation inhibition through sequence-specific RNA binding by *Mycobacterium tuberculosis* VapC toxin. *The Journal of Biological Chemistry*, 287(16), 12835–12847.
- Sheldrick, G. M. (2007). A short history of SHELX. *Acta Crystallographica Section A*, 64(Pt 1), 112–122.
- Short, F. L., Pei, X. Y., Blower, T. R., Ong, S.-L., Fineran, P. C., Ben F Luisi, & Salmond, G. P. C. (2013). Selectivity and self-assembly in the control of a bacterial toxin by an antitoxic noncoding RNA pseudoknot. *Proceedings of the National Academy of Sciences*, 110(3), E241–E249.
- Steffensen, D. A. (2008). Economic cost of cyanobacterial blooms. *Advances in Experimental Medicine and Biology*, 619, 855–865.
- Strosser, J., Ludke, A., Schaffer, S., Kramer, R., & Burkovski, A. (2004). Regulation of GlnK activity: modification, membrane sequestration and proteolysis as regulatory principles in the network of nitrogen control in *Corynebacterium glutamicum*. *Molecular Microbiology*, 54(1), 132–147.
- Thajuddin, N., & Subramanian, G. (2010). Cyanobacterial biodiversity and potential applications in biotechnology.
- Tiffert, Y., Supra, P., Wurm, R., Wohlleben, W., Wagner, R., & Reuther, J.

References

- (2008). The *Streptomyces coelicolor* GlnR regulon: identification of new GlnR targets and evidence for a central role of GlnR in nitrogen metabolism in actinomycetes. *Molecular Microbiology*, 67(4), 861–880.
- Van der Westhuizen, A. J., & Eloff, J. N. (1985). Effect of temperature and light on the toxicity and growth of the blue-green alga *Microcystis aeruginosa* (UV-006). *Planta*, 163(1), 55–59.
- Van Melder, L. (2002). Molecular interactions of the CcdB poison with its bacterial target, the DNA gyrase. *International Journal of Medical Microbiology*, 291(6-7), 537–544.
- Vanderpool, C. K., & Gottesman, S. (2004). Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Molecular Microbiology*, 54(4), 1076–1089.
- Vierke, G., Engelmann, A., Hebbeln, C., & Thomm, M. (2003). A novel archaeal transcriptional regulator of heat shock response. *The Journal of Biological Chemistry*, 278(1), 18–26.
- Whitney, P. A., Cooper T. G., & Magasanik. B. (1973). The induction of urea carboxylase and allophanate hydrolase in *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry*, 248(17), 6203–6209.
- 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.
- Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1998). Protein identification and analysis tools in the ExPASy server. *Methods in Molecular Biology*, 112, 531–552.
- Williams, K. J., Bennett, M. H., Barton, G. R., Jenkins, V. A., & Robertson, B. D. (2013). Adenylation of mycobacterial GlnK (PII) protein is induced by nitrogen limitation. *Tuberculosis (Edinburgh, Scotland)*, 93(2), 198–206.
- Wilson, D., Charoensawan, V., Kummerfeld, S. K., & Teichmann, S. A. (2007). DBD taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Research*, 36(Database issue), D88–D92. doi:10.1093/nar/gkm964.
- 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., & Jakob, U. (2000). Heating greatly speeds Coomassie blue staining and destaining. *BioTechniques*, 28(3), 426–432.
- Wozniak, R. A. F., & Waldor, M. K. (2009). A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS Genetics*, 5(3), –e1000439. doi:10.1371/journal.pgen.1000439.

References

- Yamaguchi, Y., Park, J.-H., & Inouye, M. (2011). Toxin-antitoxin systems in Bacteria and Archaea. *Annual Review of Genetics*, 45(1), 61–79.
- Yu, Z., Reichheld, S. E., Savchenko, A., Parkinson, J., & Davidson, A. R. (2010). A comprehensive analysis of structural and sequence conservation in the TetR family transcriptional regulators. *Journal of Molecular Biology*, 400(4), 847–864.
- Zhang, C. C., Durand, M. C., Jeanjean, R., & Joset, F. (1989). Molecular and genetical analysis of the fructose-glucose transport system in the cyanobacterium *Synechocystis* PCC6803. *Molecular Microbiology*, 3(9), 1221–1229.
- Zhang, C. C., Jeanjean, R., & Joset, F. (1998). Obligate phototrophy in cyanobacteria: more than a lack of sugar transport. *FEMS Microbiology Letters*, 161(2), 285–292.
- Zhang, Y. X., Xiao Kui, G., Chuan, W. U., Bo, B. I., Shuang Xi, R., Chun Fu, W. U., & Zhao, G. P. (2004). Characterization of a novel toxin-antitoxin module, VapBC, encoded by *Leptospira interrogans* chromosome. *Cell Research*, 14(3), 208–216.

Appendices

Appendix A: Primers and DNA oligonucleotides

Table A.1. Primers and DNA oligonucleotides used in this study

Primer	Sequence
Chapter Two:	
Cloning primers	
<i>glnK</i> (<i>msmeg_2426</i>) Fwd	GGAATTCCATATGATGAAGCTGATTA
<i>glnK</i> (<i>msmeg_2426</i>) Rev	CCCAAGCTTCTACAGGGCGTCG
T7 Fwd	TAATACGACTCACTATAGGG
T7 Rev	TAGTTATTGCTCAGCGGTGG
EMSA Primers	
<i>msmeg_1283/4</i> Fwd	GAACAGGAGCGGATCGACTG
<i>msmeg_1283/4</i> Rev	CATCAAACACCCGGAAGCCG
<i>msmeg_2184</i> Fwd	GGGCATACCGGCGATTG
<i>msmeg_2184</i> Rev	CGCTGCTGGTCATGGGT
<i>msmeg_2185</i> Fwd	CTGAGCATCGCCGCCAC
<i>msmeg_2185</i> Rev	CAGGTCAGATTCTCGGCG
<i>msmeg_phoH2</i> Fwd	ATGGCTAGCGACCTGCTCTGCTGTC
<i>msmeg_phoH2</i> Rev	TCAGTCACGTGGCGCTCCTAGGGGA
EMSA DNA oligonucleotides sequences	
2184_40 Fwd	GCGCCCATTGTTATCTGTCATCTGACAGGTAAGAGCTGGC
2184_40 Rev	GCCAGCTCTTACCTGTCAGATGACAGATAACAATGGGCGC
2184 Fwd_mut	GCGCCCATTGTTATAAGAAATCTAAAAGGTAAGAGCTGGC
2184 Rev_mut	GCCAGCTCTTACCTTTTAGATTCTTATAACAATGGGCGC
AmtR co-crystallisation DNA oligonucleotides sequences	
AmtR_bind Fwd	TGTTATCTGTCATCTGACAGGTAA
AmtR_bind Rev	CATTACCTGTCAGATGACAGATAA
Chapter Three:	
Cloning primers*	
<i>MAE49650_pET28bpstI</i> Fwd	AACGTCAGATGAGCAGCCTGCGTCTGC
<i>MAE49650_pET28bpstI</i> Rev	GGGAAGCTTACGAATCGGCAGTTTACGA
<i>MAE43220_pET28bpstI</i> Fwd	AACGTCAGATGCCGAATGTTATTGCAG
<i>MAE43220_pET28bpstI</i> Rev	GGGAAGCTTACCTTCGCCTGCCAGTTTC
<i>MAE49640/50 pOPIN</i> Fwd	AAGTTCTGTTTCAGGGCCCGGTTATCAAAGAAGT
<i>MAE49640/50 pOPIN</i> Rev	ATGGTCTAGAAAGCTTTAATTAAGTCCGATTCTG
<i>MAE43230/20 pOPIN</i> Fwd	AAGTTCTGTTTCAGGGCCCGATGTATCAGATTAC
<i>MAE43230/20 pOPIN</i> Rev	ATGGTCTAGAAAGCTTTAATTACCTTCGCCTGCCA
<i>slr1209/10 pET28bpstI</i> Fwd	AACGTCAGATGAATAGGAGACTGGTAGA
<i>slr1209/10 pET28bpstI</i> Rev	CCCAAGCTTACTAGGGTAATACTCTT
<i>ssr2201/slr1327 pET28bpstI</i> Fwd	AACGTCAGATGCACTCATTGTCATCACT
<i>ssr2201/slr1327 pET28bpstI</i> Rev	CCCAAGCTTCTGATTATCCATCATGG
<i>ssl2138/sll1092 pET28bpstI</i> Fwd	AACGTCAGATGACTACTGAAGCTGCCCT
<i>ssl2138/sll1092 pET28bpstI</i> Rev	CCCAAGCTTCAACCACAGGCGTTGAATCG
<i>ssl3615/sll1912 pET28bpstI</i> Fwd	AACGTCAGATGAGCACCATCGAGCAAGT
<i>ssl3615/sll1912 pET28bpstI</i> Rev	CCCAAGCTTATACGTTTCTAGGACGA
<i>ssr2962/slr1767 pET28bpstI</i> Fwd	AACGTCAGATGAATAGCTTACTGTCTCG
<i>ssr2962/slr1767 pET28bpstI</i> Rev	CCCAAGCTTGGAGTTGCTGATTCTATGA
EMSA Primers	
PCC6803 <i>slr1209</i> Fwd	AGTGCTAGAAGCCAGGGCAAAACATTACGG
PCC6803 <i>slr1209T</i> Fwd	AAGGTAATAATTTATTAAGTAGAGGTTA
PCC6803 <i>slr1209</i> Rev	ATAAAAATTCCTCCTAAATAACCCGATCAT
pUC19 mutation construct primers*	
<i>slr1209/10 left</i> Fwd	GATCGTCGACAGCACACCATGGCAATTAT

Appendices

<i>slr1209/10</i> left Rev	GATCGGATCCACTGGCAATGCAATAAC
<i>slr1209/10</i> right Fwd	GATCGAGCTCGATTGAAGCAACCGCAATTT
<i>slr1209/10</i> right Rev	GATCGAATTCTAACTTGCCACCGCTCTGT
<i>ssr2201/slr1327</i> left Fwd	GCATTCTAGATGCCATAGCGGTTGTCATAA
<i>ssr2201/slr1327</i> left Rev	GATCGGATCCGCTCAGGAAATTCCAGTGATG
<i>ssr2201/slr1327</i> right Fwd	GATCGAGCTCTCCTATCAACATTTTCAAGCTA
<i>ssr2201/slr1327</i> right Rev	GATCGAATTCGTCCTCACC GTTTACGATT
<i>ssl2138/sll1092</i> left Fwd	CGGAATTCCTCGGCACCATGAAGGTGAACAA
<i>ssl2138/sll1092</i> left Rev	CGAGCTCTACTCCTCAAATTCTTCTAAGGG
<i>ssl2138/sll1092</i> right Fwd	CGGGATCCTTGAAGCTAGTTTTCAAAGAAC
<i>ssl2138/sll1092</i> right Rev	GCTCTAGAATGGTTACCTTTTTACCTTTCC
<i>ssl3615/sll1912</i> left Fwd	CGGAATTCGGGAGGAACAAGACACCATGGA
<i>ssl3615/sll1912</i> left Rev	CGAGCTCAGTTCCTCACCATTACGCCGTAA
<i>ssl3615/sll1912</i> right Fwd	CGGGATCCTACTGTCTTTAACGACAGAAGA
<i>ssl3615/sll1912</i> right Rev	GCTCTAGATGCTCGCCGGGTAATTGGCCAT
<i>ssr2962/slr1767</i> left Fwd	CGGAATTCGTCTACCTAAAAATACCAATTAC
<i>ssr2962/slr1767</i> left Rev	CGAGCTCGTCCGCCGTGCAGCAAACACTAG
<i>ssr2962/slr1767</i> right Fwd	CGGGATCCTTGCCAACTCTGTAATCATGAC
<i>ssr2962/slr1767</i> right Rev	GCTCTAGACTAATTAGCTTTGCCGAAC TTT
pUC19_M13 Fwd	GTAAAACGACGGCCAGT
pUC19_M13 Rev	CAGGAAACAGCTATGAC

* Sequences provided for primers only where successful cloning/expression was achieved

Table A.2. List of primers used for RT-qPCR

Gene	Primer sequence	Amplicon size
Differentially expressed genes (DEG)		
sll1997_F	CAGAGCCAGACAAGCATGAG	211
sll1997_R	TGCTGACAATGGTTGAATCG	
sll1255_F	TGCCAAAAATGAGAATGCAA	178
sll1255_F	TGCTAACTCCCTTAGCCAGTG	
slr0350_F	CCTAAATGGCGAACGGTCTA	182
slr0350_R	TAGCCAGTTTCTTGGCCAGT	
slr1209_F	TTATTGCATTGCCAGTGGAG	125
slr1209_R	CCAGACTGCAATTCGGGTAT	
sll0315_F	GAGACGCTAAACGCCACAGT	100
sll0315_R	TGAGCATGACTTTAGTGCCTTG	
sll0771_F	GCCATTGTTTCTGGCATT TTT	168
sll0771_R	GATCAGGAAAGCGCAA ACTC	
sll0772_F	AGTCTGTTGGCCATTCAACC	107
sll0772_R	CAGCCATGGATTTGTACGTG	
House keeping genes (HKG)		
rrn 16s_F	CACACTGGGACTGAGACAC	190
rrn 16s_R	CTGCTGGCACGGAGTTAG	
rnpB_F	CGTTAGGATAGTGCCACAG	189
rnpB_R	CGCTCTTACCGCACCTTTG	

Appendix B: Plasmids and Bacterial Strains

Table A.3. Plasmids and bacterial strains used in this study

Bacterial Strain	Description
<i>Escherichia coli</i>	
DH10B (TOP10)	F-mcrA Δ(mrr-hsdRMS-mcrBC) Φ80 dlacZ ΔM15 ΔlacX74 deoR recA1 araD139 Δ(ara leu) 7697 galU galK rpsL endA1 nupG λ-
DH5	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17
BL21 (DE3)	F-ompT hsdSB (rB-mB-) gal dcm (DE3)
<i>Mycobacterium smegmatis</i>	
Mc ² 4517	<i>M. smegmatis</i> expression strain with T7 RNA polymerase; Km ^r
Cyanobacteria	
<i>Synechocystis</i> sp. Strain PCC6803 (PCC6803)	<i>Synechocystis</i> sp. Strain PCC6803 wild type strain
Plasmids*	
pYUB28b	<i>E. coli</i> mycobacterium shuttle vector with T7 promoter, MCS pET28b, encoding both C-and N-terminal His tags; Hyg ^r
pET28b-PstI	Modified pET28b where the NcoI site has been replaced with PstI site; C- or N-terminal His-tags; Km ^r
pUC19	<i>E. coli</i> cloning vector; Amp ^r resistance
pOPINS	<i>E. coli</i> expression vector with In-Fusion enabled pTriEx2 backbone, lacZ insert and N-His-SUMO tag; Amp ^r

* Information provided for plasmids only where successful cloning/expression was achieved

Hyg^r = hygromycin B resistance,

Km^r = Kanamycin resistance

Amp^r = Ampicillin resistance

Determination of *M. aeruginosa* Strain for Experimental Analysis

Minimal investigation has been done on cyanobacterial VapBC systems. In fact, no investigation has been done on the toxic-bloom forming members such as *M. aeruginosa*. Consequently, the first obstacle for this project was to determine the appropriate strain of *M. aeruginosa* to implement for experimental analysis. *M. aeruginosa* NIES-843 (MIC_NIES_843) was chosen because it is the only fully assembled and annotated *M. aeruginosa* genome available. The MIC_NIES_843 genome is highly plastid with 11.8 % composed of insertion sequences and miniature inverted-repeat transposable elements (Kaneko et al., 2008). The plasticity of the genome is linked to the ability of *M. aeruginosa* to bloom in a wide range of ecosystems, as a result of integration and removal of genetic elements to provide an adaptive advantage (Humbert et al., 2013). Although MIC_NIES_843 is naturally transformable there has been minimal genetic manipulation and protein expression done in the scope of this organism. It is also incredibly slow growing, with a doubling time of 1.23 days under optimal

Appendices

conditions (van der Westhuizen & Eloff, 1985) making it an inefficient laboratory organism. This strain is not available for experimental work in New Zealand and purchasing from overseas is costly and requires significant permit and ethical approval. Because of the obstacles associated with obtaining and experimenting with MIC_NIES_843, we decided to investigate their VapBC systems using *E. coli* as the host organism, a standard procedure adopted when working with difficult organisms (Bunker, McKenzie, Baker, & Arcus, 2008; Das et al., 2014; Mattison et al., 2006; Ning, Jiang, Liu, & Xu, 2013a; Sharp et al., 2012)

Appendix C: Buffers, Media and Solutions

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

Table A.4. General buffers and solutions used in this study

Reagent	Description
5 x DNA loading dye	0.05 % bromophenol blue, 0.25 % xylene cyanol, 30 % (v/v) glycerol
Fairbanks Staining solution A	0.05 % coomassie blue R-250, 25 % (v/v) isopropanol, 10 % (v/v) acetic acid
Fairbanks Staining solution D	10% (v/v) acetic acid
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 ₂ O
Native Gel Loading buffer	0.3 M Tris-HCl, 1 % bromophenol blue, pH 6.8
Resolving buffer	1.5 M Tris-HCl, pH 8.8
4 x SDS loading buffer	200 mM Tris-HCl pH 6.8, 8 % SDS, 40 % (v/v) glycerol, 0.4 % bromophenol blue, 400 mM 2-mercaptoethanol
SDS running buffer	25 mM Tris-HCl pH 6.8, 0.1 % SDS 190 mM glycine
Stacking buffer	0.5 M Tris-HCl, pH 6.8
10 x TAE	400 mM Tris-acetate, 20 mM EDTA
1 x TAE	100 ml 10 x TAE + 900 ml H ₂ O
10 x TBE	0.89 M Tris-HCl, 0.89 boric acid, 20 mM EDTA
1 x TBE	100 ml 10 x TBE + 900 ml H ₂ O
TE	10 mM Tris-HCl
Tris-EDTA-SDS lysis solution	0.1 M Tris, 0.05 M EDTA, 1 % SDS, 0.5 M NaCl
Tris-Glycine Running buffer	25 mM Tris-HCl, 250 mM glycine, pH 8.5

Appendices

Table A.5. Media used in this study.

Media	Description
BG-11 agar (1 L)	15 g agar, 3.0 g sodium thiosulphate, 10 ml 100 x BG-11 ^a , 1 ml Trace elements ^b , 1 ml Iron stock ^c , 1 ml Phosphate stock ^d , 1 ml Na ₂ CO ₃ stock ^e , 10 ml TES buffer ^f , 10 ml 1M NaHCO ₃
BG-11 agar + Sucrose (1 L)	15 g agar, 3.0 g sodium thiosulphate, 10 ml 100 x BG-11 ^a , 1 ml Trace elements ^b , 1 ml Iron stock ^c , 1 ml Phosphate stock ^d , 1 ml Na ₂ CO ₃ stock ^e , 10 ml TES buffer ^f , 50 ml 50% sucrose
BG-11 (1 L)	10 ml 100 x BG-11 ^a , 1 ml Trace elements ^b , 1 ml Iron stock ^c , 1 ml Phosphate stock ^d , 10 ml 1M NaHCO ₃
7H9	0.47 g 7H9 powder, 0.2 % (v/v) glycerol in 90 ml H ₂ O. autoclave then add 10 ml ADC enrichment and 0.05 % (v/v) Tween 80
7H10-agar	1.9 g 7H10 powder, 0.5 % (v/v) glycerol in 90 ml H ₂ O, autoclave then add 10 ml ADC enrichment and 0.05 % (v/v) Tween 80
LB-agar	1 % bactotryptone, 0.5 % yeast extract, 1 % NaCl, 15 g/L agar, pH 7.5
LB	1 % bactotryptone, 0.5 % yeast extract, pH 8.0
LBT-agar	1 % bactotryptone, 0.5 % yeast extract, 1 % NaCl, 15 g/L agar, 0.05 % (v/v) Tween 80, pH 7.5
LBT	1 % bactotryptone, 0.5 % yeast extract, 1 % NaCl, 0.05 % (v/v) Tween 80, pH 7.5
Low salt LB-agar	1 % bactotryptone, 0.5 % yeast extract, 0.5 % NaCl, 15 g/L agar, pH 7.5
Modified Sauton's	0.5 g KH ₂ PO ₄ , 0.5 g MgSO ₄ , 0.16 mM FeCl ₂ , 0.48 g citric acid, 0-30 mM NH ₄ Cl, 0.1 ml 1% ZnSO ₄ , 60 ml glycerol per litre, 5 ml 10 % Tween 80, pH 7.2
PA-0.5G	50 mM Na ₂ HPO ₄ , 50 mM KH ₂ PO ₄ , 25 mM (NH ₄) ₂ SO ₄ , 1 mM MgSO ₄ , 0.5 % glucose, 0.1 x metals mix*, 100 µg/ml each of 17 a.a (no Cys, Tyr or Met). Individual components autoclaved to sterilise
SOC	2 % bactotryptone or bactopectone, 0.55 % yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl ₂ , 10 mM MgSO ₄ , 20 mM glucose
ZYP-5052	1 % bactotryptone, 0.5 % yeast extract, 50 mM Na ₂ HPO ₄ , 50 mM KH ₂ PO ₄ , 25 mM (NH ₄) ₂ SO ₄ , 1 mM MgSO ₄ , 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose, 1 x metals mix*

^a 100 x BG-11 made by adding each of the following to 1 L of water: 149.6 g NaNO₃, 7.49 g MgSO₄.7H₂O, 3.60 g CaCl₂, 0.60 g citric acid, 1.12 ml of 0.25 M Na₂EDTA pH8.

^b Trace elements made by adding each of the following to 100 ml of water and autoclaving: 286 mg H₃BO₃, 181 mg MnCl₂.4H₂O, 22 mg ZnSO₄.7H₂O, 39 mg Na₂MoO₄.2H₂O, 8 mg CuSO₄.5H₂O, 5 mg Co(NO₃)₂.6H₂O.

^c Iron Stock made by adding 0.6 g ferric citrate to 100 ml of water and autoclaving.

^d Phosphate stock made by adding 3.05g K₂HPO₄ to 100 ml water and autoclaving.

^e Na₂CO₃ stock made by adding 2 g of Na₂CO₃ to 100 ml water and autoclaving.

^f TES buffer made by adding 22.9 g to 100 ml water, pH 8.2, filter sterilised.

*1000 x metals mix made up from the sterile stock solutions of each component to give the following concentrations: 50 µM FeCl₃ in 0.12 M HCl (filter sterile), 20 µM CaCl₂, 10 µM MnCl₂, 10 µM ZnSO₄, 2 µM CoCl₂, 2 µM CuCl₂, 2 µM NiCl₂, 2 µM NaMoO₄, 2 µM Na₂SeO₃, 2 µM H₃BO₃.

Appendix D: Gene and Protein Information

M. smegmatis *amtR* and *glnK* Information

M. smegmatis *amtR* (msmeg_4300)

Modified to include NcoI restriction site at the N-terminus and the stop codon removed to allow for a C-terminal His tag.

717 bp

```
ATGGCGACACCTCCGGCCGTGGCCGCCCCGCGGCTGGAGCAGCCGCGCAGGCCGGGG
CAGACCGCGCGCAGGAGATCCTCGACGCGGCCGCCGAACGTTCACCACCCACGGG
TACGGCAGCACCTCCACGCGCCGCATCGCCGACGAGGTGGGGGTGCGGCAGGCCTCG
CTGTATCACCACCTTCGCCACCAAGGACGACATCCTCGACGCGCTGCTGGCGGGCACC
TCGACGAGCCCCCTTGAGCTCGCACACGGCCTGCTGGGGGAGTCGGGGCCTGCGGCAC
CGCGCCTGCATGCCCTGGTGATCTACGACGCCTCACAGCTGTGCGCGGGCCGCTGGAA
CCTGGGCGCGCTGTACCTGCTGCCCCGAGTTGCGCACCGACCGGTTGCGCCGTTCCGC
CGCCGCCGCGCCGAACGTGCGCAGCGCATACCGCAGCCTGGCCGCTGCGGTTCATCGCC
GAGTGCGGCGGACCGCCCCGAGGCCGACGATCTGCCGTTCCGGCTCGTGGAGTCGGTG
ATCAACTCACGCTCCGACGACGCCGTCGTCCCTCCCGAACCAACCCTGGGTGATCGGAG
AGGGTGCGCTGCGCGTGCTCGGCTTCGACGGGGACTTTGCCGAACTCGCCGCCGCAAC
CGCGTCACGGCTCGGTGTGCGACCGCCCCGGGCGGGCCGCCCGCAAGCTTGCGGCCGC
ACTCGAGCACCACCACCACCACCACTGA
```

M. smegmatis *AmtR* with C-terminal His tag

238 a.a 25.82 kDa pI 6.92

```
MATTSGRGRPRLEQPRRPGQTAREEILDAAAELFTTHGYGSTSTRRIADEVGVRQASLYH
HFATKDDILDALLAGTVDEPLELAHGLLGESGPAAPRLHALVIYDASQLCAGRWNLGAL
YLLPELRTDRFAPFRRRRAELRSAYRSLAAAVIAECGGPPEADDLPFRLVESVINSRSDA
VVPPEQPWVIGEGALRVLGFDGDFAEALAAATASRLGVRPPGRAARKLAAALEHHHHHHH
```

M. smegmatis *glnK* (msmeg_2426) with N-terminal His tag

402 bp

```
ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGC
CATATGATGAAGCTGATTACTGCGATCGTCAAGCCGTTACGCTGGAAGATGTCAAGA
CGGGCCTGGAGCAGACGGGCATCTTGGGCATGACCGTCAGCGAGGTTTCAGGGCTACG
GGCGCCAGAAGGGCCATACTGAGGTGTACCGCGGTGCCGAGTACTCGGTGGACTTCG
TGCCCAAGGTCCGGGTGGAGGTCGTCTGACGACTCCGCCGTCGACAAGGTCGTCTG
ACGTCATCGTCCAGGCTGCGCGCACCGGGAAGATCGGTGACGGCAAGGTGTGGGTGA
GCCCCGTGGAAACCGTGGTCCGTGTTTCGACCGGGGAGCGGGGAGCCGACGCCCTGT
AG
```

M. smegmatis *GlnK* with N-terminal His tag

133 a.a 14.41 kDa pI 8.01

```
MGSSHHHHHHSSGLVPRGSHMMKLITAIKPFPTLEDVKTGLEQTGILGMTVSEVQGYGR
QKGHTEVYRGAEYSVDFVPKVRVEVVDDSAVDKVVVDVIVQAARTGKIGDGKVVWVSPV
ETVVRVRTGERGADAL
```

Appendices

Cyanobacteria VapBC operon information

M. aeruginosa VapBC systems

***MAE49640/50* with C-terminal His tag (644 bp)**

GTTATCAAAGAACTGGATCGTATTACCGTTAATCCGCAGATTTGTCTGGGTCAGCCGA
CCATTCGTGGTATGCGCATTACCGTTGTTTTGTTCTGAAACTGCTGGCAAGCCAGCTG
AGCATTCAAGAAGTTCTGGAAGCATATCCGGAAGTGAAGAAGAAGATATTCGTCAG
GCACTGAATTATGCAGCATGGGCAGTTAGCGATTATATTGTTAGCTTTACCAGCGC*A
TGA*GCAGCCTGCGTCTGCGTGCAGATGTTTCATATTAGTCCGCTGACCGTTGCCGCACT
GCGCCTGCAAGGTTATGATATTGTTTCGTACCACCGATTTTCTGCCTGCAACCGCAGCA
GATGCAGAAATTCTGGAAGTGGCACGTGTTGAAGGTCAGGTTATTCTGACCCAGGATC
TGGATTTTAGCCTGCTGGTTGCACTGAATAACTATGGTCTGCCGAGCCTGATTACCCT
GCGCCTGAGCAGCGCACGTCCGGATGTTGTTGCACAGCGTCTGCTGGATGTTCTGCTG
ACCGTGGAACCGAAGTGCAGCAATTACCATTAGTGATGATAGCGTT
CGTGTTCTGTAAGCTTGCAGGCTTGCAGGCTGAGCACCACCACCACC
ACCACTGA

*Underlined bases show overlapping region between *MAE49640* stop codon and *MAE49650* start codon.

MAE49640

77 a.a 8.68 kDa 4.25 kDa

VIKELDRITVNPQICLGQPTIRGMRITVVFLKLLASQLSIQEVLEAYPELEEEDIRQALNYA
AWAVSDYIVSFTSA

***MAE49650* with C-terminal His tag**

137 a.a 15.07 kDa 6.39 kDa

MSSLRLRADVHISPLTVAALRLQGYDIVRTTDFLPATAADAEILELARVEGQVILTQDLDF
SLLVALNNYGLPSLITLRLSSARPDVVAQRLLDVLLTVETELTEGAAITISDDSVRVRKLPI
RKLAALAEHHHHHH

***MAE43230/20* with C-terminal His tag (776 bp)**

ATGTATCAGATTACCCTGAGCCTGCCGGATGAAACCGCACTGGCACTGCATCTGACAC
CGGAACAGCTGGCACAAGAAATTAGCCTGCTGGCAGCAATTAACTGTATGAACTGG
GTAAACTGAGCAGCGGCACCGCAGCAAATCTGGCAGGTATTCCGCGTGTGTTTTCT
GAGCAAAGTGGCCGATTATGGTGTGATACCTTTCGTCTGAAAGAAGCCGAAGTATT
GAAGATCTGGCAAATGCA*TAATG*CCGAATGTTATTGCAGATACCAGCCCGATTGAT
TATCTGTATCAGACCAATCTGCTGGATCTGCTGCCGAAGTTTATGGTCAGATTATTAT
CCCGTTTAGCGTTGCACAAGAACTGGCAGTTGGTAAAGCAAGCGGTATTAGTCTGCCG
GATATTACCAGCCTGAGCTGGATTATCATTGAGCAGGCAAAAAGCGTGAGCCTGCTGC
CGCTGGTTACCGATCTGGGTAAAGGTGAAAAAGAAGTTCTGGCACTGGCAATTGAAA
TCACCGATTTTCTGGCCCTGCTGGATGATGGTCTGGCACGTCGTTATGCAAACCTGCT
GGGTATTAACTGACCGGCACCATGGGTATTCTGCTGAAAGCAAAACAGAATGGTTA
TCTGCATCGTATTGAACCGATTCTGAATCAACTGGAAGTCTGAAATTTCTGCTGGAT
GCAACCAACCGTACCAGCGTTCTGAAACTGGCAGGCGAAGGTAAGCTTGCAGGCGCA
CTCGAGCACCACCACCACCACCCTGA

*Underlined bases show overlapping region between *MAE43230* stop codon and *MAE43220* start codon.

Appendices

MAE43230

83 a.a 8.97 kDa pI 4.30

MYQITLSLPDETALALHLTPEQLAQEISLLAAIKLYELGKLSSGTAANLAGIPRVVFLSKLA
DYGVDTFRLKEAELIEDLANA

MAE43220 with C-terminal His tag

174 a.a 19.09 kDa pI 7.26

MPNVIADTSPIQYLYQTNLLDLLPNFYGQIIIPFSVAQELAVGKASGISLPDITSLSWIIQQA
KSVSLLPLVTDLGKGEKEVLALAIEITDFLALLDDGLARRYANLLGIKLTGTMGILLKAKQ
NGYLHRIEPILNQLELLKFRLDATTRTSVLKLAGEGKLAAALEHHHHHH

MAE43230/20 with N-terminal His tag and SUMO fusion protein (1088 bp)

ATGGCACACCATCACCACCATCACGGGAGCGATAGCGAAGTGAACCAGGAAGCGAA
ACCGGAAGTTAAACCGGAAGTGAAACCGGAAACCCATATTAATCTGAAAGTTAGCGA
CGGCAGCAGCGAAATCTTTTTTAAATTAATAAACCACCCGCTGCGTCGCCTGATG
GAAGCCTTTGCGAAACGTCAGGGTAAAGAAATGGATAGCCTGCGCTTTCTGTATGAC
GGCATCCGTATTCAGGCCGATCAGACCCCGGAAGACCTGGATATGGAAGACAACGAT
ATTATTGAAGCGCATCGCGAACAGATCAGCAGCGGTCTGGAAGTTCTGTTTCAGGGCC
CGATGTATCAGATTACCCTGAGCCTGCCGGATGAAACCGCACTGGCACTGCATCTGAC
ACCGGAACAGCTGGCACAAGAAATTAGCCTGCTGGCAGCAATTAAACTGTATGAACT
GGGTAAACTGAGCAGCGGCACCGCAGCAAATCTGGCAGGTATTCCGCGTGTTGTTTT
CTGAGCAAACCTGGCCGATTATGGTGTGATACCTTTCGTCTGAAAGAAGCCGAACTGA
TTGAAGATCTGGCAAATGCA*TAATG*CCGAATGTTATTGCAGATACCAGCCCGATTC
AGTATCTGTATCAGACCAATCTGCTGGATCTGCTGCCGAACTTTTATGGTCAGATTATT
ATCCCGTTTAGCGTTGCACAAGAACTGGCAGTTGGTAAAGCAAGCGGTATTAGTCTGC
CGGATATTACCAGCCTGAGCTGGATTATCATTACAGCAGGCAAAAAGCGTGAGCCTGCT
GCCGCTGGTTACCGATCTGGGTAAAGGTGAAAAAGAAGTTCTGGCACTGGCAATTGA
AATCACCGATTTTCTGGCCCTGCTGGATGATGGTCTGGCACGTCGTTATGCAAACCTG
CTGGGTATTAACTGACCGGCACCATGGGTATTCTGCTGAAAGCAAAACAGAATGGT
TATCTGCATCGTATTGAACCGATTCTGAATCAACTGGAAGTCTGAAATTTCTGCTG
ATGCAACCACCCGTACCAGCGTTCTGAAACTGGCAGGCGAAGGTTAAATTA

*Underlined bases show overlapping region between *MAE43230* stop codon and *MAE43220* start codon.

MAE43230 with N-terminal His tag and SUMO fusion protein

198 a.a 22.17 kDa pI 4.99

MAHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTPLRRLMEAF
KRQKEMDSLRLFLYDGIRIQADQTPEDLDMEDNDIIEAHREQISSGLEVLFGQPMYQITLS
LPDETALALHLTPEQLAQEISLLAAIKLYELGKLSSGTAANLAGIPRVVFLSKLADYGVDT
FRLKEAELIEDLANA

MAE43220

162 a.a 17.69 kDa pI 5.89

MPNVIADTSPIQYLYQTNLLDLLPNFYGQIIIPFSVAQELAVGKASGISLPDITSLSWIIQQA
KSVSLLPLVTDLGKGEKEVLALAIEITDFLALLDDGLARRYANLLGIKLTGTMGILLKAKQ
NGYLHRIEPILNQLELLKFRLDATTRTSVLKLAGEGN

Appendices

PCC6803 VapBC systems

***slr1209/sl1210* with C-terminal His tag (751 bp)**

ATGAATAGGAGACTGGTAGAGTCTTTGGCAGAAGCAGTTATTGCATTGCCAGTGGAG
GATTATAGTTTATTTCAAAATACCCTCATGGCAAAGATGATCAAAAAACACCAGGG
GTATGCGGTGGCCACGCTTGTATTCTGGGATACCCGAATTGCAGTCTGGACAATTATTT
CTCTTATCAATCAAGGAGGTACGGATTTGGAGTTACTGGCCGATTTTCCTGGATTAAC
TGCCTTTGATTTGTTGACCATCAAAGAATATTACCAATCTCATCAGCAGGAAATTGAT
AATCTTATTGATTACCAAGATCAAGAAAATTATC*ATGATTAA*ATTCTATAGCAACG
AAAATTTATCCCTGGGCTTAGTTAATCAACTACGTCAATTAGGTTACAGTGTTTTGACT
TCCTATGATGCAGGGAGAGCCAACCAAAGAATTACTGATTTATCTGTTTTGAAAGATG
CAACAACCTGATGGACGTTGTGTAATTACATTTAATCGAAATGATTTTATTGAATTACA
TTTGAATAGAATAAAACACAAAGGCATAATAATTTGTAAAGAAGATCGAGATCAGAT
TGGACAAGCATCAATAATTCATGATTTTTTAATTAATCAGAAAACCTTTAGATAATCGT
TTAATCAGAATATTAACCAAAATCAACCTGGATTGAAGCAACCGCAATTTACGATTA
AAGAGTATTACCCTAGTAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCTG
A

*Underlined bases show overlapping region between *slr1209* stop codon and *slr1210* start codon.

slr1209

109 a.a 12.37 kDa pI 4.35

MNRRLVESLAEAVIALPVEDYSLFQNTLMAKMIKKTPGVCGGHACIRDTRIAVWTIISLIN
QGGTDLELLADFPGLTAFDLLTIKEYYQSHQEQEIDNLIDYQDQENYHD

***slr1210* with C-terminal His tag**

142 a.a 16.46 kDa pI 9.71

MIKFYSNENLSLGLVNQLRQLGYSVLTSYDAGRANQRITDLSVLKDATTDGRCVITFNRN
DFIELHLNRIKHKGIIICKEDRDQIGQASIIHDFLINQKTLNRLIRILKQNPGLKQPQFTIKE
YYPSKLAAALEHHHHHHH

Appendices

Pentaprobe Sequence + Flanking Vector Sequence

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

922 + Flanking sequences

AGACCCAAGCTTGGTACCGGAATTCTACGAATTTTCTTTGTTTATTTCCCTTTCGCTTT
GCTTCTCTTCCCTTCGGTTCTGTTCCGTTTTACCTTGTCTTGCCTTATCTTACTTTATCTA
GAGGGCCCTATTCTATAGTGTACCTAAATGCTAGAGCTCGCT

923 + Flanking sequences

AGACCCAAGCTTGGTACTATCTTACTTTAGTTTCATTTAATTGTGTTGTACTCTCCTCT
GCGTTCACCTAGCTTAACCTTGGTTTGGCTTGATTGACTTCAGTTGCGCTCTATTCTAT
CTAGAGGGCCCTATTCTATAGTGTACCTAAATGCTAGAGCTCGCT

924 + Flanking sequences

AGACCCAAGCTTGGTACCGCTCTATTCTACTGTCCTGTGCATTCAATCGTTGAGTTCTGA
TCTAGTCTCGTCTAACCCTCCCCTGCTCCGCTGGTCTGGCCTCGCCTATCCTACCCATT
CTAGAGGGCCCTATTCTATAGTGTACCTAAATGCTAGAGCTCGCT

925 + Flanking sequences

AGACCCAAGCTTGGTACTATCCTACCCATTGGGCTCATCTGATCCATCCGGTCCCGTC
CACTCGGCTATGTTATGCTGTATTGCAGTCGTGTCGCGTCGAGCTGCCCTAATCCCACC
TCTAGAGGGCCCTATTCTATAGTGTACCTAAATGCTAGAGCTCGCT

926 + Flanking sequences

AGACCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGA
ATTCTGCAGATCCTAATCCCACCTAGCGTATCGGGTCATGTAGTGCTACGTTACGGCC
CCCGCCCGGCATCATATTATATCACCCCAGTGTAATGTGGTGTGAGGTTGGAGCATCA
CACTGGCGGCCGCTCGAGCATGCATCTAGAGGGCCCTATTCTATAGTGTACCTAAAT
GCTAGAGCTCGCT

927 + Flanking sequences

AGACCCAAGCTTGGTACGTGAGGTTGGAGTCCGACCTGGAATCTCAGCCTGACGTGCC
ATGCGGTGCGATGTACGCGCGCCACGGTATAGTATGGTACGGGATCCCGTCTAGA
GGGCCCTATTCTATAGTGTACCTAAATGCTAGAGCTCGCT

932 + Flanking sequences (932 = reverse complement to 926)

AGACCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGA
ATTCTGCAGATCTCCAACCTCACACCACATTACACTGGGGTGATATAATATGATGCCG
GGCGGGGGCCGTAAACGTAGCACTACATGACCCGATACGCTAGGTGGGATTAGGCATC
ACACTGGCGGCCGCTCGAGCATGCATCTAGAGGGCCCTATTCTATAGTGTACCTAAA
TGCTAGAGCTCGCT

Appendix E: Chapter 3 Additional Tables and Figures

Table A.6. *M. aeruginosa* *vapBC* operons and genes successfully cloned for protein expression and predicted protein molecular weight (MW).

Vector	His tag position	Operon/Gene	Expected MW (kDa)		
			VapB	VapC	VapBC
pET28b-pstI	C	<i>MAE_49640/50</i>	8.68	15.07	23.75
		<i>MAE_43230/20</i>	8.97	19.09	28.06
		<i>MAE_49650</i>		15.07	
		<i>MAE_43220</i>		19.09	
pPROEX	N	<i>MAE_49640/50</i>	12.21	13.55	25.76
pOPINF	N	<i>MAE_43230/20</i>	11.11	17.57	28.68
pOPINM	N	<i>MAE_43230/20</i>	51.67	13.55	65.22
pOPINS	N	<i>MAE_43230/20</i>	22.17	13.55	35.72

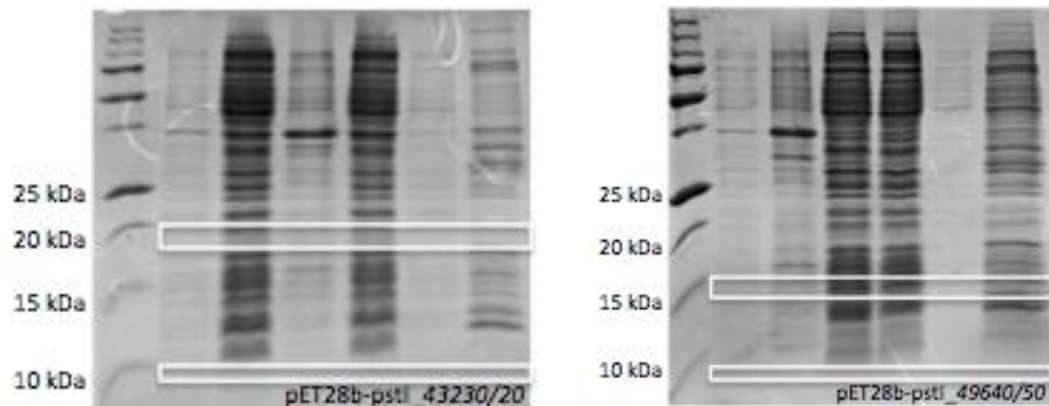


Figure A.1. SDS-PAGE analysis of *E. coli*_pET28b-pstI+*MICvapBC* small-scale expression trials. White boxes indicate expected band size. Labels: whole cell fraction (WC), soluble fraction (S), insoluble fraction (I), flow through (FT), wash (W) and nickel beads with protein bound after wash steps (B).

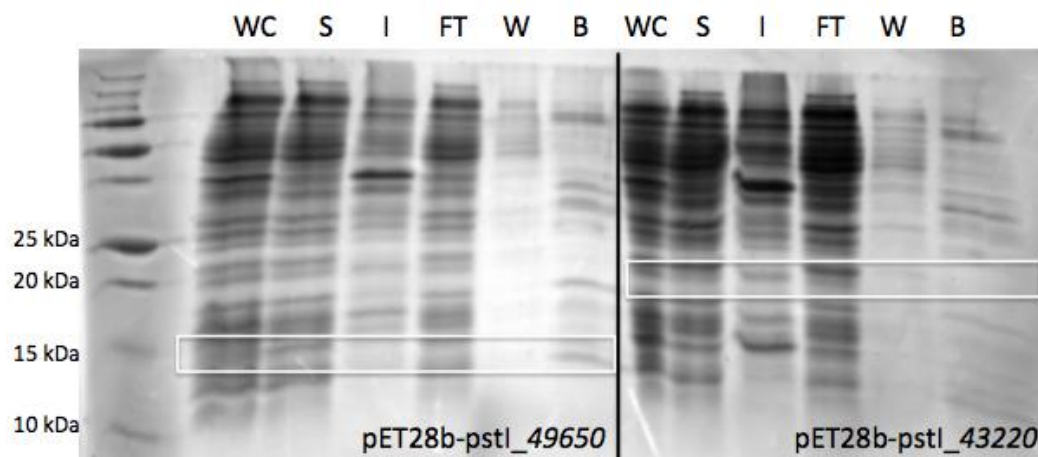


Figure A.2. SDS-PAGE analysis of *E. coli* pET28b-pstI+vapC small scale expression trials. White boxes indicate expected band size. Labels: whole cell fraction (WC), soluble fraction (S), insoluble fraction (I), flow through (FT), wash (W) and nickel beads with protein bound after wash steps (B).

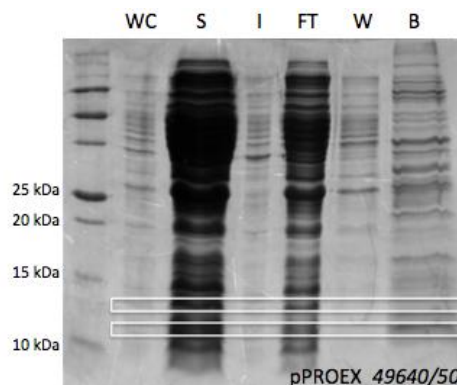


Figure A.3. SDS-PAGE analysis of *E. coli* pPROEX+vapBC_{49640/50} small scale expression trial. White boxes indicate expected band size. Labels: whole cell fraction (WC), soluble fraction (S), insoluble fraction (I), flow through (FT), wash (W) and nickel beads with protein bound after wash steps (B).

Appendices

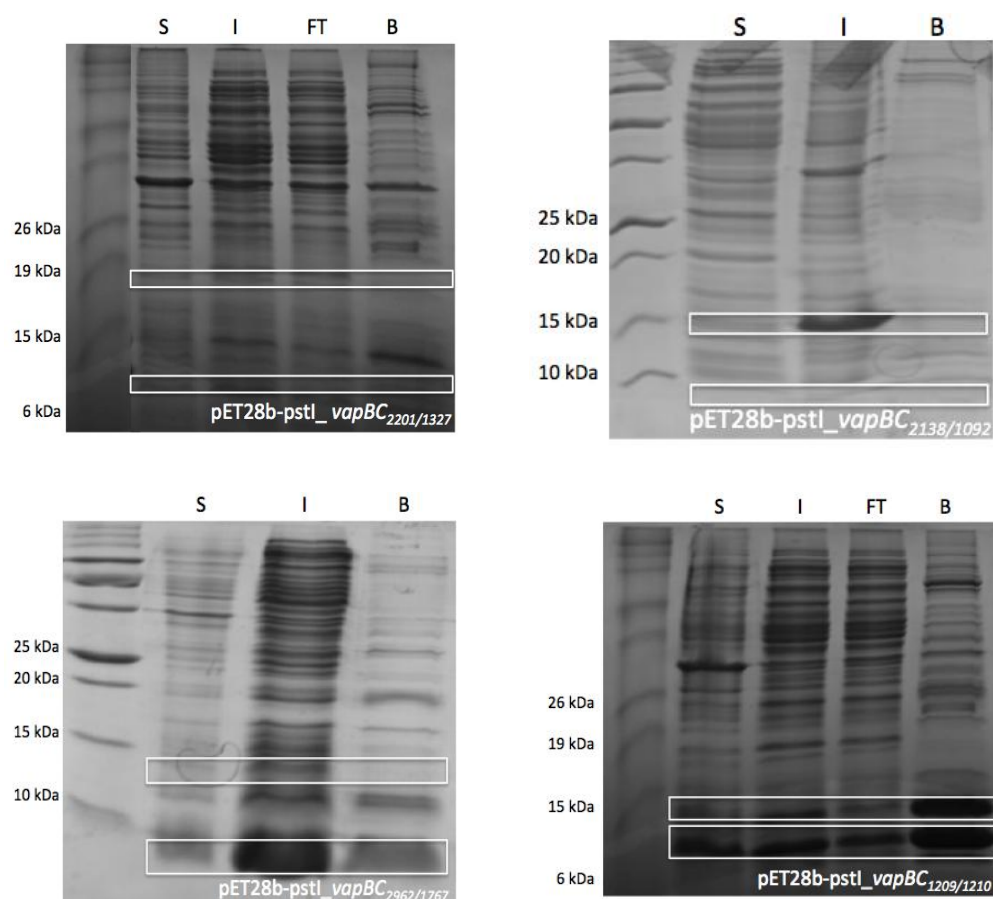


Figure A.4. SDS-PAGE analysis of *E. coli*_pET28b-pstI+vapBC (*Synechocystis*) small scale expression trials. White boxes indicate expected band size. Labels: whole cell fraction (WC), soluble fraction (S), insoluble fraction (I), flow through (FT), wash (W) and nickel beads with protein bound after wash steps (B).

