WARWICK
THE UNIVERSITY OF WARWICK

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:
http://wrap.warwick.ac.uk/1169300

Copyright and reuse:
This thesis is made available online and is protected by original copyright.
Please scroll down to view the document itself.
Please refer to the repository record for this item for information to help you to cite it.
Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

Functional and structural insights into one-component transcriptional regulators involved in methylenomycin production in *Streptomyces coelicolor* A3(2)

Rohini Ajaykumar

Supervised by: Dr Christophe Corre & Dr Allister Crow

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Life Sciences

School of Life Sciences, University of Warwick

December 2021

Table of Contents

Acknowledgements ix

Declaration x

Abstract xi

Abbreviations xii

List of Figures xiv

List of Tables xix

Chapter 1: Introduction 1

1.1 The role of *Streptomyces* in antibiotic production and discovery 1

1.2 Regulation of antibiotic production in *Streptomyces coelicolo*r A3(2) 5

1.2.1 A summary of transcriptional regulation in characterised *S. coelicolor* A3(2) antibiotic pathways 5

1.2.2 Antibiotics as ligands for transcriptional regulators of antibiotic gene clusters in *Streptomyces* 11

1.3 The close relation between heavy metal sensing and antibiotic resistance 13

1.3.1 Antibiotic (self) resistance mechanisms 13

1.3.2 Antibiotic resistance and heavy metal resistance parallels 17

1.4 Methylenomycin biosynthesis and regulation 19

1.4.1 MmA biosynthesis and regulation 22

1.4.2 MmA resistance and regulation 25

1.5 Project aims and objectives 27

1.5.1 Characterising MmyJ 27

1.5.2 Characterising MmyB 29

Chapter 2: Materials & Methods 31

2.1 Bioinformatics 31

2.1.1 Creating a phylogenetic tree 31

2.1.2 X-ray crystallography data collection and structure solving 32

2.2 Materials for experimental work 33

2.2.1 Purchased oligonucleotides 33

2.2.2 Plasmids and Strains 35

2.2.3 Buffers for protein purification work 36

2.2.4 Growth media 37

2.3 DNA and cloning work 39

2.3.1 Producing the intergenic 80 base pair region between mmyJ and mmr 39

2.3.2 Preparing short double stranded DNA 41

2.3.3 Amplifying *mmyB* from the C73-787 cosmid 41

2.3.4 Cloning *mmyB* into a pET151 vector 42

2.3.5 Preparing the pJ251 Tinsel Purple Reporter System 43

2.3.6 Creating a mScarlet-I Reporter system 45

2.3.7 Creating T18- mutant MmyJ plasmids 46

2.4 Transformations 47

2.4.1 Propagation Host Transformation 47

2.4.2 Glycerol stocks of cells 48

2.4.3 Miniprepping of DNA for Sanger sequencing 48

2.4.4 Expression Host Transformation (BL21 and C43 *E. coli* strains) 48

2.4.5 Confirmation of BL21 pET151-MmyJ 49

2.4.6 Bacterial adenylate cyclase two-hybrid (BACTH) assays 50

2.5 Protein production and purification 51
2.5.1 Small cultures to test for protein expression.....51
2.5.2 Large cultures for high protein production.....52
2.5.3 Immobilised Metal Ion Affinity Chromatography (IMAC).....52
2.5.5 Size Exclusion Chromatography (SEC) for purification 53
2.5.6 Buffer exchange of protein 54
2.5.7 Storage of pure protein.....55
2.5.8 Targeted proteomics to confirm protein samples.....55
2.5.9 X-ray Crystallography of protein 56
2.6 *In vitro* analytical work.....57
2.6.1 Size exclusion chromatography (SEC) for stoichiometry analysis.....57
2.6.2 Protein-DNA Assays (Electromobility Shift Assays).....58
2.7 Ligand Production 59
2.7.1 Growth of *Streptomyces* strains.....59
2.7.2 Organic extraction of compounds from media.....60
2.7.3 Analytical techniques for purification and detection 61
2.7.3.1 Purification by high-performance liquid-chromatography.....61
2.7.3.2 Detection of compounds by thin layer chromatography 61
2.7.3.3 Compound detection by liquid-chromatography mass spectrometry ..62
Chapter 3: Studying the transcriptional properties of MmyJ 63
3.1 Molecular basis for the control of methylenomycin self-resistance in *S. coelicolor* A3(2) 63
3.1.1. Overview of the ArsR family which are metal sensing transcriptional regulators 64
3.1.2 Non-metal binding exceptions in the ArsR family 68

3.1.3 Evolution of MmyJ from ArsR family proteins 70

3.1.4 MmyJ-like systems are present in other actinomycetes 73

3.2 Producing pure recombinant MmyJ for structural study 80

3.2.1 Expression of recombinant MmyJ 81

3.3 Evidence of MmyJ as a transcriptional protein from DNA binding 91

3.3.1 MmyJ binds within an 80 base pair region of intergenic DNA 91

3.3.2 Fragmenting intergenic DNA to find exact MmyJ binding sites 94

3.3.3 The structure of MmyJ was used to find two DNA binding sites of MmyJ ... 98

3.3.4 Conserved intergenic regions where MmyJ-like genes may bind in other actinobacteria 101

3.4 Using size exclusion chromatography to gather details into C49S MmyJ-DNA. 104

3.5 Dimerization of MmyJ 108

3.6 Testing MmyJ functionality in vivo through use of reporter systems 111

3.6.1 A pJ251 Tinsel purple reporter system 113

3.6.2 The mScarlet reporter system 116

3.7 Release of DNA-bound MmyJ by methylenomycin A and methylenomycin A intermediates 120

3.7.1 Ligand control of transcriptional proteins 120

3.7.2 Production of the predicted MmyJ ligand: Methylenomycin A 121

3.7.3 Optimising the production Methylenomycin A 129

3.7.4 Crude extracts containing mainly Methylenomycin C do not release MmyJ from being bound to DNA 131

3.7.5 Testing MmyJ binding with crude extracts containing MmA 133

3.8 Future work testing for MmyJ functioning 134

Chapter 4: Solving and analysing the structure of MmyJ 137

4.1 Crystallisation of MmyJ 137

4.2 Structure determination of MmyJ 138

4.2.1 The diffraction of C49S MmyJ 138

4.2.2 Data merging and scaling of C49S MmyJ 140

4.2.3 Molecular Replacement of C49S MmyJ 142

4.2.4 Refining the structure of C49S MmyJ 144

4.2.5 Validation of the constructed C49S MmyJ structure 148

4.2.6 Structure determination of recombinant wild type MmyJ 154

4.3 Structure of MmyJ 159

4.3.1 Overall fold of MmyJ 159

4.3.2 There are no significant structural differences between C49S and wild type MmyJ 161

4.3.3 The dimer interface of MmyJ 162

4.4 Analysis of MmyJ structure 166

4.4.1 MmyJ in comparison to other similar structures 166

4.4.2 Inferences for the ligand binding site of MmyJ 169

4.4.3 Predictions of the DNA recognition sites of MmyJ 173

4.5 Future structural work involving MmyJ 179

4.5.1 X-ray crystallisation studies with MmyJ 179

4.5.2 Mutations on MmyJ to identify key residues for functionality 181

Chapter 5: MmyB: A proposed transcriptional activator of methylenomycin A biosynthetic genes 183

5.1 Predictions of the biological role of MmyB 183

5.1.1 The *mmyB* gene and its role in methylenomycin biosynthesis 183

5.1.2 Structural insight into MmyB from bioinformatics and homology to a MmyB like protein 188

5.2 MmyB produced in *E. coli* BL21 cannot fold correctly 192

5.2.1 Cloning *mmyB* into a pET151 plasmid vector for overexpression 193

5.2.2 Production of MmyB in *E. coli* BL21 is low and does not purify well.....199

5.2.3 MmyB in *E. coli* BL21 purifies with chaperones highlighting difficulties in folding 203

5.3 MmyB can be overexpressed and purified using *E. coli* C43 205

5.3.1 Overexpression of MmyB in C43 is better than in BL21 205

5.4 Recombinant MmyB is unstable and precipitates 210

5.5 X-ray diffraction of MmyB crystals.....212

5.6 Apo-MmyB does not bind DNA and likely requires ligand binding first 214

5.7 Evidence to suggest MmyB acts as a transcriptional activator for MmA production.....216

5.7.1 Testing *mmyB* activity in *S. coelicolor* strains 216

5.7.2 Can constitutive expression of *mmyB* bypass the MmfR/MmyR/MMF regulation system?.....219

5.8 Summary of MmyB work and future work with MmyB.....221

Chapter 6: Final Discussion 224

6.1 Summary of results 224

6.2 A proposed mechanism of transcriptional regulation of antibiotic self-resistance by MmyJ 225

6.3 Impact of research and future prospects 228

Appendix 231

A: Supplementary figures for Chapter 3 231

B: Supplementary figures for Chapter 4 244

C: Supplementary figures for Chapter 5 246

References.....250

Acknowledgements

I express my deepest gratitude to both my wonderful supervisors, Dr Christophe Corre and Dr Allister Crow. Thank you both for being available to answer even my smallest queries. Your patience, guidance, and kindness has made me a far greater scientist than I could have imagined in such a short amount of time. There is always more to explore in any project like this, but I am very happy with the results produced and I hope you both are too.

My time in the lab would not have been as enjoyable or successful if not for my incredible colleagues turned friends. I would like to thank; Nazia for her help with HPLC and organic chemistry, Jack W for his help with LC-MS, Patrick for his help with TLC, Matt H for all his expertise in bioinformatics, Jack H for lending his knowledge of cloning processes and *Streptomyces* and James K for discussing problems in protein expression. You have all softened the blow of my failed experiments and heightened the joy of my successful ones, I hope I have done the same for you. Thank you to my friends from the MIBTP cohort; Jeff, Niamh, and Holly, we started this together and I can't believe how quickly we've made it to the end. My PhD experience would not have the same without you guys. Also, a big thanks to my lifelong friends outside of science; Margaux, Sancha, Amiya and Rebecca. Being surrounded by brilliant, scientific minds every day it is so hard not to get imposter syndrome and quit, but your support has been crucial in helping me persevere.

Last but by no means least, I am forever grateful to my family both close and extended for their love and blessings. To my parents, Ajaykumar and Soumini, thank you for all your sacrifices and support, without which I would not have made it this far. Thank you to my brother, Anirudh, for never doubting I would be able to do this even when I doubted myself. I hope I have done you all proud.

Declaration

This thesis was funded by the BBSRC and submitted to the University of Warwick in support of the application for the degree of Doctor of Philosophy. The work presented within this thesis is original and was conducted by the author except in the following cases:

Plasmids (pET151-MmyJ, pJ251-Tinsel Purple, mScarlet-I) and strains (W89, W81, W81+pJH2, M145) as listed in Chapter 2 were sourced from the Corre Group, University of Warwick with full credit given to the creators.

T18 and T25 MmyJ plasmids were created by Dr Jonathan Cook (Crow Lab, University of Warwick). Dr Cook also conducted the BACTH assay on the mutant T18-MmyJ plasmids.

Usage of the LC-MS machine in Chapter 3 and Chapter 5 for organic extract analysis was curtesy of Chen Lab (University of Warwick).

Crystals of MmyJ C49S and MmyB were harvested by Dr Allister Crow (University of Warwick) and sent to Diamond Light Source for X-ray crystallography. Dr Crow also created the AlphaFold structure of C49S MmyJ for molecular replacement used for C49S MmyJ structure elucidation and created a model of MmyJ bound to DNA.

MmyJ WT X-ray diffraction data was previously collected and partially processed by Professor Vilmos Fulop who kindly shared this data.

No parts of this thesis have been published thus far or submitted previously for any degree application by the author.

Abstract

In one-component regulatory systems a transcriptional regulator (TR) protein typically comprises both ligand and DNA binding domains to prevent or promote the expression of target gene(s). One component TRs are present in the methylenomycin (an antibiotic) gene cluster in *Streptomyces coelicolor* A3(2). These are involved in controlling methylenomycin biosynthesis by the TR MmyB and controlling methylenomycin resistance by the TR MmyJ. The characterisation of these two TRs was the focus of this thesis.

Bioinformatic analysis of MmyJ indicate it to be an ArsR family (a metal sensing family of TRs) protein with uncharacterised one-component MmyJ like systems suggested in other actinobacteria. DNA-MmyJ binding assays revealed MmyJ binds to two 17 base pair pseudo-palindromic regions beside one another within the intergenic region between *mmr* (MFS type efflux pump and target gene of MmyJ) and *mmyJ*. Bacterial hybrid assays showed MmyJ forming homodimers and mutation of Leu25 highlighted it as a key residue for dimer formation. X-ray diffraction of MmyJ meant its structure was solved to 1.68 Å. Structural analysis of MmyJ corroborates it to be an ArsR family protein meaning MmyJ is the first antibiotic sensing ArsR family protein structure to be solved.

X-ray diffraction was also used for crystals of MmyB however these did not diffract sufficiently for structure elucidation. Using liquid chromatography mass spectrometry, the regulation upon *mmyb* was able to be evaded in order for methylenomycin biosynthesis, highlighting the significance of MmyB. This also indicates a potential method for discovery of novel secondary metabolites through over expression of *mmyb*-like genes.

Abbreviations

Å	Ångströms
ABC	Ammonium bicarbonate
ABC	ATP binding cassette
ACT	Actinorhodin
AHFCAs	2-alkyl-4-hydroxymethylfuran-3-carboxylic acids
AMR	Antimicrobial resistance
ARRs	Atypical response regulators
BACTH	Bacterial adenylate cyclase 2 hybrid
BGCs	Biosynthetic gene clusters
BICs	Baseline ion chromatograms
BIC	Bayesian information criterion
BLAST	Basic local alignment search tool
bp	Base pair
CAA	2-chloroacetamide
cAMP	Cyclic AMP
CAP	Catabolite activator protein
CCP4	Collaborative computational project number 4
CDA	Calcium dependent antibiotic
CPK	Cryptic polyketide
CSRs	Cluster situated regulators
CV	Column volume
Da	Daltons
dASA	Difference in accessible surface area
DTT	Dithiothreitol
EIC	Extracted ion chromatogram
EMSA	Electromobility shift assay
ETE	Environment for Tree Exploration
GBL	Gamma butyrolactone
HF	High fidelity
HGT	Horizontal gene transfer
HPLC	High performance liquid chromatography
HTH	Helix turn helix
IDT	Integrated DNA Technologies
IMAC	Immobilised metal ion affinity chromatography
IPTG	Isopropyl beta-D-1-thiogalactopyranoside

iTOL	Interactive Tree of Life
kbp	Kilo base pair
kDa	Kilo Daltons
LB	Luria Bertani
LC-MS	Liquid chromatography mass spectrometry
MARE	MMF autoregulator responsive element
MATE	Multidrug and toxic compound extrusion
MEME	Multiple Em for motif elicitation
MFS	Major facilitator superfamily
Mm	Methylenomycin
MR	Molecular replacement
MUSCLE	Multi-sequence alignment
NEB	New England Biolabs
ORF	Open reading frame
PAS	Per-Arnt-Sim
PBPs	Penicillin binding proteins
RBS	Ribosomal binding site
RED	Undecylprodiginine
RND	Resistance nodulation cell division
SARPs	*Streptomyces* antibiotic regulatory proteins
SASA	Solvent accessible surface area
SDS-PAGE	Sodium Dodecyl Sulphate–Polyacrylamide Gel Electrophoresis
SEC	Size exclusion chromatography
SMM	Supplemented Minimal Media
SMR	Small multidrug resistance
TCEP	Tris-(2-carboxyethyl)-phosphine hydrochloride
TLC	Thin layer chromatography
WHO	World Health Organisation
wHTH	Winged helix-turn-helix
XRE	Xenobiotic response element
α	alpha
β	beta
ϕ	phi
ψ	psi

List of Figures

Figure 1. Key antibiotics sourced mostly from *Streptomyces* and the life cycle of *Streptomyces* 4
Figure 2. The 5 best characterized antibiotics from *Streptomyces coelicolor* A3(2)..... 6
Figure 3. The actinorhodin (*act*) antibiotic producing gene cluster with *actII* genes highlighted..... 8
Figure 4. The regulation of coelimycin A (CPK). A) At high levels of SCB1, SCB1 binds to ScbR..... 10
Figure 5. The various methods of antimicrobial resistance to antibiotics (drugs) by bacteria. 14
Figure 6. The general architecture of the main families of drug efflux pumps in bacteria. 16
Figure 7. The proposed biosynthesis of the antibiotic methylenomycin A..... 20
Figure 8. The methylenomycin A gene cluster from *Streptomyces coelicolor* A3(2) SCP1 plasmid 21
Figure 9. The regulation of methylenomycin A (MmA) from *Streptomyces coelicolor* A3(2) by MmfR and MmyR..... 24
Figure 10. A simple one-component regulation system for methylenomycin A (MmA) self-resistance in *Streptomyces coelicolor* A3(2)..... 26
Figure 11: The ribbon structure of apo-SmtB, a DNA binding protein with type 1 and 2 metal binding sites 65
Figure 12. A possible evolution model type 1 and type 2 metal binding sites in ArsR family proteins..... 66
Figure 13. Oxidised and reduced structures of HlyU with Cys38 reduced or oxidised to cysteine sulfenic acid (CSO) 69
Figure 14. Phylogeny and relation of MmyJ to other metal-sensing transcriptional proteins. 72
Figure 15. Phylogeny trees of MmyJ to closely related proteins. 77
Figure 16. Organisms identified to have a protein similar to MmyJ with genes beside it encoding for MFS transporters 79
Figure 17. Recombinant MmyJ used for expression of MmyJ. 83

Figure 18. Recombinant MmyJ protein purification by immobilized metal affinity chromatography (IMAC) specifically nickel-NTA 85
Figure 19. Recombinant MmyJ was subjected to mass spectrometry analysis 87
Figure 20. Size exclusion chromatography (SEC) of recombinant MmyJ-WT (B) and recombinant C49S MmyJ (C) against standards (A). 90
Figure 21. The intergenic DNA region between *mmr* and *mmyJ* with an 80 base pair (bp) section of it encompassing the MmyJ binding sites 92
Figure 22. Electromobility shift assay (EMSA) of C49S MmyJ binding to DNA 93
Figure 23. The 80 base pair (bp) section of DNA was further fragmented and tested for C49S MmyJ binding.....95
Figure 24. Electromobility shift assay (EMSA) of palindromic DNA and short regions of DNA C49S MmyJ may bind to 97
Figure 25. MmyJ binds at regions between the -35 promoter sites of *mmr* and *mmyJ* 99
Figure 26. MmyJ binds at separate regions encompassing either the -35 promoter sites of *mmr* or *mmyJ* 100
Figure 27. Intergenic DNA regions of organisms between their MmyJ-like genes and MFS transporter have consensus motifs.....103
Figure 28. C49S MmyJ-DNA stoichiometry was tested using analytical size exclusion chromatography. 106
Figure 29. A bacterial adenylate cyclase two-hybrid assay (BACTH assay) shows MmyJ forms dimers (A) and leucine 25 is necessary for dimer formation (B).....110
Figure 30. Theoretical basis on how a MmyJ reporter system may work *in vivo* to assess strength of MmyJ-DNA binding.....112
Figure 31. The pJ251 Tinsel Purple reporter system. 113
Figure 32. The pJ251 Tinsel Purple reporter system had to have an insert of an additional promoter for colour production 114
Figure 33. Intergenic DNA was inserted into an mScarlet reporter plasmid for testing MmyJ-DNA binding *in vivo*. 119
Figure 34. Plasmids from colonies of mScarlet with an 80 base pair (bp) *mmyJ-mmr* intergenic DNA were digested with NdeI and run on a 1% TAE-Agarose gel. 120

Figure 35. Methylenomycin gene clusters in *Streptomyces* strains used for production and extraction of Methylenomycin A (MmA).....122
Figure 36. High-performance liquid chromatography (HPLC) of samples of W89, W108 and M145.....124
Figure 37. Low resolution, liquid chromatography mass spectrometry (LC-MS) of high-performance liquid chromatography (HPLC) purified W89 samples and of crude W89, M145 and W108 samples to detect Methylenomycin A (MmA).....125
Figure 38. Thin layer chromatography (TLC) of M145 and W89 samples to indicate for MmA 126
Figure 39. Low resolution, liquid chromatography mass spectrometry (LC-MS) of W89 and M145 samples for Methylenomycin compound detection. 128
Figure 40. Low resolution, liquid chromatography mass spectrometry (LC-MS) of W89 and M145 samples for Methylenomycin compound detection 130
Figure 41. Electromobility shift assays (EMSAs) of MmyJ, DNA and ligand binding 132
Figure 42. Electromobility shift assays (EMSAs) of MmyJ, DNA and ligand binding 133
Figure 43. Crystals of recombinant MmyJ and C49S MmyJ. 138
Figure 44. C49S MmyJ crystal used for X-ray diffraction and the diffraction this crystal gave 139
Figure 45. Resolution of C49S MmyJ X-ray diffracted data. 144
Figure 46. Resolution of C49S MmyJ X-ray diffracted data at N and C terminal ends of chains. 145
Figure 47. Solvents were built into the C49S MmyJ model which may interact with protein side chains. 146
Figure 48. A graph of R values and resolution of solved MmyJ structures and published structures in the Protein Databank 148
Figure 49. Ramachandran plots for validation of residues in the C49S MmyJ model. 150
Figure 50. PROCHECK results summary for validation of the C49S MmyJ model....151
Figure 51. MolProbity results summary for validation of the C49S MmyJ model ...153

Figure 52. Ramachandran plots for validation of residues in the WT MmyJ model 156
Figure 53. PROCHECK results summary for validation of the WT MmyJ model. 157
Figure 54. MolProbity results summary for validation of the WT MmyJ model.....158
Figure 55. The structure of MmyJ. A ribbon model of the secondary structure of C49S MmyJ which is formed of 2 dimers (Chain A and Chain B) in rainbow format (A) 159
Figure 56. The structures of WT-MmyJ and C49S MmyJ overlayed over each other 162
Figure 57. Dimer interface residues that may be interacting in homodimeric MmyJ 164
Figure 58. Hydrogen bonds occurring from dimer interface residues in MmyJ 166
Figure 59. Structures of proteins most similar to MmyJ superimposed on MmyJ ..168
Figure 60. The solvent accessible surface area (SASA) of a homodimer of MmyJ...171
Figure 61. Crystal structure of *Af* ArsR bound to As(III) 173
Figure 62. The superimposed structures of ArsR and MmyJ with ligand and DNA binding regions of ArsR highlighted. 174
Figure 63. Crystal structure of homodimeric apo-NolR and DNA-bound NolR compared to apo-MmyJ in similar orientations.....175
Figure 64. A model of MmyJ binding to DNA.....178
Figure 65. The pseudo-palindromic sequences predicted to be sites of MmyJ homodimer binding.....179
Figure 66. Intergenic region of DNA between *mmyB* and *mmyY*. 184
Figure 67. There are many *mmyB*-like genes in the *Streptomyces* species 187
Figure 68. MEME analysis of intergenic regions between *mmyB* like genes and genes divergent to *mmyB*-like ones in *S. coelicolor* A3(2).....187
Figure 69. MmyB protein sequence and features. A) The *S. coelicolor* A3(2) *mmyB* gene and the codon translation into its protein sequence.....189
Figure 70. MltR homodimers bound to myristic acid 191
Figure 71. Gradient PCR to amplify *mmyB* shown on a 1% Agarose-TAE gel.....193
Figure 72. Enzymatic digest of pET151-*mmyB* on a 1% Agarose-TAE gel.....194

Figure 73. Gradient PCR on 1% Agarose-TAE of a colony containing pET151 with an *E. coli* codon optimized *mmyB* inserted. 196
Figure 74. Recombinant MmyB used for expression of MmyB 198
Figure 75. 50 mL cultures of BL21 pET151-*mmyB* was tested for expression of MmyB. 199
Figure 76. Recombinant MmyB protein purification by immobilized metal affinity chromatography (IMAC) specifically nickel-NTA 200
Figure 77. Size exclusion chromatography of recombinant MmyB. 202
Figure 78. IMAC purifications of recombinant MmyB was subjected to mass spectrometry analysis 204
Figure 79. Recombinant MmyB protein purification from BL21 and C43 strains by immobilized metal affinity chromatography (IMAC) specifically nickel-NTA. 206
Figure 80. Recombinant MmyB protein purification from C43 and BL21 strains by immobilized metal affinity chromatography (IMAC) specifically nickel-NTA including an ATP Wash. 208
Figure 81. IMAC purifications of recombinant MmyB was subjected to mass spectrometry analysis. 209
Figure 82. Size exclusion chromatography of recombinant MmyB 211
Figure 83. Crystals of recombinant MmyB were subject to X-ray crystallography ..214
Figure 84. Electromobility shift assay (EMSA) of MmyB and DNA. 215
Figure 85. The methylenomycin gene cluster in the W81 strain of *Streptomyces coelicolor*. 218
Figure 86. Low resolution, liquid chromatography mass spectrometry (LC-MS) of W81 and W81 with an extra *mmyB* containing plasmid for the detection of Methylenomycin A. 220
Figure 87. The proposed mechanism of MmyJ in *Streptomyces coelicolor* A3(2) ...226

List of Tables

Table 1: Purchased primers.....33
Table 2: Plasmids and strains. 35
Table 3: MmyJ Buffers 36
Table 4: MmyB Buffers 37
Table 5: Q5 PCR Reaction mix. 40
Table 6: TOPO cloning reaction mix. 42
Table 7: Restriction enzyme reaction mix. 43
Table 8: Ligation reaction mix). 44
Table 9: Mutagenesis PCR reaction mix 46
Table 10: BLAST search for proteins similar to MmyJ. 75
Table 11: Organisms identified to have a protein similar to MmyJ with a gene beside it encoding for MFS transporters (from Fig. 16) and the intergenic length of DNA between them 102
Table 12: Data merging and scaling results overview. 140
Table 13: Unit cell composition of X-ray diffraction data of C49S MmyJ.....143
Table 14: Final refinement R values of C49S MmyJ 147
Table 15: WT-MmyJ X-ray diffraction statistics from a merged, processed data file. 154
Table 16: Final refinement R values of WT-MmyJ 155
Table 17: The dimer interface residues and the alpha chain they are part of in a homodimer of MmyJ. 165
Table 18: The proteins in the PDB database that are structurally the most like MmyJ 167
Table 19: Genes that are *mmyB*-like and have similar synteny within the *S. coelicolor* A3(2) genome 186
Table 20: BLAST Search of MmyB.....190

Chapter 1: Introduction

1.1 The role of *Streptomyces* in antibiotic production and discovery

Antibiotics have a long and rich history with the first traces of tetracyclines (a class of antibiotics) detected to be used by ancient Egyptians to treat infection.[1] Though it was only in the early 20th Century work began to isolate compounds toxic against bacteria. In 1909 Paul Ehrlich discovered that the chemical arsphenamine (later marketed under the name Salvarsan) was toxic against the bacterium *Treponema pallidum* which causes syphilis.[2] However, at the time he had not referred to this as an antibiotic but as a 'chemotherapy' agent. More commonly acknowledged as the first antibiotic (as recognised by modern day standards) was discovered by Alexander Fleming in 1928 from *Penicillium notatum* developed into the drug penicillin by Florey and Chain.[3] This term 'antibiotic' was only coined later in 1943 by Selman Waksman to describe a chemical compound with the capability to inhibit or kill microorganisms.[4,5] Waksman also went onto discover 20 antibiotics himself which began the start of the 'golden age' of antibiotic discovery.

This period between 1950-1960 was known as the golden age of antibiotic discovery with many antibiotics discovered in this time still in use today. A large majority (around 80 %) of known antibiotics to date derive from the *Streptomyces* genus (Fig. 1A).[6]

Streptomyces is a species of Gram-positive bacteria found commonly in soil.[7,8] They belong to the phylum Actinomycetes due to their high GC rich DNA (~ 70 %) in comparison to other bacteria such as *Escherichia coli* (50 %). The *Streptomyces* growth cycle cultured on solid media is similar to that of filamentous fungi where it can branch into hyphae that are able to form vegetative mycelia extending into its surrounding environment. Aerial hyphae used for reproduction then extend to the

surface where they then release spores that disperse to find new habitable environments to colonise. This formation of the aerial hyphae is important not just as the initiation for spore formation but also for the production of secondary metabolites and other natural products (Fig. 1B).[7,9] These natural products stem from a variety of *Streptomyces* species to include antibiotics (e.g. kanamycin), antifungals (e.g. nystatin), antiparasitic (e.g. ivermectin), anticancer (e.g. bleomycin) and other industrially relevant compounds.[10,11] For the *Streptomyces* bacteria these are produced at high concentrations as a stress response to competition from other bacteria (predation and protection) but at low concentrations are thought to assist in forming symbiotic relationships with its surrounding organisms.[12–14]

Despite their exploitation for natural products there is still a lot unknown about *Streptomyces.* The two best studied species of *Streptomyces* are *Streptomyces coelicolor* A3(2)[15] and *Streptomyces griseus*[16] ; with *S. griseus* being the first *Streptomyces* species to be sourced for antibiotics (streptomycin)[4] and *S. coelicolor* A3(2) being used widely as a model organism for genetic studies[17]. The A3(2) strain of *S. coelicolor* refers to the bacteria containing both the SCP1 and SCP2 plasmids whereas *S. coelicolor* missing these plasmids is referred to as the M145 strain.[15]

First defined genetically by Vivian (1971)[18] and then later sequenced using pulsed field gel electrophoresis by Kinashi *et al.* (1993)[19], the SCP1 plasmid is a 365 023 bp linear double-stranded plasmid. While at the time of discovery this linearity of a plasmid was a unique feature, in 1994 Kinashi *et al.*[20] went on to discover linear plasmids in other streptomycetes strains after which linear plasmids were shown to be present in other actinomycetes too (e.g. Cornell *et al.* 2018)[21] . SCP1 has some unique properties and as such have been isolated in multiple forms from *S. coelicolor* A3(2); for example, SCP1 can be lost altogether in some *S. coelicolor* A3(2) strains, it can exist autonomously and can even be integrated into the chromosome in certain strains.[22] This makes for a very complex evolution of SCP1 which is still not entirely understood, especially as it has acquired large numbers of transposable elements (with ~ 80 kbp inverted terminal repeats within its DNA).[19] As such finding biosynthetic gene clusters (BGCs) on a *Streptomyces* plasmid encoding for natural

products (e.g. antibiotics) is a rarity with the majority located on the bacterial chromosome.[23] Furthermore, apart from the antibiotic resistance genes encoded within the methylenomycin gene cluster, unusually the SCP1 plasmid does not contain any other recognisable antibiotic resistance genes.[22]

While SCP1 does contain some fertility factors[18] mutant strains with loss of SCP1 still displayed low but detectable levels of fertility. Fertility in the *S. coelicolor* A3(2) strain is therefore attributed to the SCP2 plasmid as determined first by Bibb *et al.* (1977) by crossing various SCP1 and SCP2 mutants of *S. coelicolor* A3(2).[24] Similar to SCP1 it too is a low-copy-number plasmid but unlike SCP1 is circular and ~ 31 kbp in length.[24,25] Along with SCP2, a spontaneous derivative SCP2* was also isolated by Bibb *et al.* (1977) and found to have two key differences to SCP2; one being that SCP2* has a 1000-fold increase in its ability to mobilise the chromosome and the other being that phenotypically it forms more visible pocks. Pocks are circular areas of delayed growth around colonies growing with plasmid-bearing spores on a lawn of a strain lacking that plasmid.[24] The molecular basis of this difference in phenotypes is unknown with SCP2 and SCP2* also being unable to be distinguished using restriction digestion.[26] Despite SCP2* having been sequenced many of the proteins in the 34 open reading frames (ORFs) identified are still unknown.[27] Similar to SCP1, SCP2* sequencing did identify two transposable elements with it presumably being a feature of SCP2 also as integration into the chromosome may be how SCP2 plasmid DNA can be isolated from SCP2* phenotype cells.[27,28] Finally, the stability, low copy number and acceptance of large exogenous DNA inserts without structural rearrangement by SCP2 and SCP2* make it a valuable tool for cloning studies in *Streptomyces* as demonstrated first by Lydiate *et al.* (1985).[29]

The complete genomes of both *S. coelicolor* A3(2) and *S. griseus* these were published[15,16] leading to the discovery of many more novel natural products. Despite this, genomic studies infer that this is not the full range of secondary metabolites they are be able to produce.[10,30]

The tapering of discovery of natural products after the golden age is likely due to only a finite number of these being produced in bacteria that can grow in laboratory settings and have this production readily turned on. With new genome mining and heterologous pathway expression tools these BGCs capable of producing natural products are being explored with attempts being made to activate such silent BGCs for novel natural product discovery.[31–33] Understanding these BGCs and the mechanism by which they produce natural products like antibiotics can assist in finding and activating silent BGCs in the *Streptomyces* species and other actinomycetes.

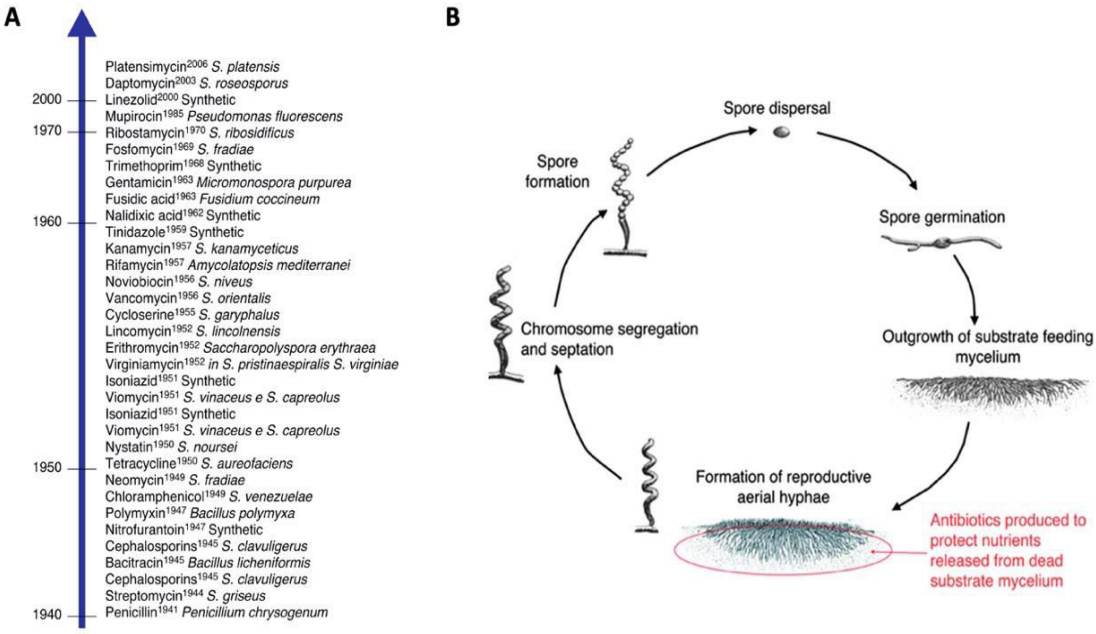


Figure 1. Key antibiotics sourced mostly from *Streptomyces* and the life cycle of *Streptomyces*. A) The key antibiotics with dates from history. As shown around 80% come from the *Streptomyces* species. Figure taken from de Lima Procópio *et al*. (2012).[17] **B)** The life cycle of *Streptomyces* grown on an agar plate in the lab. As seen, it is at the start of aerial hyphae formation for spore formation and reproduction that secondary metabolites such as antibiotics are seen to be produced. Figure taken from Seipke *et al.* (2011).[14]

1.2 Regulation of antibiotic production in *Streptomyces coelicolor* A3(2)

Most BGCs that produce natural products are not constitutively expressed (thus the issue with silent BGCs) but have their expression regulated by proteins of other genes. These regulatory genes can also be found within BGCs themselves, termed as 'pathway-specific' regulators, or at other areas within the genome and act on BGCs as 'global' regulators. Global regulators may also act upon pathway-specific regulators either directly or indirectly to control the expression of natural products of the pathway-specific BGCs.[34–36] Initially it was thought pathway specific regulators could only control the expression levels of the genes within their cluster though microarray analysis of *S. coelicolor* A3(2) revealed these pathway-specific regulators may exhibit global effects too.[37] In such case, pathway-specific regulators are now termed 'cluster-situated regulators' (CSRs).[35,36,38]

1.2.1 A summary of transcriptional regulation in characterised *S. coelicolor* A3(2) antibiotic pathways

There are 5 well studied *S. coelicolor* A3(2) antibiotic BGCs and their CSRs; actinorhodin (ACT), undecylprodiginins (REDs), coelimycin (CPK), calcium-dependent antibiotic (CDA) and methylenomycin (Mm) pathways (Fig. 2).[38] The latter (Mm) is SCP1 plasmid encoded and detailed further in Chapter 1.4 while the rest are chromosomally encoded. These CSRs encode not only genes that regulate the production but also the export of specific antibiotics.[39] It is the regulatory proteins of these pathways that will be summarised below for further discussion later on.

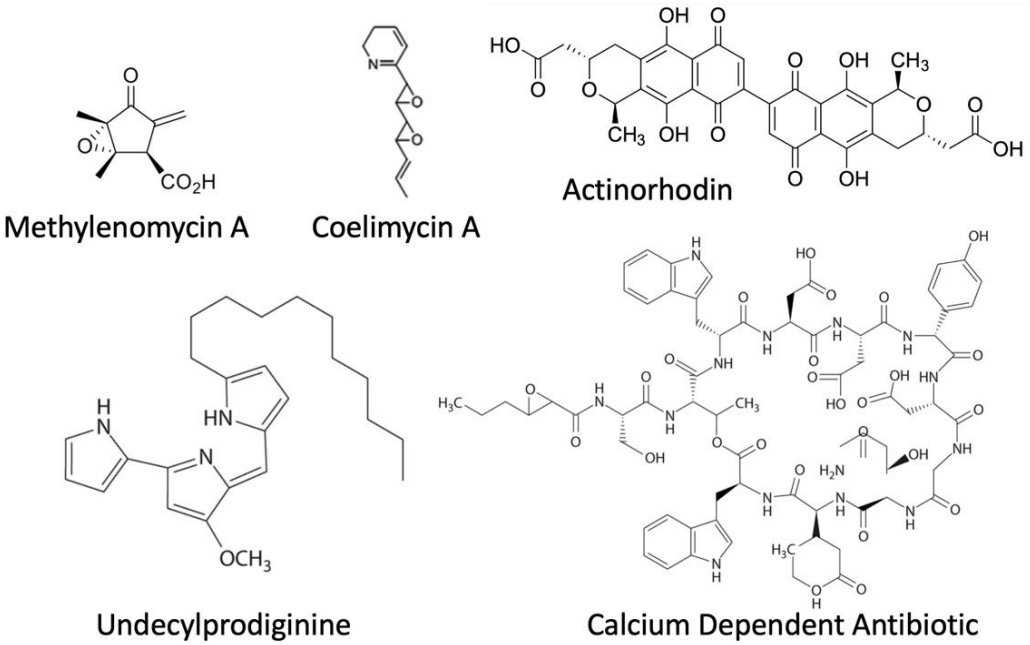


Figure 2. The 5 best characterized antibiotics from *Streptomyces coelicolor* A3(2). The structures of methylenomycin A (MmA), coelimycin A (CPK), actinorhodin (ACT), undecylprodiginine (RED) and calcium dependent antibiotic (CDA) are shown. These antibiotics are produced by biosynthetic gene clusters (BGCs) under non-consecutive control of regulators. *S. coelicolor* A3(2) contains the plasmids SCP1 and SCP2, with SCP1 encoding the BGC for MmA. The BGCs of the other 4 antibiotics are chromosomally encoded.

Actinorhodin is the antibiotic responsible for the blue colour[40] of *S. coelicolor* A3(2)[41] dependent on the pH of the environment. Its transcription is controlled by the *act* pathway which is a relatively large BGC with 22 genes for the production and regulation of actinorhodin (Fig. 3A).[42] The protein products of a large majority of the genes within the cluster have been found to be involved in the process of actinorhodin biosynthesis[43,44] but the *actII* region contains genes for regulation and self-resistance (Fig. 3B). The gene product of *actII-ORF4* is a transcriptional regulator protein for the biosynthesis of actinorhodin.[45,46] The protein has an N-terminal helix-turn-helix (HTH) motif to bind to specific regions of DNA. It is a positive regulator meaning DNA binding of the protein increases production of actinorhodin. Structurally it is similar to the DNA binding protein OmpR but was distinct enough to be a founding member of a new class of transcriptional proteins

known as *Streptomyces* antibiotic regulatory proteins (SARPs).[28,47] Its own translation is dependent on the expression of a specialised tRNA encoded by the *bldA* gene able to translate the rare codon UUA present in the mRNA of ActII-ORF4.[48]

The other regulatory protein in the subcluster is that from *actII-ORF1* (Fig. 3B). This produces the protein ActR, a transcriptional DNA-binding protein that represses expression of the promoter for the export of actinorhodin.[49] ActR binds to the operator *actO* found upstream of the gene for *actA*.[50] The gene *actA* is divergent to *actR* and encodes an efflux pump protein (Fig. 3). In fact, the gene downstream (*actB*) also encodes for a transmembrane efflux pump and as *actB*'s start site overlaps with *actA*, *actAB* is controlled by a singular promoter at the start of *actA*.[40,50,51] To summarise, this means that ActR regulates the expression of transmembrane export proteins for actinorhodin export to provide self-resistance. ActR is de-repressed from being bound to DNA by binding to late-stage intermediates of actinorhodin biosynthesis[49] to allow expression of *actAB*, thus conserving cellular energy by having self-resistance only when antibiotic is being produced.

This organisation of regulatory genes in the centre of the antibiotic BGC and having a transcriptional regulator divergent to an efflux pump gene it controls is similar to that of the methylenomycin gene cluster (Chapter 1.4).[53]

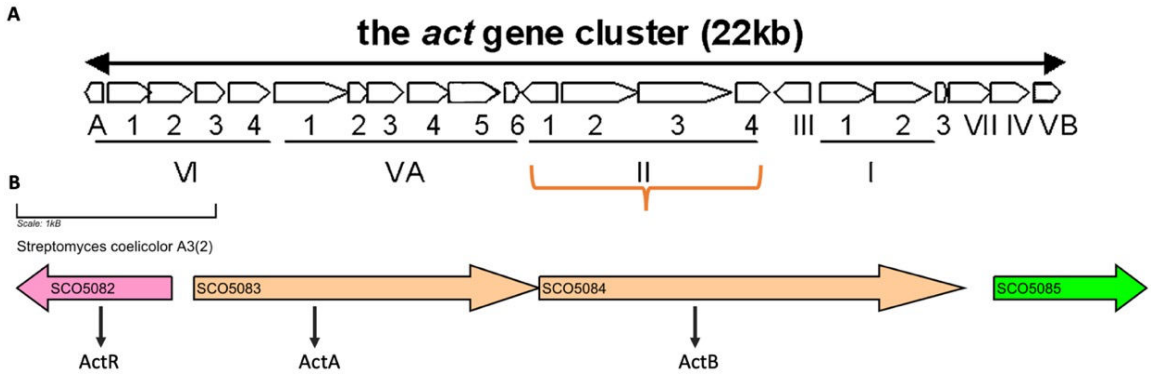


Figure 3. The actinorhodin (act) antibiotic producing gene cluster with *actII* genes highlighted. A) The act gene cluster that is responsible for the production and regulation of the antibiotic actinorhodin in *Streptomyces*. This is a large gene cluster consisting of 22 genes from the chromosome of *S. coelicolor A3(2)*. While most of the genes in the cluster relate to the biosynthesis of actinorhodin the ones in actII are relevant to actinorhodin self-resistance and actinorhodin biosynthesis regulation. Figure from Okamoto *et al.* (2009).[42] **B)** The *actII* gene cluster has 4 genes from regulation of actinorhodin. The first is a divergent gene *actR* which is a transcriptional repressor that controls expression of *actAB*.[50] Both *actAB* are genes encoding antibiotic efflux pumps.[49] The final gene in the gene cluster *actII-ORF4 (sco5085*) is a transcriptional activator for actinorhodin biosynthesis.[35] **B** created using Gene Graphics by Harrison *et al.* (2018).[52]

Undecylprodiginine (RED) production in *S. coelicolor* A3(2) is mainly under positive transcriptional control by the protein RedD[54,55] which is also a founding member of the SARP class like ActII-ORF4. The production of RedD itself is dependent on another RED protein family- RedZ. RedZ transcriptionally regulates *redD* in a positive manner but is under the influence of *bldA* for that specialised tRNA as *redZ* (but not *redD*) contains the rare UUA codon for leucine.[56,57] As well as regulating *redD* expression, RedZ is also a TetR negative autoregulator like ActR.[57] TetR family proteins are one-component transcriptional regulators involved in a diverse range of metabolic systems (carbon, nitrogen, amino acid, lipid) and those TetR proteins that regulate antibiotic efflux pumps, while well characterised, are only a minority of the family.[58]

Coelimycin (CPK) or more specifically coelimycin A is a weak antibiotic containing two epoxide rings (Fig. 2). These epoxide rings undergo spontaneous reactions with any N- acetylcysteine or glutamate in the media to the form the yellow pigmented coelimycin P1 and P2 that have no antibiotic activity.[10] The 58 kbp *cpk* gene cluster consists of 24 genes including 7 regulatory ones; *scbR, scbA, scbB, orfB, cpkO, scbR2* and *cpkN* (Fig. 4). OrfB is a histidine protein kinase homolog with its target unknown. ScbA is a synthase of a type of small hormone signalling γ- butyrolactone (GBL) molecule known as SCB1 with ScbB also involved in SCB1 synthesis.[59] SCB1 is recognised by the TetR-like protein, ScbR. Although ScbR2 has high primary sequence similarity to ScbR, ScbR2 is a pseudo-GBL receptor and cannot bind SCB1 in the same way ScbR can. Instead ScbR2 has been shown to bind to antibiotics RED and ACT indicating crosstalk between CSRs. Both CpkO and CpkN are necessary SARP proteins that function as activators for the biosynthesis of CPK. The *cpkO*, *cpkN* promoter sites along with *orfB* and *scbA* promoter sites can be bound by ScbR and ScbR2 to repress the expression of their respective genes. ScbR can also bind to the promoter of its own gene.[60,61]

The known regulation of CPK biosynthesis occurs in the following way and is summarised in Fig. 4. At low SCB1 concentrations, ScbR is bound to DNA at the promoter sites of the *cpkO*, *cpkN*, *orfB* and *scbA* genes. When SCB1 concentrations rise within the cell to a certain threshold, SCB1 binds to ScbR which subsequently releases from the DNA ScbR binds to, only to then derepress its own gene. SCB1 binding to ScbR triggers the expression of ScbA and CpkO. Expression of ScbA acts as a positive feedback loop to synthesis more SCB1 whereas increasing levels of CpkO activates CPK biosynthesis. CpkO also directly or indirectly promotes transcription of *scbR2*. ScbR2 may act as an 'off' switch for CPK biosynthesis, as it binds to the promoter sites of *cpkO, cpkN, orfB* and *scbA*. The downregulation of ScbA leads to a drop in SCB1 levels and thereby causes ScbR to repress expression of *cpkO, cpkN, orfB* and *scbA* genes. With the repression of the CPK biosynthesis activators CpkO and CpkN, CPK biosynthesis stops.[60,61]

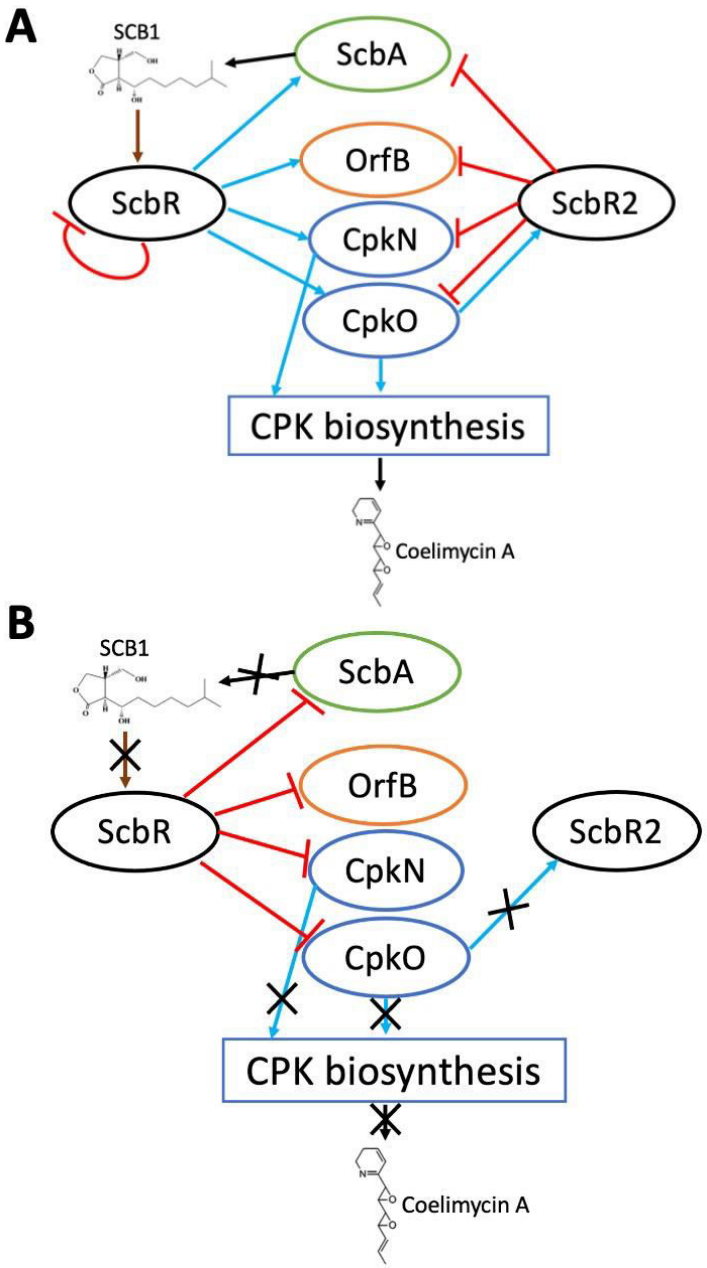


Figure 4. The regulation of coelimycin A (CPK). A) At high levels of SCB1, SCB1 binds to ScbR. This allows expression of the genes for OrfB, CpkN, CpkO, ScbA and depresses expression of the gene for ScbR. CpkN and CpkO act as activators for the biosynthesis of CPK. CpkO also causes expression of the gene for ScbR2. ScbR2 represses expression of genes for CpkO, CpkN, OrfB and ScbA. **B**) Actions by ScbR2 causes ScbA to stop producing SCB1. Low levels of SCB1 means ScbR binds to the promoters for the genes of OrfB, CpkN, CpkO and ScbA. This stops the biosynthesis of CPK and production of ScbR2. Blue arrows indicate activation of expression of genes by proteins indicated. Brown arrow indicates binding of SCB1 to ScbR. Black arrows indicate synthesis of products. Red lines represent repression of genes by the proteins indicated. Black crosses signify those pathways being stopped.

A calcium-dependent antibiotic (CDA) and its BGC was also found in *S. coelicolor* A3(2) from sequencing its genome.[15] It is an acidic lipopeptide (Fig. 2) similar to daptomycin[62] which is effective against Gram-positive bacteria in the presence of calcium ions.[63] This CDA gene cluster is 1% of the entire *S. coelicolor* A3(2) genome[62] and as expected there are many uncharacterised proteins of unknown functions in this BGC with even the complete biosynthesis pathway of CDA still to be solved. There is only one gene within the cluster implied to be part of a CSR which is CdaR. CdaR is a proposed SARP transcriptional activator[64] for the production of CDA and while it itself is suggested to be under transcriptional regulation by global regulators (such as AbsA2[64,65] and AfsQ1[66]), no evidence has yet been shown of the action of CdaR.

Global regulators can affect the production, repression or response of one or more antibiotic pathways in *S. coelicolor* A3(2).[67,68] Furthermore, these can be triggered themselves by cascades sensitive to pH, available nutrients, and other environmental factors as this has all be shown to vary production of antibiotics.[69] Though as highlighted in the 4 pathways for antibiotic production discussed above cross-talk and multiple levels of regulation is not uncommon but while this does provide a layer of complexity, insight into any of the regulatory factors in one of these clusters could be valuable in understanding similar CSRs in others.

1.2.2 Antibiotics as ligands for transcriptional regulators of antibiotic gene clusters in *Streptomyces*

The transcriptional regulators described above bind and unbind to DNA with use of organic molecules acting as ligands. Many of these TetR-like proteins such as ScbR can be regulated by small hormone-like signalling molecules, specifically GBLs and structurally similar molecules to GBLs.[70] Furthermore, these GBLs and GBL-like molecules can have indirect effects on global regulators and initiate cascades that exert their function on pathway specific CSRs.[39] Instead of GBLs it is the role of

antibiotics themselves and antibiotic intermediates that will be discussed as ligands for CSRs in their own regulatory pathway.[39,71,72]

Insight into the role of antibiotics as autoregulators can be gathered from 'atypical response regulators' (ARRs) that are CSRs for antibiotic biosynthesis.[73] ARRs are called as such due to their phosphorylation independent mechanism of action and can be involved in not only antibiotic biosynthesis regulation but also bacterial growth, development, and iron transport regulation.[56,74,75]

One of these ARRs is JadR1 (an OmpR-type regulator) which is a CSR of antibiotic jadomycin B in *Streptomyces venezuelae.* At high concentrations jadomycin B binds to JadR1 and dissociates it from being bound to DNA whereas at low concentrations JadR1 activates the expression of jadomycin B. In this way jadomycin B can function in feedback loops with JadR1 to control its own production.[73] Also in the CSR of jadomycin, is JadR (a TetR-like regulator) which works alongside JadR1 to control jadomycin production. Early jadomycin intermediates bind to JadR which releases it from its binding site of *jadY* which is involved in conversion of late intermediates into jadomycin A.[76,77]

The antibiotic chlorothricin and its intermediates from *Streptomyces antibioticus* can bind to the regulator ChlF1. ChlF1 binds to the promoter sites of genes of other proteins involved in chlorothricin biosynthesis and also the gene (*chlG*) encoding a major facilitator superfamily (MFS) transporter. In this way through ChlF1, not only does chlorothricin monitor the production of itself but also its export.[78]

Similarly, calcimycin biosynthesis in *Streptomyces chartreusis* NRRL 3882 can be regulated by calcimycin itself and its intermediates.[79] These bind to the CSR CalR3 which represses its own gene and also that of *calT* which expresses a transmembrane efflux pump protein. An increase in calcimycin therefore de-represses CalR3 allowing for expression of the protein involved in exporting calcimycin.[80]

Finally, as mentioned earlier in the RED production system, RedZ indirectly regulates RED biosynthesis (and is a direct negative autoregulator). Derepression of RedZ from binding to DNA is through recognition of RED and in this way RED can upregulate its own biosynthesis production.[73]

Clearly some antibiotics (and their intermediates) have an autoregulatory role of their own production and export, through binding to CSRs controlling their BGCs.[38,39]

1.3 The close relation between heavy metal sensing and antibiotic resistance

1.3.1 Antibiotic (self) resistance mechanisms

Antibiotic resistance falls under the umbrella term of antimicrobial resistance (AMR) where pathogens (such as bacteria) are no longer susceptible to drugs (such as antibiotics). AMR is a recognised global health crisis with estimated deaths worldwide in 2019 at around 700,000 and a projected 10 million by 2050.[81] Following these figures from the World Health Organisation (WHO) report, the UK Government set out a 5-year plan from 2019-2024[82] to tackle the issue which includes tactics such as limiting the overuse of antimicrobials, optimising the antimicrobials that are already in use and funding AMR research. Understanding resistance mechanisms can therefore assist in the development and targeting of new antimicrobials and other therapies (e.g., phage) to target drug-resistant pathogens.

Antibiotic resistance specifically, in bacteria can occur in a few ways; inactivating the antibiotic, modifying its target, decreasing uptake, and pumping it out of the cell (Fig. 5).[83,84]

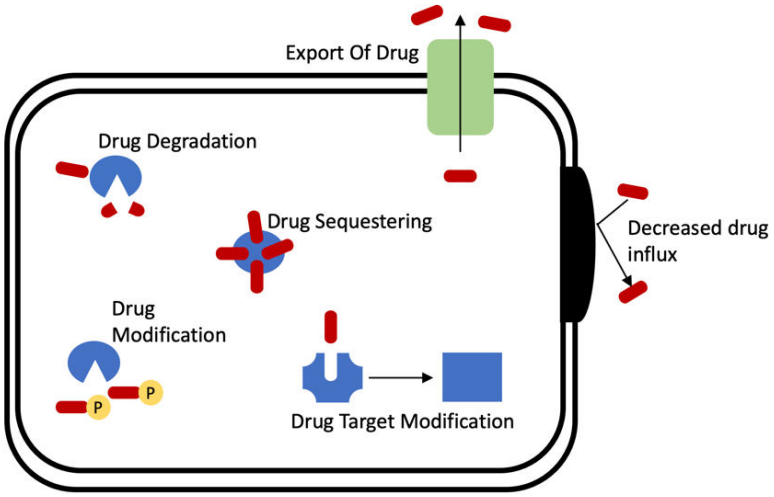


Figure 5. The various methods of antimicrobial resistance to antibiotics (drugs) by bacteria. The main ways that bacteria can acquire antimicrobial resistance to antibiotics is shown in the figure above. This can involve inactivating the drug through drug modification (phosphorylation), degradation and sequestering (such as with penicillin binding proteins). The targets for those drugs can also be modified to render the antibiotic ineffective. The make-up of the cell wall could also be modified or thickened to decrease uptake of the drug. Finally, efflux pumps to export the drugs may also be used to transport the drug out of the cell before it can function.

Inactivation of the antibiotic can be done through modification, degradation or sequestering of the drug (Fig. 5). Antibiotic modification is done through transferase enzymes adding chemical groups onto the drug to cease its efficiency. For example, *S. griseus* which produces streptomycin has self-resistance to it by expressing the enzyme streptomycin 6-phosphotransferase.[84,85] This enzyme adds a phosphate group to the antibiotic inactivating it until export and the removal of the phosphate by StrK phosphatase.[86] Other common modification groups added to antibiotics by enzymes for resistance are acetyl, phosphoryl, and adenyl groups.[87–90] Degradation of the antibiotic can also render it ineffective. The best studied antibiotic degrading enzymes are β-lactamases which degrade the β-lactam ring of antibiotics such as penicillin and cephalosporin. These enzymes use a serine in their active site as a nucleophile or metal ions and water to hydrolyse their target.[90] Despite *Streptomyces* having a wide variety of β-lactamases[91], they are not thought to have the same role in antibiotic degradation in *Streptomyces* as they do in other

antibiotic resistant bacteria.[92] Some *Streptomyces* species that produce the bleomycin family of antibiotics however do sequester the antibiotic as its produced as a means of inactivating it within its own cell.[93] This is done through use of proteins (e.g. BlmA in *Streptomyces verticillus*[94] and ZbmA in *Streptomyces flavoviridi*)[95] that bind to the antibiotic, capturing it and ceasing its activity. Around the time of export of this complex, the binding proteins are thought to be released from the antibiotic again.[96,97]

Target modification of the antibiotic is also a strategy used for self-resistance by many *Streptomyces* species (Fig. 5). As mentioned earlier resistance to β-lactam antibiotics by *Streptomyces* more likely involves the use of penicillin binding proteins (PBPs). PBPs are involved in the construction of the peptidoglycan cell wall and resistance using PBPs can be achieved by changing the number of PBPs that are available for the β-lactam antibiotic to bind to or the structure of the PBP itself.[92,98] Other target sites (e.g. ribosomal subunits, nucleic acid synthesis etc.) can be modified by resistant bacteria to reduce the effect of the antibiotic.[99–101]

Limiting the uptake of antibiotics (Fig. 5) is not a method used by *Streptomyces* species for resistance but are used by species such as *Staphylococcus aureus*. There are natural variations in the ability of bacteria to uptake antibiotics though susceptible bacteria have shown ways to decrease their ability of uptake.[102] Using *S. aureus* as an example, thickening the cell wall has provided the bacteria resistance to vancomycin by increasing the difficulty of the antibiotic to enter. This uses a combined strategy of drug uptake limitation and target modification.[103,104]

Finally, one of the most common mechanisms for antibiotic self-resistance is use of efflux pumps to export the antibiotic out of the cell. There are 5 main efflux pump families in bacteria; the ATP-binding cassette (ABC) family, the multidrug and toxic compound extrusion (MATE) family, the resistance nodulation cell division (RND) family, the small multidrug resistance (SMR) family and the major facilitator superfamily (MFS) (Fig. 6).[105,106] In Gram positive bacteria like *Streptomyces* many of these drug efflux pumps are thought to belong to the MFS family.[107] MFS family

pumps are single module pumps that can transport not just secondary metabolites but also anions and sugars through a cation symport or antiport. While they do seem to be vast in numbers, they can also have substrate specificity.[106,108] For example *Escherichia coli* (*E. coli*) utilises two separate MFS pumps for exporting macrolides (MefB[109]) and fluoroquinolones (QepA[110]).

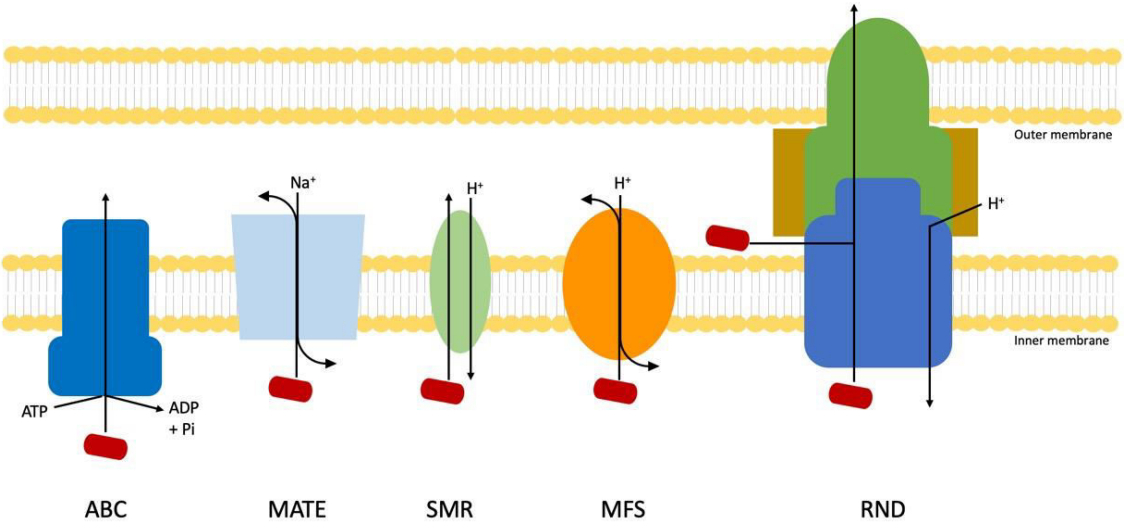


Figure 6. The general architecture of the main families of drug efflux pumps in bacteria. The ATP-binding cassette (ABC) family, the multidrug and toxic compound extrusion (MATE) family, the resistance nodulation cell division (RND) family, the small multidrug resistance (SMR) family and the major facilitator superfamily (MFS) are all families of efflux pumps used by bacteria to export drugs such as antibiotics (shown as a red pill). All but the RND family are generally single-component pumps that export drugs out from the cytoplasm.

In *S. coelicolor* A3(2) genomic analysis studies reveal 11% of the organism's genome to encode probable transporters (800 transporters) and 15% of those transporters are predicted to belong to the MFS family (120 predicted MFS family transporters).[111,112] One of the known MFS family transporters is ActA for export of actinorhodin as mentioned previously.[48]

Also discussed earlier (in Chapter 1.2.2) were other MFS encoding genes (*chlG* and *calT*) that export antibiotics out of *Streptomyces* cells.[78,80] These genes for antibiotic efflux pumps were found within the CSRs of BGCs of antibiotic production for the

specific antibiotic they export. Of those antibiotic efflux pumps characterised the majority have been shown to be under strict transcriptional control of various regulators.[113] The coupling of self-resistance to antibiotic biosynthesis is clearly a logical one as high levels of antibiotics within the cell can be toxic and the role of antibiotics for competition means it is required to be exported.[84]

1.3.2 Antibiotic resistance and heavy metal resistance parallels

Minerals and metals are a component of soil and are required for biological processes.[114] Too much however can be toxic and so bacteria have resistance mechanisms to sense and control the levels of heavy metals in the cell.[115,116] As mentioned earlier, *Streptomyces* are a dominant species of bacteria in the soil[7] and this is especially true for heavy metal contaminated soil areas[117–120]. Therefore, like other soil bacteria they too have BGCs controlling metal sensing and resistance within their cells. These metal sensing proteins tend to largely perform one of two functions: control genes for metal resistance via efflux pumps or storage of metal ions[121], or control genes for metal ion uptake[115] .

The metal sensing proteins that regulate metal resistance act via transcriptional de-repression[122] or (less commonly) activation[123], where metal binding causes the sensor protein bound to DNA to dissociate and allow transcription of the gene it controls, such as that of a metal ion efflux pump.[121] The metal sensing proteins are diverse not just in their architecture but also in the metals they can sense with some able to bind to more than just one type of metal ion.[122] This system of regulation is certainly similar to many employed by bacteria for self-resistance to the antibiotics it produces, and these metal-sensing regulators may be orthologous to antibiotic-sensing proteins or vice versa.[124]

Two of the seven major families of metal sensing transcriptional regulators (ArsR and MerR) like TetR family proteins are one-component regulators.[121] There are exceptions within these families of proteins that are not part of metal ion

regulation. For instance, exceptions within ArsR family proteins such as HlyU[125] are involved in toxin-antitoxin systems while exceptions such as SoxR[126] in the MerR family respond to oxidative stress. SoxR is interesting in that it does bind to metal ions (iron) however the metal binding motif is not necessary for DNA-binding at the *soxS* promoter. Instead, inactive SoxR is bound to DNA while also binding Fe^{2+} and in its active form (i.e., promoting expression of *soxS*) also remains bound to DNA and iron ions. However, the iron ions in SoxR's active form are oxidised from Fe^{2+} to Fe^{3+}. Therefore, the activating ligand for SoxR is not a metal ion but rather a superoxide.[126] Another notable metal sensing transcriptional regulator exception in the MerR family is BmrR.[127] BmrR is a one-component transcriptional activator that binds to certain hydrophobic cations (such as tetraphenylphosphonium and rhodamine) and activates the transcription of a multidrug resistance efflux pump (encoded by the *bmr* gene).[127] Aside from BmrR, the ligands of ArsR and MerR family members that do not bind to metal ions or are involved in metal regulation is unknown and may very well be organic molecules. As of date, there are no known proteins within the major metal sensing transcriptional regulator families that have members which can bind both metal ions and organic molecules.

These metal resistance genes much like antibiotic resistance genes can occur in mobile genetic elements.[128] Mobile genetic elements are genetic regions that can consist of plasmids, transposons and integrons for example and are able to be transferred between organisms of the same or different species. This transfer can occur in several ways though one way is by horizontal gene transfer (HGT) which can include transformation (uptake by cells of mobile genetic elements), transduction (DNA introduced into the cell by viral means) and conjugation (mobile genetic element transfer from one cell to another through a pilus).[129] HGT is a particularly common method for the transfer of antibiotic resistance genes in bacteria.[130] For heavy metal resistance too, HGT plays a major role as metal resistance genes are found in a diverse range of bacteria. Furthermore, these genes have been found on plasmids just as antibiotic resistance genes are and within transposon elements that can contain both antibiotic resistance and metal resistance genes.[131–134] This has in fact, been suggested to be leading to co-

selection of bacteria in heavy metal contaminated environments containing both antibiotic resistance and heavy metal resistance genes on the same mobile genetic elements.[135–137]

Phylogenetic analysis by Álvarez, Catalano and Amoroso (2013)[138], highlight that in *Streptomyces* heavy metal resistance cannot be refined down to a single lineage but are widespread along the *Streptomyces* phylogeny. This indicates HGT of these metal resistance genes occurring, which is unsurprising as HGT is associated with *Streptomyces* for not only adaptive genes but also housekeeping ones.[139] With both heavy metal and antibiotic resistance genes (and their regulators) being subject to HGT so readily there are likely gene products with similarities to both antibiotic and heavy metal regulators thereby acting as an evolutionary bridge between the two systems.

1.4 Methylenomycin biosynthesis and regulation

The antibiotic methylenomycin (Mm) was first characterised not from *S. coelicolor* A3(2) but from *Streptomyces violaceoruber* in two forms (A and B).[140] The B form (MmB) although seen to have antibiotic activity is thought to be an artefact of *in vitro* purification and may be a degradation product of the C form (Fig. 7).[141] Methylenomycin A (MmA) was found to be a cyclopentanone molecule active against both Gram-positive and Gram-negative bacteria at low pH's.[140] In 1976 MmA was identified as a product of the SCP1 plasmid in *S. coelicolor* A3(2)[53] and subsequent cloning studies led to the discovery of its biosynthesis and regulation genes including self-resistance genes.[142,143] This gene cluster contains 21 genes with the resistance genes for MmA flanked by biosynthesis genes (Fig. 8).[142,144]

The exact molecular mode of action of methylenomycins is still unknown. The only and most comprehensive study into the elucidation of the targets of methylenomycins comes from Idowu G. (2017). Using luciferase and β-galactosidase reporter systems containing genes involved in pathways commonly

targeted by antibiotics, isolated and purified methylenomycins were added to take luminescence (in the case of the luciferase reporter system) and blue colour intensity (in the case of the β-galactosidase reporter system) readings. Should the methylenomycin target a particular pathway, the promoter of the gene in the reporter plasmid will be induced and lead to production of the reporter protein. From these assays methylenomycins were not implied to inhibit fatty acid biosynthesis (*fabHB* lack of induction) or prevent DNA replication (*yorB* lack of induction). The *ypuA* biosensor (for inhibition of cell wall biosynthesis) was however strongly induced (stronger than the vancomycin standard) by methylenomycins. Therefore, methylenomycins likely perform by targeting the cell wall and inhibiting the biosynthesis of the cell wall in bacteria.[145]

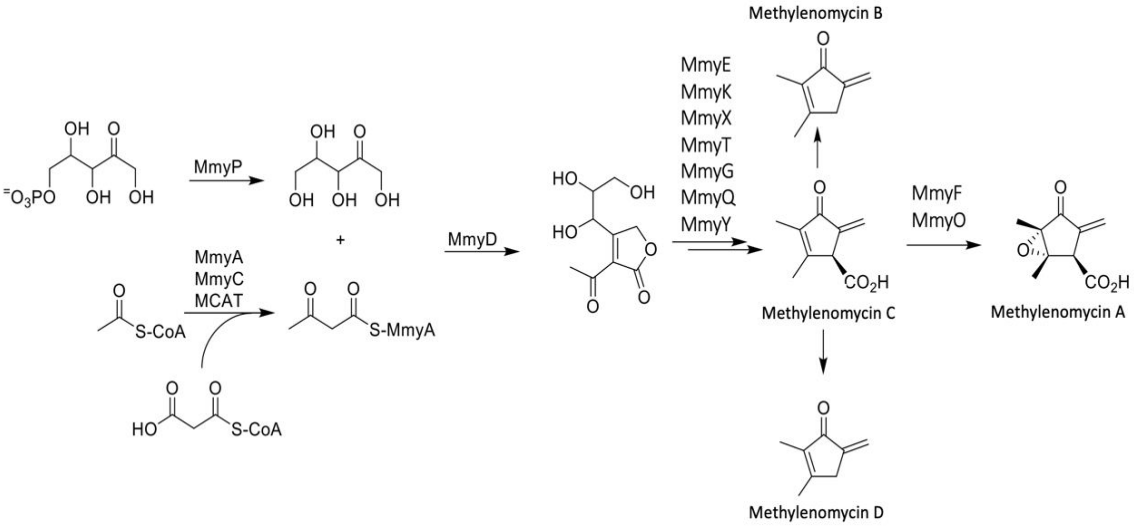


Figure 7. The proposed biosynthesis of the antibiotic methylenomycin A. The proposed biosynthesis of the antibiotic methylenomycin A (MmA). The 'Mmy' proteins catalysing MmA biosynthesis are from the gene cluster itself. Methylenomycin C (MmC) is the precursor to MmA but degrades into Methylenomycin D (MmD) or Methylenomycin B (MmB). Both MmC and MmD have no observed antibiotic activity unlike MmA which is effective against both Gram-positive and Gram-negative bacteria. Figure modified from Idowu G.A. (2017)[145].

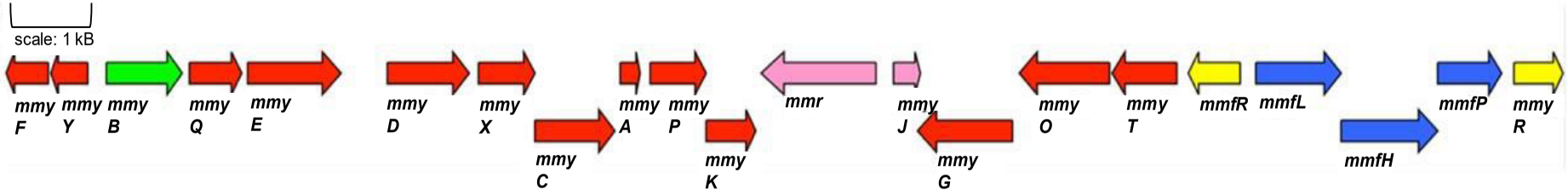


Figure 8. The methylenomycin A gene cluster from *Streptomyces coelicolor* A3(2) SCP1 plasmid. The gene cluster for the regulation and biosynthesis of the antibiotic methylenomycin A (MmA).[41] The genes in red are enzymes involved in MmA biosynthesis (Fig. 7). Those in blue (*mmfLHP*) produce a signalling hormone molecule recognised by the transcriptional regulator MmfR.[148,157] Both yellow genes are transcriptional regulators involved in biosynthesis regulation. The green gene, *mmyB*, is a transcriptional activator for MmA biosynthesis.[144] Pink genes, *mmr* and *mmyJ*, are involved in self-resistance to MmA with *mmr* encoding an MFS efflux pump and *mmyJ* being the transcriptional repressor thought to control *mmr* expression.[137,160] Figure created using Gene Graphics by Harrison *et al.* (2018).[52]

1.4.1 MmA biosynthesis and regulation

While the exact role of MmA has not yet been established a proposed pathway for its biosynthesis has been suggested (Fig. 7).[145,146] The biosynthetic genes for MmA in the cluster (in red in Fig. 8) express proteins that are involved in the process of combining a pentulose compound and an acetoacetyl compound to produce MmA at the end. The condensation of the two compounds is thought to be done by MmyD with subsequent reactions performed by the remaining biosynthetic genes. One of the intermediates observed prior to the epoxidation to form MmA is methylenomycin C (MmC).[146,147] MmC has no observed antimicrobial activity itself[145] and in high enough concentrations can degrade to methylenomycin D (MmD) before it is converted to MmA by MmyF and MmyO (Fig. 8).[146]

MmA biosynthesis is under strict regulation by three transcriptional regulators. One of these is MmyB, the proposed activator for MmA biosynthesis, which will be one of the proteins focused on in this thesis.[148] Sequencing of the SCP1 plasmid and consequently of the MmA BGC led to the discovery of *mmyB*, which indicated a protein containing an N-terminal DNA binding domain nestled within biosynthesis genes as well as highlighting the appearance of the rare TTA codon.[22] As mentioned earlier in the actinorhodin pathway the TTA codon requires the specialised tRNA from the *bldA* gene and mutation of *bldA* eliminated MmA production.[149] The requirement of *bldA* along with regulation on *mmyB* from MmfR and (possibly) MmyR implicate *mmyB* to be an important gene in the MmA pathway. The role of it being the activator for MmA production was largely explored in work by O'Rourke *et al.* (2009).[150]

RNA analysis of transcripts of *mmyB* (and other MmA regulatory genes) were taken by O'Rourke *et al.* (2009) from 12-84 hours. MmA was detectable after 48 hours which was also the first detectable mRNA transcripts of MmA biosynthetic genes (e.g., *mmyQAT*). However *mmyB* was detected at 36 hours indicating it is expressed

prior to MmA production. The TTA codon is also seen within one other MmA BGC gene- *mmfL*, which is discussed further below as being involved in the biosynthesis of a small signalling molecule.[151] The TTA codon in *mmyB* and *mmfL* was changed to CTC by O'Rourke *et al.* (2009) in *bldA S. coelicolor* mutants containing the SCP1 plasmid integrated into its chromosome. Combinations of $mmyB^{TTA}$, $mmyB^{CTC}$, $mmfL^{TTA}$ and $mmfL^{CTC}$ were tested for MmA production and as expected both codon conversions restored MmA production. Although neither of the singular codon conversions alone saw production of MmA, when $mmyB^{CTC}$ was grown next to a strain able to produce the signalling molecule *mmfL* would be involved in synthesising, MmA was detected. This highlights the importance of *mmyB* for MmA production and strongly implies it to be an activator of MmA biosynthesis.

The two other transcriptional MmA biosynthesis regulators are MmfR and MmyR (Fig. 9).

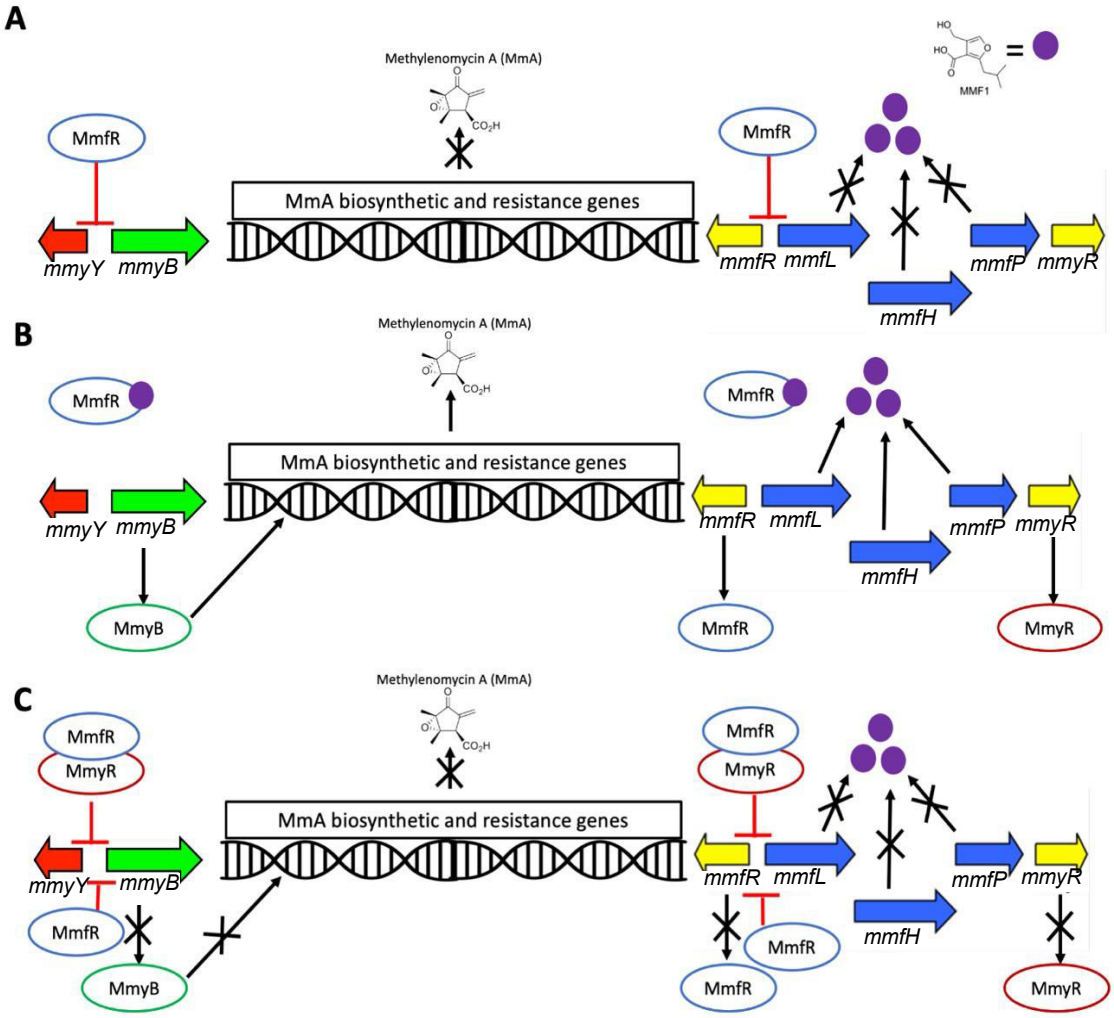


Figure 9. The regulation of methylenomycin A (MmA) from *Streptomyces coelicolor* A3(2) by MmfR and MmyR. There are 3 transcriptional regulatory proteins involved in the regulation of the antibiotic MmA; MmyB, MmfR and MmyR. MmyB is a transcriptional activator of the MmA biosynthesis that is not well characterised. Both MmfR and MmyR are transcriptional repressor proteins. **A**) Before activation of MmA biosynthesis, MmfR binds to promoter regions between *mmfR-mmfL* and *mmyY-mmyB* preventing transcription of divergent genes. **B**) When methylenomycinfurans (MMFs-purple circles) are detected in the cell at a certain threshold, they bind to MmfR and cause MmfR to release from DNA. This allows expression of *mmfLHP* which are involved in biosynthesis of MMFs, *mmfR* expression and *mmyB* expression. MmyB activates the biosynthetic genes for MmA production. MmyR is also produced. **C**) MmyR may form heterodimers with MmfR to bind to DNA at the *mmfR-mmfL* and *mmyY-mmyB* promoter regions to prevent production of MmyB, MMFs and switch 'off' MmA biosynthesis. MmfR unbound to MMFs or MmyR may also bind to DNA.

MmfR is a TetR like repressor[148] akin to the regulator ScbR[152] (Chapter 1.2.1). It too recognises small hormone molecules except in the case of MmfR this is not a GBL type molecule like SCB1 but methylenomycinfurans (MMFs)[153] which belong to the 2- alkyl-4-hydroxymethylfuran-3-carboxylic acids (AHFCAs) class of hormones[151]. When a certain threshold is reached of MMFs, they bind to MmfR and release it from being bound to DNA. The DNA recognition motifs to which MmfR binds are termed MARE (MMF-autoregulator responsive element) sites which are 18 base pair (bp) pseudo-palindromic sequences. MmfR functions as 2 homodimers with DNA clamped between them as both pairs bind to the double stranded DNA.[148,153] These MARE sites that MmfR binds to are found between *mmyY-mmyB* genes and *mmfR-mmfL*. Derepression of MmfR allows expression of the *mmyB*, *mmyY, mmfLHP* and *mmfR* genes. MMFs are produced by the *mmfPHL* genes (blue in Fig. 8).[151] The expression of *mmyB* leads to the activation of MmA biosynthesis while MMFs and MmfR keeps being produced in a feedforward type reaction (Fig. 9). [148]

The role of MmyR, the other TetR like transcriptional regulator in the MmA BGC is much less understood than MmfR. From studies of deletion of *mmyR*, the role of MmyR has been suggested to be the 'off' switch in MmA biosynthesis (Fig. 9). MmyR has strong DNA binding affinities to the promoter sites of the genes to *mmfR, mmfLHP* and *mmyB*. Again, genetic studies suggest MmyR acts as a heterodimer with MmfR however since MmfR has been shown to function as homodimers, this may also be the case with MmyR. Regardless it has been shown MmyR bound to DNA can prevent the synthesis of MmA and MMFs.[150]

1.4.2 MmA resistance and regulation

Along with biosynthesis, the cell needs to have a self-resistance strategy. In the SCP1 plasmid for MmA this is within the BGC itself. As discussed earlier, efflux pumps are a common mechanism of self-resistance (Chapter 1.3.1). The gene *mmr* in the methylenomycin gene cluster encodes for an MFS efflux pump.[15,144] Much like in the actinorhodin pathway with the predicted MFS efflux pump encoded by *actA* being controlled by the regulator ActR[49], in the methylenomycin pathway *mmr* is hypothesised to be controlled by the regulator MmyJ[148,154] (Fig. 10).

The gene for *mmyJ* is divergent to that of *mmr* and separated by a 218 bp region within which the DNA binding site(s) for the MmyJ protein is predicted to exist.[15] The action of MmyJ on *mmr* expression therefore may be akin to that of ActR on *actA*[49], where MmyJ is a transcriptional repressor for self-resistance to methylenomycin through regulation of *mmr* (Fig. 10).[148,154]

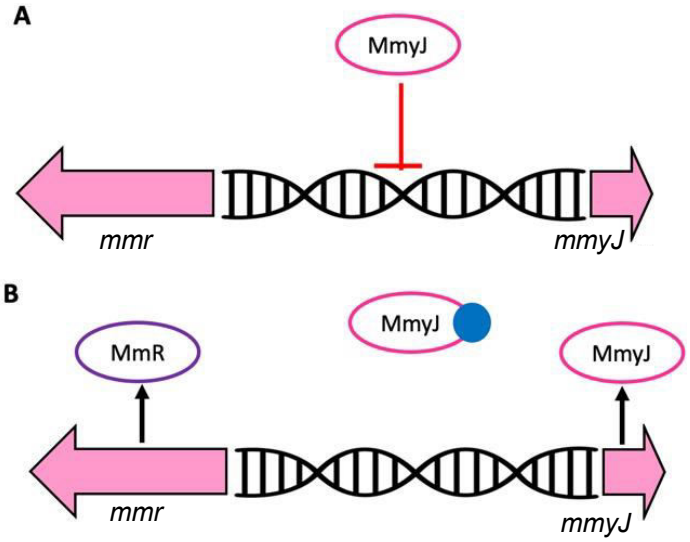


Figure 10. A simple one-component regulation system for methylenomycin A (MmA) self-resistance in *Streptomyces coelicolor* A3(2). The MmA biosynthetic gene cluster (BGC) contains two genes (*mmr* and *mmyJ*) involved in self-resistance to MmA. MmyJ is a transcriptional repressor protein that when MmA biosynthesis is off binds to unknown promoter sites between *mmyJ* and *mmr* (**A**). **B)** When MmA biosynthesis is activated either MmA or an intermediate in MmA biosynthesis likely binds to MmyJ and releases it from binding to DNA. This allows expression of *mmyJ* and *mmr*. MmR is the efflux pump that exports MmA. Fall in MmA or MmA intermediates, causes MmyJ to bind to DNA again switching 'off' self-resistance.

Information on Mmr itself is limited though its expression has been shown to only begin once there is production of MmA, possibly by MmyJ binding to MmA or an intermediate of MmA biosynthesis in order to de-repress from DNA binding and permit production of the Mmr efflux pump for methylenomycin resistance.[154] Due to the *mmyJ* and *mmr* genes being beside one another, their -35 and -10 promoter sites are also in close proximity. Therefore, MmyJ may bind at one or two sites in this intergenic region to control both *mmr* and its own expression. In this way MmyJ may be self-regulatory and self-resistance regulatory as observed with other antibiotic sensing regulators such as ChlF1[78] and CalR3[80] (Chapter 1.2.2).

Preliminary data on MmyJ suggest it is not a TetR or SARP like protein as seen in other antibiotic sensing and self-resistance CSRs but rather, MmyJ is possibly an ArsR-like protein.[155] ArsR's are typically metal regulatory proteins and as discussed previously (Chapter 1.3.2) there are similarities between antibiotic resistance and heavy metal resistance mechanisms and a high probability of co-evolution. MmyJ, if it is an ArsR like protein, may be evidence of this co-evolution.[156] Characterising MmyJ would therefore be beneficial in understanding regulation of antibiotic resistance but also in bridging gaps between heavy-metal resistance and antibiotic resistance.

1.5 Project aims and objectives

The main aims of this project (in short) were to characterise the transcriptional activator controlling methylenomycin biosynthesis production (MmyB) and the transcriptional repressor controlling methylenomycin resistance (MmyJ) to aid in the understanding of antibiotic production and self-resistance.

1.5.1 Characterising MmyJ

The key questions and objectives attempted to be addressed in this project in characterising MmyJ were:

1) Is MmyJ part of the ArsR family of metal sensing proteins despite being in an antibiotic BGC?

Initial data and bioinformatic work done previously[155] suggested MmyJ may function in a similar manner to those of the ArsR family proteins. This is unusual as the ArsR family are typically metal sensing proteins and MmyJ has not been shown to bind to metal ions but is instead likely to bind to MmA (an antibiotic).[155] The aim therefore was to solve MmyJ's structure and analyse MmyJ's functioning in order to gather viable evidence of it being an ArsR family protein or locate the family of protein that it may be a part of. This aim was met using X-ray crystallography, assays and bioinformatic studies. Meeting this aim highlighted the diversity within the ArsR family of proteins and eluded to a subclass within this family.

2) Is MmyJ a transcriptional repressor of *mmr* and itself?

The *mmyJ* and *mmr* genes are opposite one another with a 218 bp intergenic region between them containing -35 and -10 promoter sites for both genes. DNA fingerprinting done by Neal & Chater (1991) suggested MmyJ does bind within this region though the exact binding sites are unknown.[144] Thus, the aim here was to determine the exact binding site(s) of MmyJ and any stoichiometric details of this relationship. Achieving this eluded to similar MmyJ-like systems in other *Streptomyces* species. This aim was met using DNA-Protein assays and *in vivo* reporter systems.

3) Does MmyJ function in a multimeric form?

Many DNA binding proteins function as multimers and indeed all ArsR family proteins characterised to date are shown to function in a dimeric form. The aim here was to determine if MmyJ also forms dimers or multimers in order to bind DNA/ligand for functioning. This aim was met by analysing MmyJ functioning under native conditions and performing knock out mutations of residues involved in multimer formation. Understanding MmyJ functioning assisted in classifying it and understanding its role in the wider context of the cell.

4) Is the ligand for release of DNA-bound MmyJ the MmA antibiotic?

Transcriptional repressors unbind from DNA once a ligand binds to it. Initial work by Lougher M. (2015) suggested the ligand that binds to MmyJ is a methylenomycin molecule.[155] The aim was to verify this and determine which methylenomycin molecule(s) (MmA, MmB, MmC or MmD) bind to MmyJ and determine any details of this binding. This aim was met by conducting DNA-MmyJ-Ligand assays and using *in vivo* reporter systems. Determining this helps in understanding MmyJ role in the cell and adding to the body of knowledge on how antibiotic resistance pumps are regulated.

1.5.2 Characterising MmyB

The characterisation of MmyB were slightly more limited with fewer aims to comply with the time restraints of the project and were as follows:

1) Does the structure of MmyB offer insight into its functionality?

The aim here was to solve the structure of MmyB so that insight into its functioning and classification can be made. By using similar methods to MmyJ, X-ray crystallography of recombinant MmyB was attempted for structural data. Continuing on from this, a second aim was to make predictions based on MmyB's

structure as to the ligand it may bind to. This can aid in better understanding the functioning of MmyB and its family of proteins.

2) Is MmyB a transcriptional regulator?

As discussed earlier (Chapter 1.4.1) MmyB is predicted to be the transcriptional activator for MmA biosynthesis.[150] The aim therefore was to determine whether MmyB does indeed perform as a transcriptional activator and bind to DNA. If so which region or sequence of DNA, does it bind to and thus which gene(s) does it regulate? To achieve this aim, structural data insights in combination with DNA and/or Ligand-MmyB assays was conducted. Again, achieving this aim adds knowledge to the ligands that can activate antibiotic biosynthesis and functioning of this in MmyB-like systems.

3) Can the regulatory system of *mmyB* be bypassed for MmA biosynthesis?

The MmfR transcriptional regulator within the methylenomycin biosynthesis gene cluster has been well characterised by Zhou *et al.* (2021) and shown to control the biosynthesis of *mmyb*.[153] Following on from this, the aim is to determine whether this regulation upon *mmyb* can be bypassed for MmA biosynthesis. This was achieved by detection of MmA levels in cells where the MmfR regulatory system upon *mmyb* is non-functional. Determining this aim meant a similar approach can be employed to allow for bypassing of MmyB-like biosynthesis systems to aid the discovery of natural products in *Streptomyces* and other bacteria.

Chapter 2: Materials & Methods

2.1 Bioinformatics

2.1.1 Creating a phylogenetic tree

Protein sequences from NCBI were first gathered and alignments of these made using MUSCLE.[157] The aligned file was downloaded and input into the online webserver of IQ-TREE.[158] For both phylogenetic trees generated, IQ-TREE parameters entered were kept as defined on the server (auto detected substitution model, ultrafast bootstrap analysis, 99% minimum correlation coefficients and 1000 maximum iterations for branches). The best fit model chosen by IQ-TREE for Fig. 14 of all MmyJ with other ArsR family proteins was that of LG and G4 models[159,160] according to the lowest Bayesian Information Criterion (BIC) values[158]. Likewise for Fig. 15 of MmyJ and MmyJ-like proteins, IQ-TREE according to the lowest BIC values used a best fit model of JTT, F, I and G4[159–161]. The tree file generated from IQ-TREE was visualised using the online web servers Interactive Tree Of Life (iTOL)[162] and Environment for Tree Exploration (ETE)[163]. The ETE visualises the phylogenetic tree along with the aligned sequence data beside it and was used in Fig. 14, whereas Fig. 15 only has the ETE to visualise the rooted phylogenetic tree. The red numbers beside each node on the tree from ETE represents a measure of support for the node from 0 to 100 with 100 being the maximum support as computed by the bootstrapping measurements from IQ-TREE. The iTOL tree figure is unrooted (where the common ancestor is unknown) which even though is difficult to judge precise relationships between tips was given for comparison to the rooted tree. The scales in Fig. 14 and Fig. 15 indicates the branch lengths which show the evolutionary lineages changing over time. Therefore, the longer these are the greater the amount of change. This change is characterised by the number on the scale which indicates the number of amino acid substitutions per branch length.

2.1.2 X-ray crystallography data collection and structure solving

Crystals of C49S MmyJ (Materials & Methods 2.5.9) were harvested by Dr Crow (University of Warwick) in litholoops and plunge-frozen in liquid nitrogen before it was sent to the synchrotron facility (Diamond Light Source, Oxford, UK) for diffraction and analysis. Crystals of wild type (WT) MmyJ were harvested and initially processed by Professor Vilmos Fulop.[155]

Indexing and integration of reflections of the C49S MmyJ crystal was autoprocessed by the Diamond facility itself using the xia2 system[164] which utilised iMosflm[165]. This file was downloaded for further structure determination using the Collaborative Computational Project Number 4 (CCP4) Suite[166] . The CCP4 Suite was also used for structure determination from the partially processed WT MmyJ file.

For the C49S MmyJ file, 'Aimless'[166] was first used to scale and merge the integrated and indexed data. The signal to noise ratio, completeness of the data set and *R* values were all considered as quality checks prior to proceeding. Molecular replacement was used to solve the phase problem by the rendition of a model structure. This model structure of C49S MmyJ was created by Dr Crow using AlphaFold 2[167]. Phaser[168] was the CCP4 Suite programme used for molecular replacement. Molecular replacement for WT MmyJ was done using the finished C49S MmyJ structure created as a model. Both C49S and WT MmyJ structures were then built iteratively using Coot[169] and Refmac5[170] with waters and any other distinguishable ions (i.e., chloride) also fitted. Once *R* values were low enough to indicate a suitable refinement level, validation of the models were conducted.

Rampage[171] from CCP4 Suite was used to create Ramachandran plots[172] of structures while Procheck[173] assessed the stereochemistry. Finally, structures were passed through the web server MolProbity[174] for final validation assessments before being determined as valid. The details and results of this computational structure solving of MmyJ are detailed more comprehensively in Chapter 4.

2.2 Materials for experimental work

2.2.1 Purchased oligonucleotides

Table 1: Purchased primers. Primers purchased from Integrated DNA Technologies (IDT) for various methods as described.

Primer Pair Number	Method used for	Description	Forward (5' – 3')	Reverse Complement (5' – 3')
1	Sanger Sequencing	T7	TAATACGACTCACTATAGGG	GCTAGTTATTGCTCAGCGG
2	PCR	*mmyB* Gene	CACCATGGCTAGCGTCGATGGC	ACGGGATCAGGGATGGATG
3	PCR	*mmyJ* Gene	CACCGTGGCGGCACGGATCACG	TGGAGCGCCGTGCGGGCG
4	PCR	80 bp intergenic	GGTACGACAGCCATCAAATG	TACCGTGCGTTCTCACTGAA
5	Sanger Sequencing	pJ251 Tinsel Purple+ Anderson Strong	GACCCCATGCCGAACTCAG	CACGCAGAACATGCGCATTC
6	Double-stranded DNA	Anderson strong promoter	ATCGGATCCTTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGCGGATCCATC	GATGGATCCGCTAGCACTGTACCTAGGACTGAGCTAGCCGTCAAGGATCCGAT
7	MmyB-DNA EMSA	Bbox fragment	CTTCTGCTAGGGGCCTCAGACGTGGTGTTCTCAGCACCAGGGCCCCGCCGAGTTGCGGTCCAAACACCGAGGCCCCGCCGCTCTCGTAAG	CTTACGAGAGCGGCGGGGCCTCGGTGTTTGGACCGCAACTCGGCGGGGCCCTGGTGCTGAGAACACCACGTCTGAGGCCCCTAGCAGAAG
8	40 bp fragments of DNA for MmyJ-DNA EMSAs	1-40 region	GGTACGACAGCCATCAAATGTTGACGGCCATCAAAGTTTG	CAAACTTTGATGGCCGTCAACATTTGATGGCTGTCGTACC
9		41-80 region	ACAGCCGTCGTCATATGAGCTTCAGTGAGA	TACCGTGCGTTCTCACTGAAGCTCATATGACGA

			ACGCACGGTA	CGGCTGT
10		20-59 region	GTTGACGGCCATCAAAGTTTGACAGCCGTCGTCATATGAG	CTCATATGACGACGGCTGTCAAACTTTGATGGCCGTCAAC
11	15 bp fragments of DNA for MmyJ-DNA EMSAs	1-15 region	GGTACGACAGCCATC	GATGGCTGTCGTACC
12		8-22 region	CAGCCATCAAATGTT	AACATTTGATGGCTG
13		16-30 region	AAATGTTGACGGCCA	TGGCCGTCAACATTT
14		23-37 region	GACGGCCATCAAAGT	ACTTTGATGGCCGTC
15	Predicted MmyJ binding DNA region from MmyJ structure	Predicted Region 1	CCATCAAATGTTGACGG	CCGTCAACATTTGATGG
16		Control Shuffle of Predicted region 1	GGCTCACAATGTAAGTC	GACTTACATTGTGAGCC
17		Predicted Region 2	CCATCAAAGTTTGACAG	CTGTCAAACTTTGATGG
18		Control Shuffle of Predicted region 2	ATAGACGATCTGACATC	GATGTCAGATCGTCTAT
19		Full Predicted Region	CCATCAAATGTTGACGGCCATCAAAGTTTGACAG	CTGTCAAACTTTGATGGCCGTCAACATTTGATGG
20		Control Shuffle of Full Predicted Region	AAGCAAGTGGCTTCCTTACTATACATCGACAGGA	TCCTGTCGATGTATAGTAAGGAAGCCACTTGCTT
21	T18-MmyJ Mutagenesis	Q22A	CGCCATCACCCTCGCGGGCGTCCTGGAC	GTCCAGGACGCCCGCGAGGGTGATGGCG
22		L25E	CCCTCCAGGGCGTCGAAGACGCGCTGGTCGA	TCGACCAGCGCGTCTTCGACGCCCTGGAGGG
23		V29E	CTGGACGCGCTGGAAGATCCGGTGCGCC	GGCGCACCGGATCTTCCAGCGCGTCCAG
24		R34A	CGATCCGGTGCGCGCGAGCATCGTCCGGC	GCCGGACGATGCTCGCGCGCACCGGATCG

All Primers were purchased from IDT and purchased dry to be reconstituted with nuclease free water to 100 µM.

The following codon optimised *mmyB* gene was also purchased as a double stranded DNA fragment (g-block) from Genewiz:

5’ – CACC

ATGGCTAAAGACTCAACCGTTTGCTCCCCCAAACGCCAGCGCGAGGCGCTTCGCCATTTTTTACGTTCAC
GCCGTGCCCGCCTGTCTCCTGATGATGTAGGCCTTTTAGCTACTGGGCGTCGCCACACTCCTGGATTACG
TCGTGAGGAAGTGGCAGTGATTGCGGGCGTCTCCGCGTCCTGGTATACGTGGTTAGAACAGGGCCGCG
ATATTAAAGTTTCAGACGGAGTGCTGAATGCCATTAGTCAGGCCCTGCGTTTAGACGACACAGAGCGTG
CCCACCTGTATCGCCTGGCGGGAGTGAATCCTCCGCAGTCCGTGCCCGCAACTGCTGGTCAGACAGAAA
CCTCCCGCTTACAGTTGATCGTGGATGGATGGCTTCCTGCCCCAGCGTTTGTAGTGGATCGTTATTGGAA
CACGCTTGCCGCTAACCAGGCCGCACGTTCAGCGCTGGGTGTCGGGGCTGGAGACCAAAATTACCTTGC
TGCCTTCTTTACTGAACCTACTGCTCGTGCTCGCTATTTGGATTGGGACAAATTAGCTACCCGTCTGGTCG
GACAATTTCGCGTCCAAGCGGCACGCTTTCCCGAGGATCCGCGCTTCGATCGCATCGCGCGCCAGTTATG
CGCAACCGATCATGCGTTTGCTGACTTATGGGCACGCCATGAAACATGTGACACCGCTATGACTTCGGTT
CGTGTGCGCCCACCGGGTGAAGAATCAATGCGCTTCGAGCACCTTATTCTTGCACTGTTGGAGAACGCC
GACCTTCGCTTGATGTTATATATGCCACGCGGAGCCCATATCAGTGCAGAGGCGGGTTTACGTCTGGTCC
CTGCGGACCCACTGACCCGCAATTCTTTACTTCAATTGCCCGGACGTCCCATGACTTCACGCCCCGTTGAA
CGTCATATCCACCCATAATAA -3’

2.2.2 Plasmids and Strains

Table 2: Plasmids and strains. The plasmids and *Streptomyces* strains used for studies.

Plasmids				
Number	Name	Backbone	*E. coli* expression host	Reference
1	*mmyB*	pET151	TOP10	-
2	*mmyB* codon optimised	pET151	TOP10, BL21, C43	-
3	Tinsel Purple	pJ251	TOP10	Lougher M. (2015)[155]
4	Tinsel Purple with Anderson Strong	pJ251	TOP10	-
5	mScarlet-I	custom	TOP10	Capel P. (unpublished)
6	mScarlet with intergenic *mmyJ* region	mScarlet-I	TOP10	-
7	C49S MmyJ	pET151	TOP10, BL21	Lougher M. (2015)[155]
8	MmyJ	pET151	TOP10, BL21	Lougher M.

				(2015)[155]
9	T18-MmyJ	pUT18	BTH101[175]	Cook J. (Crow Lab, University of Warwick)
10	MmyJ-T18	pU18	BTH101[175]	
11	T25-MmyJ	pKNT25	BTH101[175]	
12	MmyJ-T25	pKNT25	BTH101[175]	
13	T18-Q22A MmyJ	T18-MmyJ	DH5α	-
14	T18-L25E MmyJ	T18-MmyJ	DH5α	-
15	T18-V29E MmyJ	T18-MmyJ	DH5α	-
16	T18-R34A MmyJ	T18-MmyJ	DH5α	-
Cosmid				
Number	Name		*E. coli* expression host	Reference
9	C73-787		TOP10	Corre *et al*. (2005)[146]
Streptomyces coelicolor strains				
Number	Name			Reference
10	W89			Corre *et al*. (2008)[151]
11	W108			Corre *et al*. (2008)[151]
12	M145			Kieser *et al.* (2000)[176]
13	W81			Corre *et al*. (2005)[146]
14	W81+pJH2			Aldali J. (2018)[177]

2.2.3 Buffers for protein purification work

Table 3: MmyJ Buffers. Buffers used for MmyJ purification and storage

Buffer	Components	pH
Wash	50 mM Hepes, 300 mM NaCl, 30 mM Imidazole, 10 % Glycerol	7.67
Gradient Imidazole	50 mM Hepes, 300 mM NaCl, 30-100 mM Imidazole, 10 % Glycerol	7.67
Elution	50 mM Hepes, 300 mM NaCl, 250 mM Imidazole, 10 % Glycerol	7.67
Final	30 mM Hepes, 50 mM Hepes, 5 % Glycerol	7.67

Table 4: MmyB Buffers. Buffers used for MmyB purification and storage.

Buffer	Components	pH
Wash	50 mM Hepes, 300 mM NaCl, 30 mM Imidazole, 10 % Glycerol	7.60
ATP Wash	100 mM KCl, 20 mM $MgCl_2$, 500 mM Sucrose, 5 mM ATP, 100 mM Tris, 10 % Glycerol	
Elution	50 mM Hepes, 300 mM NaCl, 250 mM Imidazole, 10 % Glycerol	7.60
Final	30 mM Hepes, 50 mM Hepes, 5 % Glycerol	7.60

2.2.4 Growth media

Luria-Bertani (LB) broth and agar plates were prepared and made available by the department of Life Sciences, University of Warwick.

Supplemented Minimal Media (SMM)[178] liquid and solid were made in the following way for a 1 L solution. All the components but the glucose (which was filter-sterilised) were autoclaved and added in a sterile environment just before use:

Casaminoacids	2 g
TES Buffer	8.68 g
Agar (only for solid media)	15 g
50 mM NaH_2PO_4 and K_2HPO_4 solution	10 mL
1 M $MgSO_4$	5 mL
50 % w/v glucose	18 mL
Trace element solution (below)	1 mL
dH_2O	Up to 1 L

Trace element solution was made containing the following metals, filter sterilised and autoclaved for use with SMM and R2YE media: $ZnSO_4$, $FeSO_4$, $MnCl_2$, $CaCl_2$, NaCl at 0.1 g/L each.

R2YE media[176,179] (solid and liquid) was made with the following components first added and autoclaved:

Sucrose	103 g
K_2SO_4	0.25 g
$MgCl_2.6H_2O$	10.12 g
Glucose	10 g
Casaminoacids	0.1 g
Agar (only for solid media)	22 g
dH_2O	Up to 800 mL

Then to 80 mL of this, separately autoclaved components of the following solutions were added just before use:

10 % yeast extract	5 mL
0.5 % w/v KH_2PO_4	1 mL
3.68 % w/v $CaCl_2.2H_2O$	8 mL
20% w/v L-proline	1.5 mL
5.73 % w/v TES Buffer pH 7.2	10 mL
Trace Element Solution	0.2 mL
1 M NaOH	0.5 mL

Lambda Broth[180] was made by autoclaving 10 g tryptone and 2.5 g NaCl with water up to 1 L. 15 g of agar was added for solid media of Lambda Broth.

Similarly, Superbroth[180] was made by adding the following together and autoclaving before use:

Tryptone	32 g
Yeast Extract	20 g
NaCl	5 g
(Agar for solid media)	15 g
1 M NaOH	5 mL
dH_2O	Up to 1 L

TYGPN[180] was made by adding the following components and autoclaving before use:

Tryptone	20 g
Yeast Extract	10 g
80 % Glycerol	10 mL
Na_2HPO_4	0.5 g
KNO_3	1 g
(Agar for solid media)	15 g
dH_2O	Up to 1 L

M9[180] media was prepared by adding the following components and autoclaved before use:

Na_2HPO_4	6 g
KH_2PO_4	3 g
NH_4Cl	1 g
NaCl	0.5 g
(Agar for solid media)	15 g
dH_2O	Up to 1 L

M63[180] media was created by adding the following components, adjust to pH 7 with KOH and autoclaved before use:

$(NH_4)_2SO_4$	2 g
KH_2PO_4	13.6 g
$FeSO_4.7H_2O$	0.5 mg
(Agar for solid media)	15 g
dH_2O	Up to 1 L

2.3 DNA and cloning work

2.3.1 Producing the intergenic 80 base pair region between *mmyJ* and *mmr*

The C73-787 cosmid (9 in Table 2) was used as the template DNA to generate the double stranded 80 base pair (bp) intergenic region of DNA between *mmr* and *mmyJ*. The PCR reaction mix (Table 5) was scaled up to 500 µl total (reaction mix described times 10) using the 80 bp intergenic region primers (4 from Table 1). Q5 buffers and polymerase were from NEB (NEB Q5® High-Fidelity DNA Polymerase).

Table 5: Q5 PCR Reaction mix. The PCR reaction mix for Q5 Polymerase reaction (NEB Protocol and reagents).

5X Q5 Reaction Buffer	10 µl
5X Q5 High GC Enhancer	10 µl
10 mM dNTPs	1 µl
10 µl Forward Primer	2.5 µl
10 µl Reverse Primer	2.5 µl
Template DNA (~ 100 ng/µl)	2 µl
Q5 High-Fidelity DNA Polymerase	0.5 µl
Nuclease-Free Water	21.5 µl

This PCR was conducted using an Eppendorf Mastercycler Nexus Gradient machine

98 °C	30 secs	
98 °C	10 secs	x35
54 °C - 60.9 °C	30 secs	
72 °C	30 secs	
72 °C	2 mins	

with the following protocol:

To show the PCR of this 80 bp DNA fragment was amplified a sample of 5 µl was taken and 3 µl of purple loading dye (NEB- Gel loading dye, purple (6X)) was added to it. A 4% agarose-TAE gel was made mixed with 4 µl of GelRed Nucleic Acid Stain 10000X to which the sample along with ladder (low molecular weight DNA Ladder-NEB) was added and was run at 100 V for 1 hour in 1X TAE buffer. Results were visualised by UV using the Syngene U:Genius 3 machine. After confirmation of the correct product, the PCR reaction mix was purified.

PCR purification was conducted with the QIAquick PCR Purification Kit by Qiagen following the instructions given in the kit.

2.3.2 Preparing short double stranded DNA

Shorter double stranded DNA were used for electromobility shift assays (EMSAs) to test for protein-DNA binding. These were prepared by purchasing primers and annealing the two strands together. This was done by mixing equal concentrations of each primer and heating at 94 ℃ (using a Star Lab Smart Instruments Mini Dry Bath) for 5 minutes and leaving overnight (~16 hours) on benchtop to cool. 'Forward' and 'Reverse complement' of primers 7-20 from Table 1 were annealed together to make double stranded DNA fragment for these EMSA studies.

2.3.3 Amplifying *mmyB* from the C73-787 cosmid

The C73-787 cosmid was used as a template to amplify the *mmyB* gene. Primers (2 from Table 1) were used with Q5 Polymerase (NEB) to make the PCR reaction mix (including a blank containing water rather than template DNA) describe in Table 5.

A gradient PCR from 65 ℃ - 73 ℃ was performed using an Eppendorf Mastercycler

98 ℃	30 secs	
98 ℃	10 secs	⎫
65 ℃ - 73 ℃	30 secs	⎬ x30
72 ℃	30 secs	⎭
72 ℃	2 mins	

Nexus Gradient in the following cycle:

5 µl of the samples were taken and 3 µl of purple loading dye (NEB- Gel loading dye, purple (6X)) was added to it. This was run on a 1% agarose-TAE gel mixed with 4 µl of GelRed Nucleic Acid Stain 10000X in 1X TAE buffer at 100 V for 1 hour with a 1 kbp DNA ladder (Quick-Load 1 kb DNA Ladder, NEB). Results were visualised by UV using the Syngene U:Genius 3 machine.

Reactions at 65 ℃, 65.75 ℃, 68.75 ℃, 72.25 ℃ and 73 ℃ (lanes 1,2,6, 11 and 12 of the cycle) were pooled together and a PCR clean up performed. This was done using a QIAquick PCR Purification Kit by Qiagen following the instructions given in the kit. This plasmid DNA was then used for cloning into the pET151 vector (detailed below) for protein expression.

2.3.4 Cloning *mmyB* into a pET151 vector

Both the PCR amplified pure *mmyB* as well as the purchased (IDT) *E. coli* codon optimised *mmyB* (~ 100 ng/µl) were cloned into a pET151 vector for protein expression. These genes have a CACC tag at the start required for this cloning. The Champion™ pET151 Directional TOPO™ Expression Kit from ThermoFisher Scientific was used for this which contained the salt solution and pET151 TOPO vector. The reaction mix detailed in Table 6 was made and stirred gently before incubating for 5 minutes at room temperature.

Table 6: TOPO cloning reaction mix. The pET151 TOPO Cloning reaction mix (from Invitrogen).

mmyB gene	2 µl
Salt Solution	1 µl
TOPO vector	1 µl
Nuclease free water	2 µl

These plasmids were transformed into *E. coli* TOP10 cells (Chapter 2.4.1) and spread onto an ampicillin resistant LB agar plate which was incubated at 37 ℃ overnight. Colonies were selected to be grown in 10 mL LB with 100 µg/mL ampicillin overnight, shaking (180 rpm) at 37 ℃. This was used to make stock colonies (Chapter 2.4.2) as well as for plasmid extraction and sending for Sanger sequencing (Chapter 2.4.3).

The pET151 with *mmyB* PCR purified from C73-787, Sanger sequencing alone was not sufficient for determination of gene insertion. Restriction digestion with EcoRV-

HF (NEB) was performed using NEB restriction enzyme buffers to make the reaction mix detailed below in Table 7.

Table 7: Restriction enzyme reaction mix. The reaction mix used for restriction enzymes (from NEB).

DNA	1 µg
10X NEB Buffer	5 µl
NEB restriction enzyme	1 µl
Nuclease free water	Up to 50 µl

This mix was incubated at 37 ℃ for 1 hour. 5 µl of purple loading dye (NEB- Gel loading dye, purple (6X)) was added to the sample and this along with a 1 kbp DNA ladder (Quick-Load 1 kb DNA Ladder, NEB) was run on a 1% agarose-TAE gel mixed with 4 µl of GelRed Nucleic Acid Stain 10000X in 1X TAE buffer at 100 V for 1 hour with. Results were visualised by UV using the Syngene U:Genius 3 machine.

2.3.5 Preparing the pJ251 Tinsel Purple Reporter System

The Anderson Strong promoter was purchased as 2 complementary strands (6 from Table), mixed in equimolar concentrations, heated to 94 ℃ (using a Star Lab Smart Instruments Mini Dry Bath) for 5 minutes and left overnight (~16 hours) on the benchtop to cool.

Then to insert this promoter into the pJ251 Tinsel Purple plasmid (3 from Table 2), both the promoter and plasmid separately were digested with BamHI-HF (New England Biolabs). The plasmid has one BamHI cut site, and the Anderson strong has 2 on either end of its DNA. The NEB protocol in Table 7 was used for this digest.

This was incubated at 37 ℃ for 1 hour for the Anderson strong promoter digest and 30 minutes for the pJ251 Tinsel purple digest. To the pJ251 Tinsel purple digest mixture, 1 µl of alkaline phosphatase (NEB Alkaline Phosphatase, Calf Intestinal (CIP)) was added to prevent self-ligation of the plasmid by dephosphorylating the ends, and the mixture was set to incubate at 37 ℃ for another 30 minutes.

Addition of the digested promoter into the digested plasmid was done at various ratios of promoter (insert) to plasmid (vector). The NEBioCalculator was used to calculate the mass and thus volumes to add using the inputs of the length of both insert (41 base pairs (bp)) and vector (3543 bp). Ratios of vector: insert was done at 0.1:1, 1:1, 1:10, 1:50.

With these the ligation reaction mixture using NEB buffers and ligase was made according to Table 8.

Table 8: Ligation reaction mix. The reaction mix used for ligating a plasmid and insertion vector (from NEB).

10X T4 NEB Buffer	2 µl
Insert DNA	various
Vector DNA	various
T4 Ligase (NEB)	1 µl
Nuclease free water	Up to 20 µl

This was incubated at room temperature (~ 20 ℃) for 30 minutes before being transformed into *E. coli* TOP10 cells (Chapter 2.4.1).

After 3 days of incubation on a LB plate with kanamycin at 37 ℃, light pink/purple colonies began to form especially on those with 1:10 and 1:50 ratios of vector: insert. These were selected for and grown in 10 mL of liquid LB with 50 µg/mL kanamycin to be used for glycerol stocks and for miniprep to extract DNA to send for Sanger sequencing (Chapter 2.4.3).

A pink/purple coloured colony containing the pJ251 Tinsel Purple plasmid with the Anderson Strong promoter inserted into it was streaked onto various plates of different media (3 rich and 2 minimal) for comparison. Lambda Broth, Superbroth and TYGPN (Chapter 2.2.4) were the 3 rich media made and the 2 minimal media were M9 and M63 (Chapter 2.2.4). These had 50 µg/mL kanamycin added and one plate of each (35-40 mL) was made and streaked with the pJ251 Tinsel Purple+Anderson strong promoter colony, along with 1 plate of LB+Kan as a control. These were incubated at 37 ℃ for 3 days also.

Different temperatures for colour changes were also tested and so plates of LB with 50 μg/mL kanamycin were streaked with the pJ251 Tinsel Purple+Anderson strong promoter colony and 1 plate each was incubated for 3 days at temperatures of: 30 ℃, 15 ℃, room temperature (~ 20 ℃) and a control at 37 ℃.

The same media (Lambda Broth, Superbroth, TYGPN, M9, M63 and LB as a control – Chapter 2.2.4) were used to test if colour change would be apparent in liquid. These were made with addition of 50 μg/mL kanamycin and a stock of the same colony was used to inoculate the medias. These were placed into a shaking incubator (180 rpm) at 37 ℃ for 3 days. Cells from these were pelleted by centrifugation of the culture at 14,000 RCF for 2 minutes.

2.3.6 Creating a mScarlet-I Reporter system

The 80 bp intergenic region of DNA was PCR purified as described above in Chapter 2.3.1. This was to be the sequence inserted into the mScarlet-I plasmid (5 in Table 2). XbaI, NdeI and ScaI are restriction enzymes that cut at a singular site in this plasmid in the desired region. The high fidelity (HF) versions of these enzymes (from NEB) were used for restriction digest of the plasmid following the reaction mixture in Table 7. Digest mixtures were incubated at 37 ℃ for 1 hour and then used for insertion of the 80 bp intergenic fragment and ligation.

Again, the NEBioCalculator was used to determine the volumes required of the insert (80 bp intergenic DNA) and the vector (3180 bp mScarlet-I plasmid). A 5:1 insert: vector ratio was selected, and volumes calculated to create a reaction mixture as detailed in Table 8. This was incubated at room temperature (~ 20 ℃) for 30 minutes before being transformed into *E. coli* TOP10 cells (Chapter 2.4.1). Transformations were spread on a plate of LB with 100 μg/mL ampicillin and left to grow overnight in a 37 ℃ incubator.

Bright pink, light pink and white colonies were selected for and used to inoculate 10 mL of liquid LB with 100 µg/mL ampicillin. These were grown overnight, shaking (180 rpm) at 37 ℃ and then used to make stock colonies (Chapter 2.4.2) as well as for miniprepping and sending for Sanger sequencing (Chapter 2.4.3).

Colonies with the 80 bp region inserted into the XbaI site was also digested with NdeI (Table 7) and run on a 1% TAE Agarose gel, visualised by UV using the Syngene U:Genius 3 machine.

2.3.7 Creating T18- mutant MmyJ plasmids

Mutants of MmyJ fused by the N-terminal to T18 fragment were created using the T18-MmyJ plasmid (9 from Table 2) as a template and mutant primers (21-24 from Table 1). The following PCR reactions were created using NEB Q5 reaction mixes and polymerase:

Table 9: Mutagenesis PCR reaction mix. The reaction mix used for mutagenesis of T18-MmyJ using Q5 polymerase and buffers from NEB.

5X Q5 Reaction Buffer	5 µl
5X Q5 High GC Enhancer	5 µl
10 mM dNTPs	5 µl
10 mM Primer	2.5 µl
Template DNA (~ 150 ng/µl)	0.5 µl
Q5 High-Fidelity DNA Polymerase	0.2 µl
Nuclease-Free Water	7 µl

As seen in Table 9, primers were not added as pairs and instead each reaction contained one primer (forward or reverse from 21-24 in Table 1). These reactions were subject to the following PCR reaction:

95 ℃	1 min	
95 ℃	1 min	⎫
60 ℃	1.5 min	⎬ x10
68 ℃	5 min	⎭
4 ℃	Hold	

Following this 12.5 µl of the forward primer reaction mixes were added to 12.5 µl of their complementary reverse primer mix along with an additional 0.5 µl of Q5 Polymerase (NEB). This new PCR reaction was subjected to the same PCR as done

previously. To then digest the template plasmid, 1.5 µl of DpnI (from NEB) was added straight to PCR reaction mixes and incubated at 37 ℃ for 2 hours. This was transformed into *E. coli* DH5α cells (Chapter 2.4.1) and miniprepped for sequencing to confirm successful mutagenesis (Chapter 2.4.3). Successfully mutated plasmids were then subject to bacterial adenylate cyclase two-hybrid (BACTH) assays (Chapter 2.4.6) by Dr Jonathan Cook (Crow lab, University of Warwick).

2.4 Transformations

2.4.1 Propagation Host Transformation

All plasmids (except those for mutagenesis, 8-12) as listed above in Table 2 were transformed into chemically competent *E. coli* TOP10 Cells (Invitrogen One Shot TOP10 Chemically Competent *E. coli*). *E. coli* TOP10 acts as a plasmid propagation strain to maintain the quality of the DNA and keep the recombination risk low.

This was done by addition of 3 µl of plasmid to 25 µl of chemically competent cells which was stirred gently and left on ice for 30 minutes. Cells were heat shocked for 30 seconds at 42 ℃ and put immediately on ice for 5 minutes. Under sterile conditions this was added to 800 µl of LB media (Chapter 2.2.4) and placed in a shaking (180 rpm) incubator at 37 ℃ for 1 hour. 100 µl of this culture was spread into a lawn onto an LB plate with the appropriate selective agar according to the plasmid's resistance. The rest of the culture was centrifuged at 14,000 RCF for 1 minute, the supernatant decanted, and the cells resuspended in the remaining supernatant. This concentrated mix was also spread onto a similar LB plate (with antibiotic). Both plates were incubated at 37 ℃ overnight before colonies were selected.

The mutant T18-MmyJ plasmids (13-16 from Table 1) were transformed with the same protocol as for TOP10 cells into *E. coli* DH5α cells.

For glycerol stocks and plasmid minipreps to confirm correct transformation into the strains and regeneration of fresh plasmid, starter culture is required. This was done by first inoculating 10mL of fresh LB containing the appropriate antibiotic (50 μg/mL Kanamycin for Tinsel Purple with Anderson strong (4 in Table 2) and 100 μg/mL Ampicillin for all others (Table 2)). Inoculation was done with a sterile loop worth of TOP10 cells (colony or stock) into 50 mL falcon tubes. These were left shaking in a New Brunswick Scientific incubator overnight (~ 16 hours) at 180 rpm at 37 ℃.

2.4.2 Glycerol stocks of cells

Glycerol stocks were made under sterile conditions by addition of equal parts starter cultures and sterilised 50% glycerol in water. These were frozen and kept in the -80 ℃ freezer.

2.4.3 Miniprepping of DNA for Sanger sequencing

For extraction of the plasmid from the TOP10 cells (and DH5α for mutagenesis of MmyJ) the 10 mL of the starter culture was miniprepped using a commercial kit (ThermoScientific GeneJET Plasmid Miniprep Kit) following the kit's directions.

5 μl of DNA was mixed with 5 μl of primer (both forward and reverse of primers 1 and 5 from Table 1) in a 1.5 mL Eppendorf and sealed tight. This was sent by the School of Life Sciences, University of Warwick to Eurofins for Sanger sequencing.

2.4.4 Expression Host Transformation (BL21 and C43 *E. coli* strains)

The plasmids for protein expression of MmyJ, C49S MmyJ and MmyB (Plasmid 1,2,7 and 8 from Table 2) were transformed into chemically competent BL21 cells (Thermo Scientific™ BL21(DE3) Competent Cells) using the same protocol as transformation into TOP10 Cells (described above).

The codon optimised MmyB (3 in Table 2) was also transformed into C43 chemically competent cells (donated by Dr Crow, Warwick University). The same protocol as for TOP10 and BL21 transformations were used.

Starter culture in LB and selective antibiotic, as above, was also made of some of these colonies and sent for Sanger sequencing using T7 promoter and terminator primers (primer 1 from Table 1). Plasmids containing the correct protein of interest sequence were made into glycerol stocks and used for further expression studies.

2.4.5 Confirmation of BL21 pET151-MmyJ

The 4 colonies of BL21 transformations with pET151-MmyJ was confirmed after Sanger sequencing by PCR too. The primers '*mmyJ* gene' (primer 3 from Table 1) were used for a Q5 Polymerase PCR reaction (with a negative control of water rather than the MmyJ plasmid) according to the reaction mix described in Table 5.

This was subjected to the following PCR programme on an Eppendorf Mastercycler Nexus Gradient machine:

98 °C	2 mins	
98 °C	30 secs	⎫
62.4 °C	45 secs	⎬ x30
72 °C	1 min	⎭
72 °C	5 mins	

Again, 5 µl of purple loading dye (NEB- Gel loading dye, purple (6X)) was added to the samples and this along with a DNA ladder (low molecular weight DNA Ladder- NEB) was run on a 1% agarose-TAE gel mixed with 4 µl of GelRed Nucleic Acid Stain 10000X in 1X TAE buffer at 100 V for 1 hour with. Results were visualised by UV using the Syngene U:Genius 3 machine.

2.4.6 Bacterial adenylate cyclase two-hybrid (BACTH) assays

These assays were completed by transforming T18 and T25 plasmids into *E. coli* BTH101 chemically competent cells (curtesy of Crow lab, University of Warwick). T18 and T25 plasmids with MmyJ either the N or C terminal (Table 2- 9-12) were created by Dr Jonathan Cook (Crow lab, University of Warwick) as the test samples. The negative controls were T18 and T25 plasmids with no protein fused on the ends and positive controls were that of a leucine zipper on each plasmid (also curtesy of Crow lab, University of Warwick). The following procedure was used for the transformation step.

First into 30 µl of BTH101 cells, 1 µl of the T18 plasmid (with and without the MmyJ/leucine zipper) and 1 µl of the T25 plasmid (with and without the MmyJ/leucine zipper) were added and mixed. This was left on ice for 45 minutes after which it was heat shocked at 42 ℃ for 30 seconds and immediately placed to recover on ice for 5 minutes. To this mix 470 µl of SOC/LB media was added and incubated for 1.5 hours at 30 ℃ at 160 rpm. After this time, 1.5 mL LB with 50 µg/mL ampicillin and 25 µg/mL kanamycin were added, and cells incubated overnight shaking (160 rpm) at 30 ℃.

Afterwards, samples were spotted onto a 'KAXI' plate. These are plates consisting of LB agar containing 50 µg/mL ampicillin, 25 µg/mL kanamycin, 40 µg/mL X-gal and 0.5 mM IPTG, made by Dr Jonathan Cook (Crow lab, University of Warwick). These were then incubated at room temperature for 3 days and imaged using an Epson scanner.

This same procedure was repeated by Dr Jonathan Cook for T18-mutant MmyJ.

2.5 Protein production and purification

2.5.1 Small cultures to test for protein expression

For MmyJ and MmyB production in BL21, small cultures were first tested for expression of protein.

Small cultures of 50 mL or 100 mL LB (Chapter 2.2.4) inoculated with 500 µl (for 500 mL) and 1 mL (for 100 mL) liquid starter culture of cells and 50 µl or 100 µl of ampicillin at 100 mg/mL were grown in a 250 mL flask at 37 ℃, shaking at 160 rpm for 3 hours or until OD_{600} reached between 0.5-0.8. Final concentration of 1 mM filter sterile IPTG was added only to the induced samples after which the temperature turned down to 20 ℃ to grow overnight.

Cultures were then centrifuged at 3219.84 RCF for 15 minutes at 4 ℃ until cells were pelleted. 15 µl of supernatant was taken, and the rest of the pellet were submerged in 'Wash Buffer' (Chapter 2.2.3). A small pinch of lysozyme from hen egg white was added and the pellet was lysed before being sonicated. Sonication was conducted using a MSE Soniprep 150 and the MSE MSS38121-114A Soniprep Exponential Titanium Probe attachment at ~15 kHz for 30 seconds with a 1-minute rest on ice for 8 rounds. This was once again centrifuged at 3219.84 RCF for 15 minutes at 4 ℃. The supernatant of this was considered the 'soluble protein fraction' and the pellet the 'insoluble protein fraction'. 15 µl of these were also taken. All 15 µl sample fractions had added 15 µl of 2X Laemmli Sample Buffer (Biorad) and final concentration of 50 mM DTT before being heated at 95 ℃ for 10 minutes. Samples and a ladder (Color Prestained Protein Standard, Broad Range (10-250 kDa) from NEB) were loaded onto a precast SDS-PAGE gel (4-20% TruPAGE gel by Merck) and run using 1X TruPAGE MOPS buffer for 40 minutes at 180 V.

The gel was stained using Instant*Blue*™ by incubating the gel in it for 15 minutes before washing off and destaining with tap water.

2.5.2 Large cultures for high protein production

MmyJ in BL21, MmyB in both C43 and BL21 were scaled up for mass production of protein.

Similar to small cultures, 1 L of LB (Chapter 2.2.4) was inoculated with 10 mL of liquid starter culture of cells and 1 mL of ampicillin at 100 mg/mL in a 2 L flask. These were grown at 37 ℃, shaking at 160 rpm for ~ 4 hours or until OD_{600} reached between 0.5-0.8. Final concentration of 1 mM filter sterile IPTG was added, and temperature dropped to 20 ℃ for overnight growth.

Cultures were then centrifuged at 1529.424 RCF for 15 minutes at 4 ℃ until the cells were pelleted. The supernatant was discarded, and cell pellets not used at the time, frozen and stored at -80 ℃. For purification of the protein, a 1 L cell pellet had a small pinch of hen egg white lysozyme added to it and 'Wash Buffer' (Chapter 2.2.3), until the volume reached ~ 40 mL. This was then lysed before being sonicated as previously using a MSE Soniprep 150 and the MSE MSS38121-114A Soniprep Exponential Titanium Probe attachment at ~15 kHz for 30 seconds with a 1-minute rest on ice in between for 10 rounds. The sonicated culture was then centrifuged at 17217.2 RCF for 20 minutes at 4 ℃. The supernatant from this was taken and immobilised metal ion affinity chromatography performed.

2.5.3 Immobilised Metal Ion Affinity Chromatography (IMAC)

The first stage of purification was done by IMAC using charged nickel. The slurry containing the Ni Sepharose beads (Ni Sepharose® High Performance by Cytiva) and soluble protein (~ 25 µl slurry per 1 mL sample) was incubated overnight in a cold room at ~8 ℃, rotating slowly. This was then purified using gravity filtration in a Pierce™ Centrifuge 10 mL columns (Thermo Scientific) at room temperature. The slurry and protein mix was added to the column (this was the 'load' fraction) and allowed to drip through (the 'flow through' fraction). Then beads were washed with

10 column volumes (CV) of 'Wash Buffer' (the flow through of this being the 'wash' fraction).

In the case of MmyB grown in C43 cells, an ATP wash step was also added. Here after the 'wash' fraction, 1 CV (10 mL) of the 'ATP Buffer' (Chapter 2.2.3- Table 4) was added with the elution fraction referred to as 'ATP Wash 1'. This was repeated with another 1 CV of the same 'ATP Buffer' and this elution was 'ATP Wash 2'.

For MmyJ and C49S MmyJ various concentrations of imidazole (30-100 mM) was trialled (Chapter 2.2.3). 1 CV of each was washed over the beads and the elution of each collected.

A final elution buffer containing 250 mM imidazole was used as an 'Elution buffer' (Chapter 2.2.3). 3 mL of this was added and incubated for 2 minutes, before being eluted off and kept on ice (this was the 'Elution 1' or 'Elution' fraction). To ensure all protein is eluted off, 5 mL (1/2 CV) of Elution Buffer was then added and the fraction ('Elution 2') collected.

15 µl of samples was taken from the appropriate fractions and had 15 µl of 2X Laemmli Sample Buffer (Biorad) added to it with also a final concentration of 50 mM DTT before being heated at 95 ℃ for 10 minutes. Samples and a ladder (Color Prestained Protein Standard, Broad Range (10-250 kDa) from NEB) was run on a TruPAGE 4-20% gel with 1X TruPAGE MOPS buffer at 180 V for 40 minutes. Gels were stained with Instant*Blue* for 15 minutes and then destained with tap water.

2.5.5 Size Exclusion Chromatography (SEC) for purification

Elution fractions from IMAC were pooled together and concentrated using an Amicon® Ultra-15 Centrifugal Filter (Merck) with a 10 kDa cut off at 4 ℃, 3219.84 RCF until concentrated to ~ 3 mL.

Size exclusion chromatography was conducted on a HiLoad 16/600 Superdex 75pg (Cytiva Life Sciences) column attached to a AKTA Purifier UPC 900 machine. The column was equilibrated with 'Final Buffer' (Chapter 2.2.3) as this was also the buffer used to run samples. First however, 1 mL of a standards mix (Low molecular weight Gel Filtration Calibration Kits-Cytiva Life Sciences) was injected, consisting of 3 mg/mL Aprotinin, 3 mg/mL Ribonuclease A, 4 mg/mL Ovalbumin and 1 mg/mL of Blue Dextran. All standards and samples were run at a flow rate of approximately 0.5 mL/min to prevent pressure exceeding manufacture's recommendation with readings at 280 nm measured.

10 µl samples of fractions containing peaks in wavelength readings were mixed with 10 µl of 2X Laemmli Sample Buffer (Biorad) as well as DTT to a final concentration of 50 mM. Samples were boiled at 95 ℃ for 10 minutes and as before run with a ladder (Color Prestained Protein Standard, Broad Range (10-250 kDa) from NEB) on a TruPAGE 4-20% gel with 1X TruPAGE MOPS buffer at 180 V for 40 minutes. Gels were stained with Instant*Blue* for 15 minutes and then destained with tap water.

Fractions containing the pure protein (MmyJ or MmyB) were taken and concentrated by a fresh Amicon® Ultra-15 Centrifugal Filter (Merck) with a 10 kDa cut off at 4 ℃, 3219.84 RCF to concentrations higher than 5 mg/mL where possible for storage and other studies.

2.5.6 Buffer exchange of protein

Where buffer exchange could not be performed with SEC (particularly for MmyB) this was done using a PD-10 Desalting Column packed with Sephadex G-25 resin (Cytiva Life Sciences). The gravity protocol was used where the column was equilibrated with 25 mL of 'Final Buffer' (Chapter 2.2.3). 2.5 mL of the sample with protein was then added and the 3.5 mL of 'Final Buffer' used to elute the buffer exchanged protein.

Protein was then concentrated as necessary with a fresh Amicon® Ultra-15 Centrifugal Filter (Merck) with a 10 kDa cut off at 4 ℃, 3219.84 RCF and used for storage or other studies.

2.5.7 Storage of pure protein

Purified protein concentrations were measured using a ThermoScientific Nanodrop 2000 Spectophotometer with the molecular coefficient (4470.00 M^{-1} cm^{-1} for recombinant MmyJ and 43430.00 M^{-1} cm^{-1} for recombinant MmyB) and molecular weight (15886.83 Daltons for MmyJ and 38134.99 Daltons for MmyB) inputted so that absorbance readings at 280 nm were used to give a concentration of the sample.

All samples were ensured to be more than 1 mg/mL for storage. For MmyJ the 'Final buffer' already contained the cryoprotectant glycerol but for MmyB in its 'Final buffer' (Chapter 2.2.3) there is no glycerol therefore, for MmyB 5 % glycerol was added before aliquots of 100 µl was made for both proteins. These were snap frozen by submerging in liquid nitrogen before being placed in a -80 ℃ freezer. Aliquots were taken as needed for single use preventing denaturation through multiple freeze-thawing cycles.

2.5.8 Targeted proteomics to confirm protein samples

For tryptic digest and targeted proteomics, samples were first run on a TruPAGE 4-20% gel with 1X TruPAGE MOPS buffer at 180 V for 40 minutes. Gels were stained with Instant*Blue* for 15 minutes and then destained with tap water overnight before bands of interest were excised.

A tryptic digest protocol from the Proteomics Research Technology Platform (Warwick University) was used for in-gel digest of proteins. Solutions for this was added just enough to cover the gel pieces. The procedure started with destaining of

the gel pieces with 50 % ethanol and 50 mM ammonium bicarbonate (ABC) for 20 minutes at 55 ℃, shaking at 650 rpm. The supernatant was decanted, and the procedure repeated 2 more times until gel pieces were clear. Next it was dehydrated with 100% ethanol with incubation at room temperature for 5 minutes shaking at 650 rpm. The supernatant was decanted again, and reduction and alkylation performed. For this, 10 mM Tris-(2-carboxyethyl) phosphine hydrochloride (TCEP) and 40 mM 2-chloroacetamine (CAA) was added and incubated at 70 ℃ for 5 minutes, shaking at 350 rpm. After decanting the supernatant more washes of ethanol and ABC as performed at the start for 3 rounds was done. Dehydration with 100% ethanol was then also performed as before. The gel was hydrated again with a solution of trypsin at 2.5 ng/µl in 50 mM ABC. This was left overnight incubating at 37 ℃.

The supernatant of this digest was collected and for further extraction gel pieces had 25% acetonitrile and 5% formic acid added to it which was sonicated in a water bath at room temperature for 10 minutes. This peptide extraction was repeated 2 more times with the supernatant collected each time. All peptide solutions were pooled and concentrated in a Speed-Vac at 45 ℃ until concentrated to 20 µl. 30 µl of 2% acetonitrile and 0.1% formic acid was then added, and samples were submitted to mass spectrometry.

Mass spectrometry was run by the Proteomics Research Technology Platform (Warwick University) with processed resulted returned for analysis.

2.5.9 X-ray Crystallography of protein

The trays for X-ray crystallography used were 96-well, 2 drop plates (MRC developed Swissci 96-Well 2-Drop Plates) along with MIDASPlus, JCSGPlus, Structure Screen 1+2 and Morpheus commercial screens (all were HT-96 from Molecular Dimensions). A sitting drop method was used with the first drop having a 2:1 ratio of protein to screen and the second drop was set to have a 1:2 ratio of

protein to screen. Trays were set up using a Mosquito LCP crystallization robot, sealed, and left at room temperature. Crystals of C49S MmyJ at 7.5 mg/mL was harvested from a tray with MIDASPlus in E7 after approximately 1 year of incubation. MmyB crystals at 6 mg/mL were harvested from a JCSGPlus screen in A2 after approximately 7 months of incubation. Both these crystals selected were harvested (by Dr Crow, University of Warwick) in litholoops and plunge-frozen in liquid nitrogen to then be sent to Diamond synchrotron (UK) for diffraction and data collection.

2.6 *In vitro* analytical work

2.6.1 Size exclusion chromatography (SEC) for stoichiometry analysis

To test for MmyJ-DNA stoichiometry the preparative grade SEC (HiLoad 16/600 Superdex 75pg (Cytiva Life Sciences)) column was used first before switching over to the analytical column (Superdex 75 Increase 5/150 GL (Cytiva Life Sciences)). The buffer used for equilibrations and for the runs was the 'Final Buffer' (Chapter 2.2.3) of MmyJ. Both columns had been connected to a AKTA Purifier UPC 900 machine with a flow rate of ~ 0.5 mg/mL for the preparative grade SEC and ~ 0.8 mg/mL for the analytical column with absorbance measurements taken at 280 nm. The DNA used was the PCR purified 80 bp intergenic fragment (as described in Chapter 2.3.1).

For preparative SEC the lowest volume able to be injected was 600 µl. Within this volume samples of MmyJ at 300 pmol was injected and MmyJ with DNA at 300 pmol MmyJ and 3 pmol DNA. Both samples were incubated for 30 minutes, rocking at room temperature before injection. Due to the large volume of the preparative column, samples became too dilute to give accurate peaks.

For the analytical SEC volumes of 300 µl were able to be injected. Samples containing 65 pmol MmyJ, 10 pmol MmyJ with 0.1 pmol DNA, and 1 pmol MmyJ

with 0.1 pmol DNA were made using MmyJ's 'Final Buffer' (Chapter 2.2.3) to bring up to volume. These were incubated, gently rocking at room temperature (~ 20 ℃) for 30 minutes before injection into the SEC column.

2.6.2 Protein-DNA Assays (Electromobility Shift Assays)

Electromobility shift assays (EMSAs) were done to test Protein-DNA interactions. First homemade 10% NATIVE-PAGE gels were made using the 1 mm Biorad system.

The recipe below was used to first make the resolving gel which was allowed to set for 10 minutes before addition of the stacking gel and comb which was then allowed to set for a further 10 minutes or until polymerised. Gels were pre-run for 2 hours in a cold room (~6 ℃) with fresh 1X TBE at 100 V before samples were loaded.

10 % NATIVE-PAGE gel: Resolving	
Component	Volume (ml)
30 % Acrylamide/Bis	3.4
1.5 M Tris-HCl pH 8.8	1.6
10 % APS (w/v)	0.1
TEMED	0.01
dH2O	4.89
TOTAL:	10

5 % NATIVE-PAGE gel: Stacking	
Component	Volume (ml)
30 % Acrylamide/Bis	1.7
1 M Tris-HCl pH 6.8	1.25
10 % APS (w/v)	0.1
TEMED	0.004
dH2O	6.9
TOTAL:	10

10% NATIVE PAGE recipe:

Relevant samples were created by addition of DNA and/or protein and/or ligand and nucleotide free water (ThermoFisher). The crude extract likely containing the ligand was reconstituted from dried W89 strains grown on supplemented minimal media (SMM) (Chapter 2.7).

The following methods were used to calculate the concentrations of each component:

DNA: DNA in µg/µl inputted into Promega Biomath dsDNA: µg to pmol calculator for DNA in pmol/µl.

Protein: (mg/mL x 1000) / weight in kDa = pmol/µl

Ligand: Crude extract ligand solutions were reconstituted in minimal water required for reconstitution and used according to dilution factor.

Samples were made to a final volume of 10 µl from the different components required and left to incubate, gently rocking at room temperature for 30 minutes. 3 µl of 2.5X NATIVE gel loading buffer (Cold Spring Harbor Protocols recipe) was added, and samples loaded onto the 10% NATIVE PAGE gels. This was run for 2 hours at 80 V in a cold room using pre-run cold gels and buffer as stated previously.

4 µl of GelRed Nucleic Acid Stain 10000X was added to 100 mL of dH_2O to stain for DNA and the gel(s) submerged for 30 minutes, gently rocking at room temperature. Gels were visualised by UV using a Syngene U:Genius 3 machine.

2.7 Ligand Production

2.7.1 Growth of *Streptomyces* strains

Strains W89, M145, W108- The W89, 108 and M145 *S. coelicolor* strains (10-14 Table 2) were grown in supplemented minimal media (SMM) in liquid form (Chapter 2.2.4) first. This was done by inoculation of 200 mL of the SMM media with 50 µl of spore stock each in a sealed 2 L baffled flask. Flasks were left shaking at 160 rpm at 30°C for 6 days before organic extraction was attempted.

These strains were also grown on solid SMM (Chapter 2.2.4) whereby for each strain, 500 mL of solid SMM was made with ~35-40 mL poured into standard petri dishes (100x15 mm) and allowed to set before 10 µl of spore stock for each plate

was spread over plates in a lawn. These were sealed with Parafilm and set in an incubator at 30℃ also for 5 days after which they were subjected to organic extraction of compounds.

Finally, these strains were also grown on solid R2YE media (Chapter 2.2.4). 500 mL of solid media for each strain was poured into plates (as described above) and 10 µl of spore stock for each plate spread into a lawn. These were sealed with Parafilm also and placed in a standing incubator at 30℃ for 2 full days before organic extraction.

Strains W81, W81+pJH2- These strains were also grown on solid R2YE media (Chapter 2.2.4). For each strain 500 mL of media was poured into petri dishes and 10 µl of spore stock spread over as a lawn. After being wrapped in Parafilm these were placed in a 30℃ for 2 full days with compounds extracted by solvent afterwards.

2.7.2 Organic extraction of compounds from media

For extraction from solid medias, the agar was first cut into roughly 1 cm x 1 cm cubes and placed in a baffled flask. Sterile water at pH 2.0 was added to completely submerge the solid media (50 mL + *w/v* of media). This was then placed in a 30℃ incubator shaking at 160 rpm overnight (~ 16 hours). The entire sample was centrifuged at 3219.84 RCF for 15 minutes before the supernatant was filtered by gravity filtration using Grade 1 Whatman Filter Paper. The flow through of this was collected for liquid-liquid organic extraction using ethyl acetate.

For strains grown in liquid media, the media after growth was acidified to pH 2.0 with HCl after which liquid-liquid organic extraction was conducted with no prolonged incubation time.

Liquid-liquid organic extraction was performed by addition of the sample into a 2 L separating funnel where ethyl acetate was added as the same volume of sample. The mixture was shaken and stood upright for 5 minutes to allow for separation of layers. With this solvent the organic layer is the top layer and so this was the layer taken.

The crude organic extract samples were placed into a Genevac Personal Evaporator EZ-2 Elite until completely dried. These were then used for analytical and detection techniques.

2.7.3 Analytical techniques for purification and detection

2.7.3.1 Purification by high-performance liquid-chromatography

High-performance liquid-chromatography (HPLC) was conducted on crude organic extracts of M145, W89 and W108 grown on SMM (as described above). These extracts were dissolved in 50:50 methanol: water and 500 µl injected into an analytical reverse-phase C18 column (Agilent TC-C18(2)) connected to the HPLC machine (Aligent Technologies 1260 Infinity). Separation was done using the solvents HPLC grade methanol (with 0.1% formic acid) and HPLC grade water (with 0.1% formic acid), going from 0-100% of the water to methanol in a runtime of 30 minutes. This same method was used for multiple runs of preparative HPLC with 1 mL injections of only the W89 sample (the same one used for analytical HPLC) with a preparative C18 column (Agilent TC-C18(2)). Wavelengths at 210 nm, 225 nm, 245 nm and 254 nm were taken for all samples run.

2.7.3.2 Detection of compounds by thin layer chromatography

Thin Layer Chromatography (TLC) was done on crude organic extracts of M145 and W89 strains (grown on SMM) prepared (Chapter 2.7.2). These were redissolved in 1 mL of ethyl acetate and spotted on an aluminium TLC, silica gel plate (Merck) cut to

size. This was placed into a TLC chamber with a solvent mixture containing 9:1 of toluene: acetic acid. Once spots had migrated the plate was air-dried and visualised using bromocresol green stain.

The stain was prepared by dissolving 0.02 g bromocresol green in 50 mL of ethanol. 0.1 M NaOH was added dropwise until the solution was blue. The TLC plate was dipped, and excess strain blotted away. Any compounds with carboxylic acid should then be able to be visualised as yellow spots on the plate.

2.7.3.3 Compound detection by liquid-chromatography mass spectrometry

The dry crude organic extract samples (Chapter 2.7.2) from all the strains were tested for contents by LC-MS, as was the dried HPLC sample (Chapter 2.7.3.1). They were all reconstituted in 50:50 methanol and water in the minimal amount of liquid needed. These were then centrifuged in spin column filters (Corning® Costar® Spin-X® centrifuge tube filters-Merck) and 500 µl of filtrate added into sample vials (Aligent part number 5190-2280) for LC-MS.

The LC-MS machine used was the low-resolution Bruker MS amaZon SL ESI machine (curtesy of Chen Lab, University of Warwick) with a reverse-phase C18 column (Agilent TC-C18(2)). The solvents used were HPLC grade water with 0.1 % formic acid and HPLC grade methanol with 0.1 % formic acid. The programme used was 5 to 95 % gradient of water to methanol over 30 minutes for each sample.

Results were collected using Bruker Daltonics HyStar 3.2 SR4 program (version 49.9).

Chapter 3: Studying the transcriptional properties of MmyJ

3.1 Molecular basis for the control of methylenomycin self-resistance in *S. coelicolor* A3(2)

MmyJ is not a totally unstudied protein as an attempt to analyse it through electromobility shift assays for functionality, protein stability work and bioinformatic analysis have been made once already. [155] Previous bioinformatic analysis revealed MmyJ to be closely related to a metal-sensing group of transcriptional regulators known as the ArsR family.[155] This was done through BLAST searching of the amino acid sequence of the unknown protein to known proteins in the NCBI database and scoring by various statistical tests to determine the closest related known proteins to unknown.

Primary structure can help to determine secondary structures such as alpha helices and beta sheets, the formation of which requires specific hydrogen bonding to occur between the carbonyl oxygen of one amino acid and the amino hydrogen of another amino acid. Alpha helices require this type of hydrogen bonding to occur every 4 amino acids which forms a curled ribbon formation with each helix turn containing 3.6 amino acids. Therefore, amino acids that cause disruptions in alpha helix formation such as proline would not be used for such secondary structures. Beta sheets likewise have hydrogen bonds occurring between lines of amino acids chains in either parallel or antiparallel formations. Beta sheets tend to contain amino acids with ring structures such as phenylalanine, tryptophan, and tyrosine due to the reduced spatial restraints a beta structure offers. The secondary structure of a protein can help give insight into its tertiary structure. For example, alpha helix structures with hydrophilic residues facing outwards from the helix may be facing an aqueous environment and be on the surface of the protein whereas alpha helices with more evenly distributed hydrophobic residues may be found buried inside the protein core or in a hydrophobic environment.[181] Therefore studying the amino acid composition of a protein and analysing closely related

proteins may give valuable insight into the protein of interest. Despite this, predicting the structure of a protein based on its amino acid sequence from bioinformatic techniques can be difficult.

Likewise, just because MmyJ is closely related to ArsR family proteins does not necessarily mean that MmyJ functions according to other ArsR family members. In that same vein BLAST searching does not show how evolutionarily similar or dissimilar MmyJ may be to ArsR family proteins. Having phylogenetic information of MmyJ could show if there might be divergences in the ArsR family and thus further imply MmyJ as part of this family of transcriptional regulators.

3.1.1. Overview of the ArsR family which are metal sensing transcriptional regulators

As discussed previously, existing work implies MmyJ is part of the ArsR family proteins. To look further into this an understanding of the ArsR family proteins is necessary.

In brief, the ArsR family of transcriptional regulators are 1 of the 7 major metal sensing families in bacteria.[182] This family contains proteins that are able to sense a large range of metal ions, (Hg^{2+}, Zn^{2+}, Cd^{2+}, Cu^{+}, Pb^{2+}, Ag^{+}, Co^{2+}) making them a diverse class of metalloregulators.[183,184] In particular actinobacteria (particularly *Mycobacerium tuberculosis* and *S. coelicolor* A3(2)) have large numbers of ArsR family proteins encoded at the DNA level, many of which have yet to be characterised fully. To have this variety of metal sensing there is also diversity in ArsR family proteins in their ligand binding structures.[122]

The family's name comes from the first discovered ArsR family protein- ArsR from *E. coli*. The *E. coli arsR* gene was found as the first gene in the operon *arsRBC* with *arsBC* encoding efflux pumps for arsenic resistance. *E. coli* ArsR is a small protein of 13 kDa, shown to function as a 26 kDa homodimer which regulates arsenic levels by

binding upstream of the *arsRBC* operon. Binding of ArsR to the *arsRBC* operon represses expression of the operon until arsenite (As^{3+}) is present. Arsenite binds to ArsR and causes the protein to be released from the DNA allowing expression of ArsBC efflux pumps to reduce arsenic levels in the cell.[185–187] A metal binding box motif formed by 6 residues (ELCVCD) is found on the α3 helix of ArsR with the 2 cysteines in the motif binding to arsenite metal ions. This α3 helix metal binding site is referred to as a type 1 metal binding site (Fig. 11).[188,189]

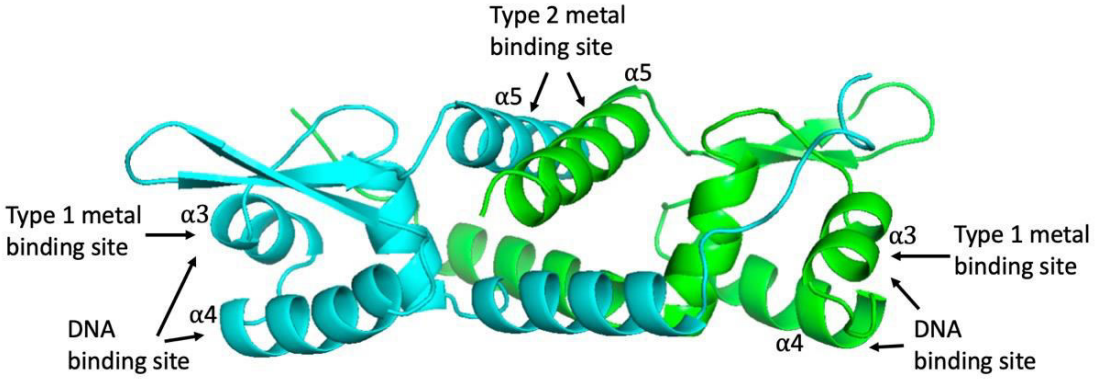


Figure 11: The ribbon structure of apo-SmtB, a DNA binding protein with type 1 and 2 metal binding sites.[190] SmtB is part of the ArsR family of transcriptional repressor proteins, which functions as a homodimer (monomers in cyan or green) to bind to DNA. It is made of 5α helices and 2β sheets in the formation of αααααββα and has 2 metal binding sites present (type 1 and type 2) though only Zn^{2+} binding at the type 2 site releases SmtB from the DNA. Type 1 sites in ArsR family proteins contain the motif ELCVCD and is found at the α3 helix.[188,189] Type 2 metal binding sites in ArsR family proteins have no exact consensus motif but metal ions are found to bind at the α5 helices.[193] One metal ion binding to each monomer is seen to be needed for DNA de-repression in ArsR family proteins. The helix-turn-helix (HTH) region formed in SmtB by α3-loop-α4 is used to bind to the major groove of the DNA.

Another ArsR family protein is SmtB [190] (Fig. 11) which was also an establishing member, meaning the family is sometimes referred to as the ArsR family. SmtB functions as homodimeric protein at 27 kDa to sense Zn^{2+} ions.[191,192] Whilst it does

have a type 1 metal binding site like ArsR this site was found to be non-functional from mutagenesis experiments of key residues in the binding motif which did not stop SmtB binding to Zn^{2+} and subsequently being released from DNA binding as predicted.[192] Instead Zn^{2+} binds at the carboxyl terminal of the α5 helices (the α5 sites) which is also known as a type 2 metal binding region,[193] which allows for conformational changes that results in release of the protein from DNA.[192,194]
A hybrid of these 2 major family members is thought to be CadC [195] which can sense 3 heavy metals- Cd^{2+}, Zn^{2+} and Pb^{2+}.[196] Like SmtB, there is a type 2 metal binding site for Zn^{2+} but rather interestingly, like in ArsR there is a type 1 metal binding site also.[197] The type 2 site in this case can have Zn^{2+} ions bound to it but this metal ion binding is non-essential to function as binding at the type 2 site does not release CadC from being bound to DNA .[198] Cysteines (C7, C58 and C60) play an important role also in the type 1 binding site for CadC like in ArsR but unlike ArsR this regulatory metal binding site is formed by the unusual feature of CadC whereby an α1 helix of one monomer crosses the dimer towards the α4 helix of the other monomer.[195] In ArsR the type 1 site is formed by the same monomer. [196] The unique feature of CadC having both type 1 and 2 sites that can bind to metal ions is suggestive of CadC as an intermediate between the ArsR and SmtB (Fig. 12).[195,199]

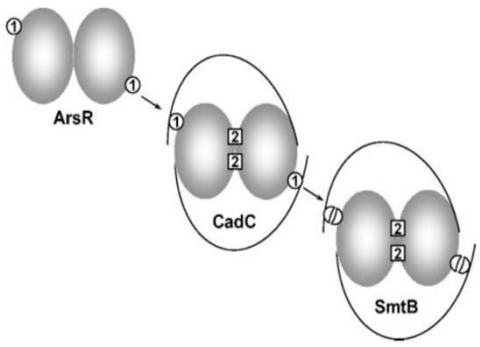


Figure 12. A possible evolution model type 1 and type 2 metal binding sites in ArsR family proteins. Type 1 metal binding sites (at the α3 sites of dimers) like that of ArsR are thought to have evolved first and then type 2 sites (at the α5 sites of dimers) which is predominantly used by SmtB. Between these is the possible intermediate CadC which can has both type 1 and type 2 sites and can utilise both. Figure taken from Ye *et al.* (2005).[195]

Taking these 3 major proteins of the ArsR family as examples, an evolution of the binding sites is seen going from use of a type 1 site in ArsR, to type 1 and 2 in CadC to only type 2 in SmtB (Fig. 12). Therefore, the ArsR family is clearly one that is diverse at least in its binding sites and so it is possible MmyJ, a non-metal sensing protein, could be part of it.

While there is diversity in the ArsR family, there are similarities that link these different proteins to the same family. All known ones thus far are transcriptional repressors. In general ArsR family genes are found upstream to the gene of the transporter for metal resistance they control[200] and the protein generally functions as symmetrical homodimers with each monomer having 5 alpha-helices and 2 beta-sheets (Fig. 11). Two of the helices create a winged typical DNA-binding motif of helix-turn-helix (HTH) which binds to the major groove of the DNA. Whereas the beta sheets of each monomer are antiparallel and in a DNA bound state forms a hairpin structure stabilising the complex through interactions with the minor groove of the DNA.[200]

In the currently accepted model of ArsR family protein functioning, the homodimer is found bound to operator/promoter DNA and ligand binding drives a quaternary structural change that stabilises low DNA-binding affinity allowing for operator/promoter transcription.[115,121] The exact DNA region of binding is not consistent within the ArsR family however for some of the metal-binding members one or two imperfect 12-2-12 inverted repeats near or overlapping the transcriptional start site of the metal-regulated gene is seen to be the binding region of the ArsR family protein.[201,202] Finally, many ArsR family proteins are considered autoregulatory meaning while they regulate the transcription of other genes, they also regulate their own transcription.[200,203]

MmyJ is a small protein (12.14 kDa) and contains an HTH motif which has allowed for DNA binding to be seen near the gene for the methylenomycin efflux pump,

similar to the binding of ArsR family proteins.[155] Furthermore, in the last few decades there have been ArsR family proteins reported to not bind metal ions but other ligands instead as will be highlighted in the next section.

3.1.2 Non-metal binding exceptions in the ArsR family

Recently exceptions to the class in terms of metal binding are coming to light. One of the better studied of these exceptions is the HlyU protein of both *Vibrio vulnificu* [204] and *Vibrio cholerae* (Fig. 13) [125] which upregulates expression of haemolysin (HlyA) required for virulence in these bacteria.[205] Despite not binding to any metal ions this protein is still classed as part of the ArsR family due to its high structural similarity to SmtB. HlyU also contains a winged HTH domain formed by α3- turn- α4 helices that interact with the major groove of DNA though in comparison to SmtB the homodimers of the HTH DNA binding domain are closer together in HlyU (27Å apart for HlyU_Vc and about 30Å for HlyU_Vv in comparison to 34Å for SmtB).[206,207]

Like other ArsR family proteins, it acts as a homodimeric transcriptional repressor binding to DNA until it can go through a conformational change that causes unbinding from the DNA. While the type 2 metal binding site is absent in HlyU, the type 1 site is partially conserved. However, this does not make it functional to bind metal ions as the 2 cystine residues that bind to metal ions in the ELCVCD motif in type 1 are substituted for Ser and Glu in HlyU. The ligand it binds is still unknown. Cysteines are still implied to be important for function as HlyU crystals grown without a reducing agent (TCEP) showed Cys38 being modified to cysteine sulfenic acid but in the reduced form remain as a free cysteine (Fig. 13).[125]

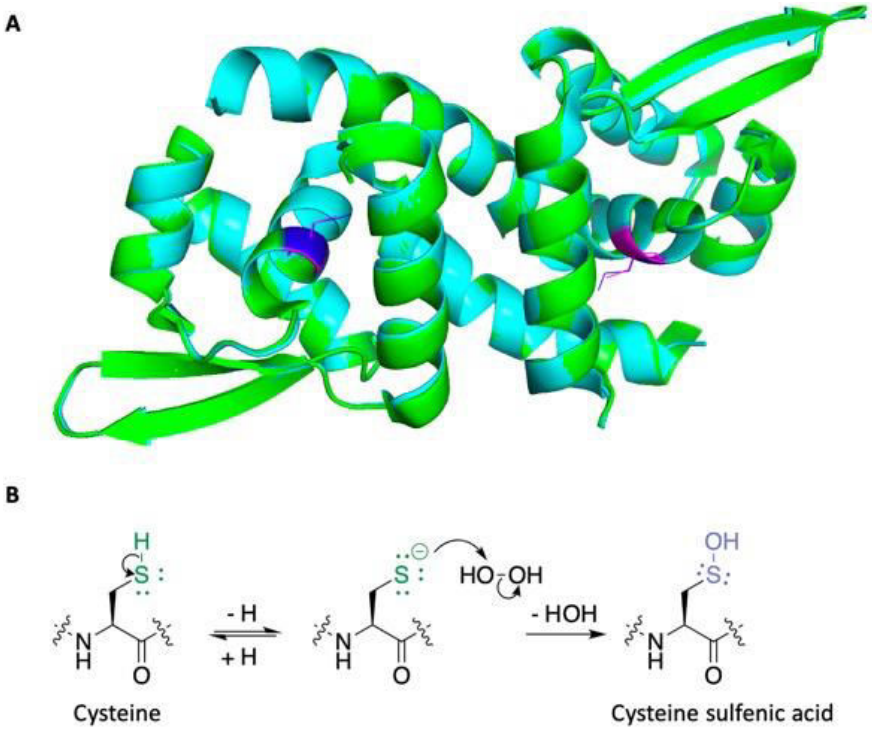


Figure 13. Oxidised and reduced structures of HlyU with Cys38 reduced or oxidised to cysteine sulfenic acid (CSO). Both oxidised (green-PDB 4K2E) and reduced (cyan- PDB 4OOI) HlyU[125] were superimposed on each other (**A**) and show no structural differences apart from Cys38 (blue in oxidised form and pink in reduced form). In the oxidised form Cys38 was seen to be a cysteine sulfenic acid. The scheme by which this occurs is shown in **B**.[209] This may indicate HlyU functions using a redox switch.

Both reduced and oxidised forms of apo-HlyU structures did not show any other structural differences (Fig. 13A) but this may be apparent during ligand or DNA binding. The observations of Cys38 in HlyU under oxidising and reducing conditions suggests HlyU acting as a transcriptional protein, binding and unbinding from DNA, may be in part controlled by a redox switch.[125,204,208]

This redox switch action by ArsR family proteins is not uncommon as another one of these exceptions to the class acts in a similar fashion. Plant pathogens *Xylella fastidiosa* and *Agrobacterium tumefaciens* express a protein involved in biofilm growth known as BigR.[209] This is another ArsR family protein that contains the typical wHTH motif and acts as a transcriptional repressor for the operon involved

in biofilm growth. Within this operon is its own gene meaning it regulates itself also.[210]

Apart from MmyJ, 2 other antibiotic linked ArsR family proteins have been proposed. DepR2 from *Streptomyces roseosporus* has been shown to act as a negative regulator to the antibiotic daptomycin by binding to the promotor of *dptEp*, a gene involved in the promotion of daptomycin biosynthesis.[211] However, the ligand that may bind to DepR2 to release it from DNA has not yet been suggested so whether the ligand is daptomycin or not remains to be seen. Similarly, BlmR from *Streptomyces verticillus* negatively regulates the biosynthesis of the antibiotic bleomycin but also autoregulates itself. The ligand BlmR binds to is also unknown but it has been shown that this is not bleomycin.[202] Both of these have the wHTH structures seen in ArsR family proteins and are considered to act as homodimers.[202,211]

3.1.3 Evolution of MmyJ from ArsR family proteins

Now that the diversity in ArsR family proteins is being seen (in terms of ligand sensing, non-consensus region of DNA binding and structural differences), it stands to reason to question where MmyJ may lie evolutionarily. The answer to this could help to understand the structure and functionality of MmyJ.

To study this, a phylogeny tree was produced (see Chapter 2: Materials & Methods section) by using the amino acid sequences of the previously discussed non-metal binding proteins, other ArsR family proteins and some non-ArsR family proteins (Fig. 14). These non-ArsR family proteins are established proteins (1 each) from the other 6 metal sensing transcriptional regulator families.[121] Results show that while BLAST searches and alignments may class MmyJ as an ArsR family protein, evolutionarily MmyJ may not be too similar to the typical ArsR family protein (Fig. 14). Interestingly MmyJ has a sister branch to PerR which is a member of the Fur family metal regulatory proteins.[212] Like the ArsR family proteins, the Fur family are

also a diverse group of metal regulators that act as co-repressors rather than derepressors. Also seen are the two mentioned implied antibiotic regulating ArsR family proteins (BlmR and DepR2) perhaps evolving at a similar time whereas the other non-metal regulating ArsR-like proteins (BigR and HylU) emerge from a sister branch to ArsR itself. Even SmtB and CadC from this figure is implied to have formed much later after ArsR itself and is as close evolutionarily to CsoR[213] (part of the CsoR family of metal sensing families in bacteria) as ArsR (Fig. 14B). Rather than implying MmyJ is not part of the ArsR family this instead seems to highlight the close relatedness of all 7 metal sensing families in bacteria. Certainly, the alignment of domains of these families (Fig. 14A) appear to be similar with one or two exceptions. Fig. 14B also shows the diversity within these metal sensing families, especially within the ArsR family.

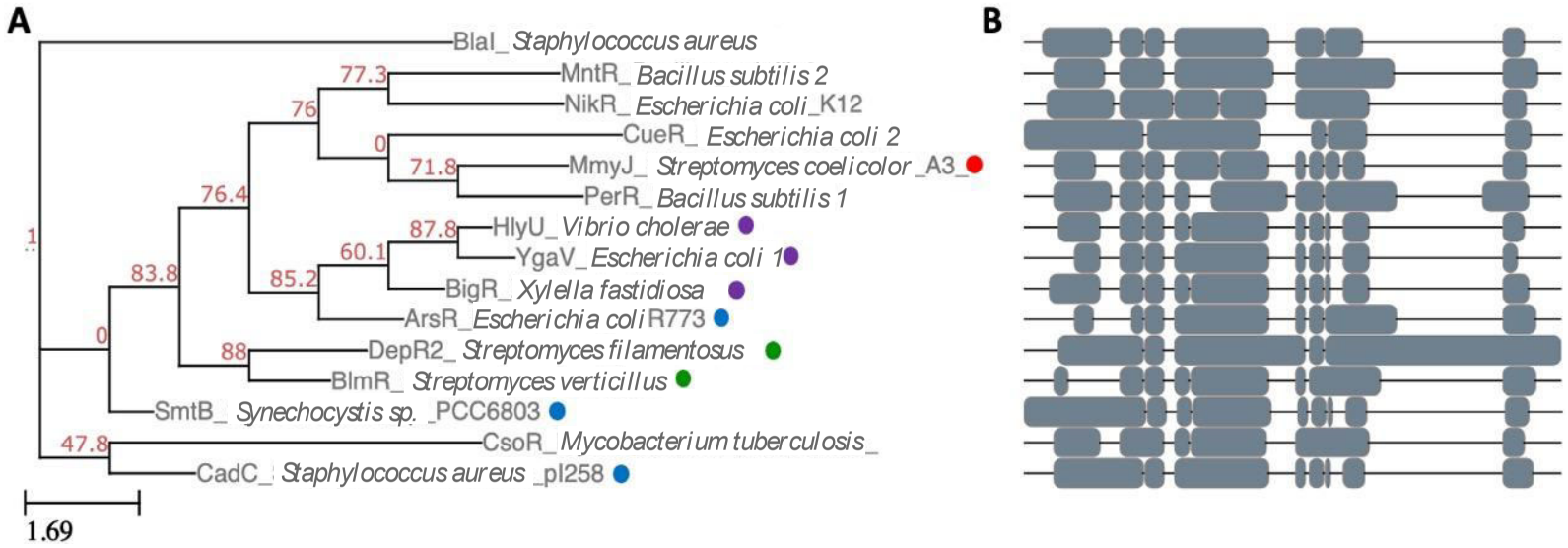


Figure 14. Phylogeny and relation of MmyJ to other metal-sensing transcriptional proteins. MmyJ has been implied to be closely related to the ArsR family group of proteins. **A**) The phylogeny tree (Chapter 2: Materials & Methods) of MmyJ (red dot), ArsR family proteins that are metal sensing (blue dots), antibiotic regulation related (green dots) and non-metal sensing (purple dots). One representative member from the other 6 metal-sensing transcriptional regulator families was also taken (no dots). **B)** The amino acid sequence alignment domains of these proteins in block form. **A** implies MmyJ is not as evolutionarily similar to ArsR family proteins as the sequence alignments show (**B**).

This figure is limited in the range of proteins being analysed and additional sequences, particularly diverse proteins within the metal sensing families, would assist in giving a more accurate view as to where MmyJ may lie evolutionarily.

Furthermore, a phylogenetic tree is built upon the alignment and analysis of the primary structure of proteins. The amino acid of these proteins does not necessarily give a precise view of the final folded protein and thus the functioning may be different to than that implied by the phylogenetic tree. Finally, gaps and errors (e.g. sequencing errors, annotation errors or redundant data)[214] may be present within the databases used which could be translated into the phylogenetic tree created, meaning some evolutionary links between proteins are over or under represented.[215] Nonetheless phylogenetic trees of proteins can be helpful in visualising aligned domain data and tracking the ancestorial roots of certain unknown proteins. This can aid in elucidation of the mechanism of action and purpose of these unknown proteins by experimental data.[216]

3.1.4 MmyJ-like systems are present in other actinomycetes

Already DepR2 and BlmR have been listed as 2 antibiotic sensing ArsR family proteins that may function like MmyJ. BlmR in particular however does not sense metals or the antibiotic bleomycin.[202] This is unlike MmyJ which is thought to be responsive to the methylenomycin antibiotic. The ligand for DepR2 has also not yet been shown.[211]

Actinomycetes (the genus to which *Streptomyces* species are part of) are known producers of antibiotics and confer resistance to them too (Chapter 1). Searching for any MmyJ-like proteins or systems that function like the theorised MmyJ system in actinomycetes may give insight into the function of MmyJ. Tracking homologous MmyJ proteins in other actinomycetes could also indicate the prevalence of MmyJ-like transcriptional regulators.

First a search for proteins similar to MmyJ through BLAST analysis was done and the top 40 proteins with the highest percentage identity taken (Table 10). The higher the percentage of identity the more similar this amino acid sequence is to that being searched and so the more significant the match should be. Many of these matches come from various *Streptomyces* species themselves and are described as 'helix-turn-helix transcriptional regulators' or 'ArsR family transcriptional regulator'. These sequences were aligned using MUSCLE (a multiple sequence alignment tool)[157] and a phylogeny tree plotted (Fig. 15). A phylogeny tree like this comparing similar proteins of various species can help to indicate the evolutionary relationship between species in regard to the protein being studied and also show any anomalies that may occur of this protein in certain species. In this case it is also particularly helpful to visualise if MmyJ is a common protein found even in many *Streptomyces* species and if the variants in other actinobacteria species are evolutionarily distant or close.

Table 10: BLAST search for proteins similar to MmyJ. Basic Local Alignment Search Tool (BLAST) was used to find proteins that are similar in amino acid sequence to MmyJ. The top 40 were taken for further analysis. Results are sorted in descending order according to the percentage sequence identity (% identity) which is indicates the similarity of two proteins. High % identity and cover show a protein that's highly similar to the query protein.

Description	Species	% Query Cover	% Identity	NCBI accession number
MmyJ	*Streptomyces coelicolor* A3(2)	-	-	AAA98340.1
MULTISPECIES: helix-turn-helix domain-containing protein	*Streptomyces*	100	99.1	WP_164272960
helix-turn-helix transcriptional regulator	*Streptomyces sp.* NRRL S-31	100	91.8979.09	WP_030742020.1
helix-turn-helix transcriptional regulator	*Streptomyces sp.* 846.5	99	77.48	WP_133995155.1
helix-turn-helix transcriptional regulator	*Streptomyces caeruleatus*	100	75.45	WP_062723279.1
ArsR family transcriptional regulator	*Streptomyces sp.* 351MFTsu5.1	99	70.27	WP_020135630.1
MULTISPECIES: helix-turn-helix transcriptional regulator	*Streptomycetaceae*	100	72.82	WP_030286058.1
transcriptional regulator	*Kitasatospora xanthocidica*	92	72.82	GHF69299.1
helix-turn-helix domain-containing protein	*Kitasatospora xanthocidica*	92	72.82	WP_189921066.1
helix-turn-helix transcriptional regulator	*Amycolatopsis roodepoortenisis*	92	75.73	WP_192746507.1
helix-turn-helix transcriptional regulator	*Amycolatopsis thermoflava*	94	73.33	WP_123685663.1
MULTISPECIES: helix-turn-helix transcriptional regulator	*Amycolatopsis*	94	72.38	WP_020419963.1
helix-turn-helix transcriptional regulator	*Amycolatopsis sp.* WAC 04182	92	73.79	WP_125680327.1
helix-turn-helix transcriptional regulator	*Streptomyces sp.* AcE210	90	73	WP_116508238.1
helix-turn-helix domain-containing protein	*Streptomyces sp.* NBS 14/10	100	63.96	WP_179285526.1
helix-turn-helix transcriptional regulator	*Streptomyces sp.*	100	63.96	NUP40258.1
helix-turn-helix domain-containing protein	*Streptomyces griseorubiginosus*	93	74.04	WP_208778111.1
helix-turn-helix transcriptional regulator	*Rhodococcus sp.* P1Y	94	74.29	WP_121115275.1
helix-turn-helix domain-containing protein	*Amycolatopsis deserti*	96	68.22	WP_191248248.1
transcriptional regulator	*Amycolatopsis deserti*	96	68.22	GHF19263.1
helix-turn-helix transcriptional regulator	*Streptomyces sp.* SA15	95	70.75	WP_095750843.1
helix-turn-helix transcriptional regulator	*Streptomyces*	100	63.06	WP_043486070.1
helix-turn-helix domain-containing protein	*Kitasatospora mediocidica*	89	70.71	WP_198045424.1
helix-turn-helix transcriptional regulator	*Amycolatopsis australiensis*	89	73.74	WP_084743248.1
DNA-binding transcriptional regulator, ArsR family	*Amycolatopsis australiensis*	89	73.74	SFW78652.1
helix-turn-helix transcriptional regulator	*Streptomyces mirabilis*	95	69.81	WP_078936472.1
hypothetical protein	*Streptomyces nachangensis*	90	68	ADC45569.1
helix-turn-helix transcriptional regulator	*Streptomyces bauhiniae*	95	69.81	WP_135785704.1
DNA-binding transcriptional ArsR family regulator	*Streptomyces sp.* 59	95	68.87	PKV36389.1
helix-turn-helix transcriptional regulator	*Streptomyces niger*	90	72	WP_052868623.1
MULTISPECIES: helix-turn-helix transcriptional regulator	unclassified *Streptomyces*	95	68.87	WP_097254997.1
helix-turn-helix domain-containing protein	*Streptomyces sp.* 3R004	99	62.73	WP_194082112.1
ArsR family transcriptional regulator	*Streptomyces sp.* TLI_55	99	61.82	WP_927272007.1
helix-turn-helix transcriptional regulator	*Streptomyces antibioticus*	99	62.73	WP_059191091.1
helix-turn-helix transcriptional regulator	*Nonomuraea candida*	96	64.49	WP_043620022.1
transcriptional regulator	*Streptomyces sp.* NBS 14/10	83	69.89	OXL26002.1
ArsR family transcriptional regulator	*Streptomyces aquilus*	99	61.82	WP_126275929.1
ArsR family transcriptional regulator	*Streptomyces sp.* BK042	99	61.82	WP_132796641.1
transcriptional regulator	*Planotetraspora silvatica*	100	62.16	GII46859.1
helix-turn-helix domain-containing protein	*Phytoactinopolyspora* halotolerans	91	62.75	WP_163738103.1

This unrooted tree (Fig. 15B) shows there to be around 6 distinct groups of these proteins similar in percentage identity to MmyJ. MmyJ itself appears to be part of one these groups with 3 other proteins that have the highest percentage sequence identity to it (Fig. 15A). This is expected and coming from other *Streptomyces* species it is likely that the *mmyJ* gene may have been transferred at some point evolutionarily soon (making them perhaps MmyJ orthologues) in comparison to the other species and possible MmyJ variants. The phylogeny tree also indicates the early ancestor from which *S. coelicolor* A3(2) MmyJ evolved from may be *Phytoactinopolyspora halotolerans* and *Nonomuraea candida*. Both are not well characterised soil dwelling actinobacteria with *P. halotolerans* as the name suggests able to tolerate high salt levels[217] while *N. candida* showed weak antibiotic activity against *Mycobacterium aurum*[218]. Discussed in Chapter 1.3 was the possible co-evolution between heavy metal resistance and antibiotic resistance; from the presence of *P. halotolerans* as a possible ancestor of *S. coelicolor* A3(2) for MmyJ evolution, antibiotic resistance regulators may also have evolved from bacteria tolerant to soil environments high in a variety of minerals and ions.

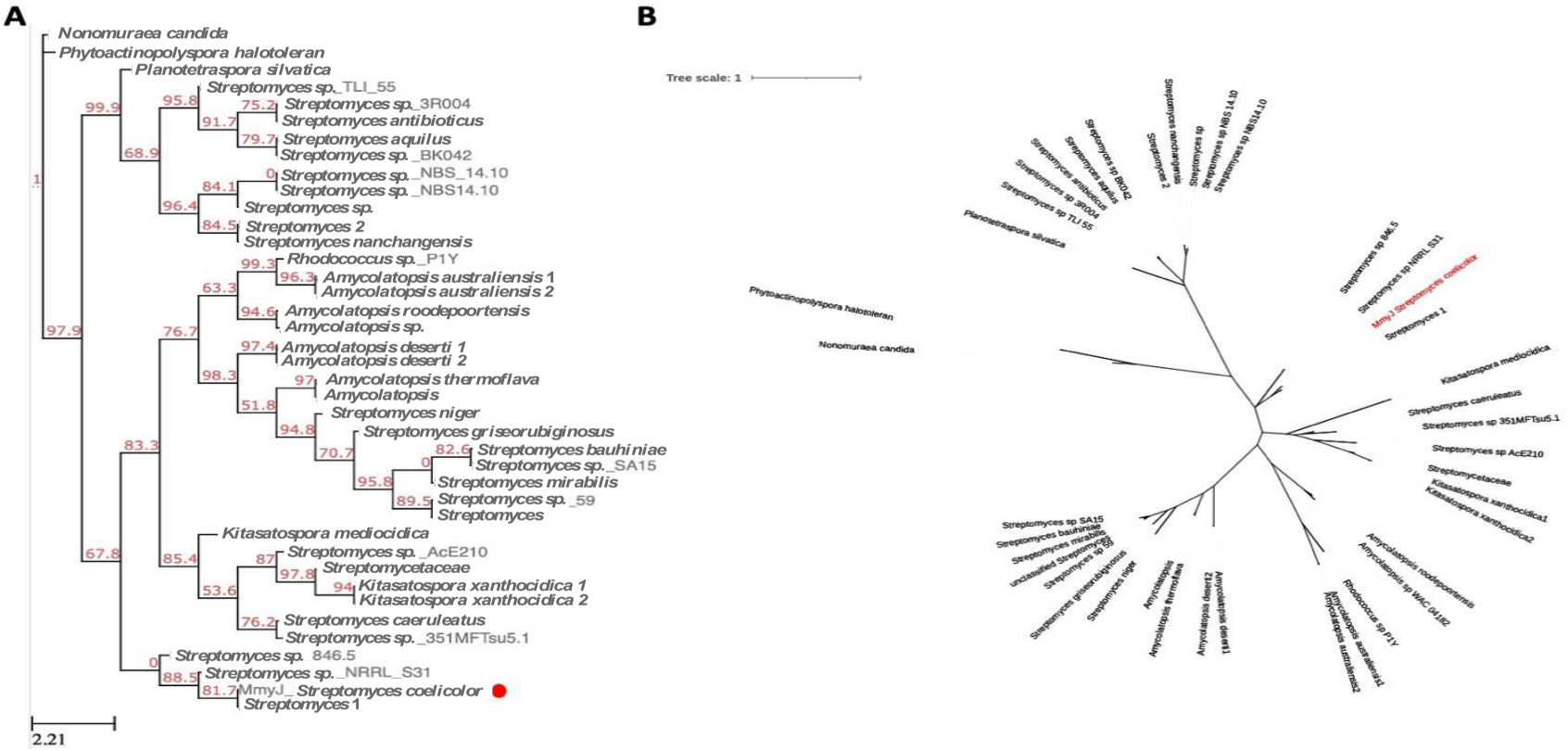


Figure 15. Phylogeny trees of MmyJ to closely related proteins. Phylogeny trees were created (Chapter 2: Materials & Methods) of MmyJ and the top 40 proteins it's similar to shown through BLAST analysis (Table 10). **A**) Shows a rooted tree (Chapter 2: Materials & Methods) and **B**) shows an unrooted tree (Chapter 2: Materials & Methods). Many related proteins were uncharacterized and had no name so the species they were identified in are instead used. MmyJ (red to (**A**) and in red (**B**)) found in *Streptomyces* seems to have similar proteins in many other actinobacteria that's not *Streptomyces* species. These similar proteins also appear form separate groups evolutionarily and the two or three other *Streptomyces* species proteins MmyJ is a part of may be MmyJ orthologues.

To identify if any of the proteins from the BLAST search are also part of *mmyJ*-like gene clusters or have divergent genes that they could be controlling, the synteny of these proteins can be considered. This in turn can highlight any bacteria that may have MmyJ-like systems.

The first 100 results from the BLAST search (Table 10) were input into the online server 'WebFlAGs'[219] to find conservation in the flanking genes between these proteins (Fig. 16). Results showed that there were many instances of divergent genes to the '*mmyJ*-like' genes and that these could be largely split into 2 categories; NADP-dependent oxidoreductase or zinc-dependent alcohol dehydrogenases (genes labelled 1 in Appendix Fig. A1) or MFS transporters (genes labelled 2 in Fig. 16). (The full list of the descriptions of these labelled genes are presented in Appendix Table A1). Mmr is annotated as an MFS transporter and so the identification of '*mmyJ*-like' genes divergent to genes described as MFS transporters indicates that these may be '*mmr*-like' genes. This synteny is seen in proteins in the top 50% of the BLAST search results (sorted by highest to lowest E-value). As expected, these MmyJ like proteins beside MFS transporters are seen mainly in *Streptomyces* species bacteria and while there are some exceptions (e.g., *Amycolatopsis* bacteria), these are all actinobacteria.

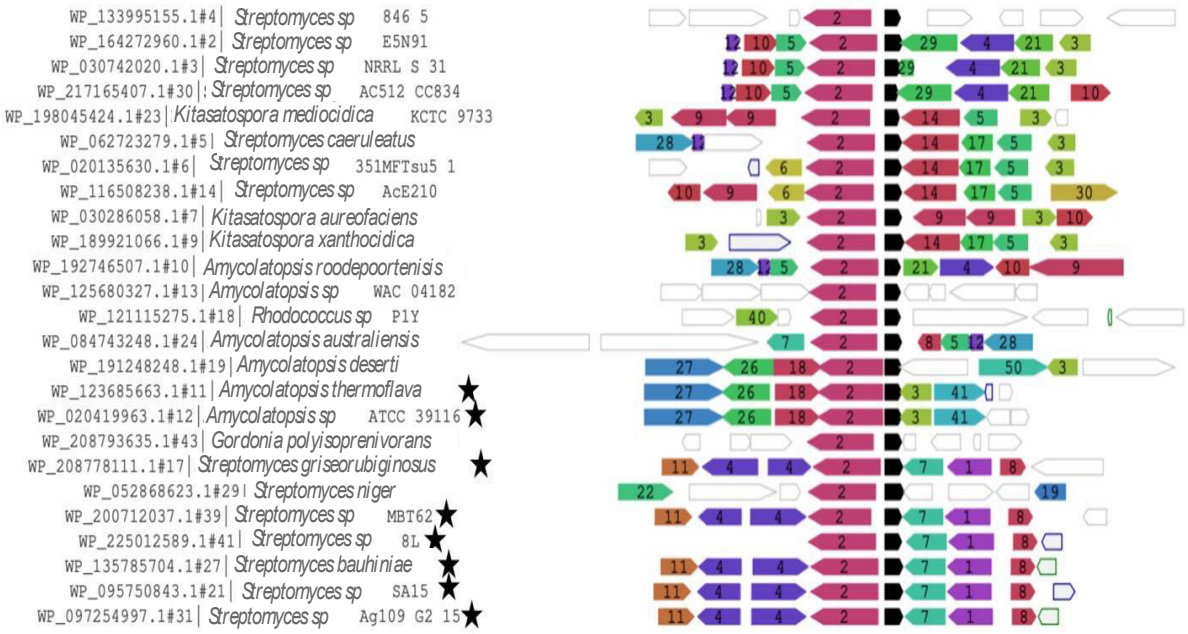


Figure 16. Organisms identified to have a protein similar to MmyJ with genes beside it encoding for MFS transporters. The top 100 results identified by protein-protein BLAST searching (Table.10) were input into WebFlAGs to find the genes flanking these *mmyJ*-like genes. Accession numbers from the NCBI database are given first, then the number this occurred in the BLAST search, followed by the species it came from and finally the flanking genes to the *mmyJ*-like gene (in black). A majority of these results contained genes beside the *mmyJ*-like one (in black) that were described as MFS transporters (labelled 2). Starred are genes on the other side of the *mmyJ*-like ones that are described as 'TetR family transcriptional regulators' (labelled 3) and 'helix-turn-helix domain-containing protein' (labelled 7), the latter of which also contains genes further along from the *mmyJ*-like one described as 'SARP family transcriptional regulators' (labelled 11). The rest of the labelled genes are given in full in Appendix Table A1.

The presence of these MFS transporter genes beside a *mmyJ*-like gene may show antibiotic resistance and regulation systems in other actinobacteria. This could be evidence of the highly transmissible nature of resistance genes between bacteria and showcase the different strategies of antibiotic resistance accumulated by even antibiotic producing bacteria. There is more variation in the type of genes on the other side of the *mmyJ*-like gene though it is interesting to note that some of these (starred in Fig. 16) have been described as transcriptional regulators which raises

the question as to why there may two transcriptional regulators beside each other. Many transcriptional regulators are found within the same gene cluster for antibiotic production and resistance, and it may be that this tight regulation is for the production of valuable secondary metabolites. Also seen in these organisms with the *mmyJ*-like gene, another transcriptional regulator gene, and the MFS transporter gene, is the presence of an SARP transcriptional regulator (labelled 11 in Fig. 16) 2 genes prior to the *mmyJ*-like one. SARPs are *Streptomyces* specific genes and several have been shown to be involved in antibiotic BGCs.[28,47,54,55,64] While these results have been limited to studying only the gene beside the MmyJ-like protein for potential resistance and resistance regulation systems, like in *S. coelicolor* SCP1 with *mmyJ* and *mmr*, these genes may be part of a larger cluster for antibiotic or other natural product production. Characterisation of MmyJ could also assist in determining if any of the *mmyJ*-like genes suggested in these actinobacteria are truly homologous to MmyJ.

3.2 Producing pure recombinant MmyJ for structural study

In order to determine the function of MmyJ, the structure and interactions of it with other molecules can be studied. Structural data which can show secondary, tertiary and quaternary structures can be useful in viewing active sites and chemical interactions occurring not only within the protein but also between other subunits and ligands of the protein. Furthermore, these may show conformational changes occurring that assist the protein in its function.

There are many ways to study the structure of a protein with the first techniques used dating back to the 1950s. By far the most common technique used to this day for atomic resolution 3D structure of proteins is X-ray crystallography. This uses pure proteins that are crystallised being subjected to an X-ray beam which causes the beam to diffract. Analysing this diffraction pattern can produce an electron density map of the crystallised protein and from there atomic bonds can be assigned to produce a 3D picture of the protein.[220]

While large parts of this process are now automated, the technique still relies on having pure crystallised protein which in itself can be challenging.

3.2.1 Expression of recombinant MmyJ

To produce pure MmyJ protein in high enough concentrations for structural analysis, an inducible *E. coli* system is used to express a recombinant variation of the MmyJ protein. This is a commonly used, economic and fast system where a plasmid with the recombinant gene containing the *lac* operator with a T7 promoter site is transformed into *E. coli.* Without allolactose being present the LacI binds to the *lacO* site and prevents transcription of the recombinant gene however when isopropyl β-D-1-thiogalactopyranoside (IPTG) is added, it mimics the action of allolactose and binds to LacI dissociating it from the DNA and allowing RNA polymerase to bind to the T7 promoter and express the recombinant protein of interest.[221]

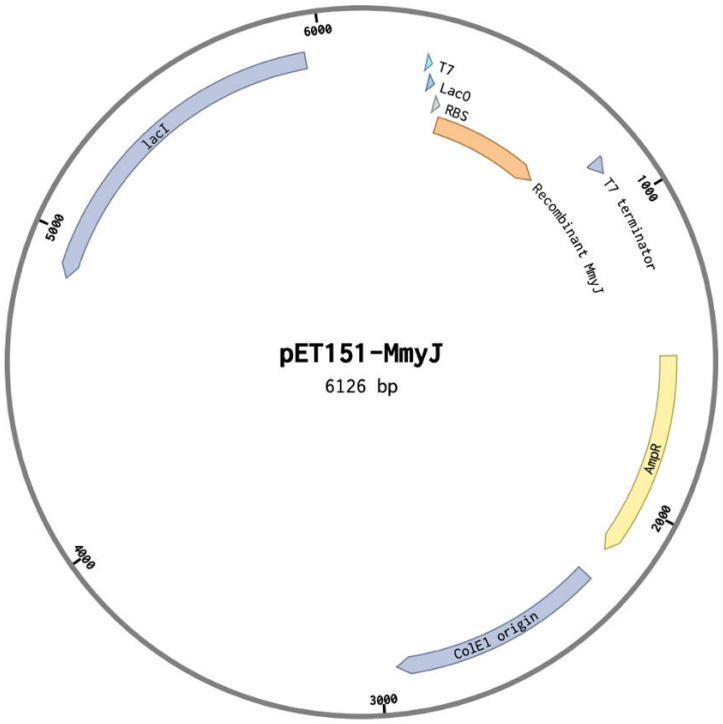
Wild type MmyJ (WT MmyJ) is a 12.1 kDa protein containing a single cysteine residue at position 49. When overexpressed this may cause disulphide bonds between monomers to be formed where they otherwise would not be. In order to avoid this, the cysteine residue can be mutated to serine, an amino acid analogous to cysteine but rather than a -SH group at the end has an -OH group.

The pET151 expression system is considered a robust plasmid for overproduction of protein especially in *E. coli* BL21 (DE3) Star cells. The Thermofisher commercial kit allows for the protein construct to be expressed containing a 6 Histidine tag, a V5 tag and a TEV site on the N-terminal (Chapter 2: Materials & Methods). This recombinant protein can therefore be purified using the fused tags as well as these tags being able to be cleaved later with use of TEV protease recognising the TEV site. This plasmid also contains an ampicillin resistance marker/gene to allow for antibiotic selection once transformed into *E. coli*.

Both WT MmyJ and C49S MmyJ were cloned into the pET151 vector by M. Lougher (2015)[155] and available for use within the Corre group (Fig. 17). From the tagged site as would be expressed by recombinant protein they are both 144 amino acids in length and an estimated molecular weight of 15756 Daltons (MmyJ WT) and 15740 Daltons (C49S MmyJ), have a pI of 6.3 and an extinction coefficient at 280 nm of 4470 $M^{-1}cm^{-1}$. The instability index of a protein is the measure of whether or not the protein will be stable in a test tube at room temperature. Proteins with instability indexes less than 40 are classified as being likely to be stable. Instability indexes of MmyJ WT and C49S MmyJ were calculated using ProtParam from Expasy.[222] The MmyJ WT construct is considered unstable with an instability index of 41.59 however the C49S MmyJ has one of 38.54 and is therefore considered more stable.

Transformation of C49S MmyJ and WT MmyJ pET151 plasmids into *E. coli* BL21 (DE3) Star was validated by colony PCR using a set of primers designed to amplify a 366 base pair (bp) region of the plasmid containing the WT or C49S *mmyJ* gene. The integrity of the expected *mmyJ* gene was also validated by Sanger sequencing (Appendix Fig. A2).

A



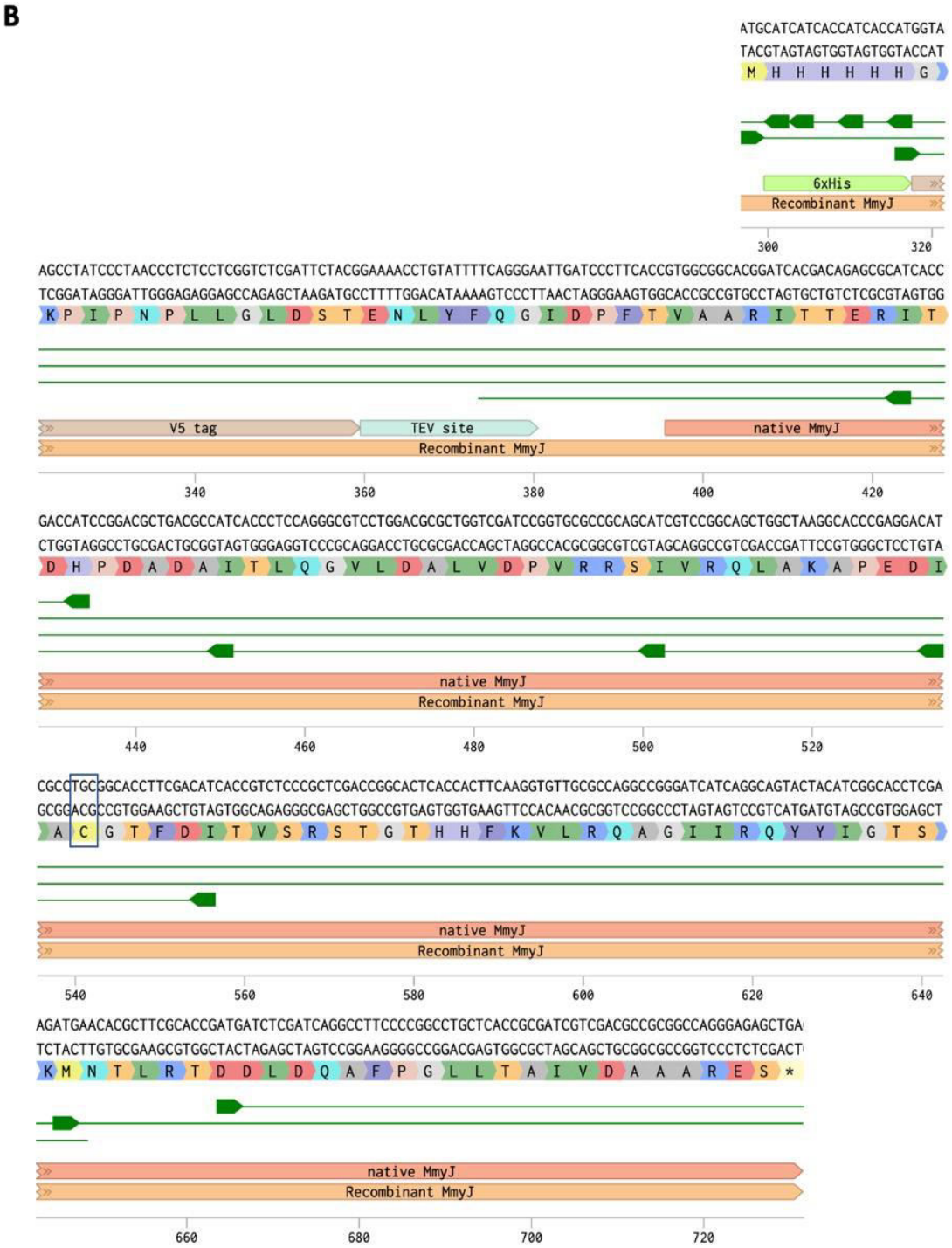


Figure 17. Recombinant MmyJ used for expression of MmyJ. A) The pET151-MmyJ plasmid was cloned by Lougher M. (2015)[155] and available for use within the lab group. The main features of this plasmid are shown in **A** with the *lac* operon for induced induction of protein production by addition of IPTG before the RBS for the recombinant MmyJ. The recombinant MmyJ being expressed consists of a 6 histidine tag (6xHis), a V5 tag, a TEV site and the native MmyJ sequence (**B**). Also shown in **B** is the only cysteine (C49) highlighted (blue box). A point mutation of this cysteine to serine (C49S) was also created by Lougher M.[155] in an identical pET151 plasmid for expression of recombinant C49S MmyJ.

3.2.2 Production of pure MmyJ protein

In order to assess the level of over-production of WT MmyJ and C49S MmyJ small scale (i.e., 50 mL) cultures were made and induced cultures were compared to non-induced cultures (Chapter 2: Materials & Methods). Fractioning the cultures into fractions of supernatant, cytosolic components, and insoluble aggregates (Chapter 2: Materials & Methods) also showed if the protein aggregates and therefore is found mostly or all in the insoluble fraction. In that case the construct may have to be redesigned or buffers fine-tuned to prevent this. Both WT MmyJ and C49S MmyJ despite WT MmyJ having a high instability index were overproduced in the soluble protein fraction (Appendix Fig. A3).

This meant production could be scaled up to 1 L cultures and purification could begin. Immobilized metal affinity chromatography (IMAC) purification with use of Ni-NTA beads that allow for the 6xHis tag of the protein to bind was first utilised (Chapter 2: Materials & Methods). Buffers were optimised (Chapter 2: Materials & Methods) through multiple trials so purification yielded in only a few strong, clean bands with one of which hoped to be MmyJ protein (Fig. 18).

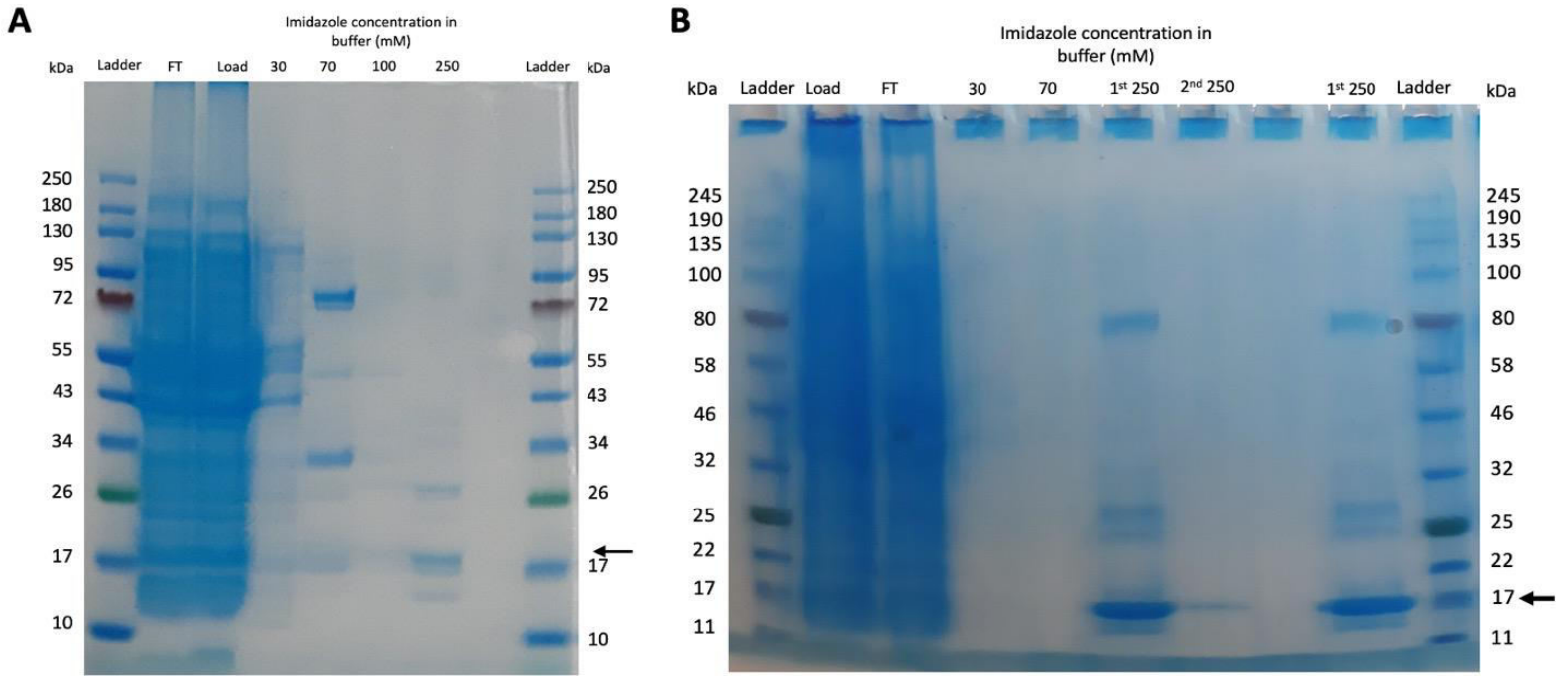


Figure 18. Recombinant MmyJ protein purification by immobilized metal affinity chromatography (IMAC) specifically nickel-NTA. A 1 L culture of both recombinant MmyJ (**A**) and C49S MmyJ (**B**) were purified through IMAC. Buffers with various concentrations of imidazole were used to elute the proteins of interest (Chapter 2: Materials & Methods). FT refers to the flow through which is the elution after loading the column. Load refers to the sample loaded onto the column. The 1st and 2nd 250 mM imidazole buffer samples refer to samples eluted by one 3 mL addition of 250 mM imidazole buffer then a second 3 mL elution of 250 mM imidazole buffer (**B**). The two samples labelled '1st 250' are the samples that were run twice to have one sample beside the ladder for easier weight determination (**B**). Samples were run on a gradient SDS-PAGE gel (Chapter 2: Materials & Methods). Black arrows indicate the likely fraction where recombinant MmyJ protein is seen which is lightly larger than the expected weight of recombinant WT-MmyJ (15.75 kDa) and of C49S MmyJ (15.73 kDa).

Consistently seen over multiple IMAC purifications of WT MmyJ and C49S MmyJ was a band at around 17 kDa (indicated by black arrows Fig. 18) and a band at roughly 80 kDa. Faint bands at 30 kDa were also detected in some trials however this band tended to fade when the amount of SDS-sample buffer was added as 2X instead of 1X (Fig. 19A compared to Fig. 18) indicating this band is possibly dimers or complexes of protein that was not being fully reduced with only 1X SDS buffer.

Although both MmyJ constructs are ~16 kDa, depending on the charges on the outside of the protein, it may run slower or faster through the SDS-PAGE gel by electrophoresis. A more negative outer surface area of MmyJ can enhance the negative charge of SDS and cause it to run faster towards the positive charge induced by electrophoresis meaning it appears at a lower kDa than expected whereas a more positive outer charge can cause it to run slower and appear greater in weight. Therefore the ~17 kDa band was thought to be the monomeric WT/C49S MmyJ but would have to be confirmed by mass spectrometry. Should these ~17 kDa bands be WT/C49S MmyJ it would stand to reason, the bands around 30 kDa which faded with additional SDS may be strong dimers of WT/C49S MmyJ. Dithiothreitol (DTT) was added to samples prior to boiling at 50 mM concentrations (Chapter 2: Materials & Methods) to reduce any disulphide bridges in samples, however only in WT MmyJ would this be possible as the only cysteine required for disulphide bond formation is mutated in C49S MmyJ. Therefore, this dimerization of the 17 kDa product is unlikely to be from disulphide bonds.

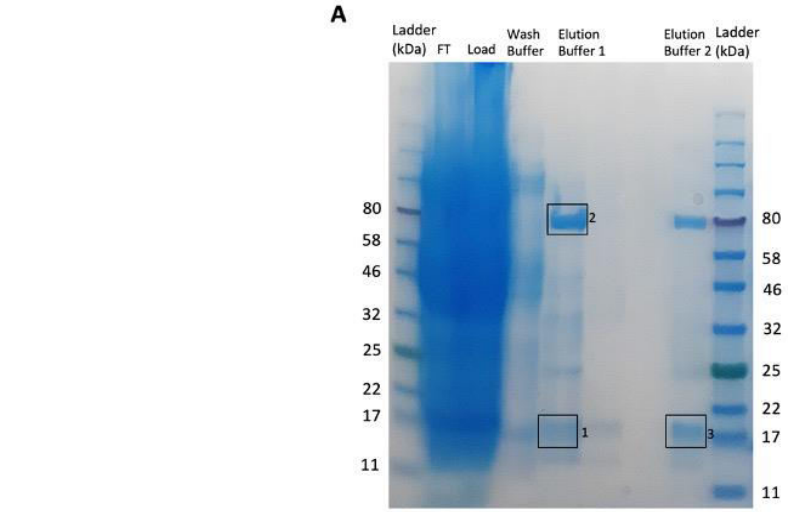
The ~80 kDa band which did not fade in intensity with addition of higher concentrations of SDS in samples and present in both purification of C49S MmyJ and WT MmyJ was suggested to be a contaminant protein.

To identify the proteins being purified in the IMAC, the 80 kDa and 17 kDa bands seen consistently during IMAC purification were excised (Fig. 19A), extracted from the gel and submitted for targeted proteomics after tryptic digest (Chapter 2: Materials & Methods). Results looking at the percentage of total spectra which is the percentage of peptides detected from the entire spectra that could be assigned

to that protein, showed detection of MmyJ peptides in the 17 kDa fractions but not in the 80 kDa fraction (Fig. 19B). This suggests the higher 80 kDa band to be a contaminant protein presumably produced by *E. coli* itself with high histidine or positively charged amino acid content on the surface causing strong binding to the

IMAC.

In order to separate this contaminant protein from the recombinant MmyJ proteins, size exclusion chromatography (SEC) was used with buffer exchange into 'MmyJ Crystallisation Buffer' (Chapter 2: Materials & Methods) done at the same



B

Probability Legend:
over 95%
80% to 94%
50% to 79%
20% to 49%
0% to 19%

Bio View:
25 Proteins in 25 Clusters

#	Visible?	Starred?		Accession Number	Alternate ID	Molecular Weight	Protein Grouping Ambiguity	S1	S2	S3
1	✓	★	acterized protein OS=Streptomyces coelicolor (strain ATCC BAA-471 / A3(2) / M145) OX=100226 GN=mmyJ PE=4 SV=10	Q7APH9_STRCO	mmyJ	12 kDa		2.7%		6.2%
2	✓		Succinate--CoA ligase [ADP-forming] subunit beta OS=Escherichia coli (strain K12) GN=sucC PE=1 SV=1	P0A836	sucC	41 kDa			1.1%	
3	✓		Ribonuclease E OS=Escherichia coli (strain K12) GN=rne PE=1 SV=6	P21513	rne	118 k...			1.8%	
4	✓		RecBCD enzyme subunit RecD OS=Escherichia coli (strain K12) GN=recD PE=1 SV=2	P04993	recD	67 kDa			0.72%	
5	✓		RNA-binding protein Hfq OS=Escherichia coli (strain K12) GN=hfq PE=1 SV=2	P0A6X3	hfq	11 kDa		1.2%		
6	✓		Outer membrane protein F OS=Escherichia coli (strain K12) GN=ompF PE=1 SV=1	P02931	ompF	39 kDa			0.72%	
7	✓		Outer membrane protein A OS=Escherichia coli (strain K12) GN=ompA PE=1 SV=1	P0A910	ompA	37 kDa			0.72%	
8	✓		NADH-quinone oxidoreductase subunit C/D OS=Escherichia coli (strain K12) GN=nuoC PE=1 SV=3	P33599	nuoC	68 kDa			1.4%	
9	✓		Malate dehydrogenase OS=Escherichia coli (strain K12) GN=mdh PE=1 SV=1	P61889	mdh	32 kDa			1.4%	
10	✓		Major outer membrane lipoprotein Lpp OS=Escherichia coli (strain K12) GN=lpp PE=1 SV=1	P69776	lpp	8 kDa		0.78%		
11	✓		Glutamine--fructose-6-phosphate aminotransferase [isomerizing] OS=Escherichia coli (strain K12) GN=glmS PE=1 SV=4	P17169	glmS	67 kDa			13%	
12	✓		Ferric uptake regulation protein OS=Escherichia coli (strain K12) GN=fur PE=1 SV=1	P0A9A9	fur	17 kDa		5.8%		3.1%
13	✓		Elongation factor Tu 2 OS=Escherichia coli (strain K12) GN=tufB PE=1 SV=1	P0CE48	tufB	43 kDa			3.2%	
14	✓		D-lactate dehydrogenase OS=Escherichia coli (strain K12) GN=dld PE=1 SV=3	P06149	dld	65 kDa			0.72%	
15	✓		CON__P35908	CON__P35908 (+1)		?	★	1.6%	2.2%	1.6%
16	✓		CON__P35527	CON__P35527		?		1.2%	1.8%	
17	✓		CON__P13645	CON__P13645		?		0.78%	1.4%	
18	✓		CON__P04264	CON__P04264		?	★	1.6%	1.8%	1.6%
19	✓		CON__P02538	CON__P02538 (+1)		?			0.72%	
20	✓		CON__P00761	CON__P00761		?		1.2%	1.1%	2.3%
21	✓		Bifunctional polymyxin resistance protein ArnA OS=Escherichia coli (strain K12) GN=arnA PE=1 SV=1	P77398	arnA	74 kDa			13%	
22	✓		ATP synthase subunit alpha OS=Escherichia coli (strain K12) GN=atpA PE=1 SV=1	P0ABB0	atpA	55 kDa			0.72%	
23	✓		30S ribosomal protein S12 OS=Escherichia coli (strain K12) GN=rpsL PE=1 SV=2	P0A7S3	rpsL	14 kDa		0.78%		
24	✓		2-dehydro-3-deoxyphosphooctonate aldolase OS=Escherichia coli (strain K12) GN=kdsA PE=1 SV=1	P0A715	kdsA	31 kDa			0.72%	
25			50S ribosomal protein L17 OS=Escherichia coli (strain K12) GN=rplQ PE=1 SV=1	P0AG44	rplQ	14 kDa				7.8%

Figure 19. Recombinant MmyJ was subjected to mass spectrometry analysis. The consistently seen bands during IMAC purification of C49S MmyJ was excised (**A**) and subjected to mass spectrometry analysis (**B**). Bands are seen usually during purification at ~17 kDa (sample 1 and 3) and at ~80 kDa (sample 2) (**A**). The peptides detected in each sample is shown in **B** by the percentage of total spectra. This is the percentage of peptides assigned to these proteins in the total spectrum seen. In the 17 kDa samples, relatively higher percentages of MmyJ peptides are seen indicating this sample likely contains mostly MmyJ. In the 80 kDa sample however no MmyJ peptides are detected, and this is likely a contaminant protein.

time.

First a run of standards was done at a flow rate of 1 mL/min (Chapter 2: Materials & Methods) in the buffer the protein samples would also be run in (Fig. 20A). A mix of known weight proteins and Blue Dextran was used. Blue Dextran is a high molecular weight (2,000 kDa+) polymer that as the name suggests has a blue appearance. The volume prior to Blue Dextran eluting is the void volume of the column, i.e. the volume that is eluted first from the column after injection which cannot contain any proteins regardless of the size of the protein.[223] For the SEC column used in these experiments the void volume was calculated to be about 70 mL. Due to the low molecular weight standards used, the difference between Blue Dextran elution and detection of the first molecular weight standard (at 44 kDa) was small (Fig. 18A). Therefore, any proteins detected between the void volume and this 44 kDa protein would need to be characterised by SDS-PAGE to understand its size which could be anything greater than 44 kDa.

One litre cultures of gravity purified IMAC elution of both WT MmyJ (Fig. 20B) and C49S MmyJ (Fig. 20C) were injected, and peaks of proteins detected at 280 nm. Major peaks were characterised through SDS-PAGE (Appendix Fig. A4).

In the WT MmyJ run the first peak comes with a small shoulder to it and is seen at around 15 minutes (Fig. 20B). This same peak but at a lesser intensity is also seen in the C49S MmyJ SEC purification (Fig. 20C). In comparison to the standard run (Fig. 14A), this retention time indicates a protein of greater size than 44 kDa. SDS-PAGE characterisation (Appendix Fig. A4) indicates this elution contains only a protein of roughly 80 kDa, which was the contaminant protein present during IMAC purification (Fig. 18 and Fig. 19).

The second peak in WT MmyJ (Fig. 20B) and the largest peak in the C49S MmyJ (Fig. 20C) purification also eluted at roughly the same retention time of 44 minutes. In comparison to the standards (Fig. 20A), this places the protein eluted from this fraction as being around or just under 44 kDa. When these eluted samples were run

on SDS-PAGE gels (Appendix Fig. A4) strong bands were seen at around 17 kDa (MmyJ as established from Fig. 19) and faint bands at double this size. Given that a 17 kDa protein during this SEC procedure would elute at a retention time closer to 70 minutes, and that while a single peak was seen on SEC, two bands were identified on the SDS-PAGE of this elution, it is highly likely this is MmyJ forming dimers.

Importantly it must be noted that C49S MmyJ cannot form disulphide bonds yet under native conditions as that in SEC purification it is seen exclusively as dimers (Fig. 20C), indicating MmyJ's affinity to form homodimers.

C49S and WT MmyJ elution from SEC purification were concentrated and used at various concentrations (5-11 mg/mL) to set up X-ray crystallography screens (Chapter 2: Materials & Methods). After 1 year of incubation, crystals of C49S MmyJ at 7 mg/mL formed and were harvested for X-ray diffraction which was used to solve its structure (Chapter 4).

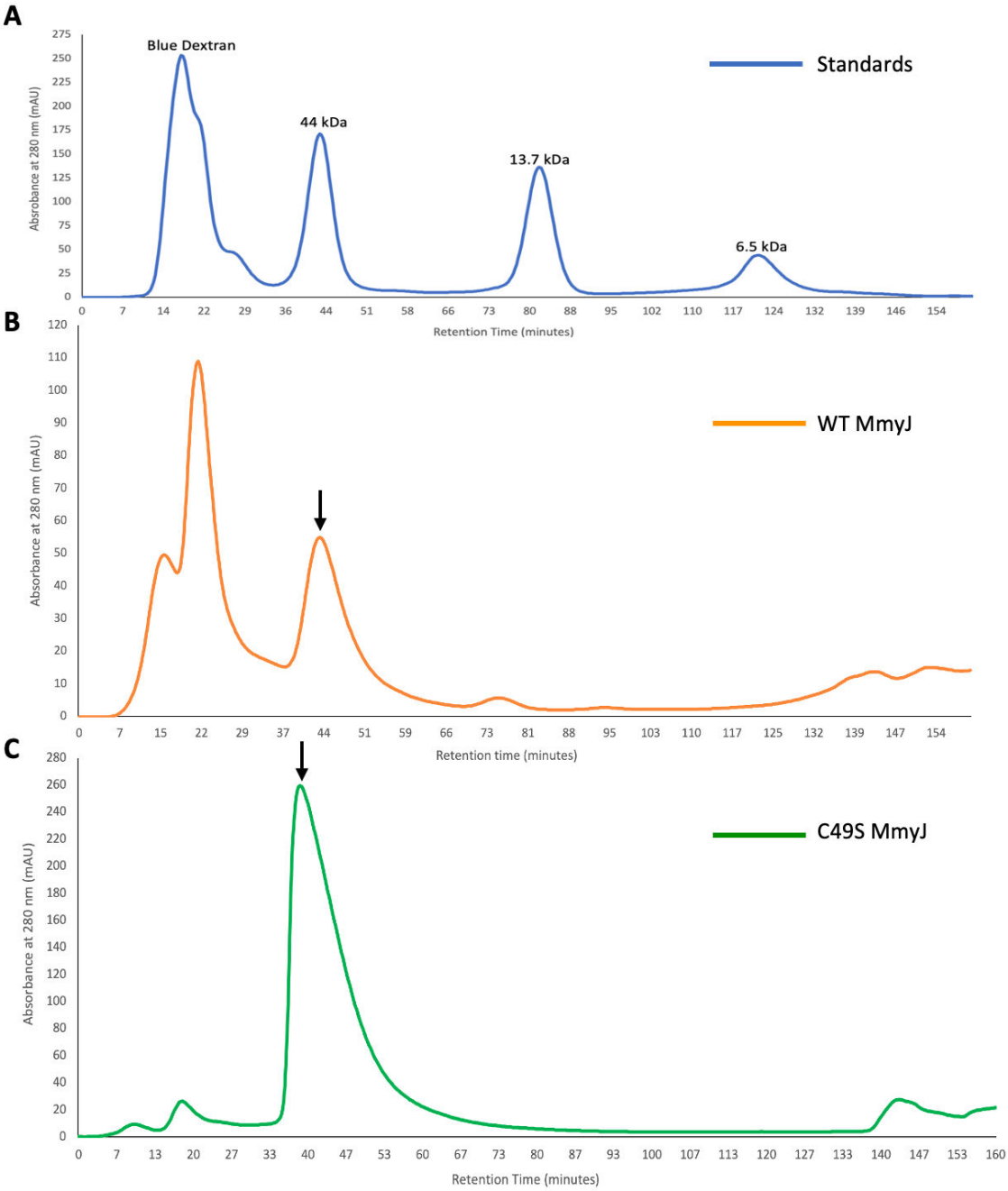


Figure 20. Size exclusion chromatography (SEC) of recombinant MmyJ-WT (B) and recombinant C49S MmyJ (C) against standards (A). IMAC purified recombinant MmyJ WT and recombinant C49S MmyJ (Fig. 18) was run through a SEC column to separate MmyJ from contaminant proteins (Fig. 13). Low molecular weight standards run on the same column with the same method (Chapter 2: Materials & Methods) is shown above (**A**) to estimate protein elution sizes. Fractions of the major peaks were also taken and run on SDS-PAGE gels to estimate protein sizes (Appendix Fig. A4). The first peak in **B** was shown to be an 80 kDa protein. The second peak in **B** and the main peak in **C** when compared to standards (**A**) are about 44 kDa but run on SDS-PAGE (Appendix Fig. A4) show a 17 kDa protein. These peaks (black arrows) are therefore likely to be (WT/C49S) MmyJ forming homodimers.

3.3 Evidence of MmyJ as a transcriptional protein from DNA binding

3.3.1 MmyJ binds within an 80 base pair region of intergenic DNA

Transcriptional proteins bind to regions of DNA to control expression levels of genes. These genes are usually ones with promoter sequences in the regions where the transcriptional proteins bind. This can prevent RNA Polymerase binding and thus switches off the expression of the target gene or transcriptional proteins binding to DNA can act as a signal for RNA Polymerase binding. Bacterial DNA has bi-directional transcription meaning two different genes on opposite ends of the DNA strands can be transcribed in close proximity to one another. Therefore, promoter sequences regions may be shared between two bi-directionally transcribed genes or spaced close together for both these genes in this intergenic region. Transcription factors can also then bind here and act to control transcription of the divergent genes.

As noted previously, *mmyJ* and the efflux pump *mmr* are divergent genes. There is a 218 bp intergenic region between the two genes. This means the -35 and -10 promoter sequence regions of both genes are encompassed within this stretch of DNA. The intergenic region and promoter sites for both *mmr* and *mmyJ* was shown by Neal & Chater (1991)[144] using high-resolution S1 nuclease mapping both *in vivo* and *in vitro* to quantify mRNA transcripts. The two major conclusions apart from the intergenic DNA sequence and promoter regions described by Neal & Chater (1991)[144] were that *mmr* and *mmyJ* both have *mmyJ* and *mmr* promoter regions are in close proximity to one another and that *mmyJ* and *mmr* transcripts could be initiated by RNA polymerase purified from *S. coelicolor* A3(2) without any specific positive regulators.

Also, seen within this region is a 13-1-13 palindromic repeat sequence and other palindromic sequences which are likely places for dimeric proteins to bind.

Previously it has been shown through DNA fingerprinting [155] that MmyJ may bind to an 80 bp fragment of DNA. The exact stoichiometry and region of binding however has not been eluded to as of yet.

To study this, the 80 bp fragment was first amplified by PCR (Chapter 2: Materials & Methods) (Fig. 21). To this DNA, various concentrations of previously purified (Fig. 20) and freshly thawed recombinant C49S MmyJ was added and run on a 10% Native PAGE gel (Chapter 2: Materials & Methods).

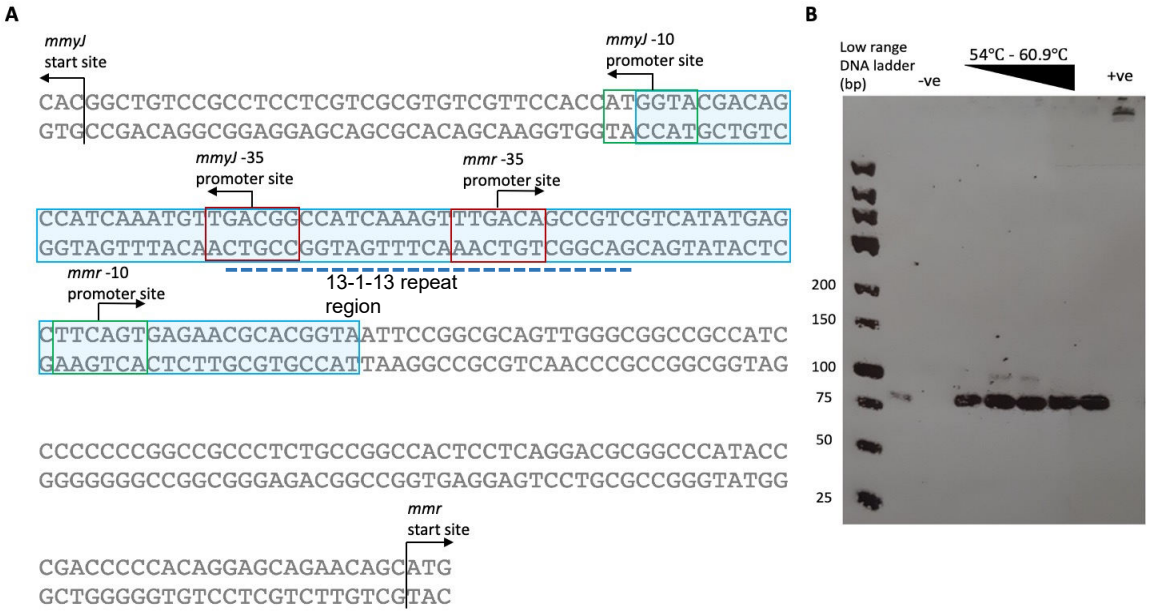
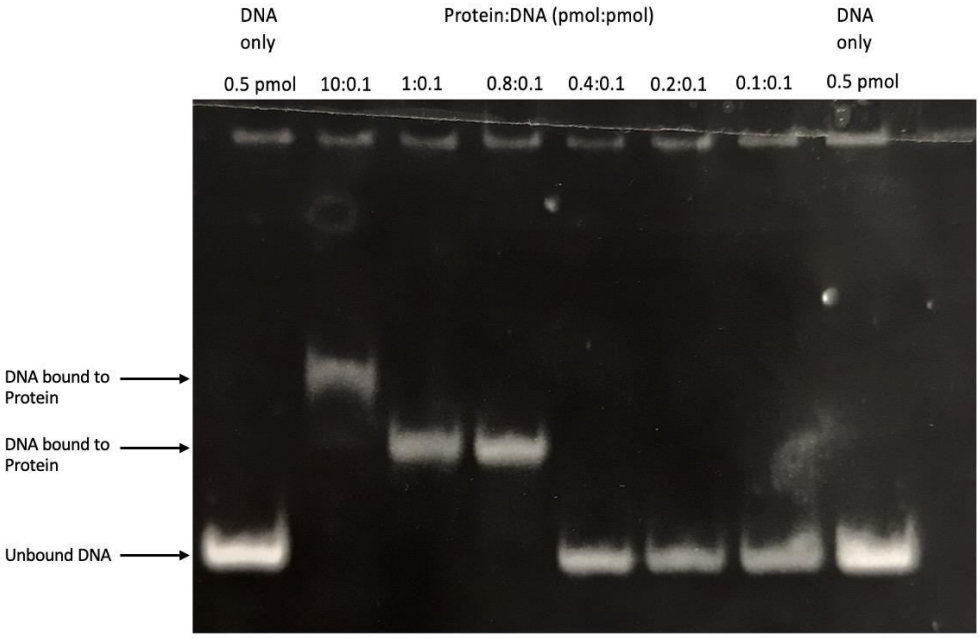


Figure 21. The intergenic DNA region between *mmr* and *mmyJ* with an 80 base pair (bp) section of it encompassing the MmyJ binding sites. A) The 218 bp intergenic region between *mmr* and *mmyJ* is shown with the 80 bp region where MmyJ is thought to bind being highlighted. Blue arrows within this region indicate the -10 and -35 promoter regions of *mmyJ* while orange arrows indicate the -10 and -35 promoter sections of *mmr*. Promoter regions determined by Neal & Chater (1991) using mRNA analysis.[144] Also shown is a 13-1-13 palindromic repeat region as a possible MmyJ binding site. **B)** This 80 bp region was amplified through gradient PCR of the *Streptomyces* C73-787 cosmid with the positive control being the cosmid itself and the negative control being water (Chapter 2: Materials & Methods)**.** Run on a 4% agarose gel, the 80 bp amplified section is clearly shown.

Fig. 22 shows the results of this experiment which when repeated gave consistent results (Appendix Fig. A5). The band shift in Fig. 22 comparing the DNA only lane and the DNA with C49S MmyJ shows a clear shift up. This indicates that C49S MmyJ is binding to the DNA thus causing it to grow heavier in weight and consequently run slower during electrophoresis. In previous studies[155] only singular band shifts were seen between MmyJ and DNA however from Fig. 22 there is clearly a second shift up when the concentration of C49S MmyJ is increased. Multiple possibilities exist to explain such results.



For example, the first shift may be one monomer of C49S MmyJ binding with the second shift being that of a secondary monomer binding to form a bound dimer. Alternatively, as C49S MmyJ is seen under native conditions (Fig. 20) to be a dimer it may be the result of one dimer bound to one region along the 80 bp DNA with a second dimer binding at a secondary location. This location may be opposite the first region or at another region either upstream or downstream to the first. Two dimers binding opposite one another was seen by another transcriptional regulator in the methylenomycin gene cluster- MmfR.[153] While MmfR does not belong to the

Figure 22. Electromobility shift assay (EMSA) of C49S MmyJ binding to DNA. Purified recombinant C49S MmyJ (Fig. 20) was incubated with an 80 base pair (bp) region of DNA (Fig. 17) at various picomolar concentrations (pmol) before being run on a 10% NATIVE PAGE gel and stained to visualise the DNA. A band above the control samples of just the DNA indicates protein binding and therefore increasing the weight and decreasing the speed of migration through the gel. At concentrations 4 times greater of protein to DNA there appears to be protein binding to the DNA as seen by the band shift. A second band shift (super shift) at concentrations of protein 100 times greater than DNA indicates further binding of protein to DNA.

ArsR family, some ArsR-family proteins have been noted to have a secondary homodimer binding site along the DNA binding region of the first site. Even as a homodimer C49S MmyJ is still relatively small and so spatially this is still a possibility along the short intergenic region. From Fig. 22 alone it is impossible to state which scenario is occurring and further studies would need to be conducted.

3.3.2 Fragmenting intergenic DNA to find exact MmyJ binding sites

C49S MmyJ is a small protein and even in a dimeric form it is unlikely to be able to encompass the entire 80 bp region. Therefore, it stands to reason the binding region is far narrower. As EMSAs have shown reliable data, the 80 bp region was

fragmented to determine the shortest length of DNA that can be bound to MmyJ (Fig. 23A).

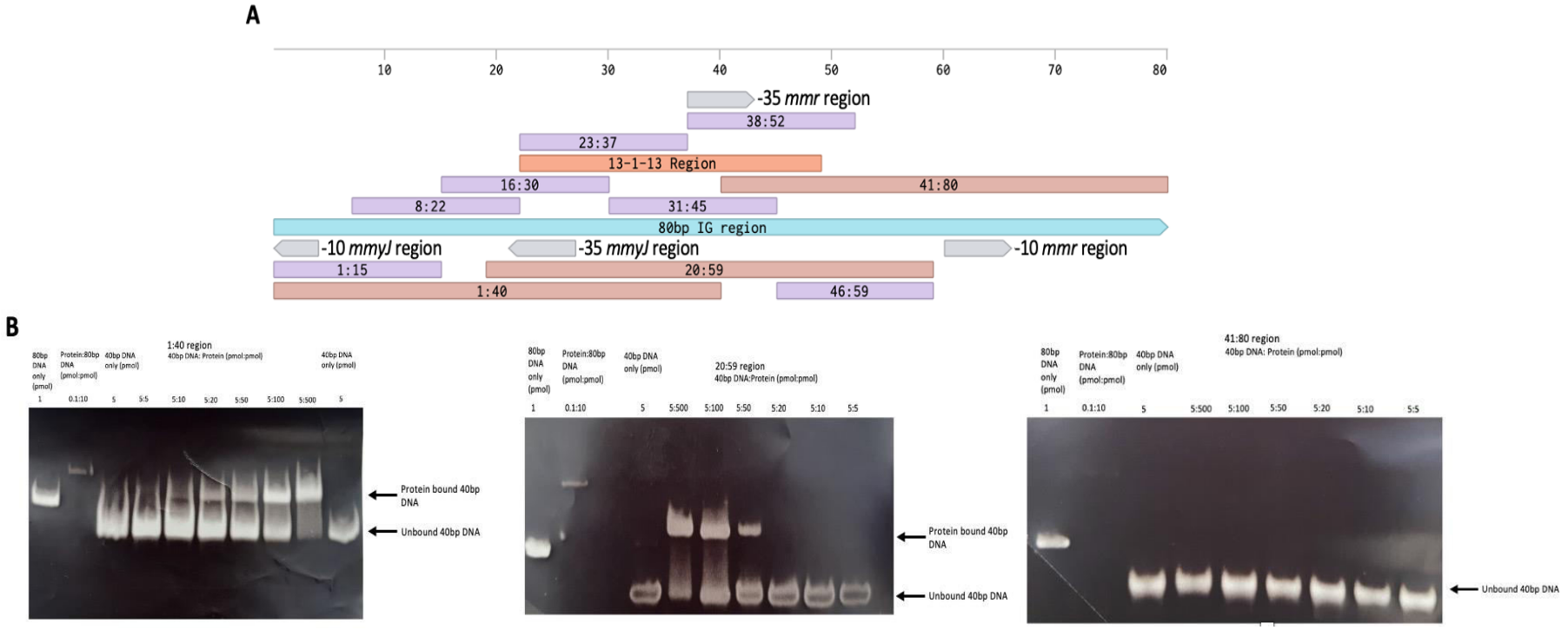


Figure 23. The 80 base pair (bp) section of DNA was further fragmented and tested for C49S MmyJ binding. A) The 80 bp of intergenic DNA shown to have C49S MmyJ bind to it (Fig. 18) was further fragmented to find the exact binding region. It was split into 3 x 40 bp sections (brown), 7 x 15 bp regions (purple) and one palindromic repeat region (13-1-13 region in orange). In grey are the promoter regions encompassed by the 80 bp section. **B**) Overlapping 40 bp regions (**A**-brown) were first tested for protein binding. In 2 of the regions (1-40 and 20-59) binding was seen through a band shift whereas no band shift or binding was seen by the 41-80 bp region. This indicates MmyJ likely binds in in that 1-59 bp section.

First, the region was split into 3 overlapping fragments through purchase of single stranded oligonucleotides which were then annealed together to create double stranded DNA (Chapter 2: Materials & Methods). Similar EMSAs to Fig. 22 were run with these DNA fragments (Chapter 2: Materials & Methods). Although results are not as clear and a higher concentration of C49S MmyJ is required there is definite binding of C49S MmyJ seen to the 1-59 bp section while no binding is seen in the 41-80 fragment (Fig. 23B). This narrows the binding region of C49S MmyJ to somewhere between the first 1-59 bp within the 80 bp area. The regions that did show DNA binding, unlike when the entire 80 bp fragment was used did not show a 'double-shift'. A singular shift was seen indicating perhaps that there is a binding site between 1-40 bp of the 80 bp fragment and a second one between 20-59 bp. It could also be a singular site that is in the range of 20-40 bp of the 80 bp fragment.

To further test if there are two different binding sites, these EMSAs were repeated by fragmenting the 1-59 bp area into overlapping 15 bp lengths with also the 13-1-13 palindromic repeat region tested (Fig. 24). Fragments of 15 bp in length were chosen based on the fragment sizes seen to be bound by other ArsR proteins. Even as homodimers these are usually in the range of 10-20 bp DNA. Some ArsR proteins are implied to bind to a 12-2-12 repeat palindromic section of DNA [201,202] and so having this 13-1-13 palindromic region was thought as a likely place for C49S MmyJ binding. However, no binding of C49S MmyJ was observed with these shorter fragments and no binding was seen with the palindromic repeat region (Fig. 24 and Appendix Fig. A6).

It may be that there is binding but it is too weak to be detected on the EMSA. In that case, a more sensitive technique could be used such as labelling the DNA fragments with a probe and using a Western Blot. No C49S MmyJ binding seen by the 15 bp fragments may be because these sections may not overlap enough and the binding site may fall in a region between 2 fragments, meaning when the C49S MmyJ homodimers try to associate they are unable to stabilise enough for detection.

As for the palindromic section of DNA, part of it falls within the 1-59 bp region already shown to have some binding property while some of it falls over this section where there was no binding seen. This may therefore imply that this region is not the exact binding site for C49S MmyJ and again it may be that C49S MmyJ can bind to the start of the palindromic region but rapidly dissociates because of instability from the rest of the DNA region.

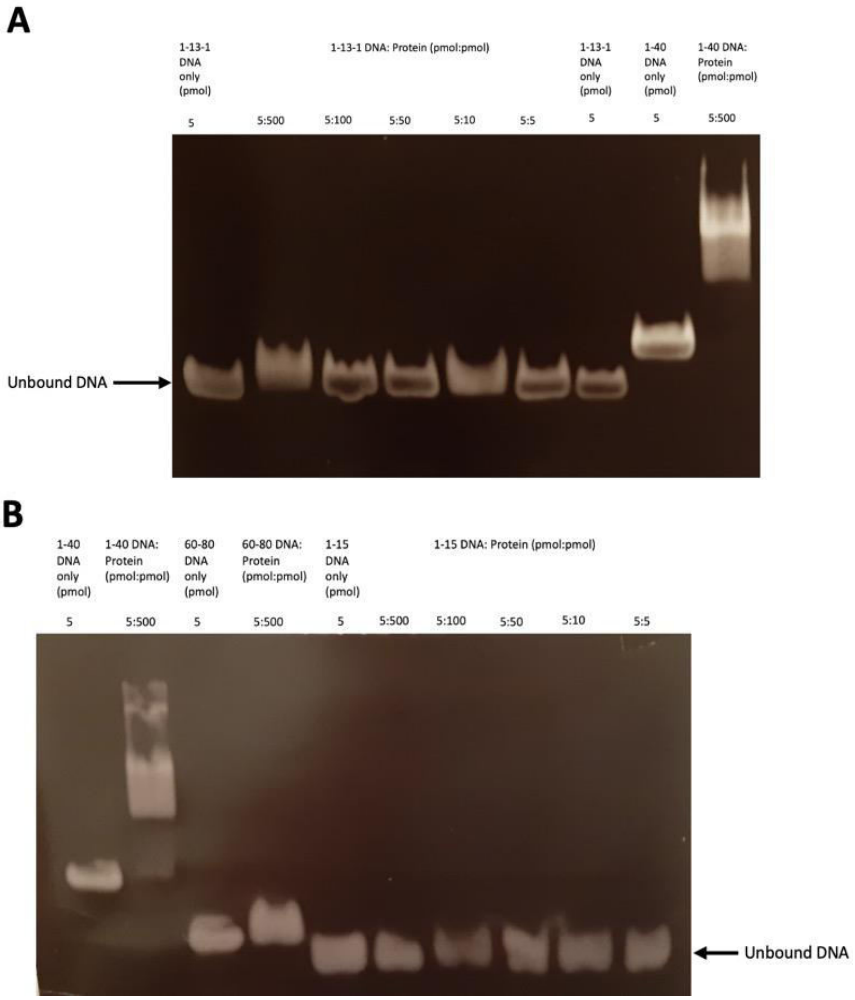


Figure 24. Electromobility shift assay (EMSA) of palindromic DNA and short regions of DNA C49S MmyJ may bind to. The 80 base pair (bp) region of intergenic DNA shown to have C49S MmyJ bind to it (Fig. 22) had been further fragmented (Fig. 23). Short 15 bp fragments of that 80 bp section and a palindromic 13-1-13 sequence in that region were tested for binding to C49S MmyJ. **A**) The palindromic region shows no protein-DNA binding. This is likely not the site at which C49S MmyJ binds to. **B**) One example of a 15 bp fragment EMSA with C49S MmyJ. This gel was representative of the other 15 bp fragments also tested (Appendix Fig. A6) where no binding of the protein to the fragment even at high concentrations of protein was seen. These may be too short section to visualise MmyJ binding through EMSAs.

3.3.3 The structure of MmyJ was used to find two DNA binding sites of MmyJ

Once the structure of MmyJ was solved (Chapter 4), the exact DNA binding site within the 80 bp intergenic region of *mmr* and *mmyJ* was able to be predicted (Chapter 4.4.3). This gave 2 pseudo-palindromic sites (CCATCAAATGTTGACGG and CCATCAAAGTTTGACAG) beside each other that MmyJ dimers may be able to bind to (Fig. 25A). One of these sites encompass the -35 promoter region for *mmr* while the other contains the -35 promoter region for *mmyJ*. Assays using C49S MmyJ akin to the ones previously done were conducted on this 34 bp region (Fig. 25B). To ensure that it is not the length of DNA or presence of DNA but the specific sequence itself that causes protein binding, DNA of the same length as being tested with the same ratios of bases but in a 'shuffled' order was purchased and used as a negative control (labelled 'Shuffle DNA').

Not only was binding of protein to DNA observed (Fig. 21B) but a double-shift was also seen similar to that when the entire 80 bp region was used (Fig. 22). The same interpretation was assumed here in that MmyJ is binding at two sites (each pseudo-palindromic region) rather than just one. From the modelling (Chapter 4.4.3) and evidence of MmyJ acting as a dimer, it is evident that one set of dimer binds at one 'palindromic' region and another to the DNA region beside it.

To test this experimentally, the same DNA-protein assays were conducted but with only the 17 bp regions this time (Fig. 26). These showed DNA binding but no double-shifts of additional binding even with higher concentrations of C49S MmyJ to that used in the entire 34 bp region of both palindromes (Fig. 25B). In such case, it is evident that one dimer binds to each of these regions and that these dimers are formed prior to DNA binding meaning MmyJ acts as a transcriptional regulator of *mmr* and is self-regulating its own expression. Furthermore, protein to DNA binding to the -35 *mmyJ* site is observed with 50 times more protein to DNA (Fig. 26A) but double this (100X) is required for the protein to bind to the -35 *mmr* site (Fig. 26B). The expression of the Mmr efflux pump is therefore likely under stricter control by MmyJ than its own expression.

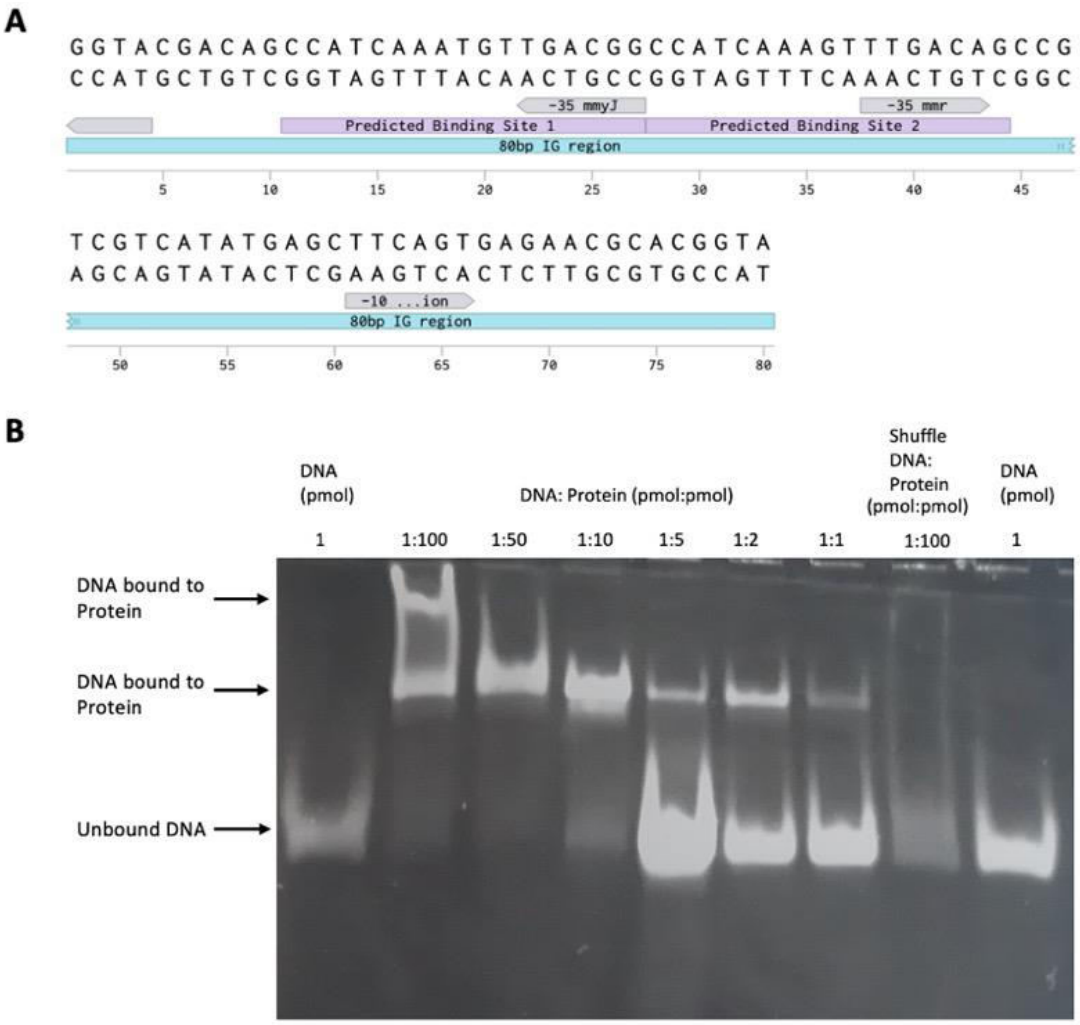


Figure 25. MmyJ binds at regions between the -35 promoter sites of *mmr* and *mmyJ*. **A)** The 80 base pair (bp) intergenic region between *mmr* and *mmyJ* was previously shown to have binding region(s) for MmyJ (Fig. 22). By solving the structure of MmyJ (Fig. 49) two regions were predicted to be the binding site of homodimeric MmyJ (Fig. 65) with one encompassing the *mmr* promoter site and the other the *mmyJ* promoter site. These are indicated in purple. **B**) Electromobility shift assays (EMSA) of this entire predicted fragment (1 and 2) were seen to be bound by MmyJ. Binding first occurs with even a 1:1 ratio concentration of protein to DNA but further binding of MmyJ to the DNA is seen with 100 times more protein to DNA. The 'Shuffle' DNA fragment was DNA with same composition of bases as DNA being tested but in a random order (order was computer generated) (Chapter 2: Materials & Methods) and therefore used as a negative control.

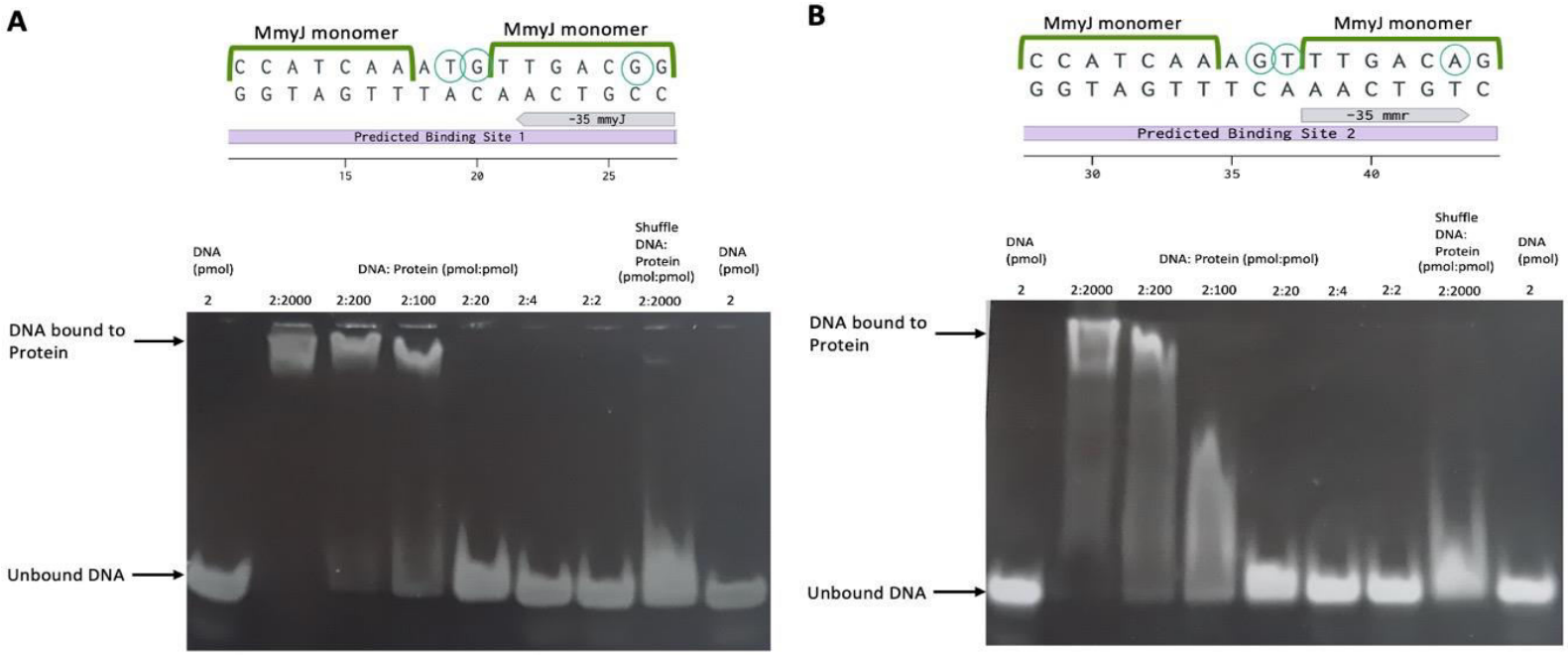


Figure 26. MmyJ binds at separate regions encompassing either the -35 promoter sites of *mmr* or *mmyJ*. A) One predicted binding site of homodimeric MmyJ which encompasses the -35 promoter region of *mmyJ*. **B)** The other predicted binding site of homodimeric MmyJ that encompasses the -35 promoter region of *mmr,* the bi-directional to *mmyJ* MFS encoding efflux pump. A fragment containing both these regions were shown to have MmyJ bind to them (Fig. 25B). Modelling from the structure of MmyJ (Fig. 65) suggested each monomer able to encompass certain bases (above) as highlighted. The two predicted regions are pseudo-palindromic and are similar itself except for the 3 bases highlight, 2 of which fall into an area not predicted to be recognized by MmyJ. Electromobility shift assays (below) indicate MmyJ can bind to each of these sites. Despite increased concentrations of MmyJ only one shift and therefore one set of binding is seen, indicative of one homodimer of MmyJ binding to each site. The site encompassing the promoter region of *mmyJ* shows MmyJ binding at 50 times protein to DNA concentration (**A**) whereas 100 times more protein to DNA is needed for the *mmr* site (**B**). The 'Shuffle' DNA fragment was DNA with same composition of bases as DNA being tested but in a random order (order was computer generated) (Chapter 2: Materials & Methods) and therefore used as a negative control.

3.3.4 Conserved intergenic regions where MmyJ-like genes may bind in other actinobacteria

Earlier (Chapter 3.1.4) looking at the synteny of *mmyJ* and *mmr* genes in other actinobacteria, a few were identified as being potential MmyJ homologs (Fig. 16). To further determine if these are MmyJ homologs and may function in a similar manner to MmyJ, the intergenic region between the *mmyJ*-like gene and *mmr*-like gene was analysed. MmyJ binds to 2 intergenic sites between *mmyJ* and *mmr* and so if these MmyJ-like proteins also bind to DNA it may do so at conserved intergenic sites also.

The intergenic regions (ranging from ~60 bp to ~300 bp) between the *mmyJ*-like gene and the MFS transporter identified from Fig. 16 (Chapter 3.1.4) are summarised in Table 11. These intergenic regions were compiled and analysed by MEME analysis to identify any consensus motifs occurring between these regions (Fig. 27), with particular emphasis on identifying any regions containing the same binding sites as shown for MmyJ in *S. coelicolor* A3(2) (Fig. 26).

The first statistically significant motif found (Fig. 27A) was present in all but one (*Streptomyces* sp. E5N91) of these intergenic regions. This motif towards the end in *S. coelicolor* A3(2) contains part of the binding site encompassing the *mmr* -35 promoter region that MmyJ binds to (Fig. 26B). This partial site is exactly the length of one monomer of MmyJ that would bind, highlighting that while there is a strong consensus for part of these MFS transporter gene regulation in the identified bacteria, they are still subject to variation and therefore stricter, more tailored control. It would be interesting to determine if these sites can be regulated by the MmyJ-equivalent proteins and whether these function as monomers, homodimers, heterodimers or as other multimers.

Table 11: Organisms identified to have a protein similar to MmyJ with a gene beside it encoding for MFS transporters (from Fig. 16) and the intergenic length of DNA between them. The MmyJ-like gene and MFS transporter gene were sought out by their accession numbers in NCBI and the intergenic region of DNA between them taken for analysis (Fig. 27). The length of this region was noted. As shown, these are mostly in the 150-250 base pair (bp) range. Highlighted in green and given are also the details for *mmyJ* itself and the MFS transporter *mmr*.

	NCBI protein accession number		
Bacteria	MmyJ-like gene	MFS Transporter gene	Length of intergenic sequence (base pairs (bp))
Streptomyces sp 846 5	WP_133995155.1	WP_133995153.1	295
Streptomyces sp E5N91	WP_164272960.1	WP_121716552.1	68
Streptomyces sp NRRL S 31	WP_030742020.1	WP_030742022.1	220
Streptomyces sp AC512 CC834	WP_217165407.1	WP_217165377.1	374
Kitasatospora mediocidica KCTC 9733	WP_198045424.1	WP_051965637.1	353
Streptomyces caeruleatus	WP_062723279.1	WP_062723280.1	264
Streptomyces sp 351MFTsu5 1	WP_020135630.1	WP_020135631.1	245
Streptomyces sp AcE210	WP_116508238.1	WP_116508236.1	258
Kitasatospora aureofaciens	WP_030286058.1	WP_030556860.1	176
Kitasatospora xanthocidica	WP_189921066.1	WP_189921068.1	265
Amycolatopsis roodepoortensis	WP_192746507.1	WP_192746508.1	158
Amycolatopsis sp WAC 04182	WP_125680327.1	WP_199745678.1	113
Rhodococcus sp P1Y	WP_121115275.1	WP_121115273.1	161
Amycolatopsis australiensis	WP_084743248.1	WP_072478164.1	183
Amycolatopsis deserti	WP_191248248.1	WP_229874807.1	53
Amycolatopsis thermoflava	WP_123685663.1	WP_123685664.1	59
Amycolatopsis sp ATCC 39116	WP_020419963.1	WP_051123585.1	66
Gordonia polyisoprenivorans	WP_208793635.1	WP_208795937.1	242
Streptomyces griseorubiginosus	WP_208778111.1	WP_208778110.1	89
Streptomyces niger	WP_052868623.1	WP_052868622.1	192
Streptomyces sp MBT62	WP_200712037.1	WP_200712051.1	295
Streptomyces sp 8L	WP_225012589.1	WP_225012588.1	201
Streptomyces bauhiniae	WP_135785704.1	WP_135785703.1	150
Streptomyces sp SA15	WP_095750843.1	WP_095750844.1	150
Streptomyces sp Ag109 G2 15	WP_097254997.1	WP_097254998.1	136
Streptomyces coelicolor A3(2)	NP_639848.1	NP_639847.1	218

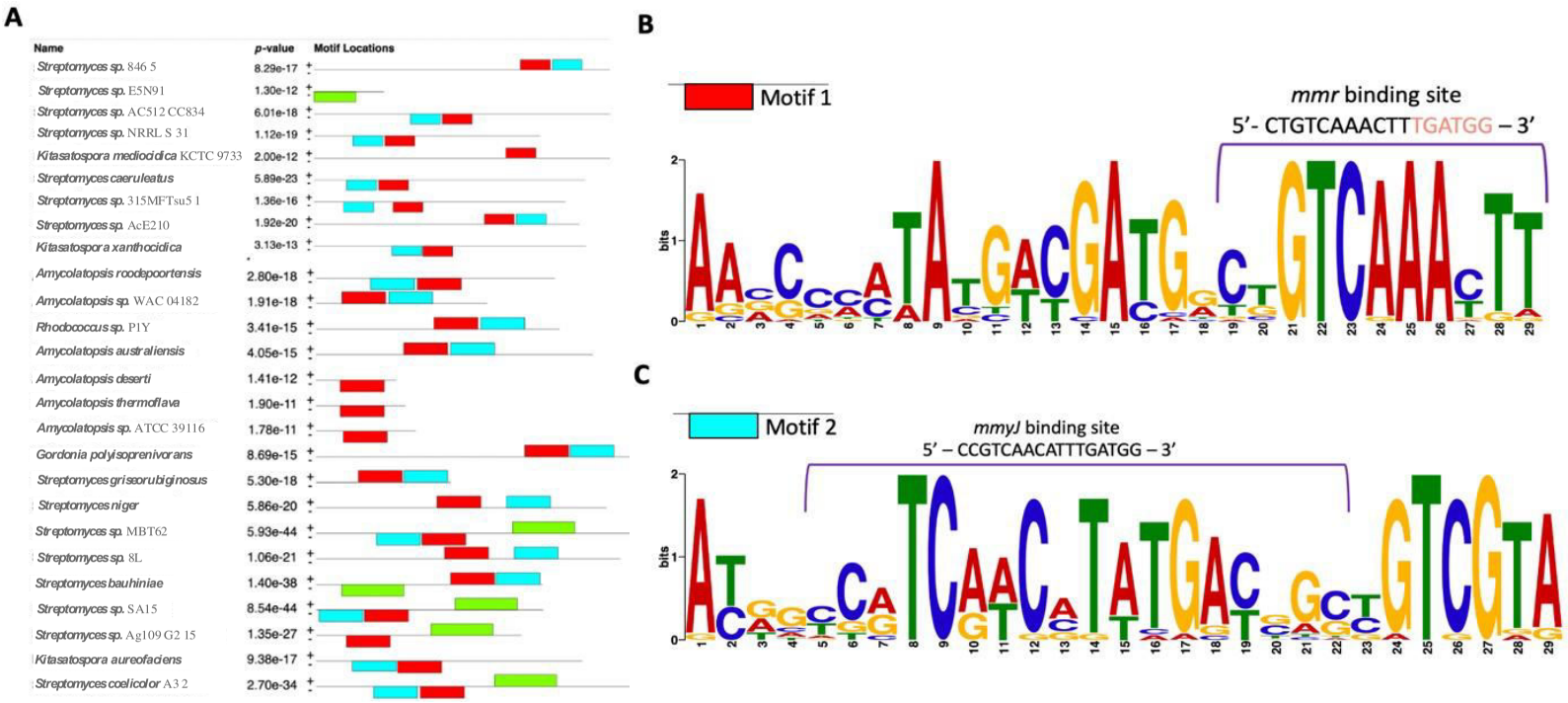


Figure 27. Intergenic DNA regions of organisms between their MmyJ-like genes and MFS transporter have consensus motifs. The intergenic regions of organisms identified as having a MmyJ-like gene bi-directional to an MFS transporter gene (Table 11) were input into MEME analysis for statistically significant motifs. **A**) Motifs identified in the intergenic DNA of each organism. All but one (*Streptomyces sp.* E5N91) had Motif 1 (red) in the intergenic section and all intergenic sections of over 80 base pairs (bp) contained Motif 2 (cyan) beside Motif 1. **B)** Motif 1 in *S. coelicolor* A3(2) contains a site corresponding to MmyJ binding to it which encompasses the -35 promoter site of *mmr* (Fig. 26B). Only one monomer of MmyJ would be able to bind to this site though. In peach are the missing bases not in the motif MmyJ was experimentally found to bind to. **C)** Motif in *S. coelicolor* A3(2) contains a site corresponding to MmyJ binding to it which encompasses the -35 promoter site of *mmyJ* (Fig. 26A). The entire binding region of this MmyJ site is seen in this motif meaning a homodimeric MmyJ could bind. Presence of both sites evidence potential MmyJ-Mmr systems in other bacteria.

The second statistically significant motif (Fig. 27B) present in a large majority of the bacteria identified (Table 11) are beside motif 1 as expected. This is because this site in *S. coelicolor* A3(2) SCP1 was seen to be where MmyJ bound to its own -35 promoter region (Fig. 26A) and while the *mmr* and *mmyJ* promoter sites are near each other, they are thought to be separately regulated by MmyJ (Fig. 26). Only this *mmyJ* promoter site of MmyJ binding was seen in this motif (Fig. 27B) and that too in its entirety. While this implies that these MmyJ-like proteins are self-regulating (as they too are described as ArsR-like proteins), it cannot be ruled out that these may be recognition sites for other regulatory proteins, perhaps even others in gene clusters they may be a part of.

While it could be argued that these are simply the -35 promoter regions that are identical, the -35 promoter regions for *mmyJ* and *mmr* are only 6 bp long and the motifs identified as protein binding sites are longer than this, indicating these motifs may be actual DNA regions recognised by transcriptional proteins. Examining experimentally whether these MmyJ-like proteins can bind transcriptionally to any of these motifs would help to validate if the identified motifs are sites of regulation thereby authenticating the presence of these MmyJ-Mmr self-resistance systems in other bacteria.

3.4 Using size exclusion chromatography to gather details into C49S MmyJ-DNA

The binding of C49S MmyJ to the 80 bp region of DNA that has been observed on native PAGE gels has 2 shift patterns discussed previously (Chapter 3.3). C49S MmyJ at high concentrations was shown to bind further to this region of DNA compared to lower concentrations indicative of a 2 binding sites or binding of one monomer then a second. As this ‘double-shift’ has not been noted before, the stoichiometry of this binding is of great interest. To do this SDS-PAGE or native PAGE gels would not be wholly suitable as these measurements would need to be taken over time.

One such method would be to use a size exclusion chromatography which was already implemented in purifying C49S MmyJ however instead of using its separating properties to distinguish between different sized proteins, in this case it was used to separate between different binding complexes of C49S MmyJ and DNA that would be of different sizes.

That is, a baseline was first made of pure MmyJ to base any other shifts that may occur on. Then different concentrations of C49S MmyJ to DNA already mixed prior can be added to note over time if they remain as a complex or if there is unbinding occur. C49S MmyJ-DNA bound complexes would be expected to be of a larger size and so should have a shorter retention time to that of C49S MmyJ alone. As measurements are taken at 280 nm, the range in which most proteins absorb at due to amino acids with aromatic rings (such as tryptophan and tyrosine) absorbing at this wavelength, only protein and no DNA would be measured here. C49S MmyJ has no tryptophan residues but does contain three tyrosine residues so can be measured at 280 nm and has an extinction coefficient at 280 nm of 4470 $M^{-1}cm^{-1}$. Therefore, peaks seen would be either C49S MmyJ-DNA complexes or C49S MmyJ alone.

This experiment was first tested on a preparative grade size exclusion column (Chapter 2: Materials & Methods) as used for in purifying C49S MmyJ however due to the large size of the column and therefore use of a greater volume of solvent, the samples were highly diluted meaning readings were deemed inaccurate (Appendix Fig. A7). A smaller analytical column of the same type was then used (Chapter 2: Materials & Methods) to gain more accurate readings on which analysis could be made (Fig. 28).

As shown in Fig. 28 with just C49S MmyJ loaded as the sample a clear, defined peak is observed at 5.8 minutes retention time. In this protein only sample with 65 pmol of C49S MmyJ protein, similar to when C49S MmyJ was being purified through size exclusion chromatography (Fig. 28), there is one clear distinct peak seen with a small shoulder prior to it. This peak has been observed with relation to the

standards as a dimer of C49S MmyJ (Fig. 18B) before. Further along (~ 10 minutes retention time) the denatured or monomeric form of C49S MmyJ is observed (also previously present in Fig. 20C).

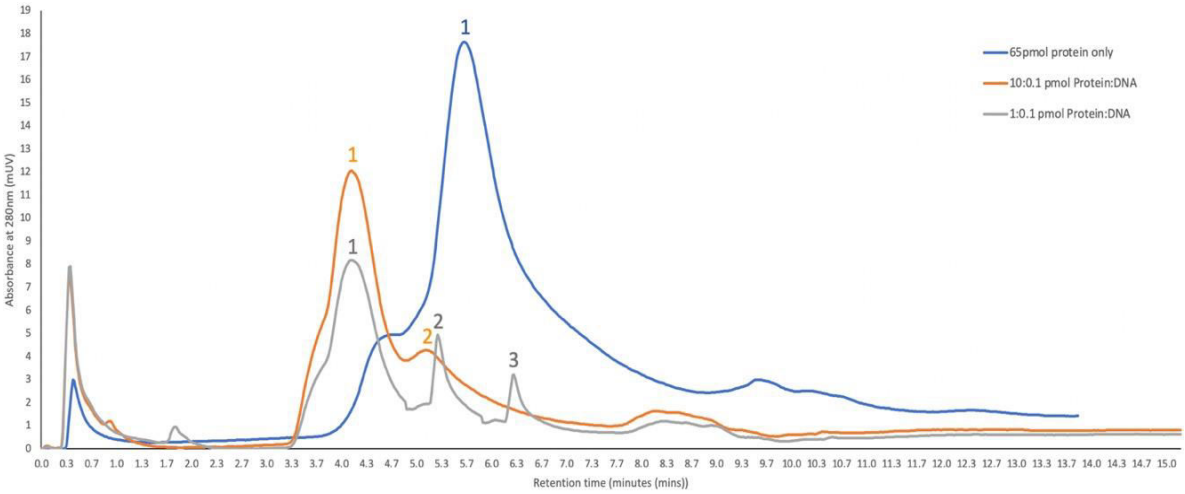


Figure 28. C49S MmyJ-DNA stoichiometry was tested using analytical size exclusion chromatography. Size exclusion chromatography (SEC) was used to see across time DNA-C49S MmyJ binding properties. An 80 base pair (bp) region of DNA was already shown to bind to C49S MmyJ (Fig. 22) but SEC was used to further investigate these properties. All samples were incubated 30 minutes prior to loading. A protein only sample (blue) at 65 picomolar (pmol) final concentration was first run showing only one peak (blue 1) as seen before during purification (Fig. 20C). A concentration of 100 times protein to DNA was then run (orange) and two peaks were seen (orange 1 and 2). This was the concentration of protein to DNA that prior showed a 'super-shift' (Fig. 22). Finally, a concentration of 10 times more protein to DNA was run (grey) with three peaks being seen (grey 1, 2 and 3). This had previously shown a singular shift (Fig. 22).

Samples of C49S MmyJ incubated with the 80 bp fragment of intergenic DNA at 100 times (100X) and 10 times (10X) higher protein concentration to the DNA (10 pmol: 0.1 pmol and 1 pmol: 0.1 pmol) were also run on this column following the same protocol as the protein only sample (Chapter 2: Materials & Methods) (Fig. 28). The distinct peaks identified in all 3 samples were numbered sequentially from the lowest retention time in the SEC column (heaviest protein size) to highest retention time (lightest protein size).

In the 100X protein to DNA sample (orange line-Fig. 28) there were 2 distinct peaks observed. The first peak (peak 1) has a small shoulder prior to it and is seen at a higher absorbance reading to the second peak (peak 2) which elutes slightly later, indicating a higher heavier protein complex is eluting at peak 1 to peak 2. In the 10X protein to DNA sample (grey line- Fig. 28), 3 distinct peaks are observed. The first peak observed in the 10X protein: DNA sample is identical to the first peak seen in the 100X protein: DNA sample. The second peak from both of these samples are observed at similar but not exactly identical retention times with the 2nd peak from the 10X protein: DNA complex seen later to the 2nd peak of the 100X protein: DNA complex. The final 3rd distinct peak in the 10X protein: DNA sample is not seen in either the 100X protein: DNA sample or the protein only standard sample.

The double stranded 80 bp DNA being tested (Fig. 21) weighs 49.47 kDa (calculated using the DNA molecular weight from The Sequence Manipulation Suite[224]). Binding of C49S MmyJ to this fragment of DNA would create a protein-DNA complex that would elute earlier than that of a protein only fraction due to the additional weight from the DNA and therefore be observed a lower retention time. Peaks 1 in both the 100X protein: DNA and 10X protein: DNA samples (Fig. 28) may therefore be that of C49S MmyJ bound to DNA. This is not only due to them being observed at the same retention time but C49S MmyJ also having been shown *in vitro* (Fig. 22) to bind to this 80 bp DNA at 10X and 100X concentrations. Protein-DNA complex formation would certainly account for the leftward shift from a shorter retention time due to increased weight observed in comparison to the protein only peak (peak 1 blue line-Fig. 28).

Peaks 2 in both the 100X protein: DNA and 10X protein: DNA samples do not align to peak 1 of the protein only sample (Fig. 28) but they also do not align with one another. Both samples contained the same reagents (but with protein and DNA at different concentrations), therefore this difference in peaks is difficult to classify as being a protein-DNA complex. Furthermore, absorbance readings of these peaks were weak (at less than 5 mUV) and cannot therefore be ruled out as artifacts of

the data. This is also true for the 3rd peak observed only in the 10X protein: DNA sample. With an absorbance reading at less than 4 mUV this too may be an artefact.

Peaks 1 in both the protein: DNA samples (Fig. 28) likely indicate protein-DNA complex formation as discussed earlier however it's unclear whether C49S MmyJ is forming multimers to bind to the DNA and/or if the C49S MmyJ if binding at multiple sites along the 80 bp DNA from this SEC study. Further peaks observed in protein: DNA samples could not confidently be assessed as data points or artifacts. Repeating these analytical SEC experiments by injecting higher volumes (at the same concentrations) of C49S MmyJ and the 80 bp DNA fragment could give higher absorbance value readings and validate peaks observed in Fig. 28. Furthermore, samples at lower protein to DNA concentrations (under 1 pmol protein: 0.1 pmol DNA) may also be useful in showing if C49S MmyJ is forming multimers to bind to DNA. There may be other explanations for the results seen in Fig. 28 however it is unclear to draw such conclusions without further repeats and testing at other various concentrations of DNA to protein.

3.5 Dimerization of MmyJ

MmyJ has already been shown to have a strong propensity to form dimers both in solution (Fig. 20) and in a crystallised form (Chapter 4). This interaction was also validated using bacterial adenylate cyclase two-hybrid assays (BACTH) (Fig. 29A).

This system is based on the catalytic domains (T18 and T25) of adenylate cyclase from *Bordetella pertussis* which are active to produce cyclic AMP (cAMP) only when in close proximity to one another. The cAMP can bind to the catabolite activator protein (CAP) which forms a complex that acts as a positive transcriptional regulator of gene transcription of β-galactosidase (via *lacZ*) in *E. coli*. Functional β-galactosidases can then cleave X-gal if X-gal is present to produce a blue coloured pigment. In this way, proteins can be fused onto T18 and T25 domains and be tested for interaction. Should these proteins interact, when colonies are spread on

LB, X-gal, IPTG (for expression of T18 and T25 plasmids) and selective antibiotics plates, colonies will turn blue.[175]

Therefore, T18 and T25 plasmids were cloned with (C49S) MmyJ at both the N and C-terminal (cloning was done by Dr Jonathan Cook (Crow lab, University of Warwick)), transformed into BACTH cells and assayed (Chapter 2: Materials & Methods). As seen in Fig. 29A, plasmids containing MmyJ turned blue indicating a positive result of MmyJ forming dimers, whereas those without MmyJ remain white. The positive control used was T18 and T25 plasmids containing leucine zippers known to form dimers.

This assay is further confirmation of MmyJ interacting as a dimer and sets a basis and standard to use to test for residues that may be important for dimer-dimer interactions. Several residues were identified at the dimer interface when the structure of MmyJ was solved that may be necessary for dimerization (Chapter 4). Out of these, 4 were chosen to be mutated for BACTH assays. These included an arginine thought to be involved in hydrogen bonding and a glutamine also involved in hydrogen bonding which were mutated to the short hydrophobic residue alanine (R34A and Q22A). Leucines were identified as conserved residues at the dimer interface of ArsR family proteins (Chapter 4) and so a key leucine in MmyJ was mutated to the charged residue glutamate (L25E). This leucine (L25 in MmyJ) is also a conserved residue in other structurally similar proteins to MmyJ (Table 18 (Chapter 4)) and primary sequence alignment in Appendix Fig. A8). Similarly, valine 29 was also identified as contributing to maintaining dimer interactions and so this was mutated to glutamate (V29E).

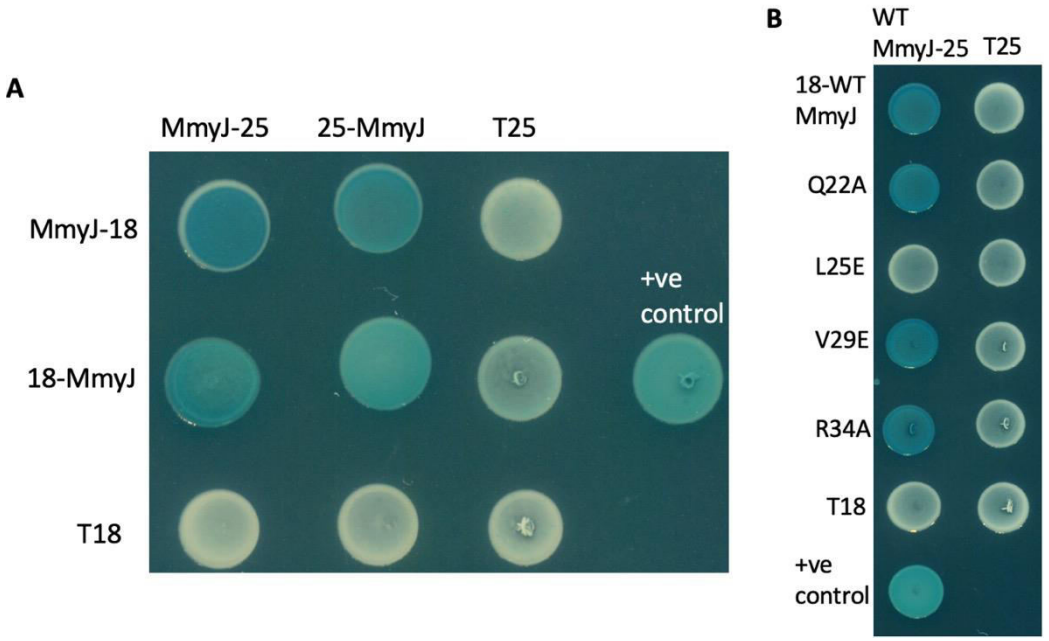


Figure 29. A bacterial adenylate cyclase two-hybrid assay (BACTH assay) shows MmyJ forms dimers (A) and leucine 25 is necessary for dimer formation (B). Catalytic adenylate cyclase domains (T18 and T25) were fused with MmyJ (cloning done by Dr Jonathan Cook, University of Warwick). Should domains interact a blue pigment is produced on plates with X-gal.[173] Should MmyJ monomers interact, the T18 and T25 subunits fused onto it will be brought together and turn colonies blue which as expected were seen, indicating MmyJ dimer formation (**A**). Negative controls with no fused MmyJ were used (3rd row and column in **A** and 6th row and 2nd column in **B**) and show no blue pigmentation. The positive control of T18 and T25 fused with a leucine zipper does appear blue too. **B**) Point mutations of residues identified at the dimer interface were made to T18-MmyJ. A BACTH assay of these were also done and with T18-MmyJ as another positive control as seen in **A**. Only L25E in **B** shows a loss of pigmentation meaning dimers cannot form due to leucine mutation. Assays done in collaboration with Dr Jonathan Cook (University of Warwick), see Chapter 2: Materials & Methods for more details.

These mutations were created separately onto T18-MmyJ plasmids (Chapter 2: Materials & Methods) and a BACTH assay was then performed by Dr Jonathan Cook (Crow lab, University of Warwick) in the same way as earlier (Chapter 2: Materials & Methods) to determine if any of these residues are necessary to maintain dimer interactions (Fig. 29B). The only residue of the 4 mutated that show a significant role at the dimer interface appears to be leucine 25. Mutation of this residue

renders the MmyJ unable to form homodimer interactions (Fig. 29B). L25 is part of the α2 helix in MmyJ (Table 17 (Chapter 4)) and as a non-polar residue is not involved in any hydrogen bonds (Fig. 58). Clearly turning this residue into a charged glutamate is enough to break the hydrophobic core of the dimer interface and prevent dimer interaction.

Dimer interaction is implied to be necessary for transcriptional functioning as evidenced from other ArsR family proteins. To test this in MmyJ, L25E MmyJ mutants could be purified, and ligand/DNA binding assays performed to highlight the necessity of a homodimer formation for functioning.

3.6 Testing MmyJ functionality *in vivo* through use of reporter systems

To corroborate interactions seen *in vitro* and to better understand the function of MmyJ natively within the *Streptomyces* species, *in vivo* work was attempted. One of the simpler measurements of *in vivo* experiments is to look at a colour change or intensity occurring within the cell when testing different conditions. The molecular interactions of MmyJ, can therefore be studied *in vivo* using reporter systems engineered to contain colour producing proteins that are expressed or repressed depending on the binding capabilities of MmyJ. While similar to *in vitro* work, the advantage of this type of *in vivo* work is that it can give quantifiable results which combined can give a more holistic view of the biophysical properties of a protein.

The experimental design of a reporter system testing DNA-MmyJ interactions *in vivo* was to first have a reporter plasmid that produces a coloured protein containing the *Streptomyces* intergenic region of DNA between the *mmyJ* and *mmr* gene (a 'DNA-reporter' construct). Transformed into a bacterial cell this construct would constitutively produce the reporter colour. Addition of a MmyJ plasmid with *mmyJ* expression under the control of an inducible expression system would be

then transformed into bacteria along with the DNA-reporter construct. Without any inducers added the bacteria would be expected to produce the DNA-reporter colour. Addition of an inducer of the *mmyJ*-plasmid would allow expression of MmyJ and should MmyJ interact with the *mmyJ-mmr* intergenic region then the coloured reporter protein from the DNA-reporter system would have its gene's expression suppressed and colony pigmentation would diminish giving evidence of MmyJ-DNA interactions *in vivo* (Fig. 30).

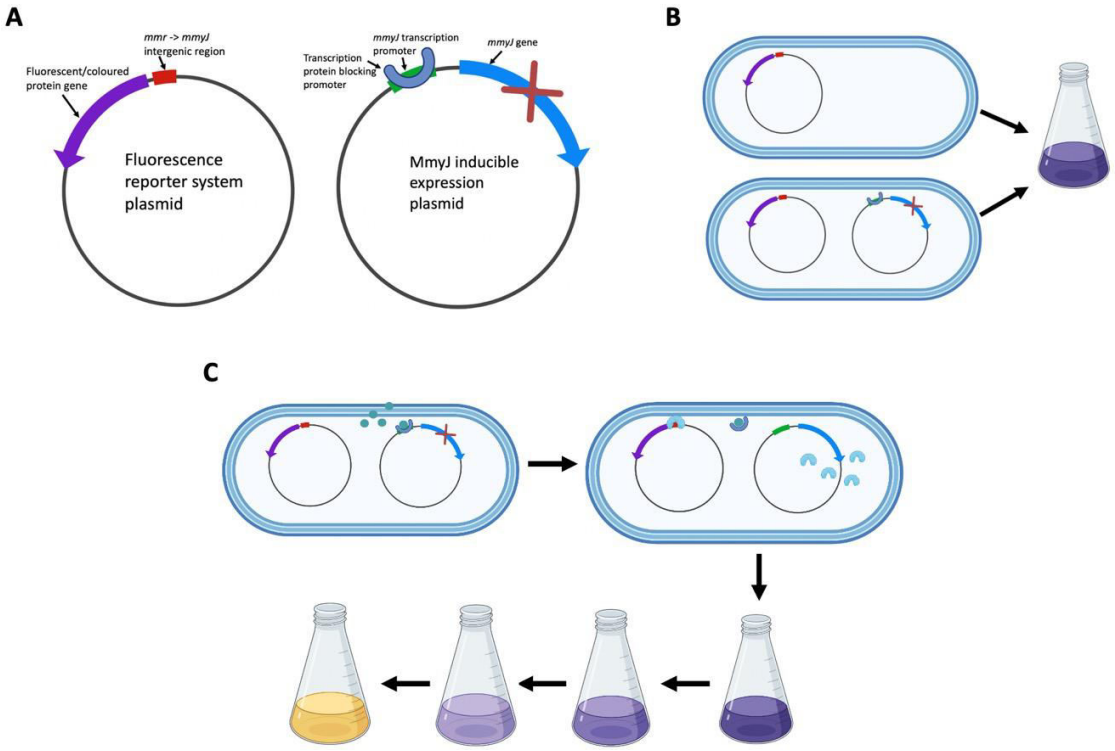


Figure 30. Theoretical basis on how a MmyJ reporter system may work *in vivo* to assess strength of MmyJ-DNA binding. The basis of *in vivo* experiments using a fluorescence (or coloured protein) reporter system to show MmyJ binding to DNA properties. **A)** First 2 different plasmids would be required**.** One plasmid would be the fluorescence producing reporter plasmid where the intergenic DNA MmyJ binds to is inserted just prior or just after the promoter of the fluorescence gene. The second would be a plasmid with an inducible system for the expression of the *mmyJ* gene. **B**) Transforming both of these plasmids into a singular bacterial cell should allow production of the fluorescence protein just as it would with only the fluorescence plasmid inserted only into a cell. **C)** Once an inducer (teal molecules) for the MmyJ plasmid has been added, the cell can express MmyJ. MmyJ would then likely bind to the intergenic DNA region inserted in the fluorescence plasmid and block expression of the fluorescence gene. Over time this would cause a decrease in fluorescence. (Figure created using Biorender.com).

3.6.1 A pJ251 Tinsel purple reporter system

Initial work done prior by Lougher M. (2015) [155] had created one such reporter system for *in vivo* analysis of MmyJ in binding/unbinding to DNA. This reporter system uses a pJ251 *E. coli* plasmid with a gene expressing tinsel purple coloured protein inserted (Fig. 31) just after the ribosomal binding site (RBS) and intergenic region of DNA between *mmr* and *mmyJ*. This plasmid was synthesised and purchased from Atum. Unfortunately, the tinsel purple colour as described commercially was not seen at all in *E. coli* TOP10 transformations.

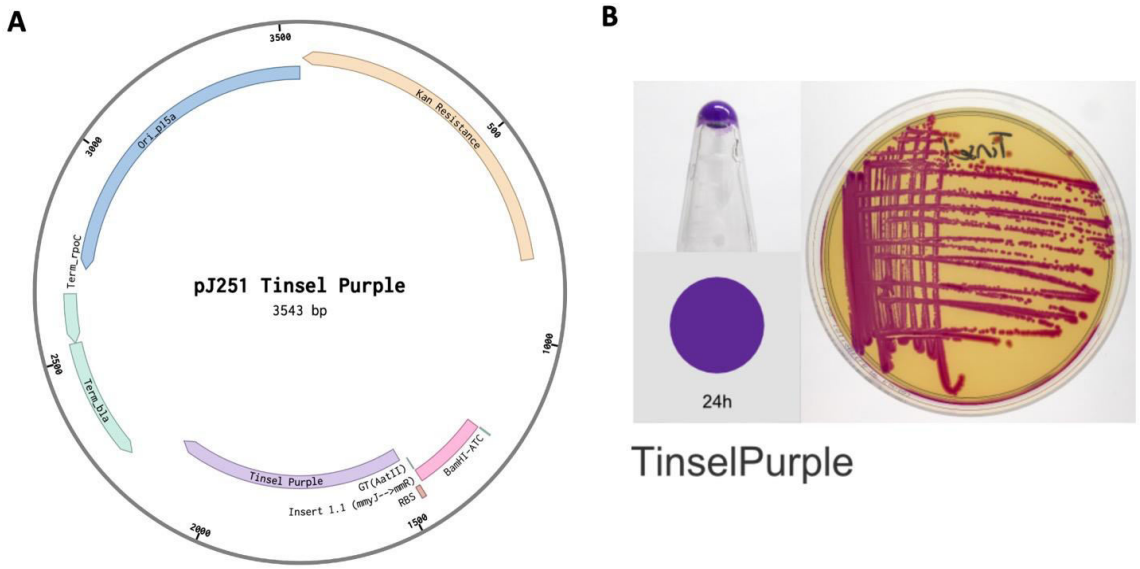
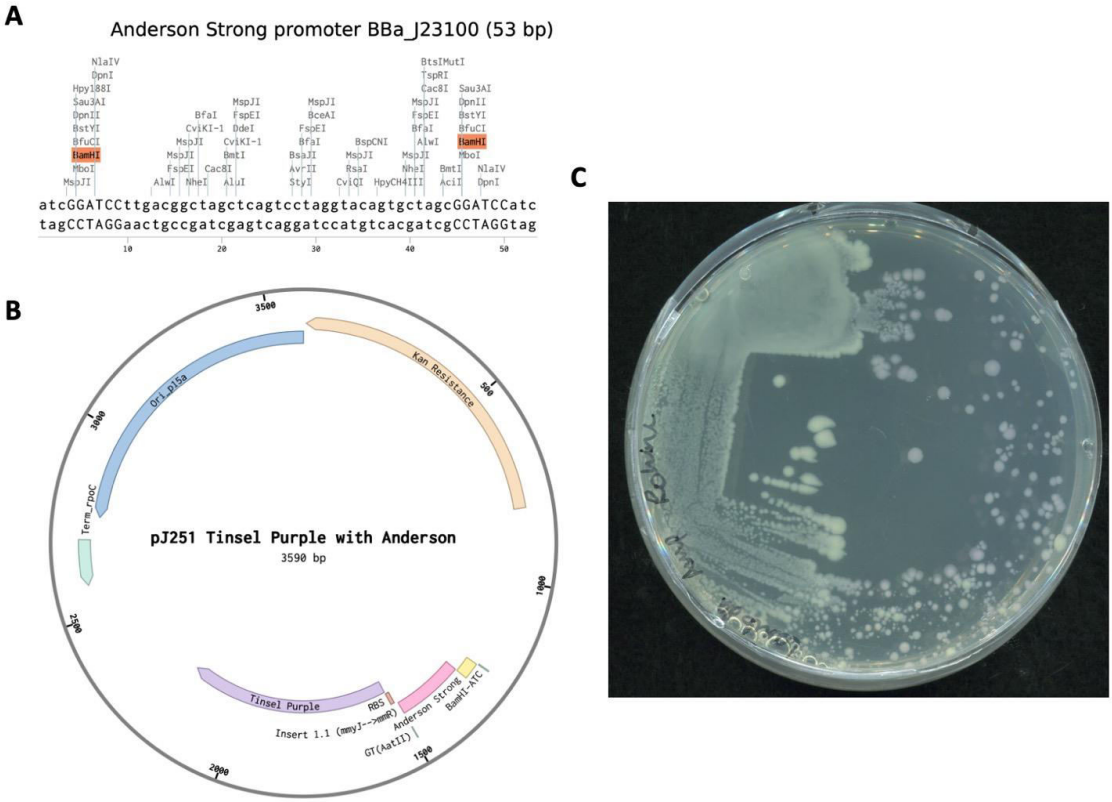


Figure 31. The pJ251 Tinsel Purple reporter system. A) This pJ251 Tinsel Purple system was designed and purchased commercially from Atum and made available for use to test MmyJ binding *in vivo*.[155] The plasmid contains the coloured protein gene (Tinsel purple) with the intergenic DNA MmyJ binds to (Insert 1.1) being inserted prior to the ribosome binding site (RBS) of the Tinsel purple gene. A BamHI enzyme site also present before the intergenic region to allow for excision or further cloning. **B**) The expected colour of Tinsel purple gene in *E. coli* after 24 hours, both as a pellet and on an LB media plate. Figure from Atum.com.

This may have been that the promoter within the plasmid for the Tinsel Purple gene was not strong enough. Using a BamHI enzyme cut site just before the intergenic *mmyJ* and *mmr* region, the Anderson Strong promoter (part BBa J23100 from iGem) was used (Chapter 2: Materials & Methods) (Fig. 32A). This promoter has been shown to be a strong *E. coli* promoter.

Once inserted and transformed, colonies of cells that grew on antibiotic-selective

agar after 1 day of incubation at 37 ℃ did not give any noticeable colour change (Chapter 2: Materials & Methods). However, after 3 days of incubation the same colonies were observed to be faintly pink (Fig. 32B). The colour of these colonies did not increase in intensity after further incubation. The faint pink/purple coloured colonies are not the same bright purple as expected (Fig. 31B).



a strong *E. coli* promoter flanked by BamHI sites (highlighted in orange) to allow insertion into plasmids. **B)** This Anderson Strong promoter (yellow) was cloned into the pJ251 Tinsel Purple plasmid just prior to the intergenic DNA region by the BamHI site (shown in teal). The cloned plasmid was confirmed by Sanger sequencing (Appendix Fig. A9). **C)** A solid LB media plate with left half of *E. coli* containing pJ251 Tinsel Purple without the Anderson strong promoter (Fig. 31) and right half containing colonies of pJ251 Tinsel Purple with the Anderson Strong promoter (**B**). Plasmids with the Anderson strong promoter show colonies of a light purple colour.

These coloured colonies were also taken for sequence analysis to identify the presence of the pJ251 tinsel purple-Anderson strong plasmid in the cell (Appendix Fig. A9).

In attempts to improve colour production different growth conditions of the cells was tested. Transformed colonies were spread onto various agar media (3 of which were rich media and 2 minimal media) as well as testing growth at 4 different temperatures. While the lighter coloured media allowed for better visualisation of the faint pink colonies due to their transparency, there were no visible difference in the intensity of the coloured colonies or a difference in colour as all colonies appeared a mix of pink and white (Appendix Fig. A10). Furthermore, colonies then grown in liquid media did not produce the any pink pigmentation as seen on plates (Appendix Fig. A10).

Liquid media cultures measuring fluorescence would be ideal as measurements can be taken over the course of a time period and to give 'real time' activity measurements. Solid media colonies could still be used to give snapshot insight of MmyJ binding capabilities to DNA. This would require the transformation of an inducible MmyJ plasmid containing a different antibiotic resistance marker to that of this Tinsel purple plasmid in order to select for colonies containing both plasmids. Due to time constraints this MmyJ plasmid could not be engineered for the proposed experiment.

Nonetheless, despite the addition of a strong promoter within the Tinsel purple plasmid there is still not a great difference in 'tinsel purple' production. This indicates a problem with the 'tinsel purple' gene itself or the expression of it. The issue of tinsel purple expression may be due the addition of the *mmr-mmyJ* intergenic region prior to the tinsel purple gene. This region while rich in GC content does not contain any start/stop codons or code for any toxic molecules that could negatively then affect transcription further downstream, however there may be unknown effects of having this region before the tinsel purple gene that is causing low tinsel purple expression.[155]

To verify that this is a problem with the addition of the intergenic region rather than an issue with the tinsel purple gene the intergenic region would need to be

removed to test if tinsel purple expression improves and subsequently produces a purple colour colony or not. This region could have been cloned out however there were no suitable restriction sites available to excise only the *mmr-mmyJ* region. Excising the *mmr-mmyJ* intergenic region in the pJ251 Tinsel purple plasmid construct may also be possible through cloning methods such as Gibson assembly. Should cloning methods fail, the gene construct could be purchased again without the *Streptomyces mmr-mmyJ* intergenic region. Alternatively, another reporter system could be tested to also determine if this in an issue with the intergenic region or expression of the tinsel purple gene or any other issues with the plasmid construct itself. Should the issue be the intergenic region DNA, shorter lengths of around 40-80 bp of this 218 bp fragment could be tested as an 80 bp region was shown *in vitro* to have MmyJ strongly binding to it (Fig. 22).

3.6.2 The mScarlet reporter system

A reporter system that uses a strong *Streptomyces* SP44 promoter[225] driving expression the mScarlet-I fluorescence gene was available in the Corre group (Appendix Fig. A11) as an alternative reporter system to pJ251 Tinsel Purple. The mScarlet-I plasmid contains the SP44 promoter followed by the plasmid RBS, after which is the *mScarlet-I* gene. This offers 3 potential sites of insertion of the intergenic *mmr -> mmyJ* through enzymatic cut sites: NdeI site for insertion just before the *mScarlet-I*, ScaI for addition just after the SP44 promoter and 2 base pairs into the RBS and an XbaI site for insertion ~20 base pairs before the SP44 promoter itself (Fig. 33A).

The full 218 bp intergenic *mmr -> mmyJ* sequence was seen as being potentially too disruptive in the reporter system as described above with the pJ251 Tinsel Purple system therefore only the 80 bp seen as functional *in vitro* was used for insertion into the mScarlet reporter system.

To test whether insertion of even this reduced sequence of DNA would disrupt the production of mScarlet-I without addition of MmyJ, the 80 bp sequence was attempted to be inserted into the XbaI site, prior to the SP44 promoter (Fig. 33B). The mScarlet-I protein (a protein with a single amino acid mutation from mScarlet) emits a bright pink colour when expressed.[226] Insertion of the intergenic *mmyJ-mmr* DNA yielded in bright pink, light pink and white colonies both on solid media agar and liquid cultures (Fig. 33C). The mScarlet plasmids from samples of white, light pink and bright pink colonies (Chapter 2: Materials & Methods) were digested with the restriction enzyme NdeI which cuts within the 80 bp fragment if it was successfully inserted into the mScarlet plasmid at the XbaI site (Fig. 34). All appeared to contain the 80 bp fragment according to Fig. 34 however this would not explain the phenotype differences seen. Due to the large difference in fragment sizes expected (3118 bp and 142 bp) digest results may have been unreliable. Therefore, the same plasmids used for restriction digestion along with 2 additional plasmids from colonies coloured white and light pink and 1 additional plasmid of a bright pink colony were sent for Sanger sequencing. Results from Sanger sequencing of the mScarlet plasmids revealed that the expected intergenic region was not detected in any of the white, light pink or bright pink colonies, highlighting unreliable and possibly unsuccessful cloning (Appendix Fig. A12).

The other 2 restriction sites (NdeI and ScaI) described above, present after the SP44 promoter was tested additionally to determine whether it is the site of insertion or the insert itself that is the issue. At both these sites, while colonies were varied in mScarlet pigmentation when sequenced no complete inserted 80 bp DNA was detected (Appendix Fig. A13).

Attempts of inserting the shorter (80 bp) intergenic *Streptomyces mmr -> mmyJ* region sequence into this mScarlet plasmid were unsuccessful so could not be used to help identify whether or not it is the intergenic *mmyJ-mmr* region of DNA that is an issue in the pJ251 Tinsel Purple system. A variation in pigmentation (white, light pink and bright pink) of colonies was produced from cloning work in the mScarlet plasmid despite no reliably successful insertion of the intergenic *mmyJ-mmr* DNA.

Without sequencing the entire plasmid of colonies exhibiting these variations in coloured pigments, the reason behind this result will remain unclear.

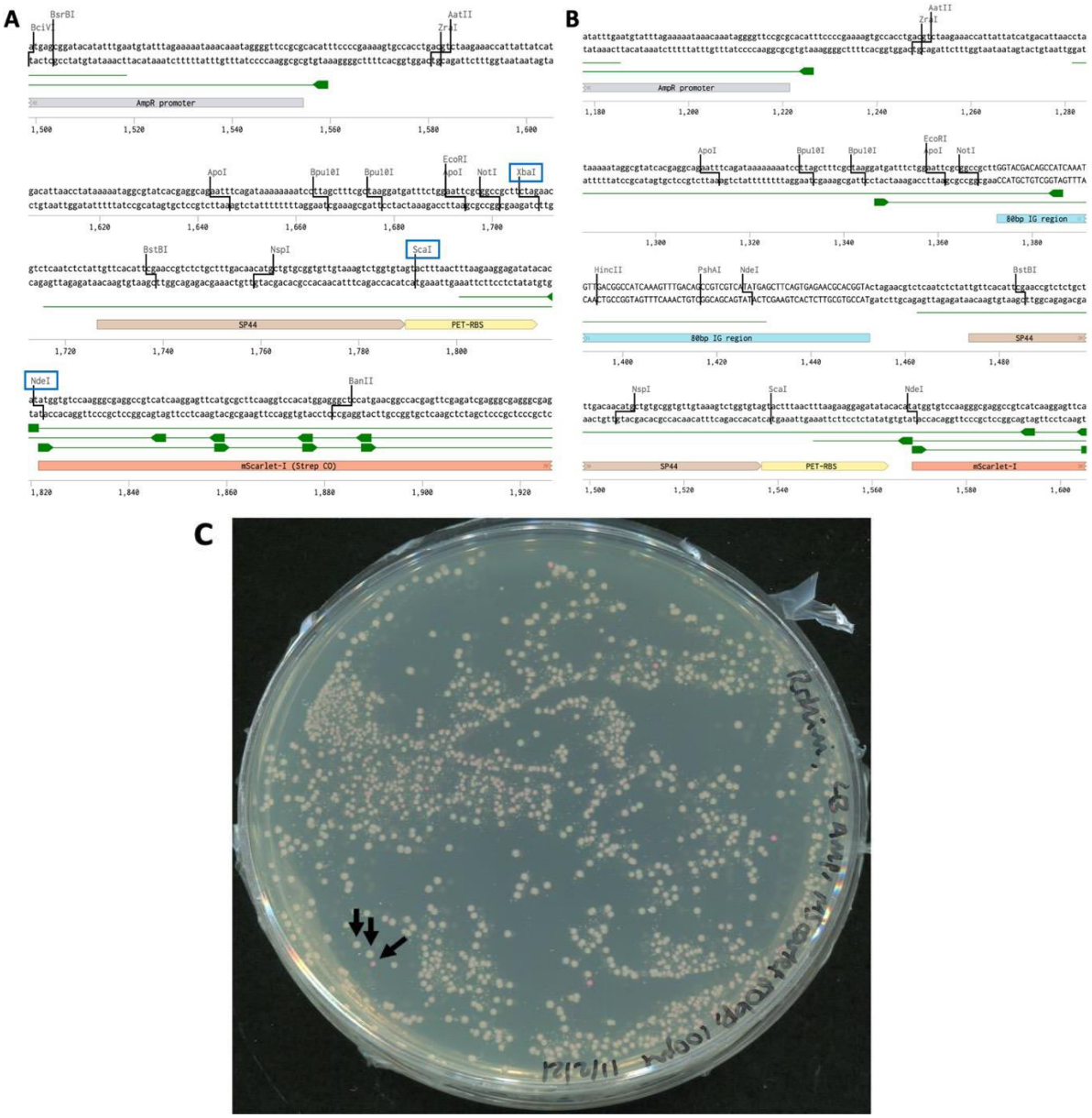


Figure 33. Intergenic DNA was inserted into an mScarlet reporter plasmid for testing MmyJ-DNA binding *in vivo*. An mScarlet reporter plasmid (Appendix Fig. A11) contains the *mScarlet-I* gene that emits a bright pink colour. A SP44 promoter site is present for increased production of the *mScarlet-I* gene. Three cut sites where restriction enzymes (in blue boxes, XbaI, ScaI and NdeI) were identified as locations for insertion of an 80 base pair *mmyJ-mmr* DNA fragment (**A**)**.** The 80 bp (Fig. 21) intergenic region of DNA for MmyJ was first cloned (Chapter 2: Materials & Methods) into the mScarlet plasmid using the XbaI site **(B**)**.** This site was chosen due to it being prior to the SP44 promoter site and limit the effect of this addition of DNA to the production of mScarlet-I protein without addition of MmyJ. **C**) Transformed TOP10 *E. coli* cells (Chapter 2: Materials & Methods) showed colonies with 3 phenotypes (shown by black arrows). Some colonies were light pink, white or bright pink (left to right, first to third arrow respectfully).

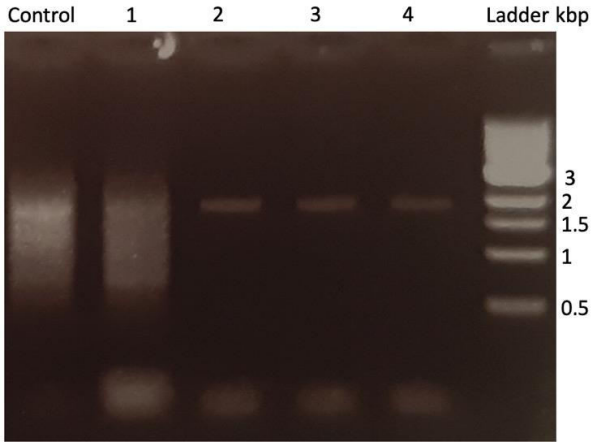


Figure 34. Plasmids from colonies of mScarlet with an 80 base pair (bp) *mmyJ-mmr* intergenic DNA were digested with NdeI and run on a 1% TAE-Agarose gel. Plasmids from bright pink (1 and 2), light pink (3) and white (4) colonies with an 80 bp *mmyJ-mmr* intergenic region of DNA attempted to be inserted into an mScarlet plasmid at an XbaI site, was digested with NdeI and was run on a 1% TAE-Agarose gel (Chapter 2: Materials & Methods). The mScarlet plasmid without the intergenic DNA (control) is 3180 bp and with intergenic DNA 3260 bp. NdeI cuts within the intergenic DNA and plasmids containing this DNA would produce DNA fragments at 3118 bp and 142 bp. NdeI cuts only once in the mScarlet plasmid without intergenic DNA (control) giving a 3180 bp. Fragments around 3 kilo bp and under 500 bp is seen in samples 1-4 but Sanger sequencing did not support the inclusion of the intergenic DNA.

3.7 Release of DNA-bound MmyJ by methylenomycin A and methylenomycin A intermediates

3.7.1 Ligand control of transcriptional proteins

Almost all, if not all, transcriptional reporters bind to not only a particular DNA sequence but also to particular ligands. For ArsR proteins these are generally metals ions however as discussed before, there are exceptions to this. [202,209,211,227] As MmyJ is thought of as a non-metal sensing or regulating ArsR protein it stands to reason the ligand for MmyJ is not a metal ion. Instead, it is proposed to bind to MmA or an intermediate of MmA. This assumption is made under the basis that MmyJ likely controls the *mmr* gene which confers the MmA resistance efflux pump and thus regulates MmA concentrations in the cell. Therefore, sensing the cellular

concentrations of MmA is necessary for this function. This may not be directly done by MmyJ of course and there could be a cascade of molecules that lead to the release of MmyJ from being bound to DNA. However, with the inclination that it's part of the ArsR family, which directly senses the ligands it helps to control, MmyJ also most likely senses MmA or an intermediate of MmA directly too.

3.7.2 Production of the predicted MmyJ ligand: Methylenomycin A

MmA must be produced and purified to a high enough yield that it can be used for assays with MmyJ. As it cannot be bought commercially it must be cultured in *Streptomyces* and purified through chromatography.

Many *Streptomyces* strains can express MmA with the SCP1 plasmid inserted (as MmA is SCP1 encoded). The *Streptomyces* used here were various *S. coelicolor* strains (a model organism for natural product production[228]) which contained the methylenomycin gene cluster with and without mutations.[145,177] The M145 strain does not contain the methylenomycin gene cluster meaning no MmA is produced and so was used as a negative control. The strain W89 contains a knockout of *mmyR* which would normally produce the transcriptional repressor MmyR that controls expression of methylenomycin biosynthesis genes as discussed in Chapter 1 (Fig. 35A). Not having this repression of MmA biosynthesis should allow for overproduction of MmA. W108 contains a knockout of the *mmyF* gene which produces the protein MmyF that assists in the epoxidation of MmC (the precursor) into MmA meaning this strain overproduces MmC (Fig. 35B). MmA and MmC can be structurally similar and so W108 can be used as a positive control to help distinguish between MmA and MmC .[145]

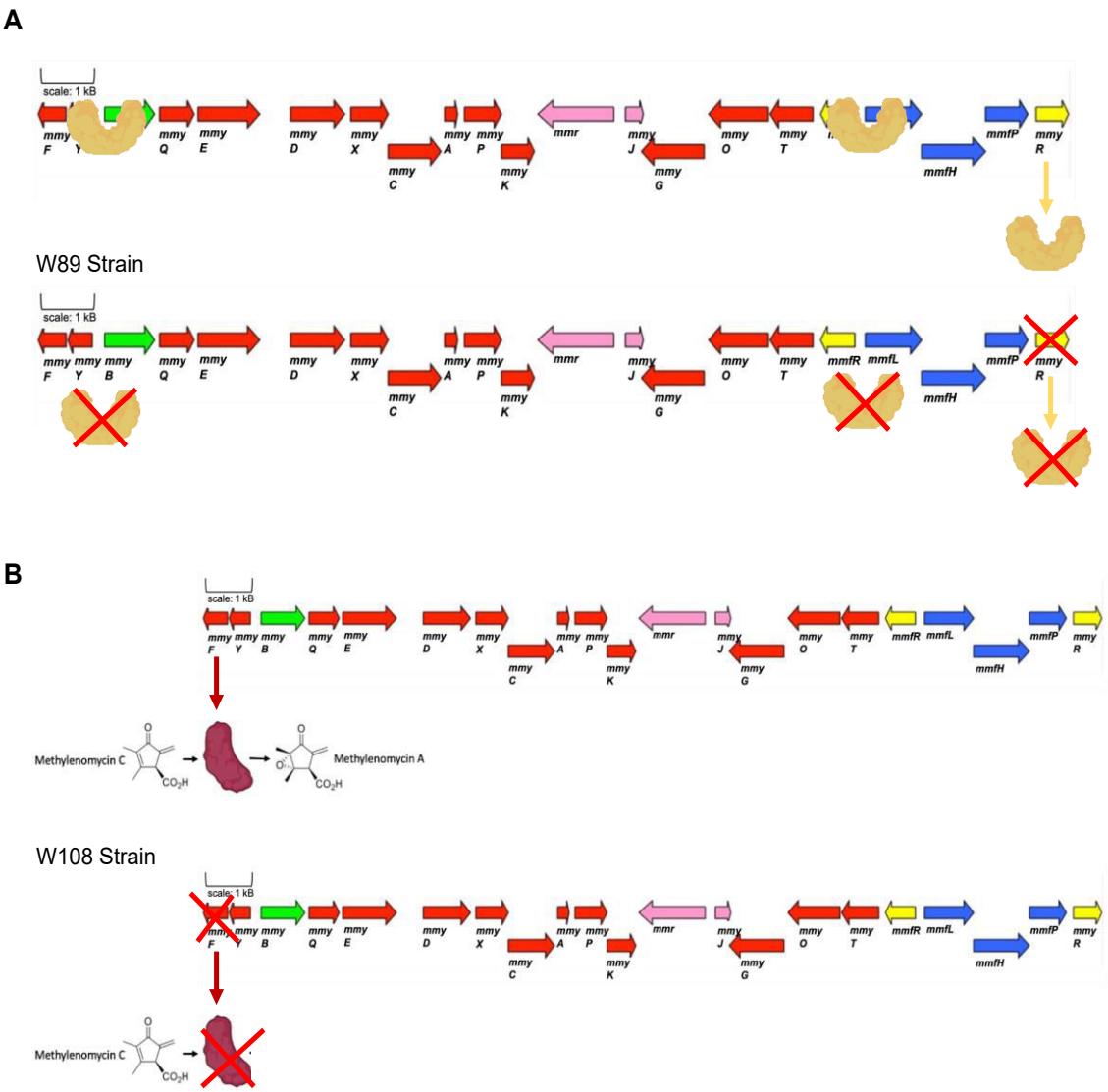


Figure 35. Methylenomycin gene clusters in *Streptomyces* strains used for production and extraction of Methylenomycin A (MmA). The M145, W89 and W108 strains were used for the production of MmA (Chapter 2: Materials & Methods). The M145 strain doesn't contain the Methylenomycin gene cluster and so is a negative control. **A)** The W89 strain is used as a sample strain that overproduces MmA. The MmyR protein acts as a transcriptional regulator for the biosynthesis of MmA. In W89, the *mmyR* gene is knocked out so MmA biosynthesis is constitutively on. **B**) MmA and Methylenomycin C (MmC) are similar and W108 overproduces MmC so can be used as a positive control. MmyF assists in the epoxidation of MmC to MmA. The W108 strain has a knockout of *mmyF* and so MmC cannot be converted to MmA.

These strains can be cultured in liquid or solid media. There is no apparent distinction between MmA production between liquid or solid media that has been shown previously, however as MmA is exported into the supernatant liquid media culturing in terms of extraction is simpler.[145] Both methods were tested with liquid media being first trialled.

Cultures of *S. coelicolor* W89, M145 and W108 grown in supplemented minimal media (SMM) were organically extracted with ethyl acetate, dried and subjected to HPLC (Chapter 2: Materials & Methods). At 225 nm (UV absorbance for MmA) a distinct peak between the W89 and the other controls was observed (Fig. 36) however subjecting this to LC-MS revealed this peak to be of 167.99 m/z (mass over charge) (Fig. 37). No other distinct peaks in the W89 strain (sample strain) and the controls were seen so other fractions from this HPLC were not taken for LC-MS. This 167.99 m/z peak likely corresponds to that of MmC rather than MmA. MmA gives a m/z of 183.06 g/mol in positive ion mode however sometimes due decarboxylation within the MS instrument a peak can also be seen at 139.00 m/z. Neither of these masses were seen however with this liquid media cultures (Fig. 37).

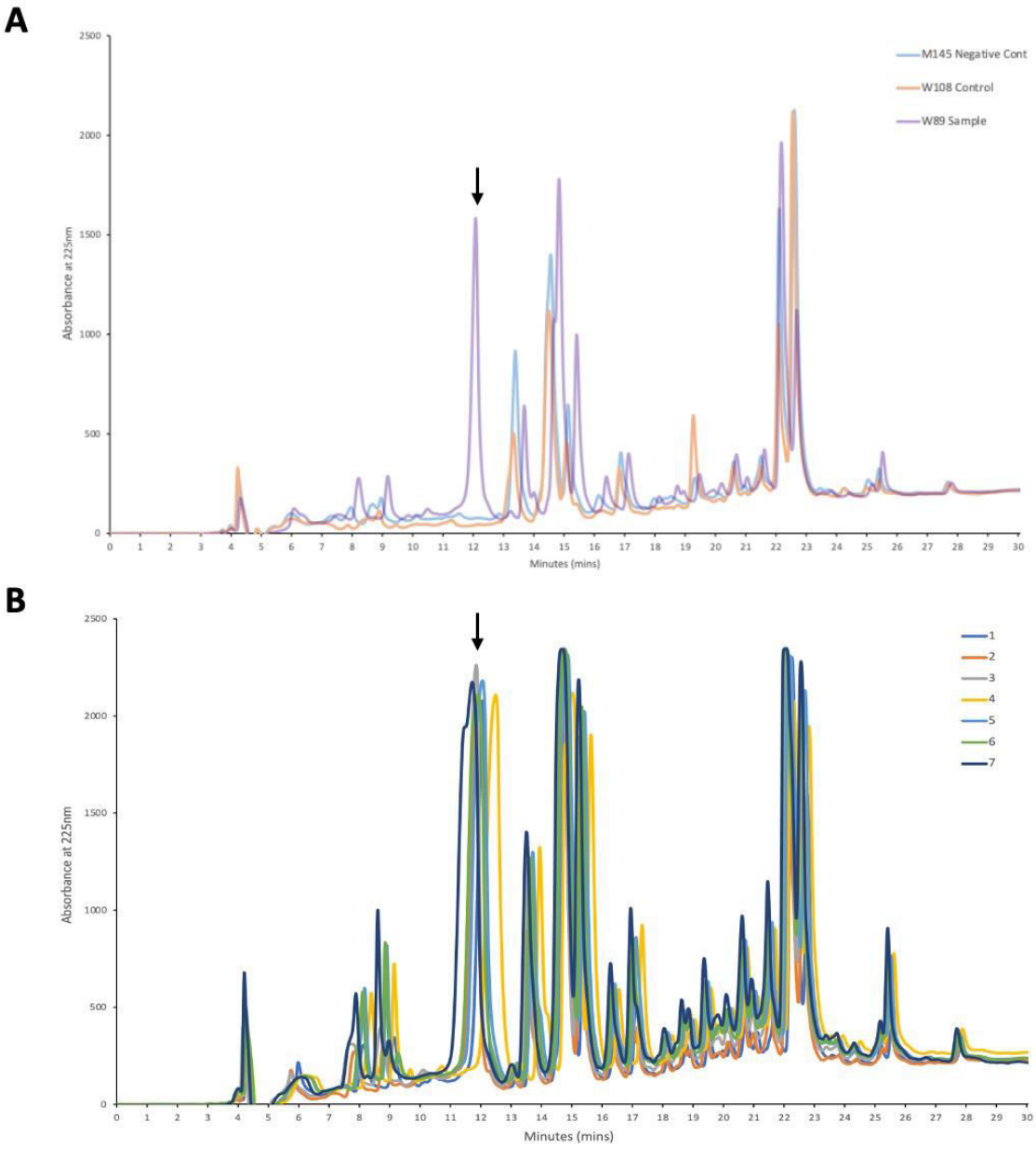


Figure 36. High-performance liquid chromatography (HPLC) of samples of W89, W108 and M145. Liquid cultures of W89, W108 and M145 were grown in supplemented minimal media (SMM) and organic compounds were extracted. To separate and detect Methylenomycin A (MmA) this was subjected first to analytical (**A**) and then multiple runs of preparative (**B**) HPLC. MmA absorbs at 225 nm and both **A** and **B** shows for absorbances at 225 nm. In **A** the W89 sample is indicated by purple line and black arrow indicates where a peak is seen in only W89 and not in the W108 or M145 controls. As M145 and W108 cannot produce MmA this is thought to be MmA. The same peak is indicated by a black arrow in **B** in multiple preparative runs of HPLC of only the W89 sample.

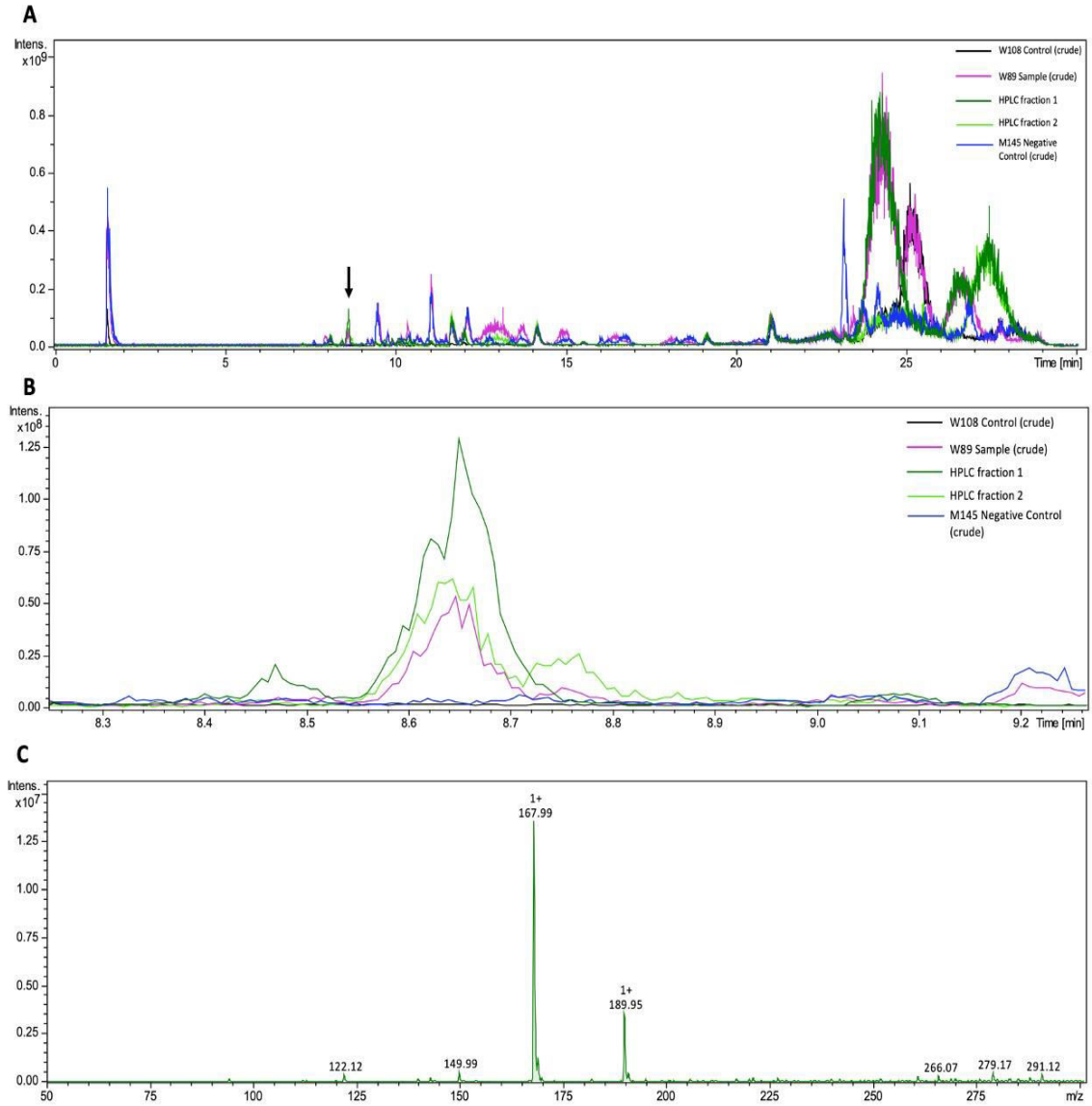


Figure 37. Low resolution, liquid chromatography mass spectrometry (LC-MS) of high-performance liquid chromatography (HPLC) purified W89 samples and of crude W89, M145 and W108 samples to detect Methylenomycin A (MmA). **A** and **B** show the baseline chromatogram of HPLC purified W89 samples (Fig. 36) and crude extracts of all three stains. Black arrow in **A** indicates a peak where there is a difference in the W89 and HPLC purified samples (dark green, light green and purple) and the W108 (black) and M145 (blue) samples. This peak is shown further in **B** and the mass is seen in **C**. The m/z shows a signal at 167.99 which is indicative of Methylenomycin C (MmC) which in positive ion has a m/z of 167.00.

Solid media was then tested. Using agar plates of SMM, the W89 (sample) and M145 (negative control) strains were cultured, and organic compounds were extracted (Chapter 2: Materials & Methods).

Another method of detection is to use thin layer chromatography (TLC). This is a relatively quick and cheap method to detect compounds using solvents and would have been a good way to check for the presence of MmA before confirming with the more time consuming and expensive mass spectrometry techniques. TLC was tested with the above SMM solid media cultures (Chapter 2: Materials & Methods) with bromocresol green used as a stain (Fig. 38B). Bromocresol green is an ideal stain for detection of carboxylic acids that MmA has as it does not require heat. Instead, the blue form of the stain in the presence of a carboxylic acid turns yellow as the pH drops (Fig. 38A).[229] TLC testing of M145 and W89 samples, did not show any distinct yellow dots to indicate the presence of a compound with carboxylic acid even though crude samples were used (Fig. 38B). Therefore, further fine tuning would be required to use TLC as an alternative or prior technique to mass spectrometry.

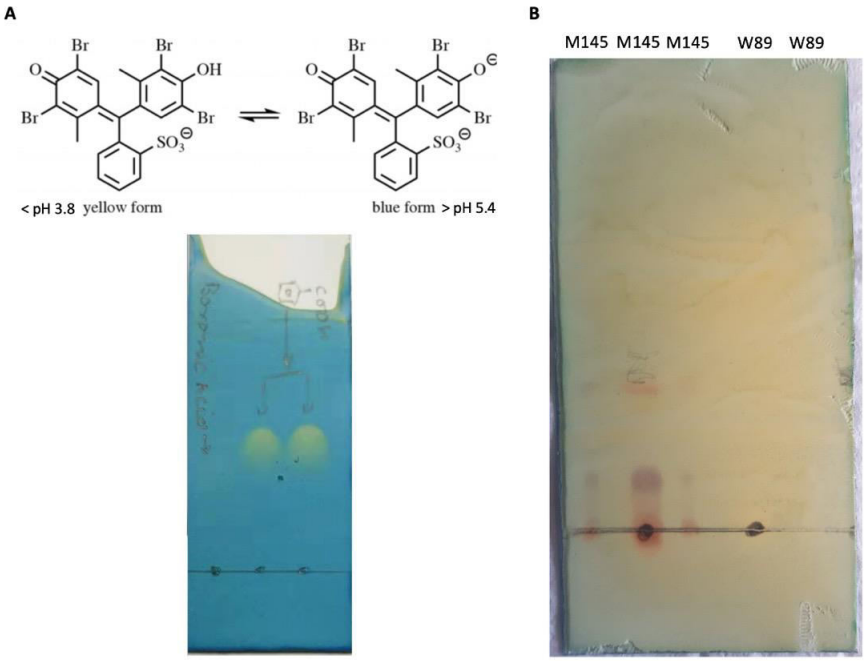


Figure 38. Thin layer chromatography (TLC) of M145 and W89 samples to indicate for MmA. **A)** TLC stained by bromocresol green can indicate for carboxylic acid compounds without use of heat. An example of that is shown in **A**. The blue form is present when the pH is above 5.4 and below 3.8, the stain turns yellow. Carboxylic acids would therefore turn the stain yellow and compounds with them would show as yellow spots on a blue TLC plate.[207] **B**) TLC was attempted on samples of M145 and W89 strains extracted from solid supplemented minimal media (SMM). Bromocresol green failed to stain effectively and so no carboxylic acid containing compounds could be detected. Further fine tuning of this method is therefore required.

Due to inconclusive results by TLC, crude extracts of solid SMM grown W89 and M145 samples were subjected to low-resolution LC-MS (Fig. 39). There appears to be production of various methylenomycin products including MmA that are not seen in the negative control (Fig. 39A). MmA is seen at 183.14 m/z and at 139.14. As this is low resolution there was a +/- 0.5 tolerance and so these masses fit into the predicted ones of MmA at 183.06 m/z and 139.00 (decarboxylated MmA) in positive ion mode (Fig. 39C). Solid agar (500 mL) was extracted for this experiment and while the LC-MS sample was only a small percentage (1%) of the crude extract, an overproduction of MmA would imply a greater level of MmA detected in comparison to the other methylenomycin gene cluster products (e.g., MmC). This method therefore could still be optimised.

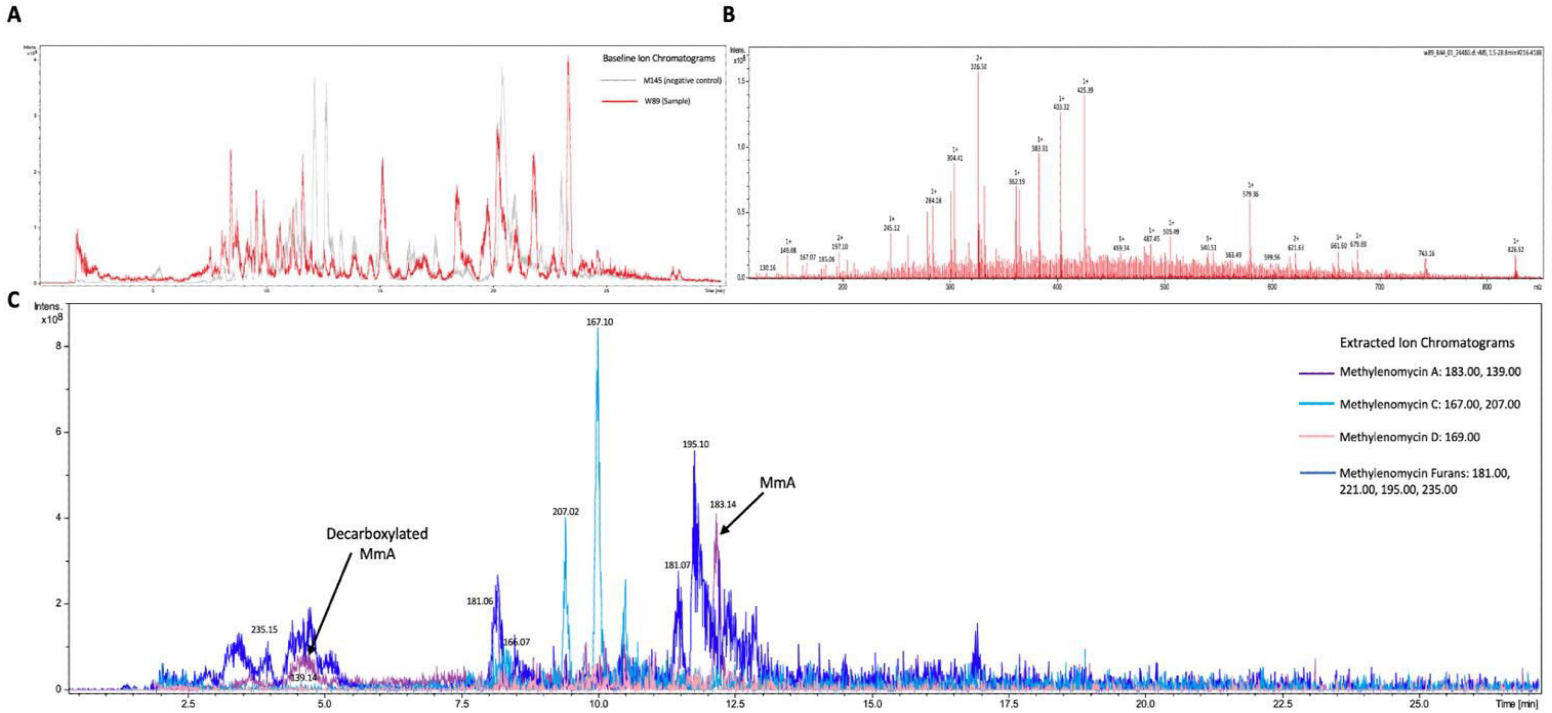


Figure 39. Low resolution, liquid chromatography mass spectrometry (LC-MS) of W89 and M145 samples for Methylenomycin compound detection. Organic compounds from M145 and W89 samples grown on solid supplemented minimal media (SMM) were subjected to LC-MS. **A**) The base peak chromatogram of both samples. The W89 sample (red) has many peaks that are different to the negative control strain sample-M145 (grey). **B**) The major masses of the entire chromatogram of W89. None of these masses are of Methylenomycin A (MmA) indicating there are other organic compounds being produced in higher amounts. **C**) The extracted ion chromatogram (EIC) of W89 for Methylenomycin compounds; MmA (purple), Methylenomycin C (MmC) (light blue), Methylenomycin D (MmD) (peach) and Methylenomycin furans (MMFs) (blue). MmC are a precursor to MmA while MmD is a degradation product and MMFs are small molecules that promote the biosynthesis of MmA. MmA can be seen in this sample in positive ion mode (m/z 183.14) and in its decarboxylated form (m/z 139.14).

3.7.3 Optimising the production Methylenomycin A

To optimise the production of MmA a different media was tested. The media was switched from the previously used SMM to rich yeast extract media (R2YE) which is also used for antibiotic production in *Streptomyces.*[230] The W89 strain and M145 control strain were grown on solid R2YE plates and after extracting the organic compounds (Chapter 2: Materials & Methods) these crude extracts were subjected to low resolution LC-MS. Comparison of the 2 strains show similar differences in peaks (Fig. 40A) which were also seen previously (Fig. 39A). When the specific extracted ion chromatogram (EIC) for MmA is taken, there is clearly a peak in the W89 sample strain but not the M145 negative control strain that appears to be of the same mass as for MmA.

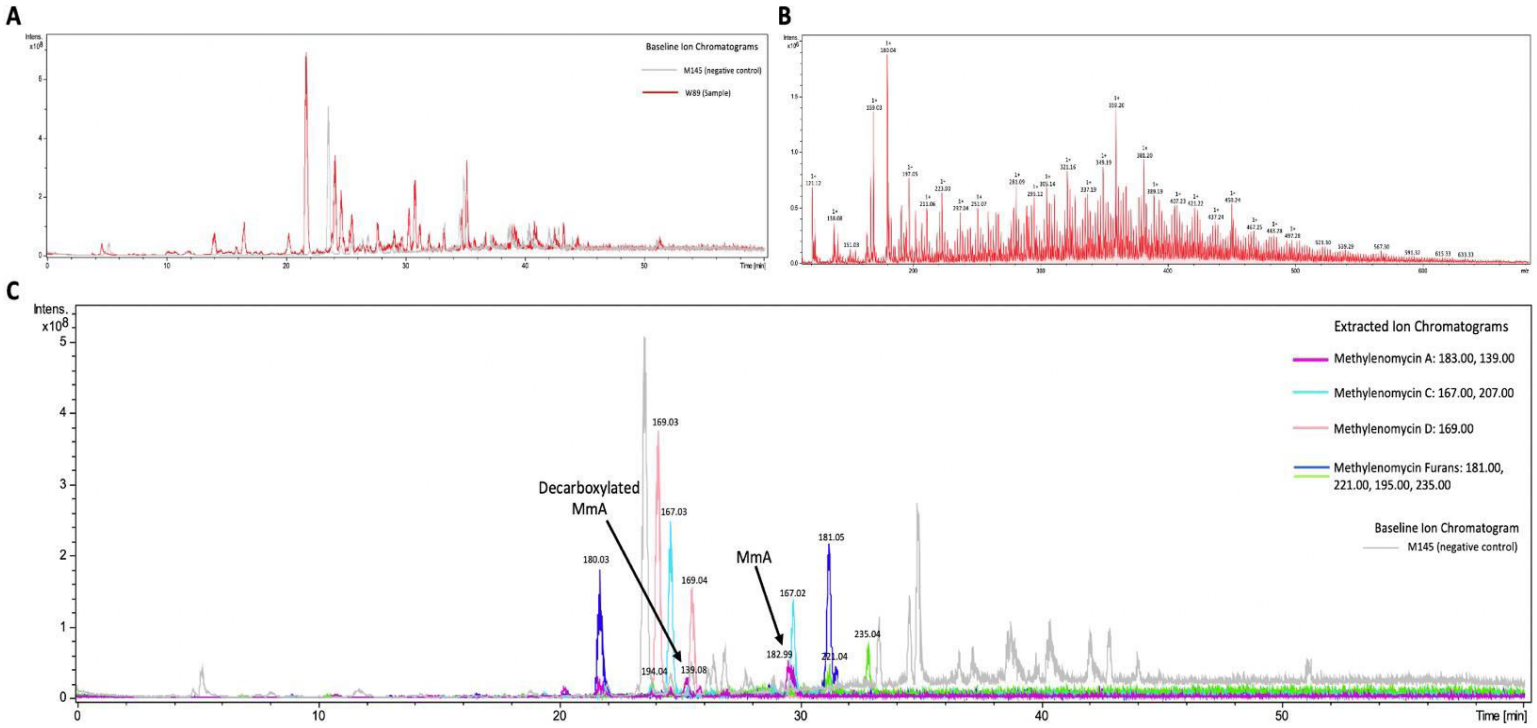


Figure 40. Low resolution, liquid chromatography mass spectrometry (LC-MS) of W89 and M145 samples for Methylenomycin compound detection. Organic compounds from M145 and W89 samples grown on solid rich yeast media (R2YE) were subjected to LC-MS. **A**) The base peak chromatogram of both samples. The W89 sample (red) has many peaks that are different and of greater intensity to the negative control strain sample- M145 (grey). **B**) The major masses of the entire chromatogram of W89. Some of these are likely Methylenomycin compounds indicating this strain is overproducing products of this gene cluster. **C**) The extracted ion chromatogram (EIC) of W89 for Methylenomycin compounds; Methylenomycin A (MmA) (pink), Methylenomycin C (MmC) (cyan), Methylenomycin D (MmD) (peach) and Methylenomycin furans (MMFs) (blue and green). MmC are a precursor to MmA while MmD is a degradation product and MMFs are small molecules that promote the biosynthesis of MmA. MmA can be seen in this sample in positive ion mode (m/z 183.14) and in its decarboxylated form (m/z 139.14). High levels of MmyD indicate the sample will need to be harvested sooner before degradation starts to occur.

This peak is fairly low in intensity however and taking the EIC's of other precursor and degradation products in the MmA production scheme there are high intensity peaks of MmC and MmyD (Fig. 40C). MmyD is a degradation product that appears when MmC is present in abundance. The large presence of these in comparison to MmA implies these cultures were perhaps extracted too late and shortening the growth time may allow for the maximum level of MmC to be converted to MmA before it then degrades into MmyD. Only 40 mL of media was used for creating this sample, indicating that R2YE is a better medium than SMM in producing MmA. Another advantage of R2YE is the shorter growth time (~ 2 days) compared to SMM (~ 6 days) for similar yields. Scaling this up and purifying through HPLC would yield high enough amounts for other analysis.

3.7.4 Crude extracts containing mainly Methylenomycin C do not release MmyJ from being bound to DNA

While pure, concentrated MmA was not able to be obtained, the crude extracts from these cultures could still be utilised for assays that test the interactions between MmyJ, DNA and ligand binding. Indeed, crude organic extracts of the W89 strain that is considered to produce MmA were tested with previously purified MmyJ and the 80 bp intergenic DNA it has been shown to bind to (Chapter 3.3.1).

Crude organic extracts of W89 grown on liquid SMM (above- Chapter 3.7.2) were dried and reconstituted with just enough water (~20 ul) to allow for complete reconstitution of compounds. These were these used in various ratios in EMSAs with MmyJ and the 80 bp intergenic DNA (Fig. 41). From the mass spectrometry data (Fig. 37) it was shown that this extract contained no (or such little it was undetectable) amounts of MmA and mainly consisted of MmC. With the picomolar scale these assays are done at only a small amount of MmA would be needed to show any unbinding. However as seen in Fig. 41 no unbinding was distinguishable.

There was likely no/insufficient concentrations of MmA within the extract to reliably draw conclusions from this experiment. With MmA there may or may not be unbinding however the concentration would need to be increased to differentiate between there being no binding due to the MmyJ-DNA complex simply not viewing MmA as a ligand and there being no binding due an insufficient amount of MmA present. Regardless, this extract contained MmC which is structurally very similar to MmA (Fig. 7). With no unbinding detected of the complex in the presence of MmC, this indicates that if MmA is the ligand for MmyJ, MmyJ is highly specific and able to distinguish MmA from structurally similar compounds. There may likely therefore be interactions with the epoxide part of MmA and the MmyJ binding

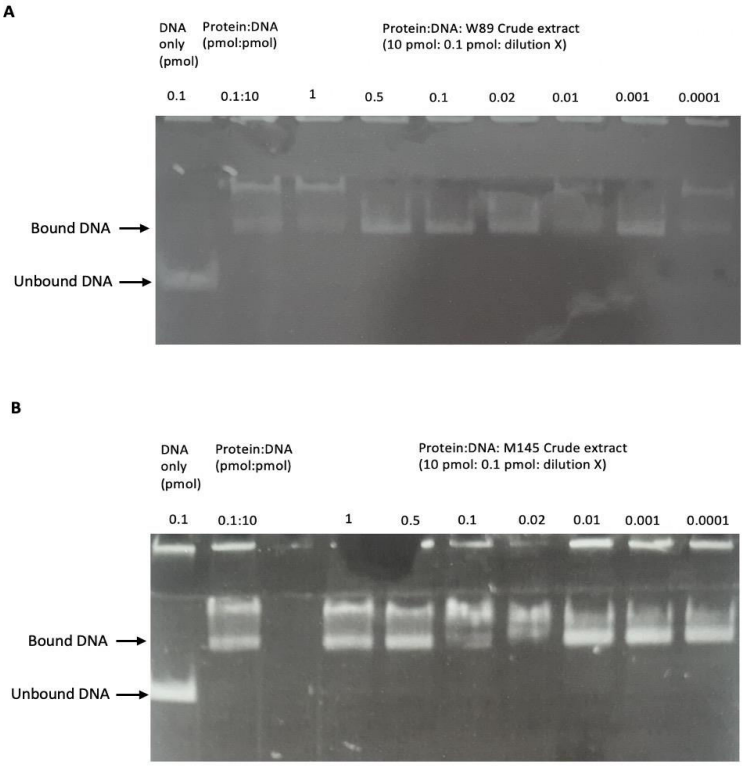


Figure 41. Electromobility shift assays (EMSAs) of MmyJ, DNA and ligand binding. MmyJ has been shown already to bind to DNA (Fig. 22) but as a transcriptional protein it also has a ligand domain to release it from DNA binding. This ligand may be a Methylenomycin compound itself. **A**) To test this a crude extract of Methylenomycin compounds (Fig. 37) was added in various dilution factors with samples of MmyJ and DNA. This extract contained a majority of Methylenomycin C (MmC). **B)** An extract with no Methylenomycin compounds was also tested to ensure other organic compounds in the sample was not producing a false positive result. From **A** no clear unbinding is seen and so MmC is unlikely to be the ligand that binds to MmyJ.

pockets.

3.7.5 Testing MmyJ binding with crude extracts containing MmA

Also tested were crude organic extracts of W89 grown on solid R2YE. These were samples as described previously that were seen to produce MmA (Fig. 40). While they could not be purified by HPLC, they were nevertheless dried and reconstituted in a small sample of water for EMSAs (Chapter 2: Materials & Methods). Unfortunately, as shown in Fig. 42 no binding was observed and in fact, even the controls failed in this experiment. This is shown by the lack of a shift in DNA even in the MmyJ protein and DNA only sample which previously consistently showed binding. This is possibly due to the freshly thawed MmyJ protein having denatured despite being in a buffer containing glycerol and storage at -80 ℃. With additional time this would be repeated with freshly purified MmyJ along with purified MmA

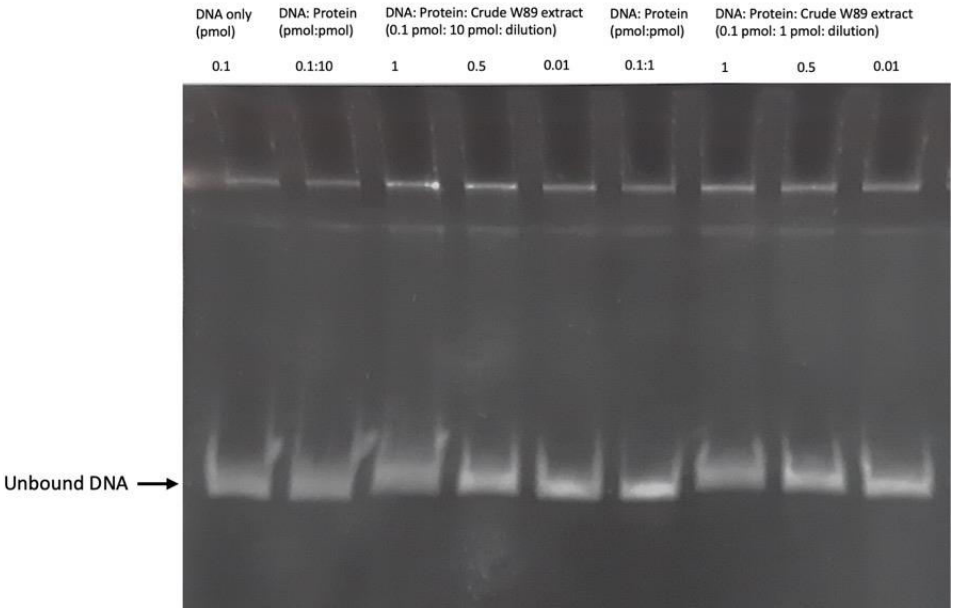


Figure 42. Electromobility shift assays (EMSAs) of MmyJ, DNA and ligand binding. MmyJ has been shown already to bind to DNA (Fig. 22) but as a transcriptional protein it also has a ligand domain to derepress it from DNA binding. This ligand may be a Methylenomycin compound itself. To test this a crude extract of Methylenomycin compounds (Fig. 40) was added in various dilution factors with samples of MmyJ and DNA. This extract contains a mix of Methylenomycin compounds including Methylenomycin A (MmA). No binding of even the MmyJ to DNA is seen however indicating issues with the protein itself. More fresh purified protein will need to be produced to repeat this.

and DNA.

3.8 Future work testing for MmyJ functioning

Bioinformatic analysis of MmyJ indicate MmyJ to be an ArsR family protein[1] (Table 10) that may have evolved from other soil bacteria tolerant to areas of high salt content (Fig. 15). Similar synteny of *mmyJ* and *mmr* genes were also highlighted in a number of actinomycetes (Fig. 16) with many of these appearing to have intergenic regions between the two genes (*mmyJ* and *mmr* equivalent) that have conserved motifs (Fig. 27). Within two of these motifs are the DNA binding sites shown for MmyJ.

DNA binding sites for MmyJ was elucidated from structural analysis of MmyJ and corroborated by EMSAs with DNA and MmyJ (Fig. 26). These are two pseudo-palindromic sites beside one another with one encompassing the -35 promoter region for *mmyJ* and the other encompassing the -35 promoter region for *mmr*. Preliminary data (Fig. 26) indicated twice as much MmyJ protein is required to bind to the -35 promoter region site of *mmr* than of *mmyJ*. This shows MmyJ having a higher affinity and thus tighter control of its own promoter region than of *mmr*. The implication of this in the wider genetic set up of MmyJ functioning is discussed later in Chapter 6.

The finer details of this interaction was attempted to be uncovered using analytical SEC (Fig. 28) however results were inconclusive and worth repeating with higher concentrations of MmyJ and DNA as well as a greater variety in ratio of both. Likewise, the action of MmyJ binding to this site and acting as a transcriptional protein was unsuccessful tested through use of reporter plasmids (Fig. 32).

A reporter system generally utilised the use of a gene encoding for a fluorescent or luminescent protein to visualise the expression levels, as was done with Tinsel Purple and mScarlet reporter systems (Chapter 3.6). However, there were unexpected variations in the colours of colonies which may be linked to the low

expression of the reporter gene. To test for the expression levels of the reporter gene, the mRNA of colonies would need to be extracted and analysed. These reporter systems were designed for *E. coli* as *E. coli* is less GC rich than *Streptomyces* making it easier to perform cloning work, as well as having a much shorter and simpler growth cycle in comparison to *Streptomyces.* However, there are reporter systems available for use within *S. coelicolor*, the most notable one being the luciferase reporter system that produces the bioluminescent protein-luciferase. The *lux* operon for luciferase was optimised for use within *S. coelicolor* by Craney *et al.* (2007)[231] and was used later by Styles K. (2016)[58] for the molecular analysis of MmfR (another transcriptional regulator in the methylenomycin gene cluster- Introduction 1.4.1). This reporter plasmid system was attempted to be utilised by Lougher M. (2015)[155] for MmyJ studies however cloning of the intergenic *mmyJ-mmr* region was unsuccessful and the *E. coli* system was favoured. Therefore, both *E. coli* mScarlet and *S. coelicolor* luciferase reporter systems could be reattempted and optimised in parallel for valuable data of MmyJ binding *in vivo*.

With the two DNA binding sites of MmyJ now known, these reporter plasmid experiments could be reattempted with the two DNA sites cloned into one plasmid each. This could be used to decipher the transcriptional strength of MmyJ to understand whether MmyJ is a strong transcriptional protein and if so if there is any variation in its binding strength between the two DNA regions.

MmA would also need to be purified in order to conduct DNA-MmyJ-MmA assays since using crude extracts yielded ambiguous data. The purification of MmA proved difficult but using a solid R2YE medium for growth of *S. coelicolor* W89 strains (Chapter 2: Materials & Methods) was discovered to be more efficient for MmA production (Fig. 40) and this method could be utilised in the future in conjunction with purification techniques like HPLC. Pure MmA would also be useful for the crystallisation of MmyJ with MmA to hopefully gain a MmyJ-ligand bound structure (discussed further in Chapter 4.5).

Finally, recombinant WT MmyJ and C49S MmyJ when overexpressed and purified showed signs of dimerization (Fig. 20B and Fig. 20C). Crystal structures of C49S MmyJ and WT MmyJ (Chapter 4) were also solved as dimers and this ability of MmyJ to form dimers was validated through bacterial hybrid assays (Fig. 29A). Also using these hybrid assays, leucine 25 was shown to be a key residue for the formation of MmyJ dimers as mutating this residue (L25E) abolished the ability of MmyJ to dimerise (Fig. 29B). All known ArsR family proteins are considered to function as dimers in order for the protein to bind to DNA. L25E MmyJ could therefore be overexpressed, purified and similar DNA binding assays as done on C49S MmyJ could be attempted to determine if MmyJ dimers are necessary for DNA binding.

Work done in the chapter has suggested some key elements in which MmyJ functions but more experiments like the ones proposed would be useful to gain a fuller picture of the action of MmyJ.

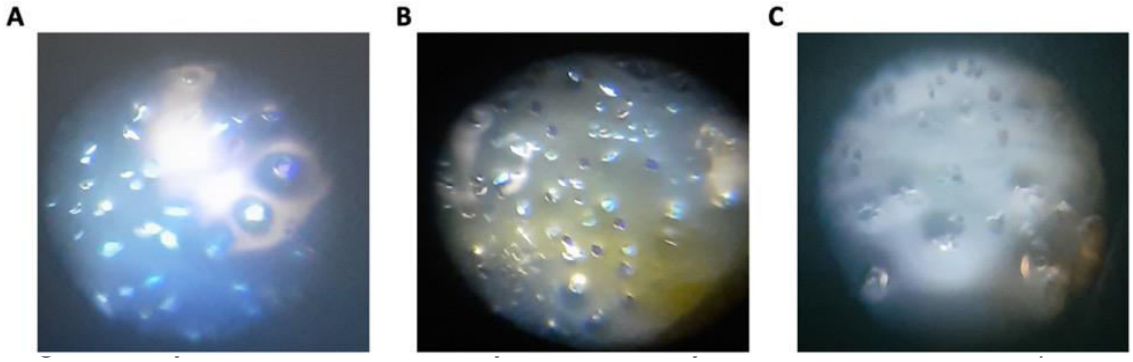
Chapter 4: Solving and analysing the structure of MmyJ

The recombinant MmyJ that was purified in Chapter 3.2 was also used for X-ray crystallography to determine the structure of the protein. This is a well-established and powerful technique that can give atomic level details of the structure of proteins from a crystallised form.

4.1 Crystallisation of MmyJ

After the purification of recombinant WT MmyJ and C49S MmyJ from an *E. coli* host (Chapter 3.2), the proteins were in a buffer containing 20 mM Hepes, 50 mM NaCl, 5% glycerol at pH 7.6 and concentrated between 5-16 mg/mL. The protein at different concentrations were then used to set up various crystallography screens (Chapter 2: Materials & Methods). After a year and a half of incubation at room temperature a few potential protein crystals of WT MmyJ and C49S MmyJ were seen (Fig. 43).

A crystal of C49S MmyJ at 7.5 mg/mL (Fig. 43) was selected for analysis that was incubated in the MIDAS+ screen (Molecular Dimensions) under reservoir conditions containing 0.1 M sodium malonate dibasic monohydrate, 0.1 M Hepes at pH 7.0 and 30% w/v poly(acrylic acid sodium salt) 2100. This crystal was harvested by Dr Crow (Warwick University) in litholoops and plunge-frozen in liquid nitrogen (Fig. 44A) before it was sent to the synchrotron facility (Diamond Light Source, Oxford, UK) for diffraction (Fig. 44B) and analysis.



C49S MmyJ at different concentrations were placed into commercial (Molecular Dimensions 96-HT) X-ray crystallography screens. After ~1 year of incubation these crystals were seen which are believed to be protein crystals due to their lack of cuboidal shape. **A)** Crystals from C49S MmyJ at 7.5 mg/mL in a MIDAS+ screen which was selected for X-ray diffraction. **B)** Crystals from a C49S MmyJ Structure Screen 1+2 at 11.49 mg/mL and **C**) crystals from a Structure Screen 1+2 of MmyJ WT at 5 mg/mL.

4.2 Structure determination of MmyJ

4.2.1 The diffraction of C49S MmyJ

This diffraction pattern or spots from X-rays diffracting off the protein are the measurements used to solve for the structure of the protein crystal. Fig. 44A shows C49S MmyJ crystals produced plenty of strong, distinct diffraction spots which gave a resolution of 1.68 Å. X-ray crystallography is a destructive method because as the crystal rotates it has X-rays beamed at it, so the resolution after a while of X-ray diffracting reduces. To account for this and build a more holistic picture of the structure of the crystal, the multiple diffraction patterns are merged, and the data is trimmed.

Collaborative Computational Project Number 4 (CCP4) is a software suite for the analysis and determination of biological structures by X-ray crystallography. This

suite contains programs designed to assist in solving for crystal structures and was used to determine the structure of MmyJ.

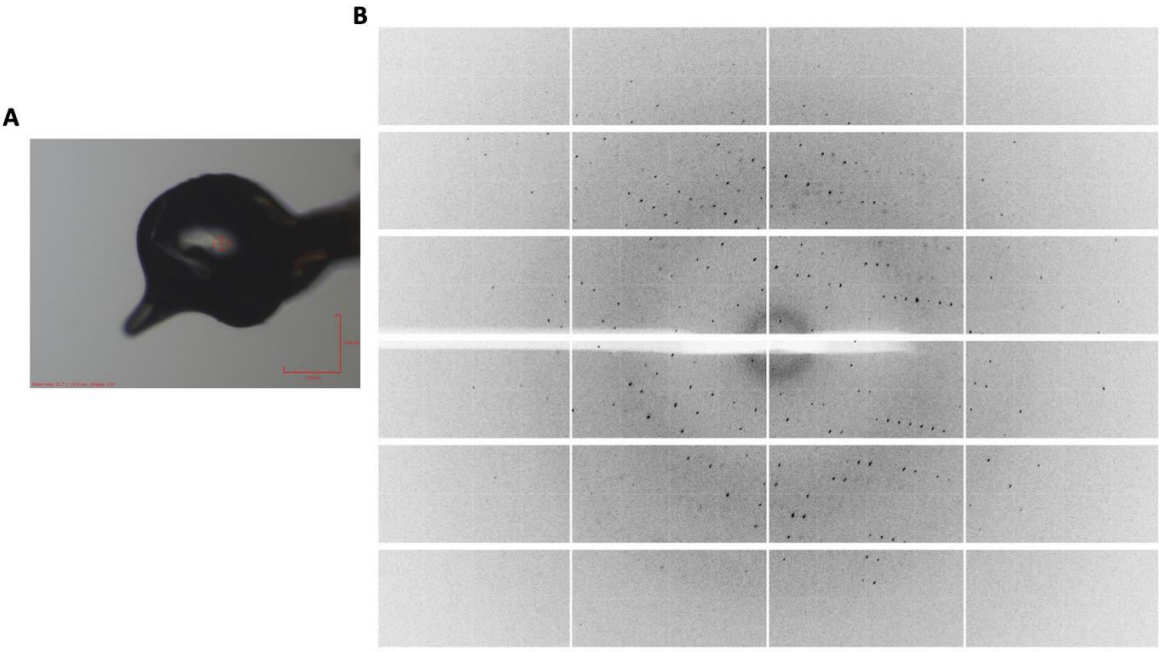


Figure 44. C49S MmyJ crystal used for X-ray diffraction and the diffraction this crystal gave. A) A crystal of C49S MmyJ at 7.5 mg/mL in a MIDAS+ screen was selected for X-ray diffraction. This was harvested by Dr Crow (University of Warwick) in litholoops and plunge-frozen in liquid nitrogen before being sent to the synchrotron facility (Diamond Light Source, Oxford, UK) for diffraction. **B)** The diffraction pattern this crystal (**A**) gave when a beam of X-rays were blasted at it. There are plenty of diffraction spots that are also distinct and strong, indicating good diffraction. This translates to the resolution of the crystal which was at 1.68 Å.

Indexing and integration are the first steps in the processing of X-ray diffracted data.[232,233] Indexing assigns integer numbers to the reflection spots (indexing the spots) which can be used to determine the geometric properties of the crystal (e.g. orientation, unit cell dimensions and symmetry). Integration then uses the intensity of these reflection spots and the background noise to be able to measure even weaker reflections. Both indexing and integration can be done by programmes in the CCP4 suite such as iMosflm[165] however the Diamond facility itself has an autoprocessing pipeline (xia2 system[164]) that uses iMosflm. This autoprocessed file from Diamond of the indexed and integrated X-ray diffraction data was downloaded for further processing using CCP4.

4.2.2 Data merging and scaling of C49S MmyJ

The first programme used on the data is 'Aimless'.[166] This scales and merges the data producing the first statistics of data quality. Scaling was mainly the aim here as it assists in making the data consistent since there can be many factors that make the diffraction intensities measured vary, such as crystal decomposition from radiation, X-ray absorption at different angles, X-ray beam intensity changes etc. Scaling can average these to show the real resolution of the data and highlight any bad regions.

Table 12 shows the overview of the results of the scaled and merged data for C49S MmyJ. As shown the resolution is 1.68 Å. This is generally considered good resolution as the length of a carbon-carbon bond is 1.5 Å and so at 1.68 Å side chains and secondary structure should be able to be confidently resolved.

Table 12: Data merging and scaling results overview. Diffraction data from C49S MmyJ after merging and scaling yielded the following table of parameters. This was done using 'Aimless' from the CCP4 Suite.[164] The key results highlight the resolution of this data was 1.68 Å, the R_{meas} 0.074, the Mean intensity 14.1 and the data shows a completeness of 99.6%. This is generally considered a suitable data set for solving the structure to a confident level accuracy.

	Overall	InnerShell	OuterShell
Low resolution limit	48.97	48.97	1.71
High resolution limit	1.68	9.05	1.68
Rmerge (within I+/I-)	0.068	0.049	0.438
Rmerge (all I+ and I-)	0.069	0.050	0.451
Rmeas (within I+/I-)	0.076	0.054	0.516
Rmeas (all I+ & I-)	0.074	0.052	0.496
Rpim (within I+/I-)	0.034	0.022	0.264
Rpim (all I+ & I-)	0.024	0.017	0.196
Rmerge in top intensity bin	0.043	-	-
Total number of observations	180526	1299	5754
Total number unique	20131	171	963
Mean((I)/sd(I))	14.1	30.6	2.0
Mn(I) half-set correlation CC(1/2)	0.999	0.999	0.912
Completeness	99.6	99.6	95.3
Multiplicity	9.0	7.6	6.0
Mean(Chi^2)	0.75	1.32	0.22
Anomalous completeness	99.0	100.0	89.1
Anomalous multiplicity	4.7	5.8	3.2
DelAnom correlation between half-sets	-0.393	-0.659	-0.118
Mid-Slope of Anom Normal Probability	0.469	-	-

The *R* factors[233,234] are measurements of the internal consistency of the data by measuring the precision of the unique reflections. Generally, data with an R_{merge} of 10% or less is considered good quality and with this data at less than 7% this is sufficient. However, the R_{merge} value can be affected by multiplicity and so the value is inflated as there are more measurements of the same intensity. Multiplicity (also known as redundancy) is an average of the times each independent reflection in the data set is observed and so having a higher multiplicity number is preferred as long as the additional reflections aren't from radiation damage. For this data multiplicity is at 9.0 and normally a multiplicity number of above 3 is considered worthy. The R_{meas} does not use multiplicity in its calculation and so gives a more realistic value of the precision of the data. For this C49S MmyJ data set R_{meas} is 7.4% indicating a high degree of self-consistency within the data and is therefore of good quality. Perhaps for merged data a better R factor to consider in precision is the 'precision-indicating merging' R factor (R_{pim}) which accounts for multiplicity and is a better indicator of the overall quality of the diffraction data. The R_{pim} for this C49S MmyJ crystal was 2.4% (Table 12) highlighting high quality data.

Another important statistic from Table 12 is the 'Mean ((I)/sd(I))' which is the average of the ratio of reflected intensity (I) to its estimated error (sd(I)). This average of signal to noise ratio is 14.2 in this data set and higher than 2 is generally considered good because at around this value of 2, there's little chance of useful data being cut out while maintaining the quality of the reflections. Finally, the completeness of the data can be considered which is the percentage of the reflections in the data set that fall under the specified resolution. The closer this is to 100% the better and for this experiment it was 99.6% for the inner shell and 95.3% for the outer shell.

4.2.3 Molecular Replacement of C49S MmyJ

Once the data has been merged, scaled and the quality checked, the phase problem must be addressed.[233] Fourier transforms are mathematical operations that can relate the diffraction pattern to the scattering pattern. These can be inverted so that the diffraction pattern can reveal the structure of the scattering object i.e., protein crystal. To compute Fourier transforms both the amplitude and the phase of the diffracted X-rays need to be known. The phase is the distance between the amplitude crests and troughs of 2 waves (the shift along the x-axis).[235] As stated earlier, it is the intensity of the diffracted X-rays that is measured (amplitude of waves squared) however in doing so the phase information is lost. Not having this phase information means Fourier transform equations cannot be solved and therefore cannot be inverted to give an electron density of the protein crystal, which is the phase problem.

One method to solve the phase problem is to do molecular replacement (MR).[236] For this a model structure is required. This model needs to have a sample protein which is structurally similar to the protein being modelled and this usually means having a high sequence identity of sample protein to protein under investigation as this will allow for an approximation of the phases. The model then needs to be positioned in the unit cell to find the one that best represents the observed diffraction pattern. Optimally positioning the known structure as closely to the unknown structure in principle requires a 6D search for both the correction orientation (3 angles) and translation (3 dimensions). This orientation issue is resolved using another mathematical function known as the Patterson function that generates Patterson maps for intensities. When the Patterson map of the diffracted data correlates to that of the model data in a certain rotation and translation, the correct position and orientation of the sample protein crystal becomes known.[233] This is known as the MR 'solution' as phases calculated from the model then serve as the first approximation of the phases of the unknown structure.

In order to solve the structure of C49S MmyJ, a computational prediction of C49S MmyJ was used as a search probe in the MR search. This search probe of C49S MmyJ was created using AlphaFold 2[167] by Dr Crow and the CCP4 program 'phaser' used for MR.[168] Calculating the correct MR solution allowed for the unambiguous determination of the space group and composition of the unit cell for C49S MmyJ (Table 13).

Table 13: Unit cell composition of X-ray diffraction data of C49S MmyJ. After molecular replacement of the diffracted C49S MmyJ merged and scaled data set (at 1.68 Å resolution), the composition of the unit cell is shown. Molecular replacement was done using 'phaser' in the CCP4 Suite[166] with an AlphaFold 2 model of C49S MmyJ to solve the phase problem. A unit cell composing of 1 or 2 (Z value in table) molecules of C49S MmyJ was tested. A dimer (2 molecules) of C49S MmyJ is most probably considered occurring in each unit cell.

Z	MW	VM	% solvent	rel. freq.	
1	12097	3.52	65.05	0.388	
2	24194	1.76	30.10	1.000	<== most probable

The space group for this data was determined as P 2_1 2_1 2_1 (space group 19) in a unit cell of the following measurements: 33.11 Å, 59.37 Å, 86.64 Å. This space group is orthorhombic, meaning all the axes are perpendicular to one another but are unequal in length.

The unit cell composition (Table 13) indicates the protein number and solvent percentage in a unit cell and the most probable of these in the data set. A monomer of C49S MmyJ (1 in Table 13) and a dimer (2 in Table 13) were both tested to see which is most probable. Results indicate each unit cell likely contains a homodimer of C49S MmyJ equalling a molecular weight of 24194 Daltons in 30.1% solvent.

4.2.4 Refining the structure of C49S MmyJ

Once initial phasing was completed using MR, the model was iteratively built using Coot[169] and Refmac5[170,234,237,238] in CCP4. The high resolution of the data is indicated on the electron density map (Fig. 39). At even higher sigma (r.m.s.d.) values of 2.09 of the x-ray scattering electrons maps, the side chains and backbone could be easily visualised (Fig. 45B).

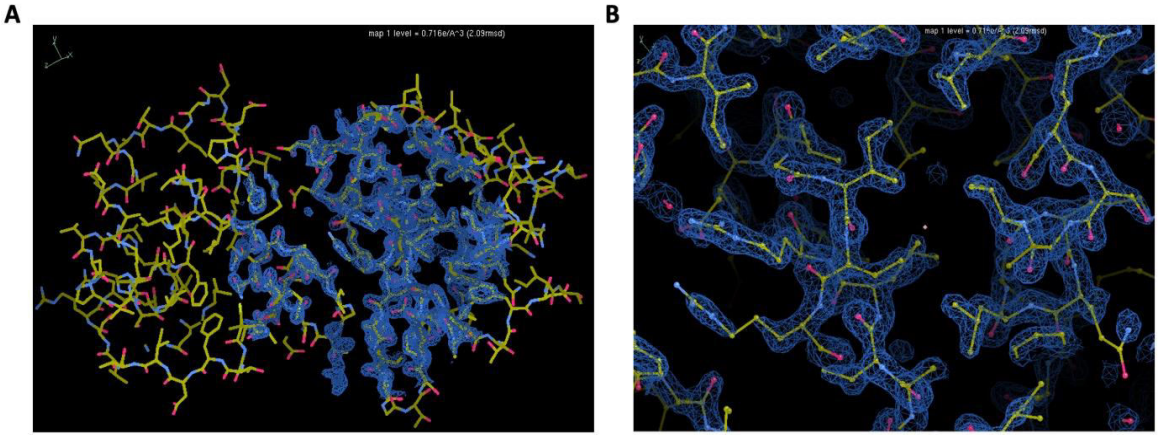
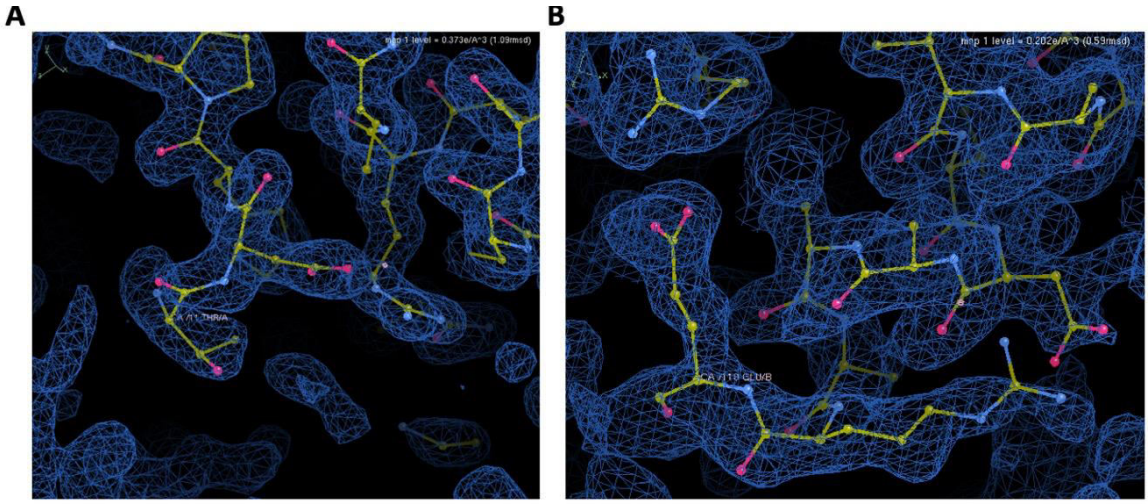


Figure 45. Resolution of C49S MmyJ X-ray diffracted data. A C49S MmyJ homodimer model was built using Coot[167] and Refmac5[233,236,237]. The structure of C49S MmyJ was fitted to a 1.68 Å electron density map. Higher contour levels (set by rmsd of the map) gives better confidence of the model. In yellow are the amino acids of the C49S MmyJ with the blue map being a 2mfofc map contoured at 2.09 sigma. **A** being an overall view of this and **B** a close-up.

This was true for most of the dimer however for each monomer (chain A and B), at the ends and starts of the sequence, resolution was weaker and so could not be built confidently meaning some residues at the C and N terminals were omitted (Fig. 46). The final model of C49S MmyJ contains two molecules per asymmetric unit and is of excellent overall quality. Each of these molecules are 111 amino acids in length. For chain A and B, the model was built from amino acids 11-110 meaning 10 residues at the start were omitted and 1 at the end. While this is not ideal the exclusion of these few residues were not considered vital to the overall structure of the homodimer and the overall secondary structure can still be visualised accurately.

Figure 46. Resolution of C49S MmyJ X-ray diffracted data at N and C terminal ends of chains. While the resolution for the majority of the C49S MmyJ model was good the quality dips at the N (**A**) and C-terminal (**B**) of each monomer of the dimeric molecule. Lower contour levels can be used to visualise more mobile parts of the structure. The map in blue is that of a 2mfofc map which was lowered to 1.09 sigma and 0.59 sigma values for visualising the N-terminal of Chain A (**A**) and C-terminal of Chain B (**B**) respectively. This meant that residues at the start and ends of each monomer had to be omitted to ensure an accurate model is built.



In addition to refinement of the C49S MmyJ homodimer, 83 crystallographically-ordered water molecules and one chloride ion were also identified. These waters may also play a role in the functioning of the protein by forming hydrogen bonds with polar residues. Most of the waters added were to the outside of the protein around areas of hydrophilicity however there were two within the structure (Fig. 47A, Fig. 47B) which may be relevant to functioning. A chloride ion was also present near the end of chain A though this was likely from the salts the protein was

crystallised in and did not appear to influence the structure of C49S MmyJ (Fig. 47C).

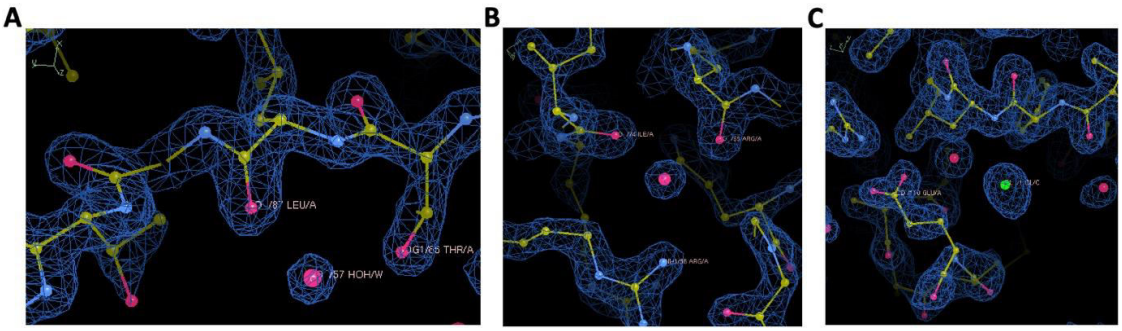


Figure 47. Solvents were built into the C49S MmyJ model which may interact with protein side chains. Additions of solvents into building the model of a protein is important not only to improve its refinement values but also to allude to any bonds occurring between solvents and side chains. Waters in particular may be forming hydrogen bonds with side chains and may be important in the functioning of the protein. Only 2 waters built (**A, B**) were shown as possibly being able to interact with MmyJ side chains. A water may interact with backbone and side chains atoms of L87 and T86 (**A**) as well as I74, R89 and R98 (**B**). Also detected was a chloride atom (**C**) though this is likely too far from the protein to cause any interactions.

To determine the quality of the refinement, R values are used. The R factor and R free values give an idea as to how well the simulated diffraction pattern calculated from the model fits to the experimentally observed diffraction pattern. The lower these values the better the fit between simulated and experimental data and thus the more accurate the refined structure is. A perfect fit would mean R values of 0 though typically these values are close to 0.20. Aiming for the lowest R factor value could introduce biases in terms of overfitting or excluding data, which can be problematic. The R free is therefore a less bias measurement of quality. The calculation for the R free is the same as the R factor however for R free values, a subset of experimental data (4.93%) is set aside, and refinement is done on the rest (95.07%). After refinement is done on the majority of the data that was not set aside, this is used to calculate how well it predicts the set aside data. An ideal model would have the R factor and R free values be the same though the R free value tends to be higher.[239]

For C49S MmyJ after refinement the R factor was 0.1882 and R free was 0.2226 (Table 14). To compare these values to structures in the PDB database at similar resolutions, the number of atoms used in the refinement (N_a) and the number of reflections (*f*) are also needed, so that a ratio of the R free and R factor can be plot against the N_a/f.[240,241] For C49S MmyJ the N_a was 1644 and the *f* was 19087. Fig. 42 shows these values of C49S MmyJ plotted with other PDB structures at various resolutions. As seen in the figure, with the resolution of C49S MmyJ being 1.68 Å the R factor and R free values of this structure fall under similar values for

Table 14: Final refinement R values of C49S MmyJ. After numerous rounds of refinement and validation, the final refinement R values are displayed below. Both R factor and R free values are around 0.2 which is typical of models built in the PDB database at around 2 Å resolution.

	Initial	Final
R factor	0.1882	0.1882
R free	0.2225	0.2226
Rms BondLength	0.0119	0.0119
Rms BondAngle	1.6250	1.6213
Rms ChirVolume	0.0790	0.0793

structures in the PDB database that are also around 1.5-2.0 Å resolution (Fig. 48). This would mean the refinement created considering the resolution of the data would be good enough to be deposited into the protein databank.

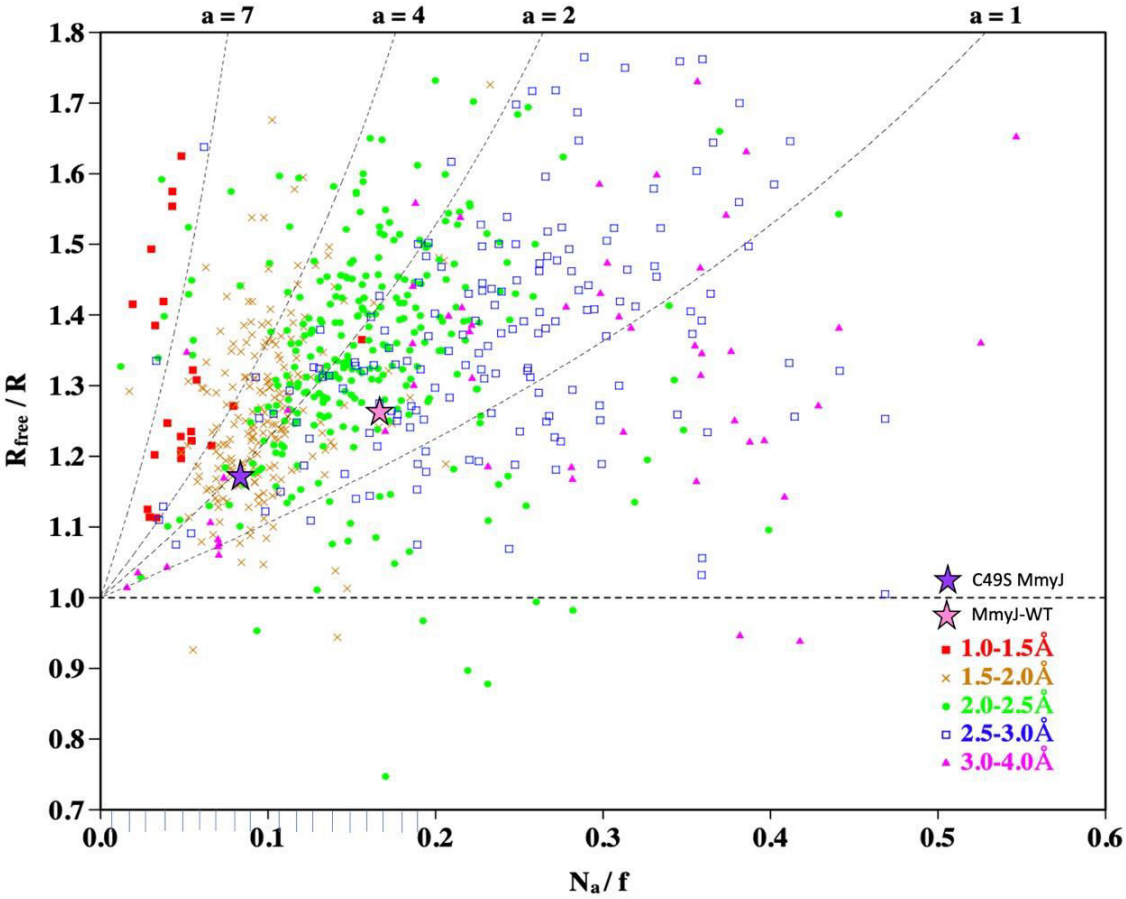


Figure 48. A graph of R values and resolution of solved MmyJ structures and published structures in the Protein Databank. The R free/ R factor values are plotted as a ratio against the number of atoms included in the refinement (N_a) and the number of reflections used (*f*). C49S MmyJ data (purple star) was at a resolution of 1.68 Å and refined until R free was at 0.226, R factor at 0.1882, N_a at 1644 and *f* at 19087. The WT-MmyJ (pink star) data was at a resolution of 2.12 Å and refined until R free was at 0.2276, R factor at 0.1774, N_a at 1665 and *f* at 9879. These R values align with structures at similar resolutions published in the protein databank. C49S and WT MmyJ values are plotted on this graph by Tickle *et al.* (1998 and 2000).[239,240]

4.2.5 Validation of the constructed C49S MmyJ structure

The R values are one way of checking the quality of the model built however there are also other validation and quality check measures available.[172–174,233] These tools and measurements can highlight areas of the model that may need to be inspected. However, all these measurements are based on other high-resolution structures

and what is considered the norm, therefore some areas of concern flagged may not necessarily be unusual for the structure of MmyJ.

One important validation tool is the Ramachandran plot.[172] This plots the phi (ϕ) and psi (ψ) angles of each amino acid in a protein sequence. Each residue has favoured regions and any outliers can be shown where further refinement may be needed. Prolines and glycines are given special allowances as proline is a 5 membered ring and so far more restricted in its angles meaning it has fewer areas on the Ramachandran plot it is allowed in.[242] Glycine on the other hand only has hydrogen as a side chain meaning it has far fewer restrictions and can occupy angles that cover a much greater area of the Ramachandran plot.[242]

While Coot has an inbuilt Ramachandran plot measurement, CCP4 was used instead with the program 'Rampage'.[171] There is little difference in the two but Rampage is generally considered a more accurate representation of the possible phi and psi angles of residues (Fig. 49). As shown the model has no outliers and all but one amino acid (Asp53) is in favourable positions though even this residue is in an allowed region.

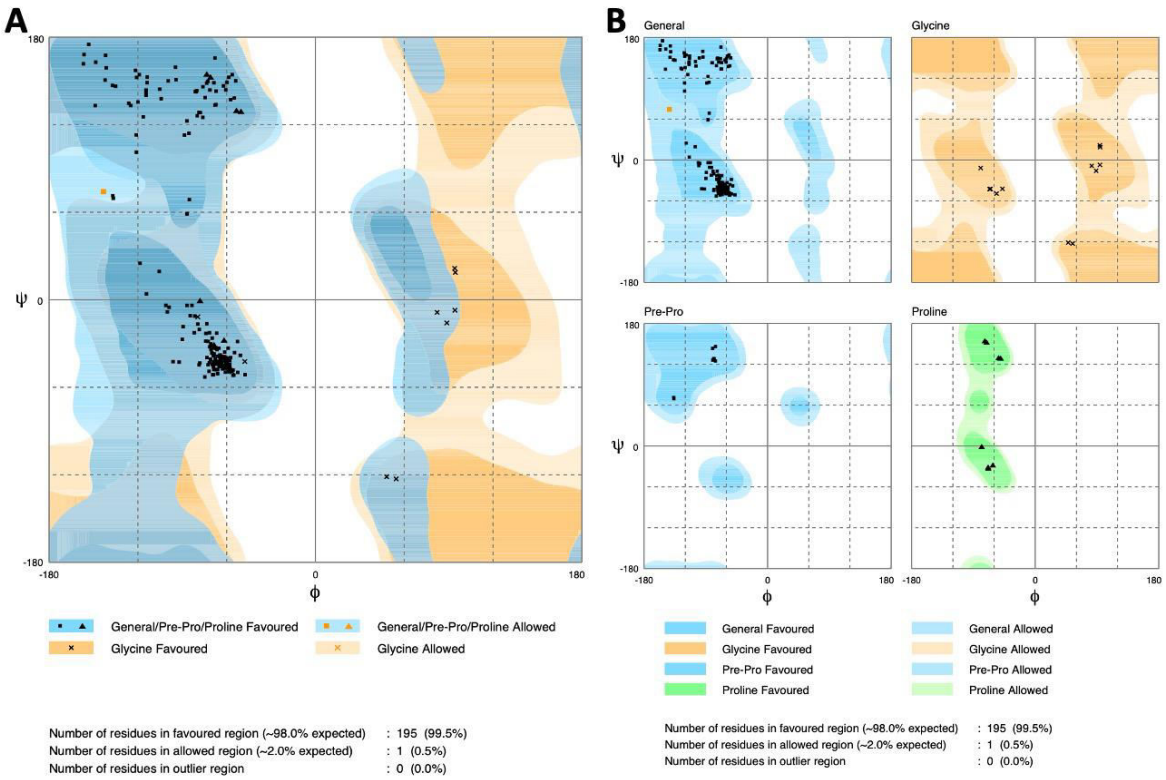


Figure 49. Ramachandran plots for validation of residues in the C49S MmyJ model. The final model after refinement was validated using Ramachandran plots. The coloured areas in maps show torsion angles of residues which are permitted and favorable to identify outliers which may be compromising the quality of the structure model. No outliers are shown for the model built of C49S MmyJ (**A**) though one was considered not in the most favorable position (**B**). This residue is not a pre-proline, proline or glycine residue which all have unique areas to be allowed in (**B**) but rather it is a more general amino acid. No outliers in the Ramachandran plot of C49S MmyJ show a valid model of the structure has likely been built.

Procheck[173] is a suite of programs that can assess the stereochemistry of the structure as a quality check. This tool can highlight any areas that may need refining further such as any atoms of a residue in too close contact with each other. Residues flagged by validation tools such as Procheck highlight 'unusual' or 'unlikely' structural features that should be inspected and rebuilt if not supported by unambiguous electron density. For C49S MmyJ, Procheck results (Fig. 50) did not find anything too unusual but did highlight the planes and bond lengths of a tiny fraction of the structure being outside the expected norms. When these were inspected and re-refined, they consistently reappeared. Given the high quality of

the data used for refinement and the clarity of the electron density map no further changes were made. There were also no close contacts ('bad contacts') found in this model by Procheck indicating all residues are forming bonds that are not clashing. The results from Procheck indicate the stereochemistry is overall very good for this structure model of C49S MmyJ.

```
 +----------<<<  P  R  O  C  H  E  C  K     S  U  M  M  A  R  Y  >>>----------+
 |                                                                             |
 |  /Users/u1408606/Documents/CCP4work/mmyJ_edit68.pdb   1.7      284 residues  |
 |                                                                             |
 |  Ramachandran plot:   94.3% core    5.7% allow    0.0% gener    0.0% disall |
 |                                                                             |
 |  All Ramachandrans:    0 labelled residues (out of 196)                     |
 |  Chi1-chi2 plots:      0 labelled residues (out of 112)                     |
 |                                                                             |
 |  Main-chain params:    6 better     0 inside      0 worse                   |
 |  Side-chain params:    5 better     0 inside      0 worse                   |
 |                                                                             |
+|  Residue properties: Max.deviation:     4.3             Bad contacts:    0  |
+|                      Bond len/angle:    4.7    Morris et al class:  1  1  2 |
 |                                                                             |
 |  G-factors           Dihedrals:  -0.01  Covalent:   0.04    Overall:   0.01 |
 |                                                                             |
 |  M/c bond lengths: 98.9% within limits   1.1% highlighted                   |
 |  M/c bond angles:  96.2% within limits   3.8% highlighted                   |
+|  Planar groups:    94.3% within limits   5.7% highlighted                   |
 |                                                                             |
 +-----------------------------------------------------------------------------+
   + May be worth investigating further.  * Worth investigating further.
```

Figure 50. PROCHECK results summary for validation of the C49S MmyJ model. Procheck[171] is a validation tool that can assess the quality of the model of a structure. Procheck was used to validate for C49S MmyJ. No areas were flagged as needing to be further investigated though a few were suggested. The bond angles and lengths for a few residues were flagged as something to check but after reviewing the structure, no further improvements could be made in those areas.

Another analysis tool for the checking of the stereochemistry of a structure is the use of the web-server MolProbity.[174,243] Similar to Procheck, MolProbity also gives an output of the geometry of the residues in the protein structure and highlights any areas for attention. When the refined C49S MmyJ structure was run through MolProbity (Fig. 51), three main areas were flagged showing some residues that had bad bonds (bond length), bad bond angles and some poor rotamers. These 'bad' bonds were shown also in the Procheck result (Fig. 50) and was unable to be improved upon. Looking at the number of actual residues this applies to (1-3) is very low and therefore is unlikely to be too significant to the overall structures. Similarly, the number of residues stated to be poor rotamers (2) is low and were

not able to be 'fixed'. All this is also taken into consideration to generate the MolProbity score. This is calculated as a log-weighted number from the clash-score, unfavoured Ramachandran percentage and the percentage of bad side-chain rotamers which gives one overall number. This number should be lower than the resolution of the structure for the structure to be considered average to good quality. The percentile score for comparison is taken from an average of PDB structures ± 0.25 Å of the resolution of the structure given and considering the resolution of C49S MmyJ is at 1.68 Å means there would be plenty of PDB structures in this range for comparison. The MolProbity score (Fig. 51) for the C49S MmyJ structure built was 1.20 which is lower than its resolution and puts it in the 98th percentile highlighting the structure built of C49S MmyJ is valid.

Summary statistics

All-Atom Contacts	Clashscore, all atoms:	3.52		98th percentile* (N=792, 1.68Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	2	1.20%	Goal: <0.3%
	Favored rotamers	154	92.22%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	195	99.49%	Goal: >98%
	Rama distribution Z-score	-0.15 ± 0.50		Goal: abs(Z score) < 2
	MolProbity score^	1.20		98th percentile* (N=8853, 1.68Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	1 / 1585	0.06%	Goal: 0%
	Bad angles:	3 / 2158	0.14%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 8	0.00%	Expected: ≤1 per chain, or ≤5%
Additional validations	Chiral volume outliers	0/259		
	Waters with clashes	10/83	12.05%	See UnDowser table for details

In the two column results, the left column gives the raw count, right column gives the percentage.
* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.
^ MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.
Key to table colors and cutoffs here:

Figure 51. MolProbity results summary for validation of the C49S MmyJ model. MolProbity[172,242] is a validation tool similar to Procheck that can assess the quality of the model of a structure. Comparable to Procheck results (Fig. 50), bad bonds and angles from poor rotamers were flagged. Again, these could not be resolved in the model through additional refinement, but it must be noted in the overall structure of a homodimeric MmyJ only a few of these residues were flagged.

4.2.6 Structure determination of recombinant wild type MmyJ

A crystal of WT MmyJ was previously diffracted by Lougher M. (2014)[155] and the diffraction data processed by Professor Vilmos Fulop but the structure could not be solved by MR or other strategies. With the structure of C49S MmyJ being solved, it could be used a basis for MR to solve WT MmyJ. The same strategy used to solve C49S MmyJ was employed to solve WT MmyJ and a structure of this was generated.

Table 15: WT-MmyJ X-ray diffraction statistics from a merged, processed data file.
Diffraction data from a WT-MmyJ that was collected by Lougher M. (2015)[155] and processed by Professor Vilmos Fulop prior to the start of this project. The key results highlight the resolution of this data was 2.11 Å, the R_{meas} 0.137, the Mean intensity 20.7 and the data shows a completeness of 99.5%. This is generally considered a suitable data set for solving the structure to a confident level accuracy. This data could not previously be used to solve for the WT-MmyJ structure due to issues with phase problem. With the structure of C49S MmyJ determined, this data set could be used, and the phase problem solved.

Low resolution limit	33.34	33.34	2.23
High resolution limit	2.11	6.69	2.11
Rmerge	0.132	0.031	0.621
Rmerge in top intensity bin	0.034	-	-
Rmeas (within I+/I-)	0.137	0.032	0.648
Rmeas (all I+ & I-)	0.137	0.032	0.648
Rpim (within I+/I-)	0.037	0.009	0.181
Rpim (all I+ & I-)	0.037	0.009	0.181
Fractional partial bias	0.000	0.000	0.000
Total number of observations	135540	4515	17389
Total number unique	10369	390	1425
Mean((I)/sd(I))	20.7	60.4	5.2
Completeness	99.5	99.3	96.4
Multiplicity	13.1	11.6	12.2
Anomalous completeness	99.2	100.0	94.9
Anomalous multiplicity	7.1	7.4	6.6
DelAnom correlation between half-sets	-0.048	0.182	-0.097
Mid-Slope of Anom Normal Probability	0.987	-	-

The merged WT MmyJ data file was at 2.12 Å resolution in a unit cell of 49.15 Å, 90.77 Å, 38.63 Å and had a space group of P 2_1 2_1 2 (space group 18) which is still orthorhombic.[155] The difference in space group and cell unit between this WT MmyJ

crystal and the C49S MmyJ crystal is due to them being different crystal forms. X-ray data collection statistics are shown in Table 15. This data was supplied by Professor Vilmos Fulop who had previously collected and processed this data using an in-house source prior to the start of this project.

WT MmyJ also was likely crystallised as a homodimer similar to C49S MmyJ. Refinement was done but again the chain length had to be trimmed due to low resolution. Chain A was built from residues 7-111 and chain B 11-111, meaning no residues at the C-terminal were omitted though a few at the N-terminal for both chains were. Once the structure was refined and waters added, the final R factor values were calculated to assess the quality of the model.

The R factor value was 0.1774 and the R free value was 0.2276 (Table 16), giving a difference of 0.0502 meaning this is acceptable considering the typical values found in protein structure databases. Also plotting these values with the number of atoms used in the refinement (1665) and the number of reflections used for refinement (9879) and comparing to other PDB database structures (Fig. 48) shows that these R values are as expected for other PBD database structures with resolutions at 2.0-2.5 Å.

Table 16: Final refinement R values of WT-MmyJ. After numerous rounds of refinement and validation, the final refinement R values are displayed below. Both R factor and R free values are around 0.2 which is typical of models built in the PDB database at around 2 Å resolution.

	Initial	Final
R factor	0.1776	0.1774
R free	0.2269	0.2276
Rms BondLength	0.0078	0.0078
Rms BondAngle	1.5137	1.5130
Rms ChirVolume	0.0693	0.0696

A final Ramachandran plot was created for this model of the WT MmyJ after refinement using Rampage[171] which found all amino acids to be within acceptable regions of the plot (Fig. 52). A few were flagged (Ile10, Arg88, Ala80 and Ser57) as being in allowed but not favoured regions but attempts to refine this made little difference so were left as is. A summary of Procheck[173] (Fig. 53) did show two

residues (Ile10 and Ala80) to be outliers of its Ramachandran plot however due to MolProbity[174,243] (Fig. 54) not indicating this to be the case along with Rampage (Fig. 52), this was noted but not considered further. Apart from that, all other Procheck results appeared satisfactory. The MolProbity results (Fig. 53) did highlight a few poor rotamers like for C49S MmyJ (Fig. 52), but this was not further investigated as all other results were fine and these unfavoured rotamers may just be a part of the structure of MmyJ. The overall MolProbity score was 1.39 which is in the 99th percentile and indicates a high-quality structure.

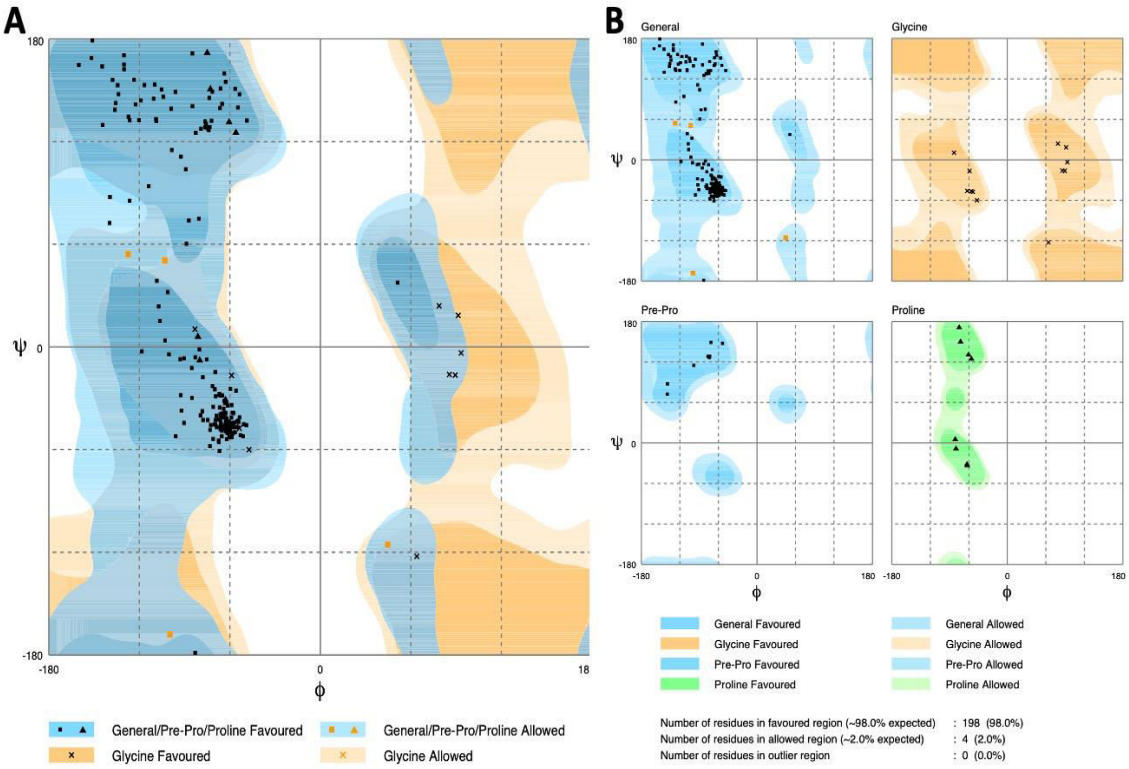


Figure 52. Ramachandran plots for validation of residues in the WT MmyJ model. The final model after refinement was validated using Ramachandran plots. The coloured areas in maps show torsion angles of residues which are permitted and favorable to identify outliers which may be compromising the quality of the structure model. No outliers are shown for the model built of WT-MmyJ (**A**) though a few were considered not in the most favorable position (**B**). These residues aren't pre-proline, proline or glycine residues which all have unique areas to be allowed in (**B**) but rather it is more general amino acids which could not be improved upon further. No outliers in the Ramachandran plot of WT-MmyJ show a valid model of the structure has likely been built.

```
 +----------<<<  P  R  O  C  H  E  C  K     S  U  M  M  A  R  Y  >>>----------+
 |                                                                             |
 |  /Users/u1408606/Documents/MmyJ_ML/MmyJML_edit24_refmac   2.1  294 residues |
 |                                                                             |
*|  Ramachandran plot:   91.2% core    8.2% allow    0.0% gener    0.5% disall |
 |                                                                             |
+|  All Ramachandrans:    2 labelled residues (out of 202)                     |
+|  Chi1-chi2 plots:      1 labelled residues (out of 115)                     |
 |                                                                             |
 |  Main-chain params:    6 better     0 inside      0 worse                  |
 |  Side-chain params:    5 better     0 inside      0 worse                  |
 |                                                                             |
+|  Residue properties: Max.deviation:     4.8             Bad contacts:    0  |
+|                      Bond len/angle:    3.9    Morris et al class:  1  2  2 |
 |                                                                             |
 |  G-factors           Dihedrals:  -0.12  Covalent:  -0.05    Overall:  -0.07 |
 |                                                                             |
 |  M/c bond lengths: 99.6% within limits   0.4% highlighted                   |
 |  M/c bond angles:  94.8% within limits   5.2% highlighted                   |
 |  Planar groups:   100.0% within limits   0.0% highlighted                   |
 |                                                                             |
 +-----------------------------------------------------------------------------+
   + May be worth investigating further.  * Worth investigating further.
```

Figure 53. PROCHECK results summary for validation of the WT MmyJ model. Procheck[171] is a validation tool that can assess the quality of the model of a structure. Procheck was used to validate for WT MmyJ. Only the Ramachandran plot was flagged as needing to be further investigated, which according to the newer program Rampage[169] (Fig. 52) did not show any outliers. A few other areas were also flagged for further review. The bond angles and lengths for a few residues were flagged as something to check but after reviewing the structure, no further improvements could be made in those areas either.

Summary statistics

All-Atom Contacts	Clashscore, all atoms:	3.16		99th percentile* (N=555, 2.12Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	4	2.37%	Goal: <0.3%
	Favored rotamers	155	91.72%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	198	98.02%	Goal: >98%
	Rama distribution Z-score	-1.49 ± 0.49		Goal: abs(Z score) < 2
	MolProbity score^	1.39		99th percentile* (N=11300, 2.12Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 1599	0.00%	Goal: 0%
	Bad angles:	1 / 2172	0.05%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 8	0.00%	Expected: ≤1 per chain, or ≤5%
Additional validations	Chiral volume outliers	0/263		
	Waters with clashes	3/87	3.45%	See UnDowser table for details

In the two column results, the left column gives the raw count, right column gives the percentage.

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.

^ MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.

Key to table colors and cutoffs here:

Figure 54. MolProbity results summary for validation of the WT MmyJ model. MolProbity[172,242] is a validation tool similar to Procheck that can assess the quality of the model of a structure. Unlike Procheck (Fig. 53) but similar to Rampage (Fig. 52) no outliers were spotted in Ramachandran plots. The only area in MolProbity highlighted for review were poor rotamers. Again, these could not be resolved in the model through additional refinement, but it must be noted in the overall structure of a homodimeric MmyJ this was the only area flagged.

4.3 Structure of MmyJ

4.3.1 Overall fold of MmyJ

The structure of C49S MmyJ was visualised using Pymol.[244] A ribbon structure of the secondary elements of the two homodimers of MmyJ is shown in Fig. 55A.

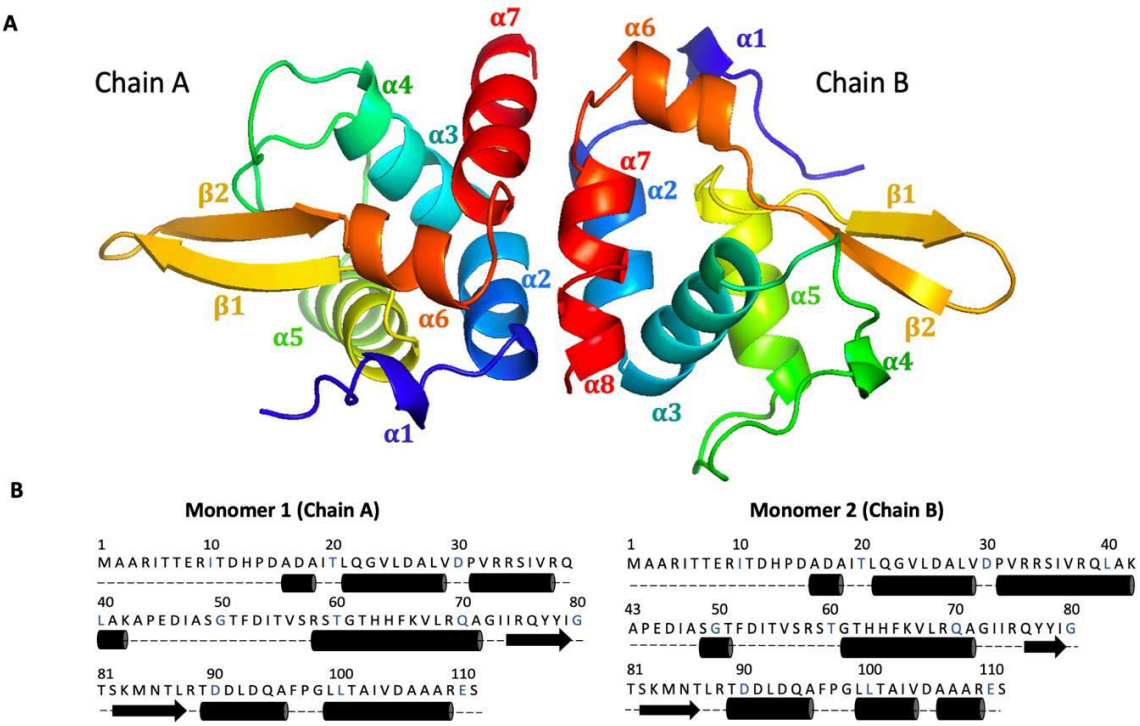


Figure 55. The structure of MmyJ. A ribbon model of the secondary structure of C49S MmyJ which is formed of 2 dimers (Chain A and Chain B) in rainbow format (**A**). The two monomers are asymmetric and have slight variances in secondary structure (**B**). Alpha helices 1, 2, 6 and beta strand 2 are all formed by the same residues in both monomers. Alpha 7 and 8 in the 'Monomer 2' are the same as alpha helix 7 in 'Monomer 1' apart from the Asp105 residue which doesn't contribute to the secondary structure. Alpha 4 has the most variance and influences the length of the 2 helices before and after it.

Each monomer is formed of 7 or 8 alpha (α) helices and 2 beta (β) strands as labelled in Fig. 55A. Both the monomers are similar to one another with some slight differences to create a pseudo-symmetrical dimer (Fig. 55B). The first 3 α-helices encompass the same residues in both monomers with α4 being the same length in

both monomers but occurring either closer or further from α3. The positioning of this α4 helix along the long stretch between α3 and α5 is the main difference between the two monomers. In the monomer with α4 further apart from α3, the α5 helix is also marginally shorter (by 2 residues). The first β-sheet in this shorter α5 monomer is also shorter than the other monomer by 2 residues, whereas the 2nd antiparallel β-sheet is the same length occurring in the same place for both monomers. Also, the same for both monomers is α6 and in the monomer with a shorter α5 helix and β-sheet the α7 helix is split in two. This split occurs with Asp105 not being part of the secondary structure in the monomer with α7 and α8, otherwise this monomer would have α7 and α8 as one helix which would be identical to α7 of the other monomer.

Alpha 1, 4 (and 8) have been labelled tentatively as alpha structures as they are very short loops of only a few residues that may or may not be relevant to functioning. Most ArsR family proteins consist of 5 α helices and 1 β sheet (in the order of: ααααββα) but MmyJ's secondary structure is arranged in the following order: αααααββαα(α). Therefore, MmyJ has ± 1 α helix (depending on the inclusion or exclusion of α1 and α4) in comparison to most ArsR family proteins but the ββ hairpin structure is present. The monomers are separate and distinct with no elements crossing over meaning dimer interactions are only formed from α2 and α7 of each chain. The alpha chains mean that there are helix-turn-helix structures in MmyJ to use for DNA binding and the β sheets at the sides likely help stabilise this interaction as they do for other ArsR proteins[122,200]. Predictions in Uniprot and by previous bodies of work[155] suggested this HTH of MmyJ to be between residues 13-107 which would encompass almost all of the protein's secondary structure. However, with the structure of MmyJ now solved, it is apparent the HTH motif likely involved in DNA binding is between residues 31-71 (α2- α5) and it is this α5 (residues 61-71) that may be binding to DNA (this is discussed further in Chapter 4.4.3).

4.3.2 There are no significant structural differences between C49S and wild type MmyJ

The only differences between C49S MmyJ and WT MmyJ proteins used was the mutation of residue 49 from cysteine to serine in C49S MmyJ. This is the only cysteine present in the MmyJ sequence and was thought to be capable of forming disulphide bonds with the cysteine of another MmyJ monomer. It was therefore originally mutated in case it might be capable of forming non-physiological disulphide bonds outside of the reducing environment of the cell. Should this be the case, the structure of C49S MmyJ and WT MmyJ should show differences.

This does not appear to be the case though as superimposing both models (Fig. 56) where C49S MmyJ is in cyan and WT MmyJ is in green, shows the two are highly structurally similar. Alignment of the two was done using Pymol and had an RMSD value of 0.457 Å. Furthermore residue 49 (labelled in Fig. 56 as Ser49 on the C49S MmyJ) is found at α4 and does not influence the structure here in this loop. It is possible that while this cysteine does not influence the structure it may be involved in ligand or DNA interactions for the functioning of the protein. As both are structurally similar, C49S MmyJ was solved to a higher resolution and so was used for the rest the analysis (Chapter 4.5).

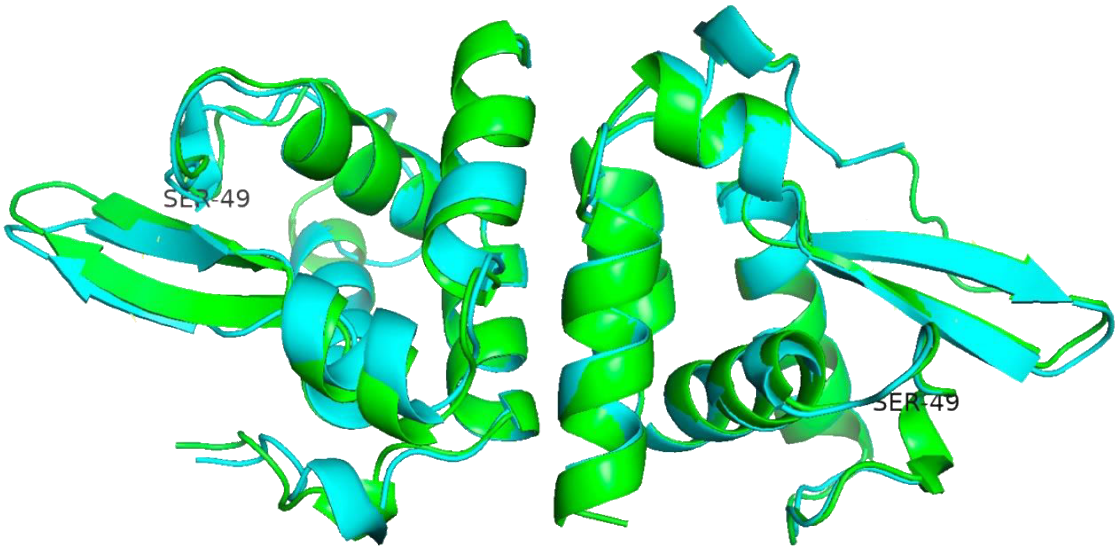


Figure 56. The structures of WT-MmyJ and C49S MmyJ overlayed over each other. Both solved structures of MmyJ (green) and C49S MmyJ (cyan) were superimposed (RMSD was 0.457 Å) to identify any structural differences that may occur from mutating the only cysteine in the sequence. No apparent differences exist, both monomers are still asymmetric and have the same differences in secondary structure.

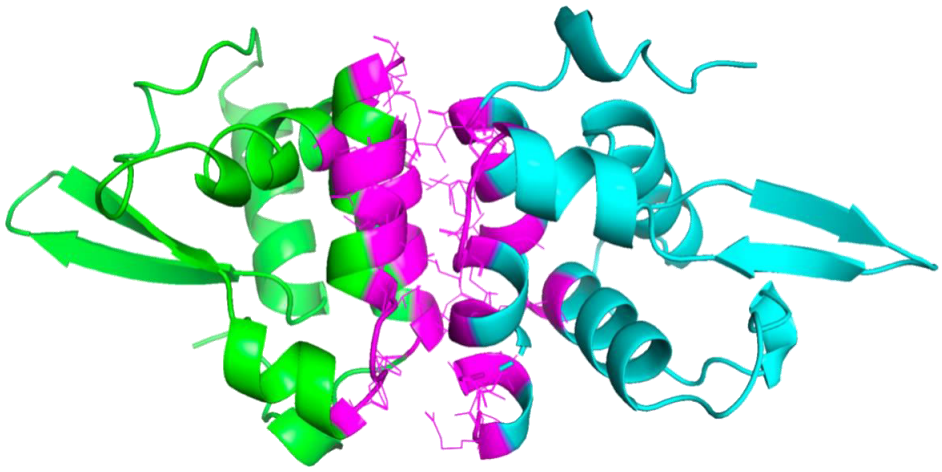
4.3.3 The dimer interface of MmyJ

The dimer interface of MmyJ appears to consist of 2 α helices (α2 and α7) top and bottom of each monomer. The specific residue interactions were modelled (Fig. 57) so those in the interface that have a dASA (difference in accessible surface area) between the monomers is over 1 Å squared. As expected, most of these interacting residues are from α2 and α7 but also a few residues from α3, one from α2 and a few others in the loop between the β sheet and α7 (Table 17). These residues were also shown by PDBePISA[245] results.

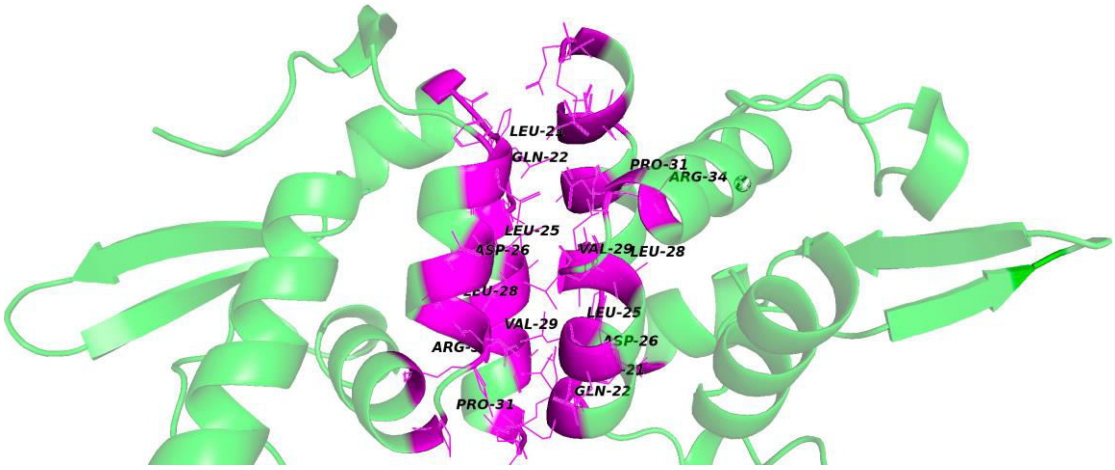
PDBePISA was used to look at the dimer interface of MmyJ and any bonding occurring there. Results showed no disulphide or covalent bonds and no salt bridges occurring but some hydrogen bonding (Fig. 58). It is the same 5 residues from each monomer that is involved in hydrogen bonding and these are; Gln22, Val29, Arg34, Arg109 and Ala95. Gln22 found at the start of α2 appears to hydrogen

bond with both Arg34 and Val29. These residues may be important in the dimer formation of MmyJ, and further work would need to be done to evaluate the importance of these residues for dimer formation.

A



B



C

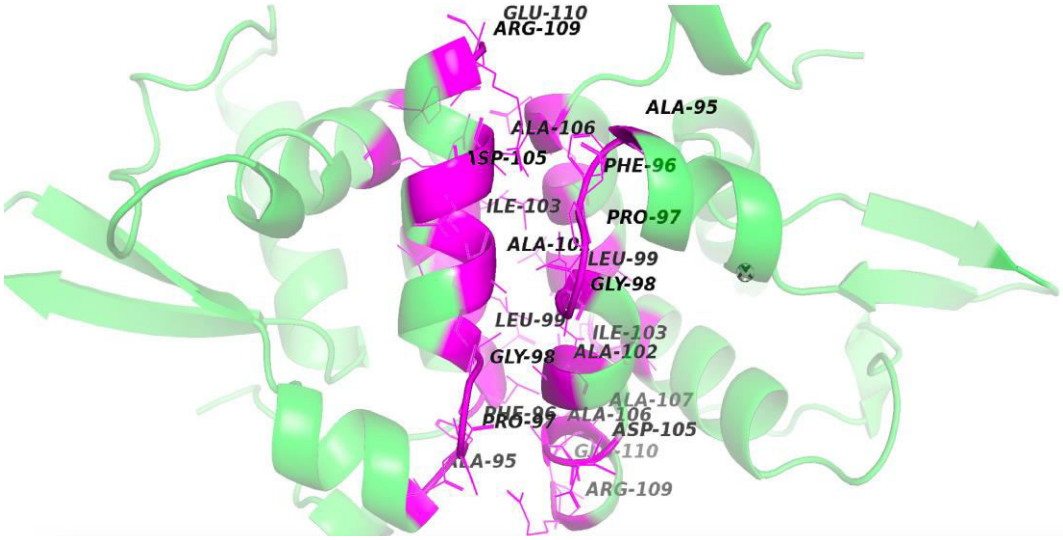
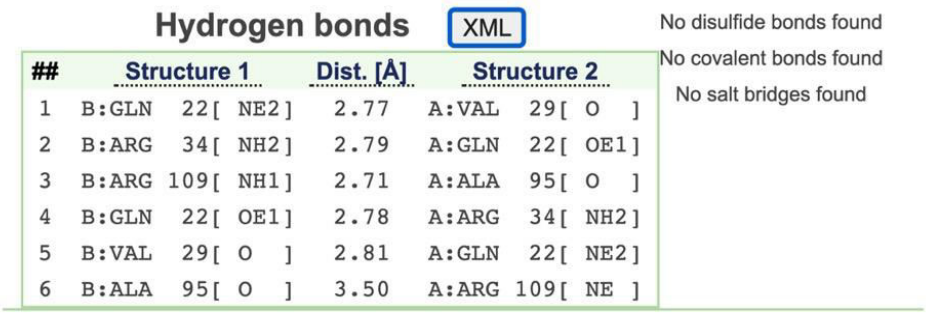


Figure 57. Dimer interface residues that may be interacting in homodimeric MmyJ. A) Residues at the dimer interface (pink) occurring along the y-axis that are in close enough proximity to interact are modelled between the 2 monomers of MmyJ (cyan and green). Interface dimers are parallel to each other and helices from each monomer are above or below one another. **B and C)** Possibly interacting residues (pink and labelled) between monomers of MmyJ (green). **B)** Residues (21-34) at the interface from the alpha helix 2-3 of each monomer. **C)** Residues (95-110) at the interface from the alpha helix 6-7 of each monomer.

Table 17: The dimer interface residues and the alpha chain they are part of in a homodimer of MmyJ. The amino acids near the dimer interface of MmyJ's that may be interacting were identified through the model (Fig. 51) and by PDBePISA[244] . As seen in the model, most residues are part of the alpha 2 or 7 helices.

Residue	Alpha chain
Leu21	2
Gln22	2
Leu25	2
Asp26	2
Leu28	2
Val29	2
Pro31	3
Arg34	3
Ala95	6
Phe96	-
Pro97	-
Gly98	-
Leu99	7
Ala102	7
Ile103	7
Asp105	7
Ala107	7
Arg109	7
Glu110	7

A



Hydrogen bonds XML

No disulfide bonds found
No covalent bonds found
No salt bridges found

##	Structure 1	Dist. [Å]	Structure 2
1	B:GLN 22[NE2]	2.77	A:VAL 29[O]
2	B:ARG 34[NH2]	2.79	A:GLN 22[OE1]
3	B:ARG 109[NH1]	2.71	A:ALA 95[O]
4	B:GLN 22[OE1]	2.78	A:ARG 34[NH2]
5	B:VAL 29[O]	2.81	A:GLN 22[NE2]
6	B:ALA 95[O]	3.50	A:ARG 109[NE]

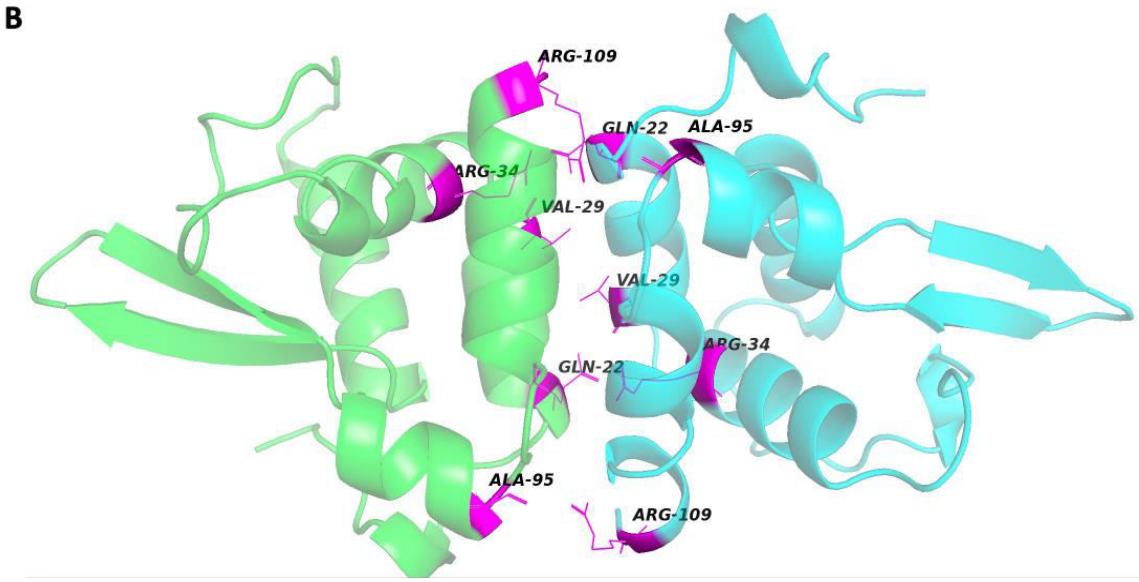


Figure 58. Hydrogen bonds occurring from dimer interface residues in MmyJ. Any bonds stronger than Van der Waals at the interface of a homodimer of MmyJ were evaluated by PDBePISA[244]. Only identical hydrogen bonds between the two residues were found (**A**) with Gln22, Arg34, Val29 and Ala95 forming said bonds. These were modelled on a homodimer of MmyJ (**B**) to show that these occur at the top, middle and bottom of the interface between both alpha helix 2 and 7 possibly providing enough strength to keep the 2 dimers together.

4.4 Analysis of MmyJ structure

4.4.1 MmyJ in comparison to other similar structures

The structure of MmyJ was compared with other structures in the PDB database to find ones that are the most structurally similar. Knowing these would help in identifying the family of proteins MmyJ may be a part of and having knowledge of the functioning of structurally similar proteins may assist in defining MmyJ's functioning.

From earlier phylogenetic and BLAST searching of the primary structure of MmyJ (Chapter 3.1) the case for MmyJ being part of the ArsR family of proteins was put forward. This is further consolidated from the most structurally similar proteins to

MmyJ that was identified (through PDBeFOLD[246]) (Table 18). Six proteins were identified with protein structures similar to MmyJ; SqrR[247], PagR[248], HlyU[125], ArsR[249], BigR[250] and TubR[251] (Table 18). These proteins are all from different bacteria with the top 5 similar to MmyJ (all except TubR) being ArsR family related proteins and all but ArsR itself being an exception to the class by not binding metal ions. As stated previously ArsR family proteins generally have their 5 alpha helices and 2 beta strands in the order of: ααααββα but some in this family (including proteins identified as similar to MmyJ) have more or less alpha and beta strands. It would therefore be appropriate for MmyJ to fall into this family of proteins as another non-metal binding exception.

Table 18: The proteins in the PDB database that are structurally the most like MmyJ. The structure of MmyJ was input into PDBeFOLD[245] to find structures with the highest identity to MmyJ. Results are organised by highest to lowest Q-score. The top 5 structures with high similarity to MmyJ are ArsR family proteins that are also considered irregular due to their ligand binding capabilities.

	Q score	P score	Z score	RMSD	N_{align}	N_g	$\%_{seq}$	N_{res}	PDB match	Match description	ArsR family?
1	0.49	5.7	7.4	2.00	82	2	15	95	6o8m	SqrR from *Rhodobacteria*	yes
2	0.49	5.4	7.2	2.02	78	4	17	86	2zkz	PagR from *Bacillus Anthracis*	yes
3	0.48	5.1	7.0	2.07	81	2	20	92	4k2e	HlyU from *Vibrio Cholerae*	yes
4	0.43	3.7	6.1	2.32	79	6	23	90	6j05	ArsR from *Acidithiobacillus ferrooxidans*	yes
5	0.42	3.1	5.7	2.38	82	4	22	99	3pqk	BigR from *Xylella Fastidiosa*	yes
6	0.28	4.4	6.6	1.80	65	4	18	111	6aht	TubR from *Bacillus Cereus*	no

The identified proteins were also superimposed onto MmyJ to visualise structural similarities and differences (Appendix Fig. B1). The structure with the most similarity (based on highest Q-score which takes into account the RMSD and therefore the 3D alignment of the structures) out of the ones identified appear to be SqrR[247] (Fig. 59B). The *Acidithiobacillus ferrooxidans* ArsR[249] (*Af* ArsR) also appeared to be structurally similar when superimposed onto MmyJ (Fig. 59A), likely from having the highest percentage sequence identity to MmyJ (Table 18). Both of these structures consist of 5 α helices and 2 β sheets, arranged in a 'typical' ArsR family structure whereby α1 and α5 of each monomer are antiparallel to each other but parallel to the same helix of the other monomer and the β sheets form a

hairpin structure. In MmyJ the beginning and end helices are parallel to each other and to each monomer therefore one helix from each monomer of MmyJ at the dimer interface is structurally misaligned with other ArsR family proteins including those in Fig. 59. For SqrR it is the helices near the C-terminal (α5 in SqrR and α7 in MmyJ) that align, while the N-terminal helices (α1 in SqrR and α2 in MmyJ) align between *Af* ArsR and MmyJ (Fig. 59). Both also have far more secondary structure compared to MmyJ with longer α helices and β sheets. Though *Af* ArsR does have a C-terminal tail important for ligand binding. Therefore, the N-terminal tail and long loops between α2 and α5 in MmyJ may also be necessary for its functionality. In SqrR the α3 helix seems to have the same break (K48, S49, V50) in secondary structure as seen in MmyJ (α4 helix) though this is the same in both monomers of SqrR unlike in MmyJ.

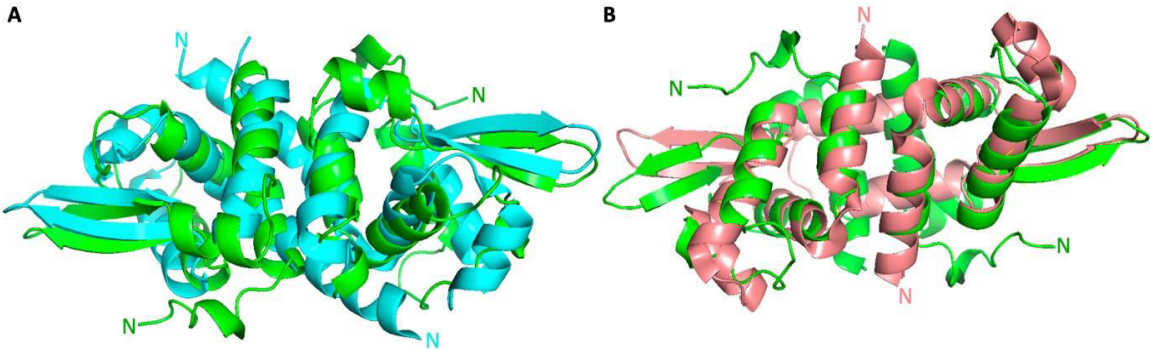


Figure 59. Structures of proteins most similar to MmyJ superimposed on MmyJ. When compared visually from Table 18 of proteins of high structural similarity to MmyJ (green), SqrR[246] (**A**-blue) and Ars from *Acidithiobacillus ferrooxidans* (*Af* ArsR)[248] (**B**-salmon) appear to the most similar. In SqrR (**A**), superimposition onto MmyJ (green) show MmyJ having a similar structure in the winged helix-turn-helix (wHTH) region towards the ends of each dimer but less so near the interface. Similarly, superimposition of *Af* ArsR (**B**) shows alpha 7 helices of MmyJ at the dimer interface being similar but very different in the alpha 2 helices at the back of the interface. *Af* ArsR also has a long flexible chain at the C-terminus whereas MmyJ has this at the N-terminus. In both SqrR and *Af* ArsR, these interface helices are antiparallel which is not seen in MmyJ. MmyJ clearly also has a more flexible structure with less secondary structure.

4.4.2 Inferences for the ligand binding site of MmyJ

The discovery of non-metal binding ArsR family proteins is increasing with even some of those structurally similar to MmyJ being non-metal binding. The function of these non-metal binding ArsR family proteins appear to be broad from; thiol regulation by SqrR[247] to S-layer protein regulation for virulence by PagR[248] to toxic compound regulation by YgaV[227] and others (as discussed in Chapter 3.1.2). These proteins have been classed as ArsR family proteins despite not binding metal ions due to their structural similarity to this class of family. Some are based only on primary sequence analysis (such as DepR2[211], BlmR[202] and EcaR[252]) which must be taken cautiously as primary sequence similarity does not necessarily match 3D structural similarity. However, others are classed as ArsR family proteins on their solved structures such as HlyU, BigR, PagR, SqrR, and YgaV, all of which except YgaV have been highlighted as structurally similar to MmyJ.

For known ligand binding of this family of proteins, one ligand binds at the same site to each monomer[115] and partial binding[253] (binding of only one ligand to a subunit of the dimer) is likely not enough to derepress the homodimer from the DNA.

There is very little structural data of ArsR family proteins bound to ligands and no non-metal ArsR family proteins bound to their ligands. For the majority of these non-metal binding ArsR family proteins there are no suggestions let alone experimental evidence as to the ligand that they may bind. The only characterised example of the residues involved in binding to a non-metal ligand of one of these proteins is SqrR.[247] Through mass spectrometry-based methods it was shown Cys41 and Cys104 are key residues involved in sulphur-species ligands in SqrR. MmyJ's ligand is thought to be MmA from earlier assays[155] of DNA-MmyJ as well as inference from metal binding ArsR family proteins. These metal binding Ars/SmtB binding proteins regulate expression of an exporter of certain metals that they themselves bind to. Similarly, as MmyJ is near the exporter controlling export of MmA it is rational to think MmA may be controlling MmyJ release from DNA.

Further evidence is required to test this and studying the residues participating in this interaction would be valuable.

To look at areas of possible interaction in MmyJ to ligands and other molecules, first the surface areas of the protein can be considered as this will highlight potential areas for binding. The solvent accessible surface (SASA) of MmyJ (Fig. 60A) shows the areas around the protein that is accessible to solvents (namely water) and are areas for binding interactions. In MmyJ this is an area of 22276.828 Angstroms^2 as calculated by Pymol. While the top of the molecule is fairly flat the bottom is seen to have two 'prongs' formed by the α5 helices sticking from underneath it (Fig. 60). Based on the SASA and the electrostatic surface potentials of MmyJ (Fig. 60B)[254], negatively charged molecules (such as DNA) can bind at either side of the MmyJ monomers or underneath of each monomer (Fig. 60). This concave pocket may be where a ligand for MmyJ can bind. As discussed earlier in Chapter 3.1.1, there are no Type 1 ("ELCVCD") or Type 2 ("DXHXXXXXXXXXXHXXE")[200] metal binding motifs present in MmyJ as there are in other metal binding ArsR family proteins therefore MmyJ is not thought to bind to metal ions but instead methylenomycin A (MmA).[155]

The Type 1 site corresponds to ligand binding occurring at the α3 region in ArsR's underneath the β-hairpin structure (like in *E. coli* R773 ArsR itself[188,189,255]) whereas Type 2 site is at the α5 area at the top of the molecule at dimer interface helices (like in SmtB[190,192,193,256]). Of course, both of these sites relate to specific motifs but binding may still occur at a similar region to these even if the motif is not for metal binding.

For MmyJ this would likely be equivalent to α7 for a Type 2-like site (Fig. 55A) but is more challenging to discern for a Type 1. The exact equivalent Type 1 site in MmyJ would be a section from α3 to α5, encompassing the short helix of α4 and regions either side of this (Fig. 60).

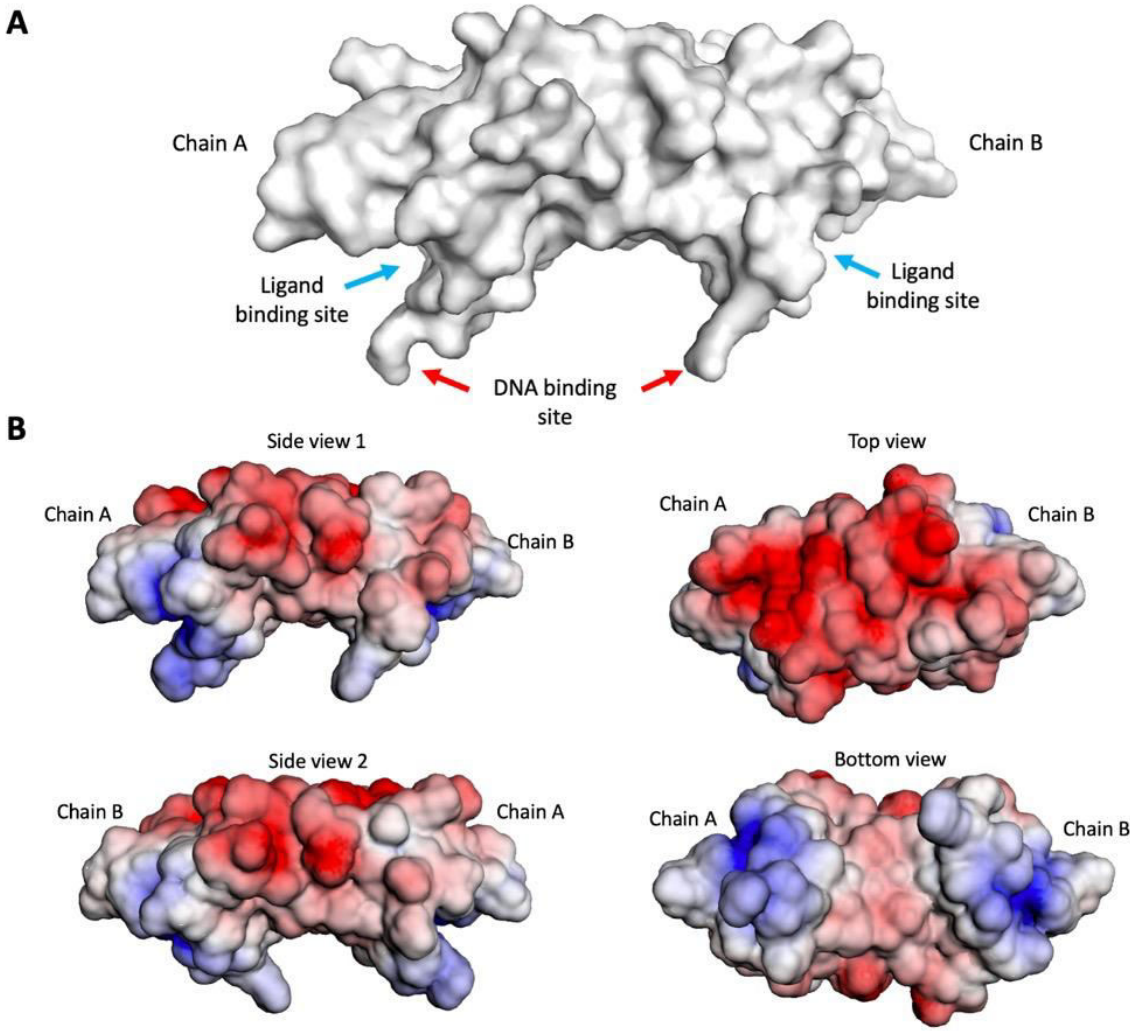


Figure 60. The solvent accessible surface area (SASA) of a homodimer of MmyJ. A) The SASA of a homodimer of MmyJ (produced by Pymol) shows the surface area of MmyJ where molecules can interact. The predicted sites for DNA and ligand binding are predicted from the SASA, similar structural proteins and the electrostatic surface potentials **(B). B)** The electrostatic surface potentials of a homodimer of MmyJ.[253] Blue areas indicate positive charges that can bind to molecules such as DNA, red indicate negative charges and white are uncharged areas. Further testing is required to support predictions.

There are exceptions of ligand binding to these sites in even metal binding ArsR family proteins themselves. In particular ArsR from *Acidithiobacillus ferrooxidans* (*Af* Arsr)[249,257] which has been shown as structurally similar to MmyJ (Table 18). While *Af* ArsR does bind As(III) it does not do so at the typical Type 1 site used by other ArsR's. Instead, it uses a region on the C-terminal tails ends of the protein

(near the α5 helix and below the α3) where cysteine resides (Cys95, Cys96 and Cys102) coordinate binding to As(III) (Fig. 61). This long C-terminal tail is not present in MmyJ nor the cysteines for this type of metal binding (Fig. 55A). The ArsR from *Corynebacterium glutamicum* (*Cg* ArsR[249]) likewise acts in a similar manner. Both of these have had crystal structures solved bound to ligand and so ligand binding residues can be confidently assessed. This variation of a Type 1 site (denoted an α3 binding site) may be present in other non-metal binding ArsR family proteins too.[125,184,247,250] This is because many of these (e.g. BigR, YgaV and SqrR) have cysteines along this site which are thought to be relevant to their functioning though no experimental data exists for the role of them.

From Fig. 54, it's apparent a Type 2-like site would not be suitable for ligand binding as this area is not solvent accessible. Instead, a Type 1-like site (or α3 site) is considered more appropriate as this region is clearly solvent-accessible. Furthermore, as stated earlier cysteines in this region are considered relevant for ligand binding in many of the non-metal binding ArsR family proteins. MmyJ has only one cysteine residue (Cys49) which is between α3 and α5 meaning this site may be used for ligand binding with the cysteine playing a role in it.

These predictions are made tentatively as there is not enough data or conservation between these non-metal binding ArsR family to propose a ligand binding site in MmyJ. Even so, while the region of ligand binding can be predicted the exact residues involved in binding to MmA cannot be. MmA is a cyclopentanone (a cyclic ketone) and so can be negatively charged by the oxygen atoms.[140,258] In Fig. 60, positively charged sites on the surface of homodimeric MmyJ were identified and in particular, cervices within each monomer of the protein are shown to be charged positively which is the predicted site of MmA binding to MmyJ. Within this site is also where the only cysteine in MmyJ is present (Fig. 55) which may be involved in ligand binding or not. To test this hypothesis, as well assess the significance (if any) of the Cys49 and to confirm predictions of MmyJ binding MmA at a Type 1-like binding site of ArsR family proteins, more experimental work such as assays and crystals of ligand bound MmyJ would need to be done.

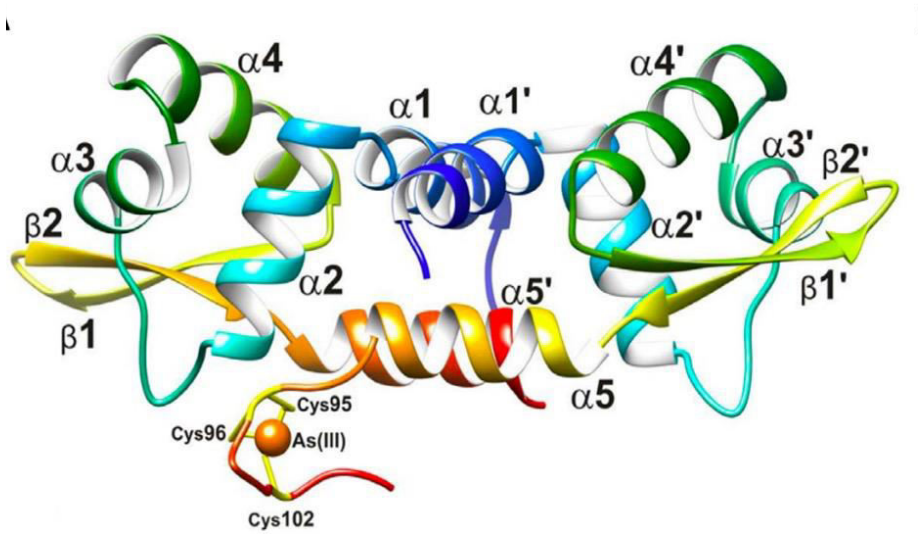


Figure 61. Crystal structure of *Af* ArsR bound to As(III).[248] Many ArsR family proteins (such as *E. coli* R773 ArsR) binds to its ligands using the 'Type 1' metal binding site which is at the α3 site.[190-192] The ligand bound crystal structure ArsR from *Acidithiobacillus ferrooxidans* (*Af* Arsr) from Prabaharan C. *et al.* (2019)[248] shows it uses 3 cysteines (C95, C96 and C102) at the C-terminal tail to bind to As(III). From the solved apo-crystal structure of MmyJ (Fig. 55A), MmyJ lacks this long C-terminal tail for ligand binding and while the N-terminal tail is long, the only cysteine in MmyJ is near the α4 helix (C49). Figure was taken from Prabaharan C. *et al.* (2019).[248]

4.4.3 Predictions of the DNA recognition sites of MmyJ

All ArsR family proteins are DNA binding as this gives them their functional qualities of being transcriptional repressors. Even before structure elucidation, MmyJ was found by functional assays (Chapter 3.3) to be DNA binding and this region was between *mmyJ* and *mmr*. With the structure of MmyJ being solved an analysis of the likely DNA-binding site of MmyJ was done as described in this section.

The DNA binding properties of ArsR family proteins come from them all having winged helix-turn-helix domains regardless of them being metal or non-metal binding. This wHTH is formed by the α3 loop αR in ArsR proteins including ArsR itself (Fig. 62). As mentioned earlier (Chapter 3.1) this HTH region binds in the major groove of the DNA while the β-hairpin structure helps in stabilisation through interaction with the DNA's minor groove. [122,200]

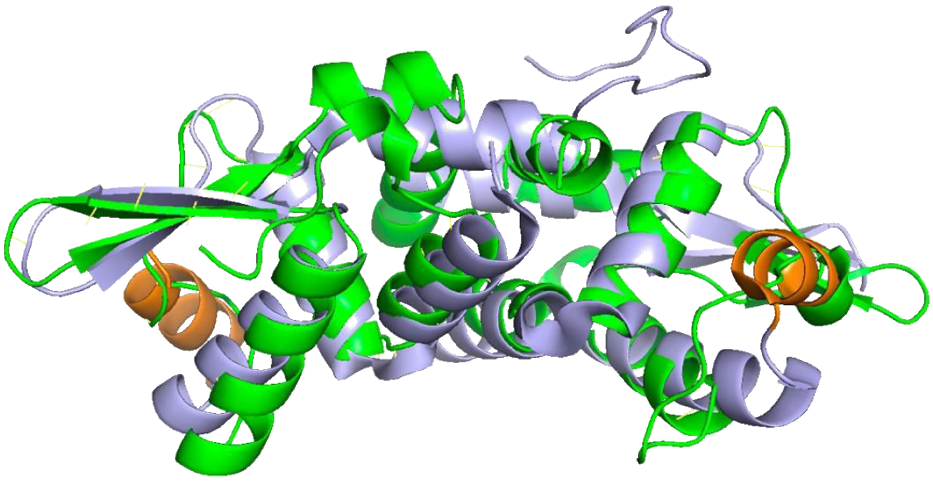


Figure 62. The superimposed structures of ArsR and MmyJ with ligand and DNA binding regions of ArsR highlighted. The structure of *Acidithiobacillus ferrooxidans* (*Af* ArsR) (in grey) was superimposed on MmyJ (green) to highlight the equivalent regions onto MmyJ where ligand/DNA binding occurs in *Af* ArsR[248]. The alpha 3 helix (orange) is the ligand binding site for Type 1 metal binding ArsR ligands (like *E. coli* R773 ArsR[190-192]). This helix is part of the helix-turn-helix region (α3 turn αR) where DNA binding occurs in all ArsR family proteins. Alpha R is akin to alpha 5 of MmyJ whereas there does not appear to be an exact equivalent in MmyJ to alpha 3.

There are only two cases so far of structural data of ArsR family proteins bound to DNA; NolR[259] (Fig. 63) and CzrR[253] with NolR being a crystal structure and CzrR being an NMR solution structure. Along with this, mutagenesis studies and modelling of ArsR family protein structures bound to DNA have also provided insights into residues involved in DNA binding. While there is no overall conservation among all ArsR family proteins for DNA binding, there is evidence that a patch of basic residues[125,189,193,202,204,227,248,250,260] in the α4 sites of each monomer (SmtB nomenclature) are involved in interactions with B-form DNA. This patch can consist of non-basic residues too including; arginines, serines, leucines and glutamines with the importance of these residues varying between proteins. For example in BlmR, Q51, Q52, R57 and R58 are valuable as mutation of these renders the protein unable to bind to DNA.[202] On the other hand, NolR bound to DNA (Fig. 63) showed Q56 as being valuable in recognising variation in DNA as these ArsR proteins bind to imperfect palindromic sites.[259] In *E. coli* ArsR His50 is required as mutation of this prevented the protein binding to DNA.[188] While salt bridges between Cys42 and

Cys108 being broken are predicted to be important for BigR to bind to DNA.[250] Therefore there are no exact conservation of residues between all ArsR family proteins for DNA binding.

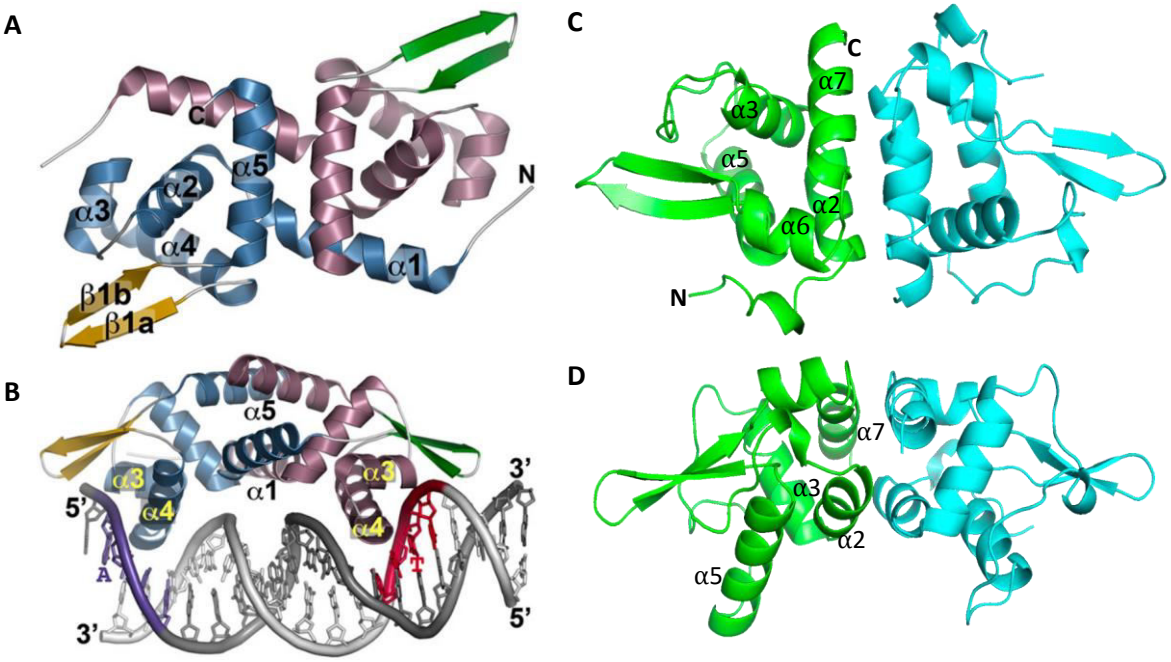


Figure 63. Crystal structure of homodimeric apo-NolR and DNA-bound NolR compared to apo-MmyJ in similar orientations. A) The crystal structure of apo-NolR and **B)** the crystal structure of the ArsR family protein NolR bound to DNA was solved by Lee *et al.* (2014).[258] As expected by ArsR family proteins, homodimeric NolR binds to the major groove of the DNA using a helix-turn-helix motif produced from its α3 and α4 helices with the beta-sheets at either end stabilising the interaction. Of the residues in NolR seen to be interacting with the DNA, it is Q56 from each monomer (from the α4 helix) that was highlighted as being able to recognise a variation in nucleotides (A or T). DNA-bound NolR has a 0.43 Å change from apo-NolR akin to a hinge where alpha 5 helices are pushed upwards. Figure taken from Lee *et al.* (2014).[258] **C and D)** The apo-MmyJ homodimers in a similar orientation to NolR. **C)** Alpha 2, 7 and beta sheets of MmyJ are in similar orientation to apo-NolR. **D)** The alpha 5 of MmyJ may interact with DNA.

Based on these inferences and the binding of DNA predicted to be at or near the α5 helices in MmyJ (Fig. 60 and Fig. 63D), key residues within this helix may be forming interactions with DNA bases. This α5 helix between the two monomers can vary by 3 residues at the start of the helix (Fig. 55) but both encompass 11 residues (G61-A71) with several within this including positively charged residues (histidine,

arginine, and lysine). H63 and H64 are found near the start of the α5 helix and one or both of these residues may play a role in interacting with the negatively charged DNA bases like in the case of *E. coli* ArsR. [188] Similarly, there is a lysine (K66) two bases down from H64 that has a long positively charged side chain that could also extend into the major groove of DNA to form ionic interactions. An arginine (R69) is present near the end of the α5 helix that may also play a role in DNA interactions like in the case of BlmR[202] and NolR[259] with a glutamine (Q70) beside it. While glutamine is not a charged residue but is instead a polar one, in the case of NolR this residue was seen to be a key amino acid for interacting with A/T bases (Fig. 63B). [259] This may also be true for MmyJ however, as both Q70 and R69 are found at the ends of α5 nearer the β-sheets, spatially they're unlikely to interact with DNA.

The size of ArsR family proteins can vary meaning the length to which they can bind DNA can also vary (even though a 12-2-12 imperfect palindromic region is suggested[200]) and so determining the angles at which binding causes the DNA to bend can be difficult. The crystal structure of PhoB bound DNA, shows the DNA bends 40° to bind to the homodimeric protein.[261] This 40° bend of DNA is also suggested in models with PagR[248] and BlmR[202]. However, in crystal structures of NolR[259] bound to DNA it's drastically reduced to only 16.8° which is also around the same (15°) as predicted in models of HlyU-DNA[125]. The distance between the α-helices binding to DNA is around 35 Å (34 Å in SmtB[193], 39 Å in NolR[259] and predicted to be around 34 Å in PagR[248]). Binding to a ligand or DNA region also induces an allosteric change in the protein.

In NolR binding to DNA caused a 0.43 Å change from apo-NolR to DNA-NolR (Fig. 63). With apo-MmyJ in the same orientation as apo-NolR (Fig. 63A,C) there seem to be structural similarities with alpha helices being present at the dimer interface and beta-strands protruding out from the sides from each monomer. There is also a degree of 'planarity' observed between both apo-structures, however the arrangement of α2-α6 in MmyJ (comparative to α2-α5 in NolR) is different. As

stated earlier DNA binding to NolR creates an allosteric change to apo-NolR whereby α3-α4 press closer toward the centre of the dimer causing the α5 and α1 helices of the dimer interface to be pushed upwards akin to a hinge (Fig. 63A,B). Therefore, based on the NolR mechanism, DNA-bound MmyJ may have identical allosteric changes where the α3 and α5 may also press closer together and cause α2 and α7 helices to be pushed upwards.

DNA-bound NolR along with free energy changes from the solution structure of CzrR give a reasonable model of the structural changes occurring when an ArsR family protein binds and unbinds to DNA. In metal binding ArsR family proteins, metal binding causes an allosteric change by disruption of hydrophobic interactions.[253] For those binding at the Type 2 site (i.e. near the dimer interface) this disruption formed by key histidine residues (His97 in CzrR and His117 in SmtB) causes a quaternary switch which spatially alters the αR (DNA binding) helices and leads to global conformation changes.[193,253] Ligand binding occurring near the Type 1 site (e.g. *E. coli* ArsR) likely causes interference with DNA binding through disruption of hydrophobic interactions in the DNA binding regions as well as the dimer interface.[249] A destabilisation of the dimer interface is therefore also considered important to cause release from the DNA. In summary, ligand binding to a DNA bound ArsR family protein causes it to go from a 'closed' structure to a more flat, 'open' one from the shifting of the wHTH region, which gives the ligand bound protein a lower affinity to bind to DNA.

The structure of MmyJ was used to model its interaction with DNA and the sites it may bind to (Fig. 64) (model created by Dr Crow). The two likely DNA-binding helices are spaced such that they could conceivably bind within the major grooves on the same face of the DNA. Whether the DNA remains straight or bends to have β-sheets interact with it also is unknown. Based on this model (Fig. 64), simple predictions of the likely DNA sequence that MmyJ could bind to were made.

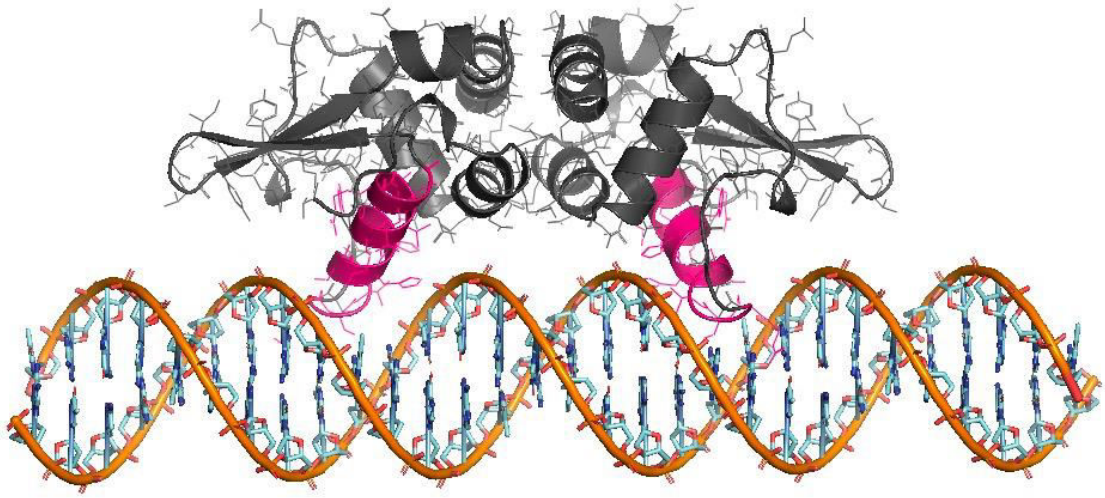


Figure 64. A model of MmyJ binding to DNA. This model of a dimer of MmyJ binding to B-form DNA was created by Dr Crow (University of Warwick, 2021). Highlighted is the alpha 5 helix of MmyJ (pink) that interact with the major groove of DNA. This model infers MmyJ is planar though allosteric changes may occur during DNA binding and the beta strands may assist in stabilising the DNA-bound complex by interacting with the minor groove. This interaction encompasses around 17 base pairs of DNA with potential regions having been found in the intergenic region between *mmyJ* and *mmr*.

Firstly, due of the symmetry of the MmyJ dimer and its disposition, palindromic sequences were expected as these form contacts with the DNA binding helix of each monomer. Also expected was a 3 or 4 bp region in the centre of the palindromic sequences that would not have contact with the MmyJ dimer. With this in mind, the intergenic region of DNA between *mmyJ* and *mmr* was inspected to find two sequence motifs consistent with the aforementioned predictions. The first DNA site consists of the sequence; CCATCAAatgTTGACGG and the second has the sequence; CCATCAAagtTTGACAG where bases in lower case for both sequences indicate DNA that would not be in contact with the MmyJ homodimer (Fig. 65). Notably both of these sequences are found directly after one another and are near identical. The only difference between the two sequences with bases that are predicted to be in contact with MmyJ residues, is a single nucleotide difference second to the end (G/A) (Fig. 65). DNA binding assays were used to test these predicted sites with results of this functional study reported in Chapter 3.3.

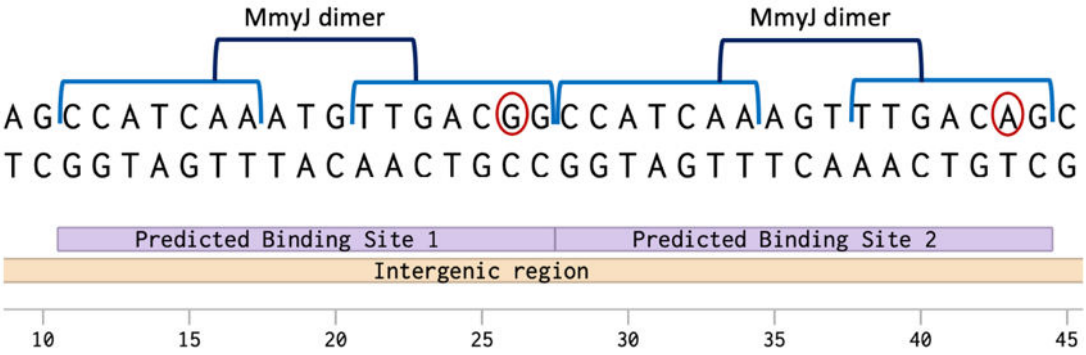


Figure 65. The pseudo-palindromic sequences predicted to be sites of MmyJ homodimer binding. By solving the structure of MmyJ and modelling this homodimer of MmyJ to B-form DNA (Fig. 64), two sequences with the intergenic region between *mmyJ* and *mmr* were found that could bind to a MmyJ homodimer. The regions were each MmyJ monomer of the dimer complex would bind to is shown by blue brackets. Both sequences are beside one and are near identical to each other except for one nucleotide (red circle) that would be in contact with a MmyJ residue. Both predicted DNA sequences were later used for electromobility shift assays to test for DNA-protein binding.

4.5 Future structural work involving MmyJ

Solving the structures of Apo C49S MmyJ and WT MmyJ has led to stronger inferences on the class and functionality of MmyJ. The structure of Apo-MmyJ already suggests it is part of a non-metal sensing group of ArsR family proteins and if this is true it would be the first structure of an antibiotic related ArsR family protein to be characterised. While this structure has given predictions on the regions of the protein used for ligand and DNA binding, experimental data is required to confirm this and the functioning of the protein.

4.5.1 X-ray crystallisation studies with MmyJ

X-ray crystallography work has been valuable already in showing the structure of MmyJ which varies from the predictions with just primary structure[155]. Continuing this MmyJ crystallised with its ligand (MmA) would be valuable. It is the WT MmyJ

that would likely be used here as the cysteine 49 may play a role in the ligand binding which when mutated could render it non-ligand binding. The structure of WT MmyJ and ligand could show unique interactions of certain residues with MmA that have never been seen before. Not only this but ligand binding may cause an allosteric change in the structure of the protein and comparing this to the apo-protein or DNA-bound protein can allow for predictions of the movement of secondary structure to allow for this. So far no allosteric changes are known of ligands binding to non-metal ArsR family proteins though zinc binding to SmtB did show a shift of 2.4 Å from the apo-form to the ligand bound form.[193] A similar change may be seen in MmyJ though it will probably be different as ligand binding in MmyJ likely occurs at a different place to SmtB. Finding the residues involved in this ligand binding could then be used to find similar motifs in other antibiotic sensing proteins and from there map the phylogeny and evolution of these types of ArsR family proteins.

The other form of MmyJ would be MmyJ bound to DNA. Crystallisation with DNA can be more difficult than crystallisation of just protein or ligand bound protein even if the protein is more stable bound to DNA because too much overhang of nucleotide bases can introduce flexible regions that make crystallisation difficult. However, with the exact range of binding of DNA predicted and confirmed by binding assays, crystallising MmyJ with DNA may well be tractable. For this it would not matter so much whether C49S or WT MmyJ is used as it has been shown functionally that the cysteine is not involved in DNA binding. Having the structure of DNA-MmyJ can help to highlight any allosteric changes occurring of the protein caused by DNA binding. It can also help to show if the full HTH is required in MmyJ for DNA binding or if as predicted (Fig. 64) only the α5 is required, as well any other overlooked structures assisting with this binding.

Having structures of both these forms of MmyJ would give a clearer view of how MmyJ functions and give finer details into the mechanism.

4.5.2 Mutations on MmyJ to identify key residues for functionality

To identify key residues that play a role in the functionality of MmyJ, these could be mutated. Already a cysteine mutation has been made and this showed no changes to the structure of the protein. The cysteine mutation also did not hinder the ability of the protein to bind to DNA showing that this residue is likely not involved in DNA binding. However, testing this mutant for ligand binding would be valuable in establishing if like other ArsR family proteins, cysteines are important for binding to a ligand.

The Type 1 metal binding site of ArsR family proteins (Fig. 62) (which although is not present in MmyJ) was discussed earlier as to have a likely equivalent site in MmyJ for MmA binding (Chapter 4.4.2). Based on this, residues between α3 and α5, including the short α4 helix, may be involved in MmA interactions and binding (residues around 43-57) (Fig. 55). MmA is a cyclic ketone molecule [140,258] with oxygen atoms that can form interactions with positively charged residues. Only residues at either end of this predicted α3-α5 site (K42 and R58) are positively charged. Towards the α5 site there are also some polar residues (T51, T55 and S57) which may also be involved with ligand interactions. However, the rest of the residues especially those nearer the α3 site appears to be largely hydrophobic or negatively charged amino acids and so are unlikely to interact with MmA. To test whether these predicted residues (C49, K42, R58, T51, T55 and S57) are important for ligand binding in MmyJ, they could be mutated to hydrophobic residues and ligand-binding assays performed.

Mutations along the α5 helix of MmyJ predicted to bind to DNA could also be made and tested to see if it binds DNA to determine the residues contacting nucleotides along the DNA. This could be conducted even without having the structure of DNA bound MmyJ.

Lastly mutating residues predicted to be involved in forming intermolecular bonds at the dimer interface may assist in determining key residues that maintain the

dimerization of MmyJ. Hydrogen bonds are the stronger than Van der Waals forces though too much disruption of these may cause problems in protein folding. However only a few of these hydrogen bonds between residues exist in MmyJ at the dimer interface (Fig. 58) and mutating these few may not prevent issues in protein folding. For this the only two residues that are charged and thus creating hydrogen bonds could be mutated as such: Q22A and R34A. Should this not be sufficient to prevent a dimer forming, further mutations may be required. Residues that are non-charged but may be important (mostly leucine's) due to their extension into the dimer interface could be replaced with amino acids that are charged. In such case, the additional mutations that could be made would be: L21E, L25E, V29E, L99E and L28E. Based on this, Q22A, R34A, V29E and L25E mutation were made and tested for (Chapter 3.5) with only L25E seen as vital for dimer formation. The other mutations proposed could also be tested in a similar manner.

The functionality of the L25E MmyJ mutant could certainly also be tested. Studies on the DNA binding and ligand binding properties of this mutant MmyJ could be conducted to indicate whether dimerization is necessary for the functioning of MmyJ. All ArsR family proteins are found as homodimers with dimerization even being shown as essential for *E. coli* ArsR to its functioning.[185] The stability could also be tested to identify if dimerization also assists in stabilising the protein and/or if a monomer can be stabilised with DNA/ligand binding.

These would all be interesting properties of MmyJ to study in the future.

Chapter 5: MmyB: A proposed transcriptional activator of methylenomycin A biosynthetic genes

5.1 Predictions of the biological role of MmyB

5.1.1 The *mmyB* gene and its role in methylenomycin biosynthesis

MmyB is the predicted activator of genes needed for production of the antibiotic MmA in the methylenomycin gene cluster (as described in Chapter 1). Its importance to the production of methylenomycin was first evidenced when an in-frame deletion of the gene resulted in a loss of methylenomycin production. This still retained production of signalling molecules that induce biosynthesis of MmA indicating that deletion of even just *mmyB* switches off MmA biosynthesis production.[150] The mechanism by which MmyB acts is still unknown however the region to which MmyB binds to DNA and thus activates methylenomycin biosynthesis has been proposed.[150] This is a 16 bp region containing 2 repeat sequences that are almost identical (CACCAGGCCCCGCCG and CACCGAGGCCCCGCCG) which has been termed 'B-boxes'. This sequence is found between the *mmyB* and *mmyY* genes which are separated by 231 bp (Fig. 66). The transcription start sites of *mmyB* and *mmyY* were described by Bentley *et al.* (2004)[22] following mRNA analysis with high resolution S1 mapping also done by O'Rourke *et al.* (2009)[150] (as previously discussed in Chapter 1.4.1). O'Rourke *et al.* (2009)[150] found there to be the same transcription start site as found by Bentley *et al.* (2004)[22] and also multiple other potential transcription start sites for *mmyb* all downstream to that initial one described by Bentley *et al.* (2004)[22]. The promoter regions for *mmyY* and *mmyB* have not as of yet been seen experimentally and those shown in Fig. 66 were predicted using the online tool BProm[262]. These same promoter regions have also been predicted bioinformatically by Styles K. (2016)[58].

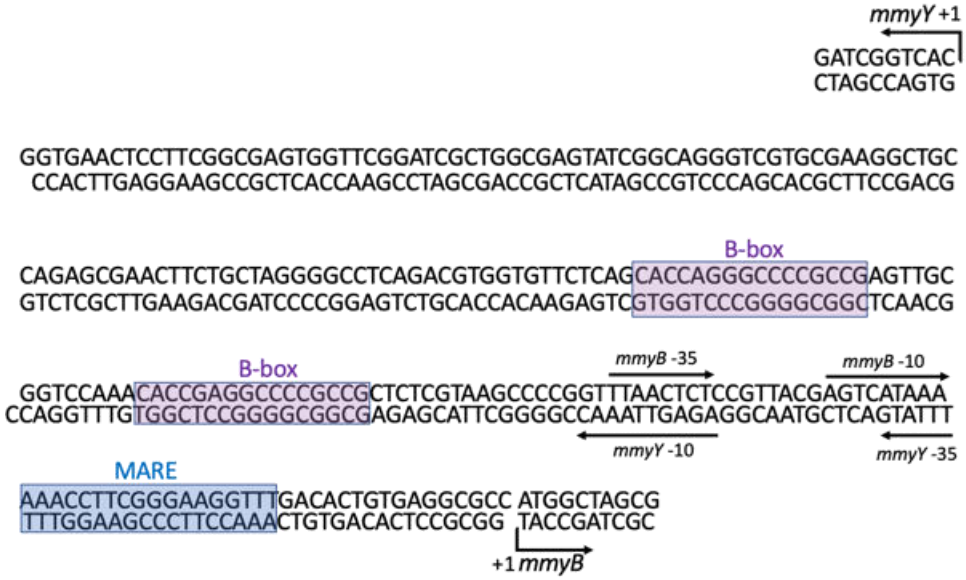


Figure 66. Intergenic region of DNA between *mmyB* and *mmyY*. The genes *mmyB* and *mmyY* are part of the methylenomycin gene cluster in *S. coelicolor* SCP1 plasmid. MmyY is thought to be involved in the biosynthesis of methylenomycin A (MmA) while MmyB is regarded as the transcriptional activator for MmA biosynthesis. The MARE sequence (in blue) is known to be the recognition motif to which MmfR (the transcriptional repressor of *mmyB*) binds.[262] In purple are B-box sequences that are proposed to be potential binding sites for MmyB.[148] The +1 transcription start sites, -35 promoter sites and -10 promoter sites for both genes are indicated. Transcription start sites of *mmyY* and *mmyB* were described by Bentley *et al.* (2004)[22] then again by O'Rourke *et al.* (2009)[150] using mRNA transcript analysis. Promoter regions were predicted bioinformatically using BProm[261] and with similar regions also predicted bioinformatically by Styles K (2016)[58].

MmyY is involved in the biosynthesis of methylenomycin so the control of expression of this gene would therefore be appropriate for MmyB. Furthermore, the operon for *mmyY* also controls the expression of the gene *mmyF,* the gene beside *mmyY* (Fig. 8), which as discussed previously is required in the final stages for coverting MmC to MmA (Chapter 1.4). The formation of these divergent operons is not unlike MmyJ which is separated by the gene it is proposed to transcriptionally control by 218 bp (Chapter 3).

Understandably, *mmyB* is not a constitutively expressed gene and is under transcriptional control itself. In the *mmyB* promoter is the MARE (MMF-

autoregulator responsive elements) sequence (Fig. 66) which is the location of binding of MmfR – another protein in the methylenomycin gene cluster. MmfR is a transcriptional repressor that exerts control over *mmyB* expression by binding to this MARE region and blocking the -35 and -10 promoter sites of *mmyB* (Fig. 66).[263] MmfR can be depressed from the DNA by binding methylenomycin furans (as discussed in Chapter 1.4).[151] This adds a further level of control of antibiotic production.

This synteny is present in other gene clusters within *S. coelicolor* A3(2) as identified by Xu *et al.* (2012) (Table 19).[264] The majority of these MmyB-like proteins in *S. coelicolor* A3(2) fall in same group as MmyB, that is they appear to have a greater relation to each other as shown by BLAST values, with only one appearing to be in a separate group with less homology to the others (Fig. 67). While MmyB proteins have been analysed the intergenic regions between them have not. These regions may contain conserved sequences that could identify areas where similar MmyB-like proteins bind to. Taking the intergenic regions between these and conducting a MEME analysis did show some conserved regions (Fig. 68) but these do not seem to be the 'B-boxes' or MARE sequence (Fig. 66) though there were some bases conserved within the group. The helix-turn-helix motif, involved in binding to DNA, in these MmyB-like proteins also showed conserved residues[264] and these conserved residues may be the ones interacting with the conserved bases in the intergenic region.

Table 19: Genes that are *mmyB*-like and have similar synteny within the *S. coelicolor* A3(2) genome. Supplementary figure from Xu *et al.* (2012).[263]

Paralog	Divergently transcribed gene	Intergenic region bp	Predicted function of gene product(s)	Comment
Family 1				
mmyB	*mmyY*	230	Nuclear transport factor 2-like protein	Methylenomycin biosynthesis
SCO0110	*SCO0111*	187	3-ketoacyl-(acyl-carrier-protein) reductase	
SCO0233	*SCO0234-SCO0235*	289	Both genes encode an SDR	
SCO0236	*SCO0237*	200	SDR	
SCO0307	*SCO0306*	123	Possible pseudogene	SCO0311 PKS; SCO0312 acyl-CoA reductase
SCO0891	*SCO0890*	102	SDR	
SCO2501	*SCO2502*	111	Similar to actinorhodin exporter ActII-2	
SCO2537	*SCO2537*	107	Similar to cephamycin transporter CmcT from *Nocardia lactamdurans*	Fused to SCO2538
SCO4680	*SCO4681-SCO4682*	45	SDR, downstream gene encodes 4-oxalocrotonate tautomerase	
SCO4944	*SCO4945*	169	MDR (alcohol dehydrogenase)	Highly conserved^
SCO6926	*SCO6925*	582	Membrane protein; gene contains TTA codon	Possibly controls lantibiotic cluster *SCO6927-6930*
SCO7140	*SCO7141*[#]	28	SDR	
SCO7706	*SCO7707*	138	Atu4866 family protein	SCO7705 similar to tetracyclin resistance oxidoreductase
SCO7767	*SCO7768-SCO7769*	130	Nuclear transport factor 2-like protein, similar to RANGDP-binding; fused to gene for GDP-sugar epimerase	Next to *SCO7766*, homolog of *actVA4*
SCO7817	*SCO7818*	117	SDR (2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase)	
Family 2				
SCO6539	*SCO6540-SCO6541*	90	DCoH domain protein; downstream gene similar to *S. purpurascens rdmD* (anthracycline biosynthesis)	Highly conserved in *Streptomyces*

[#] no shared promoter region.
^ next to ortholog of *SCO4945* in all *Streptomyces* sp, *Rhodococcus* sp, *Saccharopolyspora erythraea*, *Kineococcus radiotolerans*, *Arthrobacter aurescens*.

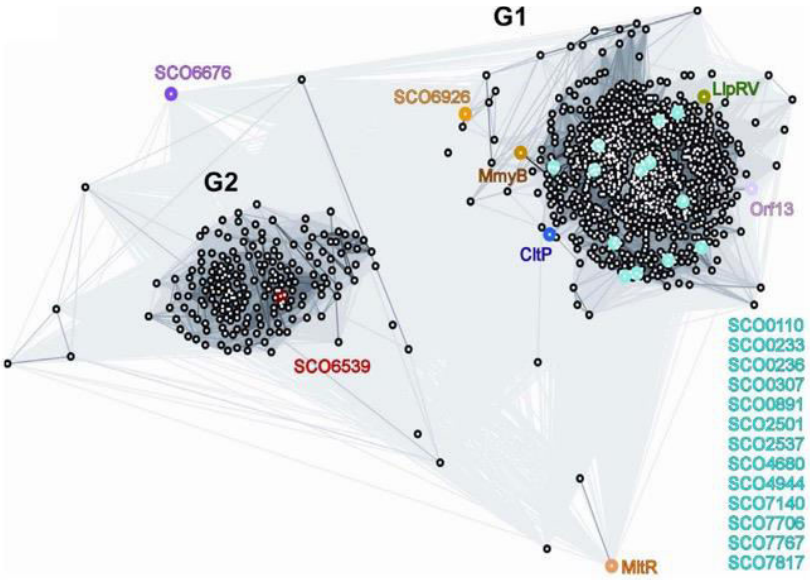
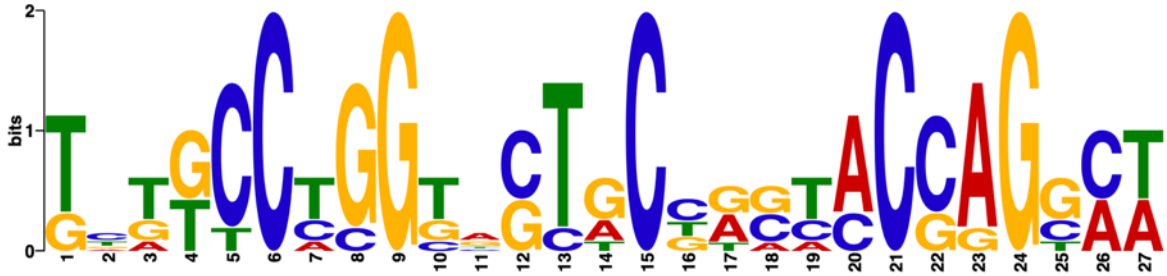


Figure 67. There are many *mmyB*-like genes in the *Streptomyces* species. MmyB-like proteins are largely part of 1 of 2 groups. Clustering analysis of MmyB-like proteins where each protein is indicated by a dot and lines indicate sequence similarity detected by BLAST searching with a black to grey gradient showing BLAST p–values of high to low significance. Some of the better characterized proteins are labelled and shown. Figure from Xu *et al.* (2012).[263]



5.1.2 Structural insight into MmyB from bioinformatics and homology to a MmyB like protein

Figure 68. MEME analysis of intergenic regions between *mmyB* like genes and genes divergent to *mmyB*-like ones in *S. coelicolor* A3(2). Genes that were *mmyB* like were identified by Xu *et al.* (2012)[263] in *S. coelicolor* A3(2) (Fig. 61B). The intergenic regions between these were taken and analyzed for any conserved motifs or regions using the MEME Suite. In the majority, there were seen to be some conservation of residues however this was not the proposed 'B-box' conserved in *mmyB* intergenic regions.

The *mmyB* from the *S. coelicolor* A3(2) SCP1 plasmid is 942 bp long which translates to 313 amino acids (Fig. 69A). This means it has a weight of 34351.76 Daltons (Da), an isoelectric point of 9.70 and an instability index of 52.05, meaning it is predicted to be unstable. Furthermore, in terms of translational control, *mmyB* contains a TTA codon which is translated by the rare leucyl-tRNA BldA.[265]

A

atggctagcgtcgatggcgtggccaaagattcgacggtgtgcagtcctaaaaggcagcgc
M A S V D G V A K D S T V C S P K R Q R
gaggcgttacgccacttcctgcggtcccgcagagcacggctttcacccgacgacgtcggg
E A L R H F L R S R R A R L S P D D V G
ttgcttgctaccggtcggcggcacacgccgggactgcgccgtgaggaagtcgcagtcatc
L L A T G R R H T P G L R R E E V A V I
gccggagtcagcgcctcctggtacacgtggctcgagcaggggagggatatcaaagtctcg
A G V S A S W Y T W L E Q G R D I K V S
gacggcgtgctgaacgcgatcagtcaggccctgagactcgatgacaccgaacgggcacac
D G V L N A I S Q A L R L D D T E R A H
ctctaccggctggctggggtcaatccgccgcagtcggtgccggcgaccgcggggcagaca
L Y R L A G V N P P Q S V P A T A G Q T
gagacgtccagacttcagctcatcgtggacggctggctgccggcacccgctttcgtcgtc
E T S R L Q L I V D G W L P A P A F V V
gaccgctactggaacaccctggccgccaatcaggcggcacggtctgcgctgggcgtggga
D R Y W N T L A A N Q A A R S A L G V G
gcgggcgaccagaactacctcgcggcgttcttcaccgagccgaccgcccgcgcccgctac
A G D Q N Y L A A F F T E P T A R A R Y
ctcgactgggacaagctggcgacccggttggtcggccagttccgtgtccaggcagcacgc
L D W D K L A T R L V G Q F R V Q A A R
tttccggaagatccgcggttcgacaggatcgctcgccagctgtgcgccaccgaccatgcc
F P E D P R F D R I A R Q L C A T D H A
ttcgccgacctgtgggccaggcacgaaacctgtgacacggcgatgacgagtgtgcgggta
F A D L W A R H E T C D T A M T S V R V
cgaccgccgggcgaggagagcatgcgcttcgagcacctgatcctggcgcttctcgagaac
R P P G E E S M R F E H L I L A L L E N
gcggatctacggctcatgctctacatgcccaggggcgcgcacatctccgcagaggccggt
A D L R L M L Y M P R G A H I S A E A G
ctgcggctggtgccagctgacccccctgacgcgaaattccctactccaactccctggtcgt
L R L V P A D P L T R N S L L Q L P G R
ccgatgacgtcacgacctgtcgagcggcacatccatccctga
P M T S R P V E R H I H P -

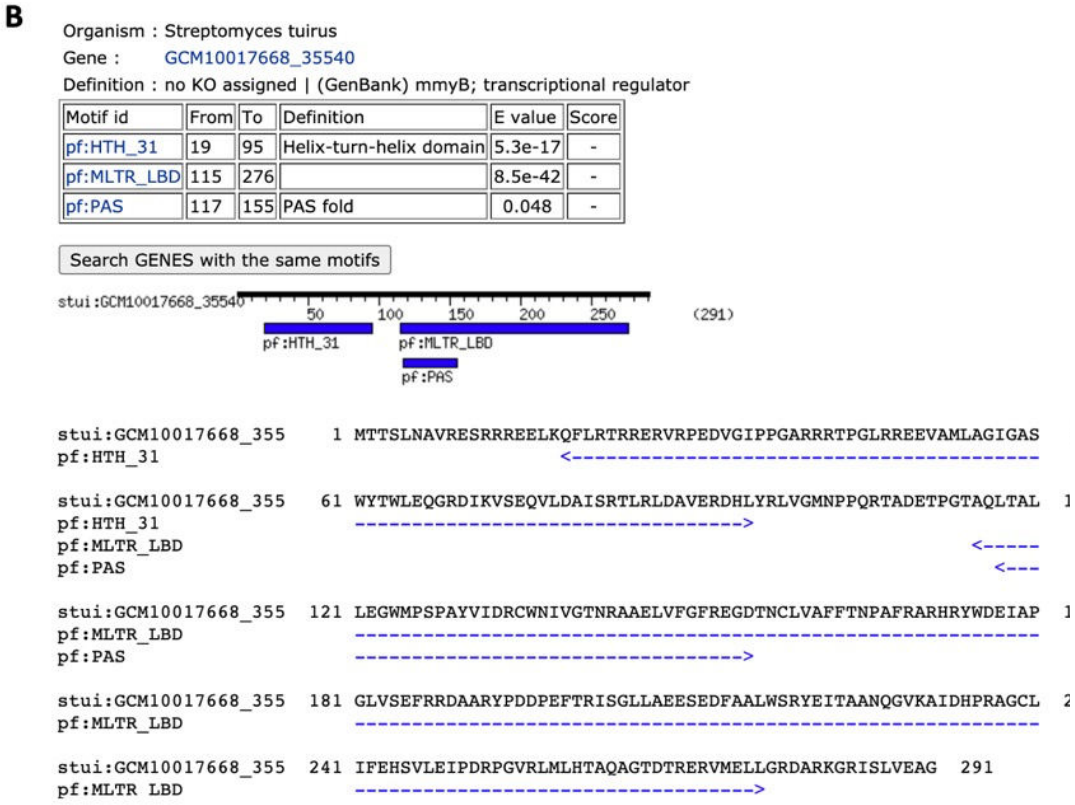


Figure 69. MmyB protein sequence and features. A) The *S. coelicolor* A3(2) *mmyB* gene and the codon translation into its protein sequence. It is 942 bp long translating into 313 amino acids. **B)** Conserved domains predicted in MmyB. Using KEGG MmyB is predicted to contain a helix-turn-helix domain (HTH) consistent with most transcriptional proteins to bind to DNA. MmyB is also thought to contain a PAS (Per-Arnt-Sim) fold which is also a feature of MltR's (a MmyB like protein) ligand binding domain (MLTR_LBD) and so this region may contain residues involved in ligand binding.

MmyB was subjected to protein BLAST analysis and the top 20 results are shown in Table 20. Using this BLAST search and analysis done previously[264], it is clear the majority of *mmyB* like genes are present in actinobacteria and more specifically in *Streptomyces*. If the role of MmyB is to be a transcriptional activator of the expression of natural product biosynthetic genes it would not be uncommon to see the majority of these in various *Streptomyces* species.

Table 20: BLAST Search of MmyB. Basic Local Alignment Search Tool (BLAST) was used to find proteins that are similar in amino acid sequence to MmyB with the top 20 being displayed. Results were sorted in descending order according to the percentage sequence identity (% identity) which indicates the similarity of two proteins. High % identity and coverage show a protein that's highly similar to the query protein. Many MmyB like proteins seem in the *Streptomyces* species and may be part of the XRE family.

Description	Species	% Query Cover	% Identity	NCBI accession number
DNA-binding protein, MmyB	*Streptomyces coelicolor* A3(2)	100	99.67	AAA98340.1
MULTISPECIES: helix-turn-helix domain-containing protein	*Streptomyces*	90	100	WP_189285527.1
putative DNA-binding protein, MmyB	*Streptomyces violaceoruber*	90	99.64	BAC76311.1
helix-turn-helix domain-containing protein	*Streptomyces rubrogriseus*	90	99.64	WP_164272945.1
helix-turn-helix domain-containing protein	*Streptomyces rubrogriseus*	90	99.28	NEC33576.1
helix-turn-helix domain-containing protein	*Streptomyces sp.* NRRL S-31	90	83.45	WP_079165416.1
helix-turn-helix domain-containing protein	*Streptomyces sp.* AC512_CC834	90	72.3	WP_217165416.1
helix-turn-helix domain-containing protein	*Amycolatopsis roodepoortenisis*	85	62.6	WP_192746516.1
helix-turn-helix domain-containing protein	*Streptomyces sp.* SA15	82	61.81	WP_095750842.1
helix-turn-helix domain-containing protein	*Streptomyces sp.* Ag109_G2-15	82	61.42	WP_097254996.1
	Streptomyces sp. 8L	82	61.02	MCA1217629.1
helix-turn-helix domain-containing protein	*Streptomyces bauhiniae*	82	60.63	WP_135785705.1
helix-turn-helix domain-containing protein	*Streptomyces sp.* MBT62	82	60.24	WP_200712036.1
helix-turn-helix domain-containing protein	*Streptomyces sp.* MBT53	82	60.24	WP_201050626.1
helix-turn-helix transcriptional regulator	*Streptomyces griseorubiginosus*	82	60.63	WP_208778112.1
MULTISPECIES: helix-turn-helix transcriptional regulator	unclassified *Streptomyces*	80	60.48	WP_223182843.1
helix-turn-helix domain-containing protein	*Streptomyces tsukubensis*	82	58.27	WP_179120106.1
helix-turn-helix domain-containing protein	*Streptomyces tsukubensis*	82	58.27	QFR94902.1
hypothetical protein B1H18_09655	*Streptomyces tsukubensis*	80	58.47	OON81143.1
transcriptional regulator with XRE_family HTH domain	*Amycolatopsis roodepoortenisis*	75	61.21	MBE1580103.1

Predictions on the structures of MmyB further indicate its role as a transcriptional protein as it contains a HTH motif necessary to bind to DNA at the N-terminal (Fig. 69B). This is also a feature of many XRE (Xenobiotic response element) family transcriptional proteins of which MmyB is thought to be a part of (indicated also by the BLAST search- Table 20). In terms of ligand binding domain predictions, MmyB has a PAS (Per-Arnt-Sim) fold domain at the C-terminal to sense small molecules

which is also a part of the 'MltR ligand binding' domain (Fig. 69B). These folds indicate binding to small ligands though the potential ligand for MmyB is unclear.

MltR (MmyB-like transcription regulator) is a distant homolog of MmyB from *Chloroflexus aurantiacus*. Its structure (PDB code 3pxp) was solved to 2.3 Å bound to a ligand, myristic acid (Fig. 70). MltR was crystallised as a homodimer and this dimer is considered relevant to the protein for DNA binding as seen in other XRE family proteins. While MltR is associated with fatty acid metabolism due to its structure being solved from co-crystallisation with myristic acid (Fig. 70), an ortholog of it in *S. griseus* was found to be involved in antibiotic control as are other closely related homologs. Therefore, the role of MltR cannot be confirmed and there is not sufficient evidence to suggest myristic acid is a true ligand for MltR. MltR is one of the only MmyB like proteins characterised to this extent to date and could provide insight into the structure of MmyB and therefore function.[264]

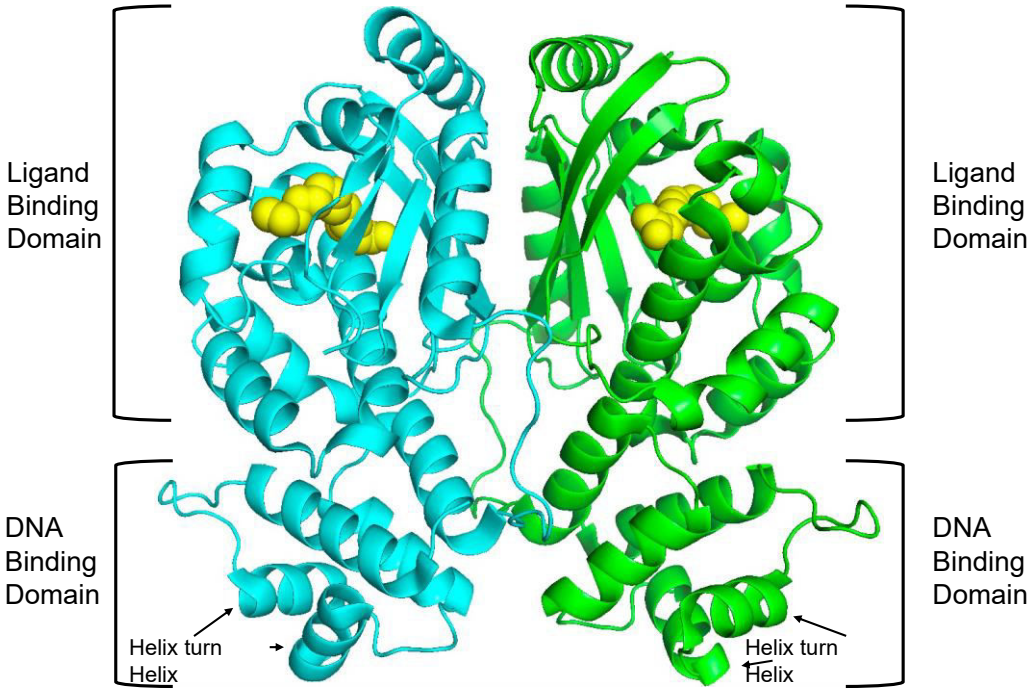


Figure 70. MltR homodimers bound to myristic acid. MltR (MmyB-like transcription regulator) was co-crystallized with myristic acid and the structure solved by Xu *et al.* (2012)[263] (pdb code: 3pxp). Shown are homodimers (each monomer coloured green or cyan) of the protein with myristic acid in yellow. The binding domains of each monomer is indicated with the helix-turn-helix (helix 3 and helix 4) for DNA binding indicated.

There are considered to be highly conserved residues between MltR and other MmyB like homologs predominantly in the ligand binding domain and in the interface between monomers.[264] This may mean these proteins function in a similar manner with the main difference being the exact small molecule ligand they bind and the region of DNA. The proposed model for MltR's functioning is that the apo-protein exists as a dimer or monomer in the cell until activated by ligand binding. Ligand binding may stabilise a MltR dimer allowing it to then bind to DNA. MmyB may also function in a similar manner where an unknown small ligand in response to cellular stress may bind to MmyB causing this complex to then bind to DNA and activate MmA production. However, inferences from MltR to MmyB must be made cautiously as the two do appear to be distinctly dissimilar (Fig. 67).

Structural and functionality data into MmyB could help to better characterise other MmyB like proteins that are homologous and aid in understanding the role of antibiotic production activators as well as the discovery of natural products.

5.2 MmyB produced in *E. coli* BL21 cannot fold correctly

The *E. coli* expression strain suggested for use with the pET151 plasmid and commonly used for large scale expression of non-toxic, non-membrane proteins is the BL21 DE3 (Star) strain (Chapter 2: Materials & Methods). The genome of this strain contains the T7 promoter under the control of the lac operon meaning expression of *mmyB* can be induced using IPTG only when cells are reaching exponential phase. This strain also does not have 2 key proteases that degrade heterologous proteins and so MmyB should theoretically be able to be overexpressed using this strain.[266]

5.2.1 Cloning *mmyB* into a pET151 plasmid vector for overexpression

The entire methylenomycin gene cluster is present in the C73-787 cosmid.[150] This includes the gene for MmyB. As *mmyB* is a *Streptomyces* based gene it is high in GC content which can make the cloning process difficult not only for PCR but for sequencing too. A high GC content primer can form non-specific binding complexes and thus make the amplification process difficult, not to mention high GC rich genomes tend to be more stable and require high temperatures for denaturing. Nonetheless, primers were able to be designed encompassing this *mmyb* gene and amplification of the gene was attempted (Fig. 71) (Chapter 2: Materials & Methods). The *mmyb* gene is 942 bp long and as shown in the gradient PCR (Fig. 71) a product close to 1 kilobase pair (kbp) was observed at a few temperatures.

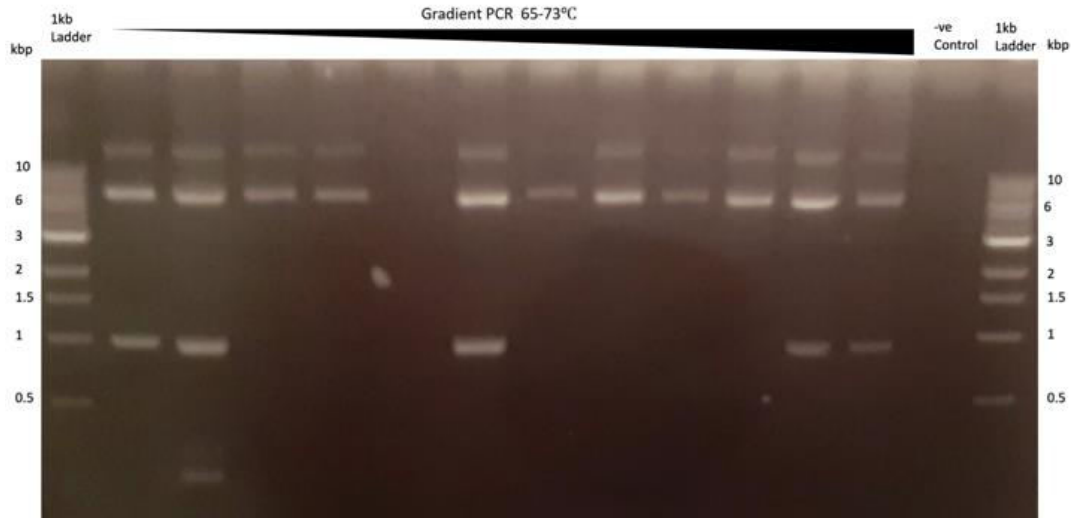


Figure 71. Gradient PCR to amplify *mmyB* shown on a 1% Agarose-TAE gel. A gradient PCR (65 ℃ - 75 ℃) was performed using Q5 Polymerase (Chapter 2: Materials & Methods) and the cosmid C73-787 as a template.[148] This *mmyB* region being amplified is 942 base pairs (bp) long. Wells 1,2,6, 11 and 12 contained the predicted product and were taken for insertion into the pET151 plasmid vector (Chapter 2: Materials & Methods)**.**

Confirmation of whether this was the correct gene was done after cloning. A gene product from the PCR at ~1 kbp (Fig. 71) was inserted into the pET151 expression vector (Chapter 2: Materials & Methods). Once cloned, this plasmid was transformed into *E. coli* TOP10 cells (Chapter 2: Materials & Methods) which retain the stability of the DNA for long term storage and give stable replications of the plasmid.

The DNA from colonies were sent for Sanger sequencing and unfortunately many showed the *mmyB* gene to be missing (Appendix Fig. C1). One potentially successful colony where the *mmyB* was detected, though the integrity was poor (Appendix Fig. C1), was taken and the gene presence was attempted to be confirmed by restriction digest (Chapter 2: Materials & Methods) (Fig. 72). The restriction enzyme EcoRV-HF was used on freshly harvested plasmids of *mmyb*-pET151 with expected band sizes at 869 bp, 1584 bp and 4237 bp. Only 2 of the 3 band sizes however were seen (Fig. 72) with the one that would contain *mmyB* missing (869 bp). This meant that this plasmid could not be confidently used to express MmyB.

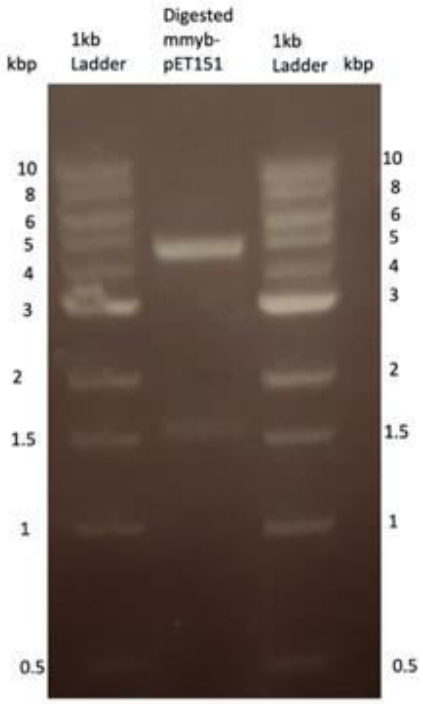


Figure 72. Enzymatic digest of pET151-*mmyB* on a 1% Agarose-TAE gel. The cloned pET151-*mmyB* plasmid using the amplified *mmyB* from Fig. 71. was digested with Eco-RV and was run on a 1% Agarose-TAE gel (Chapter 2: Materials & Methods). The predicted band sizes were 869 bp, 1584 bp and 4237 bp. A band around 5 kbp is strongly seen as is a fainter one around 1.5 kbp however no band is seen around 1 kbp which would contain the *mmyb* gene.

As an alternative to the native *mmyb* gene amplified, an *E. coli* codon optimised version was purchased from Genewiz (Chapter 2: Materials & Methods). This is the exact same amino acids that would be expressed by the native *mmyb* except the DNA codons are optimised for the tRNA present in *E. coli*. This lowers the GC

content and was therefore easier to clone into the pET151 plasmid (Fig. 73) (Chapter 2: Materials & Methods). Again, this was transformed into *E. coli* TOP10 first and the integrity of the codon optimised *mmyb* was validated by Sanger sequencing (Chapter 2: Materials & Methods and Appendix Fig. C2).

MmyB is 34.3 kDa protein however the pET151 plasmid has a histidine and V5 tag to allow for purification of the protein (Fig. 73). This makes the recombinant MmyB 38.1 kDa with an isoelectric point of 9.29 and instability index of 50.16 making it an unstable protein. The instability of recombinant MmyB is lowered from the native MmyB at 52.05, due to the tags present at the start of the protein though further tags could be added for stability. The extinction coefficient of recombinant MmyB is 43430 $M^{-1}cm^{-1}$ with 6 tryptophan and 7 tyrosine residues within it. As discussed for MmyJ the extinction coefficient is contributed to by tryptophan and tyrosine residues and this number can be used to determine the concentration of protein at 280 nm absorbance readings.

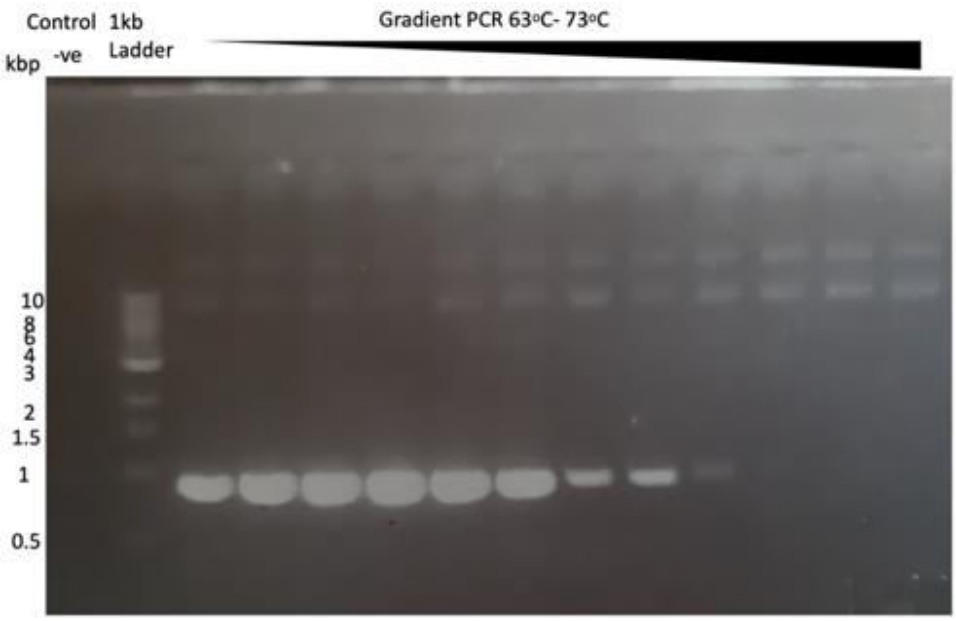
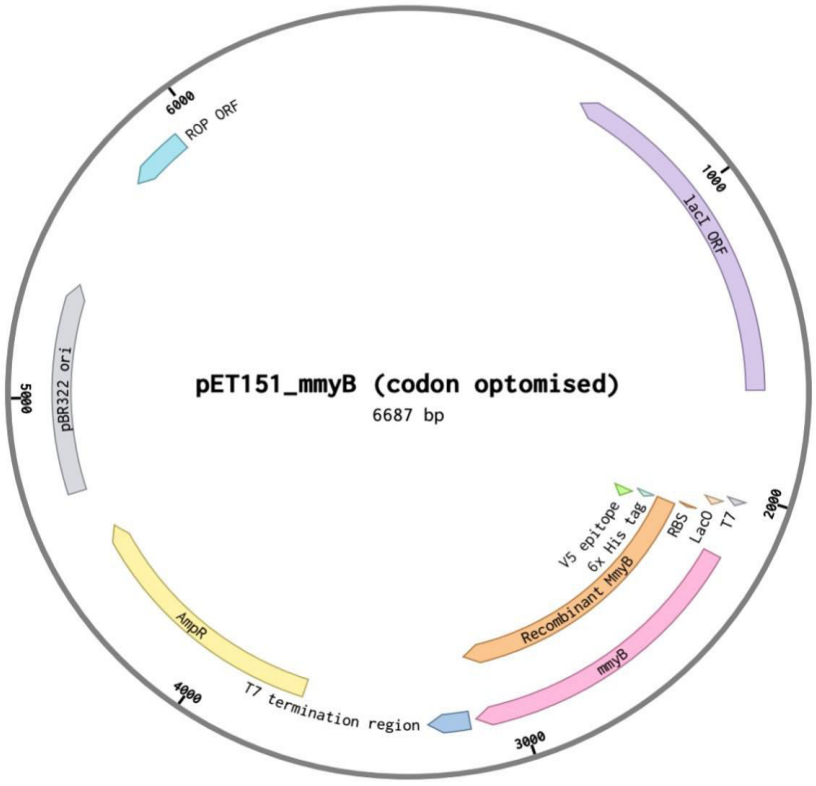
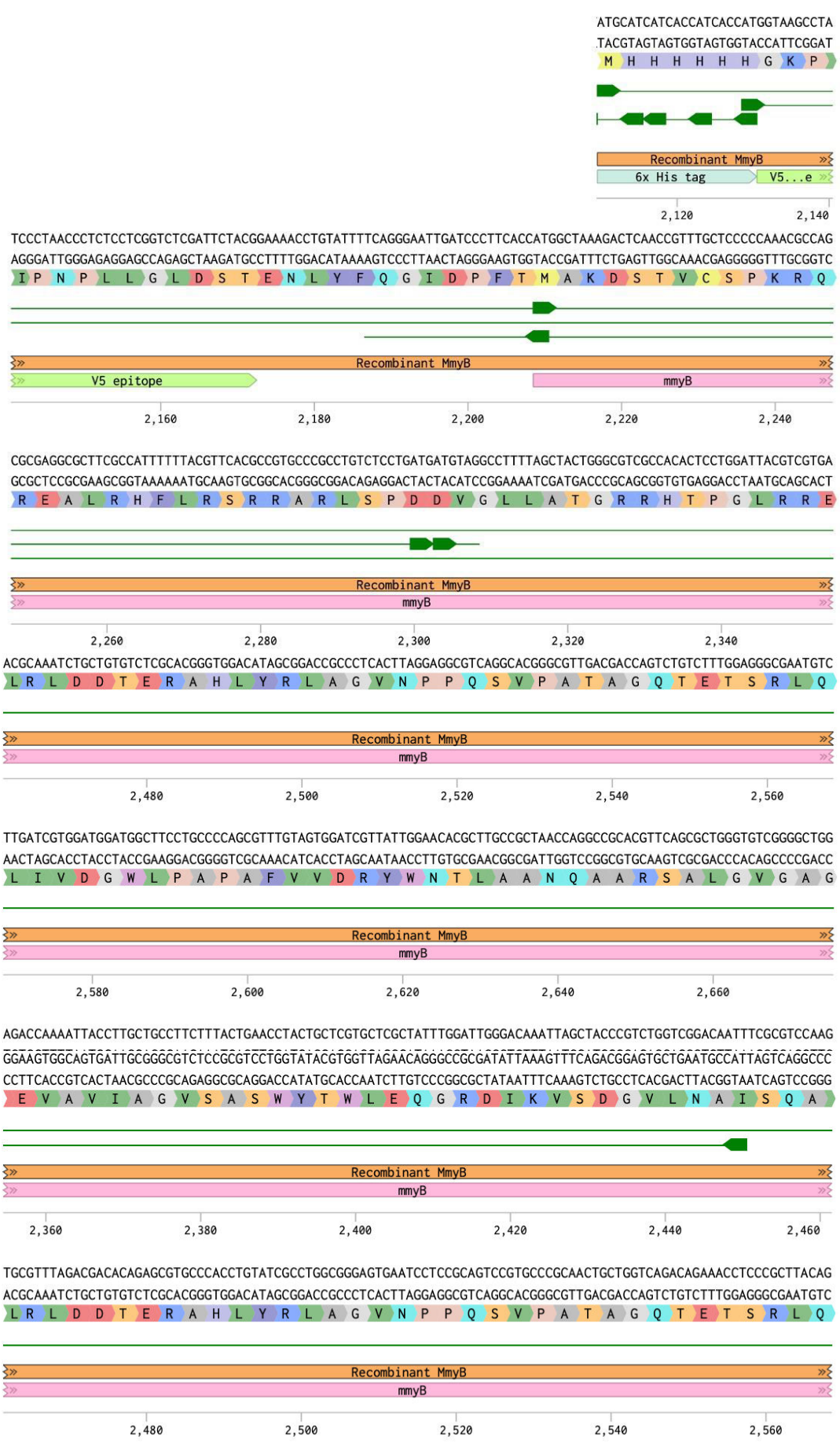


Figure 73. Gradient PCR on 1% Agarose-TAE of a colony containing pET151 with an *E. coli* codon optimized *mmyB* inserted. A pET151 plasmid was cloned using a synthetically bought *mmyB* gene with the codon optimized for *E. coli* (Chapter 2: Materials & Methods)**.** One of the colonies from an antibiotic selective LB plate was selected and subjected to colony PCR using primers for *mmyB*. The predicted gene product is 952 base pairs (bp). The negative control is the PCR with water rather than the colony DNA. Bands at ~ 900 bp is seen indicating the plasmid was cloned and inserted successfully. The plasmid from this colony was also sent for presence of *mmyB* by Sanger sequencing.

A



B


ATGCATCATCACCATCACCATGGTAAGCCTA
TACGTAGTAGTGGTAGTGGTACCATTCGGAT
M H H H H H H G K P
Recombinant MmyB
6x His tag
V5...e
2,120
2,140
TCCCTAACCCTCTCCTCGGTCTCGATTCTACGGAAAACCTGTATTTTCAGGGAATTGATCCCTTCACCATGGCTAAAGACTCAACCGTTTGCTCCCCCAAACGCCAG
AGGGATTGGGAGAGGAGCCAGAGCTAAGATGCCTTTTGGACATAAAAGTCCCTTAACTAGGGAAGTGGTACCGATTTCTGAGTTGGCAAACGAGGGGGTTTGCGGTC
I P N P L L G L D S T E N L Y F Q G I D P F T M A K D S T V C S P K R Q
Recombinant MmyB
V5 epitope
mmyB
2,160
2,180
2,200
2,220
2,240
CGCGAGGCGCTTCGCCATTTTTTACGTTCACGCCGTGCCCGCCTGTCTCCTGATGATGTAGGCCTTTTAGCTACTGGGCGTCGCCACACTCCTGGATTACGTCGTGA
GCGCTCCGCGAAGCGGTAAAAAATGCAAGTGCGGCACGGGCGGACAGAGGACTACTACATCCGGAAAATCGATGACCCGCAGCGGTGTGAGGACCTAATGCAGCACT
R E A L R H F L R S R R A R L S P D D V G L L A T G R R H T P G L R R E
Recombinant MmyB
mmyB
2,260
2,280
2,300
2,320
2,340
ACGCAAATCTGCTGTGTCTCGCACGGGTGGACATAGCGGACCGCCCTCACTTAGGAGGCGTCAGGCACGGGCGTTGACGACCAGTCTGTCTTTGGAGGGCGAATGTC
L R L D D T E R A H L Y R L A G V N P P Q S V P A T A G Q T E T S R L Q
Recombinant MmyB
mmyB
2,480
2,500
2,520
2,540
2,560
TTGATCGTGGATGGATGGCTTCCTGCCCCAGCGTTTGTAGTGGATCGTTATTGGAACACGCTTGCCGCTAACCAGGCCGCACGTTCAGCGCTGGGTGTCGGGGCTGG
AACTAGCACCTACCTACCGAAGGACGGGGTCGCAAACATCACCTAGCAATAACCTTGTGCGAACGGCGATTGGTCCGGCGTGCAAGTCGCGACCCACAGCCCCGACC
L I V D G W L P A P A F V V D R Y W N T L A A N Q A A R S A L G V G A G
Recombinant MmyB
mmyB
2,580
2,600
2,620
2,640
2,660
AGACCAAAATTACCTTGCTGCCTTCTTTACTGAACCTACTGCTCGTGCTCGCTATTTGGATTGGGACAAATTAGCTACCCGTCTGGTCGGACAATTTCGCGTCCAAG
GGAAGTGGCAGTGATTGCGGGCGTCTCCGCGTCCTGGTATACGTGGTTAGAACAGGGCCGCGATATTAAAGTTTCAGACGGAGTGCTGAATGCCATTAGTCAGGCCC
CCTTCACCGTCACTAACGCCCGCAGAGGCGCAGGACCATATGCACCAATCTTGTCCCGGCGCTATAATTTCAAAGTCTGCCTCACGACTTACGGTAATCAGTCCGGG
E V A V I A G V S A S W Y T W L E Q G R D I K V S D G V L N A I S Q A
Recombinant MmyB
mmyB
2,360
2,380
2,400
2,420
2,440
2,460
TGCGTTTAGACGACACAGAGCGTGCCCACCTGTATCGCCTGGCGGGAGTGAATCCTCCGCAGTCCGTGCCCGCAACTGCTGGTCAGACAGAAACCTCCCGCTTACAG
ACGCAAATCTGCTGTGTCTCGCACGGGTGGACATAGCGGACCGCCCTCACTTAGGAGGCGTCAGGCACGGGCGTTGACGACCAGTCTGTCTTTGGAGGGCGAATGTC
L R L D D T E R A H L Y R L A G V N P P Q S V P A T A G Q T E T S R L Q
Recombinant MmyB
mmyB
2,480
2,500
2,520
2,540
2,560

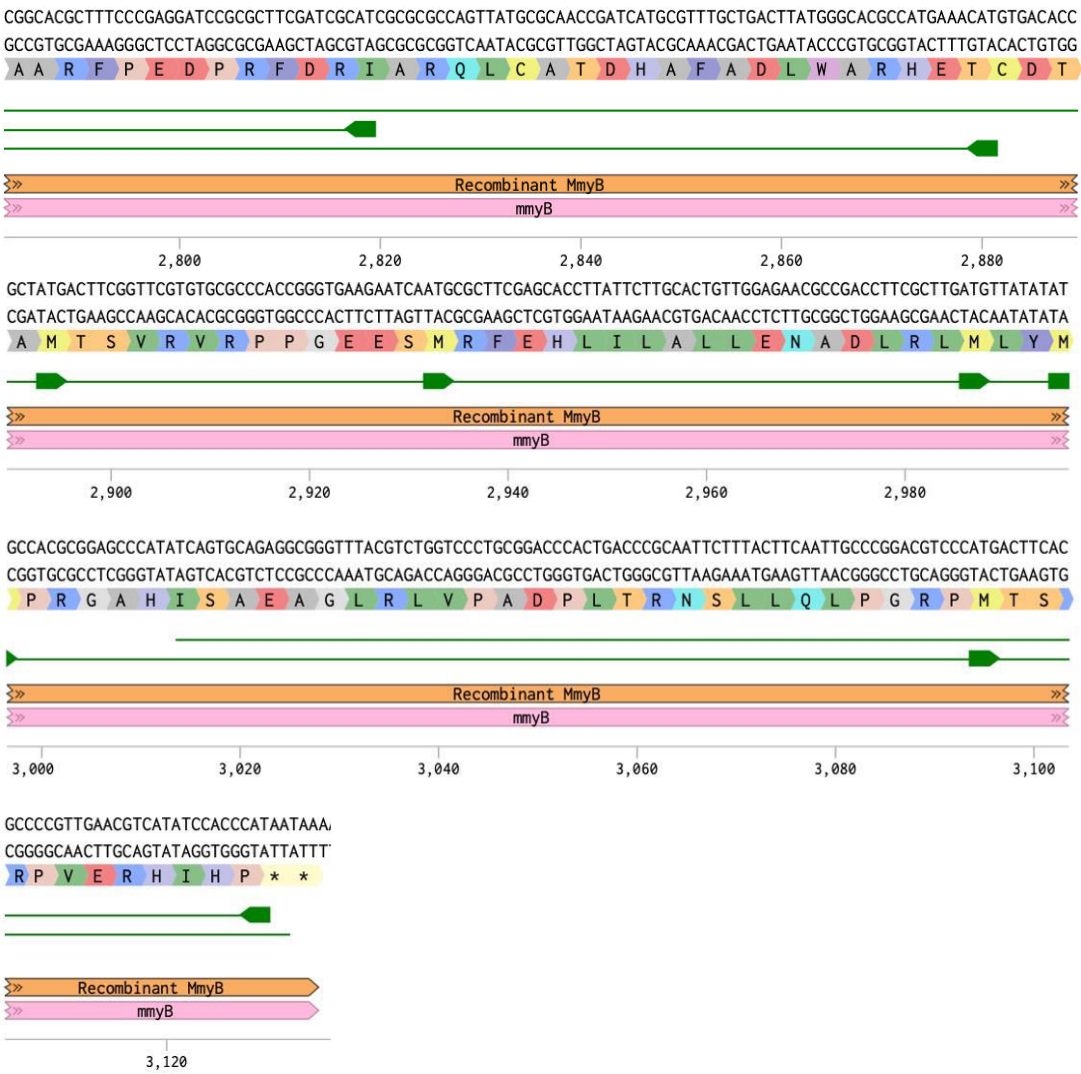


Figure 74. Recombinant MmyB used for expression of MmyB. A) The recombinant pET151-MmyB plasmid with the main features highlighted. This plasmid was created using a commercial pET151 kit and a synthetic codon-optimized *mmyB* gene (Fig. 73) (Chapter 2: Materials & Methods). This plasmid contains an ampicillin resistance (AmpR) gene to select for transformed colonies and the lac operon for induced induction of protein production by addition of IPTG. **B)** The recombinant MmyB consists of 6 Histidines (6xHis tag), a V5 tag, a TEV site and the *E. coli* codon optimized MmyB sequence. Recombinant MmyB is 340 amino acids long, has a molecular weight of 38134.99 Daltons (Da), an isoelectric point of 9.29 and an instability index score of 50.16.

5.2.2 Production of MmyB in *E. coli* BL21 is low and does not purify well

The *E. coli* codon optimised pET151-*mmyb* was transformed into *E. coli* BL21 DE3(Star) (Chapter 2: Materials & Methods) and integrity of the *mmyB* gene transformed was confirmed by Sanger sequencing (Appendix Fig. C3) to be of good quality before continuation of expression and production of MmyB.

Small scale expression trials with 50 mL cultures (Chapter 2: Materials & Methods) indicated the presence of some strongly expressed proteins (Fig. 75) however it was not clear whether any of these were MmyB. Therefore, the culture volume was increased to 1 L (Chapter 2: Materials & Methods) and purification was done first by IMAC (Chapter 2: Materials & Methods). Through this nickel purification 3 strong bands around 40 kDa, 60 kDa and 80 kDa were seen on SDS-PAGE gels (Fig. 76).

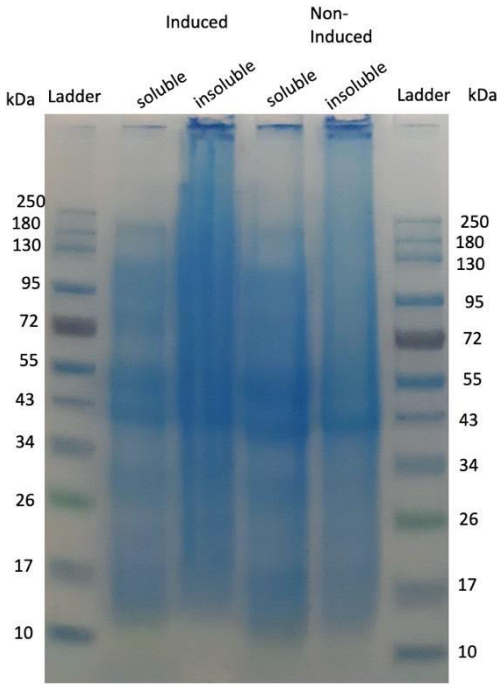


Figure 75. 50 mL cultures of BL21 pET151-*mmyB* was tested for expression of MmyB. BL21 cells with pET151-*mmyB* were grown for 3 hours at 37 ℃ and were either induced with 1 mM IPTG or not induced ('non-induced'). Cells were lysed and fractioned with samples taken from the 'soluble' fraction, containing cytosolic soluble proteins, and the 'insoluble' fraction, containing insoluble proteins and cell debris was run on a 4-20% SDS-PAGE gel. Recombinant MmyB weighs 38.15 kilo Dalton (kDa) and possible faint bands are seen around this weight.

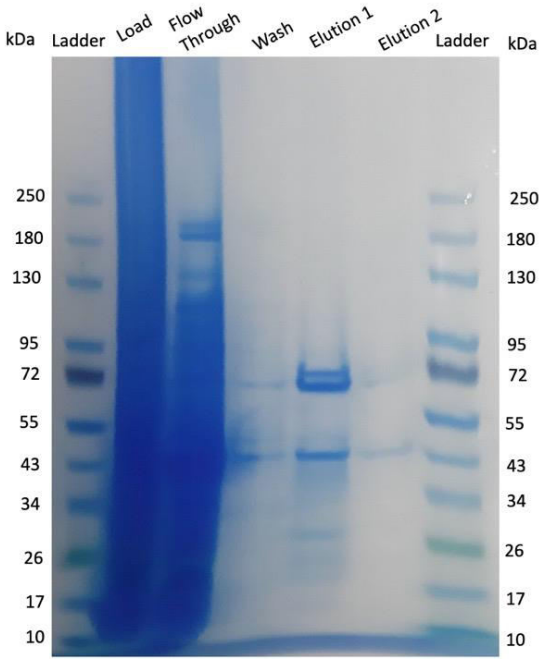


Figure 76. Recombinant MmyB protein purification by immobilized metal affinity chromatography (IMAC) specifically nickel-NTA. A 1 L culture of recombinant MmyB (Fig. 74) from a BL21 *E. coli* host was purified through IMAC. The 'Wash' buffer contained a low concentration of imidazole and the elution buffer contained a high concentration of imidazole (Chapter 2: Materials & Methods). FT refers to the flow through which is the elution after loading the column. Load refers to the sample loaded onto the column (the crude). Elution 1 is the elution after 2 minutes of incubation with 3 mL of 'Elution' buffer added. Elution 2 is the elution after further elution buffer was added (Chapter 2: Materials & Methods). Samples were run on a 4-20% SDS-PAGE gel. The recombinant MmyB has a weight of 38.15 kilo Daltons (kDa).

It is possible that recombinant MmyB runs on the SDS-PAGE slower than expected meaning it would show as a higher weight than it should be. If this is true, the 40 kDa band could be monomeric MmyB. In which case either the 60 kDa or 80 kDa band may be dimeric MmyB meaning one of those two is also likely a contaminant protein. In the expression of recombinant MmyJ an 80 kDa contaminant protein was also observed (Fig. 19) during IMAC purification. The expression systems are identical with the IMAC purification buffers also being very similar (Chapter 2: Materials & Methods) and so the 80 kDa band seen in Fig. 76 for the purification of

recombinant MmyB is likely also a contaminant protein. This could mean the 60 kDa proteins are either strong dimers of MmyB or another contaminant protein.

To separate these different sized proteins, size exclusion chromatography (SEC) was used (Fig. 77A). One distinct peak was seen with a shoulder before it and some smaller peaks after which were run on SDS-PAGE gels (Fig. 77B). Only the largest first peak from the SEC was able to be visualised on the SDS-PAGE gel which mirrors the elution band seen from the IMAC. Three protein bands are still seen in this fraction indicating perhaps that some sort of complex being formed by 3 proteins. Alternatively, it may be that these are multimeric forms of that same protein or a combination of the two.

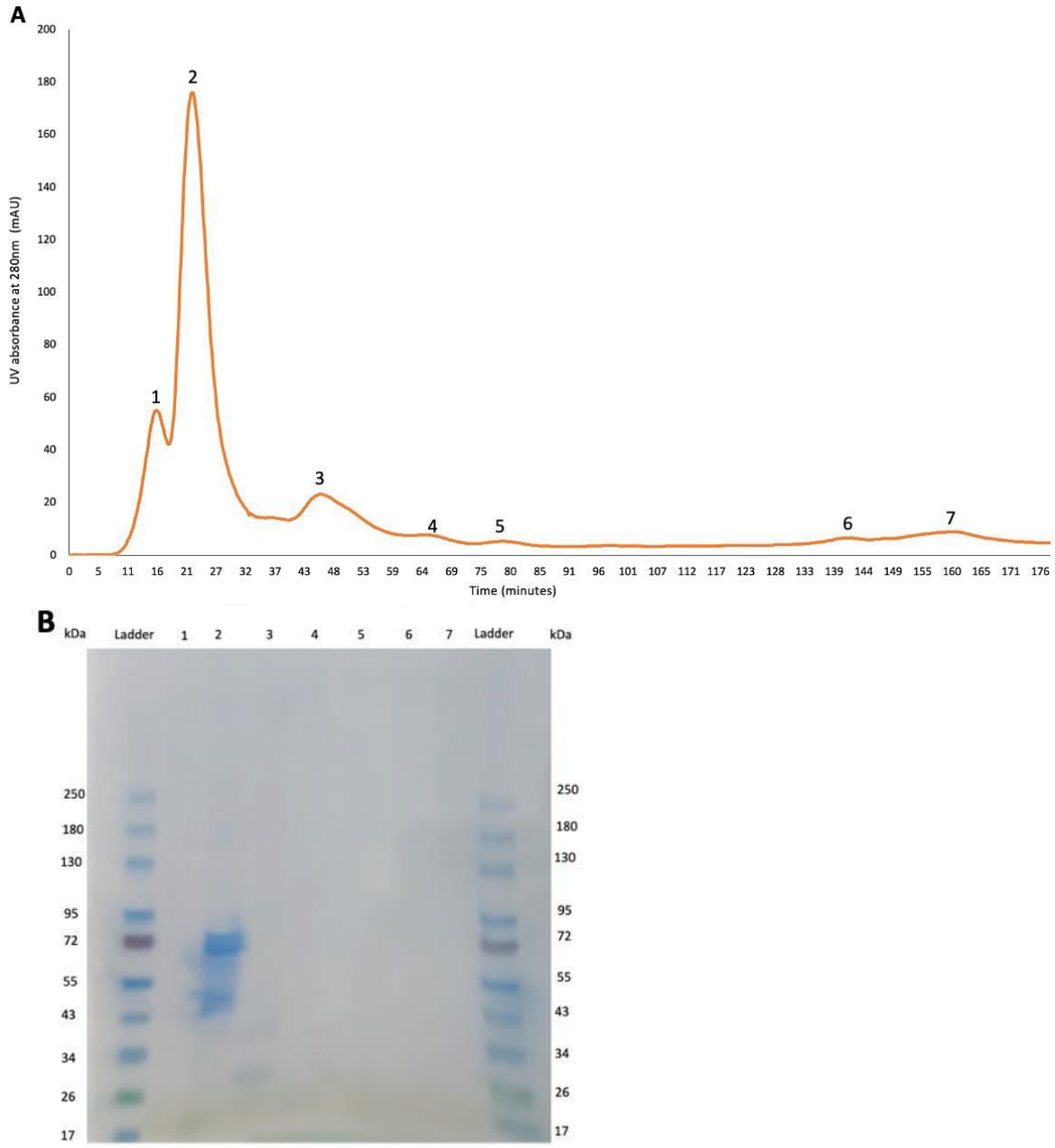
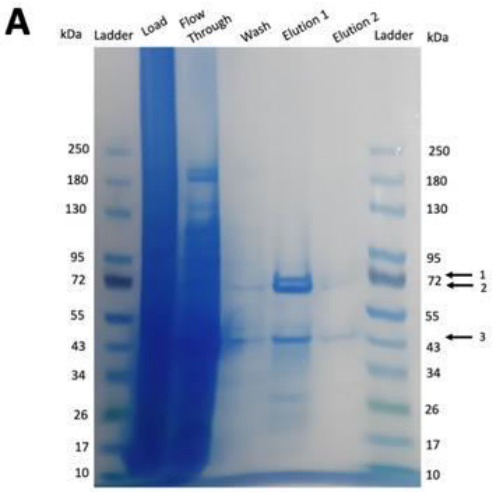


Figure 77. Size exclusion chromatography of recombinant MmyB. A) IMAC purified recombinant MmyB (Fig. 75) was run through a size exclusion chromatography column (Superdex 75pg) (Chapter 2: Materials & Methods) to separate MmyB from contaminant proteins (Fig. 76). **B)** Fractions of the major peaks were taken and run on SDS-PAGE gels to estimate protein sizes of peaks in the size exclusion (**A)**. Only the largest peak (2) was seen which consisted of the same proteins shown by the IMAC (Fig. 76) meaning these proteins are possibly aggregating together in a complex.

5.2.3 MmyB in *E. coli* BL21 purifies with chaperones highlighting difficulties in folding

To determine if the SEC purification is of one protein or a complex (and if so what the proteins in that complex are), targeted proteomics was conducted. The 3 distinctive bands seen during purification at 80 kDa, 60 kDa and 40 kDa were excised and subjected to mass spectrometry analysis (Chapter 2: Materials & Methods) (Fig. 78). The percentage of total spectra states the percentage of the total mass spectra of peptides that is assigned to a particular protein. The higher this percentage is the more likely this is the protein present in the sample.

Fig. 78 shows the percentage of total spectra of the 3 samples, with sample 1 being the 80 kDa protein, 2 being 60 kDa protein and 3 being 40 kDa protein. Starred and highlighted is MmyB. Clearly in samples 1 and 2 of the higher kDa bands there appears to be a very low percentage of peptides that could be assigned to MmyB in comparison to other proteins. In the 3rd sample at 40 kDa, there are more peptides of MmyB however it cannot be said that this sample is purely MmyB as there are peptides of other proteins in just as high or higher percentages than MmyB. Nonetheless this sample at around 40 kDa likely does contain MmyB though not in the purest form of the protein.



B

Probability Legend:
over 95%
80% to 94%
50% to 79%
20% to 49%
0% to 19%

#	Visible?	Starred?	Bio View: 147 Proteins in 139 Clusters	Accession Number	Alternate ID	Molecular Weight	Protein Grouping Ambiguity	MMYB1	MMYB2	MMYB3
								MMYB1	MMYB2	MMYB3
1	☑	☆	Cluster of 60 kDa chaperonin OS=Escherichia coli (strain K12) GN=groL PE=1 SV=2 (P0A6F5)	P0A6F5	groL	57 kDa	★	5.2%	8.3%	3.9%
2	☑	☆	Bifunctional polymyxin resistance protein ArnA OS=Escherichia coli (strain K12) GN=arnA PE=1 S...	P77398	arnA	74 kDa		4.5%	1.9%	2.0%
3	☑	☆	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] OS=Escherichia coli (strain K...	P17169	glmS	67 kDa		1.6%	1.1%	0.81%
4	☑	☆	Cluster of CON__P04264	CON__P042...		?	★	0.64%	0.64%	2.1%
5	☑	★	eptomyces coelicolor (strain ATCC BAA-471 / A3(2) / M145) OX=100226 GN=mmyB PE=4 SV=1	Q9ACS4 (+1)	mmyB	35 kDa		0.20%	0.13%	1.2%
6	☑	☆	Chaperone protein DnaK OS=Escherichia coli (strain K12) GN=dnaK PE=1 SV=2	P0A6Y8	dnaK	69 kDa	★	1.2%	0.37%	0.42%
..	☑	☆	Cluster of CON__P13645	CON__P136...		?	★	0.46%	0.36%	1.0%
8	☑	☆	CON__P00761	CON__P007...		?	★	0.22%	0.19%	0.62%
9	☑	☆	CON__P35527	CON__P355...		?	★	0.11%	0.16%	0.46%
10	☑	☆	Chaperone protein DnaJ OS=Escherichia coli (strain K12) GN=dnaJ PE=1 SV=3	P08622	dnaJ	41 kDa		0.067%	0.01...	0.40%
11	☑	☆	DNA gyrase subunit B OS=Escherichia coli (strain K12) GN=gyrB PE=1 SV=2	P0AES6	gyrB	90 kDa		0.059%	0.07...	0.35%
12	☑	☆	30S ribosomal protein S1 OS=Escherichia coli (strain K12) GN=rpsA PE=1 SV=1	P0AG67	rpsA	61 kDa		0.24%	0.13%	0.048%
13	☑	☆	Bifunctional protein PutA OS=Escherichia coli (strain K12) GN=putA PE=1 SV=3	P09546	putA	144 kDa		0.084%	0.09...	0.18%
14	☑	☆	Cluster of DNA gyrase subunit A OS=Escherichia coli (strain K12) GN=gyrA PE=1 SV=1 (P0AES4)	P0AES4	gyrA	97 kDa	★	0.15%	0.11%	0.14%
15	☑	☆	Cluster of CON__P02768-1	CON__P027...		?	★	0.14%	0.04...	0.16%

Figure 78. IMAC purifications of recombinant MmyB was subjected to mass spectrometry analysis. The consistently seen bands during IMAC purification of MmyB (Fig. 76) was excised (**A**) and subjected to mass spectrometry analysis (**B**). Bands are seen usually during purification at ~80 kDa (sample 1), ~ 60 kDa (sample 2) and at ~ 40 kDa (sample 3) (**A**). The peptides detected after tryptic digest in each sample is shown in **B** by the percentage of total spectra. This is the percentage of peptides assigned to these proteins in the total spectrum seen. Highlighted is MmyB. Low percentages of MmyB are seen in samples 1 and 2 (80 and 60 kDa) but more is seen in the 3rd sample (40 kDa). The majority of proteins in the samples seen to be proteins from *groL, arnA* and *glmS* genes.

In all 3 samples, there are consistently high amounts of peptides detected from proteins of the genes *groL*, *arnA* and *glmS*. The gene for *groL* encodes for a common *E. coli* chaperone protein GroEL which is involved in the refolding of misfolded proteins, using ATP and its co-chaperone GroES. GroES however was not detected in the mass spectrometry data. One reason for the high percentage of GroEL may be accounted for by the knowledge that GroEL forms a multimeric complex of 7 monomers in a ring stacked by another 7 monomers in a ring.[267] The gene *arnA* encodes the ArnA protein which is an enzyme involved in the modification of arabinose which attaches to Lipid A to allow for resistance to antimicrobial molecules such as polymyxin.[268] While MmyB is not an antibiotic or considered toxic, production of ArnA by *E. coli* may be indicative of *E. coli* recognising MmyB as an antibiotic related product. GlmS is also a common *E. coli* contaminant protein seen during IMAC with Ni-NTA. While this protein does not

have a high histidine content, the histidine's present are on the surface of the protein meaning it can compete with histidine tagged recombinant proteins during purification procedures.[269] Adjusting the pH and ensuring greater expression of MmyB may reduce GlmS binding to nickel during purification.

Clearly expression in *E. coli* BL21 of MmyB is possible however purification of the protein was not. This may be due to MmyB being an unstable protein and there being errors when folding causing it to form a complex with chaperones such as GroEL or it may be the BL21 system itself is not an ideal environment for MmyB. The simplest solution in such case would be to trial a different *E. coli* strain before fine tuning purification methods or redesigning the recombinant MmyB.

5.3 MmyB can be overexpressed and purified using *E. coli* C43

While MmyB is not considered a toxic protein, expression trials in BL21 did indicate a variable degree of success. It may be the case that within the BL21 strain, MmyB is being treated as a toxic molecule and thus expression levels are low. An alternative *E. coli* strain would be the 'Overexpress' C43 (DE3) cells (Merck). This C43 strain is designed specifically to allow for expression of membrane and toxic molecules from all organisms and is therefore far more robust which will hopefully assist in high expression levels of more stable MmyB.[270]

5.3.1 Overexpression of MmyB in C43 is better than in BL21

Similar to the BL21 expression cells, the *E. coli* C43 cells were also transformed with the pET151-*mmyB* that contained the codon optimised version of *mmyB* (Chapter 2: Materials & Methods). A colony containing a high integrity *mmyb* gene as determined by sequencing data (Appendix Fig. C4) was picked and used as a working stock. As conducted previously with BL21 cells, an IMAC purification was done with the C43 pET151-*mmyB* cells also. One litre cultures of BL21 and C43

pET151-*mmyB* cells were grown and purified as before with conditions kept the same for comparison of the two cultures (Chapter 2: Materials & Methods) (Fig. 79).

Black arrows in Fig. 79 indicate on the BL21 elution fraction the MmyB bands seen previously in SEC purification and the bands that were subjected to targeted proteomics (Fig. 78). The same bands can also be seen in the C43 elution fraction. While the Ni-NTA purification appears cleaner for the BL21 cell culture during the elution stage with there being less contaminant proteins, the MmyB band (~ 40 kDa) is far stronger in the C43 culture indicating a greater expression of MmyB.

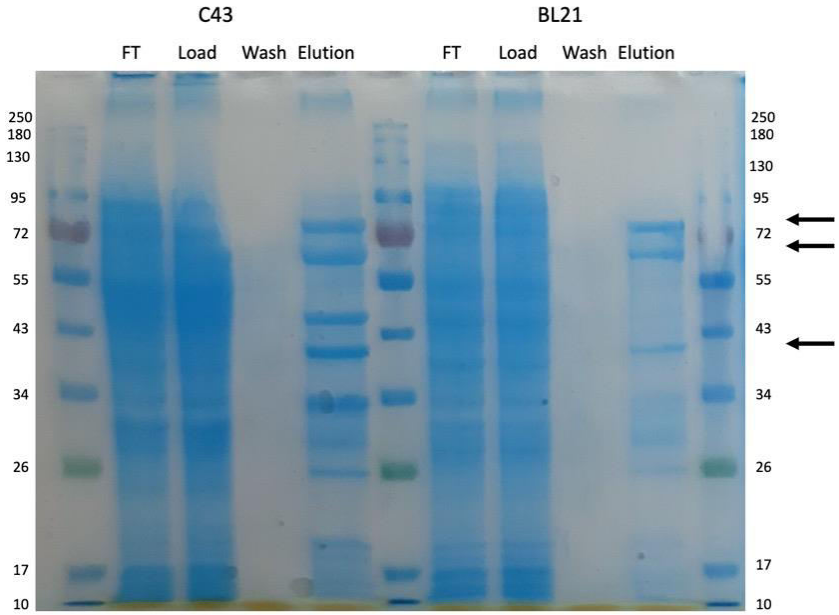


Figure 79. Recombinant MmyB protein purification from BL21 and C43 strains by immobilized metal affinity chromatography (IMAC) specifically nickel-NTA. A 1 L culture each of recombinant MmyB (Fig. 74) from BL21 and C43 *E. coli* hosts were purified through IMAC. The 'Wash' buffer contained a low concentration of imidazole, and the elution buffer contained a high concentration of imidazole (Chapter 2: Materials & Methods). FT refers to the flow through which is the elution after loading the column. Load refers to the sample loaded onto the column (the crude). Elution is the elution after 2 minutes of incubation with 3 mL of 'Elution' buffer added. Samples were run on a 4-20% SDS-PAGE gel. The recombinant MmyB has a weight of 38.15 kDa. Labelled in black arrows are the 3 bands seen previously as well in BL21 strains. The same bands are seen in C43 as well as others however the 3rd band containing more MmyB (Fig. 78) is stronger in C43 compared to BL21.

The IMAC buffers could be improved upon and optimised for the C43 cultures however as the same bands are seen in this cell type as BL21, it's unlikely that buffer optimising alone would separate out the ~ 80 kDa and ~ 60 kDa proteins.

GroEL is a chaperone and the removal of chaperones during purification was a problem faced by many others during recombinant protein purification. In such case, the first suggestion to retain the desired protein's functionality is to conduct an ATP wash step. As GroEL is a potassium using ATPase, the ATP wash should contain K^+ ions.[271] Many other chaperones use Mg^{2+} ions for their activity and while other chaperones were not detected during mass spectrometry analysis (Fig. 78), magnesium was also included in the buffer. Not only this but ArnA is repressed by bivalent cations. While the usual method is to express the protein in a BL21 (DE3) or similar strain with $MgCl_2$ added to the culture medium, addition of Mg^{2+} in the wash buffer may assist in eliminating some ArnA if it is present.[272] Considering the higher expression levels of what is likely MmyB, this is presumably enough to outcompete GlmS binding to the nickel.

Therefore, this ATP wash step (ATP buffer was based on previous chaperone studies [271,273]), was conducted within the Ni-NTA purification (Chapter 2: Materials & Methods) for both C43 and BL21 cell pellets (Fig. 80). As seen in Fig. 80 for the C43 cell type this has yielded far cleaner elution fractions with a singular prominent band around 40 kDa. However, in the BL21 pellet this was not observed, and the ATP wash has done little to gain a pure elution fraction likely indicating lower expression of MmyB and that which is expressed being bound strongly to chaperones.

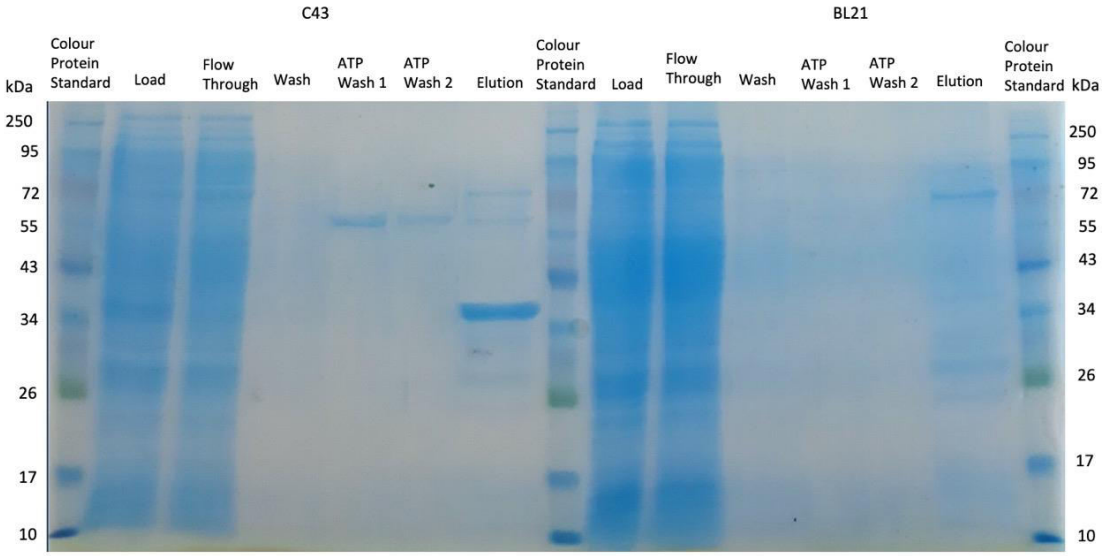


Figure 80. Recombinant MmyB protein purification from C43 and BL21 strains by immobilized metal affinity chromatography (IMAC) specifically nickel-NTA including an ATP Wash. A 1 L culture each of recombinant MmyB (Fig. 74) from *E. coli* C43 and BL21 was purified through IMAC. The 'Wash' buffer contained a low concentration of imidazole and the elution buffer contained a high concentration of imidazole (Chapter 2: Materials & Methods). FT refers to the flow through which is the elution after loading the column. Load refers to the sample loaded onto the column (the crude). ATP washes were done using a buffer with ATP in it (Chapter 2: Materials & Methods). ATP Wash 1 refers to elution after 1 CV (10 mL) of ATP Buffer and ATP Wash 2 refers to the elution after another addition of 1 CV (10 mL) of the ATP Buffer. Elution is the elution after 2 minutes of incubation with 3 mL of 'Elution' buffer added. Samples were run on a 4-20% SDS-PAGE gel. The recombinant MmyB has a weight of 38.15 kilo Daltons (kDa). A strong singular band ~ 34 kDa is seen in the C43 strain after purification.

Before further experimentation to obtain pure MmyB from C43 cells using IMAC purification with ATP washes, the prominent band seen at ~ 40 kDa was subjected to mass spectrometry to confirm presence of MmyB.

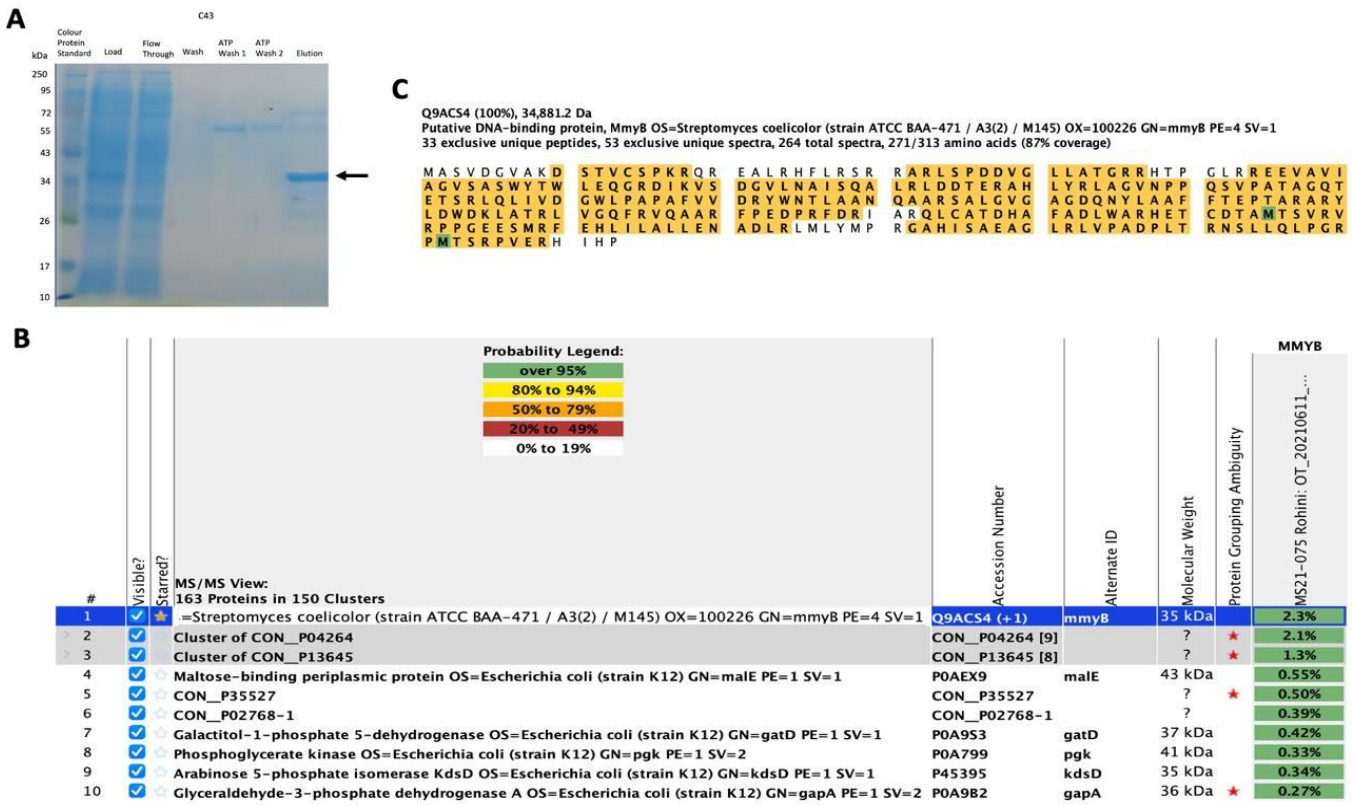


Figure 81. IMAC purifications of recombinant MmyB was subjected to mass spectrometry analysis. The band at ~ 34 kilo Daltons (kDa) seen during IMAC purification of MmyB was excised (**A**) and subjected to mass spectrometry analysis (**B**). After using ATP containing buffers for IMAC in the C43 strain (Fig. 80), a singular strong band was seen which was excised and peptides digested with trypsin. The peptides detected are shown in **B** by the percentage of total spectra in descending order. This is the percentage of peptides assigned to these proteins in the total spectrum seen. Highlighted is MmyB. Most of the sample contains MmyB. The sequence coverage (**C**) of this MmyB is high at 87% meaning the peptides identified cover the majority of the protein sequence.

Targeted proteomics on this band was conducted as done previously (Fig. 78) (Chapter 2: Materials & Methods) and results are seen in Fig. 81. The percentage of the total spectra of peptides detected this time has the greatest majority being assigned to MmyB (Fig. 81B) with minimal contamination from other proteins but more importantly with no detection of GroEL or ArnA. To ensure a select number of peptides from MmyB are not being represented disproportionately and that peptides from a large majority of the protein is seen, the sequence coverage was also viewed (Fig. 81C). This indicates that 87% of the MmyB sequence was detected making it highly likely that this fraction is almost entirely MmyB.

5.4 Recombinant MmyB is unstable and precipitates

With pure MmyB having been produced reliably in C43 cells (Fig. 80), analysis of the functionality of MmyB was conducted. Similar to MmyJ, MmyB is also predicted to function as homodimers. To test this and to further ensure the elution fraction from the IMAC purification is pure, the fraction was subjected to SEC (Chapter 2: Materials & Methods). As the fraction is run over time under non-denaturing conditions the functionality of the protein can be observed.

SEC using MmyB is shown in Fig. 82A with multiple peaks being observed. To estimate the size of these peaks and thus the proteins within them, samples of the peaks were taken and run on an SDS-PAGE gel (Fig. 82B). The last few peaks (Fig. 82A peaks 3 and 4) appear to have proteins around 40 kDa and below. Multiple bands below 40 kDa are likely fractions of denatured MmyB. The protein sizes in the first peak (Fig. 82A peak 1) could not be detected. However, as it elutes before the second peak this protein or protein complex is of a greater size than those in the second peak which means this fraction is unlikely to be pure MmyB.

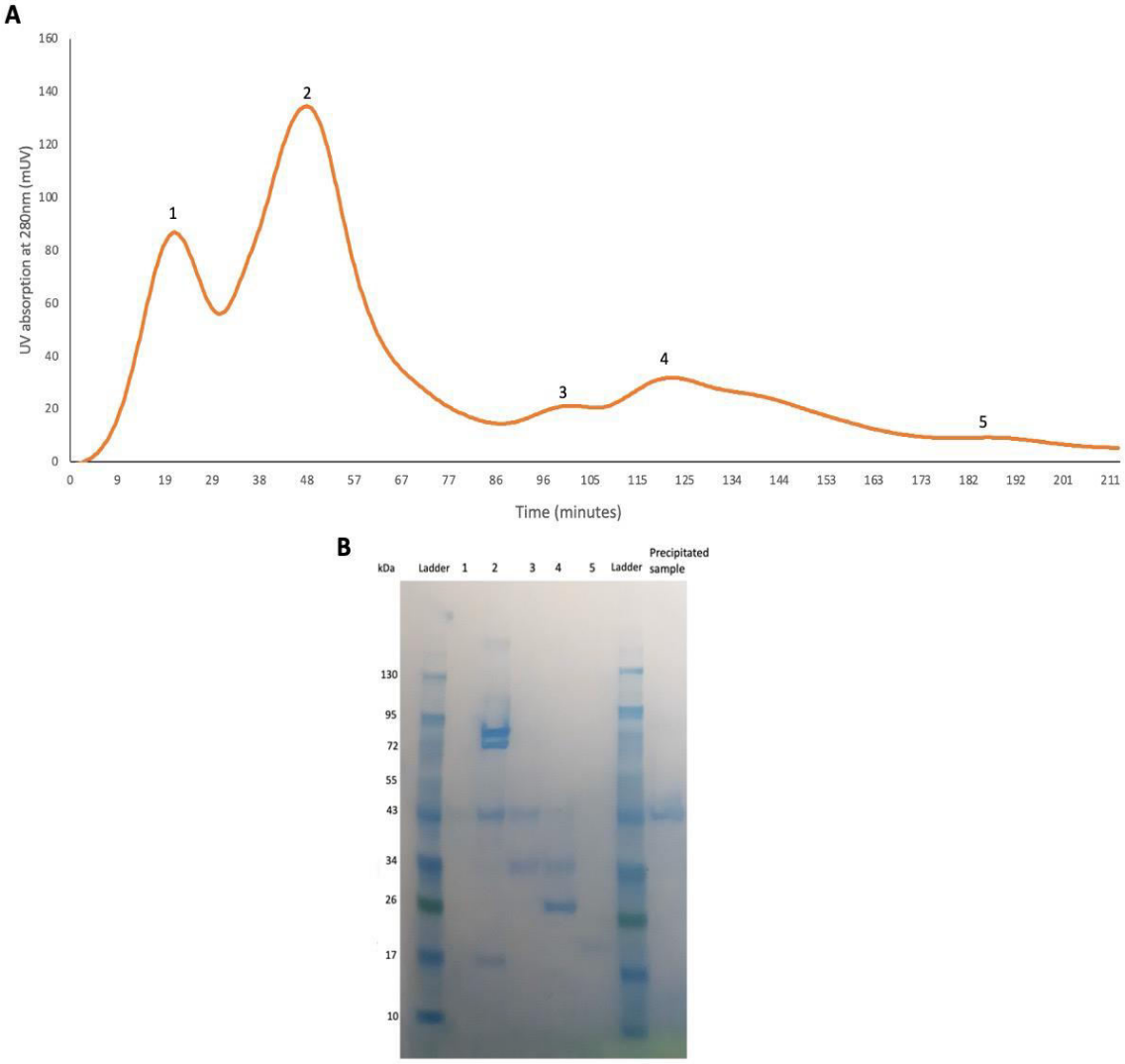


Figure 82. Size exclusion chromatography of recombinant MmyB. A) IMAC purified recombinant MmyB from a C43 strain (Fig. 80) was run through a size exclusion chromatography column (Chapter 2: Materials & Methods) to ensure MmyB is pure and for functional study. **B)** Fractions of the major peaks were taken and run on 4-20% SDS-PAGE gels to estimate protein sizes of peaks in the size exclusion (**A)**. Peak 1 and 5 were too dilute to detect on the SDS-PAGE. Peak 2 appears to consist of the same contaminant proteins seen previously (Fig. 76) with peaks 3 and 4 seemingly to be degraded proteins and peptides. This may be explained as the sample prior to loading on the size exclusion column did have a lot of precipitation ('precipitated sample') which appears to contain MmyB.

The strongest peak (Fig. 82A peak 2) clearly contains the same bands observed during IMAC purification (Fig. 79) prior to ATP washes and is therefore most probably a complex formed of chaperones and MmyB. While the ATP wash step did appear to eliminate most of these contaminant proteins (Fig. 80) it may be that they are still present and bound to some MmyB but are overrepresented in the SEC due to a reduced amount of the pure MmyB being loaded. This is due to the elution sample prior to being loaded on the SEC being centrifuged. This step did contain a large amount of precipitation. A sample of this precipitate was run also on the SDS-PAGE (Fig. 82B) and predictably a single band at ~ 40 kDa is seen which is most probably MmyB.

Purification buffers contained 10% glycerol to assist in stabilising MmyB though this is clearly not sufficient. The buffer composition to retain MmyB will therefore need finetuning. Furthering that the recombinant MmyB may need to be redesigned with additional tags to increase stability and solubility.

5.5 X-ray diffraction of MmyB crystals

Recombinant MmyB from C43 after IMAC with ATP washes (Fig. 80) was exchanged into a buffer containing 20 mM Hepes, 50 mM Hepes at pH 7.60 and concentrated to around 6 mg/mL (Chapter 2: Materials & Methods). This was then used to set up a crystal tray with the commercial JCSG+ screen (Chapter 2: Materials & Methods). After approximately 7 months of incubation crystals were seen.

Crystals thought to be protein were seen in conditions of the commercial screen containing 0.1 M sodium citrate at pH 5.5 and 20 % w/v PEG 3000 (A2 of JCSG+) (Fig. 83). This was harvested and cryo-frozen by Dr Crow which was then sent to a synchrotron for diffraction (Diamond facility in Oxford). Despite the crystal being large enough, diffraction data obtained was inconclusive.

The crystal taken likely was protein but the reason of its lack of diffraction may be due to a range of factors which would need to be investigated in the future and post-crystallisation techniques (e.g. crystal dehydration) could be trialled.[274] However before that, it would be useful to ensure that it is not an issue with the sample by taking another sample of crystals from the same tray. At the time this was not possible as there were no other suitable crystals however, crystals may arise over time. Alternatively, more pure, stable MmyB may yield protein crystals that give good diffraction data that can be used to solve for structure.

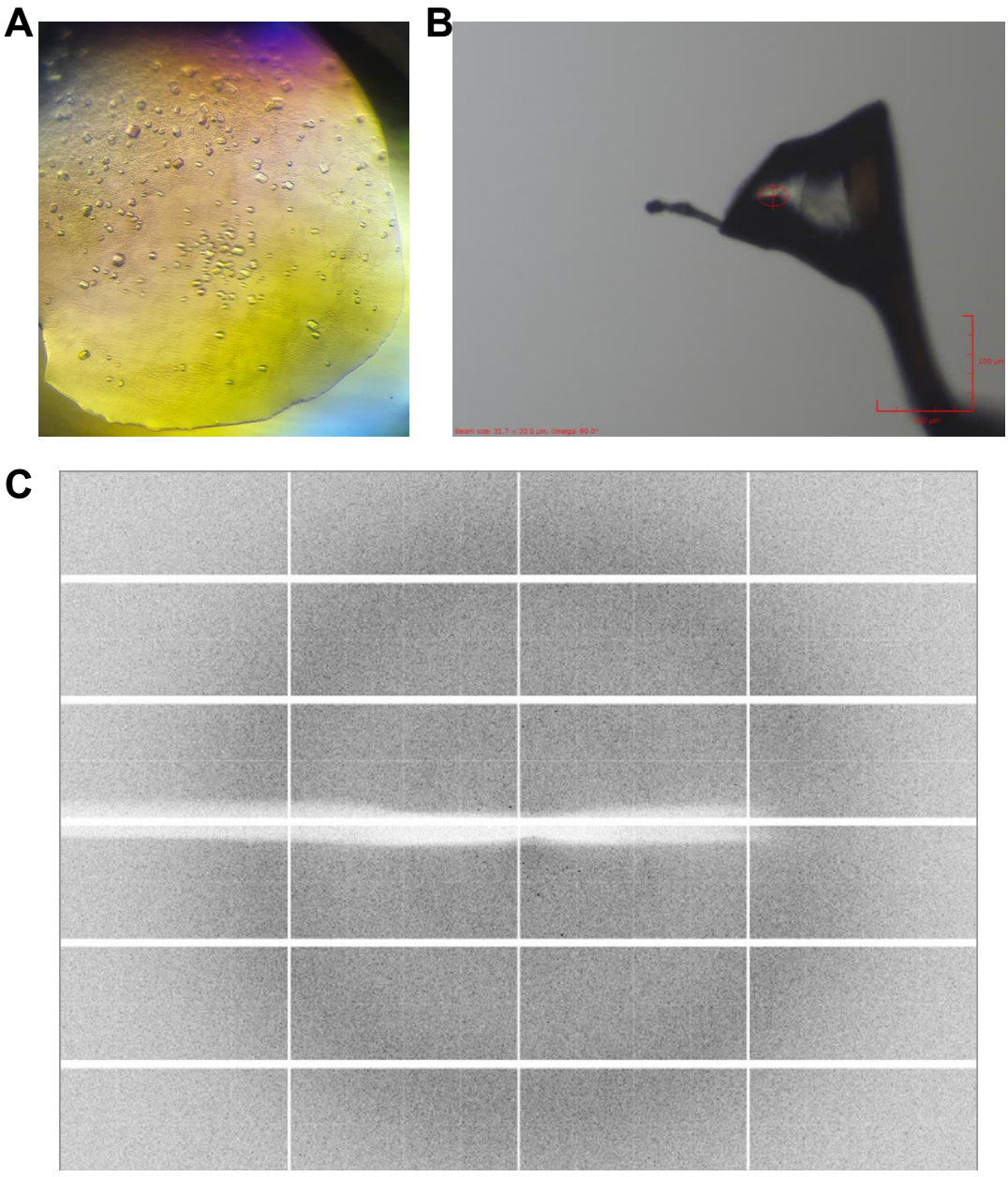
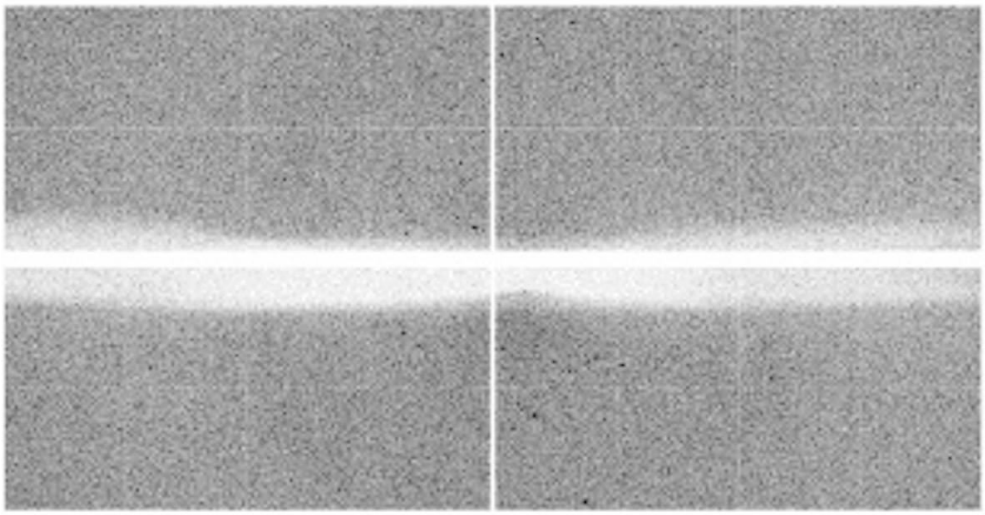


Figure 83. Crystals of recombinant MmyB were subject to X-ray crystallography. A) Crystals of recombinant MmyB purified from a C43 strain (Fig. 74) was formed using MmyB at 6 mg/mL in a commercial JCSG+ Screen (A2) after ~ 7 months of incubation at room temperature using a sitting drop vapor diffusion method. Reservoir was clear of any precipitation or crystals leading to the belief these are protein crystals. **B)** The crystal was harvested by Dr Crow (University of Warwick). The protrusion from the loop is of the 'skin' of a crystallization drop that when harvested was cryo-frozen. **C)** The X-ray diffraction data shows poor quality reflections though is reminiscent of protein crystal and not salt. **D)** A zoomed in view near the beam stop to highlight the few reflections that were seen however structure could not be solved from this diffraction data.

D



5.6 Apo-MmyB does not bind DNA and likely requires ligand binding first

MmyB is considered a transcriptional protein meaning it too is thought to bind to DNA. The exact location of DNA binding has only been suggested as a region prior to the *mmyB* gene called 'B-boxes' (Fig. 66).

The B-box region was purchased as two separate primers and annealed together (Chapter 2: Materials & Methods) in order to save on time. This region could be amplified and purified however with the high GC content of *Streptomyces*, amplification was thought to be challenging. For future work it may be useful to have a robust method of amplification and purification of this region should MmyB interact with it in order to have a more readily available supply of this DNA.

To then test if MmyB interacts with any section of the B-box DNA, an electromobility shift assay (Chapter 2: Materials & Methods) was conducted using MmyB purified as above (Fig. 80) and double stranded B-box DNA. Fig. 83 shows that although 20 times more protein was added to DNA, no shift changes and therefore no binding was detected.

This might be because the MmyB purified contains some chaperones despite best efforts to avoid this. The presence of chaperones may indicate issues with the folding of MmyB. Problems with the tertiary structure of MmyB may be then rendering it non-functional or partially functional which may affecting its DNA binding capabilities. Non-functional MmyB might therefore be an explanation as to why no DNA binding is seen in Fig. 84.

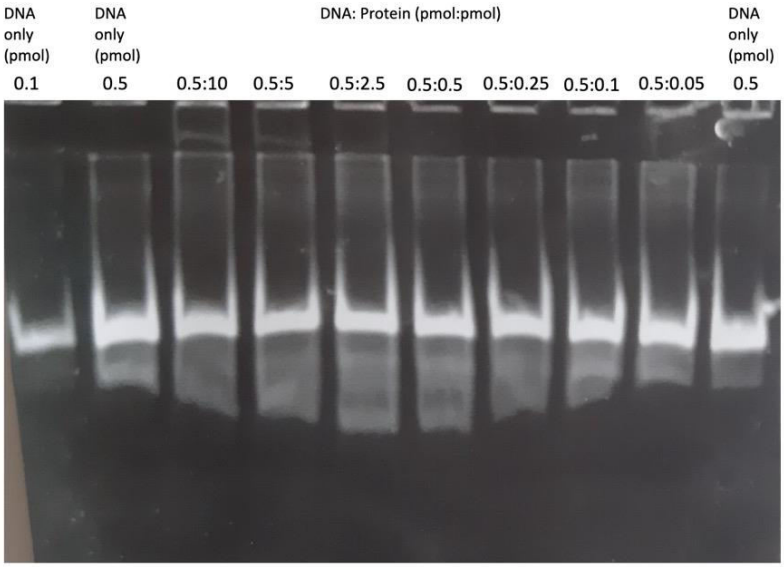


Figure 84. Electromobility shift assay (EMSA) of MmyB and DNA. Purified recombinant MmyB (Fig. 80) was incubated with the 'B-box' sequence of DNA predicted to be the binding site of MmyB (Fig. 66) at various picomolar concentrations (pmol) before being run on a 10% NATIVE PAGE gel and stained to visualise the DNA. The DNA was run alone as a control and any bands above this would indicate protein binding to DNA and thus a 'shift'. No shifts are seen indicating MmyB does not bind to DNA, at least not presumably without first binding to a ligand as many other transcriptional activators do.

There may also be another reason MmyB binding to DNA is not seen. Even with functional MmyB it might not bind to DNA without the presence of its ligand. This is because it is a transcriptional activator rather than repressor, meaning at cell homeostasis it is unlikely to be found bound to DNA. For a transcriptional activator a ligand molecule may first need to be bound to then allow it to be in the correct conformation to bind to DNA.

Unfortunately, this cannot be tested with MmyB as the potential binding ligand is unknown. It is likely this ligand may be an intermediatory product produced within the methylenomycin biosynthetic gene cluster that propels further activation of the production of the antibiotic or perhaps a molecule produced in response to cellular stress (Chapter 1), however it is impossible to say without analysis of the structure of MmyB or experimental evidence using MmyB.

5.7 Evidence to suggest MmyB acts as a transcriptional activator for MmA production

5.7.1 Testing *mmyB* activity in *S. coelicolor* strains

MmyB is predicted to be the activator for the production of MmA as evidenced from knockout gene studies that resulted in no MmA production.[148] However as discussed earlier (Chapter 1.4) *mmyB* is under transcriptional control by MmfR/MmyR/MMFs.

Strains containing knockout genes of the MmA gene cluster were created by Corre *et al.* (2008).[151] The relevant one for this study was the '*S. coelicolor* W81' strain and variant of this strain created by Aldali J. (2018)[177] which was named '*S. coelicolor* W81 with the pJH2 plasmid' (Chapter 2: Materials & Methods). For convenience the '*S. coelicolor* W81' strain work will hereby referred to as W81 and the '*S. coelicolor* W81 with the pJH2 plasmid' will be referred to as W81+*mmyB* as these better denote their functionality and purpose.

The W81 strain uses a C73-787 backbone that contains the methylenomycin gene cluster with *mmfLHP* knocked out (Fig. 85). These knocked out genes are responsible for the production of MMFs that bind to MmfR allowing MmfR to derepress from binding to DNA. As discussed previously MmfR binds at MARE sites in the promoter region of *mmyB* (Fig. 66) to control transcription of *mmyB*. Since the W81 strain cannot produce MMFs, MmfR remains bound to DNA and so *mmyB* is repressed. Theoretically this should mean no MmA is made as *mmyB* is not able to activate the switching on of MmA production.

The W81+*mmyB* strain is the W81 strain with an additional plasmid ('pJH2').[177] This plasmid contains an extra copy of *mmyb* which is under the control of the

constitutively expressed ermE* promoter. Should MmA be detected, it would suggest a bypassing of the MmfR/MmyR/MMF system for expression of *mmyB*.

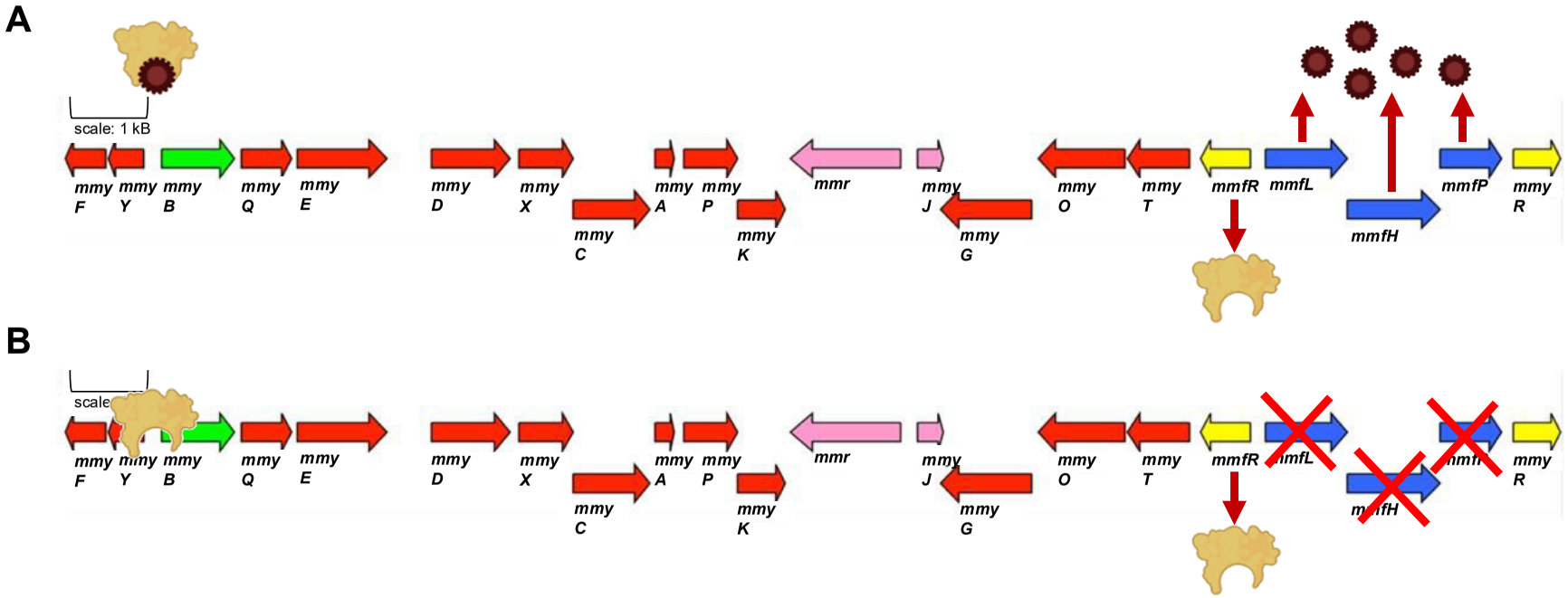


Figure 85. The methylenomycin gene cluster in the W81 strain of *Streptomyces coelicolor*. A) Normally in *S. coelicolor mmfL*, *mmfH* and *mmfP* are genes in the methylenomycin gene cluster that produce MMF signaling molecules (in red). The *mmfR* gene produces the transcriptional repressor MmfR which binds to MARE (Fig. 66) sequences, one of which is present near the start of *mmyB* and thus controls *mmyB* expression. **B)** In the W81 strain (by Corre *et al.* (2008))[151] the genes *mmfL, mmfH* and *mmfP* are knocked out meaning no MMFs are produced. This in turn stops the release of MmfR from DNA at the start of *mmyB,* halting the expression of MmyB.

5.7.2 Can constitutive expression of *mmyB* bypass the MmfR/MmyR/MMF regulation system?

Both the W81 and W81+*mmyB* strains were cultured in solid R2YE media and organically extracted using ethyl acetate (Chapter 2: Materials & Methods). Dried, crude fractions were then reconstituted in 50:50 methanol and water and subjected to low resolution liquid-chromatography mass spectrometry (LC-MS) (Chapter 2: Materials & Methods).

Extracted ion chromatograms (EICs) and base peak chromatograms (BPCs) of the two were analysed (Fig. 86). From analysis of BPCs there are already clear differences in products produced by W81 (control) and W81+*mmyB* (Fig. 86A).

As discussed previously the W81 strain is not predicted to be able to produce MmA whereas the W81+*mmyB* strain may be able to. The mass over charge (m/z) of positive ion MmA is 183.06 g/mol. Indeed, in the W81+*mmyB* strain a strong peak at this mass corresponding to MmA is seen whereas this is not present in the W81 strain (Fig. 86B).

To further corroborate this, a high-resolution LC-MS could be run and the UV data used to confirm this is MmA. While this is preliminary data, these results do emphasise the importance of MmyB. Addition of just an active *mmyB* gene to the silenced methylenomycin gene cluster switched on the gene cluster and created significant production of methylenomycin antibiotic.

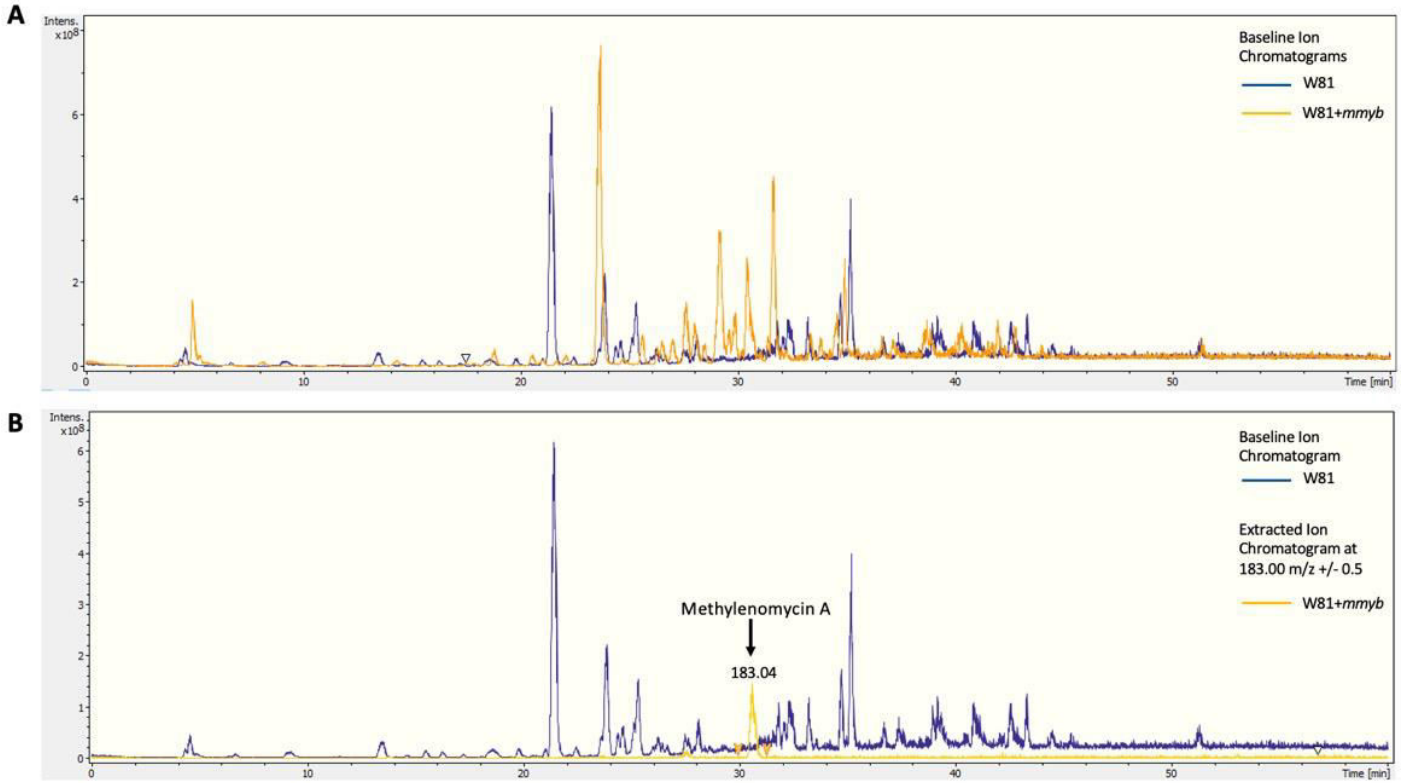


Figure 86. Low resolution, liquid chromatography mass spectrometry (LC-MS) of W81 and W81 with an extra *mmyB* containing plasmid for the detection of Methylenomycin A. *S. coelicolor* strains of W81 (Fig. 85B) and W81 with an additional *mmyB* containing plasmid (pJH2 plasmid[196] renamed 'W81+*mmyB*') was grown on R2YE media and subjected to LC-MS (Chapter 2: Materials & Methods). **A)** The base peak chromatograms (BPCs) of both strains show more compounds produced in the W81+*mmyB* strain than the W81 control strain. **B)** The BPC of W81 (blue) and extracted ion chromatogram of W81+*mmyB* at 183.06 m/z (yellow) which is the mass (in positive ion mode) of methylenomycin A. A peak at 183.04 m/z of W81+*mmyB* is seen most likely corresponding to methylenomycin A. No similar peaks are present in the W81 control sample.

5.8 Summary of MmyB work and future work with MmyB

MmyB like genes are prevalent in many actinobacteria and several of these exist within *S. coelicolor* A3(2) itself. While it has been suggested that they fall under 2 separate groups with MmyB being part of one and other uncharacterised proteins belonging to another there is still a large amount of diversity within this class (Fig. 67). Those that are genetically similar to MmyB however do have some conserved residues particularly in the helix-turn-helix region used most likely for binding to DNA. There are also many *mmyB*-like genes within the same *Streptomyces* strain (Fig. 67B) and the most similar proteins to MmyB are also from *Streptomyces* species (Table 19).

The conservation in the intergenic region in the same *Streptomyces* species with *mmyB*-like genes are not necessarily observed as expected. Regions known as 'B-boxes' are where MmyB is predicted to bind in *S. coelicolor* A3(2) with another predicted conserved region being the 'MARE' sequence. This sequence is not seen as conserved in the intergenic regions of predicted *mmyB*-like and *mmyB*-like divergent genes, however there does appear to be some conserved residues. It would be interesting to see if there is the same synteny in other species of actinobacteria that have *mmyB*-like genes and whether the intergenic regions with those contain the same conserved residues. If so, this would give a better idea as to the binding region of MmyB and also the diversity within this type of protein.

It was found that MmyB likely cannot bind to DNA within this intergenic region or to the B-box without possibly first having a ligand bound to it (Fig. 84). Alternatively, there may be another site of binding for MmyB that is not within the intergenic region between *mmyB* and *mmyY*. To find the binding region and potential ligand it would be useful to first have the structure of MmyB for modelling.

The MmyB attempted to be purified was unstable as it contained chaperone

proteins (Fig. 78) which assist in the refolding of misfolded proteins. When the majority of this was successful excluded during purification, the MmyB remaining was highly unstable and precipitated. This may be an issue with the MmyB protein itself or the conditions it was purified in.

In the first instance the stabilisation of the protein could be attempted. While glycerol was added to buffers, the concentration of salts and the pH may also need fine-tuning. To test this in a high throughput manner, thermal stability assays could be attempted. This could determine the best buffering conditions which may be enough to stabilise the MmyB. Should this not be effective the reengineering of MmyB could be considered with addition of extra solubility and stabilisation tags. Finally, if that doesn't work, purification of MmyB from growth within a *Streptomyces* host could be attempted.

Interestingly growth within the *E. coli* host with the BL21 strain did cause high expression of ArnA (Fig. 78) even in a sample containing MmyB. ArnA is typically a molecule associated with the modification of Lipid A for resistance to antimicrobials. When the strain was exchanged for one that was designed to allow expression of toxic molecules, no ArnA was detected (Fig. 81). This may indicate MmyB being recognised by *E. coli* as a toxic molecule or one that is associated with antimicrobial activity.

Finally, it was shown previously that stopping MMF production within *S. coelicolor* A3(2) halted production of MmA[150] and complementarily to that it's now been suggested having *mmyB* switched on can activate the silent biosynthetic gene cluster and allow MmA production (Fig. 86). This has promising applications such as testing other silent biosynthetic gene clusters containing MmyB-like proteins and identifying compounds produced by these genes, some of which may be novel. Of course, the organism of origin of these clusters and thus the host organism will have to be taken into consideration to note if they are compatible. Though it must be noted bioinformatic analysis has shown many *Streptomyces* species containing homologous MmyB proteins and any of these uncharacterised genes and related

gene cluster would be a good place to start. Producing plasmids containing this gene cluster alone as well as an additional plasmid with only the *mmyB*-like gene may be challenging, particularly if this cluster has additional activators and does not behave in a similar fashion to the methylenomycin gene cluster. However, it would be a relatively straight forward and suitable way to screen for the production of novel natural products.

Work done so far on MmyB has only scratched the surface of this interesting transcriptional regulator protein. Bioinformatic evidence suggests there is a vast amount of MmyB like proteins in actinobacteria and in particular *Streptomyces* species. This is unsurprising as MmyB proteins are probably involved in the activation of the production of antibiotic of which the *Streptomyces* species is famed for. Classifying and characterising MmyB would therefore be valuable to allow insight into transcriptional activators involved in the production of antimicrobials.

Chapter 6: Final Discussion

6.1 Summary of results

This thesis aimed to characterise the transcriptional repressor MmyJ for self-resistance to antibiotic methylenomycin and the transcriptional activator MmyB for the production of the methylenomycin. Both of these transcriptional factors are found within same methylenomycin gene cluster in *S. coelicolor* A3(2) (Chapter 1.4).

Work into MmyJ (Chapter 3 and Chapter 4) found a number of key results. Firstly, the structure of C49S MmyJ was solved to 1.68 Å which led to the solving of a previously diffracted (to 2.11 Å) but unsolved WT MmyJ crystal. Mutation of C49, the only cysteine in MmyJ, did not lead to any structural changes. The DNA binding function of MmyJ was also not diminished with the cysteine mutation. MmyJ's ability to bind to an intergenic region of DNA between *mmyJ* and *mmr* was previously demonstrated but the work in this thesis has identified the specific binding sites on DNA for MmyJ. These are 2 consecutive pseudo-palindromic regions one of which encompasses the -35 promoter site of *mmr* (CCATCAAAGTTTGACAG) and the other the -35 promoter site of *mmyJ* (CCATCAAATGTTGACGG). Half the concentration of MmyJ was shown to be required to bind to the promoter site for *mmyJ* than *mmr*, implying a higher affinity of MmyJ to its own promoter site and therefore tighter control of itself. Similar sites with conserved nucleotides were identified in a number of other actinobacteria with synteny of *mmyJ* and *mmr* genes. Ligand-DNA-MmyJ assays of crude extracts containing MmC but none/little MmA showed no clear de-repression of MmyJ from DNA indicating ligand specificity. MmyJ was also found to dimerise under native conditions and crystalise as a dimer. BLAST searches, phylogenetic study, and structural analysis of MmyJ strongly indicate it to be an ArsR family protein (a class of transcriptional repressors that with some exceptions regulate metal ion levels within a cell).

From Chapter 5, a recombinant MmyB was overexpressed and purified and though found to be unstable was able to crystalise. Diffraction data from the MmyB crystals were poor so structure was unable to be solved however X-ray diffraction of other MmyB crystals that may grow in time may be able to be used for structure elucidation. The transcriptional activator nature of MmyB has been established previously by O'Rourke *et al.* (2009)[148] and from this study, may have been seen through DNA-MmyB assays where recombinant MmyB was unable to bind to the DNA regions it is theorised to bind to, perhaps due to no available MmyB ligand. Again, structural analysis of MmyB may be important here for identifying potential ligands. Finally, through microbiology work it was shown that a constitutively expressed *mmyB* gene in *S. coelicolor* A3(2) with the methylenomycin gene cluster silenced, is enough to activate methylenomycin A production. A more exhaustive list of future experiments was suggested in Chapter 5.8.

6.2 A proposed mechanism of transcriptional regulation of antibiotic self-resistance by MmyJ

Combining results generated in this study, previous work, and knowledge of ArsR family proteins, a model of the function of MmyJ is proposed (Fig. 87).

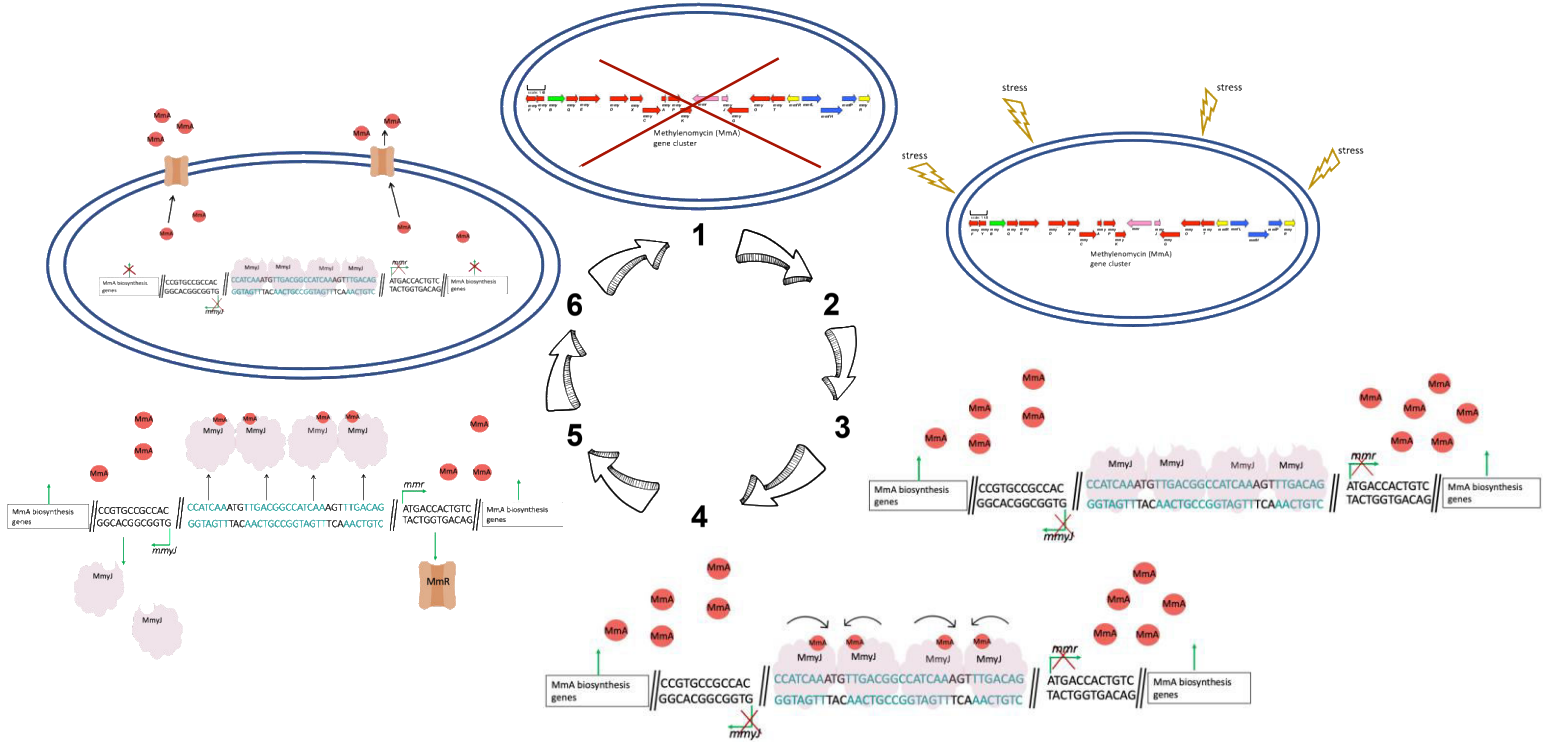


Figure 87. The proposed mechanism of MmyJ in *Streptomyces coelicolor* A3(2). 1. In a stress-free environment the methylenomycin A (MmA) gene cluster is switched off. **2**. Under stress the MmA gene cluster is activated and MmA biosynthesis begins. **3.** MmA biosynthesis genes are activated producing MmA (red dots). The -35 promoter regions of *mmyJ* and *mmr* are highlighted in teal to which MmyJ homodimers likely bind. MmyJ functions as a transcriptional repressor so at low levels of MmA remain bound to DNA. **4**. MmA concentrations rise to a threshold and MmA binds to MmyJ. Binding creates an allosteric change (black arrows) of the MmyJ homodimer. **5**. The allosteric change stabilises the MmyJ homodimer and allows for unbinding from the DNA (black arrows) thereby exposing the -35 promoter regions of *mmr* and *mmyJ* producing MmyJ and Mmr proteins. **6**. The Mmr efflux pump exports MmA out of the cell causing MmA concentrations to fall intracellularly. Less MmA means, MmyJ binds to DNA again returning the cell to its native state.

At homeostasis the *S. coelicolor* A3(2) cell produces no/little amounts of MmA meaning the expression of the methylenomycin gene cluster is not very active (1-Fig. 87). A stressor activates the methylenomycin gene cluster (2-Fig. 87) starting the biosynthesis of MmA via production MMF signalling molecules. At low levels of MmA, as a transcriptional repressor MmyJ dimers remain bound to DNA (3-Fig. 87). One dimer likely binds to the -35 promoter site of *mmr* with another dimer of MmyJ bound at the -35 promoter site of *mmyJ*. This prevents the transcription of both of these genes by restricting access for RNA polymerase. When a certain threshold concentration of MmA is reached within the cell, MmA binds to MmyJ homodimers (4-Fig. 87).

One MmA molecule likely binds to each monomer at sites near where the DNA binds, creating conformational changes within MmyJ homodimers that release it from being bound to DNA. The release of MmyJ homodimers from the each of the two DNA binding regions may occur at different times depending on the binding strength of MmyJ homodimers to each region. Also the binding affinity of MmA to MmyJ bound to the -35 *mmyJ* and -35 *mmr* promoter regions is unknown and there may be variations here that impact the overall molecular mechanism of MmyJ in the cell. In this way MmyJ homodimers bound at the *mmyJ* site may be released first from MmA interactions and then the dimers bound to the *mmr* site or vice versa.

Once MmyJ homodimers interact with MmA they are released from binding to DNA (5-Fig. 87) and the -35 promoter sites for transcription of *mmyJ* and *mmr* are exposed allowing for the synthesis of both MmyJ and Mmr. The MFS family efflux pump Mmr exports MmA out of the cell, which once the production of MmA slows or stops, causes a decrease in levels of MmA within the cell and reduces MmA available for binding to MmyJ (6-Fig. 87). With the decline of MmA concentrations within the cell, synthesis of Mmr is no longer required. Unbound to ligand, free MmyJ homodimers bind again to DNA, ceasing production of Mmr (and MmyJ) and thereby switching off the mechanism for self-resistance (1-Fig. 87). A lower concentration of MmyJ was required to bind its own -35 promoter region site

compared to the *mmr* site, indicating higher affinity of MmyJ to its own promoter site than that of *mmr.* This implies that the production of MmyJ may be repressed prior to ceasing Mmr production. One reason for this mechanism may be to ensure MmA concentration within the cell depletes sufficiently before the efflux pump is repressed. The cycle is reset and can be started again if or when the cell requires it to be.

Autoregulation in genetics is the direct modulation of gene expression from a protein or other product of a corresponding gene. The implication of MmyJ and evidence of other ArsR family proteins having this autoregulatory system in response to stimuli (e.g. ligand like metal ion or antibiotic) is a desirable feature for cells to have as it means expression of response genes are only produced in response to stimuli, thereby conserving cellular energy by providing a specific response only when necessary. Furthermore, autoregulation allows for the output levels to be finely controlled whereby the concentration of stimuli can influence gene expression and response levels, ensuring that the correct response is produced only when appropriate by levels of stimuli are present.

6.3 Impact of research and future prospects

The advances made from this study into understanding transcriptional proteins of the methylenomycin gene cluster demonstrates the significance of some of the techniques used. For instance, X-ray crystallography not only gave the structure of the MmyJ protein but solving this structure at a high resolution assisted in predicting the binding location on DNA of MmyJ. In this way, the structural and functional assay work were complimentary and gave key details into MmyJ's mode of action. Furthermore, size exclusion chromatography is generally relegated to simply purifying a protein of interest but as shown in this body of work, the separating properties of size exclusion chromatography can be useful in other ways. For example, this technique can also be valuable in understanding protein mechanisms like dimerization or protein-ligand interactions. Finally, while the

bioinformatics conducted in this study was not exhaustive, recent developments in algorithms and user-friendly web servers have allowed for connections to be made across different organisms much easier. For instance, the phylogeny and synteny of a particular protein was able to be visualised relatively efficiently leading to the observation of a few key organisms of relevance from a large dataset. No doubt, as advances are made in experimental and bioinformatic techniques the finer details of the mode of action of transcriptional regulators within the methylenomycin gene cluster will be realised more readily.

The research presented in this thesis has highlighted a potential method for the discovery of new antibiotics with only knowledge of the transcriptional activator for an antibiotic gene cluster. Cloning of a plasmid containing the proposed antibiotic transcriptional activator into a host containing the silent antibiotic biosynthetic gene cluster that activator is thought to control, and screening for organic compounds produced may show novel secondary metabolites. While this does rely on a MmyB-like system and knowledge of these antibiotic gene clusters as well as the transcriptional activators controlling them, it would be one way to unlock these gene clusters without using hormone-like signalling molecules. Overexpression and purification of these transcriptional activators has also been shown to be feasible and while the protein was unstable (a problem that may be easily remedied with use of a solubility tag), it does seem to crystalise under the right conditions within just over half a year (as demonstrated in Chapter 5). Future experiments to solve the structure of this MmyB is therefore well within reason and structural analysis may be the start of identifying a new subset of antibiotic transcriptional activators.

As for MmyJ, the transcriptional repressor for self-resistance in the methylenomycin gene cluster, evidence strongly claims it to be an ArsR family protein. MmyJ is therefore the first structurally characterised and best functionally characterised ArsR family protein that is involved in antibiotic resistance. Evidence implies that MmyJ is not the only one of its kind either with MmyJ orthologs existing in other organisms, creating the possibility of a whole new subset of ArsR family proteins represented by MmyJ. These likely orthologs would need better

characterisation themselves which would be a direction of research that could reveal the secondary metabolites exported by the gene product the MmyJ-like protein regulates. Some of these secondary metabolites or their biosynthesis genes and mechanisms may be novel.

Finally, one other direction MmyJ research could take divergent to the antibiotic resistance route would be to focus on the ArsR family protein aspect. Specifically, as MmyJ shows evidence of being a new type of ArsR family protein in controlling antibiotic self-resistance mechanisms, the evolutionary link between heavy metal resistance typical of most ArsR family proteins and antibiotic self-resistance could be investigated. Bioinformatic studies discussed in the 'Introduction' chapter (Chapter 1.3.2) speculated close links between heavy metal resistance and antibiotic resistance mechanisms which is now evidenced experimentally by MmyJ. Investigating proteins involved in antibiotic resistance further within ArsR family proteins and other heavy-metal resistance families could show how closely tied the two resistance mechanisms might be.

Appendix

A: Supplementary figures for Chapter 3

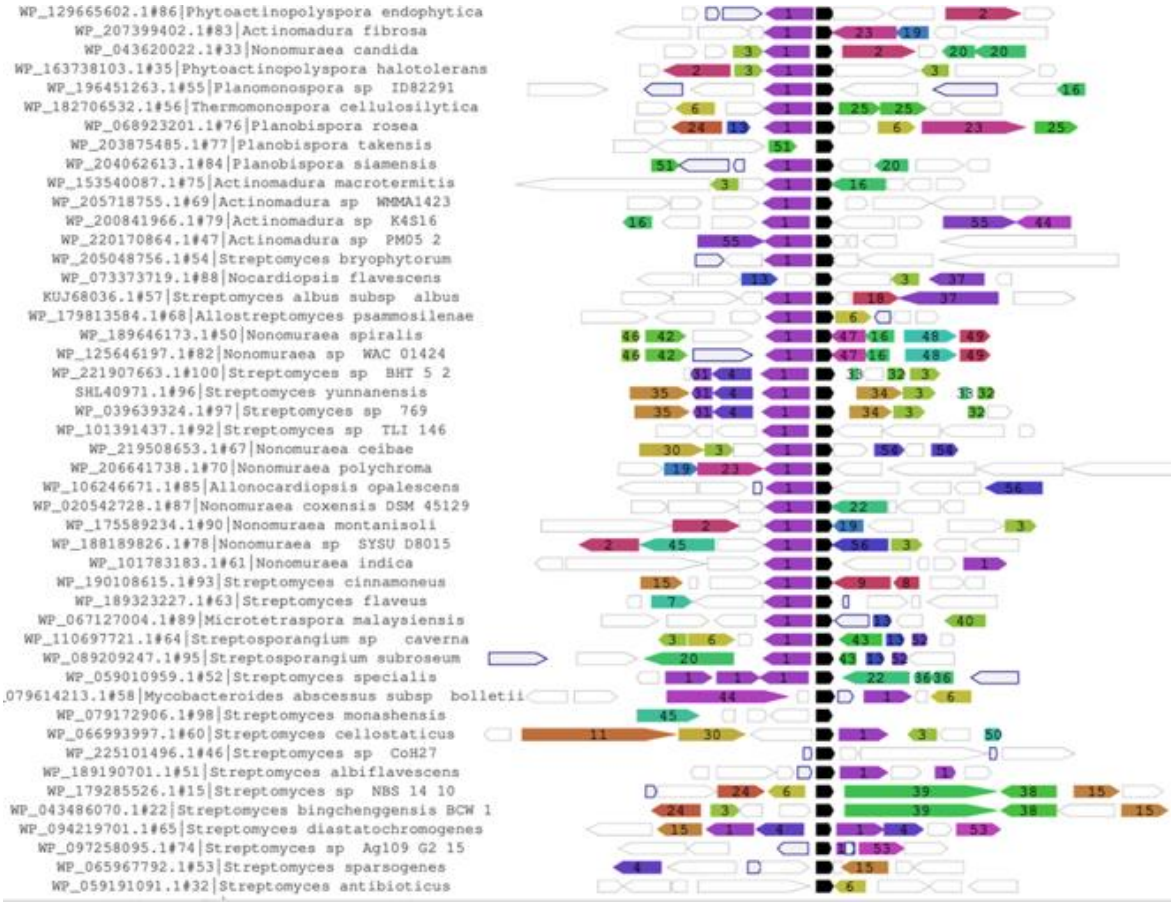


Figure A1. The WebFlAGs results of flanking genes in organisms from MmyJ-like proteins identified by BLAST. The rest of the WebFLAGs result of flanking genes (set to 4 in search) to the MmyJ-like genes identified in the top 100 results of BLAST searching. Accession numbers are given first, then the number this occurred in the BLAST search, followed by the species it came from and finally the flanking genes to the MmyJ-like gene (in black). Genes are numbered into categories. 1 is the category of genes described as zinc-dependent alcohol dehydrogenase family protein or NADP-dependent oxidoreductase. The full list is given in Table 1-Appendix.

Table A1. The complete gene descriptions of WebFlAGs results of flanking genes in organisms from MmyJ-like proteins identified by BLAST. Flanking genes from WebFlAGs results were annotated by numbers categorizing genes by their functionality. The full list of these gene descriptions according to NCBI database are given below in this table.

Gene number	Gene description
1	zinc-dependent alcohol dehydrogenase family protein/NADP-dependent oxidoreductase
2	MFS transporter
3	TetR/AcrR family transcriptional regulator
4	LLM class flavin-dependent oxidoreductase
5	AAA family ATPase
6	SDR family oxidoreductase
7	helix-turn-helix domain-containing protein
8	flavin reductase family protein
9	oxidoreductase
10	HAD-IB family hydrolase
11	AfsR/SARP family transcriptional regulator
12	acyl carrier protein
13	MerR family transcriptional regulator
14	NADH:flavin oxidoreductase
15	LysR family transcriptional regulator
16	RNA polymerase sigma factor
17	NAD(P)-binding domain-containing protein
18	helix-turn-helix domain-containing protein
19	response regulator transcription factor
20	ATP-binding cassette domain-containing protein
21	thioesterase
22	FAD-dependent oxidoreductase
23	HAMP domain-containing histidine kinase
24	aldo/keto reductase
25	MEDS domain-containing protein
26	saccharopine dehydrogenase NADP-binding domain-containing protein
27	choline dehydrogenase
28	ketoacyl-ACP synthase III
29	Gfo/Idh/MocA family oxidoreductase
30	FAD-binding oxidoreductase
31	F420-dependent oxidoreductase
32	VOC family protein
33	hypothetical protein
34	3-hydroxybutyryl-CoA dehydrogenase
35	Fucose permease
36	fluoride efflux transporter CrcB
37	PAS sensor protein
38	acyl-CoA dehydrogenase family protein
39	glycoside hydrolase family 78 protein
40	dioxygenase
41	pyridoxamine 5'-phosphate oxidase family protein
42	YihY/virulence factor BrkB family protein
43	hypothetical protein
44	alpha/beta hydrolase
45	FAD-dependent oxidoreductase
46	YccF domain-containing protein
47	zf-HC2 domain-containing protein
48	ribosome small subunit-dependent GTPase A
49	phosphoribosyltransferase

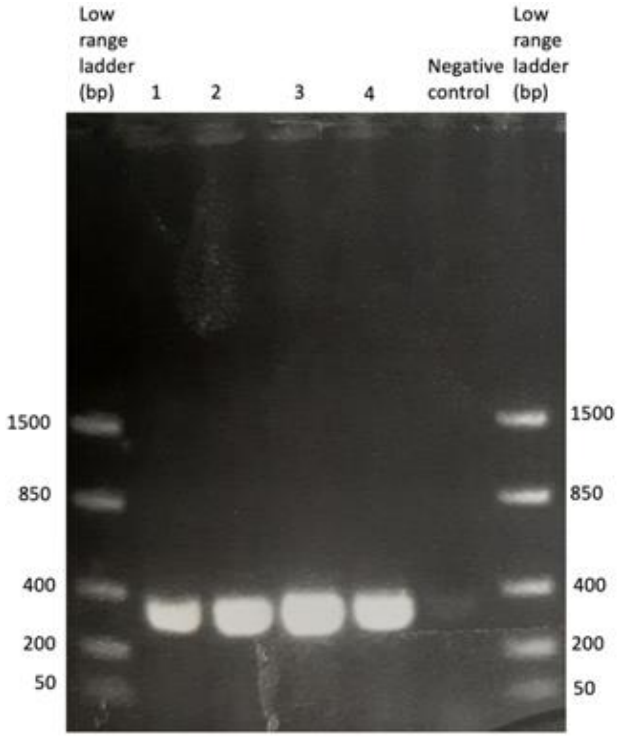


Figure A2. Transformation of pET151-MmyJ plasmid into BL21 DE3-Star cells. BL21 cells that were chemically competent were transformed with pET151-MmyJ and plated on an ampicillin resistant LB plate. 4 colonies were selected for colony PCR of a 336 base pair (bp) fragment and results visualized on a 1% Agarose gel. All four colonies as seen on the gel indicate presence of the MmyJ gene, with one used for protein production. These colonies and those of C49S MmyJ were also confirmed by Sangar sequencing to contain the correct MmyJ sequence. The sequence of both recombinant MmyJ and C49S MmyJ were checked by Sanger sequencing.

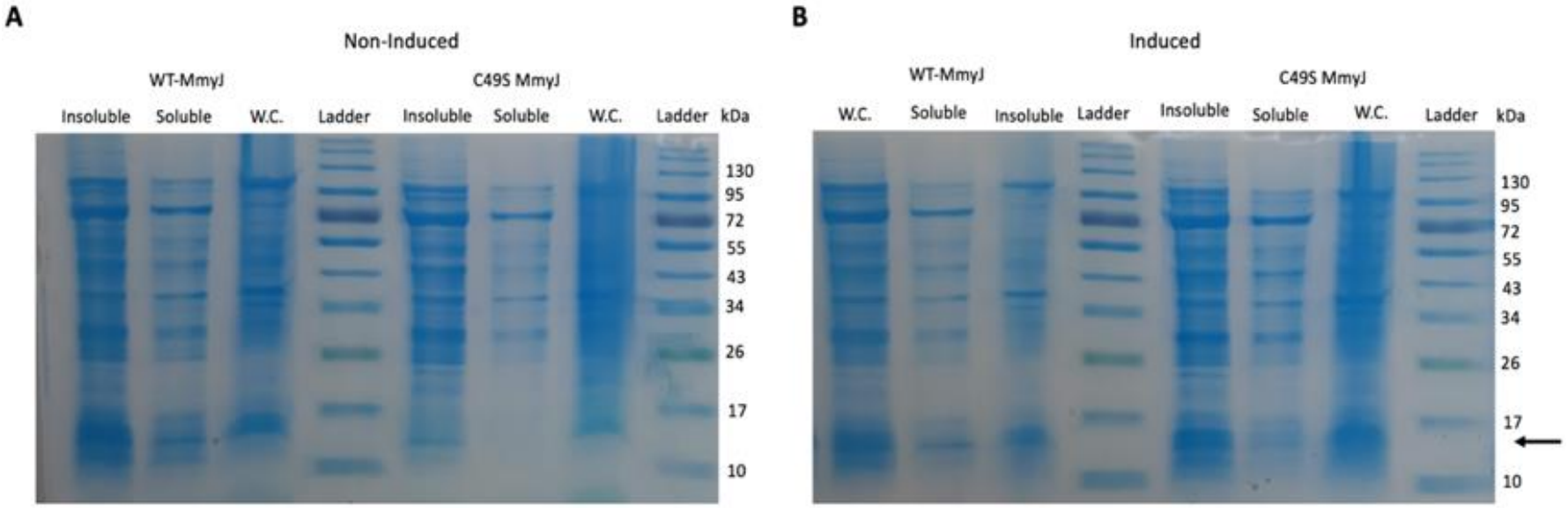


Figure A3. Protein expression trails of recombinant WT and C49S MmyJ. Small cultures (50mL) of recombinant WT and C49S MmyJ were grown and not induced (**A**) or induced with 1mM IPTG (**B**). After overnight growth a 15 µl was taken (whole cell-W.C.) while the rest was centrifuged. To this pellet the same volume of Wash Buffer (Chapter 2: Materials & Methods) as supernatant decanted was added, sonicated and centrifuged again (Chapter 2: Materials & Methods). The supernatant of this sample ('soluble' protein fraction) and pellet ('insoluble' protein fraction) were both taken. Gels were run on a 4-20% gradient TruPAGE gel (Chapter 2: Materials & Methods) and visualized. Recombinant WT-MmyJ weighs 15.75 kDa and C49S MmyJ weighs 15.73 kDa. A 17 kDa band seems to be present mainly in the induced cultures in soluble fractions (**B**) (black arrow). Cultures were scaled up to 1 L and purification done to isolate

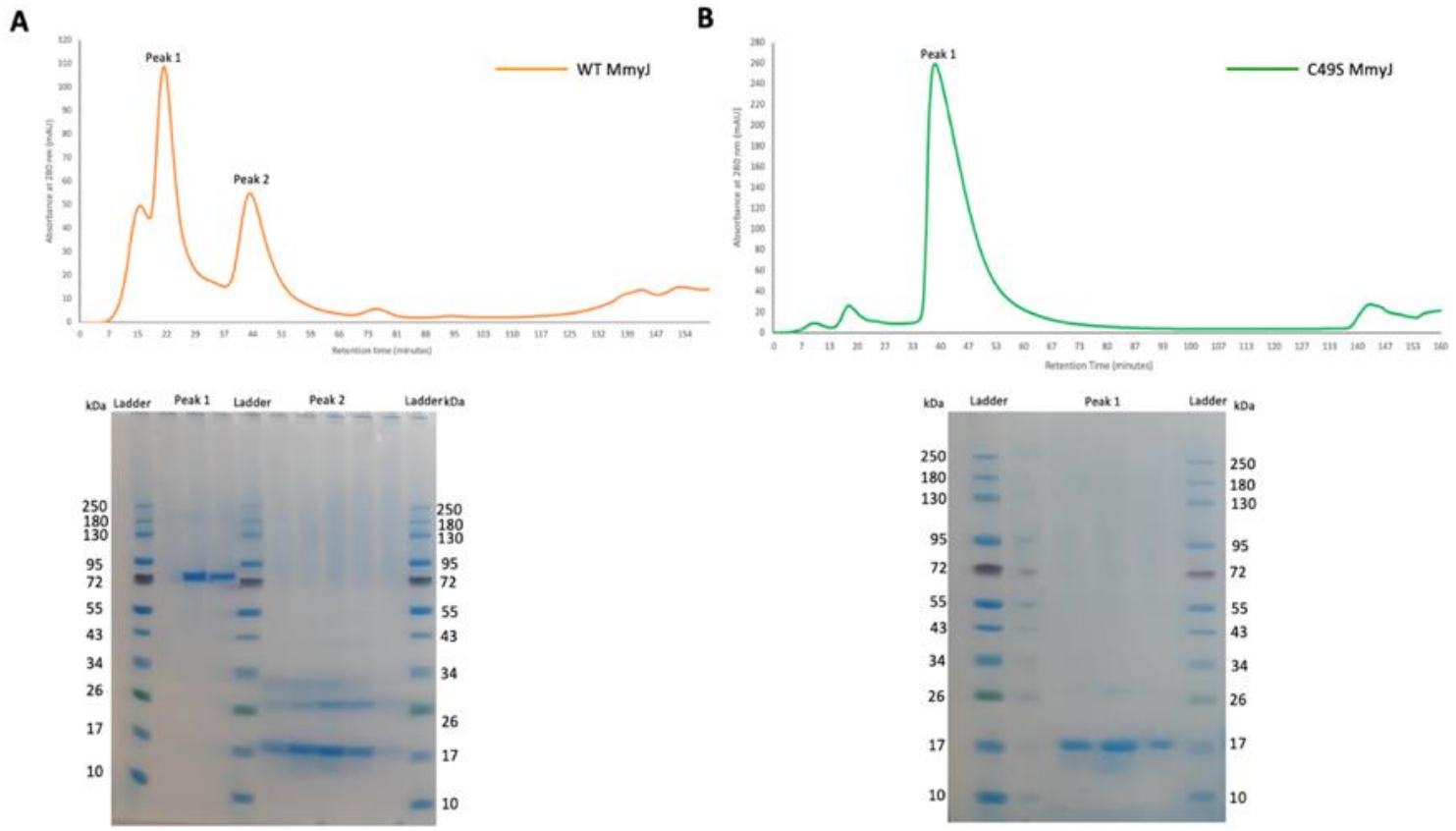


Figure A4. SDS-PAGE gels of peaks from size exclusion chromatography. A) Size exclusion graph of recombinant MmyJ WT and SDS-PAGE gel of peaks. B) Size exclusion graph of recombinant C49S MmyJ and SDS-PAGE gel of peaks. Main peaks seen are labelled and the multiple elution fractions collected encompassing these peaks were run on a 4-20% Gradient TruPAGE gels with a protein ladder (Chapter 2: Materials & Methods). In WT-MmyJ **(A)** peak 1 shows an 80 kDa protein whereas peak 2 appears to be a 17 kDa protein forming dimers. In C49S MmyJ the same peak (peak 1) also appears as a 17 kDa protein, suggesting peak 2 of MmyJ WT **(A)** and peak 1 of C49S MmyJ **(B)** are both MmyJ dimers.

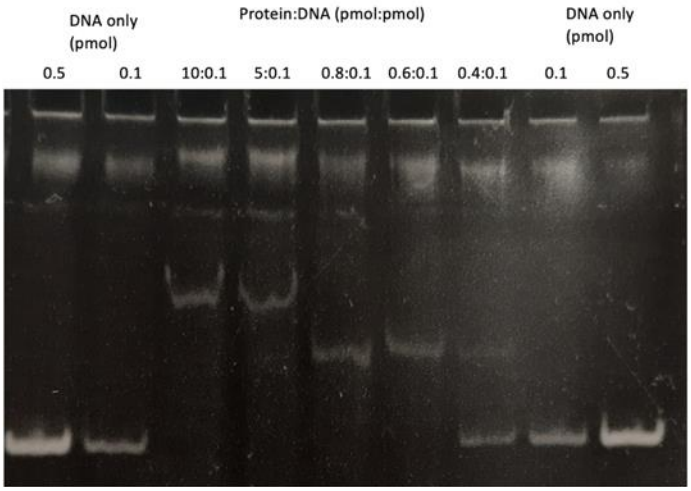


Figure A5. Electromobility shift assay (EMSA) of MmyJ binding to DNA. Purified recombinant C49S MmyJ was incubated with an 80 base pair (bp) region of DNA at various picomolar concentrations (pmol) before being run on a 10% NATIVE PAGE gel and stained to visualise the DNA. This is a repeat of the experiment of Fig. 16 with the different batch of purified C49S MmyJ and fresh DNA, on a different gel run on a different day under the same conditions. A band above the control samples of just the DNA indicates protein binding and therefore increasing the weight and decreasing the speed of migration through the gel. At concentrations 4 times greater of protein to DNA there appears to be protein binding to the DNA as seen by the band shift. A second band shift (super shift) at concentrations of protein 50 times greater than DNA indicates further binding of protein to DNA.

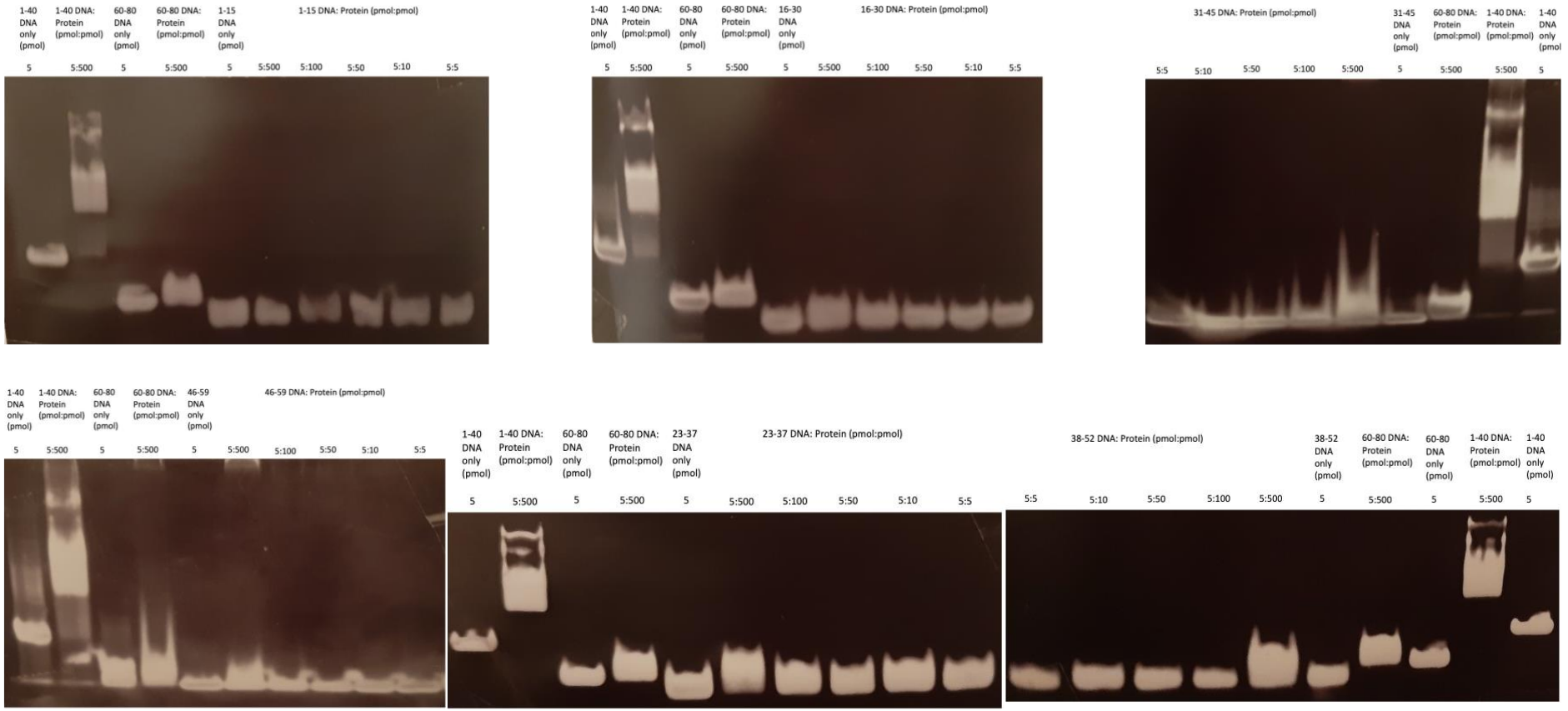


Figure A6. The 80 base pair (bp) section of DNA was further fragmented into 15 bp and tested for MmyJ binding. The 80 bp of intergenic DNA shown to have MmyJ bind to it (Fig. 16) was further fragmented to 15 bp sections (Fig. 17A in purple) find the exact binding region. No apparent binding was seen other than in the positive controls which are fragments already shown to have MmyJ bind to it.

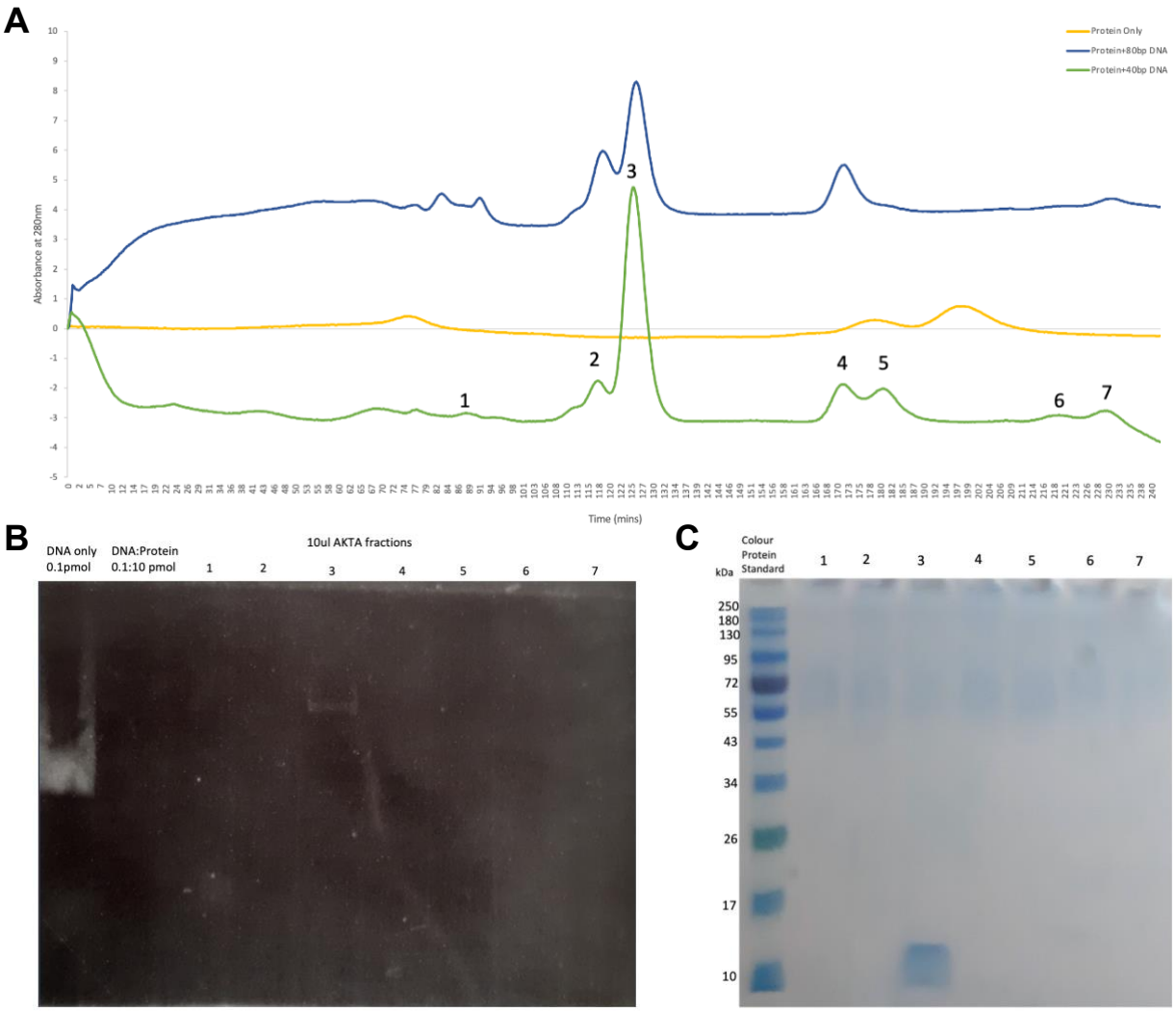


Figure A7. MmyJ-DNA stoichiometry was tested using analytical size exclusion chromatography. A) Size exclusion chromatography (SEC) was used to see across time DNA-MmyJ binding properties. An 80 base pair (bp) region of DNA was already shown to bind to MmyJ but SEC was used to further investigate these properties. All samples were incubated 30 minutes prior to loading. A protein only sample (yellow) at 65 picomolar (pmol) final concentration was first run showing only one weak peak that could be protein. A concentration of 100 times protein to DNA both the 80 bp fragment (blue) and 40 bp fragment (green) to which MmyJ binds to were then run. Some peaks were seen but when taken and run on a 4-20% TruPAGE gel (**C**) and use for an electromobility shift assay (**B**), peaks could not be distinguished clearly as MmyJ as some were too weak to assay and run on SDS-PAGE gels.

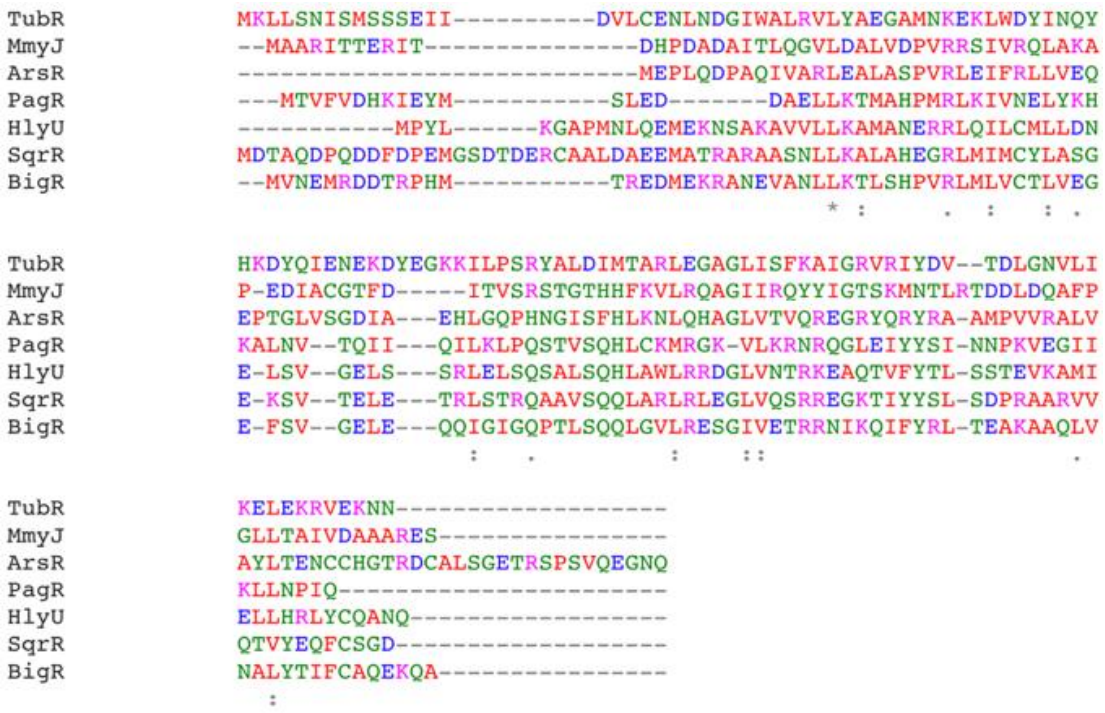


Figure A8. Structurally similar proteins to MmyJ as determined by PDBeFOLD were analyzed using MUSCLE for amino acid conservation. The amino acid sequence of proteins that were determined as being structurally similar to MmyJ were analyzed using the software MUSCLE. This aligns the sequence and places stars on residues that are conserved in all the proteins. As shown, only one leucine (L25 in MmyJ) is conserved in all the proteins and may therefore be of significance for functioning.

Figure A9. Sanger sequencing of pJ251 Tinsel purple with Anderson strong promoters. A successful pJ251 Tinsel purple plasmid cloned with the Anderson strong promoters before it. The ab1 file from sequencing was aligned against the predicted sequence in Benchling to find a perfect match.

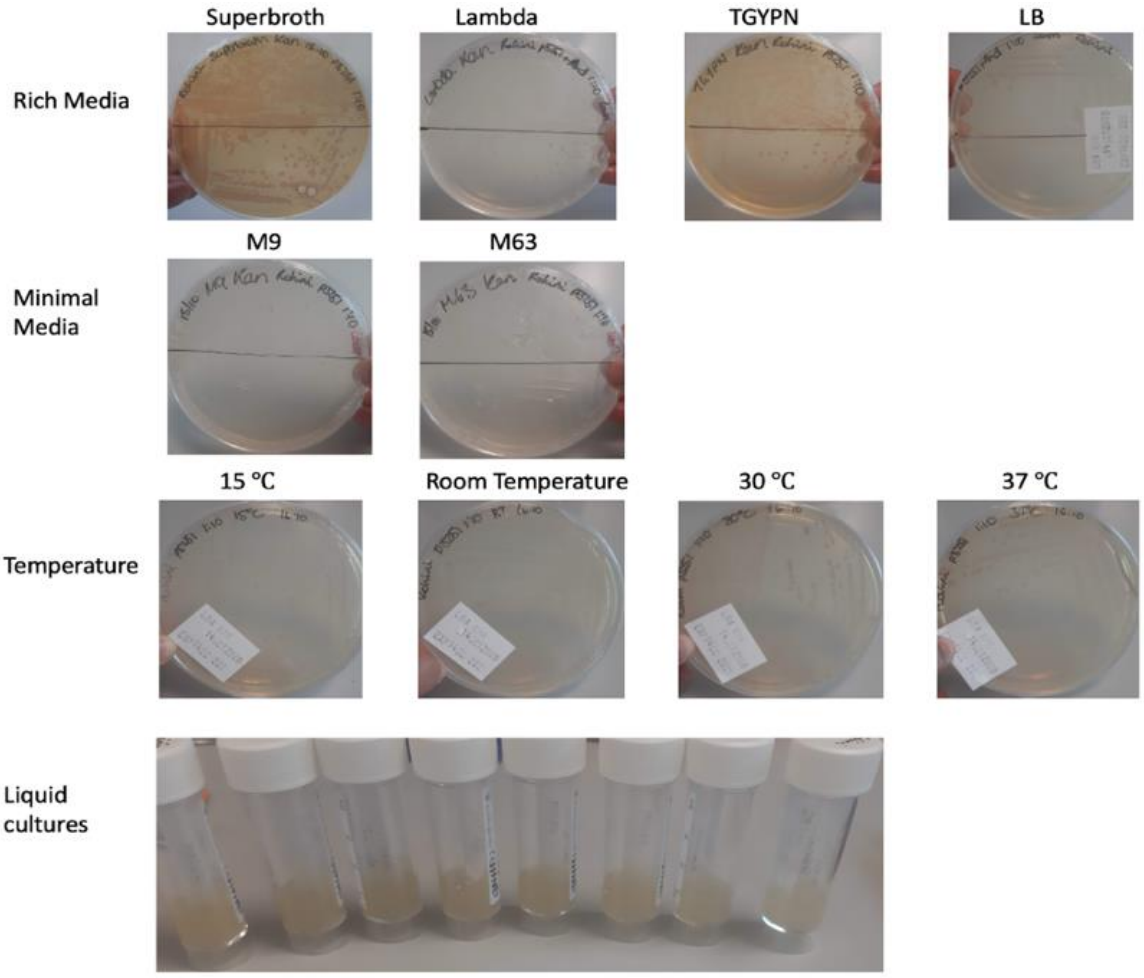


Figure A10. Different conditions of the pJ251 Tinsel purple with Anderson strong promoter plasmid colonies were tested for brighter purple pigmentation as expected. Different medias (both rich and minimal) were plated with the same colony of the pJ251 Tinsel purple (and Anderson strong promoter) plasmid TOP10 *E. coli* cells at 37℃ and showed no improvement of pigmentation. LB media was therefore used and different temperatures tested which also showed no differences in pigmentation of colonies. All plates were incubated for 3 days before visualisation. Liquid cultures of various colonies that on LB plates appear as light purple in liquid media (grown at 37℃ also for 3 days) show no light purple colour.

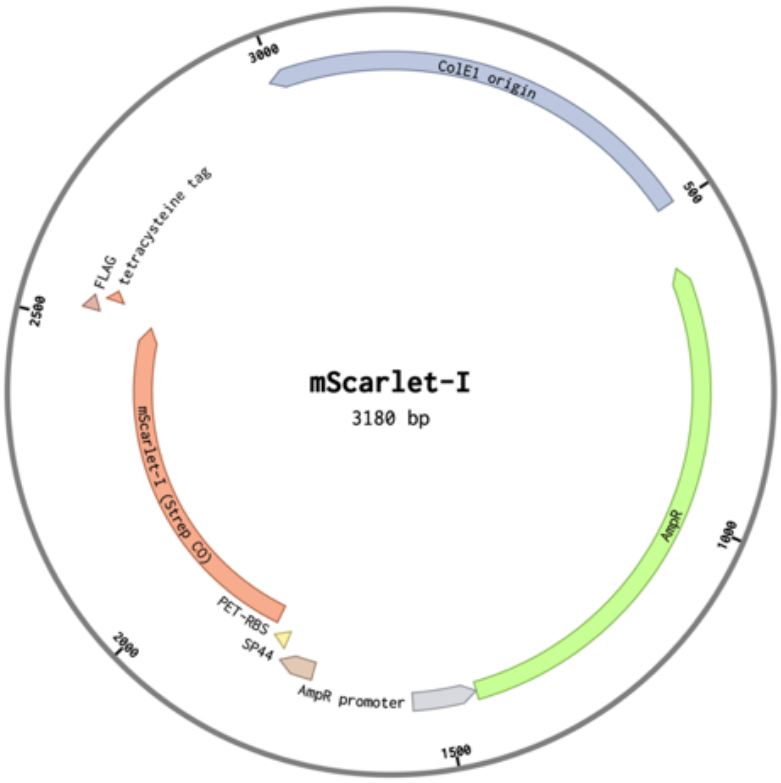


Figure A11. The mScarlet-I plasmid used for inserting *mmyJ-mmr* intergenic DNA. The mScarlet-I plasmid used to clone in DNA from the intergenic region of *mmr* and *mmyJ* (Chapter 2: Materials & Methods) contains the gene for mScarlet-I which causes emission of a bright pink pigment. Other key features are also shown.

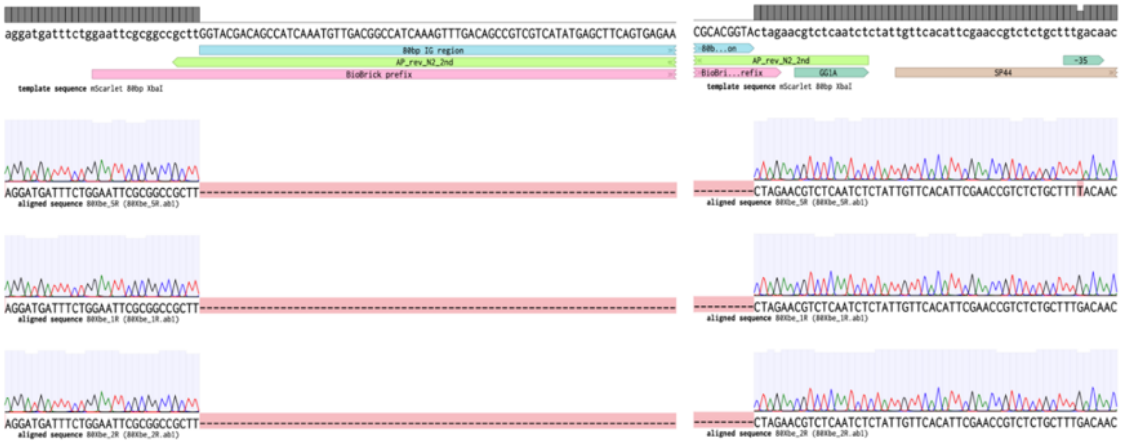
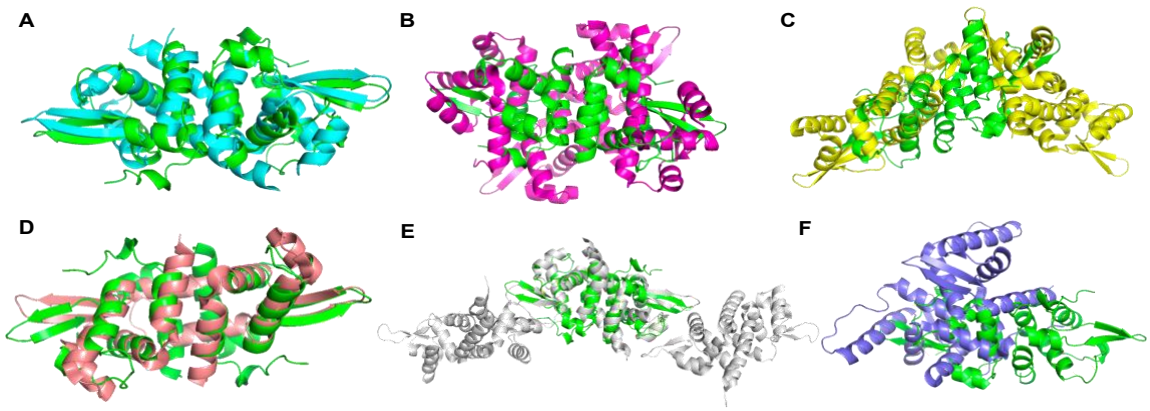


Figure A12. Sanger sequencing of mScarlet plasmid with 80 base pair (bp) DNA insert at the XbaI site. The 80 bp *mmyJ* and *mmr* intergenic DNA region was attempted to be inserted into a plasmid containing mScarlet at the XbaI cut site. Plasmids from colonies displaying bright pink, light pink and white phenotypes were digested by restriction digests first. One plasmid used for restriction digests of each of these different phenotype displaying colonies were sent for Sanger sequencing. Top to bottom shows the results from the plasmid from a bright pink, light pink and white colony respectively. None show the 80 bp DNA inserted. The ab1 file from sequencing was aligned against the predicted sequence in Benchling.

Figure A13. Sanger sequencing of mScarlet plasmid with 80 base pair (bp) DNA insert at various sites. The 80 bp *mmyJ* and *mmr* intergenic DNA region was attempted to be inserted into a plasmid containing mScarlet. Attempts were made at two other enzyme cut sites; prior to mScarlet gene after the RBS (NdeI site-**A**) and prior to mScarlet gene before the RBS (ScaI site-**B**). Multiple colonies were sequenced at different insert: vector ratios though through Sanger sequencing none appear to have worked. Three samples are given for each site. The ab1 file from sequencing was aligned against the predicted sequence in Benchling.

B: Supplementary figures for Chapter 4



Proteins identified as structurally similar to MmyJ were superimposed onto MmyJ using Pymol for better visualization of similar secondary features. In all MmyJ is represented in green. All multimers of proteins are included in this superimposition. **A** is SqrR, **B** is PagR, **C** is HlyU, **D** is *Af* ArsR, **E** is BigR and **F** is TubR. Some of these are part of the ArsR family of proteins. These superimpositions highlight that MmyJ may not be too structurally similar to these other ArsR family proteins in terms of their structural 3D folds (also shown by their low Q-score-Table 18).

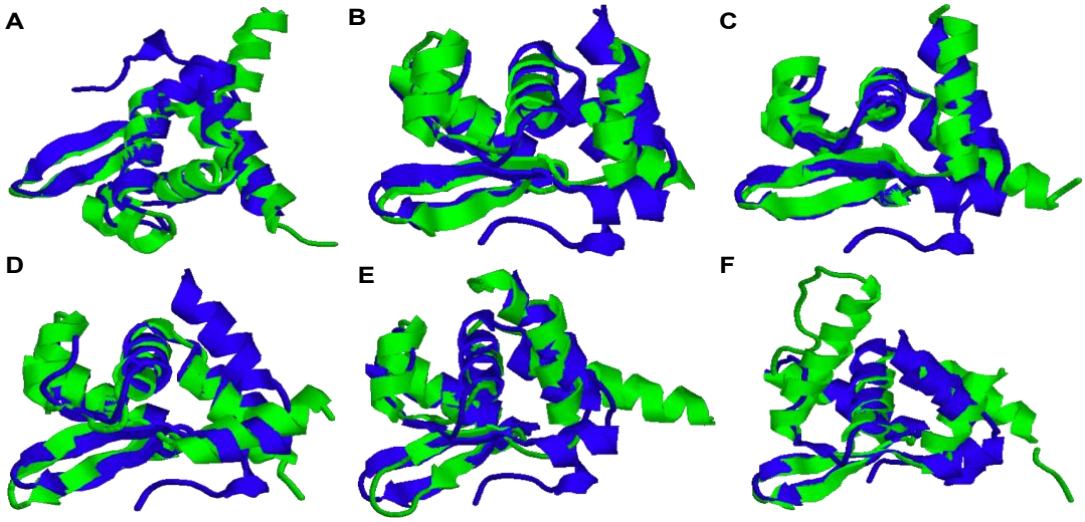


Figure B2. Superimposition of MmyJ (green) onto structurally similar proteins from PDBeFOLD (blue). Proteins identified as structurally similar to MmyJ were superimposed onto MmyJ by PDBeFold. The online viewer of PDBeFold (JSmol) was unavailable so superimpositions were visualized by RasMol. **A** is SqrR, **B** is PagR, **C** is HlyU, **D** is *Af* ArsR, **E** is BigR and **F** is TubR. These superimpositions highlight that MmyJ may not be too structurally similar to these other ArsR family proteins in terms of their structural 3D folds (also shown by their low Q-score-Table 18).

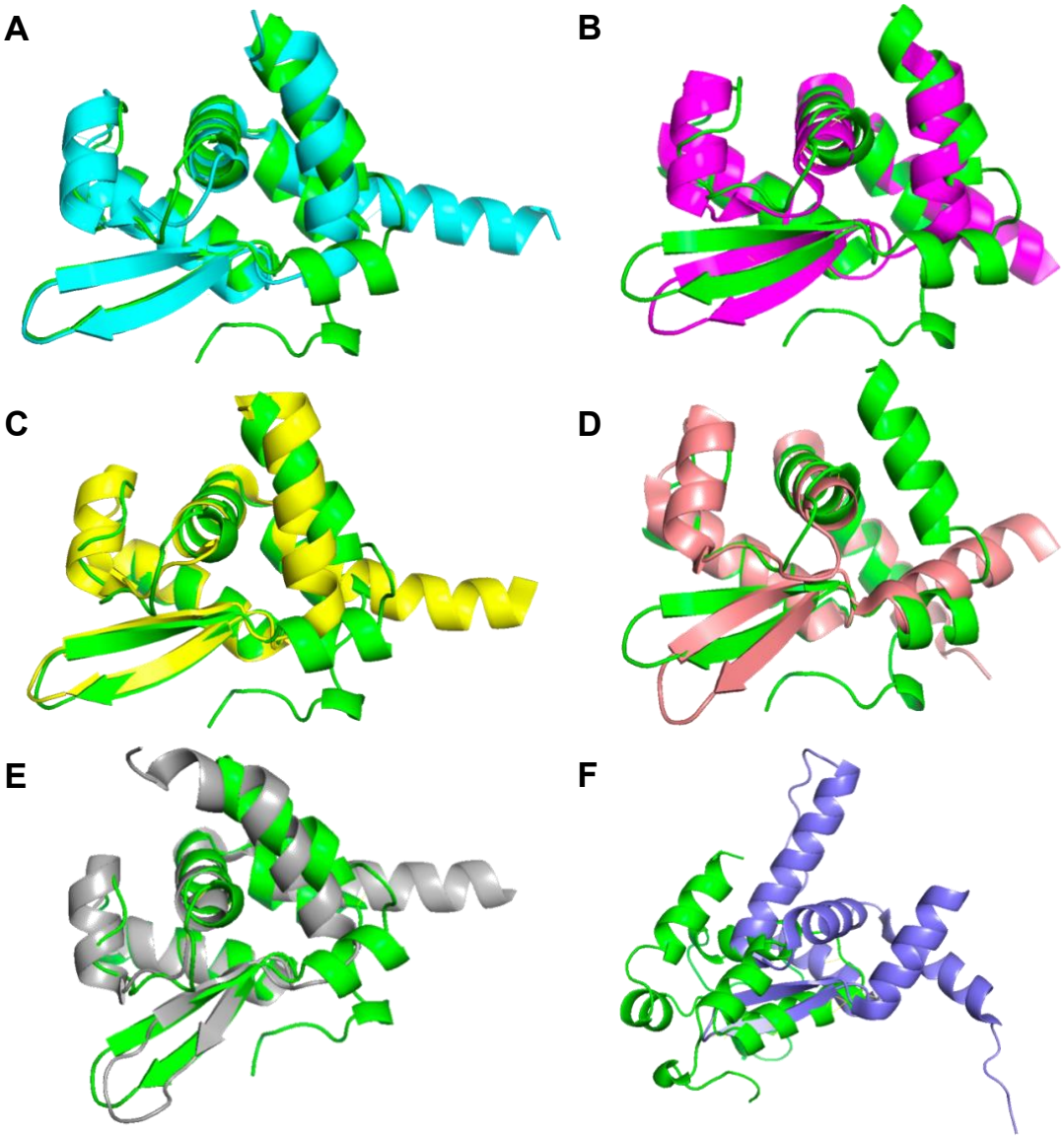


Figure B3. Superimposition of one monomer of MmyJ onto one monomer of structurally similar proteins from PDBeFOLD. 'Chain A's of proteins identified as structurally similar to MmyJ were superimposed onto one monomer of MmyJ using Pymol for better visualization of similar secondary features. In all MmyJ is represented in green. All multimers of proteins are included in this superimposition. **A** is SqrR, **B** is PagR, **C** is HlyU, **D** is *Af* ArsR, **E** is BigR and **F** is TubR. Some of these are part of the ArsR family of proteins. These superimpositions highlight that MmyJ may not be too structurally similar to these other ArsR family proteins in terms of their structural 3D folds (also shown by their low Q-score-Table 18).

Figure C1. Sanger sequencing of *mmyB* cloned into pET151. The *mmyB* gene was amplified by PCR using the cosmid C73-787 as a template and attempted to be cloned into the protein-expression plasmid pET151. Sanger sequencing was used to confirm the presence of the *mmyB* gene from various colonies. A sample of the sequencing data from various colonies taken are shown here. The ab1 file from sequencing was aligned against the predicted sequence in Benchling. While certain parts of the *mmyB* gene is seen, data is not strong and cloning proved unsuccessful.

Figure C2. Sanger sequencing of *E. coli* codon optimized *mmyB* cloned into pET151 and transformed into *E. coli* TOP10. A codon optimized *mmyB* gene for *E. coli* expression was purchased and cloned into pET151, transformed into *E. coli* TOP10 cells. Reverse and forward primers were used to confirm the presence of the *mmyB* gene. The ab1 file from sequencing was aligned against the predicted sequence in Benchling to find a perfect match.

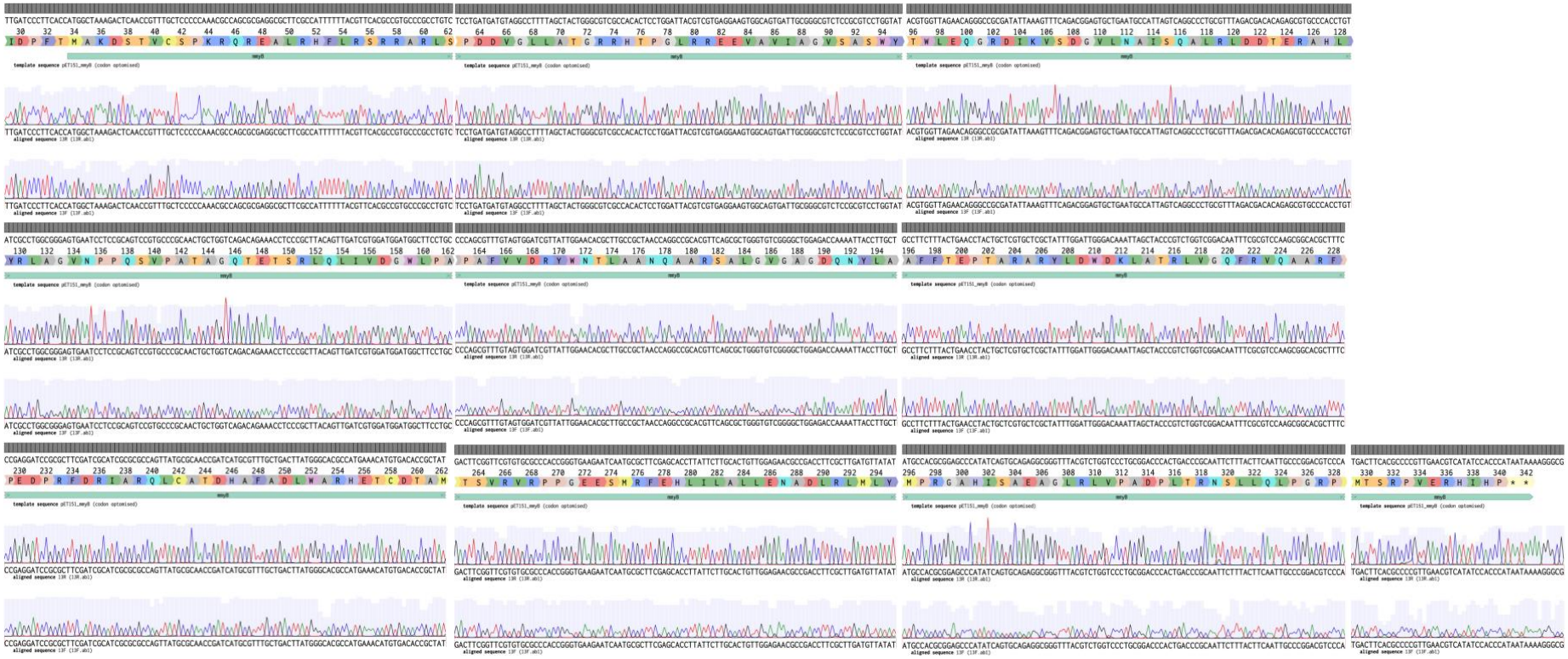


Figure C3. Sanger sequencing of *E. coli* codon optimized *mmyB* cloned into pET151 and transformed into *E. coli* BL21. A codon optimized *mmyB* gene for *E. coli* expression was purchased and cloned into pET151 and transformed into *E. coli* TOP10 cells. A colony containing the correct plasmid of these were then miniprepped and transformed into *E. coli* BL21 cells. To confirm the plasmid suffered no mutations during this transformation into BL21, Sanger sequencing was done with both reverse and forward primers. The ab1 file from sequencing was aligned against the predicted sequence in Benchling to find a match.

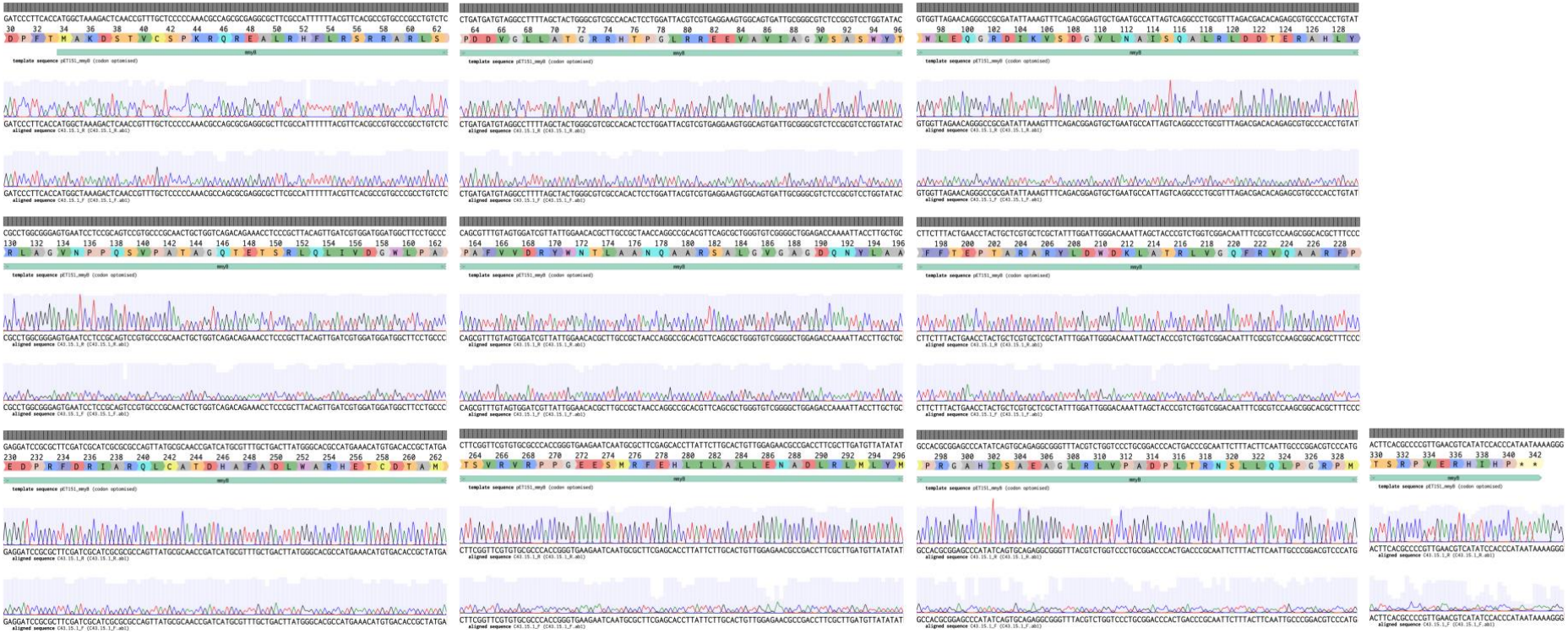


Figure C4. Sanger sequencing of *E. coli* codon optimized *mmyB* cloned into pET151 and transformed into *E. coli* C43. A codon optimized *mmyB* gene for *E. coli* expression was purchased and cloned into pET151 and transformed into *E. coli* TOP10 cells. A colony containing the correct plasmid of these were then miniprepped and transformed into *E. coli* C43 cells. To confirm the plasmid suffered no mutations during this transformation into C43, Sanger sequencing was done with both reverse and forward primers. The ab1 file from sequencing was aligned against the predicted sequence in Benchling to find a perfect match.

References

1. Bassett, E. J., Keith, M. S., Armelagos, G. J., Martin, D. L. & Villanueva, A. R. Tetracycline-Labeled Human Bone from Ancient Sudanese Nubia (A.D. 350). *Science (80-.).* **209**, 1532–1534 (1980).
2. Williams, K. J. The introduction of 'chemotherapy' using arsphenamine - The first magic bullet. *J. R. Soc. Med.* **102**, 343–348 (2009).
3. Ligon, B. L. Penicillin: Its Discovery and Early Development. *Semin. Pediatr. Infect. Dis.* **15**, 52–57 (2004).
4. Waksman, S. A. Production and Activity of Streptothricin. *J. Bacteriol.* **46**, 299–310 (1943).
5. Waksman, S. A. What is an Antibiotic or an Antibiotic Substance? *Mycologia* **39**, 565–569 (1947).
6. Watve, M. G., Tickoo, R., Jog, M. M. & Bhole, B. D. How many antibiotics are produced by the genus Streptomyces? *Arch. Microbiol.* **176**, 386–390 (2001).
7. Hopwood, D. A. *Streptomyces in Nature and Medicine: The Antibiotic Makers*. (Oxford University Press, 2007).
8. Hopwood, D. A. & Chater, K. F. Streptomyces. in *Bacillus subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics* (1993).
9. Flärdh, K. & Buttner, M. J. Streptomyces morphogenetics: Dissecting differentiation in a filamentous bacterium. *Nat. Rev. Microbiol.* **7**, 36–49 (2009).
10. Challis, G. L. Exploitation of the Streptomyces coelicolor A3(2) genome sequence for discovery of new natural products and biosynthetic pathways. *J. Ind. Microbiol. Biotechnol.* **41**, 219–232 (2014).
11. Challis, G. L. & Hopwood, D. A. Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by Streptomyces species. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14555–14561 (2003).
12. Traxler, M. F. & Kolter, R. Natural products in soil microbe interactions and evolution. *Nat. Prod. Rep.* **32**, 956–970 (2015).

13. Linares, J. F., Gustafsson, I., Baquero, F. & Martinez, J. L. Antibiotics as intermicrobiol signaling agents instead of weapons. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19484–19489 (2006).
14. Seipke, R. F., Kaltenpoth, M. & Hutchings, M. I. Streptomyces as symbionts: An emerging and widespread theme? *FEMS Microbiol. Rev.* **36**, 862–876 (2012).
15. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **417**, 141–147 (2002).
16. Ohnishi, Y. *et al.* Genome sequence of the streptomycin-producing microorganism Streptomyces griseus IFO 13350. *J. Bacteriol.* **190**, 4050–4060 (2008).
17. de Lima Procópio, R. E., da Silva, I. R., Martins, M. K., de Azevedo, J. L. & de Araújo, J. M. Antibiotics produced by Streptomyces. *Brazilian J. Infect. Dis.* **16**, 466–471 (2012).
18. Vivian, A. Genetic control of fertility in Streptomyces coelicolor A3(2): the IF fertility type. *J. Gen. Microbiol.* **64**, 101–117 (1970).
19. Kinashi, H., Murayama, M., Matsushita, H. & Nimi, O. Structural analysis of the giant linear plasmid SCP1 in various Streptomyces coelicolor strains. *J. Gen. Microbiol.* **139**, 1261–1269 (1993).
20. Kinashi, H., Mori, E., Hatani, A. & Nimi, O. Isolation and characterization of linear plasminds from lankacidin-producing streptomyces species. *J. Antibiot. (Tokyo).* **47**, 1447–1455 (1994).
21. Cornell, C. R., Marasini, D. & Fakhr, M. K. Molecular Characterization of Plasmids Harbored by Actinomycetes Isolated From the Great Salt Plains of Oklahoma Using PFGE and Next Generation Whole Genome Sequencing. *Front. Microbiol.* **9**, 1–13 (2018).
22. Bentley, S. D. *et al.* SCP1, a 356 023 bp linear plasmid adapted to the ecology and developmental biology of its host, Streptomyces coelicolor A3(2). *Mol. Microbiol.* **51**, 1615–1628 (2004).
23. Hopwood, D. A. Highlights of Streptomyces genetics. *Heredity (Edinb).* **123**, 23–32 (2019).
24. Bibb, M. J., Freeman, R. F. & Hopwood, D. A. Physical and genetical

characterisation of a second sex factor, SCP2, for Streptomyces coelicolor A3(2). *MGG Mol. Gen. Genet.* **154**, 155–166 (1977).

25. Schrempf, H., Bujard, H., Hopwood, D. A. & Goebel, W. Isolation of covalently closed circular deoxyribonucleic acid from Streptomyces coelicolor A3(2). *J. Bacteriol.* **121**, 416–421 (1975).
26. Bibb, M. J. & Hopwood, D. A. Genetic studies of the fertility plasmid SCP2 and its SCP2 variants in Streptomyces coelicolor A3(2). *J. Gen. Microbiol.* **126**, 427–442 (1981).
27. Haug, I. *et al.* Streptomyces coelicolor A3(2) plasmid SCP2*: Deductions from the complete sequence. *Microbiology* **149**, 505–513 (2003).
28. Wietzorrek, A. & Bibb, M. A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold. *Mol. Microbiol.* **25**, 1181–1184 (1997).
29. Lydiate, D. J., Malpartida, F. & Hopwood, D. A. The Streptomyces plasmid SCP2*: its functional analysis and development into useful cloning vectors. *Gene* **35**, 223–235 (1985).
30. Hutchings, M., Truman, A. & Wilkinson, B. Antibiotics: past, present and future. *Curr. Opin. Microbiol.* **51**, 72–80 (2019).
31. Moon, K., Xu, F., Zhang, C. & Seyedsayamdost, M. R. Bioactivity-HiTES Unveils Cryptic Antibiotics Encoded in Actinomycete Bacteria. *ACS Chem. Biol.* **14**, 767–774 (2019).
32. Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **13**, 509–523 (2015).
33. Bauman, K. D. *et al.* Refactoring the Cryptic Streptophenazine Biosynthetic Gene Cluster Unites Phenazine, Polyketide, and Nonribosomal Peptide Biochemistry. *Cell Chem. Biol.* **26**, 724-736.e7 (2019).
34. Takano, E., Chakraburtty, R., Nihira, T., Yamada, Y. & Bibb, M. J. A complex role for the γ-butyrolactone SCB1 in regulating antibiotic production in Streptomyces coelicolor A3(2). *Mol. Microbiol.* **41**, 1015–1028 (2001).
35. Bibb, M. J. Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.* **8**, 208–215 (2005).

36. Wei, J., He, L. & Niu, G. Regulation of antibiotic biosynthesis in actinomycetes: Perspectives and challenges. *Synth. Syst. Biotechnol.* **3**, 229–235 (2018).
37. Huang, J. *et al.* Cross-regulation among disparate antibiotic biosynthetic pathways of Streptomyces coelicolor. *Mol. Microbiol.* **58**, 1276–1287 (2005).
38. Liu, G., Chater, K. F., Chandra, G., Niu, G. & Tan, H. Molecular Regulation of Antibiotic Biosynthesis in Streptomyces. *Microbiol. Mol. Biol. Rev.* **77**, 112–143 (2013).
39. Kong, D., Wang, X., Nie, J. & Niu, G. Regulation of Antibiotic Production by Signaling Molecules in Streptomyces. *Front. Microbiol.* **10**, 1–11 (2019).
40. Bystrykh, L. V. *et al.* Production of actinorhodin-related 'blue pigments' by Streptomyces coelicolor A3(2). *J. Bacteriol.* **178**, 2238–2244 (1996).
41. WRIGHT, L. F. & HOPWOOD, D. A. Actinorhodin is a Chromosomally-determined Antibiotic in Streptomyces coelicolor A3(2). *J. Gen. Microbiol.* **96**, 289–297 (2009).
42. Okamoto, S., Taguchi, T., Ochi, K. & Ichinose, K. Biosynthesis of Actinorhodin and Related Antibiotics: Discovery of Alternative Routes for Quinone Formation Encoded in the act Gene Cluster. *Chem. Biol.* **16**, 226–236 (2009).
43. Hopwood, D. A. Genetic contributions to understanding polyketide synthases. *Chem. Rev.* **97**, 2465–2497 (1997).
44. Fernández-Moreno, M. A. *et al.* DNA sequence and functions of the actVI region of the actinorhodin biosynthetic gene cluster of Streptomyces coelicolor A3(2). *J. Biol. Chem.* **269**, 24854–24863 (1994).
45. Rudd, B. A. M. & Hopwood, D. A. Genetics of actinorhodin biosynthesis by Streptomyces coelicolor A3(2). *J. Gen. Microbiol.* **114**, 35–43 (1979).
46. Hallam, S. E., Malpartida, F. & Hopwood, D. A. Nucleotide sequence, transcription and deduced function of a gene involved in polyketide antibiotic synthesis in Streptomyces coelicolor. *Gene* **74**, 305–320 (1988).
47. Iqbal, M. *et al.* Extracting regulator activity profiles by integration of de novo motifs and expression data: Characterizing key regulators of nutrient depletion responses in Streptomyces coelicolor. *Nucleic Acids Res.* **40**, 5227–5239 (2012).

48. Fernández-Moreno, M. A., Caballero, J. L., Hopwood, D. A. & Malpartida, F. The act cluster contains regulatory and antibiotic export genes, direct targets for translational control by the bldA tRNA gene of streptomyces. *Cell* **66**, 769–780 (1991).
49. Tahlan, K., Yu, Z., Xu, Y., Davidson, A. R. & Nodwell, J. R. Ligand Recognition by ActR, a TetR-Like Regulator of Actinorhodin Export. *J. Mol. Biol.* **383**, 753–761 (2008).
50. Xu, Y., Willems, A., Au-Yeung, C., Tahlan, K. & Nodwell, J. R. A two-step mechanism for the activation of actinorhodin export and resistance in Streptomyces coelicolor. *MBio* **3**, (2012).
51. Caballero, J. L., Martinez, E., Malpartida, F. & Hopwood, D. A. Organisation and functions of the actV A region of the actinorhodin biosynthetic gene cluster of Streptomyces coelicolor. *MGG Mol. Gen. Genet.* **230**, 401–412 (1991).
52. Harrison, K. J., Crécy-Lagard, V. De & Zallot, R. Gene Graphics: A genomic neighborhood data visualization web application. *Bioinformatics* **34**, 1406–1408 (2018).
53. Wright, L. F. & Hopwood, D. A. Identification of the antibiotic determined by the SCP1 plasmid of Streptomyces coelicolor A3(2). *J. Gen. Microbiol.* **95**, 96–106 (1976).
54. Takano, E. *et al.* Transcriptional regulation of the redD transcriptional activator gene accounts for growth-phase-dependent production of the antibiotic undecylprodigiosin in Streptomyces coelicolor A3(2). *Mol. Microbiol.* **6**, 2797–2804 (1992).
55. Narva, K. E. & Feitelson, J. S. Nucleotide sequence and transcriptional analysis of the redD locus of Streptomycescoelicolor A3(2). *J. Bacteriol.* **172**, 326–333 (1990).
56. Guthrie, E. P. *et al.* A response-regulator-like activator of antibiotic synthesis from Streptomyces coelicolor A3(2) with an amino-terminal domain that lacks a phosphorylation pocket. *Microbiology* **144**, 727–738 (1998).
57. White, J. & Bibb, M. bldA dependence of undecylprodigiosin production in Streptomyces coelicolor A3(2) involves a pathway-specific regulatory

cascade. *J. Bacteriol.* **179**, 627–633 (1997).

58. Styles, K. M. Investigating Interactions Between Methylenomycin Furan Microbial Hormones And Transcriptional Repressors in Streptomyces coelicolor. (2016).
59. Sidda, J. D. *et al.* Discovery of a family of γ-aminobutyrate ureas via rational derepression of a silent bacterial gene cluster. *Chem. Sci.* **5**, 86–89 (2014).
60. Bednarz, B., Kotowska, M. & Pawlik, K. J. Multi-level regulation of coelimycin synthesis in Streptomyces coelicolor A3(2). *Appl. Microbiol. Biotechnol.* **103**, 6423–6434 (2019).
61. Bednarz, B. *et al.* Coelimycin Synthesis Activatory Proteins Are Key Regulators of Specialized Metabolism and Precursor Flux in Streptomyces coelicolor A3(2). *Front. Microbiol.* **12**, 1–17 (2021).
62. Hojati, Z. *et al.* Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from Streptomyces coelicolor. *Chem. Biol.* **9**, 1175–1187 (2002).
63. Hopwood, D. A. & Wright, H. M. CDA is a new chromosomally-determined antibiotic from Streptomyces coelicolor A3(2). *J. Gen. Microbiol.* **129**, 3575–3579 (1983).
64. Ryding, J. J., Anderson, T. B. & Champness, W. C. Regulation of the Streptomyces coelicolor calcium-dependent antibiotic by absA, encoding a cluster-linked two-component system. *J. Bacteriol.* **184**, 794–805 (2002).
65. McKenzie, N. L. & Nodwell, J. R. Phosphorylated AbsA2 negatively regulates antibiotic production in Streptomyces coelicolor through interactions with pathway-specific regulatory gene promoters. *J. Bacteriol.* **189**, 5284–5292 (2007).
66. Shu, D. *et al.* afsQ1-Q2-sigQ is a pleiotropic but conditionally required signal transduction system for both secondary metabolism and morphological development in Streptomyces coelicolor. *Appl. Microbiol. Biotechnol.* **81**, 1149–1160 (2009).
67. Santos-Beneit, F., Rodríguez-García, A., Sola-Landa, A. & Martín, J. F. Cross-talk between two global regulators in Streptomyces: PhoP and AfsR interact in the control of afsS, pstS and phoRP transcription. *Mol. Microbiol.* **72**, 53–

68 (2009).

68. Horinouchi, S. *et al.* Primary structure of AfsR, a global regulatory protein for secondary metabolite formation in Streptomyces coelicolor A3(2). *Gene* **95**, 49–56 (1990).
69. Xia, H. *et al.* The Application of Regulatory Cascades in Streptomyces: Yield Enhancement and Metabolite Mining. *Front. Microbiol.* **11**, 1–14 (2020).
70. Xu, G. *et al.* 'Pseudo' γ-butyrolactone receptors respond to antibiotic signals to coordinate antibiotic biosynthesis. *J. Biol. Chem.* **285**, 27440–27448 (2010).
71. Niu, G., Chater, K. F., Tian, Y., Zhang, J. & Tan, H. Specialised metabolites regulating antibiotic biosynthesis in Streptomyces spp. *FEMS Microbiol. Rev.* **40**, 554–573 (2016).
72. Martín, J. F. & Liras, P. Cascades and networks of regulatory genes that control antibiotic biosynthesis. *Subcell. Biochem.* **64**, 115–138 (2012).
73. Wang, L. *et al.* Autoregulation of antibiotic biosynthesis by binding of the end product to an atypical response regulator. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8617–8622 (2009).
74. Tian, Y., Fowler, K., Findlay, K., Tan, H. & Chater, K. F. An unusual response regulator influences sporulation at early and late stages in Streptomyces coelicolor. *J. Bacteriol.* **189**, 2873–2885 (2007).
75. Ulijasz, A. T., Andes, D. R., Glasner, J. D. & Weisblum, B. Regulation of iron transport in Streptococcus pneumoniae by RitR, an orphan response regulator. *J. Bacteriol.* **186**, 8123–8136 (2004).
76. Zhang, Y. *et al.* JadR*-mediated feed-forward regulation of cofactor supply in jadomycin biosynthesis. *Mol. Microbiol.* **90**, 884–897 (2013).
77. Zhang, Y. Y., Zou, Z. Z., Niu, G. Q. & Tan, H. R. jadR* and jadR2 act synergistically to repress jadomycin biosynthesis. *Sci. China Life Sci.* **56**, 584–590 (2013).
78. Li, Y. *et al.* Coordinative modulation of chlorothricin biosynthesis by binding of the glycosylated intermediates and end product to a responsive regulator ChlF1. *J. Biol. Chem.* **291**, 5406–5417 (2016).
79. Wu, H. *et al.* Cezomycin is activated by CalC to its ester form for further

biosynthesis steps in the production of calcimycin in Streptomyces chartreusis NRRL 3882. *Appl. Environ. Microbiol.* **84**, (2018).

80. Gou, L. *et al.* A novel TetR family transcriptional regulator, CalR3, negatively controls calcimycin biosynthesis in Streptomyces chartreusis NRRL 3882. *Front. Microbiol.* **8**, 1–10 (2017).
81. WHO. No time to wait: Securing the future from drug-resistant infections. *Artforum Int.* **54**, 113–114 (2019).
82. Courtenay, M. *et al.* Tackling antimicrobial resistance 2019–2024 – The UK's five-year national action plan. *J. Hosp. Infect.* **101**, 426–427 (2019).
83. Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. V. Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* **13**, 42–51 (2015).
84. Mak, S., Xu, Y. & Nodwell, J. R. The expression of antibiotic resistance genes in antibiotic-producing bacteria. *Mol. Microbiol.* **93**, 391–402 (2014).
85. Shinkawa, H., Sugiyama, M. & Nimi, O. The nucleotide sequence of a streptomycin 6-phosphotransferase gene from a streptomycin producer. *J. Gen. Microbiol.* **133**, 1289–1296 (1987).
86. Mansouri, K. & Piepersberg, W. Genetics of streptomycin production in Streptomyces griseus: nucleotide sequence of five genes, strFGHIK, including a phosphatase gene. *Mol. Gen. Genet. MGG* **228**, 459–469 (1991).
87. Ramirez, M. S. & Tolmasky, M. E. Aminoglycoside modifying enzymes. *Drug Resist. Updat.* **13**, 151–171 (2010).
88. Robicsek, A. *et al.* Fluoroquinolone-modifying enzyme: A new adaptation of a common aminoglycoside acetyltransferase. *Nat. Med.* **12**, 83–88 (2006).
89. Schwarz, S., Kehrenberg, C., Doublet, B. & Cloeckaert, A. Molecular basis of bacterial resistance to chloramphenicol and florfenicol. *FEMS Microbiol. Rev.* **28**, 519–542 (2004).
90. Wright, G. D. Bacterial resistance to antibiotics: Enzymatic degradation and modification. *Adv. Drug Deliv. Rev.* **57**, 1451–1470 (2005).
91. Viana Marques, D. de A. *et al.* Production of β-lactamase inhibitors by Streptomyces species. *Antibiotics* **7**, 1–26 (2018).
92. Ogawara, H. Penicillin-binding proteins in Actinobacteria. *J. Antibiot. (Tokyo).*

68, 223–245 (2014).

93. Sugiyama, M. & Kumagai, T. Molecular and structural biology of bleomycin and its resistance determinants. *J. Biosci. Bioeng.* **93**, 105–116 (2002).
94. Sugiyama, M. *et al.* Overproduction of the bleomycin-binding proteins from bleomycin-producing Streptomyces verticillus and a methicillin-resistant Staphylococcus aureus in Escherichia coli and their immunological characterisation. *FEBS Lett.* **362**, 80–84 (1995).
95. Rudolf, J. D. *et al.* Crystal Structure of the Zorbamycin-Binding Protein ZbmA, the Primary Self-Resistance Element in Streptomyces flavoviridis ATCC21892. *Biochemistry* **54**, 6842–6851 (2015).
96. Du, L., Sánchez, C., Chen, M., Edwards, D. J. & Shen, B. The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem. Biol.* **7**, 623–642 (2000).
97. Tao, M. *et al.* The tallysomycin biosynthetic gene cluster from Streptoalloteichus hindustanus E465-94 ATCC 31158 unveiling new insights into the biosynthesis of the bleomycin family of antitumor antibiotics. *Mol. Biosyst.* **3**, 60–74 (2007).
98. Ogawara, H. Self-resistance in streptomyces, with special reference to β-lactam antibiotics. *Molecules* **21**, 1–30 (2016).
99. Roberts, M. C. Resistance to macrolide, lincosamide, streptogramin, ketolide, and oxazolidinone antibiotics. *Appl. Biochem. Biotechnol. - Part B Mol. Biotechnol.* **28**, 47–62 (2004).
100. Redgrave, L. S., Sutton, S. B., Webber, M. A. & Piddock, L. J. V. Fluoroquinolone resistance: Mechanisms, impact on bacteria, and role in evolutionary success. *Trends Microbiol.* **22**, 438–445 (2014).
101. Lambert, P. A. Bacterial resistance to antibiotics: Modified target sites. *Adv. Drug Deliv. Rev.* **57**, 1471–1485 (2005).
102. Kumar, A. & Schweizer, H. P. Bacterial resistance to antibiotics: Active efflux and reduced uptake. *Adv. Drug Deliv. Rev.* **57**, 1486–1513 (2005).
103. Lambert, P. A. Cellular impermeability and uptake of biocides and antibiotics

in Gram-positive bacteria and mycobacteria. *J. Appl. Microbiol. Symp. Suppl.* **92**, 46–54 (2002).

104. Centers for Disease Control and Prevention (CDC). Vancomycin-resistant Staphylococcus aureus--Pennsylvania, 2002. *MMWR. Morb. Mortal. Wkly. Rep.* **51**, 902 (2002).
105. Schindler, B. D. & Kaatz, G. W. Multidrug efflux pumps of Gram-positive bacteria. *Drug Resist. Updat.* **27**, 1–13 (2016).
106. C Reygaert, W. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiol.* **4**, 482–501 (2018).
107. Elbourne, L. D. H., Tetu, S. G., Hassan, K. A. & Paulsen, I. T. TransportDB 2.0: A database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res.* **45**, D320–D324 (2017).
108. Kumar, S., Mukherjee, M. M. & Varela, M. F. Modulation of Bacterial Multidrug Resistance Efflux Pumps of the Major Facilitator Superfamily. *Int. J. Bacteriol.* **2013**, 1–15 (2013).
109. Liu, J., Keelan, P., Bennett, P. M. & Enne, V. I. Characterization of a novel macrolide efflux gene, mef (B), found linked to sul3 in porcine Escherichia coli. *J. Antimicrob. Chemother.* **63**, 423–426 (2009).
110. Cattoir, V., Poirel, L. & Nordmann, P. Plasmid-mediated quinolone resistance pump QepA2 in an Escherichia coli isolate from France. *Antimicrob. Agents Chemother.* **52**, 3801–3804 (2008).
111. Nag, A. & Mehra, S. A Major Facilitator Superfamily (MFS) Efflux Pump, SCO4121, from Streptomyces coelicolor with Roles in Multidrug Resistance and Oxidative Stress Tolerance and Its Regulation by a MarR Regulator. *Appl. Environ. Microbiol.* **87**, 1–23 (2021).
112. Zhou, Z., Sun, N., Wu, S., Li, Y. Q. & Wang, Y. Genomic data mining reveals a rich repertoire of transport proteins in Streptomyces. *BMC Genomics* **17**, (2016).
113. Sun, J., Deng, Z. & Yan, A. Bacterial multidrug efflux pumps: Mechanisms, physiology and pharmacological exploitations. *Biochem. Biophys. Res. Commun.* **453**, 254–267 (2014).
114. Leppard, S. . & Berg, J. . *Principles of Bioinorganic Chemistry*. (University

Science Books, 1994).

115. Ma, Z., Jacobson, F. E. & Giedroc, D. P. Metal Transporters and Metal Sensors: How Coordination Chemistry Controls Bacterial Metal Homeostasis. *October* **109**, 4644–4681 (2009).
116. Aljerf, L. A Gateway to Metal Resistance: Bacterial Response to Heavy Metal Toxicity in the Biological Environment. *Ann. Adv. Chem.* 032–044 (2018) doi:10.29328/journal.aac.1001012.
117. Albarracín, V. H., Ávila, A. L., Amoroso, M. J. & Abate, C. M. Copper removal ability by Streptomyces strains with dissimilar growth patterns and endowed with cupric reductase activity. *FEMS Microbiol. Lett.* **288**, 141–148 (2008).
118. Amoroso, M. J., Schubert, D., Mitscherlich, P., Schumann, P. & Kothe, E. Evidence for high affinity nickel transporter genes in heavy metal resistant Streptomyces spec. *J. Basic Microbiol.* **40**, 295–301 (2000).
119. Amoroso, M. J. *et al.* Screening of heavy metal-tolerant actinomycetes isolated from the Sali River. *J. Gen. Appl. Microbiol.* **44**, 129–132 (1998).
120. Amoroso, M. J. *et al.* Chromium accumulation by two Streptomyces spp. isolated from riverine sediments. *J. Ind. Microbiol. Biotechnol.* **26**, 210–215 (2001).
121. Giedroc, D. P. & Arunkumar, A. I. Metal sensor proteins: Nature's metalloregulated allosteric switches. *Dalt. Trans.* 3107–3120 (2007) doi:10.1039/b706769k.
122. Osman, D. & Cavet, J. S. Bacterial metal-sensing proteins exemplified by ArsR-SmtB family repressors. *Nat. Prod. Rep.* **27**, 668–680 (2010).
123. Ansari, A. ., Chael, M. L. & O'Halloran, T. V. Allosteric underwinding of DNA is a critical step in positive control of transcription by Hg-MerR. *Lett. to Nat.* **355**, 87–89 (1992).
124. Pennella, M. A. & Giedroc, D. P. Structural determinants of metal selectivity in prokaryotic metal-responsive transcriptional regulators. *BioMetals* **18**, 413–428 (2005).
125. Mukherjee, D., Datta, A. B. & Chakrabarti, P. Crystal structure of HlyU, the hemolysin gene transcription activator, from Vibrio cholerae N16961 and functional implications. *Biochim. Biophys. Acta - Proteins Proteomics* **1844**,

2346–2354 (2014).

126. Pomposiello, P. J. & Demple, B. Redox-operated genetic switches: The SoxR and OxyR transcription factors. *Trends Biotechnol.* **19**, 109–114 (2001).

127. Ahmed, M., Borsch, C. M., Taylor, S. S., Vázquez-Laslop, N. & Neyfakh, A. A. A protein that activates expression of a multidrug efflux transporter upon binding the transporter substrates. *J. Biol. Chem.* **269**, 28506–28513 (1994).

128. Bouzat, J. L. & Hoostal, M. J. Evolutionary analysis and lateral gene transfer of two-component regulatory systems associated with heavy-metal tolerance in bacteria. *J. Mol. Evol.* **76**, 267–279 (2013).

129. Juhas, M. *et al.* Genomic islands: Tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* **33**, 376–393 (2009).

130. Partridge, S. R., Kwong, S. M., Firth, N. & Jensen, S. O. Mobile genetic elements associated with antimicrobial resistance. *Clin. Microbiol. Rev.* **31**, 1–61 (2018).

131. Silver, S. *et al.* Inducible plasmid-determined resistance to arsenate, arsenite, and antimony(III) in Escherichia coli and Staphylococcus aureus. *J. Bacteriol.* **146**, 983–996 (1981).

132. Liebert, C. A., Hall, R. M. & Summers, A. O. Transposon Tn 21 , Flagship of the Floating Genome . *Microbiol. Mol. Biol. Rev.* **63**, 507–522 (1999).

133. Zhaxybayeva, O. Detection and quantitative assessment of horizontal gene transfer. *Methods Mol. Biol.* **532**, 195–213 (2009).

134. Li, L. G., Xia, Y. & Zhang, T. Co-occurrence of antibiotic and metal resistance genes revealed in complete genome collection. *ISME J.* **11**, 651–662 (2017).

135. Baker-Austin, C., Wright, M. S., Stepanauskas, R. & McArthur, J. V. Co-selection of antibiotic and metal resistance. *Trends Microbiol.* **14**, 176–182 (2006).

136. Zhang, Y. *et al.* Sub-inhibitory concentrations of heavy metals facilitate the horizontal transfer of plasmid-mediated antibiotic resistance genes in water environment. *Environ. Pollut.* **237**, 74–82 (2018).

137. Li, X. *et al.* Sub-lethal concentrations of heavy metals induce antibiotic resistance via mutagenesis. *J. Hazard. Mater.* **369**, 9–16 (2019).

138. Álvarez, A., Catalano, S. A. & Amoroso, M. J. Heavy metal resistant strains are

widespread along Streptomyces phylogeny. *Mol. Phylogenet. Evol.* **66**, 1083–1088 (2013).

139. Doroghazi, J. R. & Buckley, D. H. Widespread homologous recombination within and between Streptomyces species. *ISME J.* **4**, 1136–1143 (2010).
140. HANEISHI, T., TERAHARA, A., HAMANO, K. & ARAI, M. NEW ANTIBIOTICS, METHYLENOMYCINS A AND B. *J. Antibiot. (Tokyo).* **27**, 400–407 (1974).
141. Jernow, J., Tautz, W., Rosen, P. & Williams, T. H. Methylenomycin B: revised structure and total syntheses. *J. Org. Chem.* **44**, 4212–4213 (1979).
142. Chater, K. F. & Bruton, C. J. Resistance, regulatory and production genes for the antibiotic methylenomycin are clustered. *EMBO J.* **4**, 1893–1897 (1985).
143. Hopwood, D. A. & Kirby, R. Genetic Determination of Methylenomycin Synthesis by the SCP1 Plasmid of Streptomyces coelicolor A3(2). *Microbiology* **98**, 239–252 (1977).
144. Neal, R. J. & Chater, K. F. Bidirectional promoter and terminator regions bracket mmr, a resistance gene embedded in the Streptomyces coelicolor A3(2) gene cluster encoding methylenomycin production. *Gene* **100**, 75–83 (1991).
145. Idowu, G. A. Investigations into the biosynthesis and mode of action of methylenomycin antibiotics from Streptomyces coelicolor. (2017).
146. Corre, C. & Challis, G. L. Evidence for the unusual condensation of a diketide with a pentulose in the methylenomycin biosynthetic pathway of Streptomyces coelicolor A3(2). *ChemBioChem* **6**, 2166–2170 (2005).
147. Hornemann, U. & Hopwood, D. A. Biosynthesis of Methylenomycin A: A Plasmid-Determined Antibiotic. (1981).
148. O'Rourke, S. *et al.* Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of methylenomycin production in Streptomyces coelicolor. *Mol. Microbiol.* **71**, 763–778 (2009).
149. Chater, K. F. & Merrick, M. J. Approaches to the study of differentiation in Streptomyces coelicolor A3 (2). *Proc. Int. Symp. Genet. Ind. Microorg.* (1976).
150. O'Rourke, S. *et al.* Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of

methylenomycin production in Streptomyces coelicolor. *Mol. Microbiol.* **71**, 763–778 (2009).

151. Corre, C., Song, L., O'Rourke, S., Chater, K. F. & Challis, G. L. 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by Streptomyces coelicolor genome mining. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 17510–17515 (2008).
152. Li, X. *et al.* ScbR-and ScbR2-mediated signal transduction networks coordinate complex physiological responses in Streptomyces coelicolor. *Sci. Rep.* **5**, (2015).
153. Zhou, S. *et al.* Molecular basis for control of antibiotic production by a bacterial hormone. *Nature* **590**, 463–467 (2021).
154. Hobbs, G. *et al.* An integrated approach to studying regulation of production of the antibiotic methylenomycin by Streptomyces coelicolor A3(2). *J. Bacteriol.* **174**, 1487–1494 (1992).
155. Lougher, M. Functional and structural insights into MmyJ, An ArsR-Like Transcriptional Repressor. (2015).
156. Laskaris, P., Tolba, S., Calvo-Bado, L. & Wellington, L. Coevolution of antibiotic production and counter-resistance in soil bacteria. *Environ. Microbiol.* **12**, 783–796 (2010).
157. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
158. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
159. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320 (2008).
160. Le, S. Q., Dang, C. C. & Gascuel, O. Modeling protein evolution with several amino acid replacement matrices depending on site rates. *Mol. Biol. Evol.* **29**, 2921–2936 (2012).
161. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8**, 275–282 (1992).
162. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: An online tool for

phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

163. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
164. Winter, G. Xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
165. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 271–281 (2011).
166. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).
167. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
168. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
169. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
170. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).
171. Lovell, S. C. *et al.* Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins Struct. Funct. Genet.* **50**, 437–450 (2003).
172. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 95–99 (1963).
173. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 283–291 (1993).
174. Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. MolProbity: Structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32**, 615–619 (2004).
175. Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. A bacterial two-hybrid

system based on a reconstituted signal transduction pathway. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5752–5756 (1998).

176. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *PRACTICAL STREPTOMYCES GENETICS*. (2000).
177. Aldali, J. Investigating the role of MmyB-like pathwayspecific transcriptional activators in antibiotic production. 157 (2018).
178. Strauch, E., Takano, E., Baylts, H. A. & Bibb, M. J. The stringent response in Streptomyces coelicolor A3(2). *Mol. Microbiol.* **5**, 289–298 (1991).
179. Thompson, C. J., Ward, J. M. & Hopwood, D. A. DNA cloning in Streptomyces: Resistance genes from antibiotic-producing species. *Nature* **286**, 525–527 (1980).
180. Elbing, K. & Brent, R. Recipes and tools for culture of Escherichia coli. *Curr Protoc Mol Biol.* **125**, e83 (2019).
181. Berg, J. M., Tymoczko, J. L. & Stryer, L. Secondary structure: Polypeptide chains can fold into regular structures such as the alpha helix, the beta sheet, and turns and loops. in *Biochemistry* (W H Freeman, 2002).
182. O'Halloran, T. V. Interrelations among metal ions, enzymes, and gene expression. In: Metal Ions in Biological Systems. in *Sigel, H. and Sigel, A., Eds.* 105–146 (Marcel Dekker, 1989).
183. Campbell, D. R. *et al.* Mycobacterial cells have dual nickel-cobalt sensors: Sequence relationships and metal sites of metal-responsive repressors are not congruent. *J. Biol. Chem.* **282**, 32298–32310 (2007).
184. Liu, T. *et al.* A CuI-sensing ArsR family metal sensor protein with a relaxed metal selectivity profile. *Biochemistry* **47**, 10564–10575 (2008).
185. Xu, C. & Rosen, B. P. Dimerization is essential for DNA binding and repression by the ArsR metalloregulatory protein of Escherichia coli. *J. Biol. Chem.* **272**, 15734–15738 (1997).
186. Xu, C., Shi, W. & Rosen, B. P. The chromosomal arsR gene of Escherichia coli encodes a trans-acting metalloregulatory protein. *J. Biol. Chem.* **271**, 2427–2432 (1996).
187. Carlin, A., Shi, W., Dey, S. & Rosen, B. P. The ars operon of Escherichia coli confers arsenical and antimonial resistance. *J. Bacteriol.* **177**, 981–986

(1995).

188. Shi, W., Wu, J. & Rosen, B. P. Identification of a putative metal binding site in a new family of metalloregulatory proteins. *J. Biol. Chem.* **269**, 19826–19829 (1994).

189. Shi, W., Dong, J., Scott, R. A., Ksenzenko, M. Y. & Rosen, B. P. The role of arsenic-thiol interactions in metalloregulation of the ars operon. *J. Biol. Chem.* **271**, 9291–9297 (1996).

190. Cook, W. J., Kar, S. R., Taylor, K. B. & Hall, L. M. Crystal structure of the cyanobacterial metallothionein repressor SmtB: A model for metalloregulatory proteins. *J. Mol. Biol.* **275**, 337–346 (1998).

191. Cavet, J. S. *et al.* A nickel-cobalt-sensing ArsR-SmtB family repressor: Contributions of cytosol and effector binding sites to metal selectivity. *J. Biol. Chem.* **277**, 38441–38448 (2002).

192. Turner, J. S., Glands, P. D., Samson, A. C. R. & Robinson, N. J. Zn2+-sensing by the cyanobacterial metallothionein repressor SmtB: Different motifs mediate metal-induced protein-DNA dissociation. *Nucleic Acids Res.* **24**, 3714–3721 (1996).

193. Eicken, C. *et al.* A metal-ligand-mediated intersubunit allosteric switch in related SmtB/ ArsR zinc sensor proteins. *J. Mol. Biol.* **333**, 683–695 (2003).

194. VanZile, M. L., Chen, X. & Giedroc, D. P. Allosteric negative regulation of smt O/P binding of the zinc sensor, SmtB, by metal ions: A coupled equilibrium analysis. *Biochemistry* **41**, 9776–9786 (2002).

195. Ye, J., Kandegedara, A., Martin, P. & Rosen, B. P. Crystal structure of the Staphylococcus aureus pI258 CadC Cd(II)/Pb(II)/Zn(II)-responsive repressor. *J. Bacteriol.* **187**, 4214–4221 (2005).

196. Busenlehner, L. S. *et al.* Spectroscopic properties of the metalloregulatory Cd(II) and Pb(II) sites of S. aureus pI258 CadC. *Biochemistry* **40**, 4426–4436 (2001).

197. Busenlehner, L. S., Weng, T. C., Penner-Hahn, J. E. & Giedroc, D. P. Elucidation of primary (α3N) and vestigial (α5) heavy metal-binding sites in Staphylococcus aureus pI258 CadC: Evolutionary implications for metal ion selectivity of ArsR/SmtB metal sensor proteins. *J. Mol. Biol.* **319**, 685–701

(2002).

198. Wong, M. D., Lin, Y. F. & Rosen, B. P. The soft metal ion binding sites in the Staphylococcus aureus pI258 CadC Cd(II)Pb(II)/Zn(II)-responsive repressor are formed between subunits of the homodimer. *J. Biol. Chem.* **277**, 40930–40936 (2002).
199. Qin, J. *et al.* Convergent evolution of a new arsenic binding site in the ArsR/SmtB family of metalloregulators. *J. Biol. Chem.* **282**, 34346–34355 (2007).
200. Busenlehner, L. S., Pennella, M. A. & Giedroc, D. P. The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. *FEMS Microbiol. Rev.* **27**, 131–143 (2003).
201. Bose, M. *et al.* Identification of SmtB/ArsR cis elements and proteins in archaea using the Prokaryotic InterGenic Exploration Database (PIGED). *Archaea* **2**, 39–49 (2006).
202. Chen, H. *et al.* Negative regulation of bleomycins biosynthesis by ArsR/SmtB family repressor BlmR in Streptomyces verticillus. *Appl. Microbiol. Biotechnol.* **103**, 6629–6644 (2019).
203. Singh, V. K. *et al.* ZntR is an autoregulatory protein and negatively regulates the chromosomal zinc resistance operon znt of Staphylococcus aureus. *Mol. Microbiol.* **33**, 200–207 (1999).
204. Nishi, K. *et al.* Crystal structure of the transcriptional activator HlyU from Vibrio vulnificus CMCP6. *FEBS Lett.* **584**, 1097–1102 (2010).
205. Williams, S. G., Attridge, S. R. & Manning, P. A. The transcriptional activator HIyU of Vibrio cholerae: nucleotide sequence and role in virulence gene expression. *Mol. Microbiol.* **9**, 751–760 (1993).
206. Otting, G. *et al.* Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* **9**, 3085–3092 (1990).
207. Brennan, R. G. & Matthews, B. W. The helix-turn-helix DNA binding motif. *J. Biol. Chem.* **264**, 1903–1906 (1989).
208. Poole, L. B. The basics of thiols and cysteines in redox biology and chemistry. *Free Radic. Biol. Med.* **80**, 148–157 (2015).

209. Guimarães, B. G. *et al.* Plant pathogenic bacteria utilize biofilm growth-associated repressor (BigR), a novel winged-helix redox switch, to control hydrogen sulfide detoxification under hypoxia. *J. Biol. Chem.* **286**, 26148–26157 (2011).
210. Barbosa, R. L. & Benedetti, C. E. BigR, a transcriptional repressor from plant-associated bacteria, regulates an operon implicated in biofilm growth. *J. Bacteriol.* **189**, 6185–6194 (2007).
211. Mao, X. M., Luo, S. & Li, Y. Q. Negative regulation of daptomycin production by DepR2, an ArsR-family transcriptional factor. *J. Ind. Microbiol. Biotechnol.* **44**, 1653–1658 (2017).
212. Jacquamet, L. *et al.* Structural characterization of the active form of PerR: Insights into the metal-induced activation of PerR and fur proteins for DNA binding. *Mol. Microbiol.* **73**, 20–31 (2009).
213. Liu, T. *et al.* CsoR is a novel Mycobacterium tuberculosis copper-sensing transcriptional regulator. *Nat. Chem. Biol.* **3**, 60–68 (2007).
214. Biology, N. R. C. (US) B. on. *Bioinformatics: Converting Data to Knowledge*. (National Academies Press, 2000). doi:10.17226/9990.
215. Kapli, P., Flouri, T. & Telford, M. J. Systematic errors in phylogenetic trees. *Curr. Biol.* **31**, R59–R64 (2021).
216. Gabaldón, T. Evolution of Proteins and Proteomes: A Phylogenetics Approach. *Evol. Bioinforma.* **1**, 117693430500100 (2005).
217. Ji, Y. *et al.* Phytoactinopolyspora halotolerans sp. nov., a halotolerant actinobacterium isolated from a saline soil in Xinjiang, northwest of China. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **111**, 27–34 (2018).
218. le Roes, M. & Meyers, P. R. Nonomuraea candida sp. nov., a new species from South African soil. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **93**, 133–139 (2008).
219. Saha, C. K., Pires, R. S., Brolin, H., Delannoy, M. & Atkinson, G. C. FlaGs and webFlaGs: Discovering novel biology through the analysis of gene neighbourhood conservation. *Bioinformatics* **37**, 1312–1314 (2021).
220. Kazantsev, R. & Towles, M. X-ray Crystallography - Chemistry LibreTexts. *Chemistry LibreTexts* (2017).

221. William Studier, F., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. [6] Use of T7 RNA polymerase to direct expression of cloned genes. in *Methods in Enzymology* vol. 185 60–89 (1990).
222. Gasteiger, E. *et al.* The Proteomics Protocols Handbook. *Proteomics Protoc. Handb.* 571–608 (2005) doi:10.1385/1592598900.
223. Hagel, L. Gel filtration: size exclusion chromatography. *Methods Biochem. Anal.* **54**, 51–91 (2011).
224. Stothard, P. The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences. *Biotechniques* **28**, 1102–1104 (2000).
225. Bai, C. *et al.* Exploiting a precise design of universal synthetic modular regulatory elements to unlock the microbial natural products in Streptomyces. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 12181–12186 (2015).
226. Bindels, D. S. *et al.* MScarlet: A bright monomeric red fluorescent protein for cellular imaging. *Nat. Methods* **14**, 53–56 (2016).
227. Gueuné, H., Durand, M. J., Thouand, G. & DuBow, M. S. The ygaVP genes of Escherichia coli form a tributyltin-inducible operon. *Appl. Environ. Microbiol.* **74**, 1954–1958 (2008).
228. Hoskisson, P. A. & van Wezel, G. P. Streptomyces coelicolor. *Trends Microbiol.* **27**, 468–469 (2019).
229. Visualizing TLC Plates. https://chem.libretexts.org/@go/page/93522 (2021).
230. Borodina, I. *et al.* Antibiotic overproduction in Streptomyces coelicolor A3(2) mediated by phosphofructokinase deletion. *J. Biol. Chem.* **283**, 25186–25199 (2008).
231. Craney, A. *et al.* A synthetic luxCDABE gene cluster optimized for expression in high-GC bacteria. *Nucleic Acids Res.* **35**, 1–10 (2007).
232. Powell, H. R. X-ray data processing. *Biosci. Rep.* **37**, 1–14 (2017).
233. Rupp, B. *Biomolecular Crystallography*. (Garland Science, 2009). doi:10.1201/9780429258756.
234. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 72–82 (2006).
235. Taylor, G. L. Introduction to phasing. *Acta Crystallogr. Sect. D Biol.*

Crystallogr. **66**, 325–338 (2010).

236. Rossman, M. G. & Blow, D. M. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.* 24–31 (1962).
237. Kovalevskiy, O., Nicholls, R. A., Long, F., Carlon, A. & Murshudov, G. N. Overview of refinement procedures within REFMAC 5: Utilizing data from different sources. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 215–227 (2018).
238. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D,Biological Crystallogr.* 240–255 (1997).
239. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
240. Tickle, I. J., Laskowski, R. A. & Moss, D. S. R(free) and the R(free) ratio. II. Calculation of the expected values and variances of cross-validation statistics in macromolecular least-squares refinement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56**, 442–450 (2000).
241. Tickle, I. J., Laskowski, R. A. & Moss, D. S. Rfree and the Rfree ratio. I. Derivation of expected values of cross-validation residuals used in macromolecular least-squares refinement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54**, 547–557 (1998).
242. Gunasekaran, K., Nagarajaram, H. A., Ramakrishnan, C. & Balaram, P. Stereochemical punctuation marks in protein structures: Glycine and proline containing helix stop signals. *J. Mol. Biol.* **275**, 917–932 (1998).
243. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).
244. Schrödinger LLC. The PyMOL Molecular Graphics System. Version 1.2r3pre.
245. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).
246. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2256–2268 (2004).
247. Capdevila, D. A. *et al.* Structural basis for persulfide-sensing specificity in a

transcriptional regulator. *Nat. Chem. Biol.* **17**, 65–70 (2021).

248. Zhao, H., Volkov, A., Veldore, V. H., Hoch, J. A. & Varughese, K. I. Crystal structure of the transcriptional repressor PagR of Bacillus anthracis. *Microbiology* **156**, 385–391 (2010).
249. Prabaharan, C., Kandavelu, P., Packianathan, C., Rosen, B. P. & Thiyagarajan, S. Structures of two ArsR As(III)-responsive transcriptional repressors: Implications for the mechanism of derepression. *J. Struct. Biol.* **207**, 209–217 (2019).
250. Guimarães, B. G. *et al.* Plant pathogenic bacteria utilize biofilm growth-associated repressor (BigR), a novel winged-helix redox switch, to control hydrogen sulfide detoxification under hypoxia. *J. Biol. Chem.* **286**, 26148–26157 (2011).
251. Hayashi, I., Oda, T., Sato, M. & Fuchigami, S. Cooperative DNA Binding of the Plasmid Partitioning Protein TubR from the Bacillus cereus pXO1 Plasmid. *J. Mol. Biol.* **430**, 5015–5028 (2018).
252. Campbell, D. R. *et al.* Mycobacterial cells have dual nickel-cobalt sensors: Sequence relationships and metal sites of metal-responsive repressors are not congruent. *J. Biol. Chem.* **282**, 32298–32310 (2007).
253. Arunkumar, A. I., Campanello, G. C. & Giedroc, D. P. Solution structure of a paradigm ArsR family zinc sensor in the DNA-bound state. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 18177–18182 (2009).
254. Ali, S., Hassan, M., Islam, A. & Ahmad, F. A Review of Methods Available to Estimate Solvent-Accessible Surface Areas of Soluble Proteins in the Folded and Unfolded States. *Curr. Protein Pept. Sci.* **15**, 456–476 (2014).
255. Qin, J. *et al.* Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2075–2080 (2006).
256. VanZile, M. L., Chen, X. & Giedroc, D. P. Structural characterization of distinct α3n and α5 metal sites in the cyanobacterial zinc sensor SmtB. *Biochemistry* **41**, 9765–9775 (2002).
257. Qin, J. *et al.* Convergent evolution of a new arsenic binding site in the ArsR/SmtB family of metalloregulators. *J. Biol. Chem.* **282**, 34346–34355

(2007).

258. Jernow, J., Tautz, W., Rosen, P. & Blount, J. F. Stereospecific total synthesis and absolute configuration of (+)-methylenomycin A. *J. Org. Chem.* **44**, 4210–4212 (1979).

259. Lee, S. G., Krishnan, H. B. & Jez, J. M. Structural basis for regulation of rhizobial nodulation and symbiosis gene expression by the regulatory protein NolR. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 6509–6514 (2014).

260. Barbosa, R. L. & Benedetti, C. E. BigR, a transcriptional repressor from plant-associated bacteria, regulates an operon implicated in biofilm growth. *J. Bacteriol.* **189**, 6185–6194 (2007).

261. Blanco, A. G., Sola, M., Gomis-Rüth, F. X. & Coll, M. Tandem DNA Recognition by PhoB, a two-component signal transduction transcriptional activator. *Structure* **10**, 701–713 (2002).

262. Solovyev V, S. A. Automatic annotation of microbial genomes and metagenomic sequences. In metagenomics and its applications in agriculture, biomedicine and environmental studies (Li RE, ed). *Nov. Sci. Publ.* 61–78 (2011).

263. Zhou, S. *et al.* Structural basis for control of antibiotic production by bacterial hormones. *Nature* submitted (2020).

264. Xu, Q. *et al.* Structure of an MmyB-like regulator from C. aurantiacus, member of a new transcription factor family linked to antibiotic metabolism in actinomycetes. *PLoS One* **7**, (2012).

265. Chater, K. F. & Chandra, G. The use of the rare UUA codon to define 'expression space' for genes involved in secondary metabolism, development and environmental adaptation in Streptomyces. *J. Microbiol.* **46**, 1–11 (2008).

266. Jeong, H. *et al.* Genome Sequences of Escherichia coli B strains REL606 and BL21(DE3). *J. Mol. Biol.* **394**, 644–652 (2009).

267. Lin, Z. & Rye, H. S. GroEL-Mediated Protein Folding: Making the Impossible, Possible. *Crit Rev Biochem Mol Biol.* **41**, 211–239 (2006).

268. Gatzeva-Topalova, P. Z., May, A. P. & Sousa, M. C. Structure and Mechanism of ArnA: Conformational Change Implies Ordered Dehydrogenase Mechanism in Key Enzyme for Polymyxin Resistance. *Structure* **13**, 929–942

(2005).

269. Bolanos-Garcia, V. M. & Davies, O. R. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochim. Biophys. Acta - Gen. Subj.* **1760**, 1304–1313 (2006).

270. Dumon-Seignovert, L., Cariot, G. & Vuillard, L. The toxicity of recombinant proteins in Escherichia coli: A comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr. Purif.* **37**, 203–206 (2004).

271. Katayama, H. *et al.* Strategies for folding of affinity tagged proteins using GroEL and osmolytes. *J. Struct. Funct. Genomics* **10**, 57–66 (2009).

272. Robichon, C., Luo, J., Causey, T. B., Benner, J. S. & Samuelson, J. C. Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. coli Protein contamination after purification by immobilized metal affinity chromatography. *Appl. Environ. Microbiol.* **77**, 4634–4646 (2011).

273. Voziyan, P. A., Jadhav, L. & Fisher, M. T. Refolding a glutamine synthetase truncation mutant in vitro: Identifying superior conditions using a combination of chaperonins and osmolytes. *J. Pharm. Sci.* **89**, 1036–1045 (2000).

274. Heras, B. & Martin, J. L. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61**, 1173–1180 (2005).