

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

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/90894>

**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](mailto:wrap@warwick.ac.uk)



Investigating Interactions Between  
Methylenomycin Furan Microbial Hormones  
And Transcriptional Repressors in  
*Streptomyces coelicolor*

Kathryn Maria Styles

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

School of Life Sciences, University of Warwick

Supervisors: Christophe Corre and Vilmos Fulop

September 2016



# Table of Contents

|   |              |
|---|--------------|
| <b>LIST OF FIGURES</b> .....  | <b>VI</b>    |
| <b>LIST OF TABLES</b> .....   | <b>XII</b>   |
| <b>ACKNOWLEDGEMENTS</b> .....   | <b>XV</b>    |
| <b>DECLARATION</b> .....  | <b>XVI</b>   |
| <b>ABSTRACT</b> .....   | <b>XVII</b>  |
| <b>ABBREVIATIONS</b> .....  | <b>XVIII</b> |
| <b>1 INTRODUCTION</b> .....   | <b>1</b>     |
| 1.1 <i>Streptomyces</i> and their Natural Products .....                    | 1            |
| 1.2 Regulation of Natural Product Biosynthesis in <i>Streptomyces</i> ..... | 3            |
| 1.2.1 <i>An Introduction to Microbial Hormones</i> .....                    | 3            |
| 1.2.2 <i>The Gamma-Butyrolactones</i> .....                                 | 4            |
| 1.2.3 <i>TetR Family Transcriptional Repressors</i> .....                   | 5            |
| 1.2.4 <i>Other Microbial Hormones</i> .....                                 | 7            |
| 1.3 <i>Streptomyces coelicolor</i> – A Model Organism .....                 | 8            |
| 1.3.1 <i>S. coelicolor Genetics</i> .....                                   | 8            |
| 1.3.2 <i>Methylenomycin</i> .....   | 9            |
| 1.4 Regulation of Methylenomycin Biosynthesis .....                         | 10           |
| 1.4.1 <i>MmfR – A TetR Family Transcriptional Repressor</i> .....           | 10           |
| 1.4.2 <i>MmyR and its Role in Transcriptional Regulation</i> .....          | 14           |
| 1.4.3 <i>Analogous Systems to MmfR/MmfLHP/MmyR</i> .....                    | 16           |
| 1.4.4 <i>Other Methylenomycin Control Mechanisms</i> .....                  | 17           |
| 1.5 Introduction to Research Aims .....                                     | 19           |
| 1.5.1 <i>How Will These Research Questions Be Answered?</i> .....           | 20           |
| 1.5.2 <i>Choice of Reporter System Used</i> .....                           | 21           |
| 1.5.3 <i>Why Do We Need A Novel Inducible Expression System?</i> .....      | 22           |
| 1.6 Outline of Thesis Structure .....                                       | 23           |
| <b>2 MATERIALS AND METHODS</b> .....  | <b>25</b>    |
| 2.1 Materials and Equipment .....   | 25           |
| 2.2 Stock Solutions .....   | 30           |
| 2.2.1 <i>Media Stock Solutions</i> .....                                    | 30           |
| 2.2.2 <i>DNA Gel Electrophoresis</i> .....                                  | 31           |

|          |  |           |
|----------|--|-----------|
| 2.2.3    | <i>Phenol Chloroform Extraction Buffers</i> .....  | 31        |
| 2.2.4    | <i>Protein Purification Buffers and Solutions</i> .....                                    | 31        |
| 2.2.5    | <i>SDS-PAGE Reagents and Buffers</i> .....   | 32        |
| 2.3      | Protocols .....  | 32        |
| 2.3.1    | <i>Bacterial Cultures</i> .....  | 32        |
| 2.3.2    | <i>Vector Creation</i> .....   | 32        |
| 2.3.3    | <i>Chemical Transformation into TOP10 Cells</i> .....                                      | 35        |
| 2.3.4    | <i>Plasmid Purification</i> .....  | 36        |
| 2.3.5    | <i>Vector Screening and Sequencing</i> .....   | 37        |
| 2.3.6    | <i>Electro-Transformation of Vectors into ET12567/pUZ8002</i> .....                        | 38        |
| 2.3.7    | <i>Intergenic Conjugation to Introduce Vectors into Streptomyces coelicolor M145</i>       | 39        |
| 2.3.8    | <i>Measurement of Bioluminescence</i> .....  | 40        |
| 2.3.9    | <i>Site Directed Mutagenesis</i> .....   | 41        |
| 2.3.10   | <i>Bacterial Cultures for Protein Purification</i> .....                                   | 41        |
| 2.3.11   | <i>Collection and Purification of Secreted Proteins from Culture Media</i> .....           | 42        |
| 2.3.12   | <i>SDS-PAGE</i> .....  | 44        |
| 2.3.13   | <i>SDS-PAGE Gel Extraction Protocol for Mass Spectrometry Analysis</i> .....               | 44        |
| <b>3</b> | <b>DEVELOPMENT OF LUCIFERASE REPORTER GENE CONSTRUCTS</b> .....                            | <b>46</b> |
| 3.1      | Aims and Strategy of Investigation.....  | 46        |
| 3.1.1    | <i>Reporter Genes</i> .....  | 46        |
| 3.1.2    | <i>Regulatory System of Interest – The Methylenomycin Biosynthetic Gene Cluster</i>        | 47        |
| 3.2      | Plasmid Design and Assembly.....   | 49        |
| 3.2.1    | <i>Luciferase Vectors</i> .....  | 49        |
| 3.2.2    | <i>Vectors Containing mmfR or mmyR</i> .....   | 53        |
| 3.3      | Initial Investigation of the Reporter System.....  | 54        |
| 3.3.1    | <i>Details on the Bioluminescence Assay</i> .....  | 54        |
| 3.3.2    | <i>Preliminary Observations</i> .....  | 55        |
| 3.3.3    | <i>Timing of Measurements</i> .....  | 56        |
| 3.3.4    | <i>Investigating How and When to Add the MMFs</i> .....                                    | 57        |
| 3.3.5    | <i>Release of MmfR by Different Small Molecules</i> .....                                  | 59        |
| 3.3.6    | <i>Optimal Concentration of Inducers</i> .....   | 61        |
| 3.3.7    | <i>Summary of Strategy Chosen to Investigate MmfR/MMF/MARE Operator Interactions</i> ..... | 62        |
| 3.4      | Promoter Strength .....  | 63        |

|          |   |           |
|----------|---|-----------|
| 3.4.1    | <i>Strategy for Investigating Promoter Strength</i> .....   | 63        |
| 3.4.2    | <i>Results on Comparison of the Five Methylenomycin Cluster Promoter Strengths</i><br>64                  |           |
| 3.4.3    | <i>Comparison with Other Promoters</i> .....  | 66        |
| 3.5      | Conclusions Drawn from Preliminary Investigation and Outlook for Future<br>Investigations .....           | 68        |
| <b>4</b> | <b>MODE OF ACTION OF TRANSCRIPTIONAL REPRESSOR MMFR IN</b><br><b><i>STREPTOMYCES COELICOLOR</i></b> ..... | <b>70</b> |
| 4.1      | Aims and Strategy of Investigation .....  | 70        |
| 4.2      | Statistical Explanation of Data Handling.....   | 70        |
| 4.3      | MmfR-MARE Operator Interactions.....  | 71        |
| 4.3.1    | <i>Details on the MARE Operator Sequences</i> .....   | 71        |
| 4.3.2    | <i>Strength of MmfR Binding to the MARE operators</i> .....   | 72        |
| 4.3.3    | <i>Investigation into Release of MmfR from the MARE Operator by MMF4</i> .....                            | 74        |
| 4.3.4    | <i>Discussion of MmfR/MARE Operator Binding Data</i> .....  | 76        |
| 4.4      | MmfR-MMF Interactions .....   | 77        |
| 4.4.1    | <i>Different MMFs</i> .....   | 77        |
| 4.4.2    | <i>Different Concentrations of MMFs</i> .....   | 80        |
| 4.4.3    | <i>In Silico Analysis of MmfR Ligand Binding Pocket</i> .....   | 86        |
| 4.4.4    | <i>Effect of Site Directed Mutagenesis on the Ligand Binding Domain of MmfR</i> ...                       | 90        |
| 4.4.5    | <i>Discussion of MmfR/MMF Binding Data</i> .....  | 95        |
| 4.5      | Outlook for Further Investigations .....  | 97        |
| <b>5</b> | <b>MODE OF ACTION OF TRANSCRIPTIONAL REPRESSOR MMYR IN</b><br><b><i>STREPTOMYCES COELICOLOR</i></b> ..... | <b>98</b> |
| 5.1      | Comparison of MmyR with MmfR and Homologous Regulatory Pathways.....                                      | 98        |
| 5.2      | Aims and Strategy of Investigation .....  | 99        |
| 5.3      | MmyR-MARE Operator Interactions .....   | 100       |
| 5.3.1    | <i>Strength of MmyR Binding to the MARE Operator</i> .....  | 100       |
| 5.3.2    | <i>Investigation into Release of MmyR from MARE operators by MMF4</i> .....                               | 103       |
| 5.3.3    | <i>Discussion of MmfR/MARE Operator Binding Data</i> .....  | 105       |
| 5.4      | MmyR-MMF Interactions.....  | 106       |
| 5.4.1    | <i>Effect of MMFs on MmyR Binding at the L1 MARE Operator</i> .....                                       | 106       |
| 5.4.2    | <i>Effect of Other Potential Ligands on MmyR Binding at the L1 MARE Operator</i><br>109                   |           |
| 5.4.3    | <i>Further Investigation - Effect of MMFs on MmyR Binding at the L3 MARE</i><br><i>Operator</i> .....     | 113       |

|          |  |            |
|----------|--|------------|
| 5.4.4    | <i>Discussion of MmyR-Ligand Interactions</i> .....  | 117        |
| 5.5      | Outlook for Further Investigations.....  | 118        |
| <b>6</b> | <b>INVESTIGATION OF MMFR/MMYR SELF-REGULATION</b> .....  | <b>119</b> |
| 6.1      | Background and Strategy of Investigation.....  | 119        |
| 6.2      | Preliminary Observations.....  | 120        |
| 6.3      | Effect of the MMFs on MmfR When Under MARE Operator Regulation .....                                 | 122        |
| 6.4      | Effect of the MMFs on MmyR When Under MARE Operator Regulation .....                                 | 125        |
| 6.5      | Mathematical Modelling of the MmfR/MMF/MARE Operator Regulatory Network<br>126                       |            |
| 6.6      | Proposed Function of MmfR and MmyR In Regulating Methylenomycin<br>Biosynthesis.....                 | 129        |
| 6.7      | Conclusions and Summary .....  | 133        |
| <b>7</b> | <b>DEVELOPMENT OF A NOVEL INDUCIBLE EXPRESSION SYSTEM FOR<br/>STREPTOMYCETES</b> .....               | <b>135</b> |
| 7.1      | Aims at Strategy of Investigation .....  | 135        |
| 7.1.1    | <i>Existing Heterologous Expression Systems</i> .....  | 135        |
| 7.1.2    | <i>MmfR/MMF/MARE Operator as an Inducible Expression System in Gram-<br/>Positive Bacteria</i> ..... | 135        |
| 7.2      | <i>Streptomyces albus</i> as a Potential Host.....   | 136        |
| 7.2.1    | <i>Introduction to S. albus</i> .....  | 136        |
| 7.2.2    | <i>Luminescence Assay in S. albus</i> .....  | 136        |
| 7.2.3    | <i>Investigating potential GBL-related TetRs in S. albus</i> .....                                   | 139        |
| 7.2.4    | <i>Implications of Findings on S. albus</i> .....  | 146        |
| 7.3      | Development of Vectors for Inducible Expression.....   | 147        |
| 7.3.1    | <i>Strategy of Inducible Expression System Mechanism</i> .....                                       | 147        |
| 7.3.2    | <i>Selecting a Protein for Over Production</i> .....   | 149        |
| 7.3.3    | <i>Design of the Expression Vectors</i> .....  | 150        |
| 7.4      | Optimisation of a Novel Inducible Expression System .....  | 152        |
| 7.4.1    | <i>Strategy of Protocol Optimisation</i> .....   | 152        |
| 7.4.2    | <i>Selecting Time points for Protein Purification</i> .....  | 153        |
| 7.4.3    | <i>Considerations on Further Optimisation of the Protocol</i> .....                                  | 155        |
| 7.4.4    | <i>Work on Solid Culture and Addition of MmfR and the MMFs</i> .....                                 | 156        |
| 7.4.5    | <i>Optimisation with Protease Inhibitors</i> .....   | 160        |
| 7.5      | Discussion and Conclusions on the Work to Develop a Novel Inducible Expression<br>System .....       | 166        |
| <b>8</b> | <b>GENERAL DISCUSSION</b> .....  | <b>169</b> |

|          |   |            |
|----------|---|------------|
| 8.1      | Research Questions and Summary of Findings .....  | 169        |
| 8.2      | Answers to the Research Questions.....  | 170        |
| 8.2.1    | <i>Do all five methylenomycin cluster promoters, controlled by MmfR, have the same strength?</i> .....  | 170        |
| 8.2.2    | <i>Does MmfR bind in the same way to all three MARE operator sequences?</i> .....   | 171        |
| 8.2.3    | <i>Is MmfR release by the MMFs the same at all three MARE operators?</i> .....  | 172        |
| 8.2.4    | <i>Does MmfR respond to all five furan compounds?</i> .....   | 172        |
| 8.2.5    | <i>Do all five MMFs have the same efficacy?</i> .....   | 172        |
| 8.2.6    | <i>What are the key residues in ligand binding?</i> .....   | 173        |
| 8.2.7    | <i>How does MmyR binding to the MARE operators and the MMFs vary from that of MmfR?</i> .....   | 174        |
| 8.2.8    | <i>Are there any other ligands that MmyR may bind to?</i> .....   | 175        |
| 8.2.9    | <i>Could MmfR, MMFs and MARE operators be used as a multi-host efficient novel inducible expression system for GC rich bacteria? Would this allow the purification of recombinant proteins?</i> ..... | 176        |
| 8.3      | Impact of Data Collected .....  | 177        |
| 8.4      | Discussion and Improvements of Luciferase Assay .....   | 178        |
| 8.4.1    | <i>Discussion of Techniques</i> .....   | 178        |
| 8.4.2    | <i>Luciferase Assay Future Work</i> .....   | 181        |
| 8.5      | Discussion and Improvements of the Novel Inducible Expression System .....  | 183        |
| 8.5.1    | <i>Discussion of Techniques</i> .....   | 183        |
| 8.5.2    | <i>Novel Inducible Expression System Future Work</i> .....  | 185        |
| 8.6      | Other Relevant Future Work.....   | 188        |
| <b>9</b> | <b>SUMMARY OF RESULTS AND CONCLUSIONS .....</b>   | <b>189</b> |
|          | <b>APPENDICES .....</b>   | <b>190</b> |
|          | Appendix A. Nucleotide Sequences.....   | 190        |
|          | Appendix B. Amino Acid Sequences.....   | 191        |
|          | Appendix C. Sequence Analysis of MmfR and Homologues.....   | 192        |
|          | <b>REFERENCES .....</b>   | <b>193</b> |



## List of Figures

|  |    |
|--|----|
| Figure 1.1. Schematic representation of the streptomycete life cycle.....  | 1  |
| Figure 1.2. Key dates when antibiotics were discovered .....   | 2  |
| Figure 1.3. Chemical structures of microbial hormones from <i>Streptomyces</i> species.....  | 8  |
| Figure 1.4. Chemical structures of the antibiotic methylenomycin A and its precursor,<br>methylenomycin C from <i>S. coelicolor</i> A3(2) and <i>S. violaceoruber</i> .....  | 9  |
| Figure 1.5. Organisation of the methylenomycin biosynthetic gene cluster from the SCP1<br>plasmid of <i>S. coelicolor</i> A3(2).....   | 10 |
| Figure 1.6. Representatives from the crystal structure of the TetR family homodimer MmfR<br>from the methylenomycin biosynthetic cluster of <i>S. coelicolor</i> in its apo form.....  | 11 |
| Figure 1.7. Chemical structures of the known <i>S. coelicolor</i> methylenomycin furans .....  | 12 |
| Figure 1.8. Proposed regulation of the methylenomycin gene cluster from the SCP1 plasmid<br>of <i>S. coelicolor</i> by the TetR family member; MmfR.....   | 13 |
| Figure 1.9. Organisation of <i>mmfR/mmflHP/mmyR</i> -like clusters found in <i>S. venezuelae</i> and <i>S.</i><br><i>avermitilis</i> and their associated amino acid percentage identities compared to the <i>S.</i><br><i>coelicolor</i> cluster..... | 16 |
| Figure 1.10. Predicted motif of the intergenic auto-regulatory response element (ARE)<br>sequences from <i>S. coelicolor</i> , <i>S. venezuelae</i> , <i>S. avermitilis</i> and <i>S. hygroscopicus</i> .....  | 17 |
| Figure 1.11. Arrangement of open reading frames in the <i>luxCDABE</i> reporter system and the<br>assigned functions of individual components of this operon .....   | 22 |
| Figure 3.1. Schematic of reporter gene systems and how they can be used to study<br>transcriptional repressors and their cognate inducer ligands .....   | 46 |
| Figure 3.2. Proposed regulation of the methylenomycin biosynthetic gene cluster from the<br>SCP1 plasmid of <i>S. coelicolor</i> by the TetR family member, MmfR and cognate MMF<br>ligands.....   | 47 |
| Figure 3.3. Schematic of <i>lux</i> reporter system in <i>Streptomyces coelicolor</i> M145.....  | 49 |
| Figure 3.4. L1, an integrating <i>luxCDABE</i> reporter plasmid.....   | 50 |

|  |    |
|--|----|
| Figure 3.5. Assigned nomenclature for <i>lux</i> vectors containing different intergenic regions from the methylenomycin gene cluster, cloned upstream of <i>luxCDABE</i> .....  | 51 |
| Figure 3.6A and B. Details on the methylenomycin cluster MARE operators .....  | 52 |
| Figure 3.7. pKMS01, an integrating plasmid with <i>mmfR</i> controlled by <i>ermEp*</i> (72) .....   | 53 |
| Figure 3.8A-C. Details on and visual representations of strains containing the <i>lux</i> genes under the control of <i>mmfLp</i> .....  | 56 |
| Figure 3.9. Changes in levels of luminescence over 16 days where <i>luxCDABE</i> is under the control of <i>mmfLp</i> , compared to a negative control .....   | 57 |
| Figure 3.10. Changes in luminescence produced by the <i>lux</i> operon under the control of <i>mmfLp</i> and MmfR by adding 400 nM MMF4 at different time points .....   | 58 |
| Figure 3.11. Chemical structures of MMF2, 4 and 5 and SCB1 from <i>S. coelicolor</i> and synthetic molecule 70 .....   | 59 |
| Figure 3.12. Changes in luminescence produced by the <i>lux</i> operon under the control of <i>mmfLp</i> and MmfR in the presence 100 $\mu$ M of different microbial hormones .....  | 60 |
| Figure 3.13A and B. Luminescence produced by the <i>lux</i> operon under the control of <i>mmfLp</i> and MmfR upon the addition of different concentrations of MMF2 and MMF5 .....   | 61 |
| Figure 3.14. Schematic of luciferase assay used to investigate the strength of promoters from the methylenomycin biosynthetic cluster .....  | 63 |
| Figure 3.15. Luminescence produced by the <i>lux</i> operon when under the control of different promoters from the methylenomycin gene cluster .....   | 64 |
| Figure 3.16. Boxplot of luminescence produced by the <i>lux</i> operon when under the control of different promoters, at 72 hours growth .....   | 65 |
| Figure 4.1. Schematic showing nomenclature of <i>lux</i> vectors, based on the MARE operator being studied for a particular methylenomycin cluster intergenic regions .....  | 72 |
| Figure 4.2. Luminescence produced by the <i>lux</i> operon under the control of different methylenomycin cluster operators in the presence and absence of MmfR at 48 hours growth compared to a <i>S. coelicolor</i> M145 negative control ..... | 73 |

|   |    |
|---|----|
| Figure 4.3. Boxplot of luminescence produced by the <i>lux</i> operon under the control of different MARE operators and MmfR in the presence and absence of 100 $\mu$ M MMF4, compared to luminescence produced by <i>luxCDABE</i> under the control of the same operator but no MmfR at 48 hours ..... | 75 |
| Figure 4.4. Chemical structures of the methylenomycin furans (MMFs).....  | 77 |
| Figure 4.5. Luminescence produced by the <i>lux</i> operon when under the control of <i>mmfLp</i> and MmfR upon the addition of 100 $\mu$ M of different MMFs compared to no MMFs over time .....   | 78 |
| Figure 4.6. Bar chart of luminescence produced by the <i>lux</i> operon when under the control of <i>mmfLp</i> and MmfR in the presence of 100 $\mu$ M of different MMFs compared to no MMFs at 48 hours.....   | 79 |
| Figure 4.7. Dose-response of MMF4 over time, in strains where <i>lux</i> operon expression is controlled by <i>mmfLp</i> and MmfR.....  | 81 |
| Figure 4.8. Dose-response curve of MMF4 at the 48 hour time point, in strains where <i>lux</i> operon expression is controlled by <i>mmfLp</i> and MmfR.....  | 82 |
| Figure 4.9. Dose response curves for MMF1, 2, 3 and 5 at 48 hours growth using strains where <i>lux</i> operon expression is controlled by <i>mmfLp</i> and MmfR .....  | 83 |
| Figure 4.10. Compiled dose-response standard curves showing the relationship between the concentration of all five MMFs and the luminescence produced in strains where <i>lux</i> operon expression is regulated by <i>mmfLp</i> and MmfR, at 48 hours growth .....                                     | 84 |
| Figure 4.11. LigPlot <sup>+</sup> schematic of MmfR bound to MMF2, as shown by X-ray crystallography data .....   | 86 |
| Figure 4.12. Chemical structures of the MMFs, SCB1 and other synthetic analogues.....   | 87 |
| Figure 4.13. Positions of ligand binding residues in the MmfR amino acid sequence .....   | 89 |
| Figure 4.14. Comparison of MmfR ligand binding amino acid residues with the sequences of orthologous TFRs SgnR and SAV_2270.....  | 89 |
| Figure 4.15. Schematic highlighting the hydroxyl and amine groups of tyrosines 85 and 144 from MmfR involved in MMF binding.....  | 91 |

|  |     |
|--|-----|
| Figure 4.16. Luminescence produced by the <i>lux</i> operon when under the control of <i>mmfLp</i> and MmfR tyrosine mutants compared to wild type MmfR, in the presence and absence of MMF4 .....   | 92  |
| Figure 4.17. Boxplot of luminescence produced by the <i>lux</i> operon when under the control of <i>mmfLp</i> and MmfR tyrosine mutants, compared to wild type MmfR measured in the presence of MMF4 at 48 hours .....   | 92  |
| Figure 4.18. Standard curve for production of luminescence as a result of MMF4 binding to and releasing the wild type MmfR compared to tyrosine 85 and 144 mutants.....  | 94  |
| Figure 4.19. Bar chart showing a comparison of the binding potentials of MMF1-5 for WT MmfR, as shown by the levels of <i>luxCDABE</i> expression when under the control of <i>mmfLp</i> and WT MmfR, compared to the binding potential for MMF4 with the Y85F or Y144F MmfR mutants ..... | 96  |
| Figure 5.1. Comparison of MmfR ligand binding amino acid residues with the sequences of homologues MmyR, SgnR and SAV_2270 .....   | 99  |
| Figure 5.2. Bar chart of luminescence produced by <i>S. coelicolor</i> M145 strains at 48 hours, where <i>luxCDABE</i> expression is controlled by different MARE operators, in the presence or absence of MmfR or MmyR.....   | 101 |
| Figure 5.3. Bar chart of luminescence produced by the <i>lux</i> operon when under the control of different MARE operators and MmyR, in the presence and absence of 100 $\mu$ M MMF4 at 48 hours.....  | 104 |
| Figure 5.4. Luminescence produced by the <i>lux</i> operon when controlled by <i>mmfLp</i> and MmyR in the presence of 100 $\mu$ M MMF1-5 compared to no MMFs .....  | 107 |
| Figure 5.5. Effect on luminescence produced by the <i>lux</i> operon, when controlled by <i>mmfLp</i> and MmyR, by the addition of 100 $\mu$ M MMF1-5 at 48 hours growth.....  | 108 |
| Figure 5.6. Chemical structures of methylenomycin furans, methylenomycin A and its precursor; methylenomycin C.....  | 109 |

Figure 5.7. CCD camera images of luminescence produced by a lawn of *S. coelicolor* M145 containing the *lux* operon under the control of *mmfLp* and MmyR in the presence of methylenomycin A and C compared to a DMSO control..... 110

Figure 5.8. Luminescence produced by *luxCDABE*, under the control of *mmfLp* and MmyR, in the presence of different concentrations of methylenomycin A and C..... 111

Figure 5.9. Bar chart of the effect on luminescence produced by *luxCDABE*, under the control of *mmfLp* and MmyR, by the presence of different concentrations of methylenomycin A and C at 48 hours growth ..... 112

Figure 5.10. Luminescence produced by the *lux* operon under the control of *mmyBp* and MmyR in the absence and presence of 100  $\mu$ M MMF2, 4 or 5 compared to a positive control over 72 hours ..... 114

Figure 5.11. Bar chart of luminescence produced by the *lux* operon under the control of *mmyBp* and MmyR in the absence and presence of 100  $\mu$ M MMF2, 4 or 5 compared to a positive control at 48 hours ..... 114

Figure 6.1. Visual representation of the intergenic region located between *mmfR* and *mmfL*, as found in the 11NY and sp105 vectors..... 120

Figure 6.2. Details on and visual representations of strains containing the *lux* genes under the control of *mmfLp*, with or without the repressor proteins MmfR or MmyR under the control of *ermEp\** or *mmfRp*..... 121

Figure 6.3. Relative level of luminescence produced by the *lux* operon under the control of *mmfLp* and MmfR (under its own auto-regulatory control) in the presence and absence of 400 nM MMF4 over time..... 123

Figure 6.4. Schematic of proposed assay to measure the effects of a gradient of MMFs on the repression of the *lux* operon when *mmfR* expression is under auto-regulatory control .. 124

Figure 6.5. Model architecture of the methylenomycin biosynthetic cluster ..... 128

Figure 6.6. Schematic of proposed MmfR/MmyR regulation of the methylenomycin biosynthetic cluster ..... 132

|   |     |
|---|-----|
| Figure 7.1. A comparison of luminescence produced by the <i>luxCDABE</i> operon in <i>S. albus</i> , as regulated by the presence and absence MmfR, MmyR and MMF4 over 72 hours .....   | 137 |
| Figure 7.2. Bar chart comparing luminescence produced by the <i>luxCDABE</i> operon in <i>S. albus</i> , as regulated by the presence and absence MmfR, MmyR and MMF4 at 48 hours .....   | 138 |
| Figure 7.3. Amino acid sequence comparison of MmfR and orthologue SSHG_01258 from <i>S. albus</i> .....   | 143 |
| Figure 7.4. Phylogenetic tree (144) showing the relationship between MmfR, SSHG_01258 and other TetR family homologues across their amino acid sequences .....  | 144 |
| Figure 7.5. The genetic basis of the novel inducible expression system designed to be regulated by MmfR/MMF/MARE operator in the <i>S. coelicolor</i> M145 chromosome ....  | 148 |
| Figure 7.6. pKMS05, an expression vector containing the <i>S. mutans gbnB</i> orthologue and <i>luxCDABE</i> , both under the control of <i>mmfLp</i> .....   | 150 |
| Figure 7.7. Layout of synthetic <i>gbnB</i> insert for vectors pKMS05 and pKMS06, for use in the novel inducible expression system.....   | 151 |
| Figure 7.8A-D. Secreted proteins produced at different time points by the M145:L1 control strain compared to the strain designed to produce <i>S. mutans</i> GbnB (M145:pKMS05)..   | 154 |
| Figure 7.9A-C. Secreted proteins produced by the M145:L1 control strain compared to the strains designed to produce <i>S. mutans</i> or <i>S. enterica</i> GbnB (M145:pKMS05 and 06 respectively) in the presence and absence of MmfR and MMF5 .....  | 157 |
| Figure 7.10A-C. The effects of protease inhibitors on secreted proteins produced by the M145:L1 control strain compared to the strains designed to produce <i>S. mutans</i> or <i>S. enterica</i> GbnB (M145:pKMS05 and 06 respectively) in the presence and absence of MmfR and MMF1 ..... | 162 |
| Figure 7.11A and B. Intercellular proteins produced by the M145:L1 control strain compared to strains designed to produce <i>S. mutans</i> or <i>S. enterica</i> GbnB (M145:pKMS05 and 06 respectively) in the presence and absence or MmfR and MMF1 .....                                  | 165 |

## List of Tables

|   |    |
|---|----|
| Table 1.1. Properties of TetR family member monomers MmfR, MmyR, ScbR and ArpA ...  | 15 |
| Table 2.1. Consumables used .....   | 25 |
| Table 2.2. Equipment used .....   | 26 |
| Table 2.3. Software used .....  | 26 |
| Table 2.4. Primers used .....   | 27 |
| Table 2.5. Vectors used in luciferase reporter gene assay.....  | 28 |
| Table 2.6. Other vectors used in this research project.....   | 28 |
| Table 2.7. Parent strains used in investigation .....   | 29 |
| Table 2.8. Strains created using various vectors to be used in the luminescence assay.....  | 29 |
| Table 2.9. PCR cycle times .....  | 33 |
| Table 2.10. Restriction sites used to create vectors for this investigation.....  | 34 |
| Table 2.11. PCR cycle times .....   | 38 |
| Table 3.1. Comparison of the different methylenomycin cluster promoter sequences with strong constitutive promoter, <i>ermEp</i> * .....  | 67 |
| Table 4.1. Summary of MmfR-MARE operator binding data .....   | 76 |
| Table 4.2. A <i>t</i> -test analysis of significant changes in the luminescence produced by <i>luxCDABE</i> when under the control of <i>mmfLp</i> and MmfR in the presence of different MMFs compared to no MMFs .....   | 79 |
| Table 4.3. Binding kinetics data for MmfR with each of the five methylenomycin furans including the $B_{max}$ , $K_d$ and binding potential values for each at 48 hours growth .....  | 85 |
| Table 4.4. Summary of docking analysis of MmfR with the MMFs and other synthetic and natural analogues.....   | 87 |
| Table 4.5. A <i>t</i> -test analysis luminescence produced by <i>luxCDABE</i> when under the control of <i>mmfLp</i> and the Y85F and Y144F MmfR mutants in the presence of 100 $\mu$ M or 200 $\mu$ M MMF4 to look for significant differences in the removal of repression..... | 93 |

|  |     |
|--|-----|
| Table 4.6. Binding kinetics data for MMF4 binding wild type MmfR compared to the tyrosine 85 and 144 MmfR mutants, including the $B_{max}$ , $K_d$ and binding potential values for each at 48 hours growth .....                      | 95  |
| Table 5.1. Statistical analysis of luminescence produced by <i>S. coelicolor</i> M145 strains at 48 hours, where <i>luxCDABE</i> expression is controlled by different MARE operators, in the presence or absence of MmfR or MmyR..... | 101 |
| Table 5.2. Sequences of DNA binding regions from MmfR, MmyR and their homologues   | 103 |
| Table 5.3 Statistical analysis of luminescence produced by the <i>lux</i> operon when under the control of different MARE operators and MmyR, in the presence and absence of 100 $\mu$ M MMF4 at 48 hours .....                        | 104 |
| Table 5.4. Summary of binding data for MmyR and the MARE operators.....  | 105 |
| Table 5.5. A <i>t</i> -test analysis of the effects on luminescence produced by the <i>lux</i> operon, when controlled by <i>mmfLp</i> and MmyR, by the addition of 100 $\mu$ M MMF1-5 at 48 hours growth .....                        | 108 |
| Table 5.6. A <i>t</i> -test analysis of the effect on luminescence produced by <i>luxCDABE</i> , under the control of <i>mmfLp</i> and MmyR, by the presence of different small molecules at 48 hours growth .....                     | 112 |
| Table 5.7. A <i>t</i> -test analysis of luminescence produced by the <i>lux</i> operon under the control of <i>mmyBp</i> and MmyR in the presence of 100 $\mu$ M MMF2, 4 or 5 at 48 hours .....  | 115 |
| Table 6.1. Phenotypes of methylenomycin cluster mutants.....   | 127 |
| Table 7.1. A <i>t</i> -test analysis of the luciferase assay results collected from <i>S. albus</i> data at 48 hours.....  | 138 |
| Table 7.2. DNA binding region sequence of SSHG_01258 compared to that of MmfR and MmyR and their orthologues .....   | 140 |
| Table 7.3. Percentage identity of amino acid sequence between MmfR and MmyR homologues from <i>S. venezuelae</i> , <i>S. avermitilis</i> and <i>S. albus</i> .....   | 141 |
| Table 7.4 Percentage coverage of analysis of amino acid sequence between MmfR and MmyR and their homologues from <i>S. venezuelae</i> , <i>S. avermitilis</i> and <i>S. albus</i> .....  | 141 |



|   |     |
|---|-----|
| Table 7.5. Total score of alignment, according to NCBI BLAST, for the amino acid sequences for MmfR and MmyR and their homologues from <i>S. venezuelae</i> , <i>S. avermitilis</i> and <i>S. albus</i> ..... | 142 |
| Table 7.6. Summary of homology between TetR family members across different bacterial species .....   | 144 |
| Table 7.7. Comparison of 20 amino acid conserved DNA binding motifs in TetR family repressors with the motif from MmfR and MmyR .....   | 145 |
| Table 7.8. Details on <i>S. venezuelae</i> GbnB and its analogues .....   | 149 |
| Table 7.9. Details on synthetic GbnB orthologues predicted to be produced by the new constructs; pKMS05 and pKMS06 .....  | 151 |
| Table 8.1. Summary of data on MmfR and MmyR binding to the MARE operators .....   | 174 |
| Table 8.2. Summary of the positive and negative aspects of the <i>luxCDABE</i> reporter gene assay .....  | 179 |

## Acknowledgements

With thanks to;

My supervisors; Christophe Corre and Vilmos Fulop

Particularly Christophe for his endless patience with the many many questions I have for him

Vincent Poon, John Sidda and the rest of the Corre Group for their assistance and advice in the lab over the years

Justin Nodwell and Leslie Cuthbertson for providing the luciferase vectors

Sophie Piquerez for the use of the Photek light detector in her lab

Shanshan Zhou for her work with the MmfR mutants and complimentary gel shift assays as well as the synthesis of the MMFs

Phil Young for his advice on statistics

Cleidiame Zampronio and the proteomics facility at Warwick for their help processing protein samples for LC-MS

Jack Bowyer for his work on the mathematical modelling of the methylenomycin regulatory system

Rachael Wilkinson for her advice on protein purification

The MIBTP cohort for being a lovely bunch

My advisory panel; Richard Napier and Alexander Cameron for all their helpful advice throughout my PhD

BBSRC for funding the project

Also with very special thanks to my parents Mary and Paul and sister Bethany, Jeff Brown, Becky Johnson, Federico Botta and many others for keeping me sane during my PhD

## Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for a degree.

The research presented was carried out by the author except in the cases outlined below:

The production and analysis of MmfR mutants in Chapter 4 was carried out in collaboration with Shanshan Zhou from the Department of Chemistry.

The model architecture of MmfR/MmyR regulation found in Chapter 6 was produced by Jack Bowyer from the Department of Engineering and the analysis of this data was produced via a collaboration with this student and others.

In Chapter 7, the contribution of the WPH Proteomics RTP, Gibbet Hill Road, University of Warwick, UK should be acknowledged for the processing of the proteins purified from SDS-PAGE and help with the results from the LC-MS analysis.

No parts of this thesis have yet been published by the author, although there are a number of works currently being written.

## Abstract

An interesting insight has been developed into the roles of a paralogous pair of transcriptional regulators MmfR and MmyR in the regulation of methylenomycin antibiotic biosynthesis in *Streptomyces coelicolor*. Research involved the development and use of a luciferase reporter assay, optimised for use in GC high bacteria. MmfR belongs to the TetR-family of transcriptional repressor proteins and works as a single component system, binding to DNA at one of three methylenomycin auto-regulatory response element (MARE) operators. Here it represses transcription of five different operons until a conformational change is brought about by specific binding to one of five small signalling molecules; the methylenomycin furans (MMFs).

This investigation revealed that the five different MmfR-regulated operons have promoters of differing strengths, which is also contributed to by a variation in the strength of MmfR binding to the three MARE operator sites. Each of the five naturally produced MMF ligands were also shown to have a different efficacy for deactivating and displacing MmfR. An *in silico* analysis of the MmfR primary and tertiary structures, followed by *in vivo* mutagenesis, revealed the presence of two tyrosine residues implicated in ligand binding.

The paralogue MmyR was shown to vary in activity from that of MmfR. It showed weaker, but significant binding to only two out of the three MARE operator sites, binding with different affinities to each, and no significant removal of repression was seen in the presence of the MMF ligands.

The MmfR/MMF/MARE operator system shows promise as something that can be developed into a novel inducible expression system for use in GC high bacteria. However, whether this system can be adapted to be efficient in multiple hosts is yet to be seen, with affinity for the MARE operators from exogenous regulatory proteins predicted.

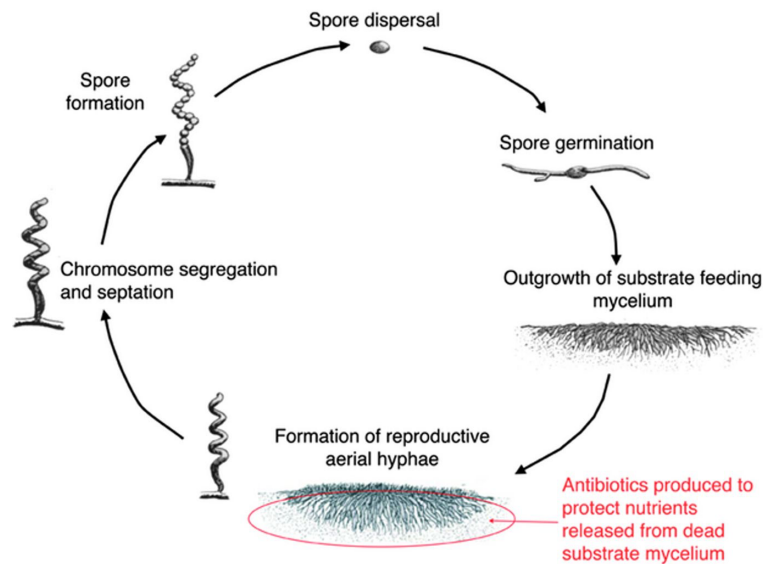
## Abbreviations

AHFCA<sub>s</sub> - 2-alkyl-4-hydroxymethylfuran-3-carboxylic acids  
AMR – antimicrobial resistance  
APS – ammonium persulphate  
BBSRC - Biotechnology And Biological Sciences Research Council  
BGC – biosynthetic gene cluster  
bp – base pairs  
CCD – charge coupled device  
CPS – counts per second  
Da – daltons  
DMSO – dimethyl sulfoxide  
dNTPs - deoxyribonucleotide triphosphates  
D-TOPO – directional TOPO  
DTP – doctoral training partnership  
*E. coli* - *Escherichia coli*  
EDTA - ethylenediaminetetraacetic acid  
FPLC – fast protein liquid chromatography  
GBL - gamma-butyrolactone  
HTH – helix-turn-helix  
ISBA – International Symposium on the Biology of Actinomycetes  
Kb – kilobase  
LB – lysogeny broth  
LBA – LB agar  
*lux* – luciferase genes *luxCDABE*  
MARE - methylenomycin auto-regulatory response element  
MIBTP – Midlands Integrative Biosciences Training Partnership  
MMF – methylenomycin furan  
Mm – methylenomycin  
MmA – methylenomycin A  
MmC – methylenomycin C  
MWCO – molecular weight cut off  
NCBI - National Center for Biotechnology Information  
NEB – New England Biolabs  
OD – optical density  
PCR – polymerase chain reaction  
QS – quorum sensing  
RBS – ribosome binding site  
RLU – relative light unit  
RPM – revolutions per minute  
SCBs – *Streptomyces coelicolor* butyrolactones  
SDM – site directed mutagenesis  
*S. coelicolor* – *Streptomyces coelicolor*  
SDS-PAGE – sodium dodecyl sulphate polyacrylamide gel electrophoresis  
SFM – soya flour mannitol  
SPR - surface plasmon resonance  
SOB – super optimal broth  
SOC – super optimal broth with a catabolite repressor  
TEMED - tetramethylethylenediamine  
TFR - TetR family regulator  
VB – Virginiae butanolide

# 1 Introduction

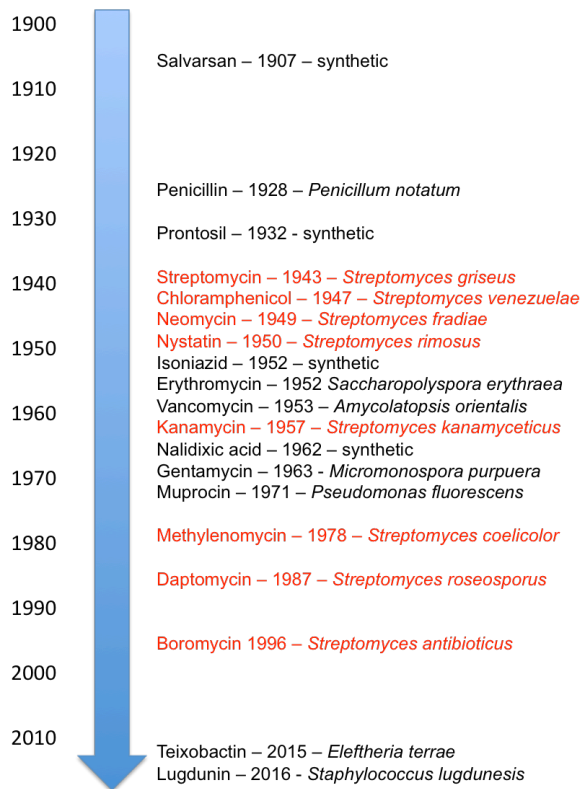
## 1.1 *Streptomyces* and their Natural Products

*Streptomyces* are Gram-positive actinomycetes found naturally in the soil. In fact, the earthy smell associated with soil comes from geosmin, a natural product produced by streptomycetes.(1) They breach the gap between bacteria and fungi with their complex mycelial life cycles (Figure 1.1). The life cycle starts with a spore, then, with the right nutrients present, this spore will germinate and form vegetative hyphae that branch into the surrounding growth media forming fungi-like mycelia. Upon nutrient depletion, non-branching sporogenic aerial hyphae will also form.(2) These aerial hyphae septate to form largely dormant unigenomic spores that can then start the cycle again. It is during this spore formation stage that grey spore pigments are synthesised, as well as a range of other natural products.(3) A natural product is any substance or chemical produced by a living organism and is often used synonymously with the term ‘secondary metabolite’.(4, 5) The natural products produced by streptomycetes total over 70 per cent of commercially available antibiotics,(6) a number of these can be seen highlighted in Figure 1.2; a timeline of antibiotic discovery and their bacterial origins. It is clear therefore that streptomycetes are of huge importance in natural product research.



**Figure 1.1. Schematic representation of the streptomycete life cycle**

*This image was taken directly from the paper by Seipke et al. from 2012 (7)*



**Figure 1.2. Key dates when antibiotics were discovered**

*Antibiotics produced by streptomycetes are highlighted in red.(8-14)*

The natural products made by *Streptomyces* include antifungals such as nystatin,(15) as well as anti-bacterials like chloramphenicol,(11) neomycin,(10) and streptomycin (9) (highlighted in Figure 1.2).(8) Streptomycetes also produce a range of other useful natural products such as the anti-parasitic ivermectin (16) as well as anti-tumour drugs, (17) immunosuppressive agents (18) and agrochemicals such as fungicides.(19) These natural products are often produced as a defence mechanism to fight other bacteria competing for resources. At sub-lethal levels however, antibiotics may be able to function as signalling molecules, benefiting otherwise susceptible bacteria and helping to maintain homeostasis in microbial communities.(20)

There is a need for new antibiotics, with resistance developing to new antibiotics within tens of years of clinical introduction, if not sooner.(21) Alexander Fleming famously warned of the inevitable risk of antibiotic resistance as early as 1945 in his Nobel Prize speech, with the widespread use of penicillin only coming about two years before and the first sulphonamide only having been commercially available since the 1930s.(22) Over-prescription and incorrect usage of antibiotics selects for resistant strains and has contributed to resistance being developed at an accelerated rate compared to that which would occur in the wild without this human interference. Interest into natural product antibiotic research declined due to the

frequent rediscovery of existing antibiotics and the development of a number of synthetic methods such as the screening of large libraries of synthetic compounds.(23-25) These synthetic methods proved to be largely unsuccessful and new strategies have again been developed for natural product discovery. The potential for discovering novel natural compounds is increasing again with the ‘genomic age’.(26)

### **Importance of Genome Mining in Natural Product Research**

Despite *Streptomyces* already being the main source of commercially available antibiotics, there is still potential for the discovery of many more, with prospective natural products lying undiscovered in currently ‘silent’ and cryptic gene clusters. These are being further investigated by genome mining.(6)

Entire genomes are scanned for sequences corresponding to hypothetical antibiotic regulatory or biosynthetic enzymes, based on sequence identity to clusters already studied experimentally. These ‘silent’ or cryptic gene clusters uncovered by genome mining often need very specific environmental conditions or inducer molecules for natural product biosynthesis to be switched on, often not present in standard laboratory conditions. For this reason, there are many hypothetical natural product gene clusters for which we know very little about the end product.(27) If the regulation of these clusters could be better understood then they could be genetically manipulated or put into a heterologous expression host (28) and the metabolites isolated. An example of a natural product that has been discovered by genome mining is coelichelin, a *Streptomyces coelicolor* iron chelator.(26, 29, 30)

As of 3<sup>rd</sup> August 2016, as many as 252 different streptomycetale strains had their complete genomes sequenced and listed on NCBI out of 8629 genomes available on this site.(31) Many more full genomic sequences are expected every year as genome sequencing becomes more economical.(27) The *Streptomyces* genus is therefore one of the most sequenced of non-pathogenic bacteria, but the true extent of the implications of such widespread analysis on the production of useful metabolites is not yet realised with many biosynthetic investigations still on-going.

## **1.2 Regulation of Natural Product Biosynthesis in *Streptomyces***

### **1.2.1 An Introduction to Microbial Hormones**

It is important for bacteria to regulate the production of antibiotics, not only to conserve cellular resources but also to prevent the antibiotic having a potentially toxic or lethal effect on the producer strain. A variety of different regulatory mechanisms are utilised to control the



production of any natural products. Streptomycete regulatory systems include transcriptional activators and repressors, the use of rare codons, sigma factors, riboswitches and receptor ligand responses as well as the use of microbial hormones.(32) Natural product biosynthesis may also be influenced by environmental factors such as pH and nutrient deprivation as well as being effected by cell density via quorum sensing (see below). In this report, it is the regulation by microbial hormones and their cognate receptors that is of particular interest.

A hormone is described as being a ‘regulatory substance produced in an organism and transported in tissue fluids such as blood or sap to stimulate specific cells or tissues into action.’(33) When applying this to the microbial level, a hormone could be described as a diffusible master switch for morphological differentiation and secondary metabolism.(34) The ability of bacteria to signal to one another using nanomolar concentrations of small molecules is well known.(35) This signalling is often influenced by a phenomenon called ‘quorum sensing’, where gene expression is regulated in response to population density and the accumulation of auto-inducing microbial hormones.(36) The higher the density of cells in an area, the greater the accumulation of signalling molecule that they produce. Once this signalling molecule reaches a threshold concentration, transcription of associated genes will be switched on. This is co-ordinated in the entire local population and has even been known to occur between species. In response to high population density therefore, streptomycetes will produce microbial hormones to signal for antibiotic production to kill competition for resources, as well as signalling to turn on the next stages in their complex sporogenic life cycle (Figure 1.1).

### 1.2.2 The Gamma-Butyrolactones

Examples of these microbial hormones in Gram-positive bacteria include the well-known gamma butyrolactones (GBLs). These are small signalling molecules involved in quorum sensing and the activation of antibiotic biosynthesis and cover a variety of similar small molecules (see Figure 1.3). GBL hormonal regulatory systems are found across a wide range of *Streptomyces* species and are formed of two parts; the enzymes for GBL synthesis as well as at least one cognate receptor.(37) Examples of these receptor-ligand systems includes A factor and its receptor ArpA in *Streptomyces griseus*,(34) SCB1 and ScbR in *S. coelicolor*,(38) VB and BarA in *Streptomyces virginiae* (39, 40) and IM-2 and FarA in *Streptomyces lavendulae* (41, 42) (see Figure 1.3 for chemical structures of these microbial hormones). In the absence of their ligand, ArpA, ScbR, BarA and FarA will normally be bound to DNA at an operator site, repressing the production of an associated natural product. The production and detection of a threshold level of their hormone ligand will cause a conformational change and the release of the repressor from a promoter site, thereby allowing the expression of natural product biosynthetic genes.

### 1.2.3 TetR Family Transcriptional Repressors

The cognate receptors for microbial hormones are often members of the TetR family of transcriptional repressors. This widespread family of transcriptional repressors includes hundreds of thousands of proteins found across a variety of genera of bacteria and archaea.(32, 43) TFRs (TetR family repressors) work as single component systems. Unlike two-component systems, such as kinase signalling pathways, the sensory and DNA binding components of these systems are located on the same polypeptide.(43) A single protein therefore receives the hormonal signal and transduces the message into a change in gene expression. These repressors are almost exclusively alpha helical and have an N-terminal DNA binding domain and C-terminal ligand binding domain. An N-terminal DNA binding domain has previously been associated with being a repressor whereas a C-terminal binding domain is often thought to relate to being an activator (note however that there are well documented exceptions to this rule).(43) TFR proteins are homodimeric meaning that for each dimer there are two identical ligand-binding pockets and two identical DNA-binding domains, usually binding to a palindromic operator sequence. A number of TFRs have had their 3D structure determined by X-ray crystallography including TetR,(44) QacR, CprB, SimR (45) and EthR (46) which has broadened our understanding of their functionality.

The common motif that connects the proteins in this family is the conserved helix-turn-helix DNA binding domain.(32) Outside of this sequence of 47 amino acids, there is no clear conservation in amino acid sequence however and the TetR family has been shown to bind to a large range of different ligands.

The TetR family of repressors are particularly common in microbes that have to adapt to changes in environment, for example soil dwelling bacteria such as *Streptomyces* or extremophiles and plant and animal pathogens. Regulation by this family of proteins is not limited to the biosynthesis of antibiotics with regulation of efflux pumps, responses to osmotic stress and the control of differentiation having been shown, amongst other targets.(32) The gene target for many TFRs is not known however. There are often repeated operator motifs, but these can be hard to predict.(43)

The complexity of regulation varies within the TetR family. An example of simple regulation is that of the repression of *tetA* by TetR, which is then released by tetracycline, thereby regulating tetracycline resistance in *Escherichia coli*.(47) However, in other systems it can be more complicated with modulation by other regulators and cross-interactions with other networks including the regulation of pathway specific activators and the repression of multiple bidirectional operators.(32) One consensus between TFRs appears to be that they are self-regulatory. For example, TetR controls the repression not only of *tetA* but also of *tetR*,

the gene for its own production. However, as with any class of proteins, there are exceptions to the rule and not all TFRs will function in the manner just described.

### **Pseudoreceptors**

GBL systems often have been found to contain two TetR family receptors, for example ScbR and ScbR2.(38, 48) ScbR2 appears to have a distinct role compared to the ScbR, the latter of which binds the *S. coelicolor* butyrolactone SCB1. This second type of protein is often thought of as a ‘pseudoreceptor’, sharing similar structures and sequence identity with the ‘real’ cognate GBL receptors but showing differences in ligand binding specificity as well as the range of DNA targets it will bind to.(49)

ScbR2 is the best studied of these pseudoreceptors.(50) ScbR and ScbR2, which share 33% identity over ~85% of their sequences, are involved in the regulation of antibiotics actinorhodin and undecylprodigiosin.(51) ScbR2 has been indicated to bind these and other end product antibiotics *in vitro* and *in vivo* as opposed to SCB1, thereby releasing its repression and influencing antibiotic production.(49, 52) ScbR has also been shown to be expressed at a different time to ScbR, further validating the evidence that it has a different role.(50)

The ligand binding domains of the GBL receptors and pseudoreceptors are not conserved. It is thought that a range of both endogenous and exogenous non-GBL ligands may bind these second repressors meaning that they may have a cross-species regulatory function.(52) It is also hypothesised that in some systems, the second receptor may only become ‘activated’ as a repressor in the presence of a cognate ligand, potentially the end product antibiotic.(50) In this way, this second repressor is thought switch off antibiotic production once enough has been produced. Alternatively, they may be able to directly or indirectly activate transcription of some genes as well as controlling the repression of others, although the mechanistics of this are not established.(50) The differences in ligand binding capacity between the true receptor and pseudoreceptor are poorly understood in many TetR systems and many inferences remain as hypotheses that are yet to be experimentally proven.

As well as differences in ligand binding, the DNA binding sequences for the pseudoreceptors have often been predicted to outnumber those for its paralogue and are also sometimes thought to be found in otherwise unrelated endogenous gene clusters, further indicating that these systems are cross-regulatory.(50)

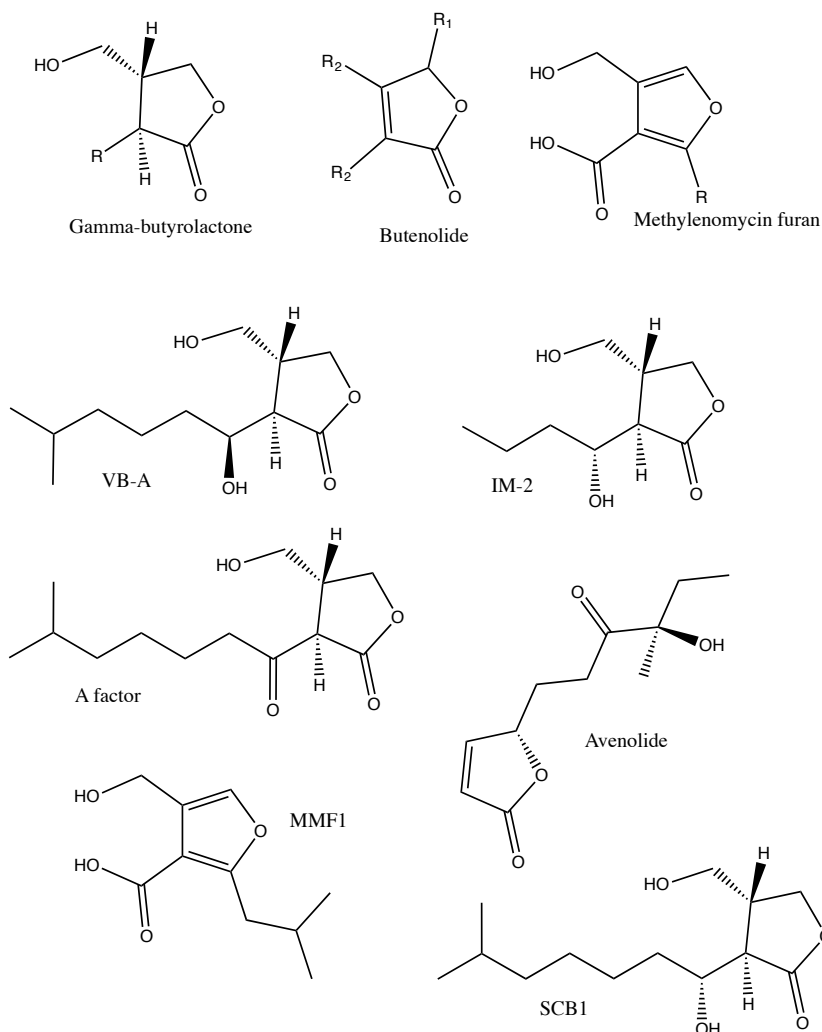
Having a second ‘pseudoreceptor’ adds an extra level of complexity and control over gene expression and understanding the role of these paralogous pseudoreceptors in the co-

ordination of antibiotic biosynthesis is key to exploiting ‘cryptic’ natural product gene clusters and increasing the yield of commercially available antibiotics.(49)

#### 1.2.4 Other Microbial Hormones

As well as these gamma-butyrolactones, other groups of microbial hormones include butenolides such as avenolide (53) and furans such as the methylenomycin furans,(54) which have also been shown to work with equivalent ligand-TetR family receptor/repressor systems. Examples of these hormones are shown in Figure 1.3.

The small methylenomycin furans (MMFs) from *Streptomyces coelicolor* A3(2) belong to a group of 2-alkyl-4-hydroxymethylfuran-3-carboxylic acids (AHFCAs) that are structurally distinct from the GBL family.(35, 54) Five known furan hormones are believed to alter the action of TetR family member MmfR (and possibly its paralogue MmyR) by inducing a conformational change and releasing the repression of the methylenomycin antibiotic biosynthetic cluster.(55) MmfR is an orthologue of the previously mentioned ArpA receptor which binds the GBL; A-factor. It is these methylenomycin furans (MMFs) from model actinomycete *Streptomyces coelicolor* A3(2) that is of particular interest in this project.



**Figure 1.3. Chemical structures of microbial hormones from *Streptomyces* species**  
*VB-A* – *S. virginiae* GBL, *IM-2* – *S. lavendulae* GBL, *A factor* – *S. griseus* GBL, *Avenolide* – *S. avermitilis* butenolide, *MMF1* – *S. coelicolor* furan, *SCB1* – *S. coelicolor* GBL.

## 1.3 *Streptomyces coelicolor* – A Model Organism

### 1.3.1 *S. coelicolor* Genetics

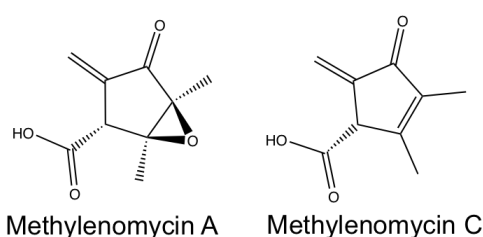
*S. coelicolor* (56) was initially chosen as a model organism as it produces red and blue pigments, a phenotype that is easy to observe and track in mutant strains.(57, 58) The name *coelicolor* comes from the Latin term *coelus* meaning ‘sky’ colour, referring to these blue pigments that this species produces.(59)

More than 10 years ago the entire *Streptomyces coelicolor* A3(2) genome was sequenced, allowing a greater understanding of the biosynthesis of crucial metabolites.(6) These soil living bacteria have a large 8 667 507 base pair single linear chromosome, the largest known bacterial genome at the time of discovery. The genome also includes two plasmids; the linear

SCP1 (356 023 bp) (60) and the circular SCP2 (31 317 bp), (61) which were also sequenced in 2004 and 2003 respectively. Of the large single chromosome, over 12% of protein coding genes are thought to be regulatory,(6) with roles in regulating morphological and metabolic changes as well as programming antibiotic synthesis via the use of microbial hormones.(32) *S. coelicolor* is known to produce at least five known antibiotics, including methylenomycin,(62) but the availability of the entire genome sequences has allowed the identification of a number of analogous pathways with unknown natural products, many of which are currently under further investigation.

### 1.3.2 Methylenomycin

Figure 1.4 shows the chemical structures of the methylenomycin antibiotic and its precursor from *Streptomyces coelicolor* A3(2). Methylenomycin A is a cyclopentanone antibiotic that acts upon both Gram-positive and Gram-negative bacteria, working optimally in low pH conditions.(63) It appears to be particularly active against the Gram-negative *Proteus*, also commonly found in soil as well as in faeces and manure.(64) *Streptomyces violaceoruber* is also known to produce methylenomycin, with 99% nucleotide identity shown in the biosynthetic cluster (65) but this strain has not been studied in this project.



**Figure 1.4. Chemical structures of the antibiotic methylenomycin A and its precursor, methylenomycin C from *S. coelicolor* A3(2) and *S. violaceoruber***

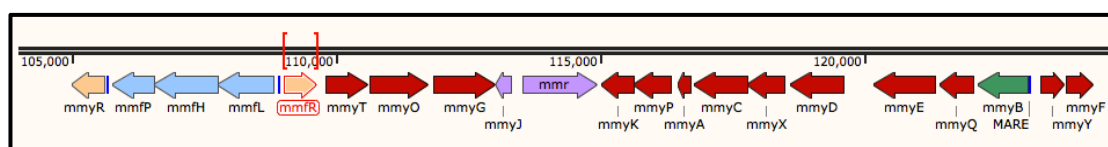
Despite being discovered in the 1970s (14) the exact mechanism of methylenomycin antibacterial action is as yet unclear and it has shown little promise clinically with high levels of toxicity to animal cells shown. In the past, this antibiotic was of particular interest as all 19 kb (54) of the 21 biosynthetic, regulatory and resistance genes are found in a cluster on the SCP1 plasmid of *S. coelicolor* (and pSV1 in *S. violaceoruber*) (66), (67), (68) (see Figure 1.5). Although extra chromosomal giant plasmids are often associated with secondary metabolism,(69) methylenomycin is one of very few known antibiotics where the whole biosynthetic cluster is entirely plasmid based. Being plasmid based meant that the genes were easily transmissible between streptomycetes and it was relatively simple to produce knockout strains.(70) This was particularly important at a time before technologies for creating knockouts had advanced to the level they have today. For studying methylenomycin

production, the SCP1-free *S. coelicolor* strain M145 could be used to study sections of the biosynthetic pathway without background interactions from the wild type system.

MmfR/MMF regulation of methylenomycin biosynthesis is analogous to a number of other antibiotic regulatory systems in other *Streptomyces* strains including *S. venezuelae* and *S. avermitilis* (see Section 1.4.3 on page 16). By further investigating this cluster that already has extensive research done on it and is partially understood, it might also be possible to shed some light on the regulation some of these other potentially harder to study biosynthetic clusters on which we have much less information.

### Methylenomycin Biosynthetic Cluster

The 21 genes of the methylenomycin cluster include *mmfR* and *mmyR*, coding for the TetR family transcriptional repressors MmfR and its paralogue MmyR, and the *mmfLHP* operon, which is responsible for making the enzymes used in the production of the cognate furan ligands for MmfR; the MMF microbial hormones. The cluster also contains *mmyJ* and *mmr*, both involved in methylenomycin resistance as well as *mmyB*, which produces a pathway specific activator. There is also a selection of methylenomycin biosynthetic genes which come together in operons; *mmyTOC*, *mmyBQEDXCAPK* and *mmmYF*.<sup>(71)</sup> The assembly of these genes on the SCP1 plasmid are shown below in Figure 1.5 with different colours denoting different types of product. Biosynthetic genes are all shown in red, repressors in peach, genes associated with resistance in purple, MMF biosynthetic genes in blue and the MmyB activator is in green.



**Figure 1.5. Organisation of the methylenomycin biosynthetic gene cluster from the SCP1 plasmid of *S. coelicolor* A3(2)<sup>1</sup>**

*Proposed functional attributions are mmfR and mmyR – transcriptional repressors, mmfLHP – methylenomycin furan biosynthetic genes, mmyJ and mmr – methylenomycin resistance, mmyB – transcriptional activator, all others – methylenomycin biosynthesis, based on the work of Chater and Bruton (65)*

## 1.4 Regulation of Methylenomycin Biosynthesis

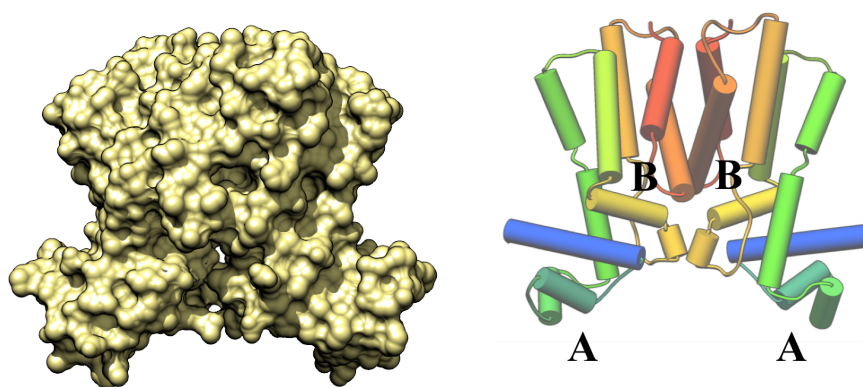
### 1.4.1 MmfR – A TetR Family Transcriptional Repressor

MmfR is a TetR family member and like them, is a homodimeric protein with a C-terminal DNA binding region and a N-terminal ligand-binding domain. As yet unpublished, the crystal

<sup>1</sup> This diagram is to scale and represents the appropriate sizes of the individual genes, it was created using the SnapGene Viewer software (72). Biotech G. SnapGene Viewer. In: Glick B, editor. 3.0.3 ed. Chicago, IL2004-2016.)

structure of MmfR has been determined, both with and without MMF2 to a resolution of 1.5Å. This reveals the molecular interactions between MmfR and MMF2 and opens up the potential for more bioinformatical analyses to be carried out on the protein (73) with a higher chance of accuracy for predictions of MmfR binding to the other four MMFs.

Figure 1.6 shows two representations of the 3D structure of this repressor protein in its apo form, based on the crystal structure obtained. The cartoon model on the right hand side shows the location of the ligand binding pocket as well as the TetR HTH DNA binding domain, labelled as B and A respectively. Each monomer in this homodimer contains one ligand binding pocket and a DNA binding domain meaning that each complete dimer will bind two MMF molecules. Work is currently being carried out to achieve a crystal structure of MmfR bound to DNA to complete the picture of its different conformations.



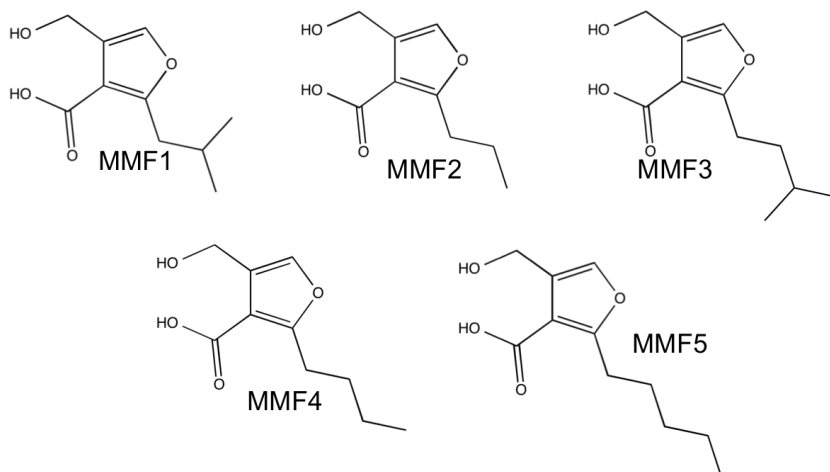
**Figure 1.6. Representatives from the crystal structure of the TetR family homodimer MmfR from the methylenomycin biosynthetic cluster of *S. coelicolor* in its apo form**  
**Left)** Filled surface model. **Right)** Cylindrical cartoon model; where A) indicates the location of the N-terminal DNA binding domain in each monomer and B) the location of the C-terminal ligand-binding pockets.

(Adapted from the work of Dean Rea et al.) (73)

### Methylenomycin Furans – Microbial Hormones

In Figure 1.7, the chemical structure of the five structurally similar methylenomycin furan signalling molecules that are thought to bind MmfR can be found. These molecules are based around a furan group; a five membered ring made of four carbons and one oxygen atom. Each MMF also has a carboxylic acid group and a hydroxyl group attached to this furan group as well as a variable region with a branched or unbranched alkyl chain of differing length.(54)





**Figure 1.7. Chemical structures of the known *S. coelicolor* methylenomycin furans**  
(Based on the work by Corre et al.) (54)

Previous research by Nicolas Malet has implied that it is the alkyl chain of these small molecules that allows the MMFs to fit the MmfR binding pocket completely and initiate the conformational change.(55) Work with synthetic analogues also indicated that an alkyl chain with a length of between three and five carbons is optimal triggering MmfR release. The hydrophobicity of this alkyl tail also means that the MMFs with longer chains are less soluble in water. The implications of the different properties of the five ligands *in vivo* are not known.

Research carried out by Peter Harrison (73) revealed that the different MMFs vary in their binding kinetics to MmfR. In particular, his work consistently indicated that MMF1 with its branched alkyl chain was best at causing MmfR release. Findings on the other four MMFs were less distinct. It is not yet clear why there are five different methylenomycin furans and what impact the choice of these ligands will have on transcription. More *in vivo* work is needed before conclusions can be made about the five molecules.

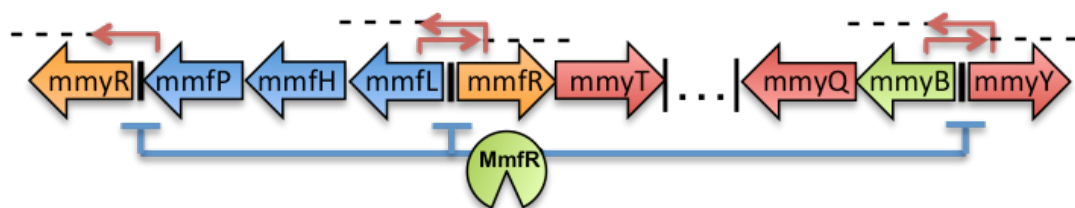
### Methylenomycin Furan Biosynthesis

Of MmfLHP, the enzymes responsible for the biosynthesis of MMF1-5, MmfL has shown 25% amino acid identity and 43% similarity over 83% of its sequence to AfsA, a butenolide synthase used in the production of the GBL signalling molecule A-factor.(71, 74) MmfL also appears to be the most critical of the three in the MmfLHP operon with MMF production ceasing in *mmfL* knockouts and low level MMF production detected when *mmfL* was present but *mmfH* and *mmfP* were absent (see the genes highlighted in blue in Figure 1.5).(54, 71) On the other hand, the production of a *mmfP* or *mmfH* knockout did not result in the termination of MMF production but just lower levels produced, the indication of which being that MmfP-like and MmfH-like proteins are present in *S. coelicolor* that can partially takes over their

roles.(54) The proposed functions of the three enzymes are that *mmfL* codes for a butenolide synthase, *mmfH* for a flavin dependent dehydrogenase and *mmfP* for a phosphatase.(75) It also appears likely that precursors to the steps catalysed by MmfLHP may be intermediates common also to the biosynthetic pathways of the GBLs.(54, 74)

### Transcriptional Regulation by MmfR

MmfR is proposed to bind to three different intergenic locations within the methylenomycin cluster, thereby blocking five different promoters and regulating the expression of the five associated operons. These intergenic protein-binding regions are known as methylenomycin auto-regulatory response elements or MAREs. A diagram of the location of these operator sites within the methylenomycin biosynthetic cluster can be found below in Figure 1.8, with the MARE operators denoted with bold vertical black lines. (Please note that this diagram is not to scale and is purely a representation to show the assembly of the gene cluster. This diagram is a reduced form the one shown earlier in Figure 1.5 and has had a section of the biosynthetic cluster removed so to only focus on the areas surrounding the MARE operators.)



**Figure 1.8. Proposed regulation of the methylenomycin gene cluster from the SCP1 plasmid of *S. coelicolor* by the TetR family member; MmfR**

*The blue lines represents the proposed location of MmfR binding operator sites and red arrows indicate the operons regulated by these operators and the direction of transcription. Proposed functional attributions are **mmfR and mmyR** – transcriptional repressors, **mmfLHP** – methylenomycin furan biosynthetic genes, **mmyT and mmyQ** – methylenomycin biosynthetic genes, **mmyB** – transcriptional activator, based on the work of Chater and Bruton (65)*

The five operons directly regulated by MmfR are *mmyR*, *mmfLHP*, *mmfR*, *mmyBQEDXCAPK* and *mmyYF*. It has been shown experimentally that there is a lag in the expression of these different operons with the transcripts for the biosynthetic genes being detected last.(71) MmfR regulation of *mmfR* and *mmfL* means that the system is auto-regulatory with both negative and positive control mechanisms (Figure 1.8) where MmfR repression influences both its own synthesis as well as the synthesis of the MMFs which will then cause its release from the MARE operator.

The MARE operator sequences at each of the intergenic regions are as follows;

```

mmfLR      5'  ATAATACCTTC  CGCAGGTATATT  3'
              3'  TATTATGGAAGG  GCGTCCATATAA  5'

mmyR       5'  AACATACCTTC  CGAGGGTATGTT  3'
              3'  TTGTATGGAAGG  GCTCCCATACAA  5'

mmyBY     5'  AAAAAACCTTCG  GGAAGGTTTGAC  3'
              3'  TTTTGTGGAAGC  CCTTCCAAACTG  5'

```

Conserved nucleotides between the three sites have been highlighted in yellow. The MARE operator sequences are 24 base pairs in length, 12 bp for each monomer, but are not a perfect palindrome despite MmfR being homodimeric. The three different intergenic MARE operator sequences do also vary quite considerably with only 13 out of the 24 bases remaining constant between the three sites, indicating possible different affinities of MmfR at each site.

There are also different -35/-10 sequences for each of the five promoters regulated by MmfR (Table 3.1) in this cluster. These different promoters may therefore have different strengths, adding yet another layer of control to this biosynthetic cluster. The variation in promoter strength between the five operons is yet to be shown experimentally however.

MmfR also has a paralogue, MmyR, which has been shown to also be involved in the regulation of methylenomycin biosynthesis.

### 1.4.2 MmyR and its Role in Transcriptional Regulation

Much less is known of the transcriptional regulation by MmyR. This TetR family member shares 35% identity with MmfR (over 54% of its sequence, covering the N terminal regions) (48) but research has suggested that it has a different mechanism of action.

MmyR has been found to not be soluble when purification has been attempted with existing expression systems in *E. coli*. It has therefore not been possible to purify MmyR for crystallography or other *in vitro* analyses. For this reason, there is a lack of evidence as to whether MmyR will also bind the MMF molecules although genomic manipulations and the creation of knockout mutants have indicated that MmyR is not sensitive to the MMFs and will not transduce their signal.(71) The lack of an accurate crystal structure also limits the bioinformatical analysis that can be done on this homologue e.g. a docking analysis is a lot less likely to be reliable.

MmyR can likely be regarded of as one of the pseudo-receptors mentioned in Section 1.2. These pseudo-receptors appear to share high levels of identity with their matching ‘real receptor’ but show differences in DNA and ligand binding properties, often binding a larger

range of DNA targets as well as sometimes acting only as a repressor in a ligand bound form.(50) Work by Choi *et al.* in 2004 into GBL receptors in non-streptomycete actinomycetes found that these pseudo-receptors often have a basic pI, usually around nine or ten. The real receptors on the other hand have a pI of around five or six.(76) Table 1.1 contains details on the molecular weight and predicted isoelectric point of MmfR, MmyR and analogues ScbR and ArpA, as calculated using the software ProtParam.(77) MmfR has a predicted pI of 6.0 compared to 5.7 for ArpA and 6.4 for ScbR thus indicating that it is a ‘true receptor’, something which matches experimental findings.(71) MmyR on the other hand has a pI of 7.8 further indicating that it is unlikely to function as MmfR does or have the same operator/ligand binding properties. This does need to be proven experimentally however as this correlation between mode of action and predicted pI does not cover all cases and should be seen as no more than an indication of potential mechanism of action. Whether MmyR will bind to alternative ligands than the MMFs or different operators to the three MARE operators is yet to be seen.

**Table 1.1. Properties of TetR family member monomers MmfR, MmyR, ScbR and ArpA**

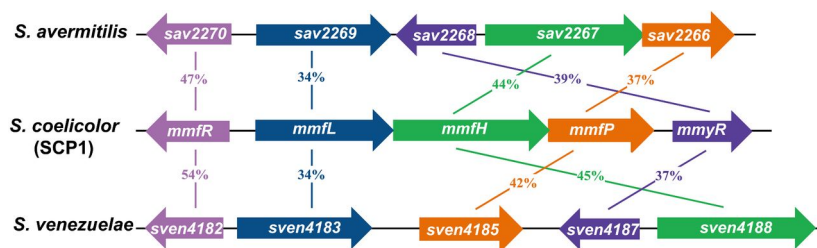
*The predicted pI of these proteins comes from the ProtParam software, according to the standard parameters set by this software.(77)*

| Protein     | Amino acid length | Molecular weight (daltons) | Theoretical pI |
|-------------|-------------------|----------------------------|----------------|
| <i>MmfR</i> | 214               | 24052                      | 5.99           |
| <i>MmyR</i> | 203               | 21883                      | 7.78           |
| <i>ScbR</i> | 215               | 23861                      | 6.38           |
| <i>ArpA</i> | 287               | 32107                      | 5.67           |

Despite the unclear ligand/operator binding properties of MmyR, knockout strains for *mmyR* produce a phenotype overexpressing methylenomycin (65, 78) indicating a key role in regulation and repression. There is no experimental evidence that MmyR will bind the three MARE operators as yet however. An *mmfR* knockout on the other hand reveals a phenotype similar to that of the wild type strain despite the repression and release of MmfR by the MMFs having been shown experimentally.(54, 55, 71) It is clear therefore that regulation by the MmfR-MmfLHP-MmyR network is a complex one.(79) One hypothesis is that MmyR plays a secondary role, stopping the positive feedback loop once enough methylenomycin has been produced rather than allowing its initial production.(54) MmfR on the other hand is the preliminary repressor, only repressing the *mmy* biosynthetic genes in the absence of the auto-inducing MMFs. More findings to support this hypothesis have been found in this project and will be discussed later. The regulation of natural product biosynthesis via MmfR/MmyR is something that is thought to be homologous in mechanism to the repression found in a number of other *Streptomyces* strains, this will now be discussed.

### 1.4.3 Analogous Systems to MmfR/MmfLHP/MmyR

The analogous pathways in other strains of *Streptomyces* with similar control mechanisms to that found in the regulation of methylenomycin also include two distinct repressor proteins as well as analogues of the *mmfLHP* cluster. For example, SgnR and GbnR are MmfR and MmyR analogues respectively, found in *S. venezuelae*.(80) SAV\_2270 and SAV\_2268 from *S. avermitilis* and SHJG\_7318 and SHJG\_7322 from *Streptomyces hygrosopicus* are also close homologues of MmfR and MmyR.(81, 82) A schematic comparison of some of these clusters is represented in Figure 1.9 with the associated percentage identities at the amino acid level displayed.

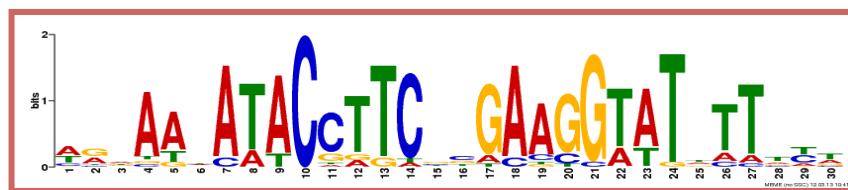


**Figure 1.9. Organisation of *mmfR/mmflHP/mmyR*-like clusters found in *S. venezuelae* and *S. avermitilis* and their associated amino acid percentage identities compared to the *S. coelicolor* cluster**

*Orthologues are represented in matching colours and approximate gene sizes shown by the size of the arrows. This figure was taken directly from the 2016 paper by Niu et al. (35).*

A sequence analysis with MEME (Table 7.3) shows that all of the transcriptional repressor homologues have a highly conserved helix-turn-helix DNA binding domains consistent with being TetR family members.(83)

These homologous transcriptional repressors not only show similarity in amino acid sequence but a comparison with the MARE operator sequence has revealed predicted intergenic auto-regulatory response elements (AREs) where these homologous TetR family members are believed to bind in their own systems.(84) This conserved 24 bp protein binding DNA sequence from the four homologous strains just mentioned is shown in Figure 1.10. As yet however, there is no experimental data confirming TFR binding to this predicted motif.



**Figure 1.10. Predicted motif of the intergenic auto-regulatory response element (ARE) sequences from *S. coelicolor*, *S. venezuelae*, *S. avermitilis* and *S. hygroscopicus***

*Schematic generated by the MEME suite (83) courtesy of Vincent Poon (84)*

The role of the pseudo-receptor appears to be conserved throughout the different strains where knockouts reveal a phenotype of natural product overproduction. For example the recent creation of a *gbnR* knockout in *S. venezuelae* resulted in the overproduction and discovery of novel natural products; the gaburedins (a family of  $\gamma$ -aminobutyrate (GABA)-derived ureas).(75) These gaburedins bear very little structural or mechanistic resemblance to methylenomycin A the implications of which are that these homologous regulatory clusters are used to control the production of a range of unrelated natural products. Homology does not extend to the BGC (biosynthetic gene cluster). The BGC target in *S. avermitilis* and *S. hygroscopicus* are as yet unknown.

Figure 1.9 also shows that, in addition to similarities between TetR receptors in these MmfR/MmfLHP/MmyR-like clusters, there are high levels of identity between MmfLHP and their orthologues giving the indication that these other strain may also produce AHFCA-like signalling molecules rather than GBLs.(54, 81) The AHFCAs may therefore be a novel and distinct class of microbial hormones found across multiple species.

#### 1.4.4 Other Methylenomycin Control Mechanisms

In addition to the interaction of furan microbial hormones and the MmfR transcriptional repressor, methylenomycin production is also known to be regulated by pathway specific activators as well the use of rare codon and low alanine levels.(63) There is also activation of methylenomycin production upon a rapid drop in pH, an artefact that is little understood.(63) These factors all come together to produce a complex regulatory network, influencing the biosynthesis of methylenomycin.

#### **Transcriptional Activators**

Methylenomycin production is believed to be dependent on the presence of the pathway specific transcriptional activator; MmyB.(71, 85) The gene for this activator is the first found in the largest operon of methylenomycin biosynthetic genes; *mmyBQEDXCAPK*, one of the operons directly regulated by MmfR (and possibly MmyR). Research has indicated that MmyB binds to pseudo-palindromic B-boxes in the methylenomycin cluster and will activate

transcription of its own operon, *mmyTOG*, *mmyY* and *mmyF*.(71) It is hypothesised that maybe the main role of the MmfR/MMF/MARE operator regulatory system is to de-repress production of MmyB rather than direct the activation of biosynthetic genes via the removal of repression at the MARE operator. MmfR therefore may only indirectly influence methylenomycin biosynthesis. Knockout *mmyB* strains will not produce methylenomycin, even when MmfR/MmyR repression was removed by the MMFs.(71) MmyB does not appear to regulate the production of the MMFs but like MmfR, is potentially self-regulatory leading to a self reinforcing model of the activation of methylenomycin biosynthesis.(71)

A bioinformatical analysis of MmyB shows similarities with Xre (Xenobiotic Response Element) family members, a family of common transcriptional regulators which, like TetR family members, have a HTH DNA binding domain.(85, 86) Members of this family will bind to DNA in homodimeric and heterodimeric forms and includes the lambda-bacteriophage Cro protein.

### TTA Codon

The gene for *mmyB* also contains the rare TTA codon thereby adding an extra level of translational regulation to the system.(87) The synthesis of tRNA for this codon (BldA) only occurs at the later stages in the complex *Streptomyces* life cycle.(88) Bld expression depends on the presence of the sigma factor BldN,(89) which itself is regulated by BldG and BldH both involved in preparing the cell for the production of aerial mycelium.(90) The TTA codon can also be found in *mmfL*, one of the genes to make an MMF biosynthetic enzyme. This use of the rare TTA codon is also found in a number of other antibiotic BGCs including those for actinorhodin and prodiginine.(71)

## Hypothetical Regulatory Mechanisms

### Sigma Factors

A number of other sigma factors may also be involved in the transcriptional regulation of this system. Promoter specificity of a polymerase can be influenced by sigma factors that interact with it and recruit the core RNA polymerase enzyme.(90) Alternative sigma factors are used to control gene expression. It is also possible to have anti-sigma factors and anti-anti-sigma factors, which exert yet more layers of control. *E. coli* only has seven sigma factors whereas *S. coelicolor* is believed to have at least 65,(90) showing the much more extensive control mechanisms found in streptomycetes. There is therefore potential that sigma factors are involved in the regulation of methylenomycin biosynthesis, however the specific sigma factors involved and the extent to which they have control is as yet unknown.

### Riboswitches

It is also possible that the methylenomycin BGC is also regulated by riboswitches. Riboswitches are sections of mRNA molecules that can directly regulate themselves in the

presence of small effector molecules (possibly the MMFs).(91) Riboswitches are particularly common in bacteria but, as with sigma factors, the experimental evidence for their involvement in regulation of translation of the methylenomycin BGC is currently lacking.

## 1.5 Introduction to Research Aims

### Research questions

*Do all five methylenomycin cluster promoters, controlled by MmfR, have the same strength?*

*Does MmfR bind in the same way to all three MARE operator sequences?*

*Is MmfR release by the MMFs the same at all three MARE operators?*

*Does MmfR respond to all five furan compounds?*

*Do all five MMFs have the same efficacy?*

*What are the key residues in ligand binding?*

*How does MmyR binding to the MARE operator and the MMFs vary from that of MmfR?*

*Are there any other ligands that MmyR may bind to?*

*Could MmfR, MMFs and MARE operators be used as a multi-host efficient novel inducible expression system for GC rich bacteria? Would this allow the purification of recombinant proteins?*

### Hypothesis

*The promoters that are predicted to be controlled by MmfR have different -35/-10 sequences so it is possible that they will have varying strengths.*

*In vivo, MmfR will bind to DNA at the MARE operator and be released upon the addition of a MMF compound.*

*The three MARE operators have different semi-palindromic sequences and so are likely to show differential binding to MmfR.*

*MmfR will respond to all of the MMFs but due to the differing length of alkyl chain between the five molecules there are likely to be differences in the binding potential of each.*

*MmyR is only produced after methylenomycin biosynthesis. It will bind to the methylenomycin cluster operators but not be released by the MMFs, thereby repressing biosynthesis when methylenomycin has been produced to conserve cell resources and to protect the bacterium from the potentially lethal effects of excess methylenomycin. Alternatively, an unknown ligand may 'activate' MmyR as a repressor.*



### 1.5.1 How Will These Research Questions Be Answered?

#### **MmfR and MmyR functionality**

To assess the binding abilities of MmfR and MmyR to the MARE operator and MMFs a newly optimised *luxCDABE* (92) reporter system was developed for use in *S. coelicolor* (see Section 1.5.2 for more details). Vectors were created to contain *luxCDABE* under the control of different methylenomycin cluster intergenic regions (including the MARE operators), allowing the investigation of promoter strength as well as the effect of different operator sites. Vectors were also created that contained *mmfR* or *mmyR* to allow the study of TetR regulation over this BGC. MMFs could then be added to this system and their effects on bioluminescence measured. This luciferase assay as well as a bioinformatical analysis of amino acid sequences of MmfR, MmyR and their homologues fulfilled the investigative demands the first six research questions in this project.

When investigating the ligand binding pocket of MmfR an *in silico* analysis of its crystal structure (73) as well as a comparison of primary structures with known homologues was done. This information was then used to perform site directed mutagenesis on these key ligand-binding residues and, using the luciferase assay, to study the effects this had on the release of MmfR from the MARE operators as well as whether MmfR function can be returned by a higher concentration of MMFs.

A comparison could then be made between MmfR and MmyR activity in methylenomycin regulation and the information brought together to explain how MmfR/MmyR act together to regulate methylenomycin biosynthesis.

#### **Developing a Novel Inducible Expression System**

The production of a heterologous expression super host was investigated by adding the *luxCDABE/mmfr* system to the genetically streamlined *Streptomyces albus* host (93) to establish whether it would be suitable to use for the inducible expression system. This then lead onto an investigation of other alternative streptomycete hosts as well as a literature review of current expression systems.

Following on from this, MmfR/MMF/MARE operator interactions were optimised so they could be adapted for use as a novel inducible expression system. Vectors were designed for the overexpression of a gene of choice to be trialled in this system and a protocol developed for the collection of recombinant proteins from *S. coelicolor*. An explanation of why a novel inducible expression system is needed can be found in Section 1.5.3.

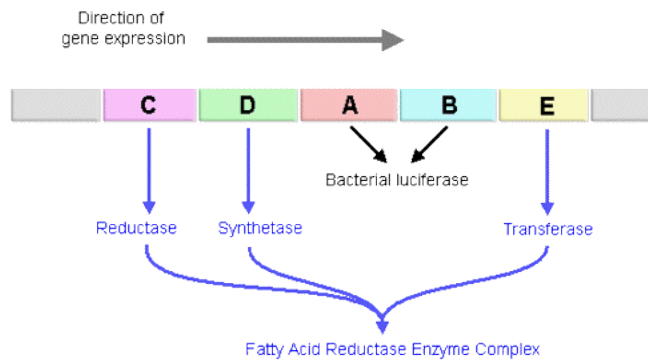
### 1.5.2 Choice of Reporter System Used

There are a number of reporter systems available and it was necessary to choose a suitable one for answering the research questions in this project. Unfortunately a number of reporter systems are not suitable for use in GC rich bacteria such as *Streptomyces*. The *lacZ* system for example, is a usually easy and sensitive reporter system and is widely used. Unfortunately, streptomycetes have an enzyme that has beta-galactosidase activity and therefore interferes with *lacZ* expression making it an unsuitable reporter system.(51) Efforts to produce *Streptomyces* knockout of this beta-galactosidase enzyme have often been unsuccessful and results in poorly growing colonies.

A beta-glucuronidase system (94) has been used to study the regulation of virginiamycin biosynthesis by BarA and its cognate ligand VB from *S. virginiae* (Section 1.2.2) in tobacco plant cells in 2006.(95) This system was then later developed for use in actinomycetes, showing promise as a viable reporter system for these bacteria.(96) This is a colorimetric assay however and therefore limited in the quantitative results that can be obtained. It was therefore decided that it would be best to utilise a reporter system that produces more quantitative results such as fluorescence or bioluminescence.

The *gpf* system has been trialled in *Streptomyces* with a degree of success. Streptomycetes will auto-fluoresce in blue light, limiting the range of colours that can be used to study different components of a system but otherwise appears to be fairly successful in these Gram-positive bacteria. The disadvantage of the technique however, is that it tends to photo-bleach rapidly, limiting the genes it can be used to study as well as posing problems if multiple readings need to be taken.(92)

A *luxAB* system from *Vibrio harveyii* has shown also a lot of potential for use in streptomycetes. The downside of this system is that *luxAB* only produce the enzyme luciferase and therefore there is the need to add a substrate to produce luminescence. This substrate may not pass through all bacterial cell walls with the same efficiency and may disrupt cell growth if overlaid onto a solid culture.(92) For this reason, the full *luxCDABE* system, which was optimised for GC high bacteria by Justin Nodwell and his team in Canada, (92) was used to investigate MmfR, MmyR, the MMFs and MARE operators further. LuxA and LuxB together form the heterodimeric luciferase protein whereas LuxC, D and E are the enzymes needed to form tetradecanal, the fatty aldehyde substrate of luciferase.(97) All other biochemicals required by this reaction are found naturally in bacterial cells allowing a self generated bioluminescent response with no need for external manipulations.(98) A diagram of the *luxCDABE* operon can be found in Figure 1.11.



**Figure 1.11. Arrangement of open reading frames in the *luxCDABE* reporter system and the assigned functions of individual components of this operon**

*This image was taken directly from the 2009 paper by Lin et al., (99)*

### 1.5.3 Why Do We Need A Novel Inducible Expression System?

Currently, there are a number of commercially available inducible expression systems. Famous examples of these are the vectors regulating gene expression via the lac operon, inducible upon the addition of IPTG. IPTG is a lactose mimic that is not enzymatically broken down like lactose and so remains at a constant level, this will bind to the lac repressor (LacI) and cause its release from DNA thereby allowing the expression of a gene of interest.(100)

Specificity of suitable expression hosts is little understood with each protein requiring a slightly different set of optimal expression conditions.(101) Many current systems are largely based in the Gram-negative *Escherichia coli* due to its fast growth, ease of culturing and well understood uses as a ‘cell factory’.(102) It is possible to optimise current expression systems to some extent to improve the expression of heterologous genes. For example genomic GC content varies widely across bacteria, ranging from anything between under 20% GC content to over 70% and so codon usage is also a key factor when designing recombinant genes.(103) There are a number of proteins that still cannot be efficiently over produced and purified using existing methods however. This can be due to a variety of complications such as physiological conditions not being suitable for the correct folding of the protein, low expression levels or because the host cannot carry out the required post-translational modifications.(101) Alternative expression systems have been developed to those in *E. coli* including systems based in yeast, other bacteria and fungi as well as those for mammalian cells.(104) These hosts will all provide slightly different conditions for protein expression, which may prove optimal for some proteins, but again these systems again cannot express all genes. There is no universal heterologous expression host. There is hope that the MmFR/MMF/MARE operator system, analogous to LacI/IPTG/lac operator can be used to

provide an alternative expression system for the overexpression of recombinant genes in the Gram-positive *Streptomyces* species. This system would hopefully be useful in the production of proteins currently not possible in existing systems.

One benefit of using streptomycetes as an expression host is their high innate protein secretion capacity.(101, 105) This has the advantage of an increased chance of the protein folding properly (106) as well as a reduced requirement for expensive purification techniques. This is therefore something that would be beneficial to include in the expression system being designed in this project.

In recent years there have been a number of systems developed for heterologous expression in streptomycetes that have shown promise. For example, the work by Noda *et al.* in 2015 showed great success with the production of streptavidin from a streptomycete host.(107) Streptavidin is originally from *Streptomyces avidinii* so seems logical therefore that it is expressed better in these GC high bacteria as conditions are likely to be closer to the native conditions needed for streptavidin production. The work by Noda *et al.* resulted in the production of a much more thermostable streptavidin product compared to those produced by *E. coli* systems, thereby expanding the potential applications of streptavidin-biotin interactions. Not all streptomycete expression systems produce a high protein yield however and much optimisation is needed. An example of the type of optimisation done includes the work by Wilkinson *et al.* who investigated improving expression systems in actinomycetes based on optimising promoters. This led to 100 times more product than when using than using wild type promoters.(108) Despite these successes however, there are still many instances where a heterologous protein cannot be purified from streptomycetes and so novel inducible expression systems are still very much in demand and it is for this reason that an additional inducible expression system is being developed in this project.

## 1.6 Outline of Thesis Structure

Chapter two specifies all of the stock solutions and protocols used to obtain results for this thesis as well as specifics on the source of all consumables used. Included in this section are details on primers used as well as lists of vectors and strains created. Further information on how these techniques developed based on experimental findings can also be found throughout the following research chapters.

Chapter three presents the optimisation of the *luxCDABE* reporter gene system for GC high bacteria for use in studying the interactions between MmfR (and paralogue MmyR) with the MMFs and the binding to the MARE operator. The chapter gives details on the assembly of

vectors for this method and how these come together to create an arrangement that can be adapted to study different aspects of the regulation of methylenomycin biosynthesis. In addition, results from the investigation into the strength of different promoters in the methylenomycin biosynthetic cluster are reported within this chapter.

Chapter four further expands on this *luxCDABE* reporter system specifically looking at MmfR as a transcriptional repressor. This chapter is divided into two main sections, the first looking at MmfR/MARE operator interactions and how binding varies at the three operators. The second is an investigation into MmfR/MMF interactions and includes details on all five methylenomycin furan ligands and their binding potentials to MmfR as well as an investigation into the MmfR ligand binding pocket and the production of mutants that were then also tested using the luciferase assay.

Chapter five follows on from the investigation into MmfR, this time looking at its paralogue MmyR. Again both interactions with the MARE operator and the MMF ligands were investigated. Due to the functionality of MmyR being different to MmfR, this chapter then goes in a slightly different direction, investigating other possible ligands for this second repressor rather than studying key residues in ligand binding.

Chapter six further explores the self-regulatory mechanism of MmfR, also using the luciferase assay. Investigations were carried out into the differences in MmfR repression and release when it is under the control of its own promoter. This chapter also briefly examines the potential of MmyR auto-regulation. This chapter is concluded with a proposed mechanism, combining the function of MmfR/MmyR in the regulation on methylenomycin biosynthesis, based on all of the investigative findings up until this point.

Chapter seven is the final investigative chapter and summarises all of the findings of chapter three and four to develop a novel expression system for use in GC rich bacteria, utilising MmfR/MMF/MARE operator interactions to induce transcription. This chapter first looks at the potential of creating an optimised streptomycete expression host followed by details on the creation of vectors for this novel expression system as well as preliminary trials into using it with *S. coelicolor* as a heterologous expression host.

Chapter eight and nine then discuss and conclude all of the findings from the previous five chapters as well as commenting on the implications of this work in wider research and explaining the possible future work that could be carried out.

## 2 Materials and Methods

### 2.1 Materials and Equipment

**Table 2.1. Consumables used**

| <i>Supplier</i>  | <i>Material</i>   |
|--|---|
| <b>Agilent Technologies</b>                                      | QuikChange Lightening SDM kit (109)   |
| <b>Biotium</b>   | GelRed™ nucleic acid dye  |
| <b>Cell projects Ltd</b>   | 2 mm short electroporation cuvettes   |
| <b>Corre group, University of Warwick, Coventry, UK</b>          | pCC4 – a pOSV566 derivative with apramycin resistance instead of ampicillin   |
| <b>Corning</b>   | 12-well plates - transparent, sterile flat bottomed tissue culture plates   |
| <b>Day-Impex Ltd</b>   | Virkon® disinfectant  |
| <b>Expedeon</b>  | InstantBlue™ Coomassie® stain   |
| <b>GE Healthcare</b>   | Ni Sepharose 6 Fast Flow  |
| <b>Invitrogen</b>  | One Shot® TOP10 Chemical Transformation kit<br>Champion™ pET Directional TOPO® Expression Kits (110)<br>Hygromycin B and UltraPure™ agarose   |
| <b>Justin Nodwell <i>et al.</i>, McMaster University, Canada</b> | Luciferase gene constructs,(92) in particular L1, 11NY and sp105 vectors shown in Table 2.5   |
| <b>Merck-Millipore</b>   | Amicon Ultra 15 Centrifugal Filter Unit MWCO 10 kDa<br>Amicon Ultra 0.5 Centrifugal Filter Unit MWCO 10 kDa<br>0.22 µm syringe filter   |
| <b>MP Biomedicals</b>  | FastDNA® SPIN Kit for Soil (111) and Lithium Chloride   |
| <b>National Diagnostics</b>                                      | ProtoGel 30% acrylamide   |
| <b>New England Biolabs (UK)</b>                                  | NEB 5-alpha Competent <i>Escherichia coli</i><br>Phusion high fidelity DNA polymerase<br>High fidelity restriction enzymes with cut smart buffer<br>Shrimp alkaline phosphatase<br>ColorPlus™ Prestained Protein Ladder, Broad Range (10-230 kDa)   |
| <b>Pernodet Group, University of Paris-Sud, France</b>           | pOSV566 vector  |
| <b>Roche Diagnostics</b>   | Expand High Fidelity PCR System   |
| <b>Sarstedt</b>  | Petri dishes and universal tubes  |
| <b>Scientific Laboratory Supplies Ltd.</b>                       | D-mannitol  |
| <b>Sigma Aldrich</b>   | Primers, agarose for electrophoresis, dialysis tubing<br>Antibiotics; apramycin, ampicillin, kanamycin and chloramphenicol<br>SIGMAFAST™ protease inhibitor cocktail tablets, EDTA-free for use in purification of histidine-tagged proteins<br>Protease inhibitor cocktail, for use in tissue culture media, DMSO solution |
| <b>Thermo Fisher Scientific - Fermentas</b>                      | FastRuler™ DNA ladders<br>GeneJET™ Plasmid Miniprep kit (112)<br>GeneJET™ Gel extraction kit (113)<br>Phusion High Fidelity DNA polymerase<br>Restriction enzymes and buffers<br>Routine PCR Taq DNA polymerase kit<br>T4 DNA ligase and PEG (114)<br>Wedge shaped spreaders and small cell scrapers                        |
| <b>VWR International</b>   | Bacteriological agar<br>Trichloroacetic acid for synthesis  |

Table 2.2. Equipment used

| <i>Supplier</i>                 | <i>Equipment</i>                               |  |
|---------------------------------|--|--|
|                                 | <i>Name</i>                                    | <i>Type</i>  |
| <b>Beckman Coulter</b>          | Centrifuge                                     | Avanti j-25  |
| <b>Bio-Rad Laboratories Ltd</b> | Horizontal electrophoresis system              | Wide Mini-Sub cell GT cell                               |
| <b>Cole-Parmer</b>              | Water bath                                     | 2 litre StableTemp                                       |
| <b>Eppendorf</b>                | Microcentrifuge<br>PCR machine<br>PCR machine  | 5424 R<br>Mastercycler Nexus<br>Mastercycler epgradient  |
| <b>Fisher Scientific</b>        | pH meter<br>NanoDrop                           | Acument basic AB15<br>NanoDrop 2000<br>Spectrophotometer |
| <b>Grant</b>                    | Water bath                                     | JB1  |
| <b>Hettich</b>                  | Centrifuge                                     | Rotina 46R   |
| <b>INFORS HT</b>                | Incubator shaker                               | Multitron  |
| <b>New Brunswick Scientific</b> | Incubator shaker                               | C24  |
| <b>Photek</b>                   | Photon counting CCD camera<br>Imaging software | HRPCS4<br>Photek IFS32                                   |
| <b>StarLab</b>                  | Heat block                                     | Mini Dry Bath Incubator                                  |

Table 2.3. Software used

| <i>Software</i>                | <i>Distributor</i>  | <i>Reference</i> |
|--------------------------------|---|------------------|
| <b>ChemDraw Pro 15.0</b>       | Perkin Elmer  |                  |
| <b>Clone Manager Basic 9</b>   | Sci-Ed Software   |                  |
| <b>Clustal Omega v.1.2.1</b>   | Conway Institute UCD Dublin                                 | (48)             |
| <b>Excel for Mac 2011</b>      | Microsoft Corporation                                       |                  |
| <b>GraphPad Prism 6 and 7</b>  | GraphPad Software   | (115)            |
| <b>Image32</b>                 | Photek  |                  |
| <b>Ligplot+ v.4.5.3</b>        | European Bioinformatics Institute                           | (116)            |
| <b>MEME v.4.10.0</b>           | National Centre for Research Resources                      | (83)             |
| <b>MacPyMOL v.1.3</b>          | Schrodinger   | (117)            |
| <b>SnapGene Viewer v.3.0.3</b> | GSL Biotech   | (72)             |
| <b>SwissDock</b>               | Swiss Institute of Bioinformatics                           | (118)            |
| <b>UCSF Chimera v.1.8.1</b>    | Resource for Biocomputing,<br>Visualization and Informatics | (119)            |

Table 2.4. Primers used

| <i>Name</i>   | <i>Sequence 5' to 3'</i>            |
|---|-------------------------------------|
| <b>Screening primers</b>  |                                     |
| T7 forward  | TAATACGACTCACTATAGGG                |
| T7 reverse/terminator   | GCTAGTTATTGCTCAGCGG                 |
| Luciferase screening forward  | AAGCCACTGAGCGGGAGCTTG               |
| Luciferase screening reverse  | GACGCTGTTGTGCGCCGAAGTTG             |
| <i>glnB</i> insert screening forward                                      | CACGATCCAAAGGAGGATGACG              |
| <i>glnB</i> insert screening reverse                                      | CACCTGCAGCCGGGCGGCC                 |
| <b>Primers for cloning intergenic regions into the <i>lux</i> vectors</b> |                                     |
| <i>mmfRp</i> forward  | GGCTGCCTTCCTTCGTGTG                 |
| <i>mmfRp</i> reverse  | AGGGGCGCTACATCTCCCG                 |
| <i>mmfYp</i> forward  | CACGATATCATCCTGCCGCGCGGTAGCC        |
| <i>mmfYp</i> reverse  | GTGGGATCCCAACGCCGAGTCCTCTCAAG       |
| <i>mmfBp</i> forward  | CCGGATATCGGTGAACTCCTTCGGCGAGTG      |
| <i>mmfBp</i> reverse  | GTGGGATCCGGCGCCTCACAGTGTCAAACCTTC   |
| <i>mmfLp</i> forward  | CACGGATCCGGCTGCCCTTCCTTCGTGTG       |
| <i>mmfLp</i> reverse  | GTGGATATCAGGGGCGCTACATCTCCCG        |
| <i>mmfYp</i> forward  | CCGGGATCCGGTGAACCTCCTTCGGCGAGTG     |
| <i>mmfYp</i> reverse  | GTGGATATCAGGGCGCCTCACAGTGTCAAACCTTC |
| <i>ermE*</i> forward  | CACGGTACCAGCTTGATGCCGGTC            |
| <i>ermE*</i> reverse  | CACGATATCGCTGACGCCGTTGGATAC         |
| <b>SDM primers</b>  |                                     |
| Y85F 1  | CAGCGCGCGAAGTGCTCCTCCACCACG         |
| Y85F 2  | CGTGGTGGAGGAGCACTTCGCGCGCTG         |
| Y144F 1   | TCCAGTCCACGAAGGGCAGGGGCAG           |
| Y144F 2   | CTGCCCCCTGCCCTTCGTGGACTGGA          |
| Y85A 1  | GGGCCAGCGCGCGCGTGCTCCTCCACC         |
| Y85A 2  | GGTGGAGGAGCACGCCGCGCTGGCCC          |
| Y144A 1   | GGTCCAGTCCACGGCGGGCAGGGGCAGC        |
| Y144A 2   | GCTGCCCCTGCCCGCGTGGACTGGACC         |
| <b>Other primers</b>  |                                     |
| <i>mmfR</i> forward ( <i>Hind</i> III restriction, to clone from pCC2)    | CACAAGCTTAAAGGAGGGCAGCCATGACGAGCG   |
| <i>mmfR</i> reverse ( <i>Not</i> I restriction, to clone from pCC2)       | CACGCGGCCGCGCGGACGCTCCCCGTC         |



Table 2.5. Vectors used in luciferase reporter gene assay

| Vector name | Genes contained  | Resistance conferred                  | Promoter       | Reference    |
|-------------|--|---------------------------------------|----------------|--------------|
| pCC4        | -  | Apr <sup>R</sup> and Hyg <sup>R</sup> | <i>ermEp</i> * | (120)        |
| pKMS01      | pCC4 with <i>mmfR</i> <sup>1</sup>                                       | Apr <sup>R</sup> and Hyg <sup>R</sup> |                | This project |
| pKMS03      | pCC4 with <i>mmyR</i>  | Apr <sup>R</sup> and Hyg <sup>R</sup> |                | This project |
| pOSV556     | -  | Amp <sup>R</sup> and Hyg <sup>R</sup> |                | (120)        |
| pKMS85      | pOSV556 with Y85F <i>mmfR</i> mutant                                     | Amp <sup>R</sup> and Hyg <sup>R</sup> |                | This project |
| pKMS144     | pOSV556 with Y144F <i>mmfR</i> mutant                                    | Amp <sup>R</sup> and Hyg <sup>R</sup> |                | This project |
| L1          | <i>luxCDABE</i> , intergenic region for <i>mmfL/mmjR</i>                 | Apr <sup>R</sup>                      | <i>mmfLp</i>   | (92)         |
| 11NY        | <i>luxCDABE</i> , <i>mmfR</i> and intergenic region for <i>mmfL/mmjR</i> | Apr <sup>R</sup>                      |                | (92)         |
| sp105       | <i>luxCDABE</i> , <i>mmjR</i> and intergenic region for <i>mmfL/mmjR</i> | Apr <sup>R</sup>                      |                | (92)         |
| L1F         | <i>luxCDABE</i> , reversed intergenic region for <i>mmfL/mmjR</i>        | Apr <sup>R</sup>                      | <i>mmfRp</i>   | This project |
| L2          | <i>luxCDABE</i> , intergenic region for <i>mmfP/mmyR</i>                 | Apr <sup>R</sup>                      | <i>mmyRp</i>   | This project |
| L3          | <i>luxCDABE</i> , intergenic region for <i>mmyY/mmyB</i>                 | Apr <sup>R</sup>                      | <i>mmyBp</i>   | This project |
| L3F         | <i>luxCDABE</i> , reversed intergenic region for <i>mmyY/mmyB</i>        | Apr <sup>R</sup>                      | <i>mmyYp</i>   | This project |
| L4          | <i>luxCDABE</i> and intergenic region containing <i>ermEp</i> *          | Apr <sup>R</sup>                      | <i>ermEp</i> * | This project |

Table 2.6. Other vectors used in this research project

| Vector name         | Extra details  | Size of vector | Resistance conferred                  | Reference   |
|---------------------|--|----------------|---------------------------------------|-------------|
| pET151              | <i>lacZ</i> , Amp <sup>R</sup>   | 5 760 bp       | Amp <sup>R</sup>                      | (121)       |
| pET151: <i>mmfR</i> | <i>lacZ</i> , <i>mmfR</i> , Amp <sup>R</sup>   |                | Amp <sup>R</sup>                      | This work   |
| pUZ8002             | RP4 derivative   | ~ 60 000 bp    | Kan <sup>R</sup>                      | (51), (122) |
| pKMS05              | L1 vector with <i>Streptococcus mutans</i> <i>glnB</i> and <i>luxCDABE</i>   | 13 047 bp      | Apr <sup>R</sup>                      | This work   |
| pKMS06              | L1 vector with <i>Salmonella enterica</i> <i>glnB</i> and <i>luxCDABE</i>  | 13 080 bp      | Apr <sup>R</sup>                      | This work   |
| pCC4                | Apr <sup>R</sup> and Hyg <sup>R</sup> , int_pSAM2 (containing attP for genomic integration), OriT, Ter, <i>ermEp</i> * (for more details see Figure 3.7) | 8 663 bp       | Apr <sup>R</sup> and Hyg <sup>R</sup> | (120)       |
| pOSV556             | Amp <sup>R</sup> and Hyg <sup>R</sup> , int_pSAM2 (containing attP for genomic integration), OriT, Ter, <i>ermEp</i> *                                   | 9 009 bp       | Amp <sup>R</sup> and Hyg <sup>R</sup> | (120)       |

<sup>1</sup> pKMS01 *mmfR* is sometimes referred to as wild type or WT *mmfR* when being compared to the mutants presents in pKMS85 and pKMS144

Table 2.7. Parent strains used in investigation

| Strain                                  | Genotype  | Reference    |
|---|---|--------------|
| <i>Streptomyces coelicolor</i> M145     | SCP1- SCP2-   | (51)         |
| <i>Streptomyces coelicolor</i> M1152    | SCP1- SCP2- $\Delta$ act $\Delta$ red $\Delta$ cpk $\Delta$ cda rpoB[C1298T])   | (123)        |
| <i>Streptomyces albus</i> J1704         | Wild type – GenBank: CP004370   | (124)        |
| <i>Escherichia coli</i> TOP10           | F- <i>mcrA</i> $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ) $\Phi$ 80 <i>lacZ</i> $\Delta$ M15 $\Delta$ <i>lacX74 recA1 araD139</i> $\Delta$ ( <i>araleu</i> )7697 <i>galU galK rpsL</i> (StrR) <i>endA1 nupG</i> | (110)        |
| <i>Escherichia coli</i> ET12567/pUZ8002 | F- <i>dam-13::Tn9 dcm-6 hsdM hsdR zjj-202::Tn10 recF143 galK2 galT22 ara-14 lacY1 xyl-5 leuB6 thi-1 tonA31 rpsL136 hisG4 tsx-78 mtl-1 glnV44</i> , pUZ8002 (Kan <sup>R</sup> )                            | (125), (122) |

Table 2.8. Strains created using various vectors to be used in the luminescence assay

The parent strain used for all of these was *S. coelicolor* M145 and vector 2 always uses the *ermEp\** promoter. 'Inducible' refers to the possibility of the strain repression being released by the MMFs, this had not been proven experimentally for all vectors at the time of creation however.

| Name given to strain | Vector 1 | Vector 2 | Type of sample         | Promoter in lux vector |
|----------------------|----------|----------|------------------------|------------------------|
| L1                   | L1       | -        | Positive control       | <i>mmfLp</i>           |
| 11NY                 | 11NY     | -        | MmfR, inducible        |                        |
| sp105                | sp105    | -        | MmyR, inducible        |                        |
| L1+pCC4              | L1       | pCC4     | Positive control       |                        |
| L1+ <i>mmfR</i>      | L1       | pKMS01   | MmfR, inducible        |                        |
| L1+ <i>mmyR</i>      | L1       | pKMS03   | MmyR, inducible        |                        |
| L1+Y85F              | L1       | pKMS85   | MmfR mutant, inducible |                        |
| L1+Y144F             | L1       | pKMS144  | MmfR mutant, inducible | <i>mmfRp</i>           |
| L1F+pCC4             | L1F      | pCC4     | Positive control       |                        |
| L1F+ <i>mmfR</i>     | L1F      | pKMS01   | MmfR, inducible        |                        |
| L1F+ <i>mmyR</i>     | L1F      | pKMS03   | MmyR, inducible        | <i>mmyRp</i>           |
| L2+pCC4              | L2       | pCC4     | Positive control       |                        |
| L2+ <i>mmfR</i>      | L2       | pKMS01   | MmfR, inducible        |                        |
| L2+ <i>mmyR</i>      | L2       | pKMS03   | MmyR, inducible        | <i>mmyBp</i>           |
| L3+pCC4              | L3       | pCC4     | Positive control       |                        |
| L3+ <i>mmfR</i>      | L3       | pKMS01   | MmfR, inducible        |                        |
| L3+ <i>mmyR</i>      | L3       | pKMS03   | MmyR, inducible        | <i>mmyYp</i>           |
| L3F+pCC4             | L3F      | pCC4     | Positive control       |                        |
| L3F+ <i>mmfR</i>     | L3F      | pKMS01   | MmfR, inducible        |                        |
| L3F+ <i>mmyR</i>     | L3F      | pKMS03   | MmyR, inducible        | <i>ermEp*</i>          |
| L4+pCC4              | L4       | pCC4     | Positive control       |                        |

## 2.2 Stock Solutions

### 2.2.1 Media Stock Solutions

#### LB medium

10 g tryptone  
5 g yeast extract  
10 g NaCl  
Make up to 1 L with distilled water  
adjust the pH to 7.0  
Autoclave

#### LB agar

10 g tryptone  
5 g yeast extract  
10 g NaCl  
15 g agar  
Make up to 1 L with distilled water  
adjust the pH to 7.0  
Autoclave

#### SFM (soya flour mannitol)

8 g bacto-agar  
8 g soya flour  
8 g mannitol  
Make up to 400 mL with tap water and  
mix together before autoclaving

#### 2XYT broth

16 g tryptone  
10 g yeast extract  
5 g NaCl  
Make up to 1 L with distilled water  
adjust the pH to 7.0  
Mix and autoclave

#### SOB media (126)

20 g tryptone  
5 g yeast extract  
0.5 g NaCl  
Make up to 1 L with distilled water,  
adjust to pH 7.5 and autoclave  
Add filter sterilised  $\text{MgSO}_4$  to a final  
concentration of 20 mM

#### SOC medium

20 g tryptone  
5 g yeast extract  
0.5 g NaCl  
Make up to 1 L with distilled water,  
adjust to pH 7.5 and autoclave  
Add filter sterilised  $\text{MgSO}_4$  and glucose  
each to a final concentration of 20 mM

#### TSB medium

17 g enzymatic digest of casein  
3 g enzymatic digest of soya bean meal  
5 g NaCl  
2.5 g dipotassium phosphate  
2.5 g glucose  
Make up to 1 L with distilled water  
adjust the pH to 7.3 (+/-0.2)

#### Antibiotics were used at the following final concentrations:

50 µg/ml apramycin  
100 µg/ml ampicillin  
25 µg/ml chloramphenicol  
50 µg/ml kanamycin  
50 µg/ml hygromycin  
25 µg/ml nalidixic acid

Autoclaving was done at 121 °C for 20 minutes and then media was stored at room temperature. Once antibiotics were added the media would be used immediately or stored in the fridge until required.

### 2.2.2 DNA Gel Electrophoresis

#### 50X TAE buffer

2 M Tris acetate, pH ~8.3  
50 mM EDTA  
In distilled water  
Filter sterilised

#### 1% gel – for one gel

1 g agarose  
In 100 mL 1X TAE buffer (40 mM Tris acetate and 1 mM EDTA)  
Heat in the microwave at full power for 90 seconds or until the agarose has dissolved  
Allow to cool slightly and add 5 µL GelRed™

### 2.2.3 Phenol Chloroform Extraction Buffers

#### Buffer I

50 mM Tris-HCl, pH 8  
10 mM EDTA

#### Buffer II

200 mM NaOH  
1% SDS

#### Buffer III

3 M potassium acetate, pH 5.5

Buffers were stored at 4 °C until needed.

### 2.2.4 Protein Purification Buffers and Solutions

#### Protein purification buffer

20 mM Tris-HCl pH 8  
100 mM NaCl  
10% glycerol

#### Improved protein purification buffer for Ni Sepharose purification

20 mM sodium phosphate (Na<sub>2</sub>HPO<sub>4</sub> and NaH<sub>2</sub>PO<sub>4</sub>)  
500 mM NaCl  
pH 7.4

#### Elution buffer

20 mM Tris-HCl pH 8  
100 mM NaCl  
10% glycerol  
200 mM imidazole

#### Improved elution buffer for Ni Sepharose purification

20 mM sodium phosphate (Na<sub>2</sub>HPO<sub>4</sub> and NaH<sub>2</sub>PO<sub>4</sub>)  
500 mM NaCl  
500 mM imidazole  
pH 7.4

Buffers were all stored at 4 °C until needed.

#### Protease inhibitors were used at the following final concentrations:

1 mM EDTA and 1:500 protease inhibitor cocktail for tissue culture from Sigma-Aldrich for culture plates.

One tablet of SIGMAFAST protease inhibitor cocktail in 100 mL 'improved protein purification buffer'.

## 2.2.5 SDS-PAGE Reagents and Buffers

### 4% SDS-PAGE stacking gel

4% ProtoGel acrylamide  
125 mM Tris-HCl pH 6.8  
0.1% SDS  
0.05% APS  
0.01% TEMED

### 12% SDS-PAGE resolving gel

12% ProtoGel acrylamide  
375 mM Tris-HCl pH 8.8  
0.1% SDS  
0.1% APS  
0.01% TEMED

### SDS-PAGE loading dye (2x)

125 mM Tris-HCl pH 6.8  
20% glycerol  
4% beta-mercaptoethanol  
0.2% bromophenol blue  
4% SDS

### SDS-PAGE running buffer (10x)

250 mM Tris-HCl pH 8.8  
2 M glycine  
1% SDS

Buffers were all stored at 4 °C until needed.

## 2.3 Protocols

### 2.3.1 Bacterial Cultures

#### *Escherichia coli*

Unless otherwise specified, *E. coli* cultures were grown at 37 °C using LB agar for solid cultures and LB media, shaking at 200 rpm for liquid cultures. Appropriate antibiotics were also added to the cultures at concentrations specified in Section 2.2.1.

#### *Streptomyces Species*

Unless otherwise specified, *Streptomyces* cultures were grown at 30 °C using SFM for solid cultures and 2xYT media, shaking at 200 rpm in baffled flasks for liquid cultures. Again, appropriate antibiotics were also added to the cultures at concentrations specified in Section 2.2.1.

### 2.3.2 Vector Creation

Much of the investigative work during this project was done *in vivo* using new strains created to contain different reporter systems and genes of interest. These new strains were achieved via the transfer of vectors containing genes of interest into a host strain.

In the case of *E. coli*, vectors were inserted via chemical or electro transformation (Sections 2.3.3 and 2.3.6). *Streptomyces* species on the other hand require DNA to be transferred from a non-methylating *E. coli* strain via intergeneric conjugation (Section 2.3.7). The vectors for *Streptomyces* were designed so that DNA integrated into the host genome rather than having an extra chromosomal plasmid, as is done in *E. coli*.

Plasmid vectors were created by placing inserts into an existing plasmid backbone via the homologous recombination of sticky ends created by the restriction digest of the vector and insert. The inserts were either a PCR product or a synthetically produced gene from the GeneArt service by Thermo Fisher.

### PCR Protocol

PCR was used to amplify the gene of interest and to add restriction sites to the 3' and 5' ends. Primer sequences can be found in Table 2.4 on page 27. The PCR reaction protocol used was as follows:

- 5  $\mu$ L 10X HF buffer
- 1  $\mu$ L 10 mM dNTPs
- 0.5  $\mu$ L 100  $\mu$ M forward primer
- 0.5  $\mu$ L 100  $\mu$ M reverse primer
- 0.5  $\mu$ L template
- 1.5  $\mu$ L DMSO
- 0.5  $\mu$ L NEB high fidelity Phusion<sup>®</sup> DNA polymerase
- Up to 50  $\mu$ L with water

**Table 2.9. PCR cycle times**

*These cycle times are for use with the NEB high fidelity DNA polymerase. Steps two, three and four are all repeated 35 times before moving onto step five, the final extension.*

| Stage                          | Temperature: | Length of time: |
|--------------------------------|--------------|-----------------|
| <b>1. Initial denaturation</b> | 95°C         | 2 minutes       |
| <b>2. Denaturation</b>         | 95°C         | 1 minute        |
| <b>3. Annealing</b>            | 45-72°C      | 1 minute        |
| <b>4. Elongation</b>           | 72°C         | 1 minute        |
| <b>5. Final extension</b>      | 72°C         | 15 minutes      |
| <b>6. Hold</b>                 | 4°C          | -               |

If this reaction did not produce a PCR product then the protocol was optimised by increasing the volume of DMSO to 2.5  $\mu$ L, varying the annealing temperature or diluting the template 1:3, 1:5 and 1:10 until a product was successfully obtained.

Unless otherwise specified, this PCR protocol was used throughout the rest of the project.

### Gel Electrophoresis

The size of PCR products as well as restriction digests and other DNA products were checked compared to a standard DNA ladder on a 1% agarose gel. This was made using the stock solutions specified in Section 2.2.2 using a 1% agarose gel in 1X TAE (40 mM Tris-acetate; 1 mM EDTA). As a standard, gels were run at 100 V for 40 minutes before being observed under UV light.

## Restriction Digest and Gel Extraction of DNA

Two restriction endonucleases were selected that each cut once each within a chosen area in the backbone vector (either pCC4, pOSV556 or L1). These same restriction sites were also added to each end of the gene inserts allowing the ligation of the two parts of the vector. Details on the restriction enzymes used can be found in the table below.

**Table 2.10. Restriction sites used to create vectors for this investigation**

| <i>Vector name</i> | <i>Genes contained</i>  | <i>Restriction sites used</i>  | <i>Insert type</i>                                     |
|--------------------|---|--------------------------------|--|
| <b>pKMS01</b>      | pCC4 with <i>mmfR</i>   | <i>HindIII</i> and <i>NotI</i> | PCR product  |
| <b>pKMS03</b>      | pCC4 with <i>myyR</i>   | <i>HindIII</i> and <i>StuI</i> | PCR product  |
| <b>pKMS85</b>      | pOSV556 with Y85F <i>mmfR</i> mutant                                | <i>HindIII</i> and <i>NotI</i> | PCR product (created using Agilent QuikChange SDM kit) |
| <b>pKMS144</b>     | pOSV556 with Y144F <i>mmfR</i> mutant                               | <i>HindIII</i> and <i>NotI</i> | PCR product (created using Agilent QuikChange SDM kit) |
| <b>L1F</b>         | <i>luxCDABE</i> , reversed intergenic region for <i>mmfL/mmjR</i>   | <i>EcoRV</i> and <i>BamHI</i>  | PCR product  |
| <b>L2</b>          | <i>luxCDABE</i> , intergenic region for <i>mmjP/myyR</i>            | <i>EcoRV</i> and <i>BamHI</i>  | PCR product  |
| <b>L3</b>          | <i>luxCDABE</i> , intergenic region for <i>myyY/myyB</i>            | <i>EcoRV</i> and <i>BamHI</i>  | PCR product  |
| <b>L3F</b>         | <i>luxCDABE</i> , reversed intergenic region for <i>myyY/myyB</i>   | <i>EcoRV</i> and <i>BamHI</i>  | PCR product  |
| <b>L4</b>          | <i>luxCDABE</i> and intergenic region for <i>ermEp*</i>             | <i>EcoRV</i> and <i>KpnI</i>   | PCR product  |
| <b>pKMS05</b>      | L1 vector with <i>Streptococcus mutans gbnB</i> and <i>luxCDABE</i> | <i>BamHI</i> and <i>NotI</i>   | Synthetic gene   |
| <b>pKMS06</b>      | L1 vector with <i>Salmonella enterica gbnB</i> and <i>luxCDABE</i>  | <i>BamHI</i> and <i>NotI</i>   | Synthetic gene   |

The concentration of DNA available for digestion was calculated using a NanoDrop and digestions carried out as specified in the protocols provided with the enzymes used. Where possible, NEB high fidelity enzymes were used. These enzymes have been optimised to work in the universal CutSmart buffer, allowing double digestion reactions and thereby increasing the yield of DNA that could be purified from a gel extraction as well as saving time.

### Example Protocol for *EcoRV* and *BamHI*

- 2.5  $\mu$ L 10x cut smart buffer
- 5 units NEB high fidelity restriction enzyme
- 0.5  $\mu$ g DNA
- Up to 25  $\mu$ L with water
- Incubate at 37 °C for an hour

### Gel Extraction of DNA

After digestion, the digestion products were then separated via gel electrophoresis on a 1% agarose gel as just described. The appropriate bands were then cut out and DNA was extracted using the Thermo Scientific GeneJET Gel Extraction Kit. The protocol provided with the kit was followed exactly until the final elution step, where 50  $\mu\text{L}$  of 65 °C water was used instead of the elution buffer and samples were incubated for 10 minutes at room temperature before a final centrifugation was carried out at the highest speed for two minutes to collect the purified DNA.

### Ligation

Ligation of the digested insert to the digested vector was carried out using a T4 ligase and 50% PEG according to the protocol provided by Thermo Scientific.(114) PEG was used as some of the restriction enzymes used e.g. *EcoRV*, were blunt cutters. The polyethylene glycol helps to increase the ligation efficiency of blunt ended DNA. The vector was trialled at a ratio of molecular weights of 1:2, 1:3 or 1:5 compared to the insert for each of the reactions to increase the chances of successful ligation.

Reaction mix:

- 45 ng vector
- X ng insert
- 2  $\mu\text{L}$  10x ligation buffer
- 2  $\mu\text{L}$  PEG (50%)
- 1  $\mu\text{L}$  T4 DNA ligase (5 units)
- Up to 20  $\mu\text{L}$  with water

After incubation at room temperature for an hour and then in the fridge overnight, 5  $\mu\text{L}$  of the ligation products were used to transform 50  $\mu\text{L}$  chemically competent TOP10 cells or 1-2  $\mu\text{L}$  to transform electro-competent ET12567 cells.

## 2.3.3 Chemical Transformation into TOP10 Cells

### Preparation of Chemically Competent TOP10 Cells

For the creation of chemically competent TOP10 cells for transformation, 10 mL sterile LB was inoculated from a single TOP10 colony and grown overnight at 37 °C shaking. In the morning this starter culture was used to inoculate 500 mL sterile LB and was again grown at 37 °C shaking, until the  $\text{OD}_{600}$  was between 0.35 and 0.40 (around three hours). The cells were then immediately placed on ice, transferring the culture to 50 mL falcon tubes. All steps from here on were kept at 4 °C and cell pellets were re-suspended as gently as possible. The cultures were allowed to chill for 20-30 minutes before centrifuging at 4000 rpm for 15



minutes at 4 °C. The supernatant was then decanted and the pellet re-suspended in ice cold sterile 100 mM MgCl<sub>2</sub>, mixing gently. (Cells could be combined in fewer tubes once re-suspended to reduce workload). Samples were then centrifuged at 3000 rpm for 15 minutes at 4 °C and the supernatant again decanted before ice cold sterile 100 mM CaCl<sub>2</sub> was added. This was then chilled on ice for 20 minutes before centrifuging at 3000 rpm and decanting the supernatant as before. The pellet was then re-suspended in ice cold sterile 85 mM CaCl<sub>2</sub> with 15% glycerol before centrifuging at 2100 rpm for 15 minutes at 4 °C. After pouring away the supernatant, the cells were re-suspended in a total of 1 mL 85 mM CaCl<sub>2</sub> with 15% glycerol and 50 µl aliquotted into pre-chilled cryovials. This was then either used immediately or flash frozen in dry ice and stored at -80 °C.

### **Transformation Protocol**

Transformation into competent *E. coli* TOP10 cells was done following a protocol from the ‘Invitrogen – Champion™ pET Directional TOPO® expression kits’ manual (page 20 from the One Shot® TOP10 Chemical Transformation Protocol).(110) Two changes were made to this protocol, firstly 2 µl of vector (instead of 3 µl) was added to 25 µl of competent *E.coli* DH5α (step 1) and LB (lysogeny broth) medium was used instead of S.O.C. medium (step 5). Transformed cultures were grown over night at 37 °C before single colonies were picked and stocks made with 50% glycerol for long-term storage at -80 °C.

### **2.3.4 Plasmid Purification**

#### **GeneJET™ Plasmid Miniprep kit**

Extraction and purification of cloned plasmids was needed to screen for the correct insert and successful vector synthesis. This was done using a GeneJET™ Plasmid Miniprep kit and associated protocol.(113) Changes to this protocol are as follows; 100 µl water at 70 °C was used instead of the elution buffer in the final step followed by incubation at room temperature for 15 minutes before the sample was centrifuged for two minutes to collect the purified DNA.

#### **Phenol Chloroform Purification of Cosmid DNA**

If plasmid DNA was not successfully recovered using the GeneJET™ Plasmid Miniprep kit, particularly a problem for larger pieces of DNA, then vectors could alternatively be extracted using a phenol chloroform protocol (see Section 2.2.3 for details on buffers and solutions). *E. coli* containing the plasmid of interest were grown at 37 °C overnight in a 10 mL culture with appropriate antibiotics. These cells were then pelleted by centrifugation at 2000 rpm for 15 minutes. The supernatant was poured away and the cell pellet re-suspended in 100 µL of solution I (50 mM Tris-HCl, pH 8 and 10 mM EDTA) before being transferred into a microcentrifuge tube.

A volume of 200  $\mu\text{L}$  of solution II (200 mM NaOH and 1% SDS) was then added to the re-suspended cells and the tubes inverted ten times. Next, 150  $\mu\text{L}$  of solution III (3 M potassium acetate, pH 5.5) was added and the tubes inverted ten times to mix the solutions. Samples were then centrifuged at full speed in a microcentrifuge for five minutes to pellet the cell matter. Immediately, 400  $\mu\text{L}$  phenol-chloroform was added to the extracted supernatant and samples vortexed for two minutes.

The samples were then centrifuged at full speed for five minutes to separate the mixture into two phases, the vector DNA should be in the upper phase. The upper phase was then transferred into a fresh microcentrifuge tube and 600  $\mu\text{L}$  ice cold 2-propanol was added. This were then left on ice for 15 minutes before samples were spun at full speed for five minutes to pellet the vector DNA. The supernatant was then removed and the pellet washed with 200  $\mu\text{L}$  ice cold ethanol before centrifuging again as before. After the supernatant was removed, the tube was left open and the pelleted DNA was left to dry at room temperature for ten minutes before the pellet was re-suspended in 50  $\mu\text{L}$  water (or 10 mM Tris-HCl, pH 8 for longer term storage).

### 2.3.5 Vector Screening and Sequencing

Newly synthesised vectors were screened both using a PCR reaction as well as a restriction digest compared to a control reaction. Sequencing was also carried out on the clones that showed most promise in the screening, as a final check for the correct product.

#### PCR Screening

For the PCR screen, primers were designed to bind either side of inserts, ideally producing products of different sizes if the gene of interest was present or absent (alternatively primers were designed so one bound within the insert and one outside). The screening PCR reaction was done following the protocol from the NEB HF Phusion<sup>®</sup> DNA polymerase described previously unless there were a large number of samples in which case a standard Taq polymerase was used, as it is less costly. The protocol for the Taq polymerase follows.

Reaction mix:

- 10  $\mu\text{L}$  5X Taq buffer ( $\text{NH}_4\text{SO}_4$ )
- 2  $\mu\text{L}$  10 mM dNTPs
- 1  $\mu\text{L}$  100  $\mu\text{M}$  forward primer
- 1  $\mu\text{L}$  100  $\mu\text{M}$  reverse primer
- 0.5  $\mu\text{L}$  template
- 1.5  $\mu\text{L}$  DMSO
- 0.5  $\mu\text{L}$  Taq DNA polymerase
- Up to 50  $\mu\text{L}$  with water

**Table 2.11. PCR cycle times**

*These cycles times are for use with the Taq polymerase. Steps two, three and four are all repeated 30 times before moving onto step five, the final extension.*

| <b>Stage</b>                   | <b>Temperature:</b> | <b>Length of time:</b> |
|--------------------------------|---------------------|------------------------|
| <b>1. Initial denaturation</b> | 95°C                | 2 minutes              |
| <b>2. Denaturation</b>         | 95°C                | 45 seconds             |
| <b>3. Annealing</b>            | 45-72°C             | 45 seconds             |
| <b>4. Elongation</b>           | 72°C                | 1 minute               |
| <b>5. Final extension</b>      | 72°C                | 15 minutes             |
| <b>6. Hold</b>                 | 4°C                 | -                      |

Again, if this PCR reaction did not produce a product then the protocol was optimised by increasing the volume of DMSO to 2.5 µL, varying the annealing temperature or diluting the template 1:3, 1:5 and 1:10 until a product for the control reaction was successfully obtained.

### **Restriction Digest Screening**

A restriction digest was also run to check for expected product sizes from a successfully created vector. Restriction endonucleases were selected to give different product sizes for a control compared to the desired ligated vector and the products of these reactions analysed by gel electrophoresis. The digestion protocol described previously in Section 2.3.2 was again used, this time scaled down to be done with only around 10 µL purified DNA and no subsequent gel extraction.

### **Sequencing**

Sequencing was done using the GATC LIGHTrun™ sequencing service, usually using the primers that had been used for the PCR screening step. Sequencing results could then be analysed for correct insertion orientation as well as mistakes that may have occurred during the PCR synthesis of the insert.

### **2.3.6 Electro-Transformation of Vectors into ET12567/pUZ8002**

TOP10 cells were used to clone plasmids and for the creation of plasmid stocks, much of the work in this project however was done *in vivo* in *S. coelicolor*. *S. coelicolor* does not accept methylated DNA and so could not be directly transformed with plasmid vectors. Instead, the purified vectors were first transformed into a non-methylating strain of *Escherichia coli*; ET12567/pUZ8002. The transformants were then used to transfer the DNA vectors to *Streptomyces coelicolor* via intergeneric conjugation.

### **Preparation of Electro-Competent ET12567/pUZ8002**

ET12567 cells with pUZ8002 were streaked out on an LB plate with kanamycin and chloramphenicol and grown overnight at 37 °C to produce single colonies. One of these

colonies was then picked and grown in 10 mL LB media with kanamycin and chloramphenicol overnight at 37 °C shaking. Of this starter culture, 200 µl was used to inoculate 10 mL fresh media with the same antibiotics. This was then grown at 37 °C until the OD<sub>600</sub> was between 0.4 and 0.6 (around four hours). This cell culture was then spun at 2000 rpm at 4 °C for 10 minutes and the supernatant removed before the cells were re-suspended in 10 ml of ice cold 10% glycerol. Cells were then spun down again (keeping them at 4 °C) and re-suspended in 5 mL of ice-cold 10% glycerol. After a final centrifugation and removal of the supernatant, cells were re-suspended in the residual liquid. These cells were now competent and either kept on ice and used immediately or stored at -80 °C for later use.

### **Electroporation Technique**

Taking care to keep the whole reaction on ice, 80 µl electro-competent ET12567 cells and 2 µl vector were added to a 2 mm electroporation cuvette. The cells were then electroporated at 2.5 kV before 1 mL ice-cold LB was immediately added. The transformants were then transferred into a microcentrifuge tube and left for an hour at 37 °C shaking. The culture was then plated out on LB agar with kanamycin and chloramphenicol to select for the demethylating pUZ8002 vector and an appropriate antibiotic to select for the vector being transformed. This was then grown overnight at 37 °C. Single colonies were picked and stocks made with 50% glycerol for long-term storage at -80 °C or immediate use in the intergeneric conjugation protocol.

#### **2.3.7 Intergeneric Conjugation to Introduce Vectors into *Streptomyces coelicolor* M145**

Unless otherwise noted, *Streptomyces coelicolor* M145 was used as the host strain for this assay. The transfer of vectors into this strain was carried out using the protocol specified in 'Practical *Streptomyces* Genetics'.(51)

Single colonies of ET12567 cells with pUZ8002 containing the relevant vector were picked and grown overnight at 37 °C shaking in LB with the appropriate antibiotics. The next morning 200 µl of this starter culture was used to inoculate 10 mL fresh media (with the same antibiotics) and this was grown at 37 °C shaking until the OD<sub>600</sub> was between 0.4 and 0.6 (around four hours). This was then centrifuged for ten minutes at 2000 rpm to pellet the cells. The pellet was then re-suspended in 10 mL LB and centrifuged again before the washing step was repeated to remove any remaining antibiotics. The cell pellet was then re-suspended in the residual LB to give a total volume of 1 mL.

A volume of 10 µl *Streptomyces* spore stock was added to 500 µl 2xYT media and the cells heat-shocked at 50 °C for ten minutes before being mixed with 500 µl of the prepared

ET12567 cells. This mixture was then serially diluted and the two strains were grown overnight together on SFM media on four different plates containing dilutions of between  $10^{-1}$  and  $10^{-4}$ . The next morning the plates were overlaid with nalidixic acid to kill the *E. coli* and apramycin or hygromycin to select for *Streptomyces* colonies contain the luciferase constructs or pCC4 vectors. This was then left to grow for three to four days, when single colonies could be collected and used to inoculate fresh plates.

Intergeneric conjugation into *S. albus* required a lower heat shock temperature of 40 °C compared to the 50 °C specified in Practical Streptomyces genetics (51) and used for *S. coelicolor*. Heat shocked cells *S. albus* were then kept at 30 °C for four hours before being combined with ET12567 strains containing pUZ8002.

### Spore Stock Production

To produce spore stocks of *Streptomyces*, four or five SFM plates with appropriate antibiotics were inoculated with a lawn of *Streptomyces*. After incubating at 30 °C for five days, around 3 mL sterile water was added to each plate. A sterile wedge shaped spreader was used to free the spores and create a suspension, which was then collected. The suspension was filtered through sterile non-adsorbent cotton wool to remove any agarose that had been picked up. The suspensions were then centrifuged at 4000 rpm for five minutes to pellet the spores. The supernatant was removed and the cells re-suspended in approximately 1 mL of sterile 50% glycerol before being stored at -80 °C.

### Genomic Extraction of Genomic DNA from *Streptomyces*

Genomic DNA extraction from *Streptomyces* was performed using the FastDNA Spin Kit for Soil from MP Biomedicals.(111) The protocol was followed exactly according to the kit.

Genomic DNA from *Streptomyces* could then be screened using the same protocols as were just described for *E. coli* in Section 2.3.5 to check for the insertion of the desired genetic material.

### 2.3.8 Measurement of Bioluminescence

Luminescence was measured using a Photek - CCD (charge coupled device) camera. Photons hit the silicon surface of the CCD chip and an electron is liberated. This then creates an electron deficient site or 'hole'. These charges are then stored until a voltage is applied and the collected charges can be shifted along, measured and converted into a digital copy of the light patterns. This provides both quantitative results of the amount of luminescence produced as well as a colour code image representing the amounts of luminescence produced by

different samples. The CCD used in this project is a HRPCS (high resolution photon counting system) and is so sensitive that it can measure single photons.

The protocol for preparing cells for measuring bioluminescence was developed as part of the experimental investigation and can be found in Chapter 3.

### 2.3.9 Site Directed Mutagenesis

Mutants were created using the Agilent QuikChange Lightning kit.(109) The protocol was followed as specified in the manual and primers were designed using the online Agilent facility.(127) Screening was carried out with standard T7 primers.

As a template for the mutagenesis the vector pET151:*mmfR* was used. The small size of this vector made the PCR aspect of the Agilent protocol easier to carry out than the pKMS01 vector also trialled. Mutant *mmfR* sequences were then cloned via PCR and inserted into the pCC4 vector. This was done according to the restriction digest sub-cloning protocol described in Section 2.3.2.

### 2.3.10 Bacterial Cultures for Protein Purification

#### Liquid Culture

Unless otherwise stated, liquid cultures of *S. coelicolor* strains containing the *gbnB* analogue were grown in baffled flasks with 2xYT media containing appropriate antibiotics at 30 °C in a shaking incubator at 200 rpm. Unless otherwise specified, these cultures were grown for 72 hours before the secreted proteins were harvested.

#### Solid Culture

Strains were grown on SFM (soya flour mannitol) media at 30 °C with appropriate antibiotics, on top of a layer of sterile dialysis tubing to make collection of secreted proteins easier. Nutrients and other small molecules should be able to pass through the tubing but not any proteins secreted by the *Streptomyces*. The dialysis tubing was cut into appropriately sized pieces and autoclaved in a glass petri dish between layers of filter paper to stop it from sticking to itself.

In later stages of protocol optimisation, plates also contained a final concentration of 1 mM EDTA to inhibit metalloproteases and were overlaid with the Sigma Aldrich Protease Inhibitor Cocktail for Tissue Culture after 24 hours growth (final concentration 1:500). For further details on the development of this protocol see Section 7.4.

Unless otherwise specified, these cultures were grown for 72 hours before the secreted proteins were harvested.

### 2.3.11 Collection and Purification of Secreted Proteins from Culture Media

#### Harvesting Secreted Proteins

To harvest the secreted proteins from liquid media, cultures were spun down and the supernatant collected before being passed through a 0.22 µm filter to remove any residual cell mass. The supernatant was then concentrated using an Amicon Ultra 15 centrifugal filter unit with a molecular weight cut off of 10 kDa. Protein concentration was measured via a Bradford protein assay to establish when the samples had been concentrated enough.

To harvest secreted proteins from the solid media, a cell scraper was used to detach cells from the dialysis tubing. The cell mass was then re-suspended in one of the protein purification buffers specified in Section 2.2.4 before being centrifuged and passed through an 0.22 µm filter to remove the cell mass. The volume of buffer used for re-suspension was kept very low meaning that further concentration of samples was not needed.

Concentrated protein samples collected could then be run directly on an SDS-PAGE gel or carried forward for further purification using nickel Sepharose or were collected via precipitation with lithium chloride and trichloroacetic acid.

#### Nickel Sepharose Purification

The protein of interest was designed with an N-terminal polyhistidine-tag and therefore was purified via nickel Sepharose purification using GE Healthcare Ni Sepharose 6 Fast Flow. The Sepharose was washed according to the bench top protocol specified by GE healthcare<sup>1</sup> using the buffers specified in Section 2.2.4. (Initially the ‘protein purification buffer’ and later the corresponding ‘elution buffer’ were used and then after optimisation of the protocol, the ‘optimised’ buffers were used, see Chapter 7.)

After the washing of the Sepharose slurry however, the GE healthcare specified protocol was not followed for the remainder of the purification. The method being used for overexpression was novel and therefore required a lot of optimisation and produced huge numbers of samples to be processed. To save both time and resources a packing column was not used, instead, as with the initial Sepharose washing steps, the protocol was continued in microcentrifuge tubes with the supernatants being collected after elution steps, as is described now.

Secreted protein samples were added to the Sepharose slurry and incubated at 4 °C for an hour to allow binding of the protein to the nickel. This mix was then spun down in a microcentrifuge at the highest speed for three minutes and the supernatant collected and

---

<sup>1</sup> [https://www.gelifesciences.com/gehcls\\_images/GELS/Related%20Content/Files/1314742967685/litdoc11002497AB\\_20110831013915.pdf](https://www.gelifesciences.com/gehcls_images/GELS/Related%20Content/Files/1314742967685/litdoc11002497AB_20110831013915.pdf)

labelled as 'unbound proteins'. The pellet was then washed once with 1 mL of the 'protein purification buffer', centrifuged and the supernatant collected as 'unbound proteins wash'. This was followed by 500  $\mu$ L elution buffer (containing imidazole) being added to the pellet, after centrifugation the supernatant collected was labelled as 'elution'. The pellet was then washed twice with 'protein purification buffer' (containing no imidazole) and the supernatants collected were labelled as 'wash 1' and 'wash 2'. The unbound proteins and elution fractions were then concentrated using an Amicon ultra 0.5 centrifugal filter unit with a 10 kDa molecular weight cut off before analysis by SDS-PAGE.

This analysis was done as a proof of concept investigation into optimal conditions for protein over production in *Streptomyces* species. For a higher level of purity to achieve a single band on an SDS-PAGE gel and potentially obtain useable proteins, a more thorough purification protocol could be used, for example FPLC (fast protein liquid chromatography).

### **Precipitation of Proteins from Solid Culture**

Proteins secreted by *S. coelicolor* growing on solid media could be purified by precipitation with lithium chloride and trichloroacetic acid according to a protocol taken from the 2006 paper by Widdick *et al.*(128)

A lawn of the *Streptomyces* strain of interest was grown on an SFM plate on top of a layer of sterile dialysis tubing for 72 hours. A cell scraper was used to collect the biomass from these plates and this was then dispersed in around 3 mL 5 M lithium chloride solution and left on ice for thirty minutes. Samples were then vortexed for two minutes and centrifuged at the highest speed in a microcentrifuge for five minutes before being passed through a 0.22  $\mu$ m filter to remove the biomass. The secreted proteins should remain in the lithium chloride solution. Trichloroacetic acid was then added to the solution to a final concentration of 20% before the solution was again incubated on ice for thirty minutes before being centrifuged at the highest speed for 15 minutes. Two phases were formed, the proteins being in the lower phase. The upper phase was removed and water added to an equivalent volume. At this point the sample then turns cloudy as the proteins precipitate. The sample was then centrifuged for 15 minutes at the highest speed to pellet the precipitated proteins. The pellet was then washed two or three times with -20  $^{\circ}$ C acetone (taking care not to disturb the precipitate too much). Samples were then air dried before being re-suspended directly into 2x SDS-PAGE loading buffer and checked via SDS-PAGE.

### **Cell Lysis for Protein Analysis**

The expression system used in this investigation was designed so that the protein of interest should be secreted from *Streptomyces*. However, it was necessary to check inside the cells to



make sure that the export tag being used was indeed working. The *Streptomyces* cells were lysed using ‘lysis matrix E’, a component usually provided with the MP Biomedicals’ FastDNA spin kit. This matrix contains ceramic and silica spheres as well as a large glass bead to allow the mechanical shearing of cells.

The pelleted *Streptomyces* cell mass collected from the solid cultures grown on dialysis tubing were added to a 2 mL lysing matrix tube followed by 1 mL sodium phosphate buffer with SIGMAFAST™ protease inhibitor cocktail (as described in Section 2.2.4). After vortexing for two minutes the suspension was allowed to settle before the supernatant was collected. The supernatant was then added to the nickel Sepharose, as described before, to check for the presence of any intracellular histidine-tagged protein.

### 2.3.12 SDS-PAGE

All SDS-PAGE cells were run using a 12% resolving gel with a 4% stacking gel using the mixtures specified in Section 2.2.5. The 12% gel was allowed to set for 30 minutes before the 4% gel was added on top.

Collected protein samples were mixed 1:1 with 2X SDS-PAGE loading dye and boiled for five minutes before being allowed to cool. Between 10 and 20  $\mu$ L stained protein sample was the loaded per gel and run at 180 V for five minutes and then 200 V for 35-40 minutes before being stained for one hour with InstantBlue™ Coomassie® stain and then washed with distilled water.

### 2.3.13 SDS-PAGE Gel Extraction Protocol for Mass Spectrometry Analysis

Proteins were extracted from SDS-PAGE gels for LC-MS analysis over two days using a protocol supplied by Dr Cleidiane Zampronio from the Proteomics Facility at the University of Warwick.

First bands of interest were cut out of the SDS-PAGE gel using a clean razor blade and placed in a microcentrifuge tube before being cut into four or five smaller pieces to increase the surface area of the slice. These gel pieces were then washed using 150  $\mu$ L 50% ethanol in 50 mM ammonium bicarbonate and incubated at 55 °C for 20 minutes shaking. The supernatant was then pipetted off and the wash repeated until the gel was de-stained (usually one or two more times).

The gel pieces were then be dehydrated by adding 100  $\mu$ L pure ethanol and incubating for five minutes at 55 °C shaking. This produced a shrunken, white gel. The ethanol was then removed and disulphide bonds reduced by adding 100  $\mu$ L 10 mM DTT in 50 mM ammonium bicarbonate and incubating for the samples 30 minutes at 56 °C shaking. The free liquid was

then removed and samples allowed to cool before 100  $\mu$ L 55 mM iodoacetamide (IAA) in 50 mM ammonium bicarbonate was added. The IAA alkylates the cysteine residues in the protein. After the IAA was added samples were incubated for 20 minutes at room temperature while being kept in the dark. The free liquid was then removed and disposed of.

The wash steps with 50% ethanol in 50 mM ammonium bicarbonate were then repeated, followed by dehydration in pure ethanol. A tryptic digest of samples was then carried out by adding 40  $\mu$ L 2.5 ng/ $\mu$ L trypsin in 50 mM ammonium bicarbonate. After allowing the gel to rehydrate for ten minutes, and additional 15  $\mu$ L 50 mM ammonium bicarbonate was added to make sure the gel was well covered and samples incubated overnight at 37 °C with shaking.

The next morning the tryptic digest was stopped by adding 100  $\mu$ L 5% formic acid in 25% acetonitrile and sonicating samples for five to ten minutes. The supernatant was then collected and this formic acid-acetonitrile step repeated three more times, each time collecting the supernatant in the same microcentrifuge tube. The combined peptide supernatants were then dried in a Speed-Vac at 40 °C for four hours before freezing, ready for LC-MS analysis.

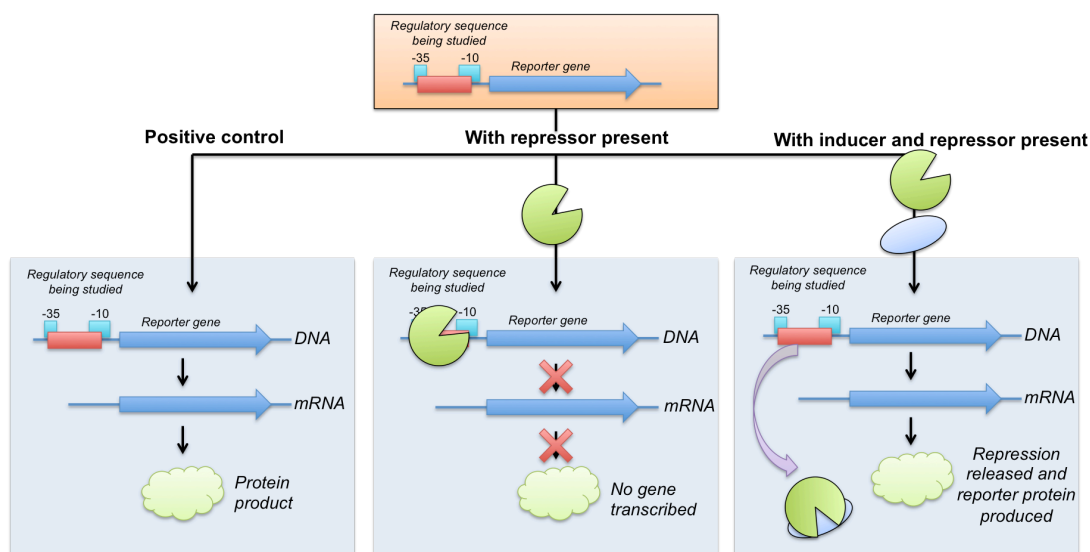
The actual LC-MS analysis was then carried out by the Proteomics Facility, according to their standard protocols.

## 3 Development of Luciferase Reporter Gene Constructs

### 3.1 Aims and Strategy of Investigation

#### 3.1.1 Reporter Genes

Figure 3.1 shows a schematic of a general reporter gene system that can be used *in vivo* to study transcriptional regulation. Here, a regulatory sequence of interest is put upstream of a reporter gene instead of its usual cognate gene. This regulatory sequence can contain -35 and -10 promoter sequences as well as operators for transcription factors. The reporter gene is designed to produce a protein that results in a measurable result, for example a coloured product or a fluorescent/bioluminescent product. The amount of reporter produced should be proportional to promoter strength and be regulated by any control mechanisms present that normally influence the regulatory sequence in the wild type system. Thus, the production of the reporter protein should be representative of the expression of the wild type gene cognate to the regulatory sequence. Repressor proteins as well as their ligand inducers can also be added to the system and changes in the production of the reporter protein observed.



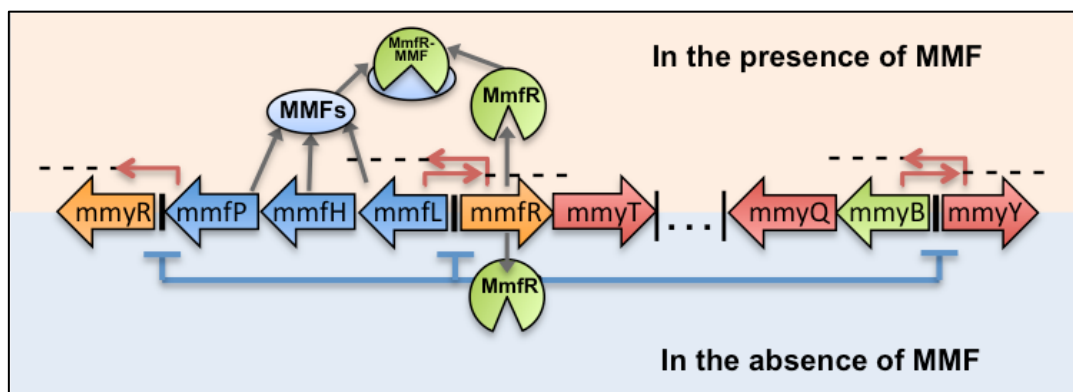
**Figure 3.1. Schematic of reporter gene systems and how they can be used to study transcriptional repressors and their cognate inducer ligands**

With the eventual aim of developing a novel inducible expression system, this investigation utilised a *Photobacterium luminescens* luciferase reporter gene assay to produce a measurable bioluminescent product under the control of the methylenomycin BGC regulatory system. This *luxCDABE* system was recently optimised to be expressed in GC high streptomycetes by Justin Nodwell and this research group.(92)

### 3.1.2 Regulatory System of Interest – The Methylenomycin Biosynthetic Gene Cluster

As discussed in the introduction, methylenomycin biosynthesis is tightly regulated by a number of different mechanisms including the use of a rare TTA codon, transcriptional repressors, transcriptional activators as well as the use of microbial hormones.(71)

The entire methylenomycin cluster is found on a 19 kb region the *S. coelicolor* SCP1 plasmid.(65) The arrangement of the five operons in this cluster and how they are thought to be regulated by transcriptional repressor MmfR and its cognate ligands, the MMFs, is shown in Figure 3.2. In this figure, operons and the direction of transcription is indicated by the red arrows and intergenic regions with MmfR-binding MARE operator sites are signified by vertical black lines. These are found between *mmfR* and *mmfL*, *mmyR* and *mmfP*, and *mmyY* and *mmyB*. This figure shows the influence of the presence and absence of the methylenomycin furans on the auto-regulation of this cluster as well as the influence over expression of other biosynthetic genes.(71) In the absence of MMF microbial hormones (shown in the blue section at the bottom of the figure), MmfR binds to DNA at the MARE operator, repressing transcription. Upon MMF binding (in the peach section at the top) there is a conformational change to MmfR, it is released from the DNA and transcription begins at the five operons. Replacing any one of the five operons with a reporter gene would allow the study of the regulation of that particular operon. (Please note that the diagram in Figure 3.2 is not to scale.)



**Figure 3.2. Proposed regulation of the methylenomycin biosynthetic gene cluster from the SCP1 plasmid of *S. coelicolor* by the TetR family member, MmfR and cognate MMF ligands**

*Horizontal blue lines indicate the location of proposed MmfR binding operator sites with the vertical black lines represent where the MARE operators can be found. Red arrows indicate the operons regulated by these operators and the direction of transcription. Proposed functional attributions are **mmfR and mmyR** – transcriptional repressors, **mmfLHP** – methylenomycin furan biosynthetic genes, **mmyT and mmyQ** – methylenomycin biosynthetic genes, **mmyB** – transcriptional activator, based on the work of Chater and Bruton (65)*

The biological function of the MmfR paralogue, MmyR is much less clear than that of MmfR and it is not known whether it will bind to any of the MARE operator sites or the MMFs. Although there is limited evidence that it would be a suitable choice for a repressor in an inducible expression system, understanding the role of this second repressor is also important in further interpreting the regulation of the methylenomycin biosynthetic cluster. For this reason, MmyR was also under investigation in this research, secondary to the work done with MmfR.

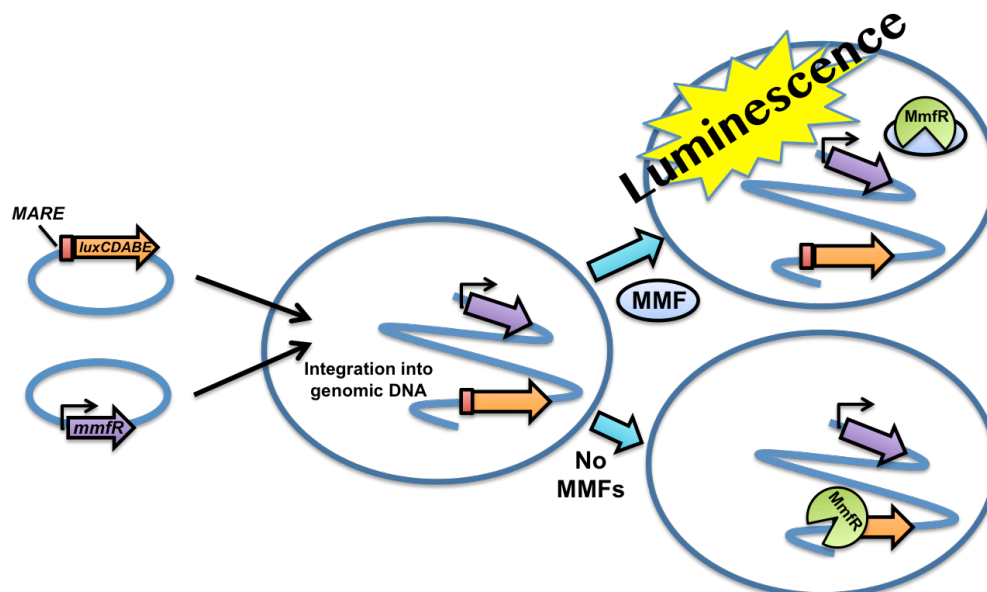
The development of a bio-luminescing reporter gene assay involved the design of two separate vectors in a strategy that will now be described. One of these vectors contained the *luxCDABE* operon under the regulation of one of the three MARE operator containing intergenic regulatory sequences (indicated by vertical black lines in Figure 3.2). A second vector contained *mmfR* (or *mmyR*) under the regulation of *ermEp\**.

### Strategy for Luciferase Reporter Gene Assay

Figure 3.3 shows a schematic of how the two vector types just described should come together to form the reporter assay in *S. coelicolor* M145 for studying MmfR. Both vectors will integrate in different positions on the *S. coelicolor* genome and are replicated along with the rest of the genome as the cell replicates.

The expression of *luxCDABE* is under the control of one of the promoters from the methylenomycin gene cluster (see Figure 3.6 for a diagram showing the positions of these promoters) as well as predicted regulation by MmfR repression at the MARE operator.

As shown in Figure 3.3, in the absence of the MMFs, MmfR are predicted to be produced and bind on to the MARE operator, thereby repressing the expression of the *lux* genes. If one of the five MMFs are then added to the system they are expected to bind to MmfR and cause a conformational change, releasing it from the MARE operator and allowing the expression of the *lux* genes. Depending on the strength of binding of MmfR to the different MARE operators and level of release by the MMFs, different levels of luminescence could be expected. Data collected can then be compared to positive and negative controls and inferences made (see Figure 3.1).



**Figure 3.3. Schematic of *lux* reporter system in *Streptomyces coelicolor* M145**

Two vectors are integrated into the *Streptomyces* genome; one containing *mmfR* (or an empty *pCC4* control vector) under the control of the *ermEp\** promoter and the other containing *luxCDABE* under the control of one of five relevant methylenomycin cluster promoters and one of three MARE operator sequences. In the absence of the furan ligands, *MmfR* will bind to the MARE operator and repress the expression of *luxCDABE*. Upon the addition of one of the methylenomycin furans (MMFs) there will be a conformational change in *MmfR* and it will be released from the MARE operator resulting in the production of luminescence via *luxCDABE* expression, producing luciferase and its substrate.

For the investigation of *MmyR*, a similar system to that shown in Figure 3.3 was used, only with pKMS03 (containing *mmyR*) used instead of pKMS01 (containing *mmfR*) (see Table 2.5). The outcome of the presence of different MARE operators and MMFs in this *mmyR* system was not known before studies began.

### Details on *luxCDABE*

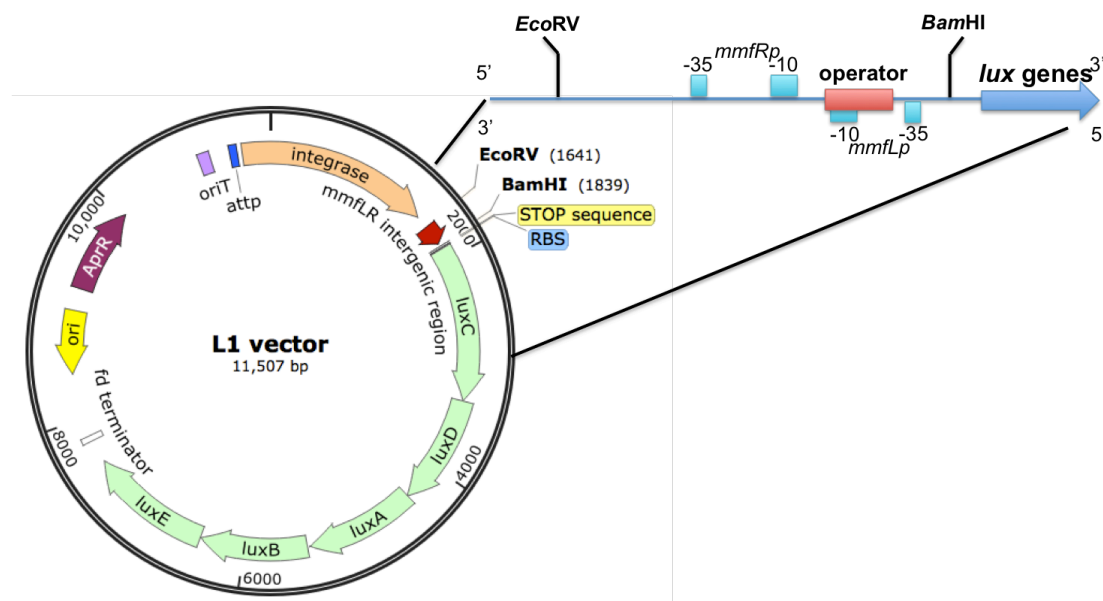
LuxA and LuxB together form the heterodimeric luciferase protein. LuxC, D and E are the enzymes needed to form tetradecanal, the fatty aldehyde substrate of luciferase.(97) All other biochemicals required by this reaction are found naturally in bacterial cells allowing a self generated bioluminescent response with no need for external manipulations.(98) Expression of the pentacistronic *luxCDABE* operon results in a blue-green light being emitted at 490 nm.(129)

## 3.2 Plasmid Design and Assembly

### 3.2.1 Luciferase Vectors

Figure 3.4 shows a diagram of a vector with a MARE operator containing intergenic sequence (Figure 3.2) inserted upstream of a 5668 bp optimised *luxCDABE* luciferase reporter gene

cluster.(92) The particular vector shown in Figure 3.4 is labelled as ‘L1’ and has the *lux* genes under the control of *mmfLp*, this was constructed by Professor Nodwell’s group in Canada and is an updated version of the pMU1 plasmid.(92) It contains the entire intergenic region from between *mmfR* and *mmfL* (labelled in the diagram as the ‘*mmfLR* regulatory sequence’). This plasmid formed the backbone for a number of other vectors produced during this project, with different intergenic regions being inserted next to the *lux* operon using the *EcoRV* and *BamHI* restriction sites shown in the figure.



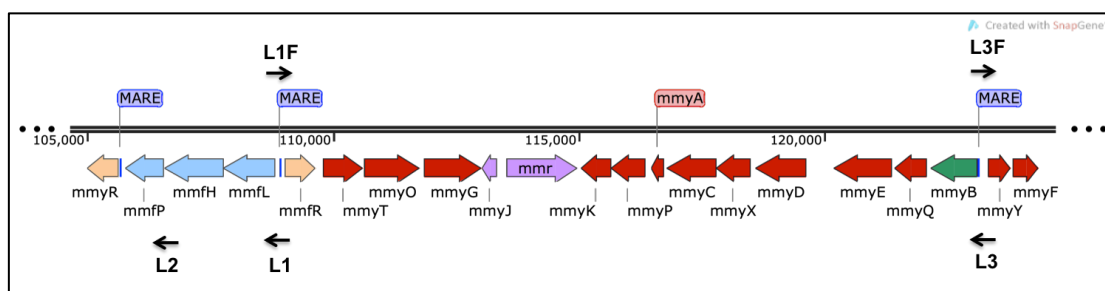
**Figure 3.4. L1, an integrating *luxCDABE* reporter plasmid**

*tfd* – transcriptional terminator from phage *fd*, *ori* – origin of replication, *AprR* – apramycin resistance cassette, *oriT* – origin of transfer, *int* and *attP* – integrase and attachment site of  $\Phi$ BT1 phage, allows integration via site specific recombination to matching *attB* site in the *S. coelicolor* genome, ***mmfLR regulatory sequence*** – intergenic region from between *mmfR* and *mmfL*, containing the ‘L1 MARE operator’ and in the direction of *mmfLp*, ***luxCDABE*** – genes for luciferase and the biosynthesis of its substrate.(92) Upstream of the *mmfLR* regulatory sequence is a **RBS** as well as a **STOP sequence** containing stop codons covering all three reading frames.

Please note that the close up version of the insert between *EcoRV* and *BamHI* restriction sites is not to scale and is purely a representation of the layout of the vector.

This vector was created by Justin Nodwell et al., optimised from work in (92).

Dependent on the orientation of the three intergenic regions inserted between the *EcoRV* and *BamHI* restriction sites, the regulation of the five different operons (shown in Figure 3.2) could be studied. This created a total of five different *lux* plasmids, labelled as L1, L1F, L2, L3 and L3F. A schematic showing the designation of this nomenclature can be found in Figure 3.5. This nomenclature is now used throughout the rest of the project.



**Figure 3.5. Assigned nomenclature for *lux* vectors containing different intergenic regions from the methylenomycin gene cluster, cloned upstream of *luxCDABE***

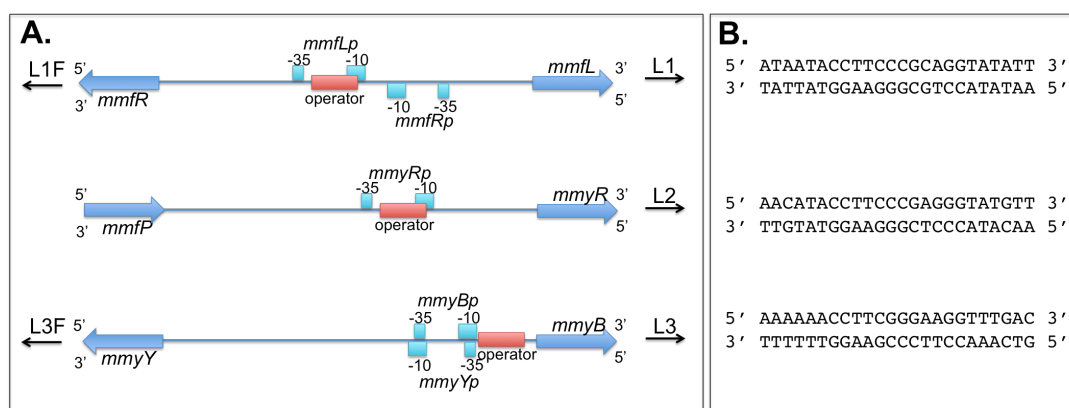
*Vertical blue lines indicate the location of proposed MmFR binding operator sites and black arrows denote the direction of transcription for each promoter. ‘F’ indicates the reverse sequence of a particular intergenic region. This diagram is to scale and based on individual gene sizes.*

*Proposed functional attributions are **mmfR and mmyR** – transcriptional repressors, **mmfLHP** – methylenomycin furan biosynthetic genes, **mmyT and mmyQ** – methylenomycin biosynthetic genes, **mmyB** – transcriptional activator, based on the work of Chater and Bruton (65)*

In this nomenclature, ‘F’ indicates the same intergenic region that has been inverted and inserted back into the vector. For example, the intergenic region containing within L1F is the exact reverse sequence of that contained in the L1 vector. MmFR is homodimeric and so affinity in either orientation would be expected to be unchanged as the 24 bp MARE operator is the same in either orientation (see Figure 3.6B for sequences).(84) However, the orientation of the insert may result in promoters of different strengths being used. Figure 3.6A shows the predicted positions of the -35/-10 promoter sequences found in each methylenomycin cluster intergenic region. In one orientation, only one of the promoters should drive the expression of the *lux* genes.

No L2F was created as there is only one promoter to study in the L2 intergenic region and so a reverse was not necessary for the study of promoter strength. If extra time had allowed however, a L2F vector could have been constructed and used as an extra negative control.





**Figure 3.6A and B. Details on the methylenomycin cluster MARE operators**

**A. Schematic of predicted MARE operator position with respect to the promoters they regulate in the intergenic regions of the methylenomycin biosynthetic cluster**

*MmfR* is predicted to bind three different MARE operators, thereby regulating five different operons by repressing five promoters.

Please note that this diagram is not to scale.

**B. Predicted MARE operator sequences**

The *MmfR*-binding MARE operators are partially palindromic 24 bp sequences and display a motif found across a number of streptomycetes. The sequences shown here correspond to the operators in the adjacent schematics in A.

Adapted from the work of Vincent Poon (84)

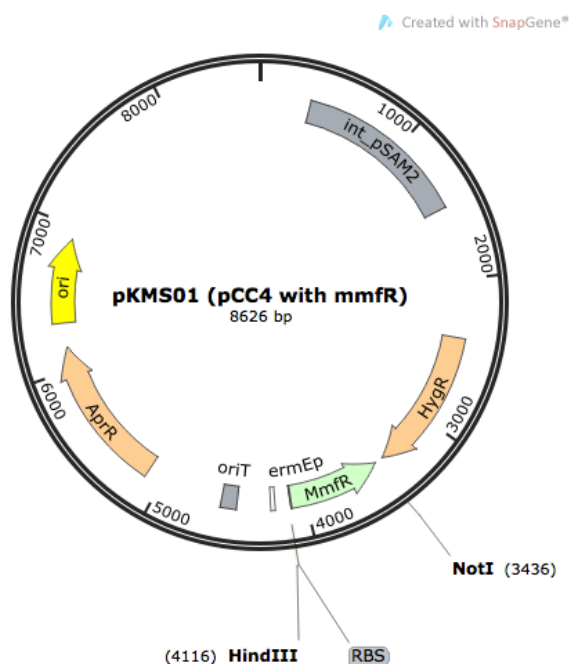
The luciferase vectors were transferred into *S. coelicolor* via intergeneric conjugation (Section 2.3.7) where upon they stably integrate into a single place the *Streptomyces coelicolor* genome via the  $\phi$ BT1 phage system. This system is based on how bacteriophages integrate into specific sites of the host genomes via site specific recombination using an integrase. The phage contains an *attP* site (Figure 3.4,) that matches an '*attB*' site in the target host where it will integrate and establish a lysogenic state.(130)

In the absence of *MmfR*, these *lux* vectors should work as a positive control. Constitutively luminescing, proportional to the strength of the promoter being analysed in that particular vector. A schematic of this type of positive control can be found in Figure 3.1.

*MmfR* and *MmyR* were added to the reporter system on a different vector, described next.

### 3.2.2 Vectors Containing *mmfR* or *mmyR*

The second type of vector used in this assay was based on pCC4 plasmids, containing the gene for transcriptional repressor MmfR or its paralogue MmyR. A schematic of the *mmfR* vector (pKMS01) can be found in Figure 3.7, showing the *Hind*III and *Not*I restriction sites used to insert this repressor gene into the pCC4 vector. The *mmyR* vector (pKMS03) was constructed in an analogous way, only with *mmyR* inserted instead of *mmfR* and using *Hind*III and *Stu*I restriction sites instead of *Hind*III and *Not*I (see Table 2.10).



**Figure 3.7. pKMS01, an integrating plasmid with *mmfR* controlled by *ermEp*\* (72)**  
*ori* – origin of replication, *int\_pSAM2* – integrase site, *hygR* – hygromycin resistance, *MmfR* – TetR family transcriptional repressor gene, *ermEp*\* – strong constitutive promoter, *oriT* – origin of transfer  
*AprR* – apramycin resistance.

It was decided that MmfR production should be under the control of *ermEp*\*,<sup>(131)</sup> a strong constitutive promoter unrelated to the methylenomycin cluster, rather than its own promoter. This should mean that MmfR will be produced at a constant rate, unaffected by its own self-regulatory action (see Figure 3.2). The alternative, self-regulation of MmfR, would mean that under the control of *mmfRp* and the MARE operators MmfR would repress its own production. The levels of this repressor would be influenced by the concentration of MMF present, making luminescence readings hard to interpret (see Chapter 6 for more information on this topic).

The self-regulation of MmyR is not known but this was also placed under the control of *ermEp*\* to ensure consistency.

The pCC4 vectors are again transferred to *Streptomyces* via intergeneric conjugation where they integrate into the genome. This integration happens using a similar system to the *attP/attB* one from the luciferase vectors but plasmids integrate at a different position in the *S. coelicolor* genome. This pSAM2 system is not phage derived but comes from an 11 kb integrative element from *Streptomyces ambofaciens*.(120) This type of mobile genetic element is only thought to be found in *Streptomyces* but has been shown to work in a similar way to the temperate phage systems.(132)

The construction of the pKMS01 and 03 vectors allowed the study of MmfR and MmyR binding to the three MARE operator sites as well as any release of the repressor upon MMF binding (see Figure 3.1 and Figure 3.3).

### 3.3 Initial Investigation of the Reporter System

#### 3.3.1 Details on the Bioluminescence Assay

Once the two vector types just described were integrated into the *S. coelicolor* genome and had been screened via PCR for successful conjugation, the levels of luminescence produced by different promoters and MmfR binding at the MARE operator and to the MMFs could then be studied. As this system had not been used before, optimisation of methods was required before a full investigation could take place.

Preliminary tests with a limited number of repeats were done to optimise the luminescence assay and find the best ways of measuring light production before a complete data set with multiple readings was taken. This saved resources such as the synthetic MMFs which are not available commercially. A number of factors needed to be determined from optimisation trials, such as time points when readings were to be taken, the optimal concentration of MMF inducer molecule and when in the assay the inducer molecule should be added. The MmfR system was used to optimise the assay protocol as there was a clearer understanding of the type of results expected compared to MmyR. Once the methodology was optimised it could then be applied to the study of MmyR.

Luminescence was measured using a Photek CCD camera and the associated software (see Methods section for details on this technique). *S. coelicolor* does not grow uniformly in liquid culture, they form large clumps of cells and will not sporulate and do not complete their full life cycle. An advantage of this luciferase system and the CCD camera was not only that it is highly sensitive but also that it could be used to take measurements from solid cultures, which should be more representative of transcriptional regulation in the wild type system and how it alters over the complex life cycle of *S. coelicolor*.

A lawn of *Streptomyces* was grown on selective SFM media for four to five days at 30°C. This formed a fresh starter culture for the rest of the study. A sterile loop was then used to inoculate the wells of a 12-well plate, each well containing 2.5 mL SFM media. Picking cells fresh from a selective SFM plate ensured that they would be luminescing optimally.<sup>1</sup>

Different sample types were spread randomly across multiple plates. The Photek CCD camera took a more sensitive reading from the centre of its ‘view’ and so by rotating the way samples were spread out, anomalies or bias were avoided. Having this random spacing of samples also avoided any bias potentially caused by increased drying of samples in outer wells or a lack of oxygen to wells in the centre of the plate.

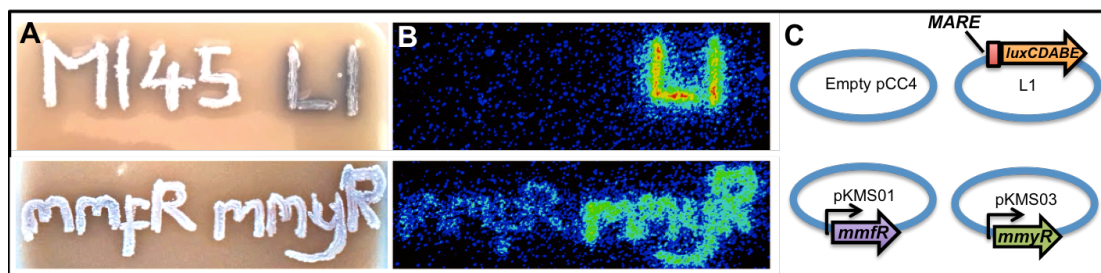
### 3.3.2 Preliminary Observations

Figure 3.8 shows an example of the type of readings taken using the Photek CCD camera. The strains used in this figure include M145 (negative control), L1+pCC4 (positive control with no repressor, labelled as ‘L1’ in Figure 3.8), L1+*mmfR* (labelled as ‘mmfR’) and L1+*mmyR* (labelled as ‘mmyR’). These were not grown in twelve well plates this one time for the sake of making this figure easy to read and label. For more details on the vectors in these strains please see Table 2.5 and Table 2.8.

Figure 3.8A shows how the strains look when grown on solid culture and the corresponding image when being measured by the CCD camera for thirty seconds is shown in Figure 3.8B. Figure 3.8C shows schematics of all the vectors used in the strains included in this figure. It can be seen that the luciferase assay in *Streptomyces* strains produces very clear levels of luminescence, easily detectable and measurable within a thirty second reading.

---

<sup>1</sup> Preliminary studies performed in this investigation showed that freezing samples and using glycerol stocks directly could compromise the amount of luminescence produced.



**Figure 3.8A-C. Details on and visual representations of strains containing the *lux* genes under the control of *mmfLp***

*Assigned nomenclature = M145 – S. coelicolor M145 negative control with no luciferase genes, L1 – positive control with luxCDABE under the control of mmfLp and the empty pCC4 vector, mmfR – strain containing L1 and pKMS01, mmyR – strain containing L1 and pKMS03*

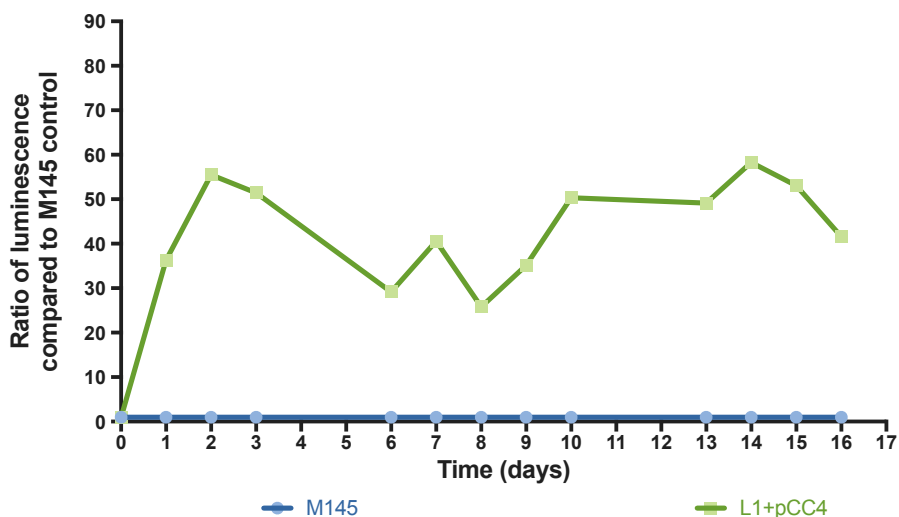
- A. Strains growing on SFM media*
- B. Luminescence seen in the Photek CCD camera during a thirty second reading by strains shown in A.*
- C. Schematic representation of reporter plasmids used*

As an initial observation from Figure 3.8, it appears that MmfR is a much better repressor than its paralogue MmyR, preventing *lux* expression a lot more efficiently (shown by the lower level of luminescence produced by the L1+*mmfR* strain versus the L1+*mmyR* strain). This was surprising as previous investigations by Sean O'Rourke *et al.* revealed that a *mmyR* knockout produces the phenotype of methylenomycin overexpression whereas a *mmfR* knockout does not.<sup>(71)</sup> Possible reasons for this inconsistency were further investigated and will be discussed further later in the report. Also of note, neither MmfR or MmyR appear to cause full repression of *luxCDABE* with both allowing some leaky expression, a detail also discussed later. As expected, the M145 negative control parent strain does not produce any measurable background luminescence whereas the positive control L1+pCC4 produced the greatest level of luminescence of the four strains.

### 3.3.3 Timing of Measurements

*Streptomyces* are slow growing bacteria, usually taking four to five days at 30 °C to get a lawn of bacteria on culture media. This is due to their complex mycelial lifecycle, which shows a number of similarities to the sporogenic fungi life cycle. A number of different time points were therefore trialled for measuring luminescence in this assay.

Figure 3.9 shows the readings collected daily over 16 days for the M145 and L1+pCC4 strains. These readings were taken to check when was best to measure luminescence. Here, and in all later trials, 0 hours refers to the time when plates have been inoculated and first placed in the incubator.



**Figure 3.9. Changes in levels of luminescence over 16 days where *luxCDABE* is under the control of *mmfLp*, compared to a negative control**

*The level of luminescence was calculated as a ratio of luminescence for L1+pCC4 compared to the negative M145 control strain (wild type *S. coelicolor* without *SCP1* or *SCP2*).*

As can be seen in Figure 3.9, there appears to be a rapid change in the amount of luminescence at around the 24-hour time point and this would then continue to increase over the next few days. After around day three there would still be measureable luminescence but it appeared to fluctuate a lot, likely to be due to the bacteria entering different stages in the cell cycle. By two weeks of growth there was still detectable luminescence but the SFM media in the 12-well plate had started to dry out and crack therefore making the results hard to compare with the initial readings.

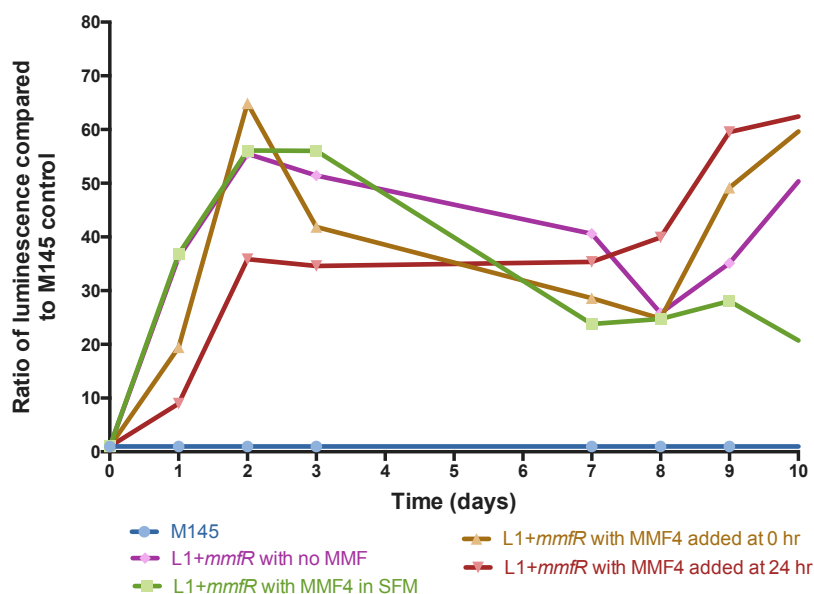
Due to these findings, more time points were trialled around 24-72 hours including measurements being taken at 0.5, 4, 13.5, 16, 17, 18, 19, 20, 21, 22, 24, 38, 40, 48, 65 and 72 hours. Readings from some of these investigations are shown later in Figure 3.10, Figure 3.12 and Figure 3.13. Based on all of these experiments, measurements at time points of 21, 24, 27, 48 and 72 hours were chosen for later tests. This allowed readings to be taken within the area of peak luminescent activity without having to take samples in and out of the incubator too often, risking both contamination and the effects of the decrease to room temperature in the Photek CCD camera.<sup>1</sup>

### 3.3.4 Investigating How and When to Add the MMFs

Tests were run to establish the effect of adding the MMFs L1+*mmfR* strains at different time points and using different protocols. The results from one of these trials are shown in Figure

<sup>1</sup> The Photek CCD camera was used at room temperature as there was no heating option with the equipment. Readings would be taken as quickly as possible, minimizing the effects of this temperature change.

3.10. For this figure, trials were run where MMFs were added at the beginning of *Streptomyces* growth (0 hours) either on top of the media or to the molten media before it had set or where the MMFs were overlaid at 24 hours, after a lawn had started to grow. This was then compared to luminescence produced by a M145 negative control strain and the relative ratio of luminescence calculated.



**Figure 3.10. Changes in luminescence produced by the *lux* operon under the control of *mmfLp* and *MmfR* by adding 400 nM MMF4 at different time points**

The level of luminescence was calculated as a ratio of luminescence produced by the M145 negative control strain. MMF4 added at 0hr and 24hr was overlaid on top of the SFM plate.

**Strains:** L1+*mmfR* – *luxCDABE* under the control of *mmfLp* (L1 vector) and *mmfR* under the control of *ermEp\** (pKMS01 vector).

As an initial observation, *MmfR* repression of the *lux* genes does indeed appear to be removed to some extent by the addition of MMF4. It was found that adding the MMFs at 24 hours growth would disrupt the surface of the lawn of growing colonies. This disruption was enough to reduce the levels of luminescence produced. It would then take a few days for the luminescence levels (and cell growth) to recover to their previous level. Figure 3.10 shows that luminescence for the strain where MMF4 was added at 24 hours is in fact lower than the non-induced L1+*mmfR* strain for the next five days until it recovers and increases again. Mixing the MMF with the SFM media appeared to produce very similar levels of luminescence to the L1+*mmfR* strain with no MMF, indicating that the furans might not diffuse through the SFM efficiently enough or that they were affected by the temperature of the molten SFM before it had set. (The SFM used had a pH of 7, previous investigations have indicated that the MMFs may diffuse better at around pH 5.)(14) Overlaying the hardened SFM with MMF4 at 0 hours shows a distinct peak in luminescence straight away. For these

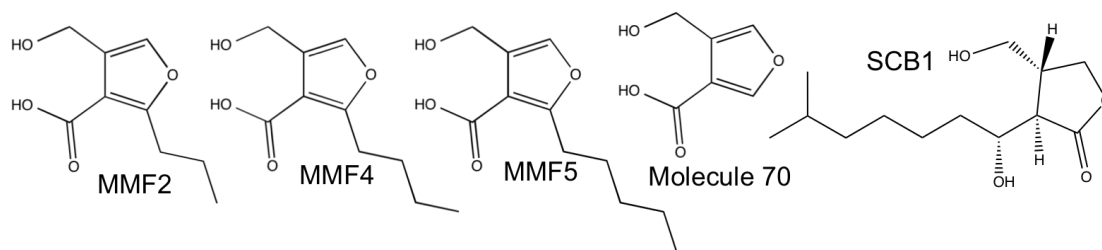
reasons, it appeared best to add the MMFs on top of the media at ‘time 0’, when the plates are also being inoculated with *Streptomyces*.

It is important to note that the concentration of MMF4 in these trials was only 400 nM. Although this produced detectable amounts of luminescence, future trials often used ten times this concentration of MMF to get a more distinct result. Also, the results from Figure 3.10 are taken over ten days, as discussed in Section 3.3.3 these later time points were not used in later investigations due to the plates drying out and the media cracking.

### 3.3.5 Release of MmfR by Different Small Molecules

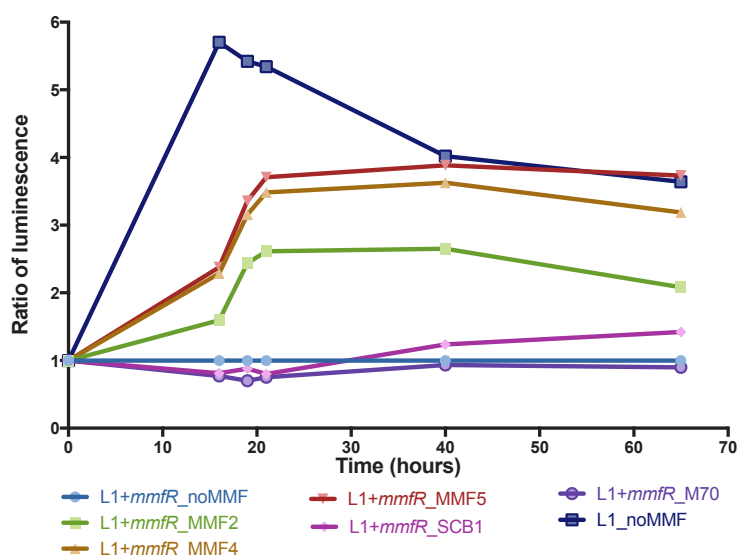
To assess whether MmfR was released from the MARE operator by the different MMFs, MMF2, 4 and 5 were added to the L1+*mmfR* strain. Levels of luminescence were then compared to a negative control of the same strain with no MMFs as well as an L1+pCC4 positive control (representing how the system looks with no repression).

Data was also collected for two other small molecules; molecule 70 and SCB1. Molecule 70 is a synthetic analogue from the work of Nicolas Malet that has the same core structure as the MMFs but lacks the alkyl chain, which has been shown to be necessary for fitting the binding pocket.<sup>(55)</sup> SCB1 is a *S. coelicolor* butyrolactone signalling molecule, known to bind to the TetR family member ScbR. Neither of these analogues were expected to bind MmfR and so function as another negative control for this study. The chemical structures of these negative controls along with MMF2, 4 and 5 can be found in Figure 3.11. Luminescence results from these studies can be found in Figure 3.12.



**Figure 3.11. Chemical structures of MMF2, 4 and 5 and SCB1 from *S. coelicolor* and synthetic molecule 70**





**Figure 3.12. Changes in luminescence produced by the *lux* operon under the control of *mmfLp* and *MmfR* in the presence 100  $\mu$ M of different microbial hormones**

**Strains used:** *L1* – positive control, *luxCDABE* under the control of *mmfLp* and empty *pCC4* vector with no repressor, *L1+mmfR* – *luxCDABE* under the control of *mmfLp* (*L1* vector) and *mmfR* under the separate control of *ermEp\** (*pKMS01* vector).

The level of luminescence was calculated as a ratio of luminescence produced by *L1+mmfR* with no MMFs present.

In Figure 3.10 there was an observable increase in luminescence by *L1+mmfR* in the presence of 400 nM MMF4 after 48 hours growth. The inducibility of *L1+mmfR* strains can again be seen in Figure 3.12 where there is a clear increase in luminescence produced in the presence of 100  $\mu$ M MMF2, 4 and 5 compared to no MMFs. At this concentration none of the MMFs achieve levels of luminescence compared to the *L1+pCC4* positive control, indicating that higher concentrations could also be trialled. As expected, molecule 70 and SCB1 give levels of luminescence very close to those for the negative control (*L1+mmfR* with no small molecule added).

### Effect of the MMFs on a Positive Control

The five MMFs were also added to the *L1+pCC4* positive control strain at a concentration of 100  $\mu$ M. Upon a statistical analysis of data (not shown here), no significant difference was found in the levels of luminescence produced in the presence of the MMFs compared to the absence for this strain. This indicates that the concentrations being used, the MMFs are not having a toxic effect of *Streptomyces* growth or regulating the methylenomycin BGC in other ways, independent of *MmfR*/*MmyR*.

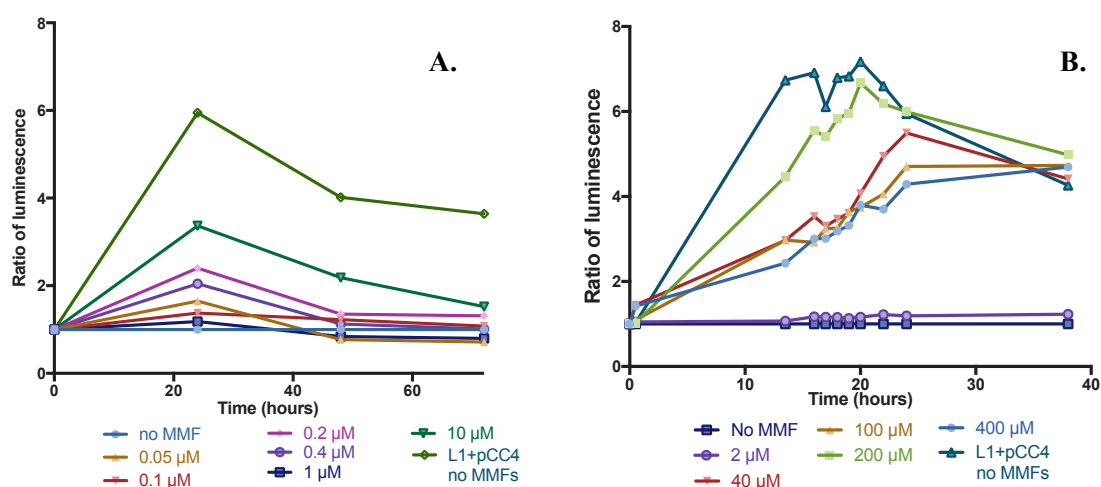
### Investigating the Release of *MmyR* by the MMFs

When equivalent trials were run with *L1+mmyR* instead of *L1+mmfR*, there was no observable increase in luminescence upon the addition of the MMFs. This and the much

lower levels of luminescence repression observed for *mmyR* strains shown earlier in Figure 3.8 indicates that MmyR has a very different role to MmfR. This is further discussed later in Section 5.

### 3.3.6 Optimal Concentration of Inducers

Following on from Figure 3.12 where it was seen that at 100  $\mu\text{M}$  MMF did not achieve levels of luminescence in line with that of the L1+pCC4 positive control, a range of different MMF concentrations were trialled to see if full release of MmfR could be attained. Results from assays where MMF2 and 5 were added at various concentrations between 0.05  $\mu\text{M}$  and 400  $\mu\text{M}$  are shown in Figure 3.13A and B.



**Figure 3.13A and B. Luminescence produced by the *lux* operon under the control of *mmfLp* and MmfR upon the addition of different concentrations of MMF2 and MMF5**

*A. Data collected for MMF2*

*B. Data collected for MMF5*

**Strain used:** *L1+pCC4* – positive control, *luxCDABE* under the control of *mmfLp* and empty *pCC4* vector with no repressor, *L1+mmfR* – *luxCDABE* under the control of *mmfLp* (*L1* vector) and *mmfR* under the separate control of *ermEp\** (*pKMS01* vector). The level of luminescence in both charts was calculated as a ratio of luminescence produced by *L1+mmfR* with no MMFs present.

Figure 3.13A shows the results from an investigation that used MMF2 at a concentration range of 0.5  $\mu\text{M}$  to 10  $\mu\text{M}$ . All of these concentrations produce more luminescence than a control without MMFs with up to 3.5 more luminescence being produced for the highest concentration compared to no MMFs being present. However, around twice as much as this is needed to achieve the same levels of luminescence as the positive control.

As can be seen in Figure 3.13B, MMF5 was trialled at concentrations between 2  $\mu\text{M}$  and 400  $\mu\text{M}$ . Concentrations of MMF5 above 200  $\mu\text{M}$  appear to no longer cause an increase in bioluminescence, possibly indicating the toxicity of the MMF compounds or a saturation of MmfR binding. A concentration of 100  $\mu\text{M}$  was chosen as a standard to use in later

investigations as it was a concentration in the middle of those that showed good levels of luminescence.

MMF2 was later trialled at concentrations in the range of 40  $\mu\text{M}$  to 400  $\mu\text{M}$  and MMF4 in the range of 2  $\mu\text{M}$  and 400  $\mu\text{M}$  (not shown here) and gave similar data sets to that seen for MMF5.

### 3.3.7 Summary of Strategy Chosen to Investigate MmfR/MMF/MARE Operator Interactions

In conclusion, the strategy designed to investigate MmfR/MMF/MARE operator interactions (Figure 3.3) and the research questions for this project proved to be functional *in vivo* in *Streptomyces*. The luciferase assay produced easily detectable levels of luminescence for the positive control (L1+pCC4) and MmfR repression of *luxCDABE* appeared to be released by the addition of three of the known MMFs at concentrations within the micromolar range. The luciferase assay was therefore deemed suitable for use in the remainder of the investigation into MmfR (an MmyR) interactions with the MMFs and MARE operators and the findings from this chapter formed the basis of much of the rest of the investigative work done in this project.

An optimised protocol was designed using data collected in this chapter, considering the best time to add the MMFs, the optimal concentration to use, which time points to take readings at as well as how to prepare bacterial cultures for analysis.

#### **Finalised Bioluminescence Protocol**

The following protocol was used to collect all future bioluminescence data. To each well of a 12 well plate, 2.5 mL SFM was added and allowed to set. The MMFs were diluted in DMSO to an appropriate concentration before being diluted one in ten with water and 10  $\mu\text{L}$  added to each well. (A 10% DMSO solution was used instead as a control when no MMFs were required, giving a final DMSO concentration of 0.0004%.) To find out the  $K_d$  and  $B_{\text{max}}$  for each MMF a concentration range of 5  $\mu\text{M}$  to 400  $\mu\text{M}$  was used. For all other tests requiring MMFs, a concentration of 100  $\mu\text{M}$  MMF was used as a standard.

After the MMFs were added, plates were inoculated with the *Streptomyces* strain of interest collected from a fresh starter culture plate and incubated at 30 °C with readings being taken at 21, 24, 27, 48 and 72 hours using the Photek CCD camera.

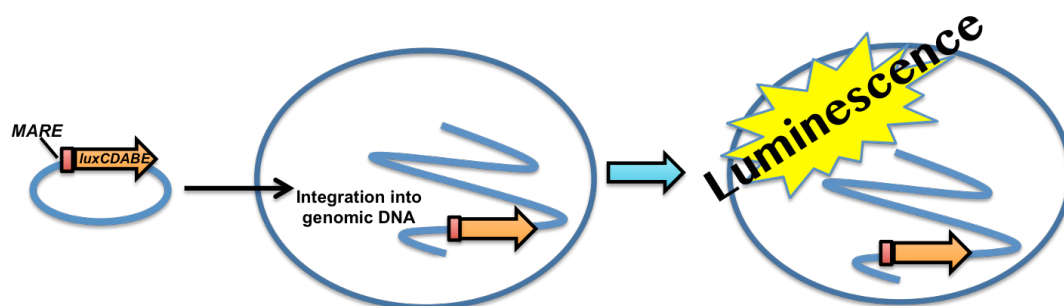
One down side of the luciferase assay was that there were large variations between results collected. For this reason, in the future multiple repeats were collected and data was again analysed as a ratio relative to a control sample.

## 3.4 Promoter Strength

### 3.4.1 Strategy for Investigating Promoter Strength

As explained earlier, there are three different MARE operator sites that MmfR is thought to bind to, regulating the expression of five different operons by blocking five promoter sequences. A diagram of the methylenomycin cluster can be found in Figure 3.5 and Figure 3.6 with an explanation of the nomenclature used in this investigation.

Each of the promoters downstream of a MARE operator sequence was tested by cloning different intergenic regions and placing them in different orientations upstream of the *luxCDABE* operon, according to the protocol explained in Section 3.2. These vectors were then studied in the absence of the transcriptional repressor MmfR or MmyR to investigate the strength of the promoter only, creating the L1+pCC4, L1F+pCC4, L2+pCC4, L3+pCC4 and L3F+pCC4 strains. No MMFs were added during this investigation and the optimised protocol laid out in Section 3.3.7 was used. A schematic of how this assay was set up can be found in Figure 3.14 (see also Figure 3.1). In the absence of MmfR/MmyR and the MMFs, the *lux* genes should be constitutively expressed and the bioluminescent output proportional to the strength of the promoter in the vector being used.

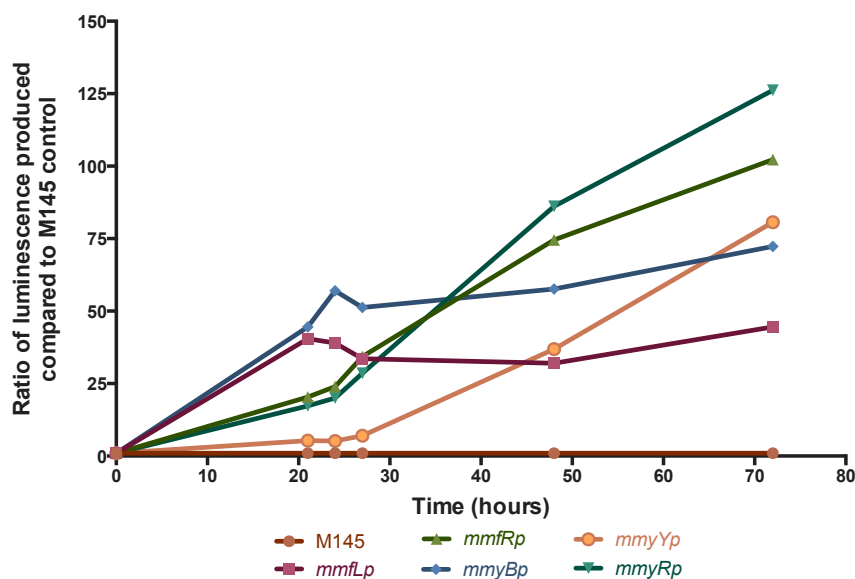


**Figure 3.14. Schematic of luciferase assay used to investigate the strength of promoters from the methylenomycin biosynthetic cluster**

*To investigate promoter strength, one vector is integrated into the Streptomyces genome; containing luxCDABE under the control of one of the five relevant promoters from the methylenomycin cluster. No MmfR or MmyR is produced by these strains; instead an empty pCC4 vector is used as a control.*

### 3.4.2 Results on Comparison of the Five Methylenomycin Cluster Promoter Strengths

Figure 3.15 shows the levels of luminescence produced over 72 hours by strains containing the five different methylenomycin cluster *lux* vectors. Figure 3.16 is a boxplot of the data from Figure 3.15 at the 72 hour time point.<sup>1</sup> This box plot includes data from 10-90% percentiles and all other results shown as outliers. For both figures, data is normalised against luminescence values for the M145 negative control.

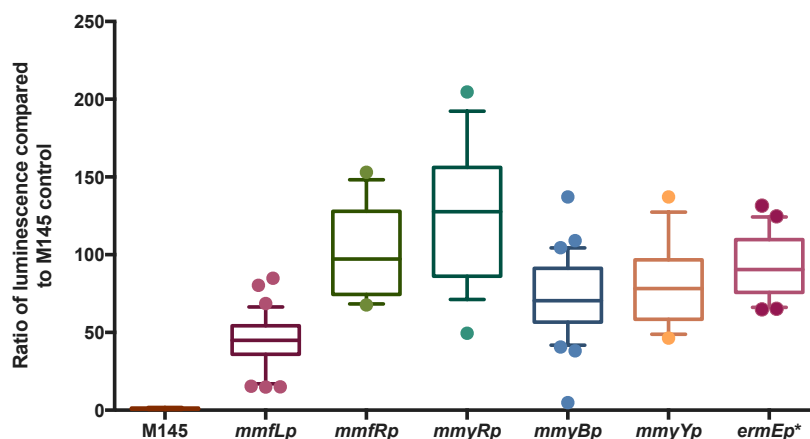


**Figure 3.15. Luminescence produced by the *lux* operon when under the control of different promoters from the methylenomycin gene cluster**

*Data normalised against luminescence values for the M145 negative control.*

**Strains used:** *M145* – negative control strain (wild type *S. coelicolor* without *SCP1* or *SCP2*), **all other strains** – contain *luxCDABE* under the control of different promoters from the methylenomycin gene cluster

<sup>1</sup> This figure also contains results from analysis of *ermEp\**, detailed later in Section 3.4.3.



**Figure 3.16. Boxplot of luminescence produced by the *lux* operon when under the control of different promoters, at 72 hours growth**

*Data normalised against luminescence values for the M145 negative control. This box plot includes data from 10-90% percentiles and all other results are shown as outliers.*

**Strains used:** *M145* – negative control strain (wild type *S. coelicolor* without *SCP1* or *SCP2*), **all other strains** – contain *luxCDABE* under the control of different methylenomycin cluster promoters

Figure 3.16 suggests the following order in terms of promoter strength;

***mmyRp* > *mmfRp* > *mmyYp* > *mmyBp* > *mmfLp***

**L2 > L1F > L3F > L3 > L1**

Another way of displaying this is;

**TetR > TetR > biosynthetic gene > transcriptional activator > MMF production**

Both *mmyRp* and *mmfRp* produce an average of more than 100 times the level of luminescence produced by the M145 control whereas *mmyYp* and *mmyBp* produce between 70 and 80 times that of the control. Even the weakest promoter, *mmfLp*, produced 45 times as much luminescence as the M145 control.

It is also interesting to point out that the three promoters which produced the highest levels of luminescence (L1F, L2 and L3F) also show some phenotypical differences from the L3 and L1 strains. In particular, the blue pigments normally seen when growing *S. coelicolor* are missing. This indicates the possible toxicity of high levels of *luxCDABE* expression or just the drain on cell resources that such high expression levels produce. The cells do however appear to grow at the same rate and still sporulate after three or four days growth and a 16S analysis of the DNA (not shown here) did reveal that it was indeed *S. coelicolor* present rather than a contaminant strain. The implications of these observations are therefore unclear.

It should be noted that at 48 hours compared to 72 hours, *mmyYp* and *mmyBp* will swap places in terms of promoter strength (otherwise the order of promoter strength remains the same). It is hard therefore to say which is stronger overall; *mmyYp* or *mmyBp*. It may be more accurate to say;  $L2 > L1F > L3F \approx L3 > L1$ . However, as the L3F strain shows the phenotypic differences associated with the stronger promoters and the L3 strain does not, it does seem that *mmyYp* is likely to be slightly stronger than *mmyBp*. Therefore the order of promoter strength displayed previously does seem the most probable.

The implications of the TetR family repressors; MmfR and MmyR having the strongest promoters are unclear. The *mmyR* gene is known to be expressed after *mmfR* (71) so possibly the high levels of expression are necessary to make enough quickly enough for it to be able to carry out its repressive role. In addition, whereas one of each of the MmFLHP enzymes can make many MMF molecules, one repressor protein dimer can only bind and block one operator site and so many more are needed. A larger number of structural proteins are needed than enzymatic ones.

When analysing this data it is also important to consider that all of the *lux* constructs used here have relied on the same RBS<sup>1</sup> for controlling translation of the *lux* genes. These results therefore show the relative amount of expression generated by a promoter rather than the absolute amount (the number of elongating polymerases per second) (133) that may be found in the wild type system. Another factor which may influence the relative versus absolute promoter strengths is the interplay of sigma factors (Section 1.4.4). It is not known whether the different genes in the methylenomycin BGC recruit different sigma factor/polymerase complexes, thereby altering the absolute expression at a particular promoter.

### 3.4.3 Comparison with Other Promoters

The luciferase system used in this investigation was fairly recently developed and its use in *Streptomyces* has been limited. For this reason, it was decided to check the system using the well-known *Streptomyces* promoter *ermEp\** as a benchmark for the methylenomycin cluster promoters. The *ermEp\** promoter is well characterised (131) and known to work as a relatively strong promoter in *S. coelicolor*, it was also the promoter used to produce MmfR and MmyR at a constant level. Table 3.1 shows a comparison of the sequence for *ermEp\** with five of the methylenomycin cluster promoters.

---

<sup>1</sup> RBS sequence used was 5' AAGGAGG 3'

**Table 3.1. Comparison of the different methylenomycin cluster promoter sequences with strong constitutive promoter, *ermEp\****

| Promoter name | -35<br>(5' to 3') | -10<br>(5' to 3') | Strength<br>(from<br>Figure<br>3.16) | Full double stranded promoter sequence (with -10 and -35 sequences underlined)                            |
|---------------|-------------------|-------------------|--------------------------------------|---|
| <i>ermEp*</i> | TTGCAT            | GAGGATCCT         | -                                    | 5' AGCT <u>TTGCAT</u> GCCGGTTCGACTCTAGAGGATCCTACC 3'<br>3' TCGA <u>ACGTAC</u> GCCAGCTGAGATCTCCTAGGATGG 5' |
| <i>mmfLp</i>  | TTGCAT            | AGGTATATT         | 5                                    | 5' CCCAT <u>TGCATA</u> AATACCTTCCCGCAGGTATATTTCT 3'<br>3' GGGT <u>AACGTAT</u> TATGGAAGGGCGTCCATATAAAGA 5' |
| <i>mmfRp</i>  | TTGCCG            | CGGTAAGCT         | 2                                    | 5' AGCTTACCGATCCCGGCTGTCTTGCAGCGGCGCAA 3'<br>3' TCGAATGGCTAGGGCCGACAGAACGTCGCGCCGTT 5'                    |
| <i>mmyRp</i>  | CTAACA            | GGGTATGTT         | 1                                    | 5' CCCACTAACATACCTTCCCGAGGGTATGTTTCCG 3'<br>3' GGGTGATTGTATGGAAGGGCTCCCATACAAAAGGC 5'                     |
| <i>mmyBp</i>  | TTAACT            | AGTCATAAA         | 4                                    | 5' CCGGTTAACTCTCCGTTACGAGTCATAAAAAACC 3'<br>3' GGCCAAATTGAGAGGCAATGCTCAGTATTTTTTGG 5'                     |
| <i>mmyYp</i>  | TTTATG            | AGTTAAACC         | 3                                    | 5' CCCCGGTTAACTCTCCGTTACGAGTCATAAAAAA 3'<br>3' GGGGCCAAATTGAGAGGCAATGCTCAGTATTTTTT 5'                     |

An *ermEp\** vector was created as before, with the *ermEp\** promoter being subcloned upstream of *luxCDABE* using L1 as a template backbone.(92) This vector has been called L4 (see Table 2.5) and was used to create the L4+pCC4 strain. This was then used to test the strength of *ermEp\** in the same way as the methylenomycin cluster promoters were tested (see Figure 3.14).

Figure 3.16 shows the results collected using L4+pCC4 compared to the data collected for the methylenomycin cluster promoters. Data from this figure indicates that *ermEp\** falls within the following order of promoter strength;

*mmyRp* > *mmfRp* > ***ermEp\**** > *mmyYp* > *mmyBp* > *mmfLp*

An observation from Figure 3.16 is that there is not a huge difference between any of the promoters with the mean luminescence produced ranging between 45 and 126 times that of the M145 control at 72 hours. Other studies involving promoter strength often involve a much larger magnitude of variation in possible promoter strengths. For example, Bai *et al.* saw a 190 fold difference in promoter strength when analysing different modular regulatory elements whereas there is less than a three fold different seen here.(134) It is conceivable that these promoters are all of similar strengths but it is also a possible indication of the limitations of the *luxCDABE* system in *Streptomyces*. Using *luxCDABE* compared to *luxAB* is beneficial because nothing needs to be added to the system. Both luciferase and its substrate are produced and so there is luminescence generated without any external manipulations. The down side of this is the possibility that the constitutive expression of *luxCDABE* is very taxing on the cell's resources and perhaps the strongest promoters are not revealing their true



strength. Despite this possible limitation however, the technique is still useful for the aim of this project; to develop a novel inducible expression system. The luciferase reporter assay can reveal which promoter is strongest and therefore most suitable for use in the inducible expression system being designed. For this reason, and the successes encountered while optimising this technique, the luciferase assay was used for more investigations into MmfR/MMF/MARE operator interactions.

A possible improvement to this protocol for the future would be to use the luciferase assay to assess other well-studied promoters. In this way, the effects of the potentially taxing expression of *luxCDABE* could be further explored. For example, the work by Bai *et al.* in 2015 could be used as a benchmark for promoter strength. In this work, hundreds of promoters were compared using a novel streptomycete *gfp* assay. It would be interesting to see how the data they collected may correlate with data produced by the system studied here. The recently developed *kasOp\** promoter (135) has been found to be ~20 times stronger than *ermEp\** and if added to the *lux* system,(134) would provide an insight into the implications of high promoter strength in an energy-demanding reporter assay.

### 3.5 Conclusions Drawn from Preliminary Investigation and Outlook for Future Investigations

#### **MmfR/MARE Operator Binding**

Investigations thus far have shown that MmfR will bind to the intergenic region between *mmfR* and *mmfL*. Using the finalised bioluminescence protocol (Section 3.3.7), MmfR binding to the two other intergenic regions between *mmyY* and *mmyB*, and *mmyR* and *mmfP* are also to be investigated.

#### **MmfR/MMF Binding**

Preliminary trials showed that MmfR repression could be released from the intergenic region between *mmfR* and *mmfL* upon the addition of MMF2, 4 or 5 (Figure 3.10 and Figure 3.12). Different concentrations of these MMFs trialled produced quantifiable results that were distinct from one another and would be suitable for testing using statistical analyses such as a *t*-test. Following on from this it was decided to trial all five MMFs using the finalised protocol described, using a range of concentration to obtain  $K_d$  and  $B_{max}$  values for each of the MMFs and from this calculate the binding potential of each. The conclusions from trialling different concentrations of MMFs in Section 3.3.6 (Figure 3.13) were particularly useful when designing experiments to investigate the  $K_d$  and  $B_{max}$  values for each of the MMFs. The

findings of the investigation into MmfR/MARE operator and MmfR/MMF binding can be found in Section 4.

### **MmyR Interactions with the MARE Operators and MMFs**

The mechanism by which MmyR works is still unclear and the potential results from further investigations are intriguing. Preliminary trials indicated that there was little MmyR binding to the intergenic region between *mmfR* and *mmfL* and the addition of the MMFs did not produce an obvious change in luminescence produced (Figure 3.8). As with the plans for the further investigation into MmfR, MmyR was also investigated using the luciferase assay trialled here (Chapter 5). This was done to further investigate potential MmyR binding at the two other intergenic regions between *mmyY* and *mmyB*, and *mmyR* and *mmfP* as well as investigating possible ligands for this ‘pseudoreceptor’ type of TFR.

### **Promoter Strength**

It was seen from the investigation carried out in Section 3.4 that the different promoters in the methylenomycin biosynthetic cluster had different strengths. In particular, the promoters for the transcriptional repressors were stronger than those for biosynthetic enzymes. This data will be useful as a benchmark and baseline when studying MmfR/MARE operator interactions adjacent to or in line with the DNA sequences for these promoters.

## 4 Mode of Action of Transcriptional Repressor MmfR in *Streptomyces coelicolor*

### 4.1 Aims and Strategy of Investigation

In this section, a more complete selection of bioluminescence results are displayed, looking specifically at the function of MmfR. The methodology used to collect this data is based on the finding of the preliminary results collected in Section 3 and the optimised protocol specified in Section 3.3.7. Vectors used to obtain results were explained previously in Section 3.2 and Table 2.8.

For each measurement there were at least 16 biological and technical repeats spread across at least three different ‘sessions’ of taking results. By taking measurements during different sessions any fluctuations due to variations in the communal incubator door being opened and different batches of media etc. should be accounted for. Data collected were analysed, finding averages and looking at the relative ratios between results as well as looking for statistical significance using a *t*-test.

The investigation was split into two main parts designed to study MmfR binding to the MARE operator or the methylenomycin furan ligands. Variables tested to study MARE operator interactions include observing the reduction in luminescence produced upon MmfR binding at the MARE operator regions as well as investigating the level of MmfR release at the different MARE operator sites by a single concentration of MMF. Variables tested to study ligand interactions included trialling different concentrations of each of the five MMFs and calculating the binding potentials for each. Following on from this, the ligand binding pocket was investigated *in silico* and mutants created that allowed the further exploration of key ligand binding residues.

### 4.2 Statistical Explanation of Data Handling

As explained earlier in Section 3.3.1, *Streptomyces coelicolor* do not grow well in liquid culture, never reaching the sporulation stage and forming large clumps of aggregated cells. For this reason, the analysis was always done using solid cultures. A downside of the technique used was that the exact number of cells was never known. This therefore meant that there was a lot of variation between repeats and a high coefficient of variance. Measuring the mass of cells on solid culture can be a very lengthy process and would not allow samples to be used for repeat measurements. For this reason, luminescence was recorded as a relative

ratio of light produced compared to the negative control for a particular sample. By calculating a ratio, the variance between the different sample types should cancel one another out.

It was also found that there is much less variability between readings at later time points (e.g. when looking at promoter strength in Figure 3.15) making these more reliable report points. More detailed statistical analyses of results collected were done with data from the 48 and 72 hour time points, when the standard deviation and coefficient of variance was lower. A possible reason for why results appeared to stabilise at the later time points could be that the cells were entering the stationary phase of growth.

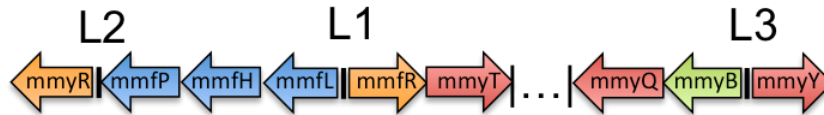
## 4.3 MmfR-MARE Operator Interactions

### 4.3.1 Details on the MARE Operator Sequences

Within the intergenic regions of the *lux* vectors used (Figure 3.6A), it is not only the promoter region that is of interest but also mainly the MARE operator sequence and the binding affinity of MmfR to it. These methylenomycin auto-regulatory response elements vary in sequence between the three known sites. The different sequences therefore can be expected to have different affinities for MmfR. The sequence of the three methylenomycin cluster MARE operators can be as follows:



The nucleotides found across all three MARE operators are highlighted in yellow. A reminder of the nomenclature used in strains created corresponding to the location of these MARE operators can be found in Figure 4.1.



**Figure 4.1. Schematic showing nomenclature of *lux* vectors, based on the MARE operator being studied for a particular methylenomycin cluster intergenic regions**

*Here the earlier nomenclature of L1 and L1F and L3 and L3F have been combined simply as L1 and L3 respectively. Please note that this diagram is not to scale.*

*For a scale diagram please see Figure 3.5.*

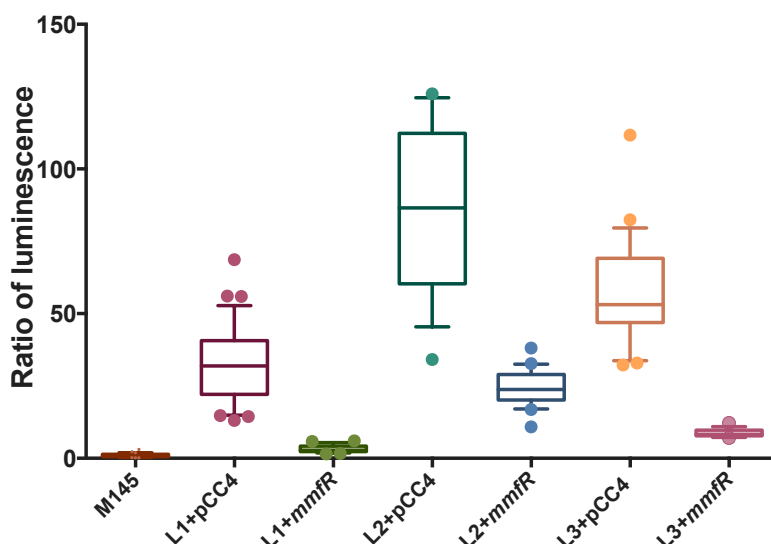
To investigate MmfR-MARE operator interactions, experiments were done with the aim of observing the reduction in luminescence produced upon MmfR binding at the MARE operator regions as well as exploring the level of MmfR release at the different MARE operator sites by a single concentration of MMF. This investigation requires the full two-vector luciferase system to be used, with both the *lux* vector and the pKMS01 vector (Figure 3.3).

MmfR is homodimeric and so affinity to the MARE operator in either orientation within a vector would be expected to be unchanged as the 24 bp MARE operator is the same in either direction (i.e. L1F should be the same as L1). However, the relative location of the promoter sequences to MmfR binding site (Figure 3.6A) may have an effect of the degree of repression that is achieved meaning that there may be differences between repression seen in L1F versus L1 and L3F versus L3 strains. In this investigation into MARE operator binding the L1, L2 and L3 vectors were used but not L1F or L3F. If more time had been allowed then L1F and L3F would also have been trialled, adding pKMS01 to create L1F+*mmfR* and L3F+*mmfR*.

#### 4.3.2 Strength of MmfR Binding to the MARE operators

Figure 4.2 shows the results of an investigation into the strength of MmfR binding to the three different MARE operators. In this figure, luminescence produced by L1, L2 and L3 vectors are compared when pKMS01 is present versus a pCC4 control with no *mmfR*. This study gave information on the level of repression that is achieved in the presence of MmfR compared to a negative control (M145) and positive control (L1+pCC4, L2+pCC4 or L3+pCC4) for a particular operator.

For Figure 4.2, all data is normalised and calculated as a ratio of the level of luminescence produced by the M145 control at 48 hours. L1+pCC4, L2+pCC4 and L3+pCC4 represent the level of luminescence expected to be produced at a particular promoter in the absence of any repression.



**Figure 4.2. Luminescence produced by the *lux* operon under the control of different methylenomycin cluster operators in the presence and absence of MmfR at 48 hours growth compared to a *S. coelicolor* M145 negative control**

Box plot includes data from 10-90% percentiles and all other results shown as outliers.

**Key:** ‘mmfR’ refers to the presence of the pKMS01 vector with mmfR under the control of ermEp\*, L1, L2 and L3 refer to the *lux* vectors with *luxCDABE* under the control of different MARE operators, pCC4 is an empty vector used as a control for pKMS01

From Figure 4.2 it can be seen that there is a 10-fold decrease in luminescence from the L1 operator upon MmfR binding compared to 3.5-fold and 6.5-fold for L2 and L3 respectively at 48 hours growth. This indicates that MmfR binds best to the L1 MARE operator, between *mmfL* and *mmfR*, followed by the L3 MARE operator, between *mmfY* and *mmfB*, and least strongly to the L2 MARE operator, between *mmfP* and *mmfR*. A similar pattern of results was also seen at 72 hours (not displayed here).

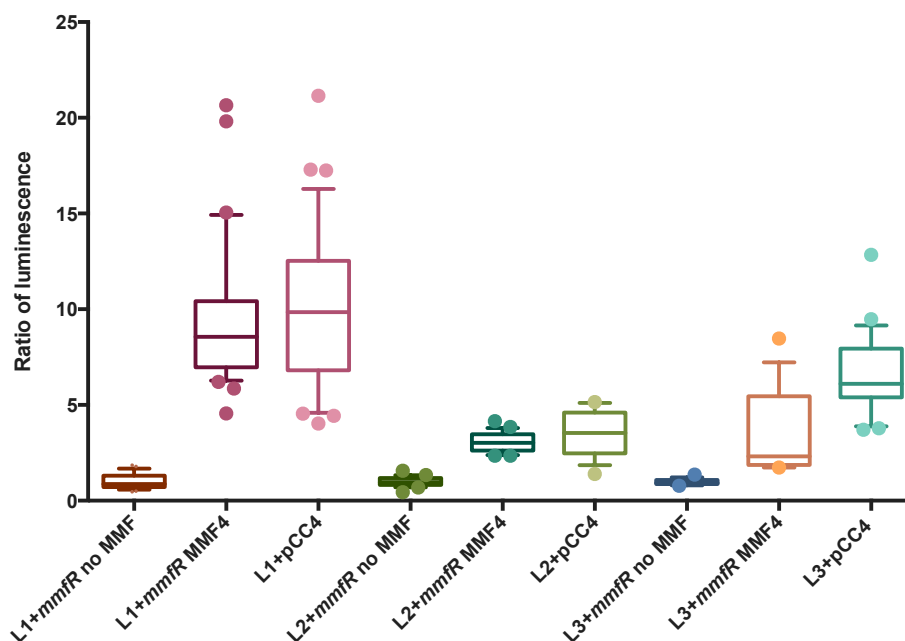
Of note is that at no operator site was there seen to be full repression of the *lux* operon (for all *mmfR* strains there was always more measurable luminescence than the level produced by the M145 negative control). At 48 hours L1+*mmfR* produced three times as much luminescence as the M145 control and for L3+*mmfR* there was almost nine times as much whereas there was over 24 times as much luminescence for L2+*mmfR* compared with the M145 control. This revealed varying degrees of apparent leakiness in the system. The level of luminescence will be influenced not only by the strength of MmfR binding but also the promoter strength at each of the sites.

The biological implications of this leakiness is unclear but it is potentially helpful in regulating methylenomycin biosynthesis and maintaining equilibrium between MmfR release and repression. If there is leakiness in the wild type system, the five operons may be expressed at a low level including the production of small amounts of MMFs by the *mmfLHP*

operon. This does not however appear to be enough to trigger the production of methylenomycin under normal laboratory conditions.(71) The *mmfR* and *mmyR* promoters are the strongest (Section 3.4), high levels of these repressors maybe enough to switch off the system until a threshold concentration of MMFs is achieved and the level of MmfR cannot keep up with the de-repression. However, there are a number of reasons why this leakiness may not actually be found in the wild type system. One is that in the wild type system, *mmyB* and *mmfL* expression are also controlled by the temporal use of a rare TTA codon, which can only be expressed in later stages of the cell cycle, when it's tRNA is available. Also, *in vivo* in wild type *S. coelicolor* A3(2), where SCP1 and the entire methylenomycin cluster and other regulatory elements are present, it is reasonable to predict that maybe the leakiness observed in this assay would be lower due to the constant self-regulatory production of MmfR (see Chapter 6 for more details). Once MmfR is under the control of its own promoter and MARE operator (rather than *ermEp\**) there is likely to be tighter regulation of the levels of this repressor. There is also potential additional assistance from sigma factors and other regulatory molecules on the regulation of this repressor when in the wild type system. The extent to which these different factors may play a role however, is unclear.

### 4.3.3 Investigation into Release of MmfR from the MARE Operator by MMF4

MmfR affinity for the MARE operator could also be studied by looking at how easily the repressor is released in the presence of the MMF at each of the three sites. A concentration of 100  $\mu$ M MMF4 was added to L1+*mmfR*, L2+*mmfR* and L3+*mmfR* and the levels of luminescence compared to the same strains with no MMFs. By comparing the results with the same strains with no MMFs, it was possible to offset the variation caused by different promoter strengths next to the different MARE operators. Data for the positive control as a ratio of each *mmfR* strain is also shown to represent maximal luminescence for a particular promoter in the absence of repression (L1+pCC4, L2+pCC4 and L3+pCC4 strains). The results from this investigation are found in Figure 4.3.



**Figure 4.3. Boxplot of luminescence produced by the *lux* operon under the control of different MARE operators and MmfR in the presence and absence of 100  $\mu$ M MMF4, compared to luminescence produced by *luxCDABE* under the control of the same operator but no MmfR at 48 hours**

*The data for the mmfR strains with MMF4 were normalised against the same strain with no MMF to give a relative ratio of 1 for the negative control. The box plot includes data from 10-90% percentiles and all other results shown as outliers. Key: 'mmfR' refers to the presence of the pKMS01 vector with mmfR under the control of ermEp\*, L1, L2 and L3 refer to the lux vectors, pCC4 is an empty vector used as a control for pKMS01 in the positive control strains.*

### Comparison with the Negative Control

Upon the addition of MMF4 to *mmfR* strains at the 48 hour time point there is an average of around between 3 and 3.3 times as much luminescence produced for the L2 and L3 operators respectively, compared to over nine times more for L1. This reveals that as well as binding most strongly to the L1 MARE operator (Figure 4.2), MmfR is also released most readily from it in the presence of its ligand.

### Comparison with the Positive Control

The results comparing MmfR/MMF4 data with the pCC4 positive controls for each MARE operator follows a similar pattern. The L1+*mmfR* strain with MMF4 achieves the closest levels of luminescence to the positive control, producing 96% of the luminescence that the positive control produced. On the other hand, L3+*mmfR* with 100  $\mu$ M MMF4 only achieves around 50% of the luminescence produced by the positive control and L2+*mmfR* achieves around 87%. This indicates that MmfR is less readily released at the L3 MARE operator, followed by L2 with L1 being the most readily release. The data from Section 4.3 is summarised in Table 4.1.



#### 4.3.4 Discussion of MmfR/MARE Operator Binding Data

Table 4.1 shows a comparison of all the results obtained from the luminescence assay on the affinity of MmfR for each of the methylenomycin cluster intergenic regions it is known to bind as well as data on the strength of promoters in each region.

**Table 4.1. Summary of MmfR-MARE operator binding data**

The data in this table summarises the findings from Figure 4.2, Figure 4.3 and Figure 3.16. Data in this table is taken from analyses done at either 48 or 72 hours growth.

| MARE operator site | Corresponding promoter | Strength of MmfR binding | MmfR release in the presence of MMF4 compared to a negative control | MmfR release in the presence of MMF4 compared to a positive control | Relative promoter strength (1 being highest) |
|--------------------|------------------------|--------------------------|---|---|--|
| L1                 | <i>mmfLp</i>           | Greatest                 | Greatest  | Greatest  | 5  |
| L1F                | <i>mmfRp</i>           | -                        | -   | -   | 2  |
| L2                 | <i>mmyRp</i>           | Weakest                  | Weakest   | Middle  | 1  |
| L3                 | <i>mmyBp</i>           | Middle                   | Middle  | Weakest   | 4  |
| L3F                | <i>mmyYp</i>           | -                        | -   | -   | 3  |

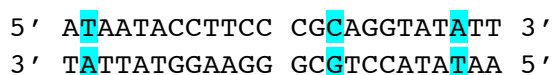
It appears, according to the comparison of results in Table 4.1, that MmyR production has the strongest promoter but is one of the ‘hardest’ to switch on with the lowest level of release of MmfR measured from the L2 intergenic region upon the addition of MMF4 compared to the same strain with no MMFs. This may explain why MmyR production has previously been shown to occur after the production of MmfR.(54)

The *mmyBY* intergenic region (L3) appears to produce results in the middle in terms of both promoter strength and MmfR binding whereas the *mmfR* and *mmfLHP* operons are quite clearly the easiest to switch on and off. The MARE operator between these two operons (L1 MARE operator) was also shown to be the least leaky, producing only three times as much luminescence in the presence of MmfR as the M145 control compared to 24 times for the L2 MARE operator. For this reason, the L1 MARE operator seems to be the most suitable choice for use in a novel inducible expression system. Following on from this, the L1 vector was also used in all future investigations in Section 4.4. L1 was chosen for this instead of L1F strains due to the phenotypic differences in the strains with stronger promoters (referred to after Figure 3.16 on page 65). It is as yet unclear what the implications of using the high strength promoters in the resource demanding *luxCDABE* system so to ensure consistent and reliable results the more ‘healthy’ looking L1 strains were used. Also, the differences in promoter strength were just over two fold between L1 and L1F, not a huge difference when one looks at the almost 200 fold difference in promoter strength see by Bai *et al.* (134) when this Chinese group studied a variety of promoters. A two-fold difference in promoter strength

therefore is not significant enough to pick one promoter over another, especially where one appears phenotypically distinct from the wild type.

The findings in Section 4.3 complement the previous gel shift assays run by Peter Harrison where the implication was that MmfR is more readily released for the *mmfRL* intergenic region (L1 MARE operator) than the *mmvBY* intergenic region (L3 MARE operator).

The results from Section 4.3 indicate the potential importance of the nucleotides that are not conserved between the three MARE operators in terms of providing the differential strength of binding to MmfR. The L1 (*mmfLR*) MARE operator was shown to bind most strongly to MmfR. Nucleotides that are found in this operator but neither of the other two MARE operators are highlighted;

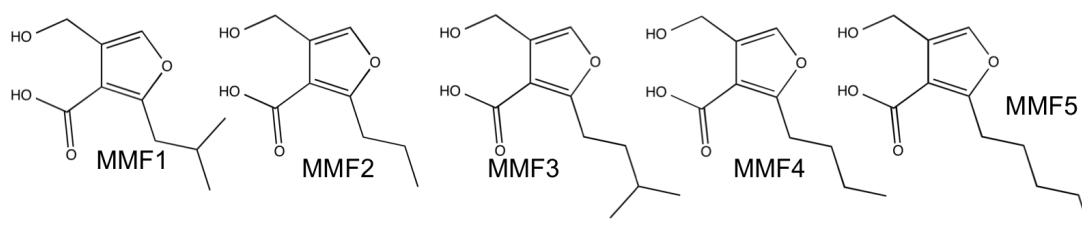


These nucleotides are potentially what give this L1 MARE operator its strength of binding and so are potential targets for future site directed mutagenesis to see how they impact the strength of MmfR binding.

## 4.4 MmfR-MMF Interactions

### 4.4.1 Different MMFs

Five methylenomycin furan ligands are naturally produced by *S. coelicolor* A(3)2. As seen in Figure 4.4, the structures of the different MMFs vary in the length and branches of the alkyl chain, and have all been shown experimentally to bind MmfR and induce methylenomycin production by previous researchers.(55)



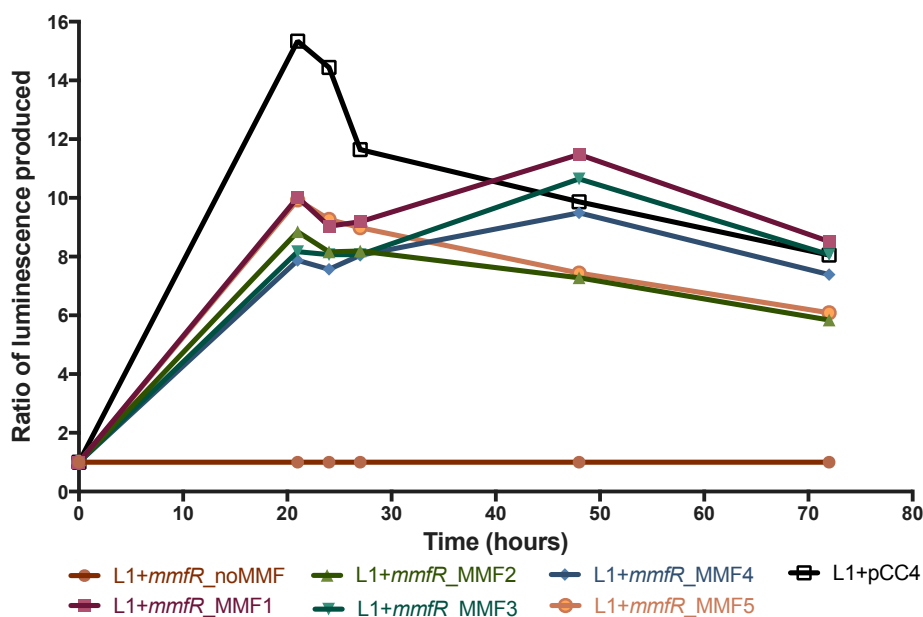
**Figure 4.4. Chemical structures of the methylenomycin furans (MMFs)**

*These five molecules have experimentally been shown to be involved in the regulation of methylenomycin biosynthesis*

When optimising the luciferase assay in Section 3.3.6, a range of different MMF concentrations of MMF2, 4 and 5 were trialled. It was found that methylenomycin furan concentrations above 5  $\mu$ M produced a detectable change in luminescence. Concentrations up

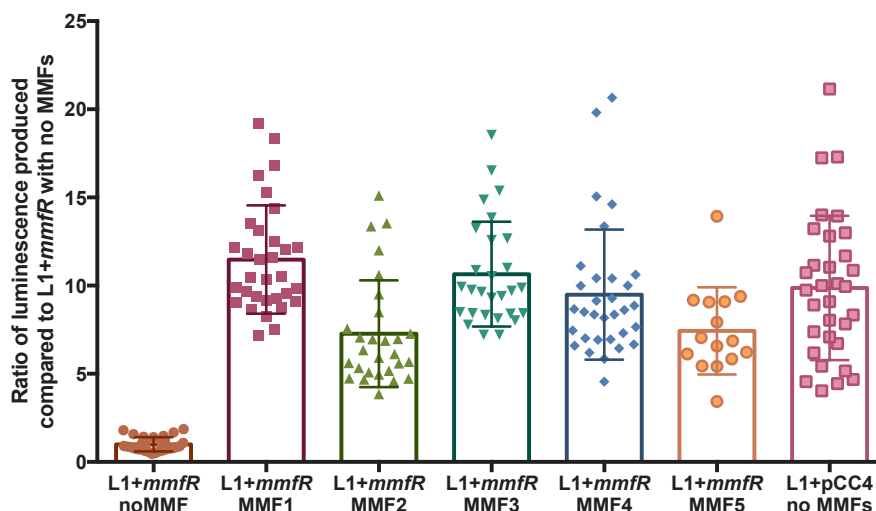
to 400  $\mu\text{M}$  were trialled and saturation of MmfR appeared to occur between 200 and 400  $\mu\text{M}$  for MMF5. From this it was decided that a final concentration of 100  $\mu\text{M}$  MMF should be used as a standard for future tests.

Each MMF was added at a final concentration of 100  $\mu\text{M}$  to the L1+*mmfR* strain and the increase in luminescence compared to the same strain with no MMFs was measured. The results from this investigation are presented in Figure 4.5 and Figure 4.6. Figure 4.5 shows the effect on luminescence of all five MMFs over five time points for 72 hours. In this figure, data is normalised against L1+*mmfR* with no MMFs and also shows a comparison with the L1+pCC4 strain (positive control, no MmfR). Figure 4.6 shows a bar chart of the data from Figure 4.5 at 48 hours only. Data from this figure is again normalised against L1+*mmfR* in the absence of MMFs. A *t*-test was run with the data collected for Figure 4.6 to establish whether there was significant release of MmfR by each of the five MMFs, this data is displayed in Table 4.2.



**Figure 4.5. Luminescence produced by the *lux* operon when under the control of *mmfLp* and MmfR upon the addition of 100  $\mu\text{M}$  of different MMFs compared to no MMFs over time**

*Luminescence produced is calculated as a ratio of the luminescence produced by the L1+*mmfR* with no MMFs, meaning that the luminescence produced by this strain has a value of one. Strains used: L1+pCC4 – positive control, luxCDABE under the control of *mmfLp*, L1+*mmfR* – luxCDABE under the control of *mmfLp* and *mmfR* under the control of *ermEp\** (*pKMS01*)*



**Figure 4.6.** Bar chart of luminescence produced by the *lux* operon when under the control of *mmfLp* and MmfR in the presence of 100  $\mu$ M of different MMFs compared to no MMFs at 48 hours

Luminescence produced is calculated as a ratio of the luminescence produced by the *L1+mmfR* with no MMFs, meaning that the luminescence produced by this strain has a value of one. Error bars show the standard deviation from the mean with all data points collected plotted. **Strains used:** same as Figure 4.5

**Table 4.2.** A *t*-test analysis of significant changes in the luminescence produced by *luxCDABE* when under the control of *mmfLp* and MmfR in the presence of different MMFs compared to no MMFs

Two tailed *t*-test with unpaired samples of equal variance used as parameters. The average induction of luminescence is calculated as a ratio of *L1+mmfR* with no MMFs (giving this sample type a value of 1).

| MMF added to <i>L1+mmfR</i> | p-value  | Significant increase? | Average induction at 48 hr (R.R.) |
|-----------------------------|----------|-----------------------|-----------------------------------|
| MMF1                        | 6.62E-26 | TRUE                  | 11.48                             |
| MMF2                        | 1.29E-15 | TRUE                  | 7.28                              |
| MMF3                        | 6.40E-24 | TRUE                  | 10.65                             |
| MMF4                        | 6.08E-18 | TRUE                  | 9.49                              |
| MMF5                        | 3.19E-17 | TRUE                  | 9.29                              |

As expected, the *t*-test analysis (see Table 4.2) showed that all five MMFs cause a significant release of MmfR from the DNA, with the *p*-value being well below 0.05 for all at 48 hours growth.

Data from Figure 4.5 and Figure 4.6 indicates that the two branched chain MMFs (MMF1 and 3) cause slightly higher levels of MmfR release than the other MMFs. Figure 4.5 shows that levels of luminescence being produced in the presence of the MMFs were anywhere between seven and 12 times that of the same strain without induction. However, despite varying

efficacy of the MMFs, at a concentration of 100  $\mu\text{M}$  none of the MMFs cause total MmfR release, with levels of luminescence still being below those produced by the L1+pCC4 strain (positive control, representing no repression).

Figure 4.6 shows the following order of efficacy for the MMF molecules at 48 hours:

**MMF1 > MMF3 > MMF4 > MMF5 > MMF2**

The biological need for five different inducer molecules with similar levels of binding to MmfR is unclear. However, an explanation may be that the MmfLHP biosynthetic enzymes are not specific enough to make a single type of MMF, thereby resulting in five similar molecules that can all bind to MmfR.

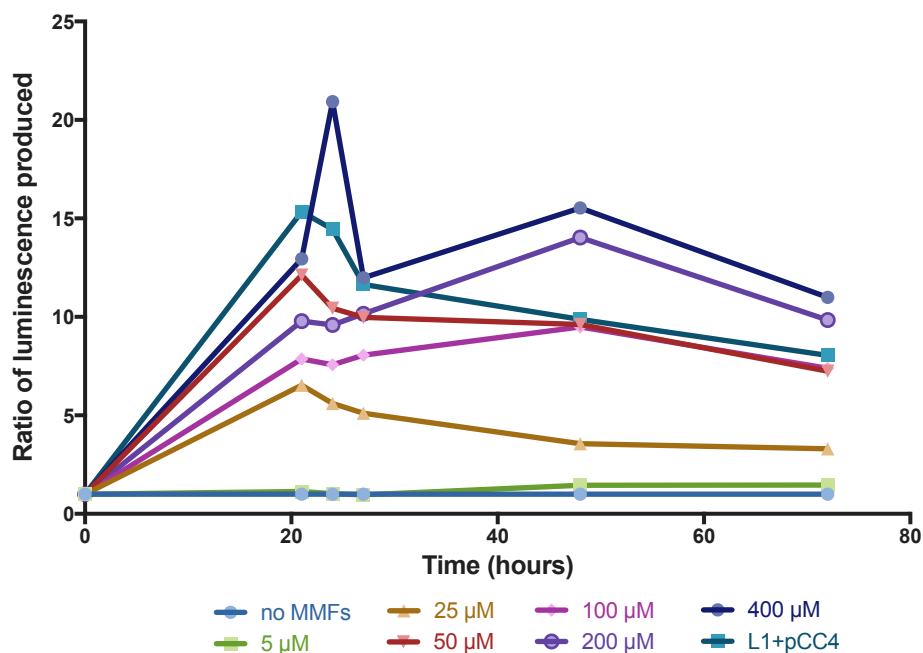
Data for 72 hours growth also showed a significant increase in luminescence in the presence of the five MMFs (data not shown here) and measurements at this time point produced the same order of efficacy by the different MMFs as was seen at 48 hours.

#### 4.4.2 Different Concentrations of MMFs

Once it was established that the addition of all five MMFs produced significant changes in the luminescence produced by the L1+*mmfR* strains and could be suitably analysed using the luciferase assay, further investigations were carried out to look at the binding affinities of each MMFs, calculating the  $K_d$  and  $B_{\text{max}}$  values from a standard curve. The  $B_{\text{max}}$  represents the maximum luminescence produced by L1+*mmfR* in the presence of the MMFs as a relative ratio of same strain with no MMFs whereas the  $K_d$  represents the equilibrium binding constant, the concentration (in  $\mu\text{M}$ ) needed to achieve half the maximum binding of the MMFs to MmfR. The equation for a standard curve is as follows;  $Y = B_{\text{max}} \times X / (K_d + X)$ , where Y is the ratio of luminescence and X is the final concentration of the MMFs.

MMF4 is used as an example of how these binding potential for each of the MMFs were determined and a full set of data is shown in Figure 4.7 and Figure 4.8. The same analysis was then done for the other four MMFs and the data summarised in Figure 4.9, Figure 4.10 and Table 4.3.

Figure 4.7 shows the data collected for the L1+*mmfR* strain in the presence of six different concentration of MMF4 over 72 hours compared to a negative and positive control. Figure 4.8 shows a standard curve produced using luminescence readings from the just 48-hour time point.



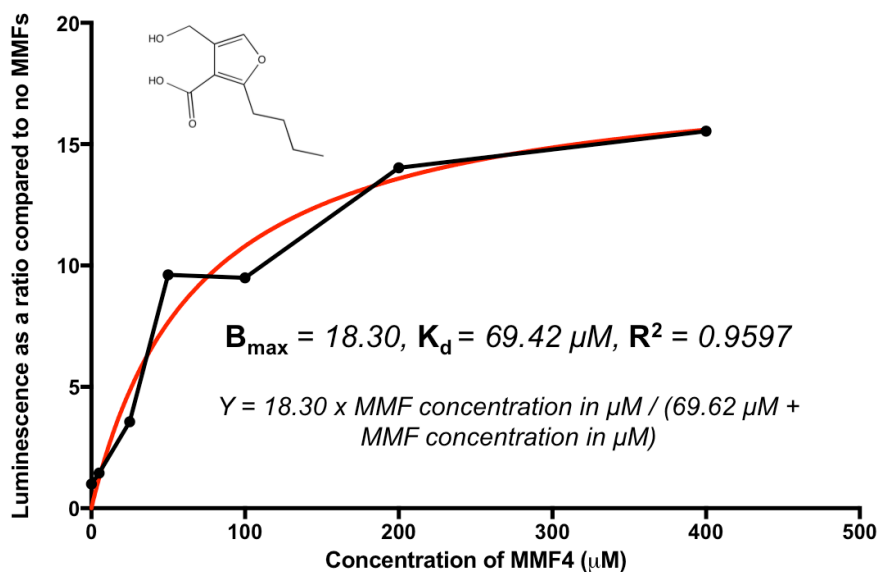
**Figure 4.7. Dose-response of MMF4 over time, in strains where *lux* operon expression is controlled by *mmfLp* and *MmfR***

*Luminescence produced is calculated as a ratio of the luminescence produced by the L1+mmfR with no MMFs. Strain used: L1+mmfR – luxCDABE under the control of mmfLp (L1) and mmfR under the control of ermEp\* (pKMS01)*

Figure 4.7 revealed that 400 µM MMF4 appeared to cause the greatest release of MmfR from the MARE operator, producing up to 20 times the luminescence of the non-induced strain. In fact, at a number of time points, 400 µM MMF4 being added to L1+*mmfR* produced higher levels of luminescence than even the positive control for the same operator sequence (L1+pCC4). It is not entirely clear why this is the case. Possibly the reason for this is the margin of error for the data collected, which does have some overlaps (not shown here). Alternatively, the MMFs may have a separate activator role as well as directly triggering the release of MmfR. Possibly the MMFs can be used to recruit sigma factors or other transcriptional regulators, but there is no experimental evidence of this being the case as yet. Further investigation would be needed to test this theory, particularly as the presence of the MMFs caused no significant change in the luminescence produced by the positive control strains (Section 3.3.5).

Data collected for final concentrations between 50 µM and 200 µM appear to be very similar, particular at later time points, indicating that MmfR/MMF binding is saturated and a threshold level of MMFs has been reached that is enough to activate the total biosynthetic pathway.

In Figure 4.7, at the time points around 24 hours, there does not appear to be a direct correlation between concentration and luminescence. Some lower concentrations appear to produce greater levels of luminescence than the higher concentrations. There was a huge deviation between these earlier results however, so this lack of correlation in some cases may just be an artefact of biological variation and the cells stabilising in growth. At 48 and 72 hours however, there is a direct correlation between MMF concentration and luminescence produced as well as there being a lower standard deviation and coefficient of variance between results which makes this data more suitable for comparisons. For this reason, these later time points were used to create a standard curve and determine the  $B_{max}$  and  $K_d$  values in Figure 4.8.



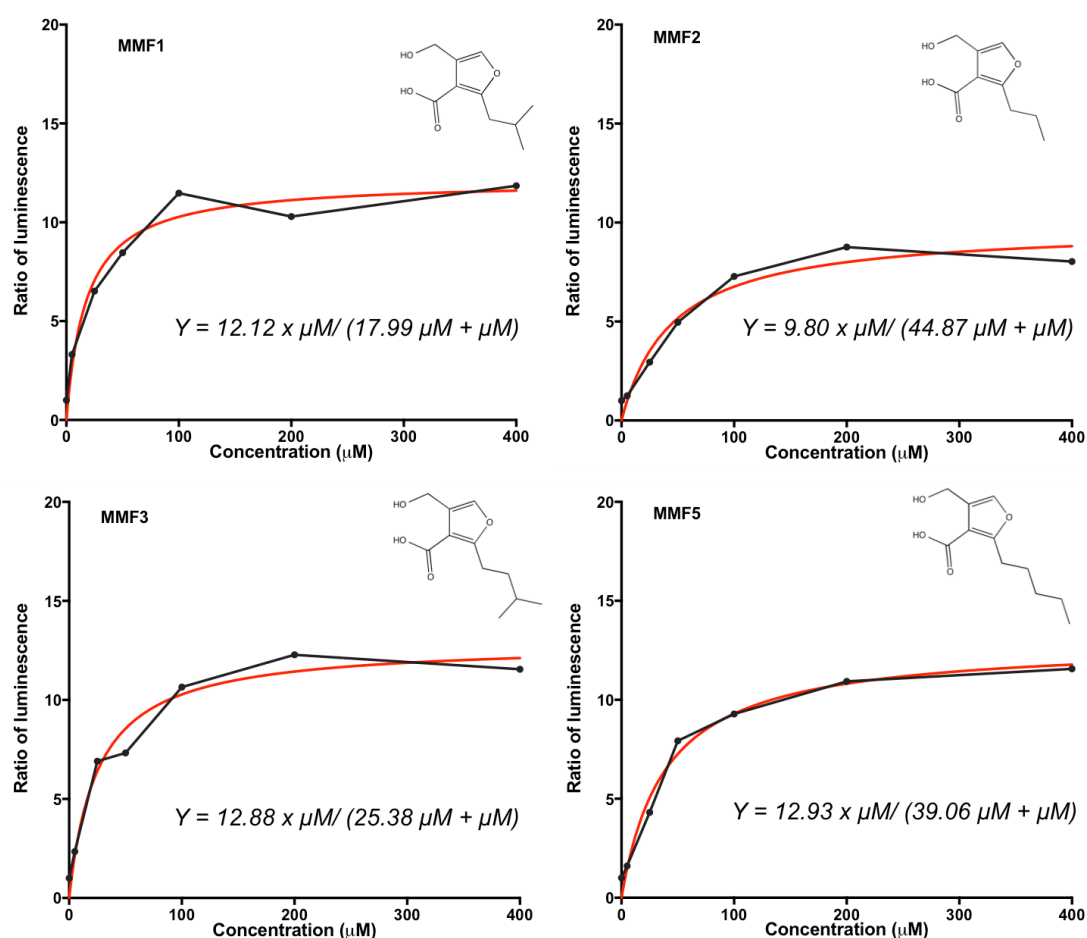
**Figure 4.8. Dose-response curve of MMF4 at the 48 hour time point, in strains where *lux* operon expression is controlled by *mmfLp* and MmfR**

*Shown in black is the actual data for each MMF with the fitted standard curve shown in red.*

**Strain used:** *L1+mmfR – luxCDABE* under the control of *mmfLp* (*L1*) and *mmfR* under the control of *ermEp\** (*pKMS01*)

In Figure 4.8, the fitted hyperbolic curve to the data is shown in red, this has an  $R^2$  value of 0.9597, indicating a very good fit of the model to the data (where 1 is a perfect fit and 0 is no relationship). The equation for this standard curve is  $Y = B_{max} * X / (K_d + X)$ , where  $Y$  is the relative ratio of luminescence at a particular concentration compared to the same strain with no MMFs. The  $B_{max}$  for this data is 18.3. In other words, the maximum amount of luminescence produced by adding high concentrations of MMF4 is 18.3 times the luminescence produced when there is no MMF added. The  $K_d$  reveals that an average of 69.42  $\mu\text{M}$  MMF4 is expected to achieve half the  $B_{max}$ .

Following on from the analysis of different concentrations of MMF4, Figure 4.9 shows the standard curves collected for MMF1, 2, 3 and 5 at 48 hours. Again, shown in black is the actual data for each MMF with the fitted standard curve shown in red. To make comparison of one MMF with another easier, the fitted standard curves for all five MMFs have been combined into a single chart, shown in Figure 4.10. The details of the  $B_{max}$  and  $K_d$  and the relative binding potential for each of the MMFs are shown in Table 4.3. The binding potential is calculated by dividing the  $B_{max}$  by the  $K_d$  and is representative of how ‘good’ each MMF is as a ligand for MmfR.(136) It is a value that combines both the availability of MmfR ligand binding pockets and the affinity each MMF has for the residues in this pocket.



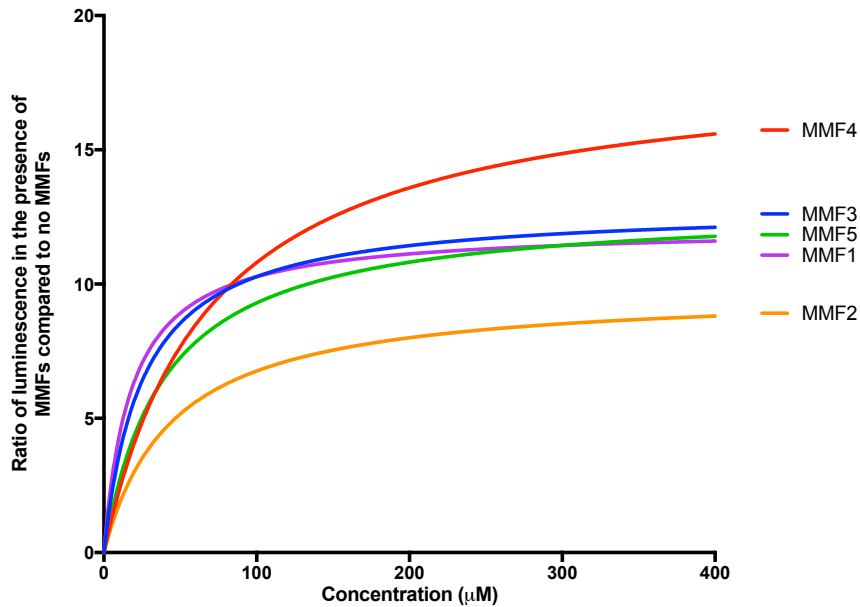
**Figure 4.9. Dose response curves for MMF1, 2, 3 and 5 at 48 hours growth using strains where *lux* operon expression is controlled by *mmfLp* and MmfR**

Shown in black is the actual data for each MMF with the fitted standard curve shown in red.

$$Y = B_{max} \times \text{micromolar concentration} / (K_d + \text{micromolar concentration})$$

**Strain used:** *L1+mmfR* – *luxCDABE* under the control of *mmfLp* (*L1*) and *mmfR* under the control of *ermEp\** (*pKMS01*)





**Figure 4.10. Compiled dose-response standard curves showing the relationship between the concentration of all five MMFs and the luminescence produced in strains where *lux* operon expression is regulated by *mmfLp* and MmfR, at 48 hours growth**

In Figure 4.9 and Figure 4.10 it can be seen that different MMFs have different  $B_{max}$  values. It is unclear why this is the case but again, one possible explanation is the possible involvement of unknown regulatory mechanisms, genetically separate to the methylenomycin gene cluster and therefore still present in the *S. coelicolor* M145 control strain which are effected by the MMFs. It should be noted that the L1+pCC4 positive control strain will give an average maximum ratio of luminescence (compared to L1+*mmfR* with no MMFs) of around 10. All of the MMFs, except MMF2, have an average  $B_{max}$  above this, further indicating that the MMFs are doing more than just releasing MmfR from the MARE operator.

If these MMFs do have an extra role you may also expect to see a difference in luminescence produced by the L1+pCC4 positive control strain in their presence. However, L1+pCC4 in the presence and absence of the five MMFs gives statistically similar luminescence readings at the time points and concentrations trialled (Section 3.3.5). How exactly this extra functionality of the MMFs works is therefore unclear. Potentially MmfR needs to be present and bound to the MMFs for them to have this extra regulatory function. Alternatively, this strain was only tested with a maximum concentration of 100 µM of each MMF and maybe a higher threshold concentration of MMFs is needed to cause a significant change in the luminescence produced by the L1+pCC4 control strain.

**Table 4.3. Binding kinetics data for MmfR with each of the five methylenomycin furans including the  $B_{max}$ ,  $K_d$  and binding potential values for each at 48 hours growth**

$$\text{Binding potential} = B_{max}/K_d$$

| <i>Analysis</i>                                    | <i>MMF1</i> | <i>MMF2</i> | <i>MMF3</i> | <i>MMF4</i> | <i>MMF5</i> |
|--|-------------|-------------|-------------|-------------|-------------|
| <b><math>B_{max}</math></b>                        | 12.12       | 9.796       | 12.88       | 18.30       | 12.93       |
| <b><math>K_d</math> (<math>\mu\text{M}</math>)</b> | 17.99       | 44.87       | 25.38       | 69.42       | 39.06       |
| <b>Binding potential</b>                           | 0.6737      | 0.2183      | 0.5074      | 0.2636      | 0.3310      |

From these findings the following order of efficacy for the MMFs is given as;

**MMF1 > MMF3 > MMF5 > MMF4 > MMF2**

This is very close to the first estimation made from preliminary work in Section 4.4.1. Only MMF4 and MMF5, which already gave very similar results, have swapped places. It can be seen from Table 4.3 that the two branched chain MMFs (MMF1 and MMF3) have the highest binding potentials. Following on from this, the MMFs with the longest straight alkyl chains have the highest binding potentials. The MMF with the lowest binding potential was MMF2, which has the shortest alkyl chain. This indicates that the bigger and longer alkyl chains are needed to best fit the MmfR binding pocket. The findings in this investigation therefore suggest that either MMF1 or MMF3 are most suitable for use when developing the novel inducible expression system.

### ***In Vitro* Work - Data from Gel Shift Assays**

Interestingly, *in vitro* work was carried out by another lab member, Shanshan Zhou, to analyse the binding of the five MMFs to MmfR. For their investigation they performed a gel electrophoresis mobility shift assay (EMSA) using purified MmfR and each of the five furan ligands. This showed a very similar pattern of binding abilities between the five furans as the *in vivo* data collected from the luciferase assay. The main differences in the data were that MMF1 and 3 gave very similar shifts to one another as did MMF4 and 5. An EMSA did not provide the resolution to distinguish between possible binding differences in these two pairs of ligands.

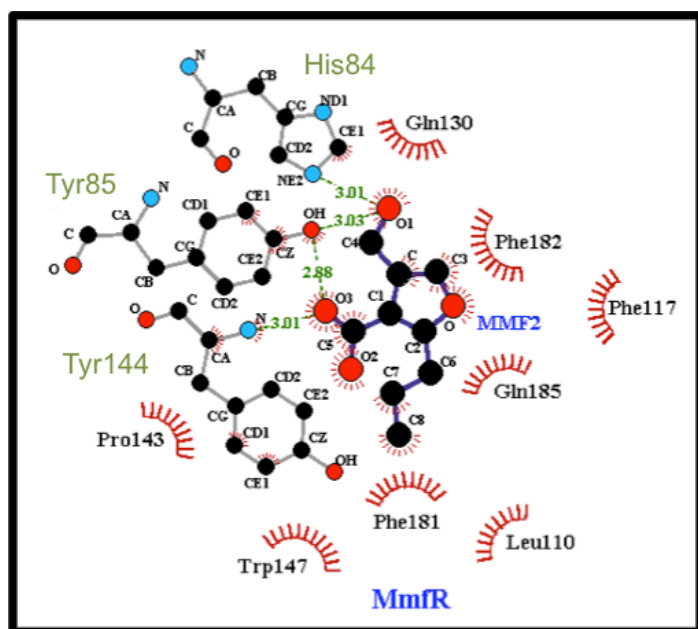
Binding profiles established from the EMSA were as follows;

**MMF1  $\approx$  MMF3 > MMF5  $\approx$  MMF4 > MMF2**

It is promising that both *in vivo* and *in vitro* data give a very similar pattern of MmfR/MMF binding affinities.

#### 4.4.3 *In Silico* Analysis of MmfR Ligand Binding Pocket

Following on from the luminescence assay investigation of MmfR/MMF binding, the amino acids directly involved in ligand binding are of interest. A crystal structure of MmfR has previously been solved both in the apo form and with MMF2 bound. (73) This allowed analysis of the ligand-binding site at an atomic resolution. Shown in Figure 4.11 is the crystal structure found for MmfR with MMF2. This was then compared with a protein:ligand docking analysis carried out using 'SwissDock' for the apo crystal structure of MmfR and MMF2 and it was found to be a close match, indicating that an *in silico* docking may be a useful and accurate way of studying MmfR/MMF binding. The docking was therefore done with the apo crystal structure of MmfR and all five MMFs as well as synthetic analogue molecules 70 and 121 and *S. coelicolor* hormone SCB1 as controls, using the SwissDock online software with the default parameters. A summary of these results is found in Table 4.4 and the structures of the ligands are displayed in Figure 4.12.



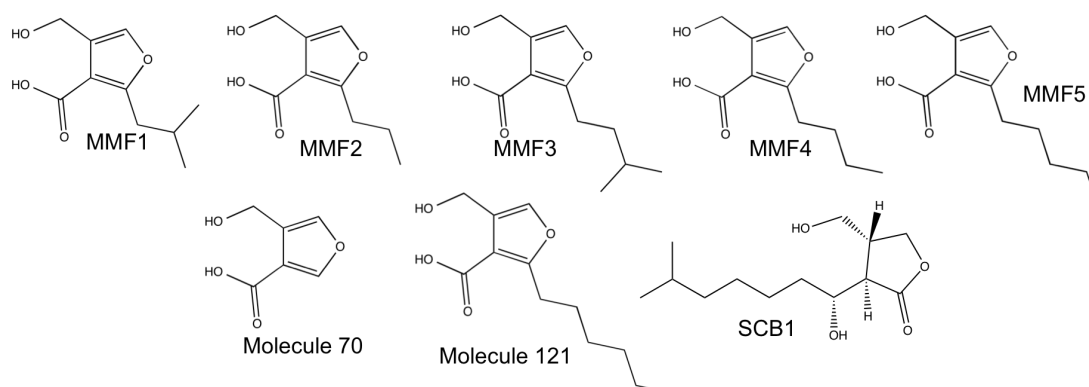
**Figure 4.11. LigPlot<sup>+</sup> schematic of MmfR bound to MMF2, as shown by X-ray crystallography data**

*The schematic created for this figure was done using the LigPlot<sup>+</sup> software, designed by Wallace et al. (116) and is based on the protein work by Rea et al. (73)*

SCB1 is a *S. coelicolor* butyrolactone signalling molecule, known to bind to the TetR family member ScbR. This SCB is structurally distinct from the MMFs (Figure 1.3 and Figure 4.12) and has been shown experimentally *in vitro* not to bind MmfR and so works as a negative control for the docking analyses (see results from Section 3.3.5). Molecule 70 is a synthetic MMF analogue that can also be used as a negative control, it has the same core structure as

the MMFs but lacks the alkyl chain, which has been shown to be necessary for inducing antibiotic production.(55) Molecule 121 is another synthetic MMF analogue, again it has the same core structure as the MMFs but its alkyl chain is one carbon longer than MMF5. The nomenclature for these synthetic molecules comes from the work of Nicolas Malet at the University of Warwick.(55)

A docking analysis was also attempted with MmyR and ScbR for comparison with MmfR but without access to crystal structures for these TetR family members the level of accuracy achieved from protein modelling was not good enough to get realistic results from a docking analysis.



**Figure 4.12. Chemical structures of the MMFs, SCB1 and other synthetic analogues**

**Table 4.4. Summary of docking analysis of MmfR with the MMFs and other synthetic and natural analogues**

*Key: (?) - Sometimes it was unclear whether a particular amino acid was involved in ligand binding, being present in some models but not all. This has been denoted with a question mark.*

| <i>Ligand</i>       | <i>Scoring function (Best <math>\Delta G</math>)</i> | <i>Amino acids involved in H-bonding</i>                 | <i>No. of H-bonds</i> |
|---------------------|--|--|-----------------------|
| <b>MMF1</b>         | -7.24  | Histidine 84 (?), Tyrosine 85, Tyrosine 144              | 4                     |
| <b>MMF2</b>         | -7.17  | Histidine 84 (?), Tyrosine 85, Tyrosine 144              | 4                     |
| <b>MMF3</b>         | -7.03  | Tyrosine 85, Tyrosine 144                                | 2                     |
| <b>MMF4</b>         | -6.86  | Tyrosine 85, Tyrosine 144                                | 3                     |
| <b>MMF5</b>         | -6.88  | Tyrosine 85, Tyrosine 144                                | 2                     |
| <b>Molecule 70</b>  | -6.43  | Tyrosine 85, Tyrosine 144                                | 2                     |
| <b>Molecule 121</b> | -6.13  | No realistic matches, Glycine 94(?)                      | 1                     |
| <b>SCB1</b>         | -6.33  | No realistic matches, Arginine 87 (?), Aspartate 120 (?) | 2                     |

It appears that tyrosines 85 and 144 are consistently involved in hydrogen binding with the furan ligands in the models and so are likely to be crucial to the conformational change

leading to the release of the MmfR protein from the MARE operator sequence of DNA. Histidine 84 may also play a role but was not constantly found between different models and so seems to be less important. There are also a number of other amino acids that are consistently involved in hydrophobic interactions with the ligand in the MmfR binding pocket (shown in Figure 4.13).

*In silico*, molecule 121 appears to not bind MmfR, despite it being shown to induce methylenomycin production in previous *in vivo* experimentation by Nicolas Malet,(55) with the implication from his research that molecule 121 can cause the conformational change in MmfR that will release it from the DNA. This disparity indicates that the simulations from this *in silico* analysis are only estimations and cannot be taken as final proof of what occurs *in vivo*. The lack of flexibility of MmfR and the ligand molecule during the docking is the most likely cause of the anomaly.

Molecule 70 is lacking the alkyl chain usually found in the MMFs. *In vivo* it has been shown not to induce methylenomycin production. The docking shows that molecule 70 can still bind to the active region of MmfR but it would seem that with the lack of alkyl chain, these hydrogen bonds alone would not be enough to induce a conformational change to MmfR. *In vivo*, there would also be water molecules present in this docking and so potentially such a small molecule would not bind to the active site. As expected, the docking done with SCB1 showed no convincing binding models to MmfR.

Results from the luciferase assay done to study MmfR/MMF binding indicated that MMF1 and 3 worked best to release MmfR from the MARE operator sites. The *in silico* analysis did not show exactly the same pattern however. As just mentioned, there are a number of limitations of the *in silico* docking analysis and so the *in vivo* data is therefore likely to be more representative of the wild type MmfR/MMF binding profiles. The docking analysis is however, useful in narrowing down the selection of residues for mutagenesis to further study the MmfR ligand binding pocket.

The 11 amino acids involved in hydrogen binding and hydrophobic interactions between MmfR and the MMFs have all been highlighted in Figure 4.13 to show how they are spread across the primary structure of MmfR. It was then possible to compare these residues to motifs found in MmfR orthologues.

```

> MmfR [Streptomyces coelicolor A3(2)]
MTSAQQPTPFVAVRSNVPRGPHPPQERSIKTRAQILEAASEIFASRGYRGASVKDVAERVGMTKGAVYFHF
PSKESLAIIVVEEHYARWPAAMEEIRIQGFTPLETVEEMLHRAAQAFRRDDPVMQAGARLQSERAFIDAEL
PLPYVDWTHLLEVPLODAREAGQLRAGVDPAAAARSLVAAFTGMQHVSDNLHQRADIMERWQELRELMFF
ALRA
    
```

**Figure 4.13. Positions of ligand binding residues in the MmfR amino acid sequence**  
 Yellow – hydrogen bonds, Green – hydrophobic interactions  
 This is based on the crystal structure of MmfR/MMF2, solved by Dean Rea et al. (73)

A number of other systems in different *Streptomyces* strains encode TetR family proteins, orthologous to MmfR including SAV\_2270 from *S. avermitilis* and SgnR from *S. venezuelae*. The presence of the two probable key ligand binding tyrosine residues in these streptomycete homologues would further increase the likelihood that these amino acids are indeed involved in ligand binding and therefore should be studied further.

The key MmfR residues in ligand binding are spread across the entire amino acid and so entire sequence alignment and comparison was carried out between MmfR and orthologues SAV\_2270 and SgnR. The results from this alignment are displayed in Figure 4.14. Matches to the MmfR polypeptide sequence are shown in blue with mismatches shown in yellow.

|             |  |
|-------------|--|
| <b>MmfR</b> | ---MTSA--QQPTPFVAVRSNVPRGPHPPQERSIKTRAQILEAASEIFASRGYRGASVKDV  |
| SAV_2270    | MDVMSSERNQSTRLPAGTTFDDTAHLKQORAIRTRGTILNAAAAAFATDGFPOVTIKDI    |
| SgnR        | -----MATPRSOPKQERARRTKVHILQSAAELFAERGYATVTLQDV                 |
|             | * * : * : : * : : * : : *                                      |
| <b>MmfR</b> | AERVGMTKGAVYFHFPSKESLAIIVVEEHYARWPAAMEEIRIQGF-TPLETVEEMLHRAA   |
| SAV_2270    | ADGAEMTKGAVYFHFNPKEALAVAVLEEYRRMQEAVNGALEHGDPTSPTTVDDVMRRLA    |
| SgnR        | AERAEMTKGAVYFHYTNKEALAVAVVQEHYARWPEILKGAEGDHA-EPFDMLTAVLDTVT   |
|             | * . ***** * : . * : * * : : : . : : :                          |
| <b>MmfR</b> | QAFRRDDPVMQAGARLQSERAFIDAELPLPYVDWTHLLEVPLOD----AREAGQLRAGVD-  |
| SAV_2270    | RAFHEDEVFIHAGARLQIERPYIKAELPVPYVGT LKVLTELLDQ----CRTAGNLPKSTK- |
| SgnR        | RAFARDIVVQAGARLQIERALIDAELPEPYVGTWEDYLTRLIAE----ARDAGQLRDGVE-  |
|             | : : * . : * . * : : * . * : : . * : : .                        |
| <b>MmfR</b> | -----PAAAARSLVAAFFGMQHVSDN-LHQRA--DIMERWQELRELMFFALRA----      |
| SAV_2270    | -----PEALARALGSAVFGAQHISWV-LNDRE--DIVERVEEIIDAFVPLH-----       |
| SgnR        | -----PRAAARVLVSAFFGMQHISDV-LSGRS--DLTERYEELRTVLLLEGLRR----     |
|             | * . . : * : . : * . : : * : : : :                              |

**Figure 4.14. Comparison of MmfR ligand binding amino acid residues with the sequences of orthologous TFRs SgnR and SAV\_2270**  
 Blue – matches with MmfR, Yellow – mismatches with MmfR  
 Amino acid sequences used: MmfR – methylenomycin cluster transcriptional repressor, SgnR – MmfR homologue from *Streptomyces venezuelae*, SAV\_2270 – MmfR homologue from *Streptomyces avermitilis*

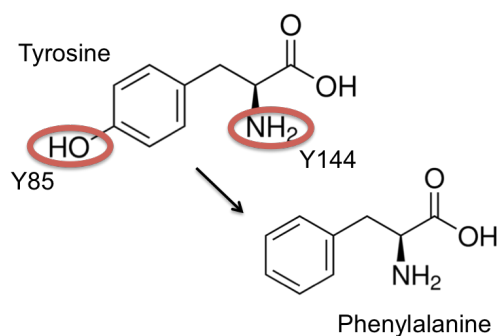
It appears that tyrosine 85 and 144 are conserved across the three orthologous amino acid sequences further suggesting that they are indeed crucial to ligand binding and are appropriate targets for site directed mutagenesis to further demonstrate their role in ligand binding.

A number of the amino acids shown to be involved in hydrophobic interactions with the MMF ligand for MmfR are also conserved in its orthologues, especially SgnR. It is possible therefore that these orthologues will interact with MMF-like ligands as the similarities in their ligand binding pockets mean that they could provide similar interactions with the acyl chain of the MMFs.

#### 4.4.4 Effect of Site Directed Mutagenesis on the Ligand Binding Domain of MmfR

Site directed mutagenesis was carried out on the *mmfR* gene to create mutants with either Y84 or Y144 converted to a phenylalanine residue in the MmfR protein using the Agilent QuikChange Lightning Site Directed Mutagenesis Kit. Figure 4.15 shows details of which atoms from the MmfR tyrosine residues are involved in MMF binding and how this will be effected in the phenylalanine mutants.

As can be seen in this figure and Figure 4.11, it is the OH group of Y85 which interacts directly with the MMF molecule and the NH<sub>2</sub> group of Y144. The NH<sub>2</sub> group will still be present in the same position for the Y144F mutant. The conversion of tyrosine to phenylalanine will also not considerably change the size of the binding pocket and so it is expected that little change will be seen from this mutation. The NH<sub>2</sub> group will be present in all amino acids so it would be very hard to create a mutant that will properly check for the function of Y144 in ligand binding. The Y144F mutant therefore worked as a negative control compared to the results from the Y85F mutant. In the Y85F mutants, it is expected that a more considerable effect on ligand binding will be seen due to the absence of the key OH group. In this Y85F mutant, the similarity in size between tyrosine and phenylalanine and the presence of the benzene ring should minimize the effects of changing the size and conformation of the binding pocket, allowing the analysis of just the hydroxyl group and its role in ligand binding.



**Figure 4.15. Schematic highlighting the hydroxyl and amine groups of tyrosines 85 and 144 from MmfR involved in MMF binding**

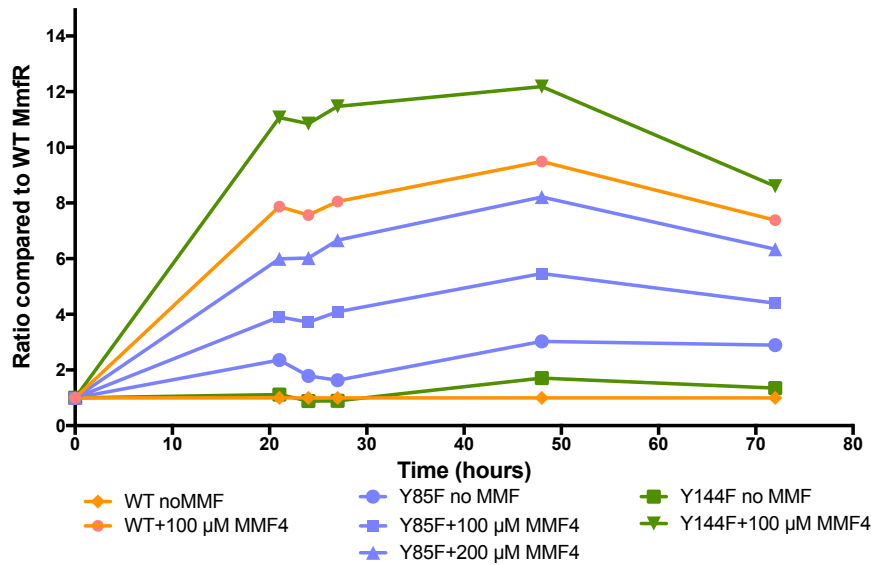
*This schematic is based on the crystal structure of the MmfR/MMF2 complex. (73)*

Once the *mmfR* sequence was mutated in *E. coli* to now code for phenylalanine instead of tyrosine, the sequence was inserted into the pCC4 vector where the wild type *mmfR* was normally inserted, creating pKMS85 and pKMS144. This could then be integrated into the *Streptomyces coelicolor* M145 genome along with the L1 *luxCDABE* vector, allowing the mutants to be analysed using the luciferase reporter assay. This assay was carried out as with the wild type MmfR and a final concentration of 100  $\mu$ M or 200  $\mu$ M MMF was added to cultures and the luminescence produced compared to no MMFs measured. The results collected from the initial trials with the Y85F and Y144F mutants can be found in Figure 4.16 and Figure 4.17. Figure 4.16 shows the luminescence produced at five time points over 72 hours by L1+WT*mmfR*,<sup>1</sup> L1+*mmfR* Y85F and L1+*mmfR* Y144F in the presence and absence of MMF4.<sup>2</sup> Figure 4.17 is a bar chart of data from Figure 4.16 at the 48-hour time point only. A *t*-test analysis of data can be found in Table 4.5 with L1+WT*mmfR* and 100  $\mu$ M MMF4 being compared to the Y85F and Y144F mutants with MMF4 to look for significant differences in ligand binding properties.

<sup>1</sup> This is the same strain as was referred to earlier as simply L1+*mmfR*.

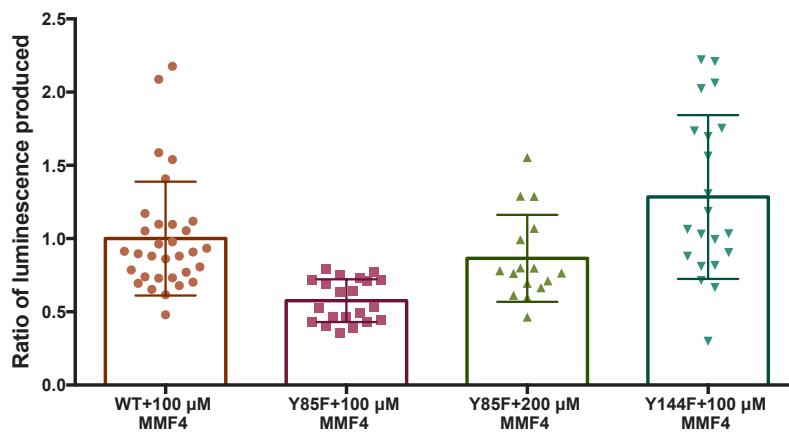
<sup>2</sup> MMF4 is used as it has been synthesised on a larger scale than the other MMFs and so was more readily available.





**Figure 4.16. Luminescence produced by the *lux* operon when under the control of *mmfLp* and MmfR tyrosine mutants compared to wild type MmfR, in the presence and absence of MMF4**

Luminescence produced is calculated as a ratio of the luminescence produced by the *L1+WTmmfR* with no MMFs. **Strains used:** *L1+WTmmfR*, *L1+mmfR Y85F* and *L1+mmfR Y144F* – *luxCDABE* under the control of *mmfLp* (*L1*) and *mmfR* under the control of *ermEp\** producing either the wild type MmfR, a Y85F mutant or a Y144F mutant (*pKMS01*, *pKMS85* and *pKMS144* respectively).



**Figure 4.17. Boxplot of luminescence produced by the *lux* operon when under the control of *mmfLp* and MmfR tyrosine mutants, compared to wild type MmfR measured in the presence of MMF4 at 48 hours**

Luminescence produced is calculated as a ratio of the luminescence produced by the *L1+mmfR* with no MMFs, meaning that the luminescence produced by this strain has a value of one. Error bars are shown as the standard deviations of data with all data points collected also shown on the chart. **Strain used:** same as Figure 4.7

**Table 4.5. A *t*-test analysis luminescence produced by *luxCDABE* when under the control of *mmfLp* and the Y85F and Y144F MmfR mutants in the presence of 100  $\mu$ M or 200  $\mu$ M MMF4 to look for significant differences in the removal of repression**

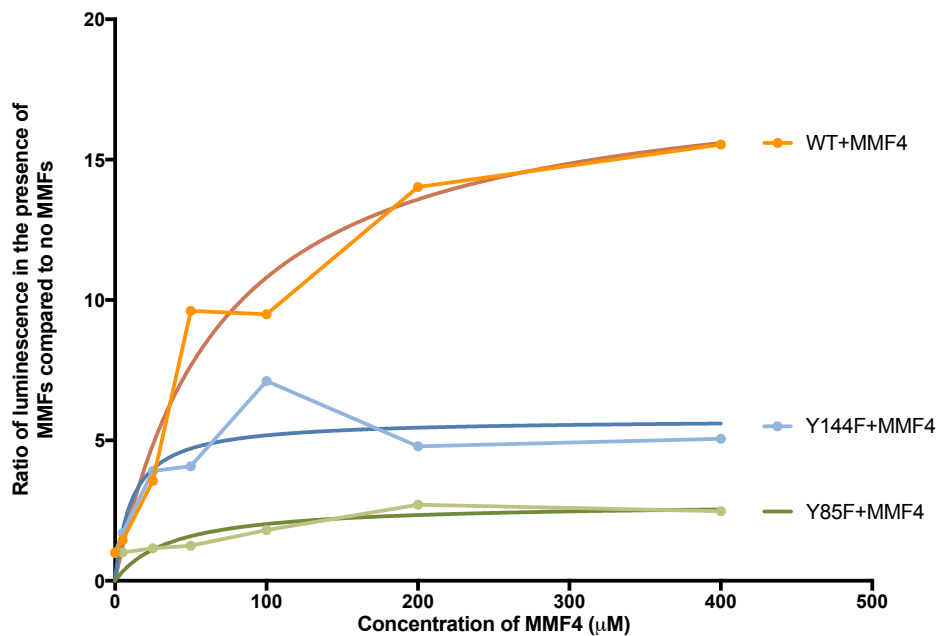
| <i>Strain used</i>         | <i>p-value</i> | <i>Significant difference?</i> | <i>Average induction at 48 hr compared with L1+WTmmfR and 100 <math>\mu</math>M MMF4 (R.R.)</i> |
|----------------------------|----------------|--------------------------------|---|
| Y85F and 100 $\mu$ M MMF4  | 1.60E-5        | YES                            | 0.576   |
| Y85F and 200 $\mu$ M MMF4  | 2.30E-1        | NO                             | 0.865   |
| Y144F and 100 $\mu$ M MMF4 | 3.36E-2        | YES                            | 1.284   |

The Y85F strain did show a lower level of de-repression by MMF4 compared to the wild type strain. Only at twice the level of MMF4 did the Y85F strain produce statistically similar levels of induction to those seen in the wild type. As predicted, the Y144F mutation did not cause a significant reduction in release of MmfR by MMF4. Interestingly it did actually appear to produce significantly higher levels of luminescence upon the addition of 100  $\mu$ M MMF4, with more than 125% of luminescence of the wild type strain with MMF4 at 48 hours (see Figure 4.17). Whether this mutation has indeed ‘optimised’ the binding pocket and made it more sensitive is as yet unclear but would be an interesting idea to investigate further in the project. It was an artefact also identified by Shanshan Zhou when running an *in vitro* gel shift assay, further indicating that a version of MmfR with improved ligand binding has indeed been produced.

Another observation from Figure 4.16 is that neither the Y85F or Y144F mutants appear to be as good at repressing luminescence as the wild type MmfR at the *mmfLR* intergenic region with both L1+*mmfR* Y85F and L1+*mmfR* Y144F producing higher levels of luminescence L1+WT*mmfR* in the absence of the MMFs. Y144F appears to be a better repressor than Y85F however. Although the mutations were to the ligand binding pocket, they are close to the dimer interface and so could be also effecting the overall structure of the protein and therefore its DNA binding properties.

Data from Figure 4.17 in particular, reveals large standard deviations and huge overlaps in the error bars. A *t*-test did show that the variation in the different sets of results were statistically significant but it is unclear whether from this data alone, reliable conclusions can be made about the activities of the mutant MmfRs compared to the wild type. For this reason it was decided that a range of concentrations of MMF4 would be trialled and the  $B_{max}$  and  $K_d$  values derived in the hope of achieving some more distinct differences between samples. To obtain the  $B_{max}$  and  $K_d$  values, MMF4 was added at the same ranges of concentrations as were added to the wild type MmfR strains in Section 4.4.2.

The standard curves collected for MMF4 binding to L1+*mmfR* Y85F and L1+*mmfR* Y144F compared to earlier data collected for L1+WT*mmfR* with MMF4 are shown in Figure 4.18. Using this data, the  $B_{max}$ ,  $K_d$  and binding potential of MMF4 to each of the mutants was calculated. These values have been compared to those from the wild type MmfR in Table 4.6.



**Figure 4.18. Standard curve for production of luminescence as a result of MMF4 binding to and releasing the wild type MmfR compared to tyrosine 85 and 144 mutants**

*Strain used: L1+WTmmfR, L1+mmfR Y85F and L1+mmfR Y144F – luxCDABE under the control of mmfLp (L1) and mmfR under the control of ermEp\* producing either the wild type MmfR, a Y85F mutant or a Y144F mutant (pKMS01, pKMS85 and pKMS144).*

**Table 4.6. Binding kinetics data for MMF4 binding wild type MmfR compared to the tyrosine 85 and 144 MmfR mutants, including the  $B_{max}$ ,  $K_d$  and binding potential values for each at 48 hours growth**

$$\text{Binding potential} = B_{max}/K_d$$

| L1+ <i>mmfR</i> strain | $B_{max}$ | $K_d$ ( $\mu\text{M}$ ) | Binding potential |
|------------------------|-----------|-------------------------|-------------------|
| WT and MMF4            | 18.3      | 69.42                   | 0.26              |
| Y85F and MMF4          | 2.8       | 37.51                   | 0.07              |
| Y144F and MMF4         | 5.8       | 11.13                   | 0.52              |

As can be seen in Table 4.6, the binding potential of the Y85F MmfR mutant to MMF4 is much lower than the wild type binding to the same ligand. This data further confirms that the hydroxyl group of the tyrosine residue in position 85 of MmfR is likely to be necessary for ligand binding and the resultant conformational change causing its release from the MARE operator. The Y144F mutant on the other hand appears to have a binding potential of almost double that of L1+WT*mmfR* with MMF4. It again appears therefore that this mutation has optimised the repressor, reducing the amount of MMFs needed to achieve de-repression. Further investigation is needed to establish whether the same would be seen for all five MMFs.

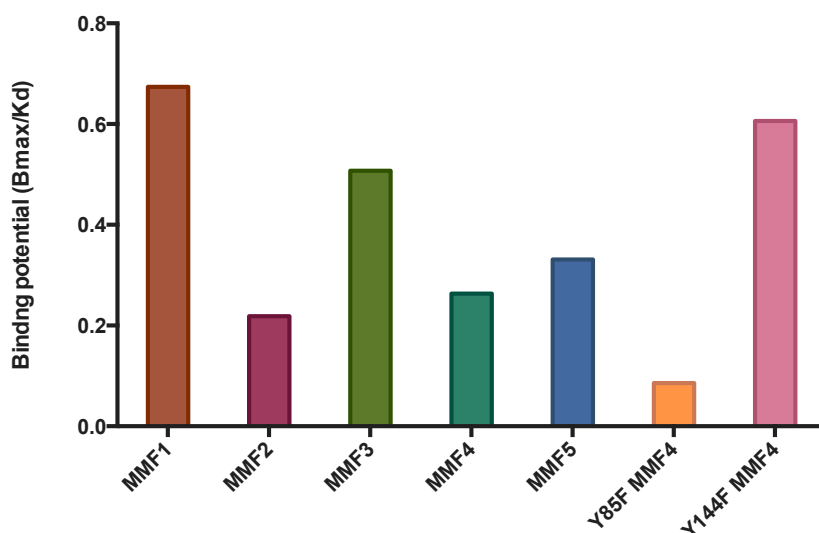
It is also of note that both the Y85F and Y144F mutants have a lower  $B_{max}$  than the wild type. This indicates that for both of these mutants, there were differences in the level of MmfR release from the MARE operator that can be achieved the presence of the MMFs at a saturating concentration. The differences in DNA binding properties of the mutants is still poorly understood however.

### Alanine Mutants

Alanine mutants of the tyrosine residues were also created by another group member and were used to carry out gel shift assays. It appears that despite the  $\text{NH}_2$  group being present in alanine, the change the shape of the binding pocket by the smaller amino acid was enough to significantly alter MMF binding in the Y144A mutant. This therefore gives further evidence to tyrosine 144 being key to ligand binding as well as the tyrosine in position 85.

#### 4.4.5 Discussion of MmfR/MMF Binding Data

Figure 4.19 summarises the data found on MmfR/MMF binding using the luciferase reporter gene assay, combining both the analyses done on the binding affinities of the five MMFs to the wild type MmfR as well as the MmfR ligand binding pocket tyrosine mutants. Unless otherwise stated, this figure refers to the wild type *mmfR* strains.



**Figure 4.19.** Bar chart showing a comparison of the binding potentials of MMF1-5 for WT MmfR, as shown by the levels of *luxCDABE* expression when under the control of *mmfLp* and WT MmfR, compared to the binding potential for MMF4 with the Y85F or Y144F MmfR mutants

*Binding potential* =  $B_{max}/K_d$       *Strains used:* **L1+WTmmfR** – *luxCDABE* under the control of *mmfLp* and wild type *mmfR* under the control of *ermEp\** (*pKMS01*), **L1+mmfR Y85F** – *luxCDABE* under the control of *mmfLp* and *mmfR* with a mutation to tyrosine 85 under the control of *ermEp\** (*pKMS85*), **L1+mmfR Y144F** – *luxCDABE* under the control of *mmfLp* and *mmfR* with a mutation to tyrosine 144 under the control of *ermEp\** (*pKMS144*). Unless otherwise stated, L1+WTmmfR is used for all data points.

Results from the luciferase assay revealed that there are detectable changes in *lux* gene expression with concentrations of the MMFs as low as 5  $\mu$ M, with the  $K_d$  values ranging between 18 and 70  $\mu$ M for the five molecules. Saturation of MmfR appears to occur sometime after around 200  $\mu$ M, and varies between the particular ligands. The binding potentials varied between the different MMFs, with the branched alkyl chains providing the best efficacy.

Four out of the five MMFs had a calculated  $B_{max}$  bioluminescence reading greater than the maximal reading for the positive control (L1+pCC4) indicating that they may have more of a dose effect that just releasing MmfR and may work as some kind of activator. This is something which would be exciting to investigate further, potentially with a more high throughput assay than the luciferase one used here.

An *in silico* analysis of MmfR and its homologues indicated that there are two key residues involved in ligand binding in MmfR, that of tyrosines in amino acid positions 85 and 144. This was indicated to be consistent in binding across all five MMFs and mutants created for these residues provided an interesting set of data.

To summarise the findings from Section 4.4, the following binding potentials were calculated for all of the samples tested (see Figure 4.19);

**MMF1 > Y144F MMF4 > MMF3 > MMF5 > MMF4 > MMF2 > Y85F MMF4**

The binding potential of the Y85F MmfR mutant to MMF4 is much lower than the wild type binding to the same ligand, or indeed any of the other furans. The Y144F MmfR mutant on the other hand appears to have increased release from the MARE operator in the presence of MMF4 when compared to the wild type. This mutant also appears to have greater binding potential to MMF4 than the wild type strain does to either MMF2, 3 or 5.

Although the Y144F mutant appears to be more sensitive to the MMFs, it was not deemed suitable for later use in the novel inducible expression system due to its decreased repressive activities. These decreased repressive activities were also seen for the Y85F mutant, possibly as a result of the ligand binding residues selected for mutation being close to the dimer interface of MmfR and therefore are potentially having an effect on MmfR conformation and consequently, DNA-binding ability.

## 4.5 Outlook for Further Investigations

In this chapter MmfR has been shown to cause repression at the three known methylenomycin gene cluster MARE operators as well as being released upon the binding of all five MMFs, in line with the hypotheses for this investigation. When stating these hypothesis, it was explained that the role of the MmfR paralogue, MmyR was much less understood. Knockouts of this protein, a potential pseudo MMF receptor, produce the phenotype of methylenomycin overproduction.(71) It is therefore clear that it has a repressive role. The DNA binding sequences and ligands of this second type of TFR are ambiguous however. The next stage of this investigation into methylenomycin cluster repressor/ligand interactions is to use the luciferase assay to study the role of paralogue, MmyR. Many of the methods used in Chapter 4 were also used for this investigation. For example, an investigation into strength of MARE operator binding could be done in the same way with this alternate repressor.

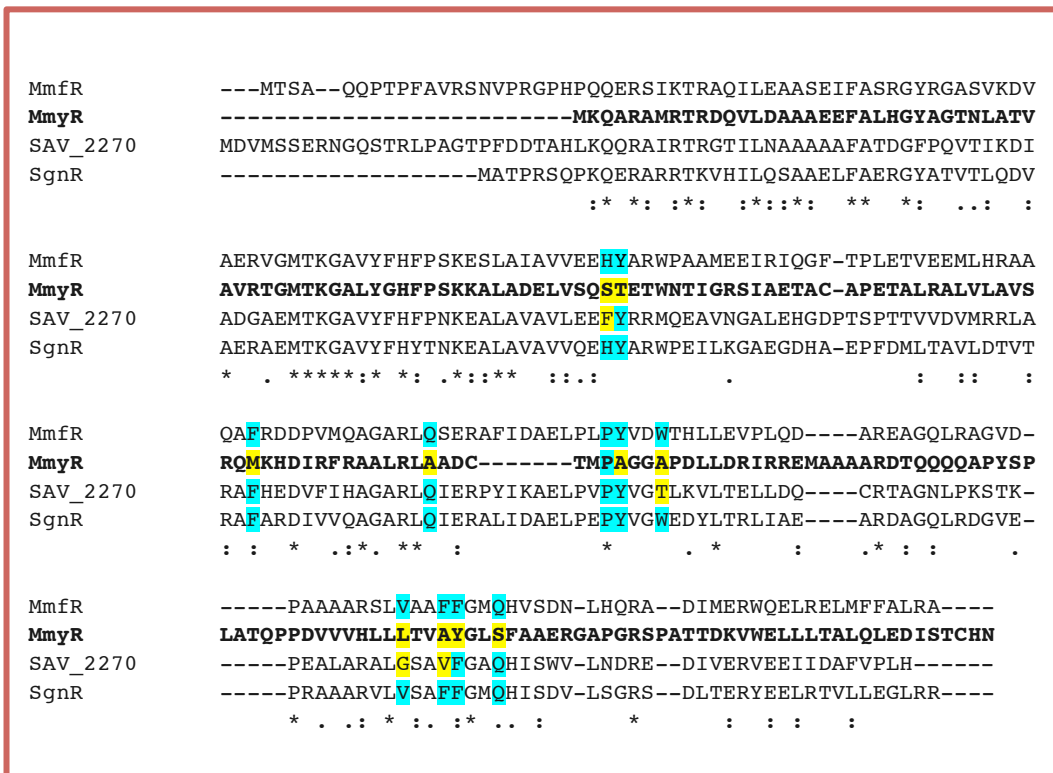
## 5 Mode of Action of Transcriptional Repressor MmyR in *Streptomyces coelicolor*

### 5.1 Comparison of MmyR with MmfR and Homologous Regulatory Pathways

Before the investigations in this project, the ligand and DNA binding properties of MmyR were unknown and it was thought to be a protein akin to the GBL pseudoreceptor type TetR family members. An investigation of identity and similarity between MmfR and MmyR reveals that they share 35% identity and 56% similarity over 110 amino acids (just over 50% of their sequences) (see Table 7.3, Table 7.4 and Appendix C). Unsurprisingly the area where most identity is shared covers the N-terminal TetR family HTH domain. MmyR also shares 47% and 39% identity across 58% and 94% of its amino acid sequence with its orthologues GbnR and SAV\_2268 from *S. venezuelae* and *S. avermitilis* respectively compared to 36% and 50% identity seen with their paralogues, SgnR and SAV\_2270 (across 72% and 34% of its sequence). Despite these general similarities however, there are key differences in the ligand (Section 5.4) and DNA binding (Section 5.3.1) regions of MmfR and MmyR.

Figure 4.14 shows an entire amino acid sequence alignment and comparison carried out for homologues MmfR, MmyR, SAV\_2270 and SgnR. Residues identified as being found in the ligand binding pocket of MmfR in Section 4.4.3 have been highlighted. Matches to the MmfR polypeptide sequence are shown in blue with mismatches shown in yellow.

It appears that the tyrosines in positions 85 and 144, believed to be involved in hydrogen binding the ligand, are conserved across all of the MmfR orthologues but are absent in the paralogous MmyR and its own orthologues (not shown here). Preliminary trials with the luciferase assay indicated that MmyR is not released upon the addition of MMFs to the media, it is not surprising therefore that MmyR is not shown to include either of these tyrosine residues or the majority of the residues involved in hydrophobic interactions. This and the low levels of repression by MmyR seen in the luciferase assay indicate that MmyR works in a very different way to MmfR. These differences are under investigation in this chapter, also using the luciferase assay.



**Figure 5.1.** Comparison of MmfR ligand binding amino acid residues with the sequences of homologues MmyR, SgnR and SAV\_2270

**Blue** – matches with MmfR, **Yellow** – mismatches with MmfR

**Amino acid sequences used:** MmfR – methylenomycin cluster transcriptional repressor, MmyR - methylenomycin cluster transcriptional repressor, SgnR – MmfR homologue from *Streptomyces venezuelae*, SAV\_2270 – MmfR homologue from *Streptomyces avermitilis*

## 5.2 Aims and Strategy of Investigation

Previous investigations have shown that recombinant MmyR is not soluble when purified from standard *E. coli* expression systems and it has therefore not yet been possible to purify it for crystallisation. Without currently being able to purify this protein, the types of analyses that can be run on it are more limited, for example a gel shift assay cannot be run. Without the 3D structure of this protein *in silico* docking analyses are also much less reliable. The bioluminescence assay therefore offers a very useful technique for studying MmyR *in vivo*.

Again, the two vector system was used (Figure 3.3) with *mmyR* added in a vector under the control of *ermEp\** (pKMS03) instead of *mmfR* (pKMS01) in addition to the *lux* vector with *luxCDABE* under the control of one of the MARE operators containing methylenomycin cluster intergenic regions. As was done for MmfR, this investigation was split into two main parts. The first was to investigate MmyR/MARE operator binding, looking at the three known MmfR binding sequences, this time with MmyR. The second part of the investigation looked at MmyR/MMF binding. This also included an investigation into other potential cognate



ligands for MmyR, in line with the differential ligand binding properties of the pseudo GBL-receptors, analogous to this protein.

In addition to the *in vivo* reporter gene analysis of MmyR activity, a number of *in silico* amino acid sequence analyses were also carried out in an attempt to better understand the differences between MmyR and the paralogous MmfR and which motifs in their amino acid sequences may be responsible for any different functionality seen.

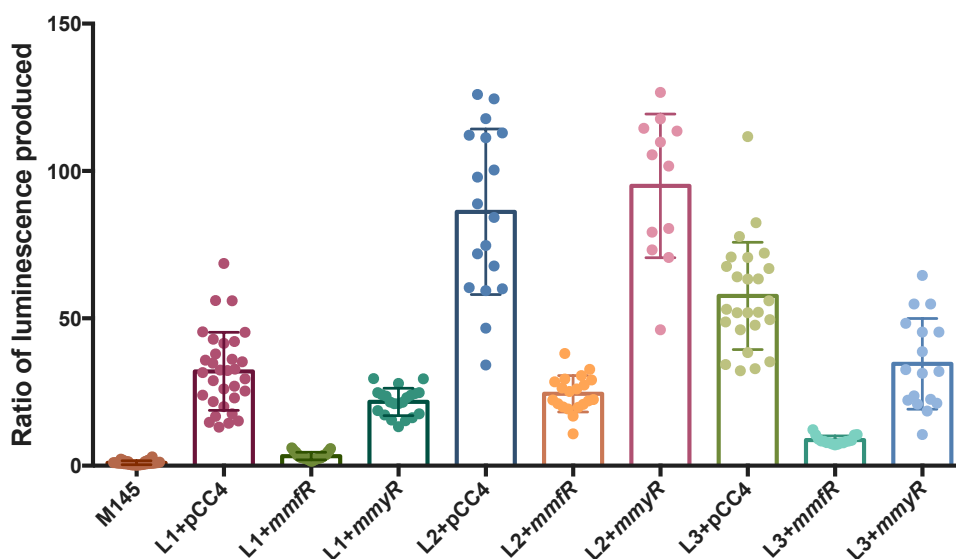
## 5.3 MmyR-MARE Operator Interactions

### 5.3.1 Strength of MmyR Binding to the MARE Operator

In the previous chapter, MmfR was shown to cause repression of the *lux* operon in the presence of all three predicted 24 bp MARE operators, here we are investigating if this is also the case for MmyR. The same nomenclature for strains is used here as were followed in Chapters 3 and 4.

Figure 5.2 shows a bar chart of data collected of luminescence produced by *Streptomyces* strains at 48 hours in the presence of MmyR compared to no repressor at the same MARE operator (pCC4 positive control) and the M145 negative control. If MmyR were binding to a particular MARE operator then a reduction in luminescence would be expected for that strain compared to the positive control strain with no repressor. The luminescence produced by the M145 control strain represents how the system looks in the absence of *luxCDABE* expression. Data on MmfR repression at each MARE operator is also included in Figure 5.2 as a reference point of what significant repression in this luciferase reporter system may look like.

The data from Figure 5.2 was then used to perform a *t*-test to check if the level of repression caused by MmyR at each MARE operator was a statistically significant level of repression or down to natural fluctuations in bioluminescence. The results from this *t*-test are found in Table 5.1. The p-values for the *mmyR* and *mmfR* strains were calculated against the results for the L1, L2 or L3 + pCC4 positive control strains (offsetting the influence of promoter strength on data). The mean and median luminescence shown in Table 5.1 was calculated as a ratio, relative to the luminescence produced by the M145 control.



**Figure 5.2.** Bar chart of luminescence produced by *S. coelicolor* M145 strains at 48 hours, where *luxCDABE* expression is controlled by different MARE operators, in the presence or absence of MmfR or MmyR

*Luminescence produced is calculated as a ratio of the luminescence produced by the M145 negative control. Error bars are shown as the standard deviations of data with all data points collected also shown on the chart.*

*Strains used: M145 – negative control strain, L1, L2 and L3 + pCC4 - luxCDABE under the control of mmfLp, mmyRp or mmyBp with empty pCC4, L1, L2 and L3 + mmfR – luxCDABE under the control of mmfLp, mmyRp or mmyBp and mmfR under the control of ermEp\* (pKMS01), L1, L2 and L3 + mmyR – luxCDABE under the control of mmfLp, mmyRp or mmyBp and mmyR under the control of ermEp\* (pKMS03).*

**Table 5.1.** Statistical analysis of luminescence produced by *S. coelicolor* M145 strains at 48 hours, where *luxCDABE* expression is controlled by different MARE operators, in the presence or absence of MmfR or MmyR

*Strains used: same as Figure 5.2. Luminescence produced is calculated as a ratio of the luminescence produced by the M145 negative control. Results for t-test were calculated against the results for the L1, L2 or L3 + pCC4 positive control strains.*

| Strain  | Mean luminescence produced | Median luminescence produced | t-test results of data compared to L1, L2 or L3 + pCC4 | Significant binding at the MARE operator? |
|---------|----------------------------|------------------------------|--|---|
| M145    | 1.0                        | 0.86                         | -  | -   |
| L1+mmfR | 3.2                        | 2.81                         | 6.8E-17  | YES                                       |
| L1+mmyR | 21.7                       | 21.59                        | 1.2E-03  | YES                                       |
| L1+pCC4 | 32.0                       | 31.94                        | -  | -   |
| L2+mmfR | 24.4                       | 23.81                        | 1.9E-11  | YES                                       |
| L2+mmyR | 94.9                       | 103.60                       | 3.9E-01  | NO  |
| L2+pCC4 | 86.2                       | 86.54                        | -  | -   |
| L3+mmfR | 8.7                        | 8.34                         | 1.1E-13  | YES                                       |
| L3+mmyR | 34.6                       | 31.95                        | 1.1E-04  | YES                                       |
| L3+pCC4 | 57.7                       | 53.09                        | -  | -   |

As can be seen from the results of the *t*-test (Table 5.1), there was indeed a significant level of repression by MmyR at two of the three MARE operator sites; the L1 and L3 MARE operators. The levels of repression at both these sites are much lower than MmfR binding however with between 3.9 and 6.8 times more luminescence being produced by the *mmyR* strains compared to the *mmfR* strains at the L3 and L1 MAREs respectively.

The *t*-test indicated that at the MARE between *mmyR* and *mmfP* (L2 MARE operator) there is not a significant change in luminescence when MmyR is present compared to when it is not. It is interesting that MmyR does not appear to bind to the MARE operator that regulates its own expression, meaning that it is potentially not self regulatory in the way that MmfR is (see Chapter 6). MmfR does show significant binding to the L2 MARE, but it is the weakest binding of the three, indicating that this operator has the least regulation.

To further compare findings from Table 5.1 and Figure 5.2 with the data collected on MmfR in last chapter; MmfR binds to the L1 intergenic region with the greatest affinity showing an average of 90% reduction in luminescence, followed by L3 with an 85% reduction and then L2 at 72% reduction compared to a positive control. MmyR on the other hand appears to bind best to the L3 intergenic region, seeing a 40% reduction in luminescence compared to the positive control without repressors followed by the L1 intergenic region strains which see an average reduction in luminescence of 32%.

To understand the differences in MmfR and MmyR binding to the MARE operators it is necessary to check their amino acid sequences. Both MmfR and MmyR have a predicted DNA binding motif within their TetR type HTH domains. Within this 20 bp region, there were found to be conserved similarities and differences between MmfR and MmyR and their individual orthologues from *S. venezuelae* and *S. avermitilis*. Table 5.2 shows a comparison of these DNA binding motifs in the six homologues. There is a conserved GAVYFH sequence found in MmfR, SgnR and SAV\_2270 whereas MmyR and its orthologues have an alternative conserved GALYGH sequence. These conserved two amino acid variations are the potential cause of the different DNA binding profiles. This would make an interesting target for site directed mutagenesis, to see if the DNA binding profiles of MmfR and MmyR can be exchanged by altering only these two residues.

**Table 5.2. Sequences of DNA binding regions from MmfR, MmyR and their homologues**

| <i>Name</i>     | <i>Sequence (20 aa)</i>      |
|-----------------|------------------------------|
| <b>MmfR</b>     | SVKDVAERVGMT <b>KGAVYFHF</b> |
| <b>SgnR</b>     | TLQDVAERAEM <b>TKGAVYFHY</b> |
| <b>SAV_2270</b> | TIKDIADGAEM <b>TKGAVYFHF</b> |
| <b>SAV_2268</b> | NLQNIADRIRL <b>TKGALYGHF</b> |
| <b>GbnR</b>     | NLADITARTGL <b>TKGALYGHF</b> |
| <b>MmyR</b>     | NLATVAVRTGMT <b>KGALYGHF</b> |

Having seen the limited repressive activity of MmyR on the methylenomycin cluster MARE operators it begs the question of how a *mmyR* knockout brings about the phenotype of methylenomycin over production when the *mmfR* knockout does not (Table 6.1). Based on the findings of there being a conserved and different DNA binding motif in MmyR and its orthologues to MmfR, it would not be unrealistic to conclude that maybe MmyR binds additional, as yet unknown, DNA targets and regulates methylenomycin biosynthesis by another means. It may repress less strongly than MmfR but have many more genetic targets.

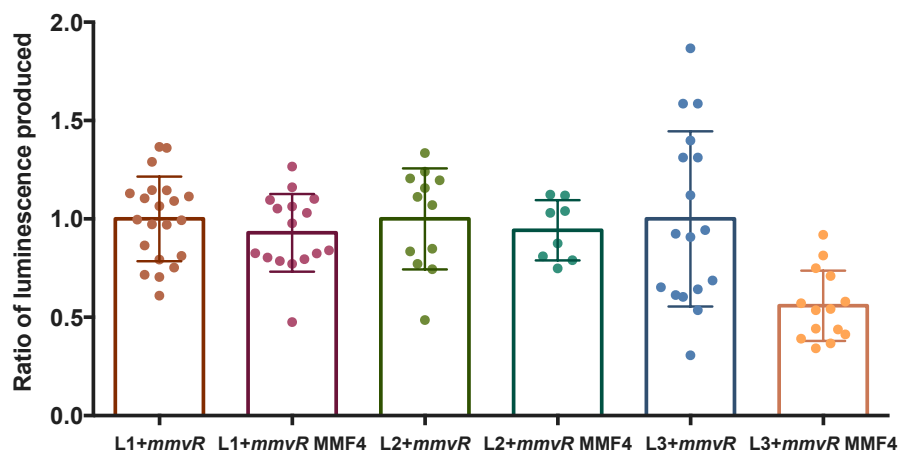
Another possible theory is that it is only the strength of the *mmyR* promoter (L2 strains, see Section 3.4) and therefore the amounts of MmyR produced that allow repression of the methylenomycin cluster by MmyR to take place. In this system, MmyR levels are limited by the weaker *ermEp\** promoter and so the full level of repression expected *in vivo* in the wild type system may not be seen. An alternative hypothesis is that maybe MmyR is only active as a repressor in a ligand bound form rather than the apo form.

The luciferase assay was therefore again utilised to try and further understand the complex relationship between MmyR and methylenomycin regulation. An initial investigation was done in the same way as was done for MmfR in the previous chapter, looking for the release of MmyR by a single concentration of MMF4 at each of the MARE operators.

### 5.3.2 Investigation into Release of MmyR from MARE operators by MMF4

Preliminary research (Section 3.3.5) indicated that MmyR will not bind to the MMFs and be released from the DNA as MmfR is. MmyR also lacks the two tyrosine residues present in the ligand binding pocket of MmfR that are believed to form hydrogen bonds with the furan ligands. However, the potential release of MmyR by the MMFs did need to be checked to confirm the preliminary findings using the fully optimised luciferase assay. This assay was carried out in the same manner as the investigation into MmfR in Section 4.3.3. However, reduction as well as increase in luminescence was looked for this time, in case MmyR was only working as a repressor in a ligand bound form.

Figure 5.3 shows the luminescence from L1+mmyR and L3+mmyR strains in the presence of 100  $\mu$ M MMF4 compared to the same strains without MMF4. This data was then checked with a *t*-test to see if there was any significant change in the amount of luminescence produced in the presence of MMF4, the results of which are displayed in Table 5.3. The L2+mmyR strain was also included in this study as a control, no significant MmyR binding to the L2 MARE operator was seen and therefore no removal of repression would be expected to be seen either.



**Figure 5.3. Bar chart of luminescence produced by the *lux* operon when under the control of different MARE operators and MmyR, in the presence and absence of 100  $\mu$ M MMF4 at 48 hours**

*The data for the mmyR strains with MMF4 were normalised against the same strain with no MMF to give a relative ratio of 1 for each negative control. The error bars for this data show the standard deviations. Strains used: L1, L2 or L3 + mmyR – luxCDABE under the control of mmfLp, mmyRp or mmyBp and mmyR under the control of ermEp\* (pKMS01)*

**Table 5.3 Statistical analysis of luminescence produced by the *lux* operon when under the control of different MARE operators and MmyR, in the presence and absence of 100  $\mu$ M MMF4 at 48 hours**

*Strains used: same as Figure 5.3. The data for the mmyR strains with MMF4 were normalised against the same strain with no MMF to give a relative ratio of 1 for each negative control.*

| Strain  | Mean ratio of luminescence produced | <i>p</i> -value | Significant change in luminescence? |
|---------|-------------------------------------|-----------------|-------------------------------------|
| L1+mmyR | 0.93                                | 0.3122          | NO                                  |
| L2+mmyR | 0.94                                | 0.5756          | NO                                  |
| L3+mmyR | 0.56                                | 0.0016          | YES (significant decrease)          |

As expected, there was no significant change in the L2+*mmyR* strain in the presence of MMF4. The same was also found for L1+*mmyR*. The results were more noteworthy for L3+*mmyR* however, where a significant decrease in luminescence was observed in the presence of MMF4. This is inline with the hypothesis that MmyR is active in its ligand bound form and indicates that MmyR binds better to the L3 MARE operator in the presence of MMF4. However the change is not large, with L3+*mmyR* with MMF4 still producing an average of 19 times as much luminescence as the M145 control (compared to nine times more for L3+*mmfR*<sup>1</sup>). This decrease in luminescence may still not be sufficient to make a significant impact on the expression of methylenomycin biosynthetic cluster. The true implication of this result is therefore as yet unclear. Due to a lack of data on the structure of MmyR it is unfortunately not possible to determine the 3D ligand binding pocket to try and interpret these results better. MmyR binding at the *mmyB/mmyY* intergenic region was therefore further investigated in Section 5.4.3 on page 113 using the luciferase assay.

### 5.3.3 Discussion of MmfR/MARE Operator Binding Data

Table 5.4 summarises the findings found of MmyR binding and release from the three methylenomycin cluster MARE operators. It can be seen that MmyR binds best at the operator between *mmyY* and *mmyB* (L3) and this binding is appears to strengthen upon the addition of 100  $\mu$ M MMF4. Binding is also shown to the L1 MARE operator but with no significant change upon the addition of MMF4. No binding to the L2 MARE operator was detected either in the presence or the absence of the MMFs.

**Table 5.4. Summary of binding data for MmyR and the MARE operators**

| <i>MARE operator site</i> | <i>Corresponding promoter</i> | <i>Strength of MmyR binding</i> | <i>Significant change in MmyR binding upon addition of MMF4</i> | <i>Relative promoter strength (1 being the highest)</i> |
|---------------------------|-------------------------------|---------------------------------|---|---|
| L1                        | <i>mmfLp</i>                  | Weakest                         | NO  | 5   |
| L1F                       | <i>mmfRp</i>                  | -                               | -   | 2   |
| L2                        | <i>mmyRp</i>                  | No binding shown                | NO  | 1   |
| L3                        | <i>mmyBp</i>                  | Greatest                        | YES (increased binding)   | 4   |
| L3F                       | <i>mmyYp</i>                  | -                               | -   | 3   |

When comparing the sequences of the three individual operators it was found that there is almost 80% identity between the L1 and L2 MARE operator but each of these only share 62-67% identity with the L3 MARE operator (for L1 and L2 respectively), according to an alignment (137) run with T-Coffee;

<sup>1</sup> See Section 4.3

```

L1      5' ATAATACCTTCC CGCAGGTATATT 3'
        3' TATTATGGAAGG GCGTCCATATAA 5'

L2      5' AACATACCTTCC CGAGGGTATGTT 3'
        3' TTGTATGGAAGG GCTCCCATACAA 5'

L3      5' AAAAAACCTTCG GGAAAGTTTGAC 3'
        3' TTTTGGGAAGC CCTTCCA AACTG 5'
          * * * * * * * * * * * * * *
    
```

It seems strange therefore, that it is the L2 MARE operator, rather than L3, which has the more distinct MmyR binding profile. This weaker binding and repression at the L2 MARE operator is also something that is seen for the data collected on the *mmfR* strains in the previous chapter however, emphasising the potential importance of the two residues that are conserved between the L1 and L3 MARE operators (highlighted in yellow) but not present in L2 in interacting with the HTH DNA binding domain of MmyR (and MmfR).

The investigation in Section 5.3.2 gave evidence of possible MmyR/MMF interactions and there was a interest in investigating this further. L1+*mmyR* strains were first used to trial all of the MMFs (Section 5.4.1) to allow a direct comparison with data from L1+*mmfR*. The potentially more important L3+*mmyR* strains were then investigated with the MMFs, the data from which can be found in Section 5.4.3.

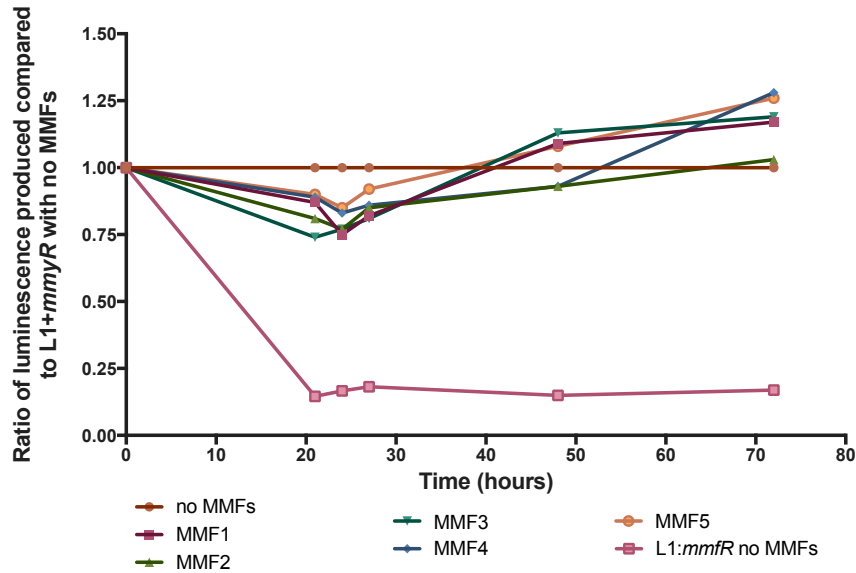
## 5.4 MmyR-MMF Interactions

### 5.4.1 Effect of MMFs on MmyR Binding at the L1 MARE Operator

As reported in Section 5.3, it is clear that 100  $\mu$ M MMF4 does not cause the release of MmyR from the L1 MARE operator. It seems unlikely that any of the other MMFs will bind to and release MmyR from the DNA at this site either. However, there is limited experimental data to prove that this is the case and so all were therefore analysed, particularly as some MMFs have a greater relative binding potential than others. The structural differences between the five MMFs mean that the possibility of MmyR binding to some MMFs, even if it was not seen in the case of MMF4, cannot be discounted. MMF1, 2, 3 and 5 (Figure 5.6) were all added at a concentration of 100  $\mu$ M to the L1+*mmyR* strains and the luminescence assay carried out in the same way as it was done for L1+*mmfR* strains in Section 4.4.

The findings from measuring luminescence at five time points over 72 hours can be found in Figure 5.4. This data was then analysed as before using a bar chart to look results at from 48 hours only, this can be found in Figure 5.5. A *t*-test was then used to determine if there are

any statistically significant results, the results of which can be found in Table 5.5. Results for L1+*mmfR* are also included to represent what more significant repression in this system might look like.

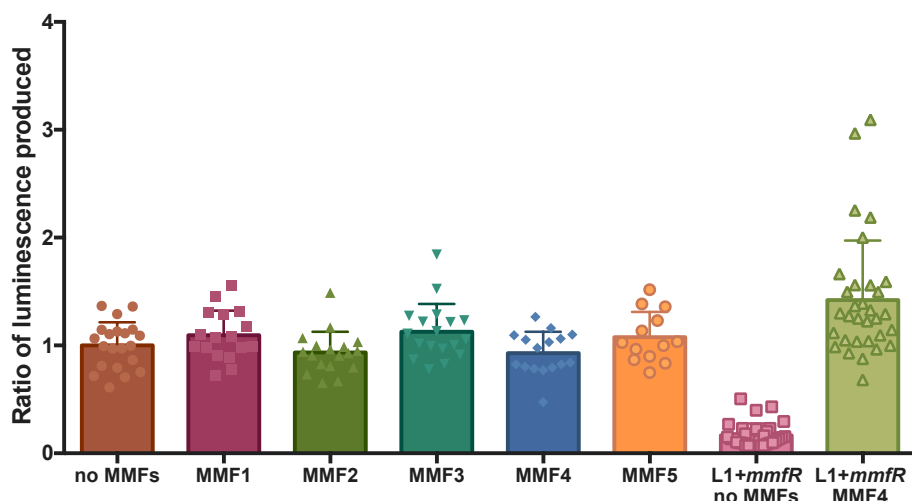


**Figure 5.4. Luminescence produced by the *lux* operon when controlled by *mmfLp* and MmyR in the presence of 100  $\mu$ M MMF1-5 compared to no MMFs**

Average light production is calculated as a relative ratio of luminescence produced by L1+*mmfR* with no MMFs (giving this sample a value of 1). **Strains used:** L1+*mmfR* – *luxCDABE* under the control of *mmfLp* and *mmfR* under the control of *ermEp\** (*pKMS01*), L1+*mmyR* – *luxCDABE* under the control of *mmfLp* and *mmyR* under the control of *ermEp\** (*pKMS03*). Unless otherwise specified, all data points are for the L1+*mmfR* strain.

As can be seen in Figure 5.4, there are some fluctuations in luminescence produced by L1+*mmfR* in the presence of the MMFs compared to no MMFs present but results using different furans never vary significantly from the negative control (L1+*mmfR*, no MMFs). The levels of luminescence produced by L1+*mmfR* are distinctly lower than those for the L1+*mmyR* strain in the presence of the MMFs at all time points however. This indicates that the MMFs are not causing MmyR to act as a repressor in a ligand bound form at the L1 MARE operator. Figure 5.5 and Table 5.5 also support this.





**Figure 5.5. Effect on luminescence produced by the *lux* operon, when controlled by *mmfLp* and MmyR, by the addition of 100  $\mu$ M MMF1-5 at 48 hours growth**

Average light production is calculated as a relative ratio of luminescence produced by *L1+mmyR* with no MMFs (giving this sample a value of 1). **Strains used:** Same as Figure 5.4. Unless otherwise specified, all data points are for the *L1+mmyR* strain.

**Table 5.5. A *t*-test analysis of the effects on luminescence produced by the *lux* operon, when controlled by *mmfLp* and MmyR, by the addition of 100  $\mu$ M MMF1-5 at 48 hours growth**

Average light production is calculated as a relative ratio of luminescence produced by *L1+mmyR* with no MMFs (giving this sample a value of 1). The *p*-value was also calculated based on *L1+mmyR* with no MMFs. Data for *L1+mmyR* is included as a comparison to represent how more significant repression look in this kind of assay.

| MMF added to <i>L1+mmyR</i> | <i>p</i> -value | Significant difference? | Average light production at 48 hr (R.R) |
|-----------------------------|-----------------|-------------------------|---|
| MMF1 (100 $\mu$ M)          | 0.2008          | FALSE                   | 1.09                                    |
| MMF2 (100 $\mu$ M)          | 0.3266          | FALSE                   | 0.93                                    |
| MMF3 (100 $\mu$ M)          | 0.1041          | FALSE                   | 1.13                                    |
| MMF4 (100 $\mu$ M)          | 0.3122          | FALSE                   | 0.93                                    |
| MMF5 (100 $\mu$ M)          | 0.3327          | FALSE                   | 1.08                                    |
| <i>L1+mmyR</i> (no MMF)     | <0.0001         | TRUE                    | 0.15                                    |

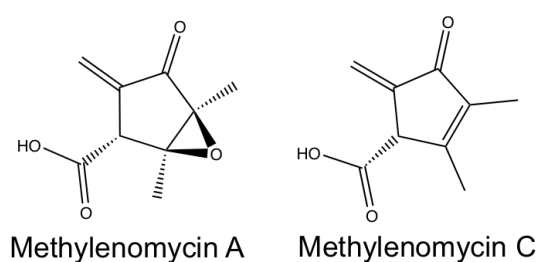
The *t*-test results shown in Table 5.5 confirm that there is no significant release or extra binding of MmyR upon the addition of the five MMFs when looking at measurements for the *L1* MARE operator. Without there being any significant effects on luminescence by adding the MMFs to the *L1+mmyR* strains, no further analyses were done to try and work out the  $K_d$  and  $B_{max}$  values for this strain.

Following on from the work with MmyR with the five MMFs, a further investigation was carried out to look for other potential cognate ligands for MmyR, in line with the hypothesis that this pseudo GBL-receptor analogue may bind alternative ligands to MmfR.

#### 5.4.2 Effect of Other Potential Ligands on MmyR Binding at the L1 MARE Operator

Previous studies of MmyR homologues such as ScbR2 have shown that this pseudo GBL receptor will bind to and be released by endogenous antibiotics, potentially a form of cross regulation between pathways.(138) It would appear logical that either an MMF or the end product of the methylenomycin biosynthetic cluster; methylenomycin A, or one of its precursors may bind to MmyR. In this case however, with limited repression seen in the apo form, MmyR may only be 'activated' as a repressor by the cognate antibiotic. Binding to methylenomycin A or a precursor could be a method to prevent the over production of the antibiotic, something that would be potentially toxic to the streptomycetes as well as a drain on cell resources if the antibiotic is not needed anymore. As yet, there has been no experimental evidence of this negative feedback loop.

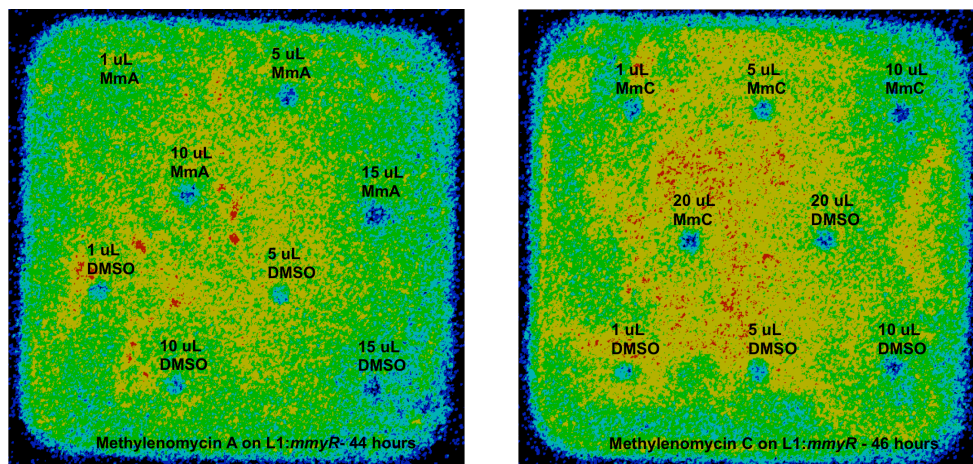
Both methylenomycin A and the precursor methylenomycin C (Figure 5.6) trialled with the L1+mmyR strain to look for any significant change in luminescence produced. This could be either an increase in luminescence if these small molecules cause the release of MmyR or a reduction in luminescence if they are 'activating' MmyR as a repressor.



**Figure 5.6. Chemical structures of methylenomycin furans, methylenomycin A and its precursor; methylenomycin C**

Methylenomycin A and C have lengthy purification procedures which do not yield large amounts of product. For this reason, preliminary trials were run where the methylenomycins were added in different concentrations to sterile filter paper discs on top of a lawn of L1+mmyR, in an effort to decrease the volumes of these small molecules needed. From this, the aim was that more thorough trials could be designed and run when more was known about optimal concentration of the potential ligands. The methylenomycin molecules were diluted in DMSO and added at a range of concentrations that have been shown by other researchers

in the group to not be toxic to *Streptomyces*. Figure 5.7 shows these L1+mmyR plates with methylenomycin C and A added.

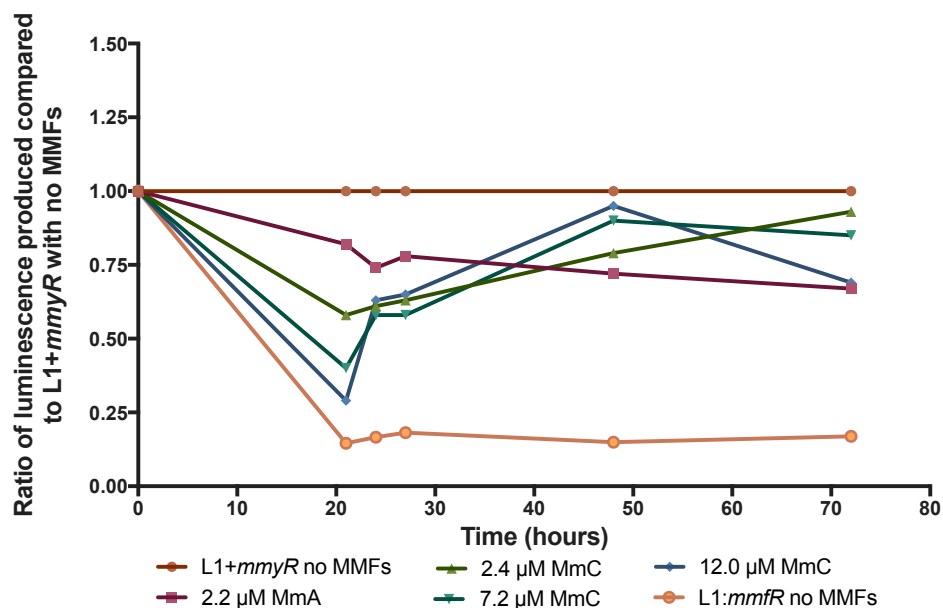


**Figure 5.7. CCD camera images of luminescence produced by a lawn of *S. coelicolor* M145 containing the *lux* operon under the control of *mmfLp* and MmyR in the presence of methylenomycin A and C compared to a DMSO control**

*Methylenomycin A and C were used diluted in DMSO at a concentration of 5.5 mM and 6.0 mM respectively and added to sterile paper dots on a lawn of L1+mmyR. Equivalent amounts of DMSO were added as a negative control.*

As can be seen in Figure 5.7, there does not appear to be any obvious difference in luminescence when adding methylenomycin A or C when compared to a DMSO control. However, it was very hard to spread an even lawn of *Streptomyces* over such a large plate, meaning that any small changes in luminescence may be missed due to a non-uniform lawn of streptomycetes present. This investigation was therefore carried forward to tests involving quantitative data taken from 12-well plates as was done in all previous investigations with the luciferase assay. The data for this investigation can be seen in Figure 5.8 and Figure 5.9. Due to the lack of methylenomycins available however, the number of trials that could be run were limited and not every possibility could be tested.

Figure 5.8 compares luminescence produced by the L1+mmyR strain in the presence and absence of methylenomycin A or C compared to the L1+mmfR strain over five time points in 72 hours.



**Figure 5.8. Luminescence produced by *luxCDABE*, under the control of *mmfLp* and MmyR, in the presence of different concentrations of methylenomycin A and C**

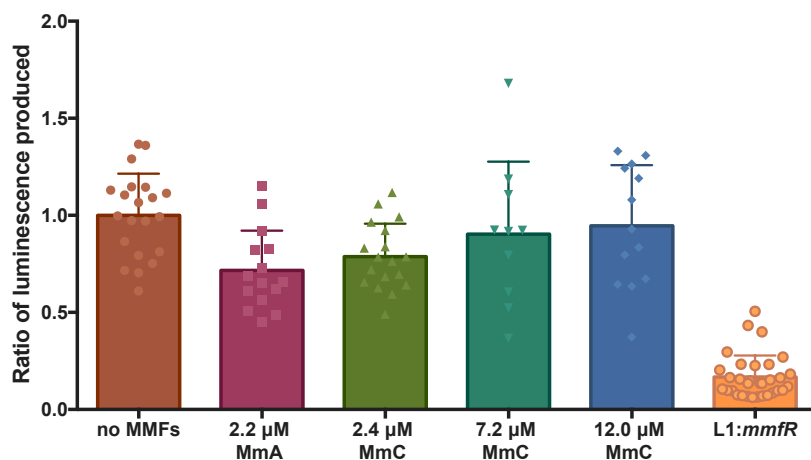
Average light production is calculated as a relative ratio of luminescence produced by L1+mmyR with no potential ligand added (giving this sample a value of 1).

**Strains used:** L1+mmfR – *luxCDABE* under the control of *mmfLp* and *mmfR* under the control of *ermEp\** (*pKMS01*), L1+mmyR – *luxCDABE* under the control of *mmfLp* and *mmyR* under the control of *ermEp\** (*pKMS03*). Unless otherwise specified, all data points are for the L1+mmyR strain.

The data collected from the methylenomycins show more promise in having interactions with MmyR at the L1 MARE operator than the MMFs. As can be seen in Figure 5.8, there is a general decrease in the levels of luminescence in the presence of methylenomycin A and C compared to those seen for the MMFs in Figure 5.4.

The effects of methylenomycin C and A on luminescence appear greatest in the first 27 hours but this fluctuates a lot, with different concentrations swapping in position in terms of greatest level of light produced. It was found in earlier investigations (see Chapter 3 and 4) that readings from the 48 and 72 hour time points had smaller standard deviations and the cell cultures appeared to have stabilised in growth. The methylenomycins were added at a concentration below the predicted lethal level but will likely still cause stress to the cells and so potentially disrupt their growth. This may explain why a general dip in luminescence was seen at 21 hours in Figure 5.8 before increasing again over the next few days. As a control, methylenomycin C was added to L1+pCC4 strains (data not shown here), this showed no significant difference in luminescence produced when compared to the same strain in the absence of methylenomycin C during the first four time points. There was however a decrease in luminescence at the 72-hour time point, possibly due to cell death caused by the antibiotic. For these reasons, a statistical analysis run from MmyR/methylenomycin data at the 48 hour

time point, to avoid this potential cell death and the larger deviations at the earlier time points as well as allowing direct comparisons with all other data collected in previous chapters. The results from this analysis can be found in Figure 5.9 and the corresponding Table 5.6.



**Figure 5.9. Bar chart of the effect on luminescence produced by *luxCDABE*, under the control of *mmfLp* and MmyR, by the presence of different concentrations of methylenomycin A and C at 48 hours growth**

*Average light production is calculated as a relative ratio of luminescence produced by L1+mmyR with no potential ligand added (giving this sample a value of 1).*

*Strains used: Same as Figure 5.8. Unless otherwise specified, all data points are for the L1+mmyR strain.*

**Table 5.6. A *t*-test analysis of the effect on luminescence produced by *luxCDABE*, under the control of *mmfLp* and MmyR, by the presence of different small molecules at 48 hours growth**

*Average light production is calculated as a relative ratio of luminescence produced by L1+mmyR with no MMFs (giving this sample a value of 1). p-value was also calculated based on L1+mmyR with no MMFs. Data for L1:mmfR is included as a comparison to represent how more significant repression may be seen in the assay.*

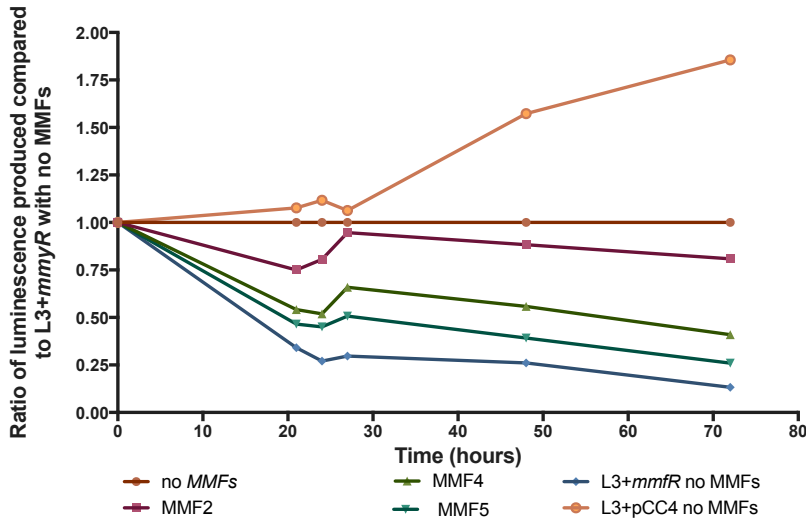
| MMF added to L1+mmyR | p-value | Significant difference? | Average light production at 48 hr (R.R) |
|----------------------|---------|-------------------------|---|
| MMF1 (100 μM)        | 0.2008  | FALSE                   | 1.09                                    |
| MMF2 (100 μM)        | 0.3266  | FALSE                   | 0.93                                    |
| MMF3 (100 μM)        | 0.1041  | FALSE                   | 1.13                                    |
| MMF4 (100 μM)        | 0.3122  | FALSE                   | 0.93                                    |
| MMF5 (100 μM)        | 0.3327  | FALSE                   | 1.08                                    |
| MmA (2.2 μM)         | 0.0003  | TRUE                    | 0.72                                    |
| MmC (2.4 μM)         | 0.0017  | TRUE                    | 0.79                                    |
| MmC (7.2 μM)         | 0.3645  | FALSE                   | 0.90                                    |
| MmC (12 μM)          | 0.5558  | FALSE                   | 0.95                                    |
| L1:mmfR (no MMF)     | <0.0001 | TRUE                    | 0.15                                    |

Figure 5.9 and the corresponding *t*-test analysis from Table 5.6 reveal that both methylenomycin A and C will cause a significant change in luminescence produced by L1+*mmyR*. However, this was only the case for the lowest concentration of methylenomycin C, with the higher ones showing no significant effect on luminescence produced by L1+*mmyR*. Equivalent trials were also run with methylenomycin A but there was not enough compound available to get the full set of repeats and so these are not presented here. Nonetheless, the few trials that were run did indicate the same pattern, where only the lowest concentration of methylenomycin A caused a significant change in luminescence produced.

The change produced even by the lowest concentrations of methylenomycin A and C are not large however, with a reduction in luminescence of less than 30% for L1+*mmyR* compared to the same strain with no compound being present. This compares to presence of MmfR (L1+*mmfR*) producing 85% less luminescence than the L1+*mmyR* strain (Table 5.6). It is still possible that methylenomycin A or C do cause the ‘activation’ of MmyR as a repressor but results are currently inconclusive and will likely remain so until the effect of methylenomycin concentration and its mechanism of antibiotic action is understood better.

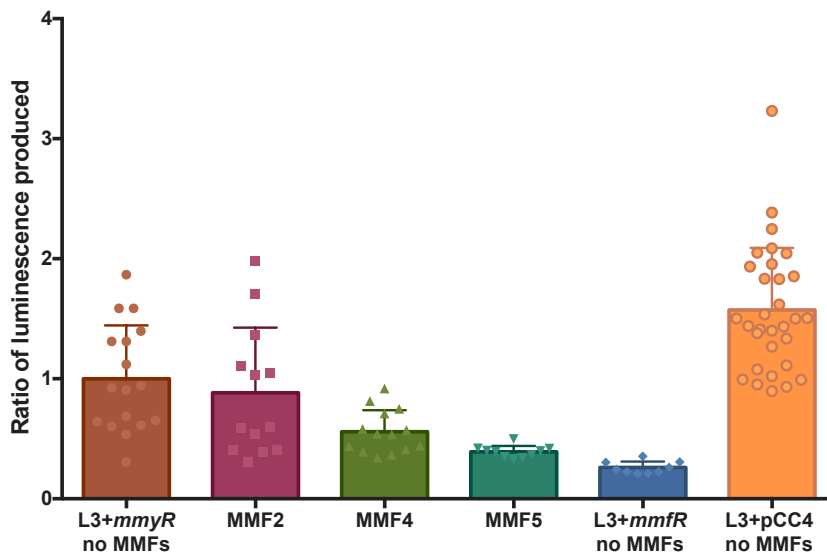
### 5.4.3 Further Investigation - Effect of MMFs on MmyR Binding at the L3 MARE Operator

In Section 5.3, MMF4 was shown to cause a significant reduction in luminescence produced by the L3+*mmyR* strain, something not seen at either of the other MARE operators. It was therefore decided to investigate further into this result, adding 100  $\mu$ M MMF2 and MMF5 to L3+*mmfR* and inspecting for changes in luminescence. The assay was carried out using the same methods as were used for L1+*mmyR* strains and results are shown in Figure 5.10, Figure 5.11 and Table 5.7. Data has also been compared to luminescence produced by L3+*mmfR* (representing more significant repression in this assay) and the positive control; L3+pCC4.



**Figure 5.10. Luminescence produced by the *lux* operon under the control of *mmyBp* and MmyR in the absence and presence of 100  $\mu$ M MMF2, 4 or 5 compared to a positive control over 72 hours**

Average light production is calculated as a relative ratio of luminescence produced by L3+mmyR with no MMFs (giving this sample a value of 1). **Strains used:** L3+mmfR – luxCDABE under the control of mmyBp and mmfR under the control of ermEp\* (*pKMS01*), L3+mmyR – luxCDABE under the control of mmyBp and mmyR under the control of ermEp\* (*pKMS03*), L3+pCC4 – luxCDABE under the control of mmyBp, no repressors. Unless otherwise specified, all data points are for the L3+mmyR strain.



**Figure 5.11. Bar chart of luminescence produced by the *lux* operon under the control of *mmyBp* and MmyR in the absence and presence of 100  $\mu$ M MMF2, 4 or 5 compared to a positive control at 48 hours**

Average light production is calculated as a relative ratio of luminescence produced by L3+mmyR with no MMFs (giving this sample a value of 1).

**Strains used:** Same as Figure 5.10. Unless otherwise specified, all data points are for the L3+mmyR strain.

**Table 5.7. A *t*-test analysis of luminescence produced by the *lux* operon under the control of *mmyBp* and MmyR in the presence of 100  $\mu$ M MMF2, 4 or 5 at 48 hours**

*Average light production is calculated as a relative ratio of luminescence produced by L3+mmyR with no MMFs (giving this sample a value of 1). The p-value was also calculated based on L3+mmyR with no MMFs. Data for L3+mmfR is included as a comparison to represent more significant repression seen in this kind of assay.*

| <i>MMF added to L1+mmyR</i>         | <i>p-value</i> | <i>Significant difference?</i> | <i>Average light production at 48 hr (R.R)</i> |
|-------------------------------------|----------------|--------------------------------|--|
| <b>MMF2 (100 <math>\mu</math>M)</b> | 0.5222         | FALSE                          | 0.88   |
| <b>MMF4 (100 <math>\mu</math>M)</b> | 0.0016         | TRUE                           | 0.56   |
| <b>MMF5 (100 <math>\mu</math>M)</b> | 0.0002         | TRUE                           | 0.39   |
| <b>L3+mmfR (no MMF)</b>             | <0.0001        | TRUE                           | 0.26   |

The findings of the *t*-test analysis in Table 5.7 showed that both MMF4 and MMF5 caused a significant decrease in luminescence of the L3+*mmyR* strain, with up to a 61% reduction in luminescence seen compared to no MMFs. This compares to a reduction in luminescence of 74% for L3+*mmfR* versus L3+*mmyR* indicating that the presence of MMF5 brings MmyR repression almost to the level of that seen for MmfR. MMF2 did also cause a slight decrease in luminescence but it was not enough of one to be classed as statistically significant. Therefore, the greater the binding potential of the MMF to MmfR the greater the apparent decrease in luminescence it brought about in the L3+*mmyR* strain as well as the lower the *p*-value. A higher concentration of MMF2 therefore may reveal more significant results if trialled.

It should be noted that due to time constraints, L3+pCC4 had not been trialled with all five of the MMFs as a control during the preliminary investigations in Chapter 3. However, when trialled with L1+pCC4, none of the MMFs produced a significant change in luminescence at a 100  $\mu$ M concentration. It cannot be said with certainty that this is also the case with the L3+pCC4 strain but it reduces the chances that the MMFs are bringing about a change in luminescence for the L3+pCC4 strain in the absence of MmyR/MmfR. There is also currently no experimental evidence of this extra MMF transcriptional regulation. Therefore, until further research is done, MMF interactions with MmyR rather than any other transcriptional control are the most reasonable explanation for the changes in luminescence seen in Table 5.7 and Figure 5.11.

Based on the analysis from Table 5.7 and Figure 5.11 therefore, it is possible that at least some of the MMFs can bind to MmyR to an extent and, in combination with the DNA binding sequence at the L3 MARE operator, cause MmyR to work better as a repressor. It has



been shown by amino acid sequence analysis that MmyR does not have the same ligand-binding pocket as MmfR (Figure 5.1). However, if MmyR was indeed binding the MMFs at the L3 MARE operator, it appears to lead to a conformational that may help it better bind the DNA instead of its release. It is not inconceivable therefore that the ligand-binding site could vary considerably between MmfR and MmyR and yet they can still both interact with the MMFs, with the differences in the structure of the binding pocket resulting in the opposite effect that the MMFs have on each.

When looking at the levels of repression achieved by L3+*mmyR* in the presence of the MMFs, there is never more than a 60% reduction in luminescence compared to when no MMFs are present. The impact of this leaky repression in the wild type system is not known. It must not be forgotten however, that the *mmyR* promoter was shown to be the strongest of the five studied in the methylenomycin cluster (Figure 3.16), indicating that relatively high levels of MmyR are produced in the absence of repression. In this assay, MmyR repression is limited by the strength of the *ermE\** promoter. MmyR also appears to not regulate itself at the L2 MARE operator, potentially leading to even greater levels produced in a wild type system compared to MmfR, which can repress its own production.

If the MMFs are promoting the binding of MmyR at the L3 MARE operator then the implications would be the repression of the *mmyBQEDXCAPK* and *mmmYF* operons, where *mmyB* codes for a pathway specific transcriptional activator and all other genes code for enzymes thought to be used in methylenomycin biosynthesis. If there are greater levels of MmyR produced in a wild type system than this synthetic system then there may be total repression of the production of MmyB, which could be enough to stop the entire biosynthetic pathway from being expressed.

In Section 5.3, it was found that MmyR caused the greatest repression at the L3 MARE operator, followed by the L1 MARE operator, with no binding shown at the L2 operator. It is not totally clear why MmyR did not also work as a better repressor at the L1 operator in the presence of the MMFs. It is possible that the weaker binding of MmyR at 24 bp L1 MARE operator sequence means that the addition of 100  $\mu$ M MMFs was not enough to cause a significant change in luminescence. If this is the case, a higher concentration of MMFs being added to the L1+*mmyR* may reveal higher levels of repression. Alternatively, when MmyR binds to the L1 MARE operator it may do so in a conformation that makes it harder for the MMFs to enter or interact with its ligand-binding pocket. The operator at the L3 intergenic region shares less than 63% identity (137) with the one at the L1 intergenic region, it is possible therefore that MmyR is in a slightly different conformation when bound at each of these sites.

#### 5.4.4 Discussion of MmyR-Ligand Interactions

The results from the investigation into MmyR ligands revealed that there is no significant binding to any of the five MMFs in the L1+*mmyR* strain at a concentration of 100  $\mu$ M. The data collected on methylenomycin A and C as possible ligands was inconclusive with only the lowest concentrations giving a significant decrease in luminescence produced for the L1+*mmyR* strain. At this point, it is unclear why the higher concentrations did not show the same effects.

At the L3 MARE operator, the addition of the 100  $\mu$ M MMFs did appear to cause a significant reduction in luminescence produced, indicating the 'activation' of MmyR as a repressor in the presence of some of the MMFs (particularly those that showed a higher MmfR binding potential). To further understand the exact role of MmyR however, supplementary investigations are needed.

A number of papers have indicated that in systems with two repressors such as MmfR/MmyR, the second repressor binds a much wider variety of DNA targets as well as ligands.(50) It is unclear how many different DNA sequences MmyR may be able to bind to and what its exact set of receptor-ligand interactions are. The regulatory role of MmyR appears to be much more complicated than that of MmfR. In particular, finding out more about the structure of MmyR would be very helpful. As mentioned before, it has not been possible to purify MmyR from standard *E. coli* expression systems available. However, there are some close homologues of MmyR from other *Streptomyces* strains that have shown promise in terms of solubility and during crystallisation trials. A broader understanding of the role of one of the MmyR homologues would greatly assist in finding out about potential ligands for MmyR as well as understanding the conformational change these ligands bring about. These results would also hopefully shed some light on how and when MmyR will bind to the MARE operators or other DNA sequences.

As well as work on the structure of MmyR, if larger amounts of methylenomycin A and C could be obtained, some more trials could be run, possibly with lower concentrations of these molecules to look for more conclusive results than those found in this investigation. It would also be helpful to test L3+*mmyR* with the remaining MMF1 and MMF3 ligands to check whether these two furans also help to activate MmfR as a repressor. These two ligands had the greatest binding potential to MmfR and so could be expected to also perform well with MmyR. The reason these extra trials with MMF1 and 3 were not run were due to time limits on the project as well as lack of availability of the ligands at the time when experiments were being run.

## 5.5 Outlook for Further Investigations

Following on from this investigation of MmyR, it would be interesting to compare the data gained here with that obtained for MmfR and together, draw some conclusions about the overall regulation of methylenomycin biosynthesis, in particular looking at the combined regulation of their own genes as well as other biosynthetic operons. Of particular interest are the positive and negative feedback loops that MmfR and MmyR may create through the regulation of *mmfLp*, *mmfRp* and *mmyRp*. Through these feedback loops it appears that *S. coelicolor* A(3)2 can retain silence in the methylenomycin gene cluster until needed and then switch it off again once no longer required. A further investigation was carried out in the following chapter, including the collection of new luminescence data as well as an analysis of data already collected from Chapters 3, 4 and 5. From this investigation, there is hope that a more complete picture of regulation can be achieved, something that will then be used in the development of a novel inducible expression system.

## 6 Investigation of MmfR/MmyR Self-Regulation

### 6.1 Background and Strategy of Investigation

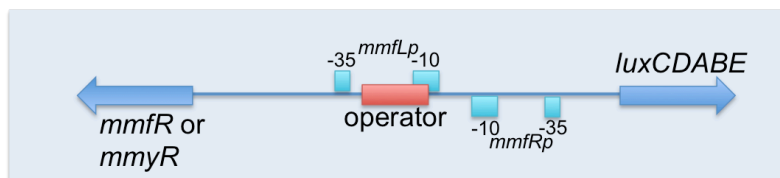
This chapter follows on from the work done with MmfR and MmyR in the previous three chapters and their binding affinities to the three MARE operators and five MMFs. It appears clear that MmfR and MmyR have different roles in regulation and come together to create a complex regulatory network, responding differently to the MARE operators and MMFs.

The five operons thought to be directly regulated by MmfR are *mmyR*, *mmfLHP*, *mmfR*, *mmyBQEDXCAPK* and *mmyYF*. In Chapter 4, MmfR was shown to bind to the three operators between these operons and be released upon the addition of any of the five MMFs. Like other TetR repressors therefore, MmfR appears to control its own expression as well as influencing the expression of the methylenomycin biosynthetic genes (directly or indirectly). MmyR on the other hand showed weak binding to only two of the three operators and showed no significant release by the MMFs (Chapter 5). Interestingly, MmyR did not show any significant binding to the operator that controls its own expression, indicating that unlike MmfR, it is not directly auto-regulated. Although no significant release of MmyR was seen in the presence of the MMFs, there did however seem to be some interaction of the MMFs with MmyR at the L3 MARE operator, with an increased repression of *luxCDABE* seen.

The two vector luciferase reporter systems described in Figure 3.3 were used to study MmfR and MmyR regulation with the transcriptional repressors being produced under the control of the constitutive promoter *ermEp\** at a constant rate. To study the auto-regulatory action of MmfR, a vector that contained *mmfR* under the control of its native promoter was utilised. This vector was constructed by Justin Nodwell and his team and named as 11NY (see Table 2.5 and Table 2.8).<sup>(92)</sup> MmyR was also further investigated via another equivalent vector named sp105, also constructed by Justin Nodwell and his team. Both 11NY and sp105 have the same sequence as the L1 vector (shown previously in Figure 3.4) except for the addition of *mmfR* or *mmyR*, upstream of the L1 intergenic region (Figure 6.1). These variations are represented in the diagram in Figure 6.2C.

The diagram in Figure 6.1 highlights the intergenic region (found between *mmfR* and *mmfL*) that is contained in the L1, 11NY and sp105 vectors. It can be seen that in these vectors, *luxCDABE* is under the control of *mmfLp* and *mmfR* or *mmyR* under the control of *mmfRp*. The expression of both *luxCDABE* and *mmfR* or *mmyR* were also predicted to be regulated by any repression of the MARE operator.

Both MmfR and MmyR were shown to cause significant repression of the reporter genes in strains that contained the L1 MARE operator. Therefore, in both 11NY and sp105 containing strains, MmfR or MmyR should hypothetically be able to regulate their own expression.



**Figure 6.1. Visual representation of the intergenic region located between *mmfR* and *mmfL*, as found in the 11NY and sp105 vectors**

Please note that this diagram is not to scale and purely a representation of the approximate position of the -35/-10 sites and MARE operator.

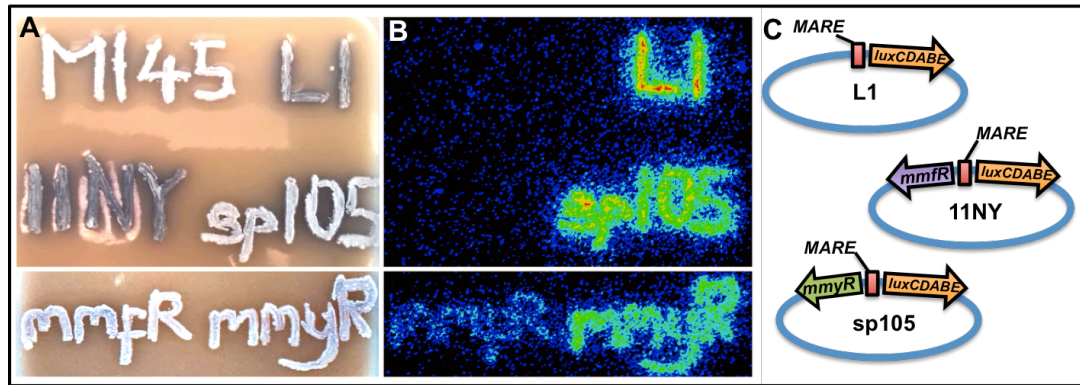
The 11NY and sp105 vectors are from the work of Nodwell et al. (92)

By studying repression and possible release of MmfR and MmyR by the MMFs from strains containing the 11NY and sp105 vectors and comparing it to data collected for *mmfR* and *mmyR* under the control of *ermEp\** it was possible to draw further conclusions about the complex regulation of the methylenomycin biosynthetic cluster and the significance of TetR family auto-regulation.

## 6.2 Preliminary Observations

Figure 6.2<sup>1</sup> shows initial readings taken for the strains containing 11NY and sp105 compared to M145, L1+pCC4 (labelled as 'L1' in Figure 6.2), L1+*mmfR* (labelled as 'mmfR') and L1+*mmyR* strains (labelled as 'mmyR'). In this figure, M145 works as the negative control whereas L1+pCC4 is the positive control, with no repressor present. Figure 6.2A is a photo of how these strains look when grown on solid culture and the corresponding image of the same plates when being measured by the CCD camera for thirty seconds is found in Figure 6.2B. Figure 6.2C on the other hand shows the schematics of key reporter vectors from the strains included in this figure.

<sup>1</sup> This figure is based on the earlier Figure 3.8 which contained data on M145, L1+pCC4, L1+*mmfR* and L1+*mmyR* only



**Figure 6.2. Details on and visual representations of strains containing the *lux* genes under the control of *mmfLp*, with or without the repressor proteins MmfR or MmyR under the control of *ermEp\** or *mmfRp***

Assigned nomenclature = **M145** – *S. coelicolor* M145 negative control with no luciferase genes, **L1** – positive control with *luxCDABE* under the control of *mmfLp* and the empty *pCC4* vector, **11NY** – *luxCDABE* under the control of *mmfLp* and *mmfR* under the control of *mmfRp*, **sp105** – *luxCDABE* under the control of *mmfLp* and *mmyR* under the control of *mmfRp*, **mmfR** – contains *L1* and *pKMS01*, **mmyR** – contains *L1* and *pKMS03*

- A. Strains growing on SFM media
- B. Luminescence seen in the Photek CCD camera during a thirty second reading by the strains shown in A
- C. Schematic representation of reporter plasmids used

In Figure 6.2B, when *mmfR* expression is under the control of *mmfRp* (in the 11NY strain), MmfR appears to more completely repress the expression of the reporter genes as opposed to when it is controlled by *ermEp\** (in the L1+*mmfR* strain). Firstly, Figure 3.16 indicated that *ermEp\** is a weaker promoter than *mmfRp*, meaning that without the presence any other regulatory mechanisms, a higher level of MmfR could potentially be expected in 11NY strain compared to L1+*mmfR*, thereby causing a greater level of repression. However, *mmfRp* was not considerably stronger than *ermEp\** (showing a 1.1 fold increase in luminescence) and hence would not be expected to produce such a pronounced difference in luminescence as was observed in Figure 6.2. Secondly, it is also unlikely that the tighter repression of the *lux* genes is due to MmfR being able to work as an activator at its own promoter but repress other genes. The differences in luminescence for strains containing the 11NY vector compared to *pKMS01* are therefore an indicator of the benefits of the self-regulatory control of *mmfR* by MmfR at the MARE operator. As soon as MmfR is released in the 11NY strain there will be no repression of *mmfR* expression and so more MmfR will be made, ensuring a constant level of repression. It is this tight control over MmfR production, allowing synthesis of this repressor as and when needed, that appears to be the likely cause of more complete repression achieved in the 11NY strain compared to when *mmfR* is under the control of the unrelated *ermEp\**.

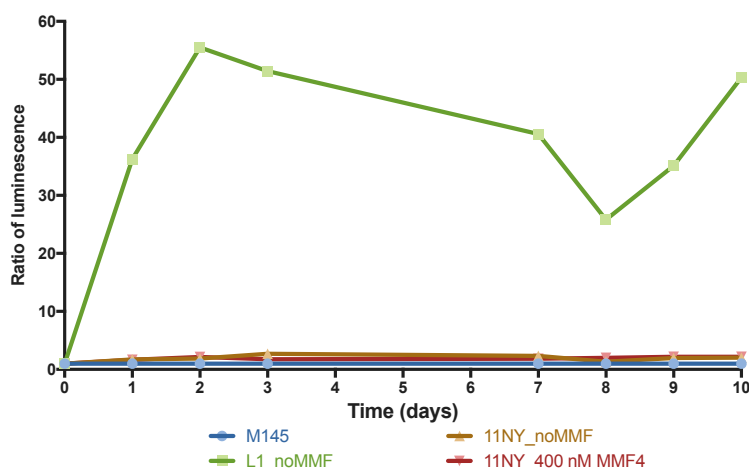
On the other hand, when under the control of *mmfRp*, MmyR (in the sp105 strain) appears to repress luminescence to a lesser degree than when MmyR production is controlled by *ermEp\** (in the L1+*mmyR* strain). Put simply, this phenotype indicates that MmyR under self-regulation does not result in improved repression of the L1 MARE operator.

The *mmyR* strain phenotypes seen in Figure 6.2 help confirm the observations made in Section 5.3 where it was found that MmyR caused the greatest repression at the L3 MARE operator, followed by the L1 MARE operator, with no binding indicated at the L2 MARE operator. If MmyR will not bind to the L2 MARE operator then it will not regulate *mmyRp* and its own expression. As there was no repressive advantage observed for having *mmyR* under an auto-regulatory control system it is possible that this second repressor is not auto-regulatory. MmyR may be regulating itself in some way, but it is evidently different to the direct control mechanism MmfR exerts on *mmfR*.

It should be noted however that these results are not quantitative and purely based on visual observations and so may not hold true when further analysed quantitatively.

### 6.3 Effect of the MMFs on MmfR When Under MARE Operator Regulation

In Chapter 4 it was seen that when any of the five MMFs were added to the L1+*mmfR* strain, there was an apparent release of MmfR and an increase in *luxCDABE* expression. MMF4 was added to the 11NY strain and the effects on luminescence observed. The results from this assay can be found in Figure 6.3. Data for 11NY was compared to the M145 negative control and L1+pCC4 positive control.



**Figure 6.3. Relative level of luminescence produced by the *lux* operon under the control of *mmfLp* and MmfR (under its own auto-regulatory control) in the presence and absence of 400 nM MMF4 over time**

*The level of luminescence was calculated as a ratio compared to that of the M145 strain*

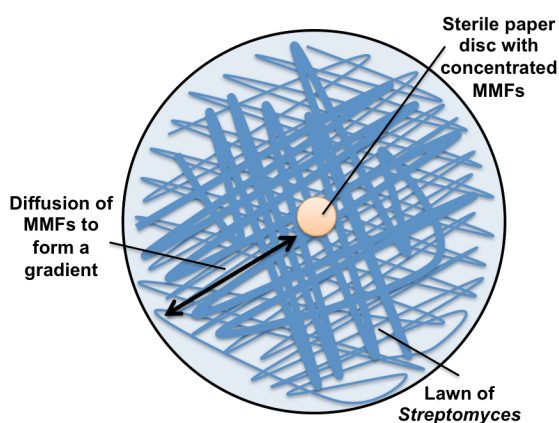
The 11NY strain was more representative of the wild type system than L1+*mmfR*, i.e. *mmfR* is being controlled by its native promoter rather than *ermEp\**. Despite this, it can be seen from Figure 6.3 that the expression of the *lux* genes was not inducible upon the addition of MMFs in the 11NY strain. In the presence of MMF4, the levels of luminescence remain very close to those produced by the M145 negative control. The reason for this is again likely to be the auto-regulatory nature of MmfR. Upon the release of MmfR from the operator by MMF4, there is only a short window of *lux* expression before more MmfR is also made, repressing the expression of the luciferase genes almost immediately. Finding this very short window of luminescence was not experimentally practical.

A concentration of 400 nM MMF4 was used for the 11NY data collected, this is the same concentration as was used for the experiment displayed in Figure 3.10 where there was an observable increase in luminescence by L1+*mmfR* in the presence of MMF4. As the L1+*mmfR* strain has *mmfR* under the control of the constitutive *ermE\** promoter, it means that there are much longer lasting effects of de-repression by the MMFs. MmfR is produced at a relatively constant rate, where expression is presumed to be unaffected by the addition of the MMFs and is independent of its own repressive activity.

Trials where the concentration of the MMFs was increased did not show any more release of MmfR in the 11NY strains (data not shown here).



It is unclear how these findings on the lack of observable effect of the addition of the MMFs in the 11NY strain is representative of what actually occurs in the wild type system. Studies with *mmfLHP* mutants have shown that the addition of MMFs will induce the production of methylenomycin (Table 6.1).(71) There are a number of possible explanations on how this release of MmfR occurs in the wild type system when it was not seen in this reporter system. The first is that there was a small window of MmfR release (and therefore increased luminescence) but due to readings only being taken every 24 hours, this period of *luxCDABE* expression was missed. A small window of expression may be enough to result in the expression of *mmyB* and thereby switching on the entire biosynthetic cluster. Another possible explanation is that in the wild type system it is a very specific threshold concentration of the MMFs needed to release MmfR. A concentration too far above or below this specific threshold level may not alter the MmfR/MMF/operator feedback loops in a way that results in methylenomycin production. This precise threshold concentration of MMFs is not known but if more time was available, a gradient of different MMF concentrations could be trialled with the 11NY strain to check this hypothesis (see Figure 6.4). A final consideration as to how MmfR is released in the wild type system is that other factors may also be at play, altering the window of time during which MmfR is released. In the M145 strain used in this assay, none of the methylenomycin BGC is present except those genes added in the *lux* vectors and so any extra regulation by this cluster would also be lacking.



**Figure 6.4. Schematic of proposed assay to measure the effects of a gradient of MMFs on the repression of the *lux* operon when *mmfR* expression is under auto-regulatory control**

*The lux operon is under the control of mmfLp and the LI MARE operator. If there is a narrow threshold window of MMF concentration that will induce the strains containing the 11NY vector, this will be seen as a circle of luminescence on the plate.*

## 6.4 Effect of the MMFs on MmyR When Under MARE Operator Regulation

When MMF4 was trialled with the sp105 strain, no obvious change in luminescence was observed (data not shown here). This is consistent with the data reported in Section 5.4 where the addition of MMF4 produced no significant changes in luminescence for strains containing the L1 intergenic region with *mmyR* under the control of *ermEp\**. Having *mmyR* under the control of the L1 MARE operator and *mmfRp* would not be expected to change this phenotype as there should be no change in the ligand-binding pocket of MmyR.

Although MmyR may not directly control the expression of its own gene, there is a possibility that it has an indirect regulatory effect on its own activity. This hypothesis relates to putative interactions with the MMFs and will now be discussed.

In Chapter 5, a reduction in *lux* expression was observed for *mmyR* strains in the presence of 100  $\mu$ M MMF2, 4 and 5 at the L3 MARE operator. It appears therefore that in some cases, the MMFs may improve the repressive ability of MmyR. This however was a property not seen for MmyR binding at the L1 MARE operator. It is possible that weaker binding of MmyR seen at the 24 bp L1 MARE operator sequence is not strong enough for the addition of 100  $\mu$ M MMFs to cause a significant change in luminescence. If this is the case, a higher concentration of MMFs added to the L1+*mmyR* may reveal higher levels of repression. Alternatively, MmyR may bind to the two operator sites in a slightly different conformation. The L1 and L3 MARE operators share less than 63% identity (137) and it is possible that this difference could have enough of an effect on the MmyR tertiary structure to alter the way that the MMFs can enter its ligand-binding pocket.

There is a biological explanation for why the wild type system may be set up so that at the L1 MARE operator, MMFs would not enhance MmyR binding. This is due to the genes actually being regulated at the L1 MARE site; *mmfR* and *mmfLHP*. In particular, *mmfLHP* which code for MmfLHP, the enzymes used in the assembly of the MMFs.

If the MMFs could bind to MmyR at the L1 MARE operator and make it a better repressor, it would create a negative feedback loop. A MMF-induced enhancement of MmyR repression would switch off the production of more MMFs. This decrease in the concentration of the MMF ligands would then lead to MmyR becoming a less efficient repressor again and a return the system to the over-production of the MMFs. This would then become a feedback cycle alternating between enhanced and reduced MmyR activity, something that would clearly not be productive whilst trying to switch off methylenomycin production. It is therefore not logical for MmyR to only be able to optimally switch off MmfLHP production

at the L1 MARE operator in the presence of the MMFs. The more MmyR represses these genes, the more it would be down regulating its own effects and thus preventing further repression. The hypothesised different conformations of MmyR therefore may make it indirectly auto-regulatory, based on its resultant control over the levels of MMFs produced, which are possibly needed for its full repressive activity at the L3 MARE operator.

This inference is entirely hypothetical however, and purely based on there results collected in this projects investigation. It does appears however that the different MARE operator sequences are more important in controlling the promiscuous effects of MmyR than MmfR, the latter of which appears to be less selective and have a more similar role at each operator site.

It is very hard to shed light on the exact role of MmyR, which proves to be much more elusive and complex than MmfR regulation. One area that may help develop hypotheses about MmyR activity (and MmfR) is that of a mathematical model of the methylenomycin regulatory system.

## 6.5 Mathematical Modelling of the MmfR/MMF/MARE Operator Regulatory Network

During this PhD, work has been carried out in conjunction with the Department of Engineering at the University of Warwick where a PhD student, Jack Bowyer, was mathematically modelling the methylenomycin regulatory system. Here Bowyer *et al.* created five different architectures of possible regulatory systems and then compared the output of the models with a range of laboratory generated experimental findings.<sup>(79)</sup> This collaboration involved extensive discussion between the Department of Engineering and the School of Life Sciences as well as the sharing of raw data to be compared with the mathematical models. From this analysis, one particular architecture was chosen as being the most representative of the MmfR/MmyR regulation. This architecture accounts for reversible MmfR and MmyR binding at the MARE operators, where expression of these two regulators is controlled within the methylenomycin cluster.

Other alternative architectures trialled has a number of variable components including the regulation of MmfR by a constant and external parameter (mirroring the control by *ermEp\**, investigated in this thesis) or having MmfR only being able to bind to the MMFs in a MmfR/DNA complex rather than in its apo form when it is free in solution.

The experimental data used to test these architectures focussed on the seminal methylenomycin paper by Sean O'Rourke *et al.* from 2009.(71) In this paper, a number of mutants were created for methylenomycin cluster elements, for example *mmfR* and *mmyR* knockouts. The genetic and metabolic profiles for these mutants were then assessed by O'Rourke *et al.*, shedding light on the role of the individual genes of the 19 kb methylenomycin cluster. A summary of the findings of these assays can be found in Table 6.1. Here it can be seen that methylenomycin production increases drastically in the *mmyR* mutants but not *mmfR* mutants. The mutants for transcriptional activator MmyB on the other hand, show repressed methylenomycin production but increased levels of the MMFs. Only mathematical models that presented these phenotypes as outputs when altering different parameters were deemed to be representative of wild type MmfR/MmyR regulation.

**Table 6.1. Phenotypes of methylenomycin cluster mutants**

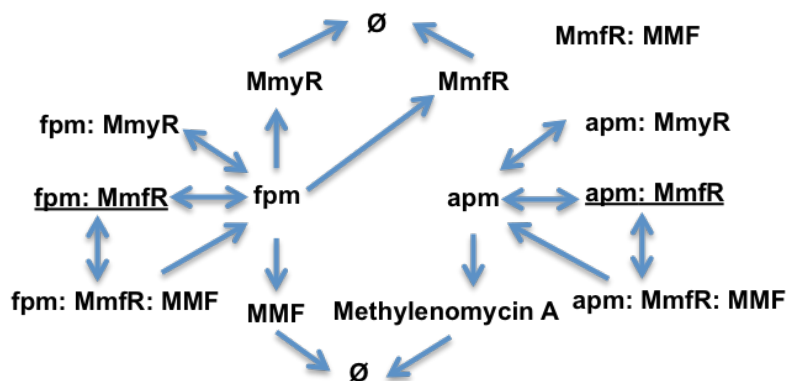
*The phenotypes displayed were reported by Sean O'Rourke et al.(71)*

*Mm = methylenomycin*

| Strain | Gene deletion                              | Mm production | MMF production | MMF responsive |
|--------|--|---------------|----------------|----------------|
| J1506  | -  | +             | n/a            | n/a            |
| J2629  | $\Delta mmyR::aac(3)IV$                    | +++           | n/a            | n/a            |
| J2650  | $\Delta mmfP::scar$                        | ++            | n/a            | n/a            |
| J2642  | $\Delta mmfH::scar$                        | -             | -              | +              |
| J2643  | $\Delta mmfL::scar$                        | -             | -              | +              |
| J2635  | $\Delta(mmfp \rightarrow mmfL)::aac(3)IV$  | -             | -              | +              |
| J2636  | $\Delta mmfR::aac(3)IV$                    | -/+           | n/a            | n/a            |
| J2637  | $\Delta(mmyR \rightarrow mmfR)::aac(3)IV$  | +++           | n/a            | n/a            |
| J2638  | $\Delta(mmyR \rightarrow mmfL)::aac(3)IV$  | -             | -              | ++             |
| J2639  | $\Delta(mmfp \rightarrow mmfR)::aac(3)IV$  | -/+           | -              | -              |
| J2641  | $\Delta mmyR::aac(3)IV, \Delta mmfR::scar$ | +++           | n/a            | n/a            |
| J2644  | $\Delta mmyB::scar$                        | -             | +              | -              |

In all the model systems created by Jack Bowyer, the methylenomycin BGC was split into two parts. The first is the *apm* or antibiotic producing module, this encompasses *mmyTOG*, *mmyBQEDXCAPK* and *mmyYF* and is controlled by the L3 MARE operator. The second module is the *fpM* or the furan producing module, this includes *mmyR*, *mmfR* and *mmfLHP* which are controlled by both the L1 and L2 MARE operators.

The diagram in Figure 6.5 shows the model architecture that was shown to produce results closest to those in the wild type system. This figure was adapted from a conference paper written by Jack Bowyer *et al.* in 2016 to go with his research findings.(79) In this model, MmfR is produced by the *fpM* in an auto-regulatory fashion and is bound to the *fpM* and *apm* in the 'resting state'. Both MmfR and MmyR are modelled to bind and regulate the two modules but only MmfR was modelled to bind and be released by the MMFs.



**Figure 6.5. Model architecture of the methylenomycin biosynthetic cluster**

*apm* – antibiotic producing module, *fpm* – furan producing module,  $\emptyset$  - degradation

Reversible and irreversible reactions are represented by double and single ended arrows respectively.

Underlined elements denote those with initial non-zero concentrations.

The modelling from this architecture is very simplified compared the wild type system however. Although the chosen model architecture does not contradict any of the findings from this thesis, it combines the L1 and L2 MARE operator regulation as well as not factoring in the TTA codons or any other potential transcriptional control. It also does not consider all potential MmyR ligand interactions, alternative DNA binding sequences or *mmyR* regulation that is separate from the *fpm*. In addition, the modelling did not consider the findings found on variations in operator binding for MmyR compared to MmfR as well as the differing binding potentials of the five MMFs. Finally, the initial construction of each alternative architecture was also largely based on assumptions present at the beginning of the work e.g. that the MMFs will only bind MmfR when it is bound to the MARE operators and that MmyR will not be released by the MMFs. It is therefore not possible to draw firm conclusions from this work, instead the model produced offers an indication of the regulation in the methylenomycin network and helps to point researchers in the right direction when planning future experiments.

Despite its limitations however, the chosen model architecture does suggest that the self-regulatory function of MmfR is key to producing results in line with those for the wild type. Models that showed constant *mmfR* expression under the control of separate regulatory regions did not match experimental data. It was only when *mmfR* is under the control of the *fpm* that the system was closest to the phenotype of wild type one. The model also predicted that the MMFs would not bind to MmfR when it was free in solution, they would only interact when MmfR was bound to the *apm* or *fpm*, then causing its release upon a conformational change.

Models where MmyR was released by the MMFs did not produce results in line with experimental data. This matches the findings of Section 5.4, where no significant de-repression of *luxCDABE* was seen in the presence of any of the MMFs for the L1+*mmyR* strain. (De-repression was not seen in any of the MMFs trialled in the L2+*mmyR* or L3+*mmyR* strains either.) However, the models investigated by Bowyer *et al.* did not consider an increased binding of MmyR in the presence of the MMFs and so no conclusions can be drawn on this aspect of regulation using this model.

The modelling also excluded the feasibility of architectures where MmyR was released from the *fpm* upon binding to methylenomycin. The experimental findings from this thesis using the luciferase assay into MmyR/methylenomycin binding were inconclusive (Section 5.4.2). The model indicates that this hypothesis can most likely be discounted but some future experimentation may be helpful with validating this aspect of the chosen architecture.

There are plans in the future to use findings from this thesis, including those collected in the luciferase assay on MMF  $K_d$  and  $B_{max}$  values and the binding affinities of MmfR and MmyR for each of the MARE operators, in a more rigorous validation of the currently proposed mathematical model. The inclusion of this extra data should help to give the model more capacity for the conclusions that can be drawn from it and increase the likelihood of it being biologically representative and reliable.

By combining all the information collected via this mathematical model as well as the experimental findings from this and the previous three chapters it is possible to bring together the findings on MmfR repression in combination with those of MmyR and propose how they may work together to regulate methylenomycin biosynthesis.

## 6.6 Proposed Function of MmfR and MmyR In Regulating Methylenomycin Biosynthesis

Figure 6.6 is a schematic showing a visual representation of the proposed regulation of the methylenomycin biosynthetic cluster by MmfR and MmyR based on all experimental evidence revealed in this project. Parts A-G of this figure are referred to in the following text, where this proposed pathway is explained.

In Section 4.3, MmfR was shown to bind best at the L1 MARE operator, controlling the expression of itself and that of *mmfLHP*. MmfR will also bind and repress at the L2 and L3 MARE operators but with less strength, thereby regulating the expression of *mmyR* and the operons starting with *mmyB* and *mmyY* (A). There is some leakiness in the system at this point

but MMF levels remain below the threshold needed to release MmfR and result in methylenomycin biosynthesis.

Upon production, or detection by quorum sensing, of a threshold level of the MMFs and the consequent release of MmfR from all three operators (B), the five operons will be expressed and methylenomycin biosynthesis begins.

The release of repression from the *mmfLHP* and *mmfR* intergenic region results in the production of the enzymes needed to produce the MMFs as well as MmfR. Two MmfR monomers are needed to repress at one MARE operator site, however each MmfL, H and P produced will result in the synthesis of thousands of MMFs molecules generating a positive feedback loop in MMF over-production (C). Although MmfR is also being produced it is continuously released by the MMFs.

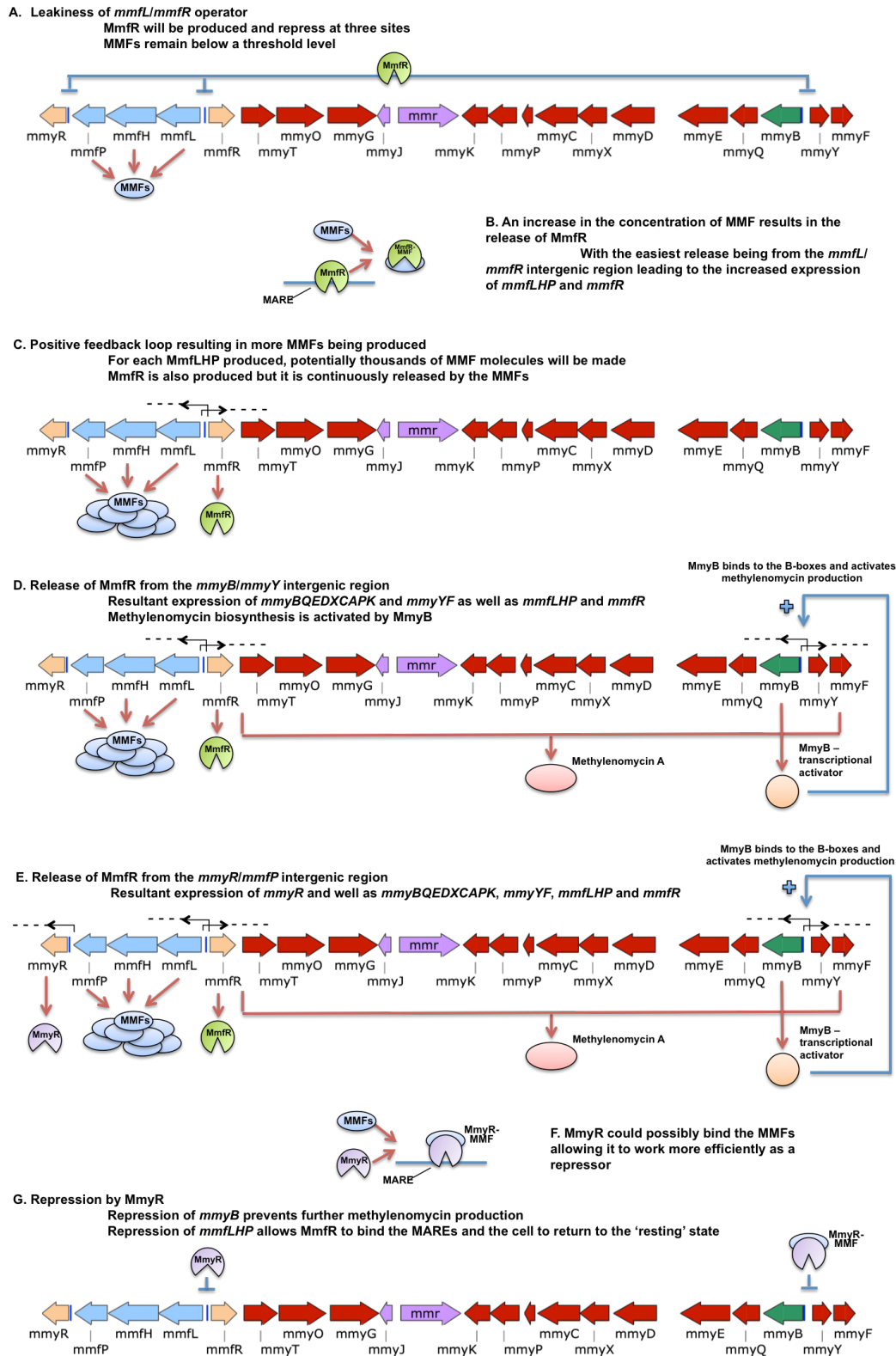
The research in this investigation has implied that the repression of *mmfR* and *mmfL* at the L1 MARE operator is most easily released (B), followed by *mmyB* and *mmyY* and then finally *mmyR* (Table 4.1). This potentially causes lag in the expression of these operons with a potentially later expression of the *mmyB* and *mmyY operons* (D) followed closely by the *mmyR* gene (E) (see also later discussion on page 133).

Once MmyR has been produced, my suggestion is that it will then bind the MMFs (F), allowing it work more fully as a repressor in combination with the right genetic sequence for DNA binding. MmyR/MMF binds the L3 MARE operator to regulate the production of transcriptional activator MmyB and biosynthetic enzyme MmyY and the operons they come from, starting to switch off methylenomycin production once it is no longer needed. As mentioned in the introduction, it is hypothesised that the main role of MmfR/MmyR/MMF regulation may be to control the expression of *mmyB*, this proposed pathway is in line with this hypothesis therefore.

MmyR also appears to cause some repression at the L1 MARE operator (further regulating *mmfR* and *mmfLHP*) but this does not appear to be influenced by the presence of the MMFs at the concentrations trialled (G). At this point MmfR is no longer significantly repressing the production of MmyR due to its release by the MMFs, and MmyR does not regulate itself at the L2 MARE operator leading to huge amounts of this second repressor being produced. Due to this lack of *mmyR* repression it is likely that once expression has begun, MmyR production will not be switched off again until the entire biosynthetic pathway has been turned off and MMF production/detection has ceased.

As MMF levels drop again, repression by MmfR will increase (and MmyR decrease). MmfR has been shown to have least affinity for the L2 MARE operator and so it is possible that this is the last place to be repressed, allowing *mmyR* expression (and therefore its repressive activity) to continue for longer than the expression of other elements of the BGC, returning the methylenomycin cluster to its 'resting state' (A).





**Figure 6.6. Schematic of proposed MmfR/MmyR regulation of the methylenomycin biosynthetic cluster**

*This schematic is based on the findings of the luciferase assay utilised in this project.*  
**Assigned functions:** *mmfR* and *mmyR* – transcriptional repressors, *mmfLHP* – methylenomycin furan biosynthetic genes, *mmyJ* and *mmr* – methylenomycin resistance, *mmyB* – transcriptional activator, *all others* – methylenomycin biosynthesis

There are of course still gaps in this proposed pathway and it does not account for the TTA codon, sigma factors or any potential extra signalling roles of the MMFs or any other possible transcriptional regulation. However, this proposed pathway does not contradict the mathematical model produced and combines all of the experimental findings of the luciferase assay.

Research by O'Rourke *et al.* indicated that transcripts of *mmyR*, *mmfL* and *mmfR* were produced before *mmyB* which in turn was produced before any of transcripts for biosynthetic genes.(71) This does show slight differences to the proposed pathway here, which is based on ease of MmfR release from each operator. In particular, the presence of *mmyR* transcripts early on is something not proposed in Figure 6.6. It is unclear however whether these transcripts observed by O'Rourke *et al.* were due to the leakiness of the system rather than expression having been actively switched on and whether a threshold level of MmyR is needed before it can have its regulatory function. The presence of a *mmyR* transcript also does not mean that it is binding to the MARE operators and is exerting its repressive activity, especially as it has been indicated to have weaker binding than MmfR and therefore may take more time. Alternatively, there may be extra translational regulation of the system that is as yet unknown which could explain these differences.

An extra comment on the findings of O'Rourke *et al.* is that the release of MmfR from the L2 and L3 MARE operators seen in Section 4.3 was at times very similar or the order reversed (depending on whether compared to its negative or positive controls), so it is possible that there is not the lag in the production of their associated genes shown in Figure 6.6, and steps D and E therefore may be either combined or reversed. However, the proposed pathway still remains even if MmfR has the same strength of release at the L2 and L3 sites. If *mmyR* is expressed alongside the *mmyB* and *mmyY* operons or even before them, the enzymes and activators will catalyse many more reactions than the structural MmyR which can only repress one site at a time. Only a short window of de-repression of the *mmyB* and *mmyY* operons may be needed for the activation of methylenomycin biosynthesis, before MmyR accumulation and resultant repression occurs to a significant level and prevent the production of more enzymes.

## 6.7 Conclusions and Summary

It is clear from this investigation that methylenomycin regulation is very complex. In the wild type system, MmfR production and its release by the MMFs will be influenced by not only its own auto-regulation but also MmyR repression, quorum sensing and the presence of

endogenous and exogenous MMFs as well as potential interactions of sigma factors and other regulator proteins that are yet to be discovered. Although the findings of this chapter do not show the whole picture of MmfR auto-regulation it is clear that when under the control of its wild type operator and promoter, much tighter regulation of gene expression is obtained. MmyR on the other hand does not appear to have a direct effect on its own expression. The role of MmyR is still much less well understood than that of MmfR.

Moving forward from the luminescence work and onto the development of a novel inducible expression system, it appears clear that as we much better understand the role of MmfR, it is a much more suitable component for an optimised inducible system. Of the three MmfR binding intergenic regions, the L1 region (between *mmfL* and *mmfR*) appears to be most suitable, with the greatest repression by MmfR seen as well as the greatest increase in luminescence produced upon the addition of a single concentration of MMFs (Figure 4.3). Of the five MMFs, the furans with the branched alkyl chains; MMF1 and 3, appeared to have the greatest binding potential to MmfR at the L1 MARE operator (Figure 4.19) and are therefore obvious choices as ligands to induce the system that will be developed. Work was moved forwards to start developing this novel inducible expression system for use in streptomycetes. In this system the MmfR/MARE operator regulatory sequence was placed upstream of a gene of interest. The auto-regulation of MmfR was chosen to be excluded from the system, due to the poorly understood nature of how MmfR is released by the MMFs when under this control, instead regulation by *ermEp\** was used, as is described and explained in the following chapter.

## 7 Development of a Novel Inducible Expression System for Streptomyces

### 7.1 Aims at Strategy of Investigation

#### 7.1.1 Existing Heterologous Expression Systems

There are a number of well-known and validated *E. coli* expression systems that are utilised for controlling the production of a huge variety of proteins.(102) *E. coli* is not always suitable for every protein of interest however, with a number of polypeptides proving to be insoluble or difficult to purify in these Gram-negative bacteria. It can be hard to predict the conditions a specific protein needs for the synthesis of an active product, with problems encountered in proteins folding into the correct conformation, poor expression levels as well as an inability to carry out the necessary post-translational modifications. It is unsurprising therefore that the same system cannot be used to achieve the successful purification of soluble proteins in all cases. Expression systems have been developed for a number of different types of organism, increasing the variety of conditions present in hosts and broadening the number of recombinant genes that can be expressed but again this does not cover all cases. The Gram-positive *Streptomyces* have shown promise as heterologous expression hosts, with the possibility of improved expression of genes from other GC high or Gram-positive bacteria.(139, 140) There is hope that the MmfR/MMF/MARE operator system, analogous to LacI/IPTG/lac operator, could possibly be used to provide an alternative inducible expression system for the overexpression of recombinant genes.

#### 7.1.2 MmfR/MMF/MARE Operator as an Inducible Expression System in Gram-Positive Bacteria

In the previous four chapters, the MmfR/MMF/MARE operator system from the methylenomycin cluster of the SCP1 plasmid of *S. coelicolor* has been investigated. It has shown promise as an inducible expression system in terms of promoter strength as well as removal of MmfR repression by the furans ligands. Using these findings, research has been carried out in two main areas, firstly an investigation into the choice of a *Streptomyces* host suitable for an inducible expression system. Secondly, preliminary work was undertaken to start to develop an inducible expression system that can be used to control the production of recombinant proteins that are otherwise difficult to obtain e.g. because they are toxic to the host.

To turn the modified *lux* system into a novel inducible expression system that can be adapted to study and produce recombinant proteins of interest, using a strain with a reduced genome could be advantageous, to decrease the chances of background interactions with the system as

well as metabolically streamlining the host to conserve resources for the over production of the protein of interest. Up until now, all of the investigations into MmfR/MMF/ operator using the luciferase assay were done using *S. coelicolor* M145. This is a genetically reduced derivative of the wild type and model organism *Streptomyces coelicolor* A3(2). The M145 variant was developed via the removal of the SCP1 and SCP2 plasmids. Conveniently, all of the methylenomycin cluster, including all biosynthetic, regulatory and resistance genes are found on the SCP1 plasmid. This made M145 a suitable host strain for the luciferase reporter gene assay with no background interactions from the methylenomycin cluster being present in this strain. In particular, the absence of the native *mmfR*, *mmvR* and *mmfLHP* were particularly beneficial. Components of the methylenomycin cluster were added as and when needed. For the development of the novel inducible expression system, a further investigation was carried out into whether an even more streamlined host could be achieved, trialling *Streptomyces albus* as a potential superior expression host.

## 7.2 *Streptomyces albus* as a Potential Host

### 7.2.1 Introduction to *S. albus*

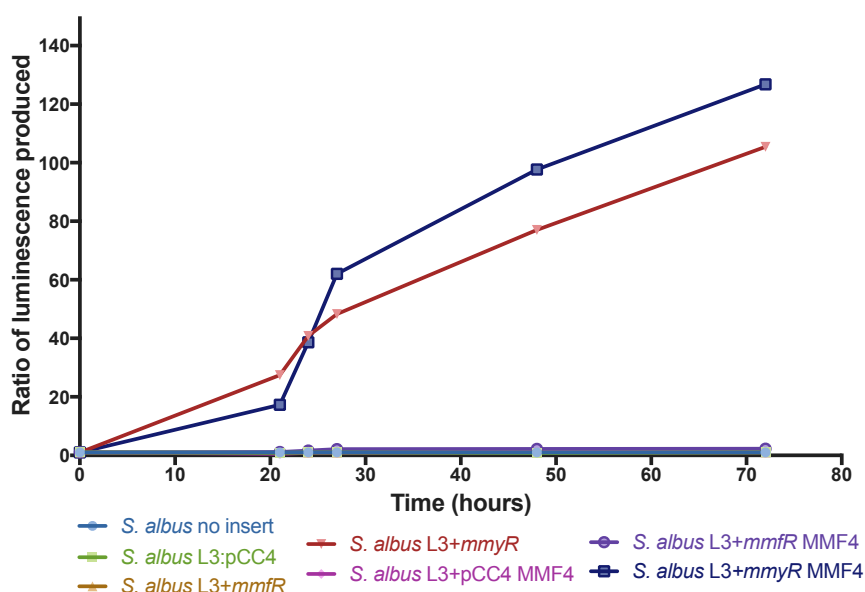
*S. albus* has one of the smallest known genomes of any in the streptomycete genus at only 6.8 Mb.(27, 93) This strain provides a very interesting case study when looking at phylogenetic relationships and the evolution of genetic elements due to the natural removal of any apparently unneeded genetic material from the genome. *S. albus* has recently started to be widely studied with the potential of it being used as a premium host for heterologous expression of natural products.(93, 141) In this report, genomic, transcriptomic and *in vivo* analyses have been carried out on *S. albus* strain J1074 to better understand how it can be used as a super host and whether there will be any background interactions with MmfR/MMF/MARE operator from native gene expression.

### 7.2.2 Luminescence Assay in *S. albus*

To check *S. albus* for suitability as an expression host, the previously used luciferase reporter gene assay was transferred over to this strain. For this strain to be a suitable host for the MmfR/MMF/MARE operator inducible expression system, results collected for the *lux* strains created would need to be akin to those collected for *S. coelicolor* M145. Comparable results would indicate that the MmfR/MMF/MARE operator system works in *S. albus* as well as *S. coelicolor* without any background interactions from existing *S. albus* networks.

The L3 vector (containing *mmvBp*) used in the earlier luciferase assay as well as *mmfR* and *mmvR* (in pKMS01 and 03) were all integrated into *S. albus* creating *S. albus* L3+*mmfR* and

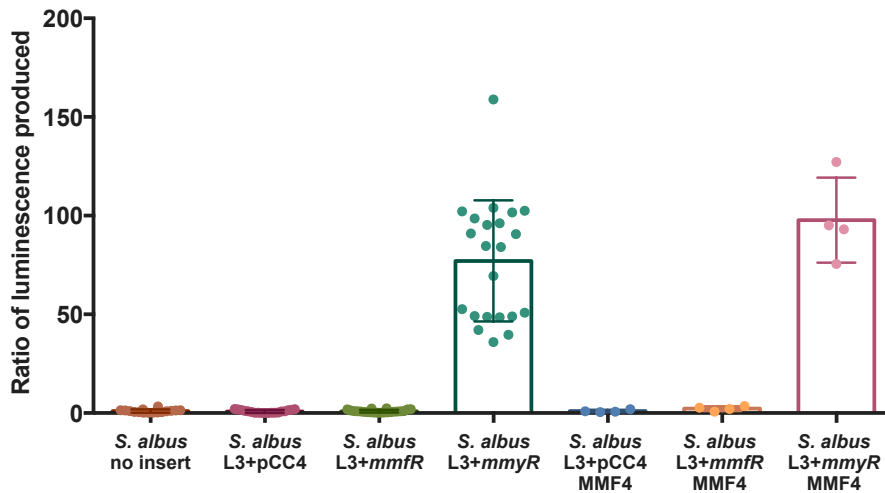
L3+mmyR strains.<sup>1</sup> These were then analysed via the measurement of luminescence produced using the Photek CCD camera in the same way that the *S. coelicolor* M145 strain was in the previous four chapters. Measurements were again taken at 21, 24, 27, 48 and 72 hours growth and the luminescence compared to a negative control strain with no *luxCDABE* insert and a positive control strain with no repressor (L3+pCC4). The findings of this investigation are shown in Figure 7.1 and Figure 7.2. Figure 7.1 shows the luminescence produced at the five time points over 72 hours for all samples and Figure 7.2 shows a bar chart that compares luminescence at just the 48 hour time point. A *t*-test analysis was then run with data from Figure 7.2, the results of which can be found in Table 7.1.



**Figure 7.1. A comparison of luminescence produced by the *luxCDABE* operon in *S. albus*, as regulated by the presence and absence MmfR, MmyR and MMF4 over 72 hours**

Average light production is calculated as a relative ratio of luminescence produced by the *S. albus* negative control with no insert (giving this sample a value of 1). **Strains used:** *S. albus* – wild type negative control strain, *S. albus* L3+pCC4 - *luxCDABE* under the control of mmyBp and pCC4, *S. albus* L3+mmfR – *luxCDABE* under the control of mmyBp and mmfR under the control of ermEp\* (pKMS01), *S. albus* L3+mmyR – *luxCDABE* under the control of mmyBp and mmyR under the control of ermEp\* (pKMS03)

<sup>1</sup> This nomenclature is the same as that which was used for the equivalent investigation in *Streptomyces coelicolor* M145, see Chapter 3.



**Figure 7.2. Bar chart comparing luminescence produced by the *luxCDABE* operon in *S. albus*, as regulated by the presence and absence MmfR, MmyR and MMF4 at 48 hours**  
*Average light production is calculated as a relative ratio of luminescence produced by the S. albus negative control (giving this sample a value of 1). Strains used: same as Figure 7.1*

**Table 7.1. A *t*-test analysis of the luciferase assay results collected from *S. albus* data at 48 hours**

*Average light production is calculated as a relative ratio of luminescence produced by S. albus with no insert (giving this sample a value of 1). The p-value was also calculated based on S. albus with no insert. Strains used: same as Figure 7.1*

| Strain                       | p-value  | Significant difference? | Average light production at 48 hr (R.R) |
|------------------------------|----------|-------------------------|---|
| <i>S. albus</i> L3+pCC4      | 5.06E-01 | FALSE                   | 0.82                                    |
| <i>S. albus</i> L3+mmfR      | 8.27E-01 | FALSE                   | 0.95                                    |
| <i>S. albus</i> L3+mmyR      | 8.64E-12 | TRUE                    | 77.07                                   |
| <i>S. albus</i> L3+pCC4 MMF4 | 9.87E-01 | FALSE                   | 1.01                                    |
| <i>S. albus</i> L3+mmfR MMF4 | 2.27E-02 | TRUE                    | 2.25                                    |
| <i>S. albus</i> L3+mmyR MMF4 | 1.32E-13 | TRUE                    | 97.72                                   |

It was found that the MmfR/MMF/MARE operator inducible expression system in *S. albus* was not comparable to that in *S. coelicolor* M145 indicating that *S. albus* is not a suitable heterologous expression host for this particular expression system. As can be seen in both Figure 7.1 and Figure 7.2, the *S. albus* L3+pCC4 positive control and L3+mmfR both appear to have almost entirely repressed levels of luminescence. The L3+mmyR strain on the other hand produces high levels of luminescence both in the presence of the MMFs and without. The findings of the *t*-test in Table 7.1 show that there is a significant increase in luminescence produced by the L3+mmfR strain in the presence of MMF4 but this is only minimal with twice as much luminescence being produced as is for the control strain with no *lux* insert.

This is still nowhere near close to the levels of luminescence produced by *S. albus* L3+*mmyR* or the levels of induction seen for the equivalent system in *S. coelicolor*.

A possible interpretation as to why the L3+pCC4 and L3+*mmfR* strains produce no luminescence is that a protein from the *S. albus* genome could be causing repression at the MARE operator sequence. The L3+*mmyR* strain still produces high levels of luminescence and the conclusion inferred from this is that MmyR can bind genetic elements in the *S. albus* genome and thereby repress the expression of this native TetR that might otherwise bind the MARE operator.

Where it is possible that a native protein from *S. albus* is binding to the MARE operator sequence and preventing the expression of the luciferase genes, it appears unlikely that this protein also has the correct binding pocket for the MMFs and for this reason the addition of 100  $\mu$ M MMF4 causes little or no induction of *lux* expression. This potential native *S. albus* TetR, homologous to MmfR/MmyR, is discussed further in the following paragraphs.

It is also necessary to consider that, despite the plasmid inserts in *S. albus* being checked by PCR and the ex-conjugants gaining the selective apramycin and hygromycin resistance from the vectors inserted, it is still entirely possible that the inserted genes are not being expressed properly in *S. albus*. As mentioned previously, *S. albus* is known to have a streamlined genome due to genetic reshufflings and deletions of ‘unnecessary’ genes, this is something which may have occurred to the L3, pKMS01 and pKMS03 inserts after they had been screened by PCR. In hindsight it may have been helpful to run extra screenings of the ex-conjugants during and after the luciferase assays to check for maintenance of the insert. Until this has been investigated further therefore, the analyses just discussed should be studied with caution.

### 7.2.3 Investigating potential GBL-related TetRs in *S. albus*

#### **Discovery of SSHG\_01258**

A literary and database investigation was carried out to look in the *S. albus* genome for a potential native TetR family member that would bind to the methylenomycin cluster MARE operator sequences. Zaburanyi *et al.* described the sequencing of the streamlined *S. albus* genome, where duplicates and redundant genes have efficiently been removed (93) meaning that this strain has a smaller genome than even many artificially reduced streptomycete strains. In particular, they noticed that no butenolide synthase or genes associated with the production of GBLs/AHFCA were present in the *S. albus* genome.(93) However, there is still a gene for a predicted GBL-binding TetR family member, that has not been lost during chromosomal rearrangements.(93) A BLAST sequence analysis revealed only one significant MmfR homologue, identified as SSHG\_01258, showing over 40% identity with the *S.*



*coelicolor* MmfR over 86% of its sequence (Table 7.3 and Figure 7.4). As no gene name was given in the Zaburanyi paper it can only be assumed that this is the GBL binding protein that they were referring to. No obvious homologue of MmyR could be found.

Through an analysis of the primary structure of SSHG\_01258 compared to MmfR and other TFRs, this potential MARE operator-binding protein could be better understood. In particular, the ligand and DNA binding domains were of particular interest to try and explain better the results collected in Figure 7.1 and Figure 7.2.

### TetR Family DNA Binding Motif of SSHG\_01258

The amino acid sequence analysis in Section 5.3.1 revealed that within the predicted 20 bp DNA binding region of MmfR, there was a conserved GAVYFH sequence found in MmfR and its orthologues from *S. venezuelae* and *S. avermitilis* whereas their paralogues showed an alternative conserved GALYGH sequence. Table 7.2 shows a comparison of the DNA binding motifs from these proteins with that of the hypothetical *S. albus* protein. It appears that SSHG\_01258 shows the conserved GAVYFH sequence found in MmfR and homologues SgnR and SAV\_2270. It is therefore plausible that SSHG\_01258 is indeed binding to the methylenomycin cluster operator site contained in the L3 vector and is repressing *lux* gene expression, analogous with the stronger binding to the MARE operator seen for MmfR compared to MmyR.

**Table 7.2. DNA binding region sequence of SSHG\_01258 compared to that of MmfR and MmyR and their orthologues**

| Protein name | p-value | Sequence (20 aa)     |
|--------------|---------|----------------------|
| SSHG_01258   | 6e-19   | SVNDISARSGRTSGAVYFHY |
| MmfR         | 8e-21   | SVKDVAERVGMTKGAVYEHF |
| SgnR         | 4e-20   | TLQDVAERAEMTKGAVYEHY |
| SAV_2270     | 3e-18   | TIKDIADGAEMTKGAVYEHF |
| SAV_2268     | 1e-19   | NLQNIADRIRLTKGALYGHF |
| GbnR         | 3e-18   | NLADITARTGLTKGALYGHF |
| MmyR         | 4e-18   | NLATVAVRTGMTKGALYGHF |

A further analysis of the seven homologues just discussed was carried out, looking at percentage identity as well as alignment scores. Results of these findings can be found in Table 7.3, Table 7.4 and Table 7.5.<sup>1</sup> It can be seen that there is a high level of identity across all seven amino acids, where MmfR shares 47% and 54% identity with its orthologues from *S. venezuelae* and *avermitilis* respectively, across 98% of their sequences. The *S. albus* homologue falls just below this with 41% similarity across 86% of its sequence. The BLAST

<sup>1</sup> Sequence similarities can be found in ‘Appendix C. Sequence Analysis of MmfR and Homologues’

alignment score for SSHG\_01258 (Table 7.5) also indicates that it falls within the same bands as SgnR and SAV\_2270 when compared to MmfR with all having scores well over 100.

As may be expected from TetR family members, the regions of highest percentage identity between the homologues falls across the DNA binding domain within the TetR type HTH motif. This therefore adds further to the possibility of all these homologues binding the same MARE operator sequences.

**Table 7.3. Percentage identity of amino acid sequence between MmfR and MmyR homologues from *S. venezuelae*, *S. avermitilis* and *S. albus***

|            | MmfR | MmyR | SgnR | GbnR | SAV_2270 | SAV_2268 | SSHG_01258 |
|------------|------|------|------|------|----------|----------|------------|
| MmfR       | 100  | 35   | 54   | 28   | 47       | 33       | 41         |
| MmyR       | 35   | 100  | 36   | 47   | 50       | 39       | 36         |
| SgnR       | 54   | 36   | 100  | 31   | 53       | 34       | 43         |
| GbnR       | 28   | 47   | 31   | 100  | 35       | 37       | 56         |
| SAV_2270   | 47   | 50   | 53   | 35   | 100      | 33       | 35         |
| SAV_2268   | 33   | 39   | 34   | 37   | 33       | 100      | 33         |
| SSHG_01258 | 41   | 36   | 43   | 56   | 35       | 33       | 100        |

**Table 7.4 Percentage coverage of analysis of amino acid sequence between MmfR and MmyR and their homologues from *S. venezuelae*, *S. avermitilis* and *S. albus***

| Query sequence | Alignment sequence |      |      |      |          |          |            |
|----------------|--------------------|------|------|------|----------|----------|------------|
|                | MmfR               | MmyR | SgnR | GbnR | SAV_2270 | SAV_2268 | SSHG_01258 |
| MmfR           | 100                | 51   | 92   | 66   | 85       | 75       | 89         |
| MmyR           | 54                 | 100  | 72   | 58   | 34       | 94       | 53         |
| SgnR           | 98                 | 75   | 100  | 72   | 91       | 81       | 90         |
| GbnR           | 72                 | 61   | 61   | 100  | 68       | 96       | 19         |
| SAV_2270       | 98                 | 32   | 83   | 61   | 100      | 74       | 77         |
| SAV_2268       | 75                 | 91   | 78   | 68   | 78       | 100      | 72         |
| SSHG_01258     | 86                 | 47   | 86   | 25   | 75       | 71       | 100        |

**Table 7.5. Total score of alignment, according to NCBI BLAST, for the amino acid sequences for MmfR and MmyR and their homologues from *S. venezuelae*, *S. avermitilis* and *S. albus***

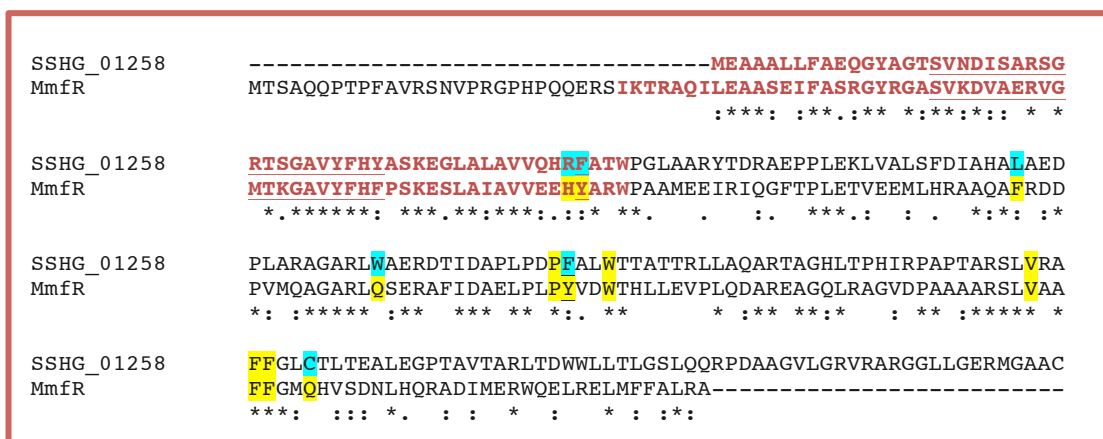
**Red** -  $\geq 200$ , **Pink** - 80-200, **Green** - 50-80, **Blue** - 40-50

The NCBI BLAST score is based on the standard parameters provided by the BLAST software. (31, 142, 143)

| Query sequence | Alignment sequence |      |      |      |          |          |            |
|----------------|--------------------|------|------|------|----------|----------|------------|
|                | MmfR               | MmyR | SgnR | GbnR | SAV_2270 | SAV_2268 | SSHG_01258 |
| MmfR           | 433                | 73   | 210  | 61   | 179      | 88       | 144        |
| MmyR           | 73                 | 413  | 72   | 95   | 62       | 90       | 64         |
| SgnR           | 197                | 72   | 401  | 68   | 165      | 69       | 140        |
| GbnR           | 63                 | 98   | 73   | 401  | 52       | 79       | 46         |
| SAV_2270       | 177                | 71   | 174  | 58   | 448      | 62       | 85         |
| SAV_2268       | 88                 | 105  | 75   | 78   | 59       | 412      | 55         |
| SSHG_01258     | 132                | 69   | 140  | 46   | 79       | 48       | 410        |

### Ligand Binding Domain of SSHG\_01258

In Figure 7.2, there did not appear to be any pronounced release of *lux* repression in the presence of MMF4. An analysis of the SSHG\_01258 sequence compared to MmfR specifically looking at the ligand binding pocket, shown in Figure 7.3, reveals that SSHG\_01258 does not contain the two tyrosine residues predicted to be key to MMF ligand binding in MmfR. There are instead two phenylalanine residues. Previous work with the bioluminescence assay (Section 4.4.4) showed that although the Y144F mutant appeared to improve ligand binding, when working with an MmfR Y84F mutant, a single tyrosine replacement with phenylalanine was enough to significantly reduce ligand binding. This and the fact that only 56% of the nine residues involved in hydrogen binding the MMFs are conserved between MmfR and SSHG\_01258 indicates why the addition of MMF4 did not release the repression of *luxCDABE* in *S. albus* (Figure 7.2 and Table 7.1). The small amount of luminescence induction seen in the presence of MMF4 was likely to be due to MmfR having bound the MARE operator in some cases whereas the lack of further induction is presumed to be an indication of the presence of SSHG\_01258 at the MARE operator, which is then not released by the MMFs. To confirm this hypothesis, *sshg\_01258* could be cloned into the pCC4 vector and added to the *lux* system and the effects on repression with and without the MMFs observed in *S. coelicolor* M145.



**Figure 7.3. Amino acid sequence comparison of MmfR and orthologue SSHG\_01258 from *S. albus***

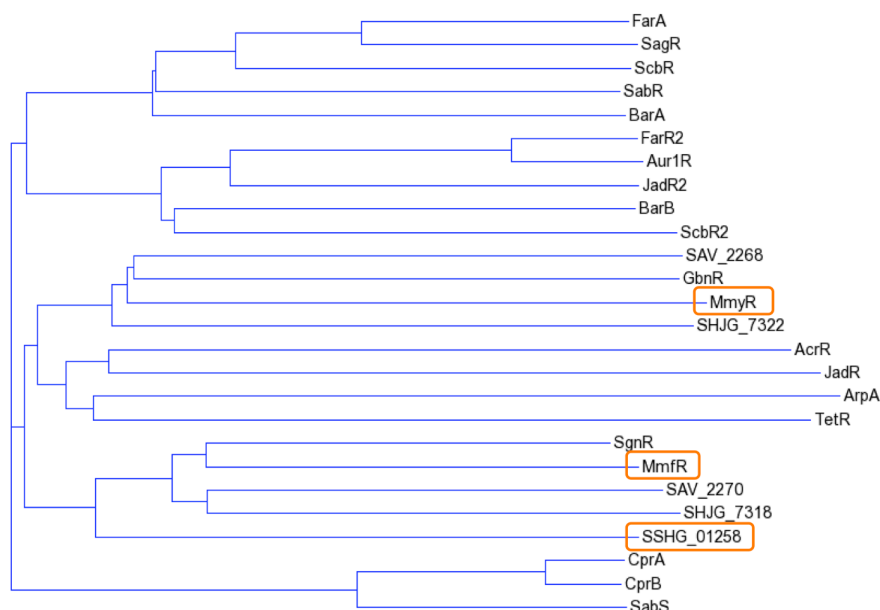
*TetR HTH domain shown in red and predicted DNA binding residues are underlined. Y85 and Y144 involved in hydrogen binding the MMFs in MmfR have also be underlined. Highlighted in yellow are all the residues shown to be involved in ligand binding in MmfR. The equivalents in the SSHG\_01258 homologue are also highlighted, with yellow indicating a match and blue indicating a mismatch.*

Following on from this analysis of the DNA and ligand binding domains of SSHG\_01258 it was of interest to see how this protein fits into the larger picture of TetR phylogeny and whether there are any other homologues with similarity of DNA binding motifs to MmfR, shedding light on the possible interactions of alternative repressors with the methylenomycin cluster operator sites. This was of interest in terms of choosing other possible hosts for the MmfR/MMF/MARE operator inducible expression system.

### Sequence Comparison of SSHG\_01258 with Other MmfR-Like and MmyR-Like Proteins

A phylogenetic analysis was carried out with a much larger array of TetR family members, particularly from streptomyces, and to include the GBL binding receptors mentioned in the introduction. The findings of these analyses can be found as a phylogenetic tree in Figure 7.4 with the homologies summarised in Table 7.6.

Following on from this investigation, the 20 bp DNA binding motifs from MmfR and MmyR were searched for in each of the 24 homologues analysed in Figure 7.4. The alignments of these sequences are presented in Table 7.7.



**Figure 7.4.** Phylogenetic tree (144) showing the relationship between MmfR, MmyR, SSHG\_01258 and other TetR family homologues across their amino acid sequences

**Table 7.6.** Summary of homology between TetR family members across different bacterial species

| <i>Protein name</i> | <i>Strain</i>                             | <i>Homologue</i> |
|---------------------|---|------------------|
| MmfR                | <i>Streptomyces coelicolor</i> A3(2)      |                  |
| MmyR                | <i>Streptomyces coelicolor</i> A3(2)      |                  |
| SgnR                | <i>Streptomyces venezuelae</i> ATCC 10712 | MmfR             |
| GbnR                | <i>Streptomyces venezuelae</i> ATCC 10712 | MmyR             |
| SAV_2270            | <i>Streptomyces avermitilis</i> MA-4680   | MmfR             |
| SAV_2268            | <i>Streptomyces avermitilis</i> MA-4680   | MmyR             |
| SHJG_7318           | <i>Streptomyces hygroscopicus</i>         | MmfR             |
| SHJG_7322           | <i>Streptomyces hygroscopicus</i>         | MmyR             |
| SSHG_01258          | <i>Streptomyces albus</i> J1074           | MmfR             |
| CprA                | <i>Streptomyces coelicolor</i> A3(2)      |                  |
| CprB                | <i>Streptomyces coelicolor</i> A3(2)      | CprA             |
| ScbR                | <i>Streptomyces coelicolor</i> A3(2)      |                  |
| ScbR2               | <i>Streptomyces coelicolor</i> A3(2)      |                  |
| TetR                | <i>Escherichia coli</i>                   |                  |
| ArpA                | <i>Streptomyces griseus</i>               | TetR             |
| AcrR                | <i>Pantoea ananatis</i> LMG 20103         | TetR             |
| JadR                | <i>Streptomyces venezuelae</i> ATCC 10712 | TetR             |
| JadR2               | <i>Streptomyces venezuelae</i> ATCC 10712 | ScbR2            |
| SabR                | <i>Streptomyces acidiscabies</i>          | ScbR             |
| SabS                | <i>Streptomyces acidiscabies</i>          | CprA             |
| BarA                | <i>Streptomyces virginiae</i>             | ScbR             |
| BarB                | <i>Streptomyces virginiae</i>             | ScbR2            |
| SagR                | <i>Streptomyces aureofaciens</i>          | ScbR             |
| Aur1R               | <i>Streptomyces aureofaciens</i>          | ScbR2            |
| FarA                | <i>Streptomyces lavendulae</i>            | ScbR             |
| FarR2               | <i>Streptomyces lavendulae</i>            | ScbR2            |

**Table 7.7. Comparison of 20 amino acid conserved DNA binding motifs in TetR family repressors with the motif from MmfR and MmyR**

| Name       | Sequence              | Exact motif match? |                  |        |        |        |          |
|------------|-----------------------|--------------------|------------------|--------|--------|--------|----------|
|            |                       | (MmfR)<br>GAVYFH   | (MmyR)<br>GALYGH | GAVYGH | GALYFH | GALHFH | No Match |
| MmfR       | SVKDVAERVGMTKGAVYFHF  | X                  |                  |        |        |        |          |
| MmyR       | NLATVAVRTGMTKGALYGHF  |                    | X                |        |        |        |          |
| SgnR       | TLQDVAERAEMTKGAVYFHY  | X                  |                  |        |        |        |          |
| GbnR       | NLADITARTGLTKGALYGHF  |                    | X                |        |        |        |          |
| SAV_2270   | TIKDIADGAEMTKGAVYFHF  | X                  |                  |        |        |        |          |
| SAV_2268   | NLQNIADRIRLTKGALYGHF  |                    | X                |        |        |        |          |
| SHJG_7318  | TMLDVAELSGMTKGAVYFHF  | X                  |                  |        |        |        |          |
| SHJG_7322  | NLQRVAEANLTKGALYAHF   |                    |                  |        |        |        | X        |
| SSHG_01258 | SVNDISARSGRTSGAVYFHY  | X                  |                  |        |        |        |          |
| CprA       | SLSEIVAAGVTKGALYFHF   |                    |                  |        | X      |        |          |
| CprB       | TLSEIVAAGVTKGALYFHF   |                    |                  |        | X      |        |          |
| SebR       | TITEILKVAGVTKGALYFHF  |                    |                  |        | X      |        |          |
| SebR2      | SLTMISSRAGVSNLALHFHF  |                    |                  |        |        | X      |          |
| TetR       | No motif match        |                    |                  |        |        |        | X        |
| ArpA       | No Motif match        |                    |                  |        |        |        | X        |
| AcrR       | SLADVSAASVTRGAIYWF    |                    |                  |        |        |        | X        |
| JadR       | No Motif match        |                    |                  |        |        |        | X        |
| JadR2      | KLSAISSGAGVSPGALHFHF  |                    |                  |        |        | X      |          |
| SabR       | TIAMVLERSAVTKGALYFHF  |                    |                  |        | X      |        |          |
| SabS       | SLSDIVEHAQVTKGALYFHF  |                    |                  |        | X      |        |          |
| BarA       | TVAEILSRASVTKGAMYFHF  |                    |                  |        |        |        | X        |
| BarB       | SLTAISNSAGVSNLALHFHF  |                    |                  |        |        | X      |          |
| SagR       | TISEILSEAGVTKGALYFHF  |                    |                  |        | X      |        |          |
| Aur1R      | TLSMISVGAGVSPGALHFHF  |                    |                  |        |        | X      |          |
| FarA       | TISEILTVAAGVTKGALYFHF |                    |                  |        | X      |        |          |
| FarA2      | TLSMISVGAGVSPGALHFHF  |                    |                  |        |        | X      |          |

The most closely related receptors of those analysed to MmfR were SgnR, SAV\_2270, SHJG\_7318 and SSHG\_1258 of which SSHG\_01258 appears to be the least related to MmfR, as shown by the clades on the phylogenetic tree in Figure 7.4. If SSHG\_01258 is indeed binding to the MARE operator of the L3 vector then it could be expected that possibly SgnR, SAV\_2270 and SHJG\_7318 are also able to bind, due to their even closer sequence similarity to MmfR (Table 7.3) and conserved GAVYFH DNA binding domain. It would be interesting to clone *sav\_2270*, *sgnR* or *shjg\_7318* into the pCC4 vector and see if it can indeed bind the *S. coelicolor* MARE operators using the *lux* assay.

MmyR has shown significant binding to some of the MARE operators indicating that the conserved GALYGH motif is also viable in interacting with some methylenomycin cluster operators. It is therefore possible that SAV\_2268 and GbnR are also capable of binding the L1 and L3 MARE operators.

As well as the MmfR-type GAVYFH and MmyR-type GALYGH DNA binding domain there were also commonly found to be GALYGH and GALHFH motifs conserved in the homologues. A GAVYGH motif on the other hand was not found, despite its similarities to the other motifs, which may reveal more about which residue combinations are key to DNA binding. ScbR, ScbR2, CprA and CprB contain either the GALYGH and GALHFH sequences. These four proteins come from *S. coelicolor* and do not appear to have interfered with the earlier luminescence assay indicating that they do not bind to the MARE operators. It seems likely therefore that their orthologues, with the same DNA binding motif, will not bind to the MARE operators either. However, these interactions are not fully understood yet so these possibilities cannot be ruled out.

SHJG\_7322 and BarA do show a motif with some identity to the ones from MmfR and MmyR but this is not as conserved as other motifs and so the inference is that they are also less likely to bind the MARE operators than some of the other homologues. TetR and its orthologues ArpA, AcrR and JadR on the other hand do not have a DNA binding motif resembling the ones from MmfR or MmyR and can likely be ruled out as potential interactors with the methylenomycin cluster operators.

Despite the assertions that the TFRs containing the MmfR-type GAVYFH or MmyR-type GALYGH motifs are more likely to interact with the MAREs whilst the other proteins are not, this has not been shown experimentally and therefore cannot be confirmed. The DNA binding region will be influenced by more than the block of 20 amino acids analysed here. For example, the amino acids found at the dimer interface of all these homodimers will effect the conformation of the DNA binding region. It is therefore possible that some of the proteins containing the MmfR or MmyR type motifs may not bind to the MARE operators whereas some of the other motifs may sometimes bind the MARE operators. The cut-off for amino acid sequence identity resulting in binding to the same DNA operators is not known and, without crystal structures of the proteins or an *in vivo/in vitro* analysis, would be very hard to predict.

#### 7.2.4 Implications of Findings on *S. albus*

The findings on SSHG\_01258 reveal more about the potential use of MmfR/MMF/MARE operator as an inducible system in a variety of hosts. It is likely that the *sshg\_01258* gene from *S. albus* would need to be inactivated before this strain would be viable as an expression host.

*S. venezuelae* was also a potential host for the expression system due to its ability to complete its life cycle and sporulate in liquid culture. For large-scale protein purification, liquid culture

is preferable to solid culture. The findings from working with *S. albus* indicates that for *S. venezuelae* to be used as an expression host, the *sgnR/gbnR/sgnLHP* cluster would likely need to be removed to prevent interference with the MARE operator from SgnR, GbnR or their cognate ligands. Even if these MmfR homologues were removed from potential host strains however, it is unclear as to whether even more distantly related receptors may also interact with the pathway. Therefore, due to time limitations on developing this novel inducible expression system, it was decided to continue working with *Streptomyces coelicolor* as the system has already been shown to work in this strain without detectable cross-repression.

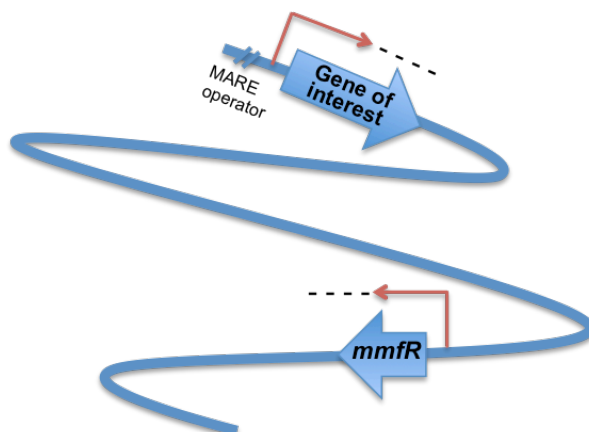
At a later date, there are hopes to further optimise *S. coelicolor* M145 to create a more streamlined host via gene deletion. (Section 8.5.2). There are hopes that a reduced *S. coelicolor* M145 strain, in addition to being more suitable for gene expression, the phenotypic differences sometimes observed in the luminescence assay may also be reduced as the cells can focus their resources on the expression of a more select number of genes.

## 7.3 Development of Vectors for Inducible Expression

### 7.3.1 Strategy of Inducible Expression System Mechanism

Based on all of the findings of the bioluminescence assay it was decided that the MmfR/MMF/MARE operator system has potential for development as the control mechanism in a novel inducible expression system. Synthetic expression vectors were designed to work in a similar way to the luciferase reporter gene assay, relying on two different vectors (Figure 3.3). One of these vectors was designed to include one of the MARE operators in front of a gene of interest (instead of *luxCDABE*) which was then integrated into the *S. coelicolor* M145 genome along with pKMS01; *mmfR* on the pCC4 vector (a vector derived from pOSV556) (120) under the control of *ermEp\**. A schematic of how this will look once in the *Streptomyces* genome can be found in Figure 7.5. MmfR is expected to be produced constitutively, binding to the MARE operator and repressing the expression of the downstream gene. The addition of the MMFs to the system should then switch on the production of the protein of interest.





**Figure 7.5. The genetic basis of the novel inducible expression system designed to be regulated by MmfR/MMF/MARE operator in the *S. coelicolor* M145 chromosome**

The intergenic region between *mmfR* and *mmfL* (containing the L1 MARE operator and in the direction of *mmfRp*) was selected for the expression system as it showed the strongest MmfR binding as well as the easiest release of repression upon the addition of the MMFs. In terms of selecting which MMFs to use, both MMF1 and MMF3 (those with branches alkyl chains) were shown to have the highest binding potentials to MmfR and so are the logical choice as the inducer in the inducible expression system.

An additional feature that was desirable for this inducible expression system was the introduction of a secretion signal for the protein of interest. This signal would allow recovery of the protein from the culture supernatant without having to lyse the *Streptomyces* cells. Not only does this reduce the number of proteins present in the initial sample to be purified but also streptomycetes are much harder to lyse than *E. coli*, thus avoiding this potentially tricky step in purification is preferential. There are two main bacterial export systems, the Tat (twin arginine translocation) system (145) and the Sec system.(146) The Sec system is universally conserved across bacteria and archaea and is usually the main route of protein export for cells. This pathway exports proteins in their unstructured state and is used to move proteins involved in a large number of different processes. The Tat system on the other hand is unusual as it exports proteins in their fully folded state (147) and is therefore of particular interest in developing this expression system. This system has fewer substrates and is found in fewer types bacteria than the Sec system (146) but the *Streptomyces* genus are the largest known users of this Tat pathway, with *S. coelicolor* having more than 100 Tat substrates.(128, 148) (Although this means that there are a number of naturally exported proteins that will be collected alongside the protein of interest, the number obtained will still be many less than would be found from lysing the same cells and collecting intracellular proteins.)

Both the Sec and more recently the Tat pathway have shown successes for use in expression systems.(149) In particular the Tat pathway was successfully used in *Streptomyces lividans* for the collection of proteins with a Tat signal peptide added.(150, 151) The aim therefore is to trial this export pathway alongside regulation by MmfR/MMF/MARE operator in *S. coelicolor*.

### 7.3.2 Selecting a Protein for Over Production

To test out the novel inducible expression system there was a range in choices of genes that could be expressed. Our group has been interested in characterising urea synthetases, with particular interest in orthologues of *S. venezuelae* GbnB from *Salmonella enterica* and *Streptococcus mutans*. For this reason, these two proteins were chosen as the focus of the expression system being designed.

*S. venezuelae* contains a gene cluster annotated as *sgnLHP/sgnR/gbnR*, which is homologous to *mmfLHP/mmfR/mmyR* from *S. coelicolor*. In *S. venezuelae* these genes are thought to regulate the expression of *gbnABC*, the biosynthetic and export genes for the recently discovered gamma-aminobutyrate urea natural products; the gaburedins. GbnA is a glutamate decarboxylase which is thought to produce GABA whereas GbnB works as an ATP-dependent enzyme belonging to the acyl-CoA synthetase family and GbnC as a gaburedin exporter protein.(75)

GbnB shares between 27% and 31% identity with its analogues from *S. mutans* and *S. enterica* respectively, across 96% of their sequences. Like GbnB, both analogues are predicted to be acyl-CoA synthetases (AMP-forming)/AMP-acid ligases. Based on their similarities, it can be predicted that possibly both analogues may also be involved in natural product biosynthesis, hence the desire to purify and characterise them. Details on the properties of the amino acid sequences for all three of these AMP-binding proteins can be found in Table 7.8.

**Table 7.8. Details on *S. venezuelae* GbnB and its analogues**

*Details on the predicted pI are from ProtParam(77). Strains: Streptomyces venezuelae ATCC 10712, Salmonella enterica subsp. enterica serovar Schwarzengrund str. CVM19633 and Streptococcus mutans GS-5*

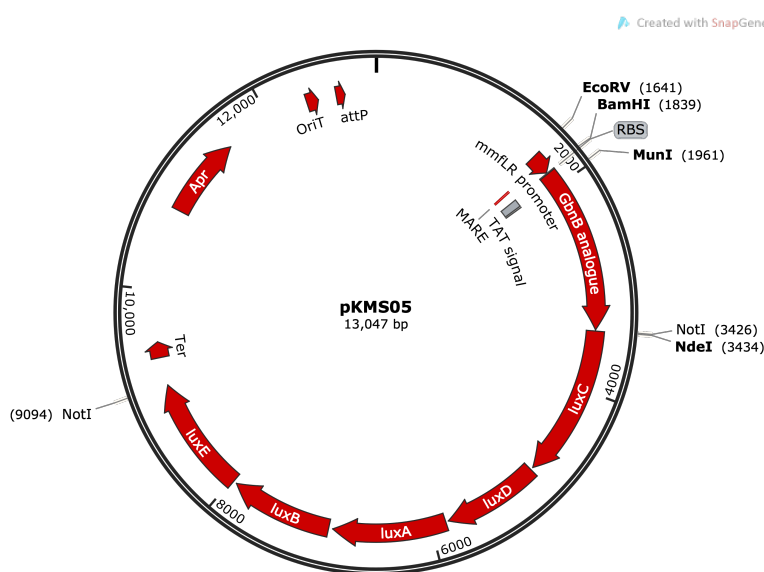
| Strain               | Amino acid length | Molecular weight (daltons) | Theoretical pI |
|----------------------|-------------------|----------------------------|----------------|
| <i>S. venezuelae</i> | 532               | 57642                      | 5.17           |
| <i>S. mutans</i>     | 487               | 55627                      | 8.8            |
| <i>S. enterica</i>   | 498               | 56347                      | 6.15           |

As yet, neither of these GbnB-like proteins from *S. enterica* and *S. mutans* have been successfully purified from existing commercially available expression systems in *E. coli*. If

they can successfully be over produced and purified from *S. coelicolor* M145 therefore it would certainly be helpful in showing the usefulness of such an expression system. Expression vectors were therefore designed to include one or the other of these genes.

### 7.3.3 Design of the Expression Vectors

In addition to the already engineered pKMS01 (containing *mmfR* under the control of *ermEp\**), two more vectors were created for use in the expression system. These are called pKMS05 and pKMS06 for *S. mutans* and *S. enterica gbnB* respectively (see Figure 7.6, Table 2.5 and Table 2.6) and contain the analogues under the control of the *mmfL/mmFR* intergenic region. To make these vectors, the already existing L1 vector was used as a backbone.



**Figure 7.6. pKMS05, an expression vector containing the *S. mutans gbnB* orthologue and *luxCDABE*, both under the control of *mmfLp***

*This vector contains the *mmfL/mmFR* intergenic region meaning that the expression of the gene insert is under the control of *mmfLp* and the L1 MARE operator.*

The *gbnB*-like genes were inserted into the L1 vector downstream of the methylenomycin cluster intergenic region. Depending on the restriction enzymes used, the *lux* genes could either be cut out or left downstream of the insert. The *lux* genes were left in the vectors while the system was being optimised as an extra control that could be easily observed. Luminescence produced should reflect when the gene of interest was also being expressed. Once the system has been shown to be working, the extra *luxCDABE* genes can then be cut out using the *NotI* restriction sites and the vector re-ligated (Figure 7.6).

The *gbnB* analogue insert was created synthetically via the GeneArt service by Thermo Fisher and included a number of extra features, which will now be explained. A diagram of the layout of the insert can be found in Figure 7.7.

*Bam*HI | RBS | ATG | TAT signal | 6 Histidine | *Mun*I | *gbnB* gene | *Not*I | *Nde*I |

**Figure 7.7. Layout of synthetic *gbnB* insert for vectors pKMS05 and pKMS06, for use in the novel inducible expression system**

*Bam*HI, *Mun*I, *Not*I and *Nde*I all refer to the restriction sites added in these positions. RBS indicates the ribosome binding site.

Firstly the insert was designed to include a codon optimised *gbnB* gene, specific to *Streptomyces* codon usage, producing genes with 65% and 58% GC content for *S. enterica* and *S. mutans gbnB* respectively, compared to the average genomic GC content of 72% for *S. coelicolor*. Between the start codon and the rest of the gene sequence, a polyhistidine tag was added to allow easy nickel Sepharose purification of the protein. A Tat signal was also added next to the Histidine tag. The Tat signal was a twin arginine repeat translocation pathway signal (Section 7.3.1), with the specific sequence taken from the work carried out by Palmer *et al.* (146) giving the following amino acid sequence;

TKPVVPSGVSRRGFLGGSLGVAGAVLLAA

The specific conserved Tat pattern within this sequence has been underlined. This sequence was also codon optimised for *S. coelicolor* expression.

Also added to the synthetic insert was a synthetic *Streptomyces* ribosome binding site (5' AAGGAGG 3') as well as a number of different restriction sites. These restriction sites were designed so that different components of the insert could be cut out if needed. For example, allowing the choice of leaving the *luxCDABE* genes in the L1 host vector. The *Mun*I and *Not*I restriction sites can be used to easily swap in and out alternative genes for over expression into this vector. The full sequences of the synthetic inserts used are in Appendix A and B.

Details on the properties of the GbnB-like proteins that are predicted to be purified from strains expressing pKMS05 and 06 can be found in Table 7.9. The increased molecular weights compared to the earlier Table 7.8 are down to the polyhistidine tag and export signal being added to the peptide sequence.

**Table 7.9. Details on synthetic GbnB orthologues predicted to be produced by the new constructs; pKMS05 and pKMS06**

*Details on the stability and the predicted pI are from ProtParam (77).*

| GbnB analogue      | Amino acid length | Molecular weight (daltons) | Theoretical pI | Predicted stability |
|--------------------|-------------------|----------------------------|----------------|---------------------|
| <i>S. mutans</i>   | 524               | 59410                      | 9.03           | Stable              |
| <i>S. enterica</i> | 535               | 60156                      | 6.54           | Unstable            |

After the synthetic gene had been synthesised, cloned into the L1 vector and the construct checked via sequencing and restriction digests, the new vectors were transferred into *S. coelicolor* M145 by intergeneric conjugation. This new inducible expression system was then tested and optimised for purification of the proteins of interest.

## 7.4 Optimisation of a Novel Inducible Expression System

### 7.4.1 Strategy of Protocol Optimisation

Research into protein over production in streptomycete systems is much more limited than that for equivalent *E. coli* expression systems. For this reason, optimisation of the method had to begin with the very basics. First the streptomycete system was trialled to see at which time point any exported proteins might be present. It was also necessary to try different types of media, for both liquid and solid cultures as well as investigating different carbon sources present in the media. In addition to this, different buffers were trialled in the nickel Sepharose purification of the protein as well as the addition of protease inhibitors to cultures. An outline of the methods used can be found in Section 2.3.11.

Many of the trials were run initially with only the vector created for the *S. mutans* GbnB orthologue (pKMS05). This protein was predicted by ProtParam to have greater stability than the *S. enterica* equivalent (see Table 7.9). Although this is only a fairly arbitrary value, there was a desire to reduce the workload of trialling both vector types and so only one was selected for use primarily and there was no other clear reasoning for choosing one over the other. Initial trials were also run with no MmfR present (pKMS01 was not added). The inducible aspect of the system was only added once optimisation of protein over production has yielded some results. The strains containing pKMS05 or 06 are referred to as M145:L1 and M145:pKMS05 throughout the rest of the chapter.

Protein samples collected were processed and then checked using SDS-PAGE for the presence of a band in the expected position. As can be seen in Table 7.9, a protein of around 60 kDa is expected for both orthologues. A number of the extracellular protein samples collected were also further purified using their poly-histidine tag, based on nickel Sepharose affinity (see Section 2.3.11). As many samples were being processed it was not practical for high levels of purification to be obtained for each sample. To save both time and resources, a packing column was not used, instead the nickel Sepharose protocol was carried out in microcentrifuge tubes with the supernatants being collected after each of the elution steps. This of course will result in the presence of a larger number of non-specific proteins being collected in the elution fractions as well as the protein of interest. This should be taken into

consideration when looking at all gels of purified proteins in these preliminary optimisation trials. For a higher level of purity to achieve a single band on an SDS-PAGE gel and potentially obtain useable proteins, a more thorough purification protocol could later be used, for example FPLC (fast protein liquid chromatography).

Before any protein recovery trials were carried out, the strains were first checked for luminescence to determine whether the expression system vectors were integrated into the *Streptomyces* genome and were indeed being expressed. It was found that the cells containing pKMS05 and 06 were luminescing at all time points checked.

#### 7.4.2 Selecting Time points for Protein Purification

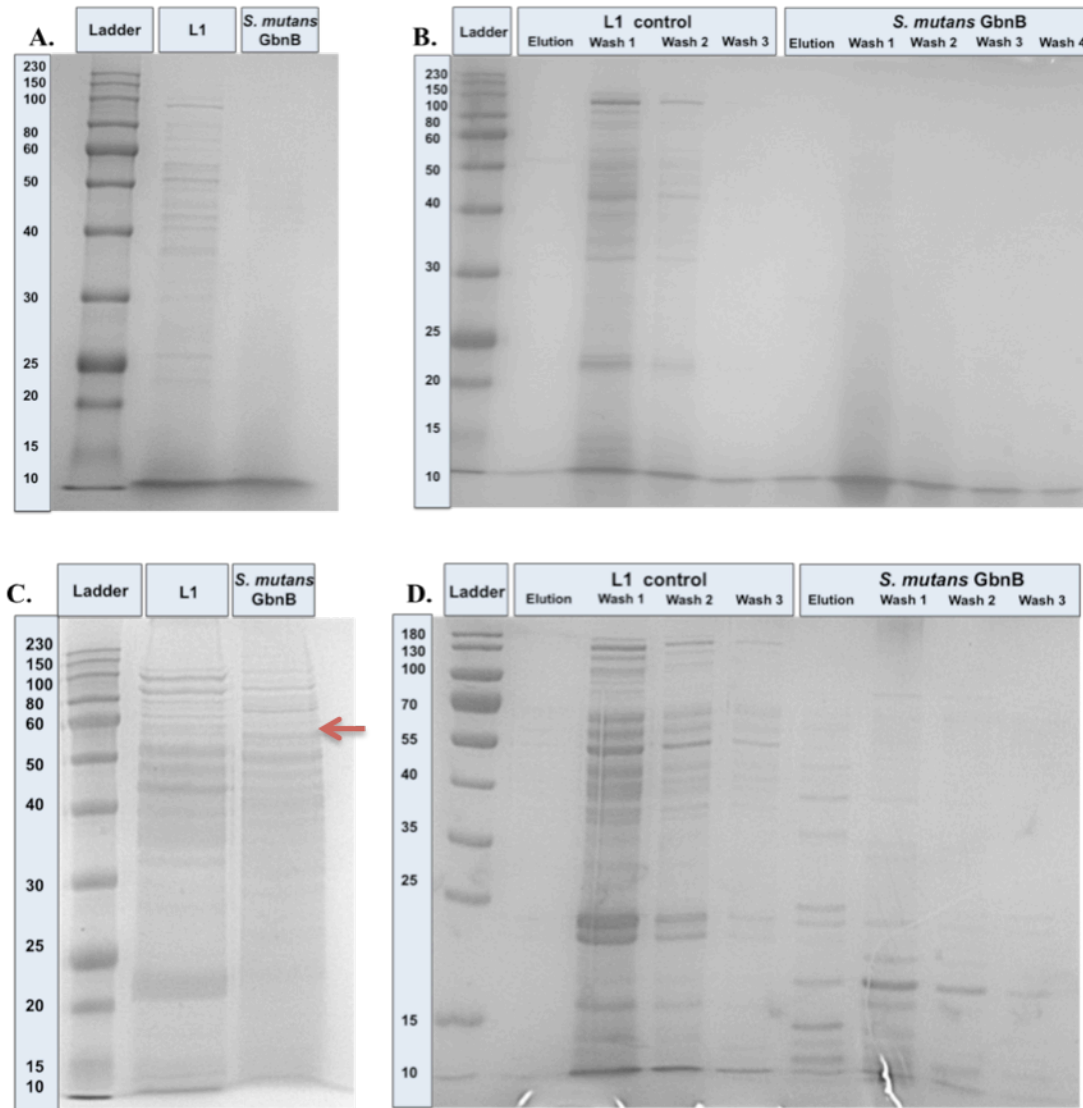
Whereas *E. coli* will often produce a protein of interest in the first 24 hours of growth, *Streptomyces* have a much slower doubling time of 2.2 hours (51, 152) compared to 20 minutes for *E. coli*.(153) The heterologous expression of some antibiotics in *Streptomyces* may take up to two weeks to achieve a good yield (152) and so it is likely that measurable heterologous expression in streptomycetes will occur after the 24 hour time point. Based on a literary review, it was expected that protein overproduction in the streptomycetes may peak anywhere within the first 72 hours of growth, although longer trials were possible for this assay if necessary.(101)

*S. coelicolor* M145 with pKMS05 (*S. mutans gbnB*) were grown in 2xYT media in baffled flasks. The supernatant was then collected at 24, 48 and 72 hours growth before being concentrated in a centrifugal column and run out using SDS-PAGE. Figure 7.8A-D show the results of such collections. Expression levels were compared to a control sample of L1 with no *gnbB* insert, which represents the levels of exported proteins normally produced by *S. coelicolor* M145.

At the 24-hour time point, very few exported proteins were seen so no extra purification methods were utilised. At both 48 and 72-hour collections however, the proteins present in the supernatant were analysed as well as proteins ‘purified’ on nickel Sepharose<sup>1</sup>. Both the elution and washing stages of the protein purification have been shown on the SDS-PAGE gels in Figure 7.8B and D.

---

<sup>1</sup> During the nickel Sepharose purification, the buffers described in Section 2.2.4 as ‘protein purification buffer’ and ‘elution buffer’ with 200 mM imidazole were used.



**Figure 7.8A-D. Secreted proteins produced at different time points by the M145:L1 control strain compared to the strain designed to produce *S. mutans* GbnB (M145:pKMS05)**

*All samples are collected from concentrated 2xYT growth media supernatant*

- A. 24 hours growth
- B. 48 hours growth, purified with nickel Sepharose
- C. 72 hours growth
- D. 72 hours growth, purified with nickel Sepharose

As can be seen in Figure 7.8A, there were very little exported proteins seen at 24 hours. At 48 hours there are more detectable proteins in the supernatant of the M145:L1 and M145:pKMS05 cultures but very few bands were observed after nickel Sepharose purification of either samples. The 72-hour time point shows most promise in terms of total protein yield but at no time point is there a really distinct band in the position of the expected molecular weight, even after nickel Sepharose purification. For example, Figure 7.8C and D show bands at the size of the expected product but they are no more distinct than any of the others that can

be seen in these samples. It is promising however that there are some bands produced by the *gbnB* strain which are not obviously present in the samples produced by the L1 control, these extra bands are all smaller than 60 kDa and therefore may represent the GbnB-like proteins degraded into smaller polypeptides. The indication from this data is that an optimisation of growth conditions may be needed to obtain the protein of interest.

Based on observations from Figure 7.8, all optimisation trials from here on were run for at least 72 hours before exported proteins were collected and checked via SDS-PAGE.

### 7.4.3 Considerations on Further Optimisation of the Protocol

Cultures from the investigation represented in Figure 7.8A-D were being grown in 2xYT, a variant on LB media with increased levels of tryptone and yeast extract. A limitation of this media is that it lacks any real carbon source (154), which may be detrimental to protein overproduction. For this reason, the following medias and supplements were also trialled; 2xYT, LB, TSB (0.25% glucose), 2xYT with 3% mannitol, 2xYT with 3% glucose, 2xYT with 5% glucose, 2xYT with 3% glycerol, 2xYT with 3% fructose, 2xYT with 3% sucrose.

The data for this investigation is not shown here but there was no clear impact on the size of the band at ~60 kDa in the presence of any of the different media or supplements. One feature that was particularly distinct in these results however, was the presence of very large bands at the bottom of the SDS-PAGE gel, possibility indicating streptomyces cell lysis in liquid media. This cell lysis would cause the internal as well as exported proteins became mixed and the release of many proteases, therefore explaining the lack of larger molecular weight bands seen elsewhere on the gels. Reducing this cell lysis is therefore of interest when optimising this protocol.

Widdick *et al.* (128) have reported that in liquid media, energy demanding systems or systems that put stress on a cell can result in the lysis of streptomyces. It is possible that having the *luxCDABE* system in addition to the expression system both working in *Streptomyces* could be causing the cell lysis observed by Widdick *et al.* (128). The maintenance of the *lux* genes in the L1 vector when creating pKMS05 and 06 was to act as a second marker for gene expression, easier and quicker to measure than protein export. The implications this was to have on the health of the cell were not known and in the future, when time allows, this extra marker should maybe be removed or replaced with something less energy demanding.

In the paper by Widdick *et al.*, an alternative method of purifying exported proteins from *S. coelicolor* was described, this time using solid media where the streptomyces appeared to grow better even if they had been shown to lyse in liquid media.(128) This approach was therefore adopted.



#### 7.4.4 Work on Solid Culture and Addition of MmfR and the MMFs

Using solid culture has the added benefit of not only potentially reducing cell lysis but also allowing the *S. coelicolor* M145 strains to enter the sporulation stage of their life cycle, not normally possible in liquid media. It also means that cultures can be grown on smaller scales, reducing the demand for high volumes of MMF inducers.

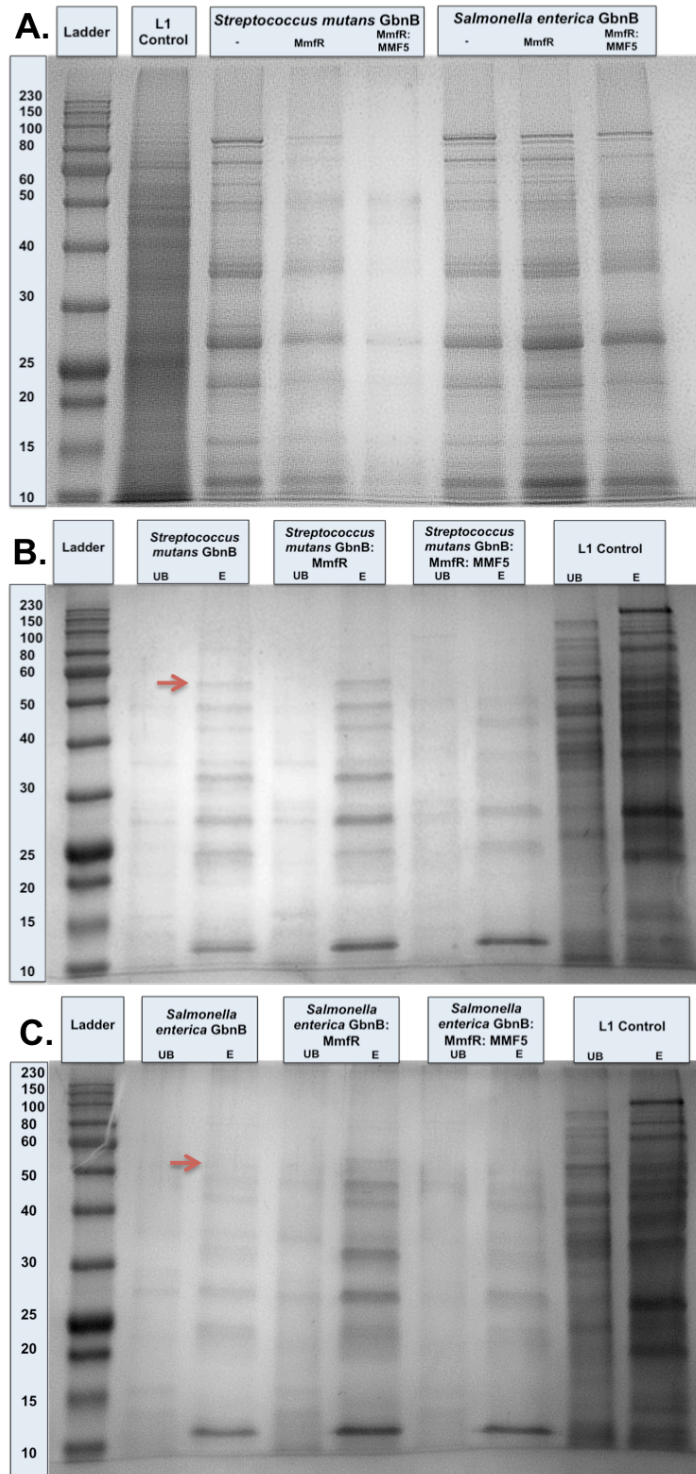
When using solid media, exported proteins were collected by washing the cell mass that had been grown on dialysis tubing. This sterile tubing had been placed on top of the solid soya flour mannitol growth media and allowed the passage of nutrients from the medium to the bacteria, while keeping proteins produced on top of the tubing for easy harvesting. These harvested proteins often did not need to be concentrated further and could either be used straight away in nickel Sepharose purification or be precipitated out by lithium chloride/trichloroacetic acid precipitation (see Section 2.3.11). Both these sample types could then be analysed by SDS-PAGE. The lithium chloride/trichloroacetic acid precipitation protocol follows the one specified by the Widdick paper exactly except that they used cellophane instead of dialysis tubing (and did not utilise the alternative nickel Sepharose purification).

For the trials with solid media, both the *S. mutans* and the *S. enterica* GbnB-like proteins were tested and pKMS01 (containing *mmfR*) was added to the system. MMF5<sup>1</sup> was then added to check the inducible aspect to the expression arrangement and for any changes in protein over production.

Figure 7.9A shows the lithium chloride and trichloroacetic acid precipitation of secreted proteins from M145:L1, M145:pKMS05 and M145:pKMS06 in the presence and absence of MmfR (from pKMS01) and MMF5 on an SDS-PAGE gel. This gel shows all of the precipitated exported proteins collected, with no selection for the histidine-tagged analogues. Figure 7.9B shows the nickel Sepharose purification of secreted proteins from M145:L1 and M145:pKMS05 (producing *S. mutans* GbnB) in the presence and absence of MmfR and MMF5 and Figure 7.9C shows the same for pKMS06 (*S. enterica* GbnB). During the nickel Sepharose purification, the buffers described in Section 2.2.4 as ‘protein purification buffer’ and ‘elution buffer’ with 200 mM imidazole were again used.

---

<sup>1</sup> MMF5 was used rather than one of the MMFs with a higher binding potential due to availability of the furans that had been synthesized at the time of this trial. The more efficient MMF1 and 3 were then used again when they became available.



**Figure 7.9A-C.** Secreted proteins produced by the M145:L1 control strain compared to the strains designed to produce *S. mutans* or *S. enterica* GbnB (M145:pKMS05 and 06 respectively) in the presence and absence of MmfR and MMF5

All samples collected after 72 hours growth on SFM media.

UB = unbound proteins E = eluted proteins

- A. *S. mutans* and *S. enterica* GbnB - LiCl/TCA precipitation of secreted proteins
- B. *S. mutans* GbnB - nickel Sepharose purification of secreted proteins
- C. *S. enterica* GbnB - nickel Sepharose purification of secreted proteins

### General Observations

For the gel shown in Figure 7.9A it was possible to see a faint band in the position of ~60 kDa in all the pKMS05 and 06 strains (although this is not always apparent on the scans of these gels). This band does not appear to be present in exactly the same position for the L1 control strain but the band is by no means distinct in any of the pKMS06 or 05 samples.

The samples in Figure 7.9A contain precipitations of all of the many exported proteins from *S. coelicolor*, without any nickel Sepharose selection for the GbnB analogues, so are expected to show multiple bands. However, in a very efficient expression system it would be hoped that a bolder band may be seen here at 60 kDa. This therefore does indicate that this system is not yet working as desired and more adjustments are necessary. As was mentioned previously, it is known that *luxCDABE* are being expressed in these strains and so it is expected that the *gbnB*-like genes are also being expressed.

In Figure 7.9B and C, after nickel Sepharose purification the band at ~60 kDa which was seen in Figure 7.9A remains (see red arrows) but it is still not any more distinct than the eight or nine others also seen. When compared to the L1 control it can be seen that some of these bands are also present in the control but a few do seem to be in a slightly different position. When considering these gels, it is interesting that all of the bands in the nickel Sepharose purified samples appear to be 60 kDa or smaller and, although there isn't such a distinct band at the bottom if the gel like those seen when trialling different carbon sources, there is still an indication of protein degradation. If the GbnB-like proteins are being proteolytically cleaved there will still be the histidine-tag attached to their N-terminals, resulting in a purified protein with a smaller molecular weight. On solid media there does not appear to be the cell lysis seen in liquid cultures and so the majority of internal *Streptomyces* proteases are likely being kept separate from the protein of interest. There are however still the proteases that are being released by *S. coelicolor* to assist with aspects of its normal sporulation cycle. Limiting cell lysis will not directly reduce the activity of these exported proteases and so the use of protease inhibitors was implemented, with the aim that this may result in the production of a more distinct 60 kDa band.

To further understand the nature of the proteins being purified on nickel Sepharose, samples from Figure 7.9B were sent for LC-MS analysis. Three of the bands at ~60 kDa, 27 kDa and 12 kDa were cut out of the gel, digested and processed according to the protocol specified in Section 2.3.13. Samples were processed by the Proteomics Facility at the University of Warwick. Unfortunately however, the results from this analysis were not conclusive and largely showed proteins associated with those native to the *Streptomyces* genus.

It is also of note that the bands on the gel for the *S. enterica* GbnB (Figure 7.9C) were less defined than those for *S. mutans* GbnB, especially around 60 kDa. As was seen in the earlier Table 7.9, the *S. enterica* GbnB is predicted to be an ‘unstable’ protein whereas the *S. mutans* GbnB classified as being ‘stable’. This is a possible explanation therefore of why *S. mutans* GbnB will give more promising bands on a gel therefore. However, the classification of ‘stable’ versus ‘stable’ is based on calculations using the ProtParam standard set of parameters and therefore is unlikely to directly representative of expression in *S. coelicolor*. There are many other factors that may have also contributed to the differences between the gels for these analogues including the level of completion of successful protein folding achieved and the possibility that the intracellular pH of *Streptomyces* is more optimal for one protein than the other.

### Inducibility of the System

Figure 7.9 shows the first results from adding the MMF-inducible MmfR to the expression system. A number of the bands in Figure 7.9B and C are very faint so it is not yet possible to comment on the effect of adding *mmfR* to the expression system (via pKMS01). In particular, better samples need to be collected from these strains in the presence of the MMFs as these gels are particularly indistinct and the total amount of proteins harvested (including wild type proteins) was low. The production of gels with proteins from a greater number of culture plates may result in the bands on these gels being more distinct due to an increased protein yield.

MmfR binding to the L1 MARE operator has been shown to be leaky (Section 4.3) and so some level of the production of the GbnB-like proteins in the repressed *mmfR* strains was also expected to be seen. This does appear to potentially be the case for the gels in Figure 7.9, with a number of similar bands being seen when the pKMS01 was added compared when it is not. However, the levels of all protein bands produced from these gels appears to be very similar indicating that there is little difference in expression the presence of MmfR. The level of similarity between the sample types suggests that when under the control of *ermEp\**, MmfR does not result in enough repression for the system to be properly controlled in an inducible manner. Whether the current MmfR/MMF/MARE operator system is indeed too leaky for use in this system is not clear however or whether the similar bands could be a result of all proteins seen being native proteins and therefore unrelated to the expression vectors and the desired recombinant proteins.

### 7.4.5 Optimisation with Protease Inhibitors

A number of the proteins potentially secreted by *Streptomyces* are proteases, so adding protease inhibitors throughout the growth of cultures could increase the yields of the proteins of interest.

When using standard expression systems, such as those using the *E. coli* host BL21\*, protease inhibitors are generally only needed to be present in buffers once cells have been lysed. These optimised expression hosts often have their genes for secreted proteases knocked out and therefore only when intracellular proteases are released is protease inhibition needed. These protease inhibitors therefore usually have a short half-life as purification procedures will often follow immediately. This standard inhibitor half-life of less than 24 hours at temperatures above 4 °C brought some challenges when selecting the right inhibitor cocktail for use in this assay. Secreted proteases were likely being released throughout growth of the cells, along with the desired GbnB-like protein and so would need constant inhibition over 72 hours. A protease inhibitor cocktail specifically designed for tissue cultures and which remains active for at least 48 hours at 37 °C was identified and tested, adding it to plates after 24 hours growth. This protease inhibitor cocktail was also checked to make sure that at the concentration used, none of the components were known to interact with nickel Sepharose. No predicted interactions were found.

Protein samples were prepared for both nickel Sepharose purification and lithium chloride/trichloroacetic acid precipitation from M145:L1, M145:pKMS05 and M145:pKMS06 in the presence and absence of MmfR and MMF1 from solid cultures as before. This time however, the solid culture plates contained 1 mM EDTA (to inhibit metalloproteases) and at 24 hours growth were overlaid with the protease inhibitor cocktail for tissue culture to a final dilution of 1:500.<sup>1</sup> These plates were then allowed to grow for a further 48 hours before the secreted proteins were harvested.

The buffer for nickel Sepharose purification had also been altered for this stage in the trials. Not only was the protein purification buffer now used with an extra protease inhibitor cocktail but instead of the Tris-HCl buffer, a sodium phosphate buffer with higher sodium chloride levels was used.<sup>2</sup> This change was made due to concerns over the influence that temperature has on the pH of Tris-HCl buffers and whether at 4 °C, the pH of the buffer was too close to the theoretical pI of the proteins trying to be purified. Having a pH close to that of the pI of a protein has been indicated to have a negative effect on protein purification by other

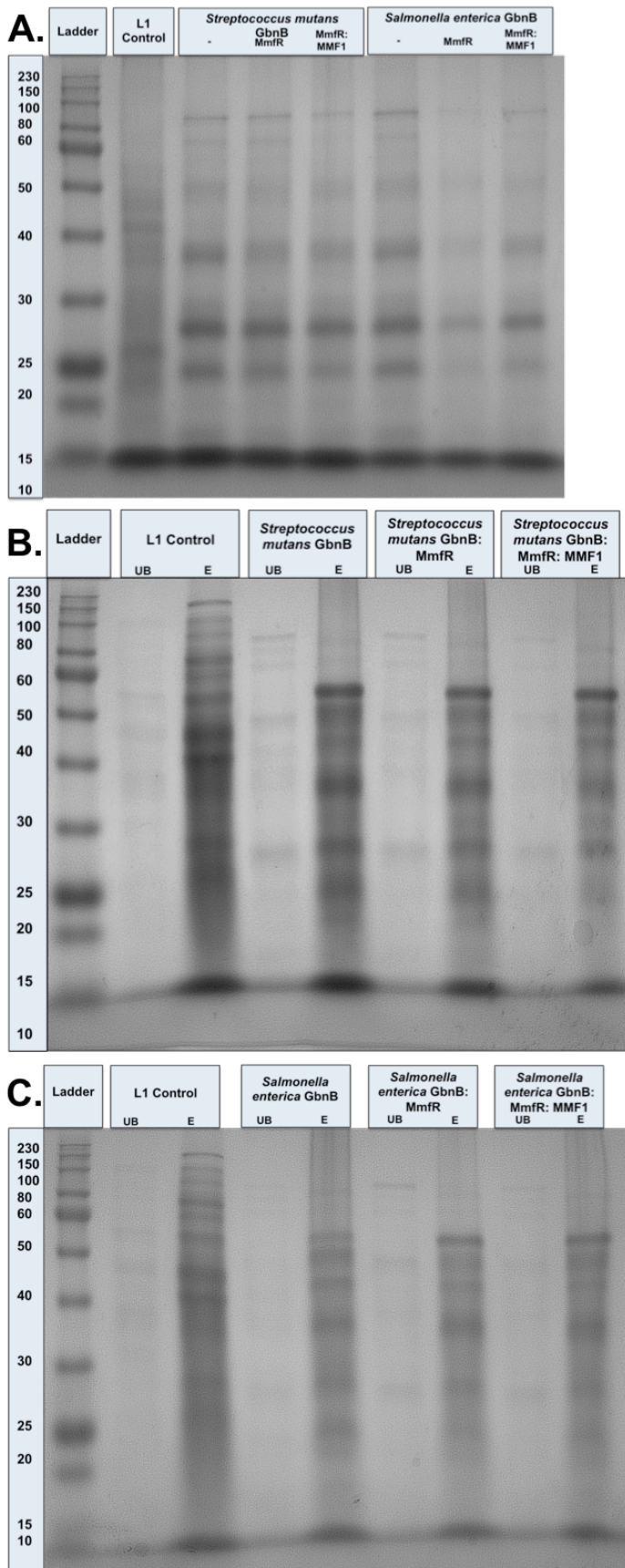
---

<sup>1</sup> See Section 2.2.4 for more details

<sup>2</sup> See 'Improved protein purification buffer for Ni Sepharose purification' in Section 2.2.5.

researchers. A corresponding sodium phosphate elution buffer with increased imidazole concentration to 500 mM compared to the 200 mM used before was also utilised. The chosen optimised sodium phosphate buffers were produced following guidelines that came with the nickel Sepharose kit.

Results from these investigations into the influence of protease inhibition in protein purification can be found in Figure 7.10A-C, with results from the lithium chloride/trichloroacetic acid precipitation found in Figure 7.10A and the nickel Sepharose purification of *S. mutans* and *S. enterica* GbnB in Figure 7.10B and C respectively.



**Figure 7.10A-C. The effects of protease inhibitors on secreted proteins produced by the M145:L1 control strain compared to the strains designed to produce *S. mutans* or *S. enterica* GbnB (M145:pKMS05 and 06 respectively) in the presence and absence of MmFR and MMF1**

All samples collected after 72 hours growth on SFM media containing 1 mM EDTA. Plates were overlaid with Sigma Aldrich protease inhibitor cocktail for tissue culture at 24 hours growth. SIGMAFAST protease inhibitor cocktail was also added to the protein re-suspension buffer.

UB = unbound proteins  
E = eluted proteins

**A.** *S. mutans* and *S. enterica* GbnB - LiCl/TCA precipitation of secreted proteins

**B.** *S. mutans* GbnB - nickel Sepharose purification of secreted proteins

**C.** *S. enterica* GbnB - nickel Sepharose purification of secreted proteins

As can be seen in Figure 7.10B in particular (nickel Sepharose purification of *S. mutans* GbnB), there was a marked increase in the boldness of the band at around 60 kDa for all of the pKMS05 strains. There are still some other bands present, likely to be due to the level of purity achieved from the way the nickel Sepharose samples were processed. The 60 kDa band is however now the largest of those seen. The same can also be seen for the strains containing pKMS06 (*S. enterica gbnB*) although this protein is still giving less distinct results. The increase in intensity of the band size ~60 kDa might reflect a halting or reduction in protease degradation of the protein of interest, increasing its yield. The L1 control also appears to not have some of the bands that are present in the samples from the strains containing the expression vector on each gel, increasing that prospect that the GbnB-like proteins are produced.

Despite the successes increasing the prominence of the band consistent with the GbnB-like protein molecular weight however, there are still concerns as to whether enough MmfR is present to efficiently repress and switch off the expression of the gene of interest in the absence of the MMFs. In Figure 7.10A, B and C the profiles shown on the gels for both expression vectors looked very similar in the presence and absence of MmfR and the MMFs. This therefore casts some doubt on the band being seen at 60 kDa being for the GbnB-like protein. For this reason, a number of the bands were cut out of the gels shown in Figure 7.10B and C and again analysed by LC-MS. The bands at the expected size of around 60 kDa were cut out for both the *S. mutans* and *S. enterica* samples, as well as a number of prominent smaller bands from the *S. mutans* sample (from ~50, 43, 37 and 26 kDa) to check for protein degradation. Most LC-MS samples collected were from the *S. mutans* GbnB producing strains as these bands were more distinct. Samples were also always taken from the positive controls containing the expression vectors (pKMS05 or 06) with no MmfR repression due to it not being clear how well MmfR was working in the system as a repressor. The contribution of the WPH Proteomics RTP, Gibbet Hill Road, University of Warwick, UK should be acknowledged for the processing for these samples and help with the analysis of the results.

As might be expected, the main hits from the LC-MS analysis data returned were again for *Streptomyces* proteins. For none of the samples were there detectable *Photobacterium luminescens* associated proteins indicating that the luciferase and the enzymes for its substrate were remaining inside the streptomycete cells.

Both *S. venezuelae* GbnB and its two analogues belong to the acyl-CoA synthetase AMP-binding family of enzymes. Any AMP binding enzymes found in the bands analysed for LC-MS would therefore indicate the potential presence of the protein of interest. The six bands analysed showed sequence similarity to very few non-actinomycete proteins and the results

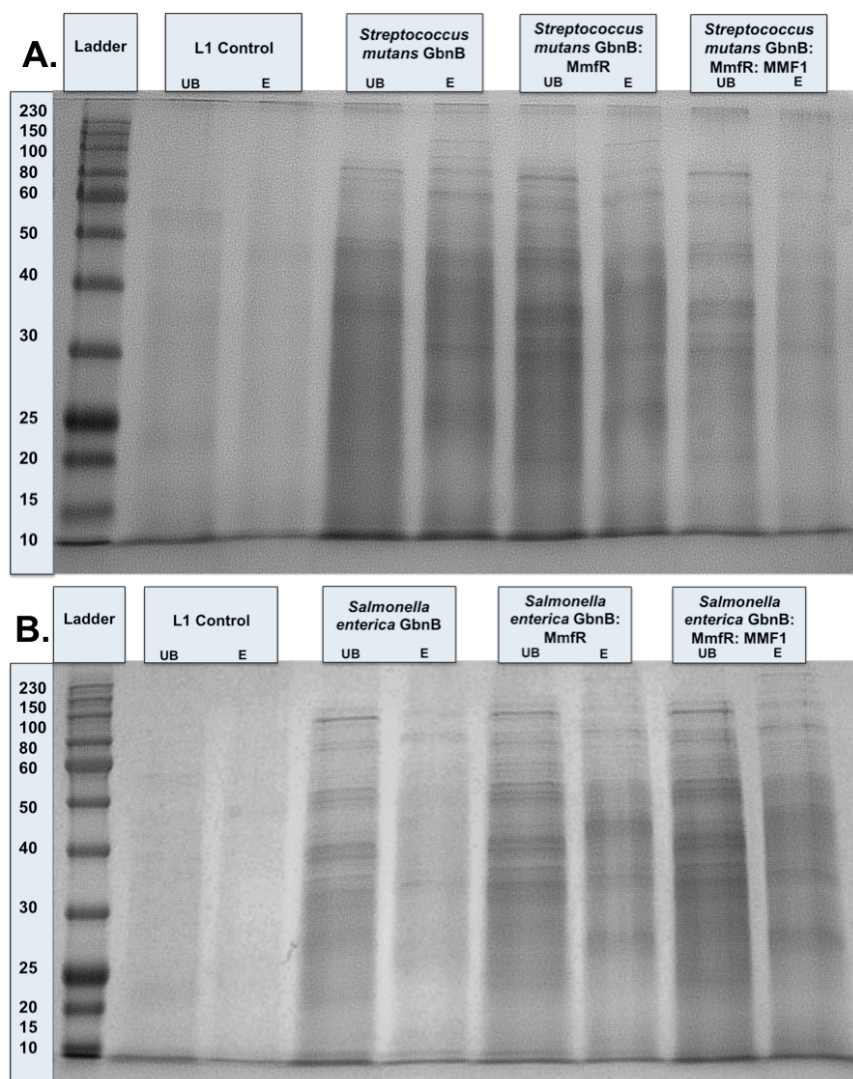


were largely inconclusive in terms of detecting the presence of the two GbnB-like proteins. Some of the bands showed similarity with a sugar ABC transporter substrate-binding protein from *S. enterica* (A0A0W5EVT2\_SALCE), unfortunately this protein does not bear much resemblance to the family of proteins that the GbnB analogues come from. There are also at least six possible amino acids detected in various bands that are Acyl-CoA binding enzymes from the *Streptomyces* species. These could possibly in fact be the Acyl-CoA synthetases being looked for due to the high identity found between the analogues. Unfortunately however, no conclusions can be made from this data on the successes of the successful export and purification of the GbnB analogues. With further purification of protein samples, the bands collected may hopefully yield more relevant results with less interference from exported proteins native to *S. coelicolor* M145.

### Checking Inside Cells for GbnB Orthologues

As expression vectors were designed to attach a Tat signal to the protein of interest and label it for export out of the cell, all trials up until now have been run looking for proteins in the supernatant of liquid cultures or by washing the cell mass of solid cultures. To make sure that the protein of interest was properly being exported, the cell mass that had been collected for the experiments in Figure 7.10 was lysed and the cell contents checked for the GbnB orthologues. After cell lysis, the samples were processed in the same way as the exported proteins had been previously, using the optimised buffer with protease inhibitor cocktail present and purifying via nickel Sepharose purification. These purified proteins were then analysed using SDS-PAGE, images of these gels are shown in Figure 7.11A and B.

As explained before, it is known that some expression of the pKMS05 or 06 inserts is occurring, as luminescence is seen when checked for the Photek camera. Assuming the upstream *gbnB* analogues are also being expressed it is therefore a case of finding out whether a corresponding protein has been produced and if so, where this may be; whether it is a case of successful folding and export but in low amounts, the folded proteins remaining inside the cell with failed export or alternatively, incorrect folding and degradation of the protein analogues.



**Figure 7.11A and B. Intercellular proteins produced by the M145:L1 control strain compared to strains designed to produce *S. mutans* or *S. enterica* GbnB (M145:pKMS05 and 06 respectively) in the presence and absence of MmfR and MMF1**

*A* – *S. mutans* GbnB      *B* – *S. enterica* GbnB

All samples collected after 72 hours growth on SFM media containing 1 mM EDTA. Plates were overlaid with Thermo Fisher protease inhibitor cocktail for tissue culture at 24 hours growth. SIGMAFAST protease inhibitor cocktail was also added to the protein re-suspension buffer. Intracellular proteins were collected using the method specified in Section 2.3.11.

Unsurprisingly, there are many more bands seen for internal cell proteins collected than for the exported ones studied earlier. Even after nickel Sepharose purification there were no distinct bands that would have indicated the accumulation of a particular protein in the imidazole elution fraction. This appears therefore to suggest that it is unlikely that there is significant accumulation of folded GbnB analogues inside the cell which are not being exported. This does leave the possibility of the *gbnB* genes are not being expressed at all or are being expressed but incorrect folding results in protein degradation within the cell. However, due to the presence of bands of the correct molecular weight seen in Figure 7.10,

low expression levels might explain why the protein of interest was not seen previously, something that could be improved with further optimisation of the culture protocol as well as a more rigorous purification system to achieve fewer non-specific bands. Whether these GbnB analogues can be purified as functional enzymes however, is yet to be seen.

## 7.5 Discussion and Conclusions on the Work to Develop a Novel Inducible Expression System

Whereas the luminescence data collected on MmfR and MmyR helped draw conclusions and offered a fairly complete picture of their regulatory activity, the work creating a novel inducible expression system did not reach such completion with much optimisation left to be done. This chapter utilises a number of technologies that had not been previously used in this project and so learning them took time. Time limitations meant that optimisation could not be carried out to a level where the inducible expression system was ready for use by others with alternative recombinant proteins. The research did however provide some unexpected insights into the regulatory activity of MmfR/MmyR.

Although the attempt to develop *S. albus* as an optimised heterologous expression host was not successful, these investigations did provide extra information on the regulatory natures of MmfR and MmyR and how these relate to the activity of homologues. Specifically, the investigation into *S. albus* as a potential host revealed the possibility of MmfR/MmyR homologues interacting with the MARE operator *s*. From the *in silico* and reporter gene assay analysis carried out in Section 7.2 it is not unreasonable to predict that possible MmfR will be able to bind alternative operators in different streptomyces hosts as well as MmfR orthologues possibly being able to bind the methylenomycin cluster operators.

Despite these findings being very interesting however, the cut off point for sequence identity in the DNA binding region of MmfR, MmyR and their homologues where they will bind the MARE operators and other operators is not known. This therefore casts doubt on the MmfR/MMF/MARE operator system being used in strains of streptomyces which have MmfR homologues with particularly high sequence identity, with the possibility of the presence of alternative MmfR binding sequences as well as other TetR family members interacting with the MARE operators in the inducible system. As a result, a wider analysis of MmfR homologues and MARE operator-like sequences is needed to further understand the interactions suggested in Section 7.2.

The developing inducible expression system shows promise as something that is worthwhile continuing to optimise. Results from SDS-PAGE gels indicate that the previously unpurified GbnB-like proteins may have been successfully produced by this streptomyces system with bands produced at the expected size but this is currently at a fairly low level. The *luxCDABE* operon, located downstream of the *gbnB* analogues being expressed and successfully resulted in luminescence being produced, so it is conceivable that the *gbnB* analogues are also being expressed. However, it is still not entirely clear whether the analogues are successfully being exported or folding into their native conformation, with the results from the LC-MS analysis not offering any clear indication of the presence of *S. enterica* or *S. mutans* peptides. Hopefully with the use of a more rigorous purification technique, there will be more distinct bands seen for the histidine-tagged proteins. Should more time have been available for this project, a more thorough purification protocol, such as FPLC would also have been attempted to achieve a reduced number of bands.

In terms of the inducible nature of the system, this is not something that has yet been shown experimentally, with strains containing pKMS01 giving no clear differences to those containing only pKMS05 or 6. However, it is known from the luciferase assay in Section 4.4 that the release of MmFR will be induced upon the addition of any of the five MMFs and so this system should hypothetically be adaptable for the regulation of other genes. The MmFR/MMF/MARE operator system could be tested further using an alternative reporter assay such as *gusA* or through the use of an antibiotic resistance gene, where the detection of repression and its consequent release would validate the inducibility of the system.

The luminescence trials showed that there was leakage of *luxCDABE* expression when the MmFR repressor was under the control of the *ermE\** promoter. Since the development of *ermEp\** in 1985 (131) as a strong constitutive promoter, much more research has been carried out into stronger promoters that can be used in these GC high bacteria. If *ermEp\** was replaced with another stronger promoter such as *kasOp\** (134) in pKMS01, a higher level of MmFR would likely be produced and therefore more full repression at the methylenomycin cluster promoter of the expression vectors may be seen.

The majority of the work done with optimising the expression system was done using solid cultures. For large-scale industrial production of soluble proteins of course, liquid cultures are preferable due to their ability to be scaled up to a larger extent. With the removal of the energy demanding *lux* genes and some further fine-tuning, it is hoped that liquid culture without the cell lysis will give a useful system.

Testing the novel inducible expression system on two proteins that had never successfully been purified was always going to be challenging and in hindsight, a control vector

overexpressing a gene known to be purifiable from streptomycete systems should have also been created. This would have made results from optimisation trials easier to interpret.

The future work required to develop this expression system has been discussed further in Section 8.5.

## 8 General Discussion

### 8.1 Research Questions and Summary of Findings

In the introduction the following research questions and hypotheses were laid out:

#### Research questions

*Do all five methylenomycin cluster promoters, controlled by MmfR, have the same strength?*

*Does MmfR bind in the same way to all three MARE operator sequences?*

*Is MmfR release by the MMFs the same at all three MARE operators?*

*Does MmfR respond to all five furan compounds?*

*Do all five MMFs have the same efficacy?*

*What are the key residues in ligand binding?*

*How does MmyR binding to the MARE operator and the MMFs vary from that of MmfR?*

*Are there any other ligands that MmyR may bind to?*

*Could MmfR, MMFs and MARE operators be used as a multi-host efficient novel inducible expression system for GC rich bacteria? Would this allow the purification of recombinant proteins?*

#### Hypothesis

*The promoters that are predicted to be controlled by MmfR have different -35/-10 sequences so it is possible that they will have varying strengths.*

*In vivo, MmfR will bind to DNA at the MARE operator and be released upon the addition of a MMF compound.*

*The three MARE operators have different semi-palindromic sequences and so are likely to show differential binding to MmfR.*

*MmfR will respond to all of the MMFs but due to the differing length of alkyl chain between the five molecules there is likely to be differences in the binding potential of each.*

*MmyR is only produced after methylenomycin biosynthesis. It will bind to the methylenomycin cluster operators but not be released by the MMFs, thereby repressing biosynthesis when methylenomycin has been produced to conserve cell resources and to protect the bacterium from the potentially lethal effects of excess methylenomycin. Alternatively, an unknown ligand may 'activate' MmyR as a repressor.*

An attempt was made to answer these research questions via the use of a luciferase reporter gene assay, *in silico* amino acid and gene analyses, the creation of repressor mutants and optimisation trials of an MmfR/MMF/MARE operator based inducible expression system in *Streptomyces*.

Findings from the luciferase assay indicated the relative strength of binding at the different MARE operators for both MmfR and MmyR as well as their affinities for the five MMFs. It should be noted however that the relevance of these findings in the larger picture of the wild type system is still not fully understood. The MmfR/MMF/MARE operator system showed promise as a component of a novel inducible expression system, the development of which was met with varying levels of success.

In Section 8.2, each individual research question has been presented and a summary of the findings made for each question displayed. (An analysis of data collected can also be found previously at the end of each individual research chapter.) The impact of this data is then discussed in Section 8.3. The methods used to perform the luciferase assay as well as developing the novel inducible expression system have also been assessed and comments made on possible improvements in Section 8.4.1 and 8.5.1. Following on from this, a discussion of suggested future work is also presented in Section 8.4.2 and 8.5.2. A final summary of all the findings from this project can then be found in Chapter 9.

## 8.2 Answers to the Research Questions

### 8.2.1 Do all five methylenomycin cluster promoters, controlled by MmfR, have the same strength?

It was shown that the five promoters<sup>1</sup> in the three different intergenic regions were of different strengths (Figure 3.16). The luminescence assay indicated that the strength of promoter goes in the following order; *mmyRp* > *mmfRp* > *mmyYp* > *mmyBp* > *mmfLp*. The range in levels of luminescence by the unrepressed positive controls was between an average of 44.6 and 126.2 times the levels produced by the M145 negative control at 72 hours. This represents only a 2.8 fold difference in luminescence between all five promoters, whereas for example, Bai *et al.* saw a 190 fold difference in promoter strength when analysing different modular regulatory elements.(134) It is unclear whether this small range in promoter strengths is indeed representative of the wild type promoter strength or whether the resource demanding nature of the *luxCDABE* system is a limiting factor on measurable luminescence.

---

<sup>1</sup> See Figure 3.5 and Figure 3.6 for the location of these promoters within the methylenomycin BGC

It is interesting that *mmfR* and *mmyR* have the strongest promoters. For each site that is repressed, two MmfR or MmyR monomers are needed. For the enzymes and other regulators coded for by other operons, one single protein molecule should be able to exert its effects on multiple targets. It therefore appears logical that their promoters are not as strong as those for *mmfR* and *mmyR*.

### 8.2.2 Does MmfR bind in the same way to all three MARE operator sequences?

Research indicated that MmfR binds differently to each of the MARE operators. MmfR was shown to bind best at the L1 MARE operator, controlling the expression of its own gene and *mmfLHP*. For the L1+*mmfR* strain there was shown to be a 10-fold decrease in luminescence compared to the positive control L1+pCC4 (Figure 4.2) at 48 hours growth when MmfR was under the control of *ermEp\**. MmfR will also bind and repress at the L2 and L3 MARE operators but with less strength. There was seen to be a 6.5 fold decrease in luminescence for L3+*mmfR* compared to L3+pCC4 and a 3.5 fold decrease for L2+*mmfR* compared to L2+pCC4 thereby influencing the expression of *mmyR* and the *mmyBQEDXCAPK* and *mmyYF* operons. A similar pattern of results was also seen at 72 hours.

For all *mmfR* strains there was always more measureable luminescence than the level produced by the M145 negative control, indicating that there was not full repression of the *lux* operon. At 48 hours L1+*mmfR* produced three times as much luminescence as the M145 control and for L3+*mmfR* there was almost nine times as much whereas there was over 24 times as much luminescence for L2+*mmfR* compared with the M145 control. This revealed varying degrees of apparent leakiness in the system, which will not only be influenced by the level of MmfR binding but also the promoter strength at each of the sites.

As was seen in Chapter 6, the self-regulatory nature of MmfR will also influence the level of leakiness seen in the system. When under the regulation of the L1 intergenic region and *mmfRp* instead of *ermEp\**, MmfR appeared to more tightly regulate the luminescence produced. Therefore, the levels of leakiness seen in the earlier luminescence assays (Figure 4.2) are likely to be greater than that which would be present in the wild type system, where *mmfR* would be naturally under the control of *mmfRp*. It is likely however that in the wild type system, the same pattern will be seen with the L2 MARE operator still being the most leaky followed by the L3 MARE operator and then the L1 MARE operator, due to the predicted relative strength of MmfR binding at these sites.



### 8.2.3 Is MmfR release by the MMFs the same at all three MARE operators?

As well as variations in promoter strength and strength of MmfR/operator binding, there also seems to be variation between the ease of release of MmfR repression from the different MARE operator sites. This was seen when MMF4 was trialled at a single concentration of 100  $\mu$ M in the L1, L2 or L3 based strains (Figure 4.3). L2+*mmfR* and L3+*mmfR* both start with already high levels of luminescence and produce an average of three to 3.3 times as much luminescence upon addition of MMF4 respectively, compared to nine times more for L1+*mmfR* at 48 hours when compared to the same strains without MMFs.

When compared to a pCC4 positive control strain (representing no repression) the L1+*mmfR* strain with MMF4 produces 96% of the luminescence that L1+pCC4 produced. On the other hand, L3+*mmfR* with 100  $\mu$ M MMF4 only achieves around 50% of the luminescence produced by its positive control and L2+*mmfR* achieves around 87%. This indicates that compared to the positive control, MmfR is less readily released at the L3 MARE operator, followed by L2, with L1 being the most readily release. (By comparing the values for *mmfR* strains with MMF4 with their own individual positive and negative control strains, the effects of different promoter strengths should also have been offset in these datasets.)

### 8.2.4 Does MmfR respond to all five furan compounds?

All five MMFs (Figure 4.4) produced a significant increase in luminescence for the L1+*mmfR* strain, containing the *mmfL/mmfr* intergenic region when analysed using a *t*-test (Figure 4.6 and Table 4.2). At a concentration of 100  $\mu$ M, there was seen to be an increase in luminescence of between 7.3 and 11.5 times that of the L1+*mmfR* in the absence of any MMFs. This is between 23.6 and 37.2 times the levels of luminescence produced by the M145 negative control.

### 8.2.5 Do all five MMFs have the same efficacy?

The  $B_{\max}$  and  $K_d$  values were calculated for each of the MMFs for MmfR at the L1 intergenic region, from this it was also possible to calculate the binding potential of each when under the particular conditions found in the luminescence assay. It is apparent that the relative binding potential for each MMF did vary, with the values ranging from 0.22 to 0.67, a three-fold difference.

This data produce the following order of efficacy:

**MMF1 > MMF3 > MMF5 > MMF4 > MMF2**

It appears that the MMFs with the branched alkyl chains (Figure 4.4) work the best at releasing MmfR from the L1 intergenic region followed by the MMFs with the longest straight alkyl chains.

These calculated binding potentials represents the efficacy of each MMF for MmfR in causing its release from the *mmfL/mmfR* intergenic region. It is therefore possible that it will vary at the L2 and L3 intergenic regions. When binding the different operators, MmfR might have a slightly different conformation as a result of the interactions with different DNA sequences, therefore slightly altering the way the MMFs enter and interact with the binding pocket. Unfortunately there was not time to test the full range of concentrations at every MARE operator site.

### 8.2.6 What are the key residues in ligand binding?

Based on an *in silico* analysis of MmfR and observations from its crystal structure with MMF2, two tyrosine residues in positions 85 and 144 were selected as being likely to be involved in ligand binding *in vivo*. The *mmfR* gene had point mutations made to alter these amino acids in MmfR to phenylalanine residues (Figure 4.11). These mutants were added to a pCC4 vector and which was then put into *Streptomyces* strains containing *luxCDABE* under the control of *mmfLp* (L1 strains), allowing the analysis of the ligand-binding activities of these mutants via the luciferase assay.

The Y85F mutant appeared to be released by MMF4 to a significantly lower level than the wild type MmfR. Even at a 200  $\mu$ M concentration of MMF4 this mutant did not result in the same levels of luminescence as were induced in the wild type MmfR with 100  $\mu$ M MMF4 (Figure 4.17). This indicates that this tyrosine residue is indeed key to ligand binding.

The Y144F mutant on the other hand was not expected to show any difference in ligand binding as it is the amine group that is involved in binding the MMFs, and this remained the same in the mutant. Rather unexpectedly however, Y144F with 100  $\mu$ M MMF4 appeared to produce around 125% the luminescence of the wild type strain with the same concentration of MMF4. This indicates the possibility that this change to the ligand-binding pocket may have optimised furan binding, however more *in vitro* tests are needed to prove whether this is indeed the case.

It is also of note that both the mutants appeared to produce slightly less repression of *luxCDABE*. Despite being mutants for the ligand binding pocket, the mutations lie close to the dimer interface of MmfR and so may influence the overall structure of MmfR and therefore also its DNA binding properties.

### 8.2.7 How does MmyR binding to the MARE operators and the MMFs vary from that of MmfR?

Table 8.1 shows a comparison of data collected on MmfR and MmyR in terms of strength of binding to the MARE operators.

Despite MmyR knockouts having been shown to overproduce methylenomycin, the luminescence assay showed a much lower level of repression by MmyR compared to MmfR (Figure 3.8). MmyR showed the greatest level of luminescence repression at the L3 MARE operator followed by the L1 MARE operator. No significant reduction in luminescence was detected in L2+*mmyR* compared to L2+pCC4 however, indicating that MmyR does not bind the L2 MARE operator.

**Table 8.1. Summary of data on MmfR and MmyR binding to the MARE operators**

*The data in this table summarises the findings from Section 4.3 and 5.3*

| MARE operator site | Corresponding promoter | Strength of binding (no MMFs present) |            | MmfR release in the presence of MMF4 compared to a control |                  | Significant change in MmyR binding upon addition of MMF4 | Relative promoter strength (1 being the strongest) |
|--------------------|------------------------|---------------------------------------|------------|--|------------------|--|--|
|                    |                        | MmfR                                  | MmyR       | Negative control   | Positive control |  |  |
| L1                 | <i>mmfLp</i>           | Greatest                              | Weakest    | Greatest   | Greatest         | NO   | 5  |
| L1F                | <i>mmfRp</i>           | -                                     | -          | -  | -                | -  | 2  |
| L2                 | <i>mmyRp</i>           | Weakest                               | No binding | Weakest  | Middle           | NO   | 1  |
| L3                 | <i>mmyBp</i>           | Middle                                | Greatest   | Middle   | Weakest          | YES (increased binding)                                  | 4  |
| L3F                | <i>mmyYp</i>           | -                                     | -          | -  | -                | -  | 3  |

To compare findings on MmyR relative to the data collected on MmfR; MmfR appears to bind the L1 intergenic region with the greatest affinity showing a average of 90% reduction in luminescence compared to L1+pCC4, followed by L3 with an 85% reduction and then L2 at 72% reduction compared to their own pCC4 positive controls. MmyR on the other hand sees a 40% reduction in luminescence for L3+*mmyR* compared to L3+pCC4 and an average reduction in luminescence of 32% for the L1+*mmyR* strain compared to L1+pCC4. It is clear therefore than in their apo forms and under the control of the *ermEp\** promoter, MmfR binds more strongly to the MARE operators than MmyR.

In Table 8.1, it seems clear that the L2 intergenic region has the least regulation by the TFRs; MmfR and MmyR. MmfR best regulates the L1 intergenic region, controlling the production of itself and MmfLHP whereas MmyR appear to best regulate *mmyBQEDXCAPK* and *mmyYF*. It was hypothesised in the 2009 paper by O'Rourke *et al.* (71) that the main role of

MmfR may be to regulate MMF production and the main role of MmyR to regulate the transcriptional activator MmyB. This premise is therefore in line with what was suggested by the data here.

How MmyR actually carries out its repressive role appears to be complex, with much lower levels of repression seen than those produced by MmfR. One theory was that MmyR may need to be bound to a ligand for it to be able to bind the MARE operator and repress transcription. MmyR lacks both the tyrosine 85 and 144 residues previously associated with hydrogen binding with the furan ligand in MmfR (Figure 4.14). Unsurprisingly therefore, luminescence in the L1+*mmyR* strains did not show any significant induction in the presence of any of the MMFs at a concentration of 100  $\mu$ M (Table 5.6). There was however an indication that MmyR will work better as a repressor in the presence of the MMFs when bound at the L3 MARE operator. The addition of 100  $\mu$ M MMF5 to L3+*mmyR* appeared to bring the levels of luminescence closer to those seen for *mmfR* strains, showing 25% of the luminescence seen for L3+pCC4 (compared to 15% for L3+*mmfR* with no MMFs). Future trials with MMF1 and 3 will hopefully bring about even greater levels of repression, particularly as they had the highest binding potentials when binding MmfR.

It is not clear why no such difference in luminescence was seen in the presence of the MMFs at the L1 MARE operator. However at this MARE operator there was seen to be weaker binding of MmyR and it is possible that only with a higher concentration of MMFs that a significant change in the levels of repression may be seen. Alternatively, the differences in the MARE operator sequences between the L1 and L3 sites may slightly alter the conformation of MmyR in a way that adjusts how the MMFs can bind. Again, this is something that would be interesting to look at in future trials.

In summary, MmyR appears to have very different binding affinities for both the MMFs and the MARE operators when compared to MmfR. It appears to significantly interact with the methylenomycin system, but in a very different way to MmfR. Having two repressors with different roles and affinities thereby appears to allow tighter control of methylenomycin regulation.

### 8.2.8 Are there any other ligands that MmyR may bind to?

In other GBL binding systems there is often a TetR family GBL receptor as well as a paralogous ‘pseudoreceptor’ which has different ligand binding properties as well as a greater promiscuity of DNA targets. As MmyR is analogous to some of these ‘pseudoreceptors’ it seemed logical that it may bind an alternative ligand than the MMFs.

Methylenomycin A and its precursor methylenomycin C were trialled to see if they caused a

significant change in luminescence in the L1+*mmyR* strain. Results from this assay were inconclusive with the lower concentration of the methylenomycins producing greater changes than the higher concentrations. Despite controls being run, it was unclear whether these artefacts were a result of the toxicity of the compounds in the cultures or truly representative of a change in repressive activity of MmyR. However, when a number of different architectures were trialled in mathematical models of the methylenomycin regulatory system, models where MmyR was released upon binding methylenomycin did not match the phenotypes seen in previously collected experimental data. This therefore indicates that it is unlikely that MmyR binds methylenomycin A or C. It would be helpful to analyse this further with the luciferase assay however due to the limitations of the mathematical modelling performed (Section 6.5).

The potential of an alternative MmyR ligand cannot be excluded but as yet there are no firm hypotheses on what these molecules may be.

### 8.2.9 Could MmfR, MMFs and MARE operators be used as a multi-host efficient novel inducible expression system for GC rich bacteria? Would this allow the purification of recombinant proteins?

In terms of developing a novel inducible expression system, the *lux* vectors have certainly shown promise with high levels of expression produced clearly measurable levels of luminescence, with up to an 18 times increase in luminescence once the MMF inducer has been added (Figure 4.7). The *lux* genes can easily be replaced with a gene of interest to be over expressed, under the control of MmfR/MMF/MARE operator, which should hypothetically be able to be controlled in the same way as the *lux* genes were. Two separate GbnB analogues from *S. mutans* and *S. enterica* were inserted between the L1 intergenic region and *luxCDABE* in the L1 vector for use as a novel inducible expression system.

The optimisation of the inducible expression system proved to be challenging however, with little known about how to purify proteins from *S. coelicolor* compared to the better-known *E. coli* systems. In particular there were a number of challenges that were encountered when trying to develop a heterologous expression super host meaning that trials were done in *S. coelicolor* M145 instead, a strain which has limitations when trying to grow it in liquid culture. While optimising this system there appeared to be a good improvement in protein yield of secreted protein of the expected molecular weight when using a selection of protease inhibitors, indicating that protein degradation may have a significant impact on protein yield. There is hope that with even further optimisation and more extensive purification of the proteins of interest that a continued improvement in protein yield may be seen.

The presence of a band of the right size for the GbnB analogues does indicate that this actinomycete system may allow the purification of recombinant proteins from GC rich bacteria. This however still needs to be confirmed with further tests on the purified proteins to identify them as the desired GbnB-like proteins, as the LC-MS analysis done was largely inconclusive.

The development of an optimised heterologous expression host was met with very limited success. It did however, help to shed some light on the cross-species promiscuity of GBL-like receptors, through the *in vivo* and *in silico* work done with *S. albus*. The *lux* system was added to *S. albus* with the potential of it being used as a host for the novel inducible expression system due to its very small, streamlined genome size. It was found however that luminescence for the L3+pCC4 positive control was repressed (Figure 7.1). A BLAST search for MmfR homologues found the TetR family member SSHG\_01258, which shared over 40% identity with MmfR (across 86% of its sequence) and was potentially binding to the MARE operator and repressing *mmyBp*. No significant MmyR homologue could be found however.

This indicates an extra hurdle when developing a multi-host efficient inducible expression system. TetR family members share a homologous helix-turn-helix DNA binding region which sometimes may share enough sequence identity to bind to one another's target DNA sequences. This poses a limitation when transferring the MmfR/MMF/MARE operator expression system between hosts that also contain homologues with high sequence identity to MmfR/MmyR. Even weak binding to the MARE operator by native host receptors could be enough to considerably interfere with an inducible expression system and its regulation. Of course, with the advancing and increasing availability of many bacterial genomes, an examination could be run in potential expression hosts for MmfR homologues. However, the cut off for sequence identity that would result in these analogues binding the methylenomycin cluster MARE operators is as yet unknown.

In summary, the MmfR/MMF/MARE operator system still shows promise as an inducible expression system in streptomycetes. However, much more optimisation is needed before it could be widely used.

### 8.3 Impact of Data Collected

There is hope that the research presented in this thesis can shed some light on the biosynthesis of some other natural products from strains of *Streptomyces* that contain homologues of the MmfLHP/MmfR/MmyR system. For example, tyrosine 85 and 144 (believed be used in ligand binding) are conserved in other homologous proteins indicating that other systems may

also use furan molecules as ligands. Some of these homologues are from otherwise silent and less well understood pathways and so have the potential of revealing otherwise undiscovered antibiotics if manipulated in the correct way.

In particular this research has revealed more about the functioning of TetR family pairs analogous to GBL receptors and pseudoreceptors and how they have distinct roles in regulating secondary metabolism. It is apparent that whereas the GBL-like receptors often have similar and predictable mechanisms of action where they bind a hormonal ligand and then are released from a DNA operator, the paralogous ‘pseudoreceptors’ are much harder to predict the function of. Despite the knockouts of these pseudoreceptors often bringing about the same phenotype of overproduction of the cognate secondary metabolite, the mechanism by which they achieve repression appears to vary hugely. Sometimes the pseudoreceptor will only be released by a different, non-GBL ligand, possibly the cognate natural product. The findings of this thesis on the other hand, indicate that MmyR and MmfR may both bind the MMFs but this will bring about a completely different effect on MmyR, that of increased repression compared to a release of repression that these ligands bring about for MmfR. There are also quite distinct differences in the affinity of MmfR versus MmyR for the DNA operators. All this information is helpful when trying to switch on secondary metabolite production for silent gene clusters, with a broadened selection of possible alternatives to be trialled when trying to manipulate the activity of MmyR analogues.

The research into the DNA binding domains of MmfR, MmyR and their homologues (Table 7.7) revealed a number of common motifs, which could also be of relevance when manipulating TetR family binding to a chosen DNA operator. This does need further investigation but could prove to be very useful when seeking to control other biosynthetic clusters as well as engineering the novel inducible expression system. In particular, the work on MmfR/MmyR may help to better understand the regulation by SAV\_2270 and SAV\_2268 in *S. avermitilis* and SHJG\_7318 and SHJG\_7322 in *S. hygrosopicus*, of which the natural product they regulate is unknown. These homologues shared the identical DNA binding motifs to those found in MmfR and MmyR and so possibly will reveal similar DNA binding profiles to those in the methylenomycin regulatory system studied here.

## 8.4 Discussion and Improvements of Luciferase Assay

### 8.4.1 Discussion of Techniques

The chosen luminescence assay provided easily detectible results quantitative results which were possible to collect in a 30 second reading making this assay quick to obtain data from.

The Photek CCD camera was sensitive enough that small changes in luminescence could be detected for concentrations of the MMFs as low as 5  $\mu\text{M}$ . A summary of the positive and negative aspects of the *luxCDABE* assay has been presented in Table 8.2.

**Table 8.2. Summary of the positive and negative aspects of the *luxCDABE* reporter gene assay**

| <i>Positive aspects</i>  | <i>Negative aspects</i>  |
|--|--|
| Allowed the use of solid cultures  | Energy taxing nature of the <i>luxCDABE</i> operon resulted in phenotypic differences of some cells                                  |
| Multiple readings over time was possible as this was a non-destructive assay   | It was not practical to measure the cell mass from solid cultures  |
| Did not need to add substrates for luciferase, everything was made in the system   | Large deviations between results meaning that a number of repeats needed to be taken   |
| Easy manipulation of the system to study alternative operators, promoters and repressors via the sub-cloning and manipulation of vectors | The software for the Photek camera was old so transfer of data into a readable format for other computers could be a lengthy process |
| Very sensitive CCD camera  |  |
| Quantitative and easily detectable results   |  |

In particular, this luciferase assay allowed the use of solid culture, where *S. coelicolor* grows better, completing its complex sporogenic life cycle. The use of solid culture also tends to be less susceptible to contamination while manipulating samples to take multiple readings. In addition to this, solid cultures were smaller in volume than the liquid ones trialed and so were less demanding on the synthesis of purified MMFs, which are not commercially available. The measurement of luminescence from solid cultures using the Photek CCD camera was also non-destructive, allowing the collection of repeat measurements from the same samples over time.

Liquid cultures proved to be impractical to use for this assay. The large clumps that *S. coelicolor* form in liquid culture meant that there was huge variation between results collected. Also, without being able to complete its sporogenic life cycle, it is unclear how results collected from liquid cultures may be representative of methylenomycin regulation in the wild type system, with unknown interplay from extra transcriptional control such as sigma factors and riboswitches possible that may be dependent on different stages in the life cycle. The use of baffled flasks needed for *Streptomyces* growth in liquid culture also meant that samples had to be transferred into a different plate for measurements by the Photek camera, increasing the chances of contamination. It was also not practical to add the sample back into the baffled flask after the luminescence had been measured, meaning that the assay was destructive and reduced the volume of the culture each time a reading was taken. Other strains of *Streptomyces* such as *S. venezuelae* do grow better in liquid culture but as was seen in



Section 7.2, analogous TFRs may interfere with the system and so knockouts would need to be created. Using an alternative host is therefore not as simple a transition as may be desired.

A downside of using solid culture was that the exact mass of cells was not known for each sample. Methods to measure the mass of solid cultures are often disruptive to the lawn of bacteria or be destructive to the sample and result in contamination and so would have prevented repeat readings being taken for the same cultures over time. Also, when inoculating plates directly from glycerol stocks (which were of a known concentration), luminescence production appeared to be effected by storage of the stocks at -80 °C. For this reason, a lawn of *Streptomyces* was grown on selective SFM media for four to five days at 30°C and this formed a fresh starter culture for the rest of the study. Picking cells fresh from a selective SFM plate ensured that they would be luminescing optimally for data collection but the concentration of cell mass was not known. To overcome the problem of not knowing the exact cell density, multiple repeats were done for each sample type and luminescence was calculated as a ratio, relative to a control sample. This control samples would be under the same variations in the inoculation of individual cultures and so should cancel out any effects of variation in cell mass and can be used to normalise data.

It was also found that there is much less variability between readings at the later time points likely to be due to the stability brought about by cells were entering the stationary phase of growth. These 48 and 72 hour time points were therefore more reliable time points to study in detail, using boxplots and *t*-tests, due to a lower standard deviation and coefficient of variance. If these assays were to be done again, it is likely that the decision would also be taken to reduce the number of readings taken, removing the 21 and 27 hour time point readings and just collecting data at 24, 48 and 72 hours. The earlier time points were initially used as there was a lot of change in luminescence earlier on in growth, however the deviations in luminescence during this exponential phase make a reliable analysis of results difficult. Reducing the number of readings taken around 24 hours would have the benefit of decreasing any impact of temperature changes when taking samples out of the 30 °C incubator. Although samples are removed from the incubator for less than ten minutes at a time, it is unclear what the impact of initially doing this every three hours was.

Another positive aspect of the luciferase assay was that it was easy to adjust and adapt for the study of different operators, promoters and repressors. Through the use of restriction digests and sub-cloning, different methylenomycin cluster intergenic regions could be studied and MmfR mutants were easily added to the system. This allowed the creation of a large array of different *lux* strains, permitting the analysis of a range of variables, brought together in different combinations.

Using *luxCDABE* compared to *luxAB* is beneficial because nothing needs to be added to the system. Both luciferase and its substrate are produced and so there is luminescence generated without any external manipulations. Anything extra needed to produce this bioluminescence is already present in the *lux* host strain. The downside of this self-contained system is the likelihood that the constitutive expression of *luxCDABE* is very taxing on the cells resources and perhaps for example, the strongest promoters are not revealing the true extent of their strength when tested with this assay. This is something that would be helpful to investigate further, looking at the different metabolic profiles of cells with phenotypic differences compared with those presenting the wild type phenotype of pigments produced.

In conclusion, the luciferase assay proved to be very useful in the study of methylenomycin regulation in *S. coelicolor*. As with any reporter gene assay there were both positive and negative aspects of collecting results in this manner. The main downside of this technique was the variation between results. With more information on the cross-species affinity TFRs have for each others operator sites, this system could be developed for use in streptomycete strains that grow better in liquid culture. This would allow the calculation of cell mass in each sample and would hopefully reduce the deviations between data sets. However, until this is achieved, the luciferase assay would likely be suitable as it is in future investigations.

#### 8.4.2 Luciferase Assay Future Work

The luciferase assay provided a lot of useful data on the regulation of methylenomycin biosynthesis and so would be used again to test more variables involving MmfR, MmyR and the MMFs. In particular, there is a desire to collect data on a number of other controls to produce a better-rounded assessment of biosynthetic regulation. There is hope that some of these trials will be completed during a short post-doctorate project to be carried out after the conclusion of this report and will be done again using the protocols specified in Section 3.3.7 and Figure 3.3. These potential trials will now be discussed.

A control test was run where the five MMFs were tested on the L1+pCC4 strain to look for significant changes in luminescence, in the absence of MmfR or MmyR. In the light of research indicating that the MMFs may have extra activator roles as well as causing the release of MmfR from the MARE operator (Section 4.4.2) it seems apparent that a full set of trials should also be run with L2+pCC4 and L3+pCC4 with all five MMFs. Preliminary trials were run with these two strains in the presence of a single concentration of some of the MMFs but this was not prioritised and a full data set was not collected. It would maybe even be useful to trial a range of concentration of the MMFs with these positive control strains to assess whether there is a calculable  $K_d$  or  $B_{max}$  value. It is possible that the MMFs may be used to recruit other transcriptional regulators, located at different loci to the methylenomycin

cluster and so would be able to alter levels of luminescence even in the absence of MmfR/MmyR. Alternatively, the MMFs may be involved in the activation of a riboswitch in some of the methylenomycin cluster genes.

Another area that would be interesting to explore with the luciferase assay would be a study into the strength of other streptomycete promoters. In this way, the effects of the potentially taxing expression of *luxCDABE* could not only be further explored but also a promoter could be selected for the over expression of *mmfR*. Something that would be particularly useful in the development of the novel inducible expression system where more complete repression of recombinant genes by MmfR is desired (see Section 8.5). The *gfp* assay work by Bai *et al.* from 2015 could be used as a benchmark for promoter strength in this particular assay. In their work, Bai *et al.* developed the *kasOp\** promoter (135) which was found to be ~20 times stronger than *ermEp\** and so is of potential interest as a strong constitutive promoter. As before, alternative promoters could be inserted upstream of the *luxCDABE* genes and the relative bioluminescence produced observed in the same manner as was done in Section 3.4. This could also then be compared to the work by Bai *et al.* to see if the same fold difference in luminescence for different promoters was seen in their equivalent *gfp* assay, thus indicating any limitations of *luxCDABE* expression.

If more time allowed the L1F+*mmfR* and L3F+*mmfR* strains would also be further explored and compared to their L1 and L3 counterparts to investigate the influence of promoter position compared to the MARE operator in the reversed intergenic regions. The strength of MmfR/MmyR binding should be the same in either orientation due to the MARE operator remaining the same but the proximity of the MARE to the corresponding promoter may have an effect on the ‘leakiness’ of the system.

Investigations into MmfR/MMF/MARE operator interactions were more extensive than those done for MmyR. There is interest in further investigating the role of the lesser-understood MmyR, in particular looking at the effect of different concentrations of the MMFs on this repressor at the different MARE operators. During investigations in Section 5.4 there was a lack of evidence of a significant effect on luminescence by 100  $\mu$ M for L1+*mmyR* and so no trials were run to find out the  $K_d$  and  $B_{max}$  values for *mmyR* strains, in an attempt to conserve MMF stocks. However, since then possible MmyR/MMF interactions have been shown at the L3 intergenic region, showing more promise for the production of significant and utilisable results if trialled with a full range of MMF concentrations. These trials were not run previously due to the finalisation of the lab project but it would be very interesting to carry out these tests in the near future. The aim of these further trials would be to investigate

whether higher concentrations of the MMFs could be used to achieve ‘full’ repression at the L3 intergenic region or possibly a significant change in repression at the L1 intergenic region.

Another area that would be fascinating to explore using the luciferase assay is the cross-strain promiscuity of the DNA binding domain of TetRs analogous to MmfR. Either *sav\_2268* or *sgnR* could be subcloned downstream of *ermEp\** in pCC4 and added to the system along with the L1, L2 or L3 vectors. The level of repression achieved by these homologous repressors at the different intergenic regions, if any, could then be determined. In addition to looking at DNA binding properties, these analogues share high sequence identity in their ligand binding pockets to MmfR and so this assay set up could also be used to investigate whether SAV\_2268 and SgnR can bind to and are released by the furan ligands *in vivo*.

Finally, the investigation into the auto-regulation of MmfR and possible MmyR was only in its preliminary stages before this project concluded. There are many more investigations that could be carried out using the sp105 and 11NY strains, including the trial of a gradient of MMFs to see if luminescence could be induced at a specific threshold level. It would also be interesting to alter the 11NY and sp105 vectors to include alternative intergenic regions in an auto-regulatory system.

## 8.5 Discussion and Improvements of the Novel Inducible Expression System

### 8.5.1 Discussion of Techniques

Investigations into the development of a novel inducible expression system were divided into two main parts; research into the potential use of *S. albus* as heterologous super host and the optimisation of the MmfR/MMF/MARE operator inducible expression system in a *S. coelicolor* M145 host. This investigation came after a selection of data from the luciferase assay had been obtained. For this reason, this section of the investigation was much more constrained by time as it came nearer the end of the project. These methods required much optimisation and could not be developed as far as being a useable heterologous expression system that could be employed in multiple hosts. However, the data collected did form a good basis of preliminary investigations, which can then be developed further in the future. For this reason there are large overlaps between technique discussion and improvements and planned future experimentation compared to the discussion of the luciferase assay technique.

#### **Creation of an Optimised Heterologous Expression Host**

The luciferase assay used to investigate the MmfR/MMF/MARE operator system in *S. albus* has already been discussed in Section 8.4.1 but there are some further improvements that

would be beneficial with regards to this particular strain. In particular, it was a concern that some of the findings of the luciferase assay in Section 7.2.2 were due to problems with *luxCDABE* expression in *S. albus*, possibly as a result of the genome rearrangements this strain undergoes. It would potentially be beneficial therefore to regularly re-check the *S. albus* genome for the presence of the luciferase inserts should the assay be done again, to confirm that the results seen were not due to impeded genomic *lux* maintenance.

### **Optimising the Novel Inducible Expression System**

When developing the inducible expression system for *S. coelicolor* M145 a number of different variables were trialled including the use of liquid versus solid media, different carbon sources, harvesting proteins at different time points as well as the use of protease inhibitors. The use of protease inhibitors showed promise in the optimisation of the technique and so is something that should certainly continue to be used over the 72 hours of *Streptomyces* growth in future trials. One major way in which the results of these optimisation trials may be improved is through the use of more extensive purification methods. During the investigation in Section 7.4, a very basic nickel Sepharose purification technique was used to provide a more high throughput system to test assorted variables. However, a more distinct band at the expect position on the SDS-PAGE gel may be achieved with a more refined method such as fast protein liquid chromatography FPLC. This semi-automated technique would hopefully reduce the number of non-specific bands seen in the purified samples and make results easier to analyse.

Also, despite SDS-PAGE bands being analysed using LC-MS, the results from this were at times hard to analyse and inconclusive, due to the presence of so many potential polypeptides. An additional technique such as Western blotting therefore may be of use when looking for the presence of the histidine tagged recombinant GbnB-like proteins.

Further improvements could also be made to the design of the expression vectors being used, with the benefit of hindsight and lessons learnt while optimising the expression system. For example, there was no spacer included between the TAT export signal and the 6xhis-tag. It became apparent later that the natural cleavage of the TAT signal might also cleave off some of the histidine tag, resulting in a potentially lower binding affinity for the nickel Sepharose. To remove this possible problem, either a spacer of random DNA could be added between the TAT signal and histidine tag or the histidine tag could be made longer so that any cleavage would still leave at least six histidine residues.

In addition to this, there was no cleavage site added for the removal of the histidine tag. Although this was not a problem in the optimisation trials, for the expression system to be adaptable for a wide range of different proteins and uses it would be helpful to add a cleavage

site. Another obvious improvement that will need to be made to the expression vectors is the removal of *luxCDABE*. These genes were left in the vectors created to allow expression to be tracked via luminescence measured with the Photek CCD camera as well as via the more lengthy process of checking proteins via SDS-PAGE. However, due to the resource demanding nature of these *lux* genes, it is likely that protein expression is not optimal while they are still in the system and so will be removed, using the restriction sites included in the vector for this very reason, before any future trials.

## 8.5.2 Novel Inducible Expression System Future Work

### **Creation of an Optimised Heterologous Expression Host**

In addition to *S. albus*, there are a number of other possible streptomycetes that could be developed as heterologous expression super hosts. One of these is the reduced *S. coelicolor* strain, M1152.(123) Like *S. coelicolor* M145, this strain is lacking the SCP1 and SCP2 plasmids but also has a number of extra genes removed, making this strain even more streamlined and therefore potentially a better host. Due to M1152 being a reduced version of M145, there is little chance of any unexpected background interactions not encountered in the work done for this project. It is not currently the perfect host however. It was found recently that the M1152 host still contains some biosynthetic genes from its gamma butyrolactone pathway but lacks repressor ScbR2 and therefore overexpresses these GBLs.(155) It has been experimentally shown that SCB1-3 (*S. coelicolor* butyrolactones 1-3), involved in coelimycin regulation, are over produced in the M1152 strain as well as five novel SCBs (SCB4-8).(155) This is a drain on the cells resources as well as adding the potential of these unregulated signalling molecules interacting with other pathways. In the M145 strain, these SCBs are under their normal wild type regulation by both ScbR and ScbR2 and so this overproduction has not been shown to be a problem.

Before M1152 can be used as a heterologous expression host, this problem ideally needs to be solved and three genes of interest removed from the genome; *scbR*, *scbA* and *scbB*. ScbA and ScbB are involved in the biosynthesis of the SCBs. There are a number of techniques that could be used to achieve these gene knockouts. An example of a technique for gene manipulation which has recently grown in fame is that of CRISPR/cas9.(156) Unfortunately this technique has not shown successes when trialled in *S. coelicolor* by other lab members at Warwick, due to potential toxicity caused by unrepaired double stranded breaks in the genome and inefficient transfer of DNA via conjugation. It is unfortunate that such a powerful technique cannot be used for this purpose, at least not without some major improvements.

Mutagenesis techniques in streptomycetes often employ gene targeting, where homologous recombination is used to change a plasmid-based endogenous gene, followed by transfer into

*Streptomyces* via conjugation. Problems often come about at the stage of conjugal transfer of plasmids from *E. coli* into *Streptomyces*. Techniques recently developed by Netzker *et al.* earlier in 2016 (157) may be able to assist with solving these problems. This research group developed an optimised conjugation protocol by looking at different media, antibiotic concentration, temperature and calcium ion concentration. It is therefore suggested these newly optimised conjugal transfer techniques are used to remove *scbR*, *scbA* and *scbB* from *S. coelicolor* M1152, possibly using PCR targeting. PCR targeting is a protocol specified in the ‘Redirect technology: PCR-targeting system in *Streptomyces coelicolor*’ manual (158) designed by Bertolt Gust, Tobias Kieser and Keith Chater from the John Innes Centre. This protocol is based on a well known mutational technique developed by Datsenko and Wanner in 2000 for use in *E. coli* (159) but has been adapted for *S. coelicolor*. The procedure involves the excision of specific genes from a streptomycete cosmid in exchange for a PCR product of a selectable marker with added short homologous arms. This is done in a strain of *E. coli* with a high recombination and mutation rate (due to the presence of genes for  $\lambda$  RED recombination). These modified cosmids are then transferred to *S. coelicolor* via conjugation and the mutation integrated into the genome via homologous recombination of large stretches of native DNA in the cosmid.

These techniques could also possibly be used for the removal of *sgnR/gbnR/sgnLHP* from *S. venezuelae* or *sshg\_01258* from *S. albus* with the prospect that removing these genes may prevent any interference with MmfR/MMF/MARE operator in these alternative hosts, hopefully allowing the cross-species adaptability of the inducible expression system. *S. venezuelae* has been shown to sporulate in liquid cultures and liquid cultures have the potential to produce larger volumes of biomass and therefore a greater potential protein yield meaning that this is an attractive alternative host.

In addition to work to be done in the laboratory, a more widespread amino acid analysis with a detailed look at predicted TetR family protein functions would also be beneficial. Any extra information obtained on the cross-species affinity TFRs for each other's operator sites would assist in the development of the MmfR/MMF/MARE operator system for use in alternative streptomycete strains e.g. *S. lividans*. DNA binding promiscuity between transcriptional repressors could also be investigated *in vitro* using a gel shift assay or surface plasmon resonance (SPR) to test for receptor affinity for specific DNA sequences and cross-talk with the MmfR/MARE operator system.

### **Optimising the Novel Inducible Expression System**

It was seen in Section 7.4.4 that the addition of pKMS01 (*mmfR*) to strains containing pKMS05 or 06 did not result in any obvious reduction in protein synthesis. This is possibly

because the bands seen are unrelated to *luxCDABE* or the GbnB analogues, and are therefore not controlled by MmfR repression. Alternatively it could be due to the ‘leakiness’ of repression at the L1 intergenic region. The L1 intergenic region was selected for use in the inducible expression system as MmfR showed the strongest binding at this site, it appears however that even this small amount of leakiness may still result in significant levels of recombinant gene expression. To fully optimise the system, it would be preferable that *mmfR* be put under the control of a stronger promoter than *ermEp\** with the hope that higher levels of MmfR will result in more complete repression. There was evidence in Section 6.2 when under the control of its own promoter, levels of MmfR repression appeared greater. This indicates that it is a problem with the levels of MmfR present rather than its repressive ability that allows significant leaky gene expression. Unfortunately the native methylenomycin promoters such as *mmfLp* are unsuitable for controlling *mmfR* expression due to their self-regulatory control and lack of observable induction by the MMFs under laboratory conditions. Instead there are a variety of strong constitutional promoters that could be trialled, for example the *kasOp\** promoter which was indicated to produce 20 times more *gfp* fluorescence than when the same gene was under the control of *ermEp\** (See also Section 8.4.2).(135) Again using the luciferase assay, the promoter that is shown to achieve the best levels of MmfR repression could then be used in the novel expression system to make it truly inducible.

An additional adjustment that would also be interesting to explore is the use of a different Tat export signals. Li *et al.* presented research on Tat export in *S. coelicolor* A(3)2 in 2005 (148) where details were given of the signal peptides for 129 possible Tat substrates, as predicted by their software; TATscan. It is possible that if a different one of these was used, an improvement in protein yield may also be seen. The specific efficiency of different signal peptides is not known and so one was picked at random for this assay but it may not be as efficient as other signal peptides. A range of alternative Tat signals could be sub-cloned into pKMS05 or 06 and expression levels assessed.

When designing the expression system, the two GbnB analogues were selected due to the interest by the research group at Warwick and GSK into their purification. However in hindsight, selecting two proteins that had never before been successfully purified was not a good place to start with a technique that needed so much optimisation. It would be beneficial instead to also try over expressing a gene that is known to be extractable from current systems available for GC high bacteria such as the gene for that of streptavidin.(107) This would then provide a control for optimisation trials. It is possible that a better indication of the presence of the GbnB-like proteins was not achieved because the system was not optimal for protein



folding or any other factors involved in the successful collection of a soluble protein for these specific polypeptides rather than a flaw in the actual expression system.

## 8.6 Other Relevant Future Work

As mentioned in the introduction, the exact mechanism of methylenomycin activity is as yet unknown. There are plans to investigate this further in the future. A selection of *B. subtilis* reporter strains have been developed which contain five different promoters attached to the firefly luciferase gene.(160),(129) The levels of luminescence produced by each strain in the presence of an antibiotic indicate what its biological target is. Once this target has been established, further research can be carried out looking for greater depth of understanding on interactions with this target. Another researcher within Warwick University is currently carrying out this work.

In addition, work is also being carried out to purify a MmyR orthologue from *Streptomyces avermitilis*; SAV\_2270. This orthologue has been indicated to be soluble when purified with commonly used *E. coli* systems, meaning that crystallisation may be possible. This would help to shed light on the functionality of these ‘pseudoreceptor’ type proteins. If the crystal structure of SAV\_2270 could be obtained, the structure of MmyR may also be possible to be modelled based on this analogue, allowing a greater understanding of key amino acids in DNA and ligand binding including opening the possibility of more accurate docking analyses with potential ligands. Previous attempts to model MmyR based on MmfR have as yet proved to be unsuccessful, with not enough homology of structure to produce viable docking analyses.

## 9 Summary of Results and Conclusions

To conclude all of the observations made during the experimentation carried out for this project, an interesting insight has been developed into the roles of the paralogous pair MmfR and MmyR in the regulation of methylenomycin biosynthesis.

The better-understood MmfR was previously predicted to bind three different intergenic regions at a 24 bp methylenomycin auto-regulatory response element site, thereby regulating the expression of five different operons. This investigation revealed that these five different operons have promoters of differing strengths. The promoters for *mmfR* and *mmyR* appeared to be the strongest whereas the promoters for other enzymes and regulators were weaker. In addition to this, there is a variation in the strength of MmfR binding to the three MARE operator sites as well as the level of release of this repressor that can be induced by a single concentration of MMF4 ligand. Each of the five MMF ligands were shown to have a different efficacy for MmfR when it was repressing at the *mmfL/mmfr* intergenic region, with the furans that have the branched alkyl chains proving to be the most efficient ligands. An *in silico* analysis of the MmfR primary and tertiary structure revealed the presence of two tyrosine residues thought to form hydrogen bonds with the furan ligands. A mutational analysis of these residues indicated that they are indeed involved in ligand binding and may also effect DNA binding due to their close proximity to the dimer interface.

The activity of MmyR varied from that of MmfR. This paralogue showed significant binding to two out of the three MARE operator sites that MmfR was indicated to bind, but showed no significant binding to the intergenic region that contains the promoter for its own gene. As with MmfR, there was variability in affinity of MmyR for each of the two operator sequences it did appear to bind. MmyR showed no significant release in the presence of the MMFs from either site, however a level of significant improvement in repression was detected for MmyR at the *mmyY/mmyB* intergenic region in the presence of all the MMFs trialled. No conclusive data was collected on the possibility of MmyR binding to methylenomycin A or C but the possibility of other alternative ligands cannot be ruled out.

The MmfR/MMF/MARE operator system shows promise as something that can be developed into a novel inducible expression system, although much optimisation is necessary. In particular the use of protease inhibition throughout culture growth appeared to be key to the recovery of proteins. However, whether this system can be adapted to be efficient in multiple hosts is yet to be seen, with affinity for the MARE operators from exogenous proteins predicted.

## Appendices

### Appendix A. Nucleotide Sequences

#### Intergenic regions

MARE operator sequences shown in bold

##### *mmfL\_mmfR (194 bp) – L1*

GGCTGCCCTTCCTTCGTGTGTGCGGGCCCTGCGGGCCCATGCTGGTGCGACCCGGGTGCGCACGGAAACCCATTG**CATAATACCTT**  
**CCCGCAGGTATATTTCTCTCGGTCAGCTTACCGATCCCGGCTGTCTTGCGAGCGCGCAAGCCAGCCGGTGGTCCCGTACGAGGACA**  
 CACCGGGAGATGTAGCGCCCT

##### *mmfP\_mmyR (150 bp) – L2*

ATCCTGCCGCGCGGTAGCCGTGCTGCCTCCACTTTTGGCCGATGACTGGGACATCGTCCACGTGCGCCGACCGCCCCACT**AACA**  
**TACCTTCCCGAGGGTATGTTTCCCGGGCCGTTGGCTCACGACCTTGAGAGGACTCGGGCGTTG**

##### *mmyY\_mmyB (229 bp) – L3*

GGTGAACCTCCTTCGGCGAGTGGTTCGGATCGCTGGCGAGTATCGGCAGGGTCTGCGAAGGCTGCCAGAGCGAACTTCTGTAGGG  
 GCCTCAGACGTGGTGTCTCAGCACAGGGCCCCCGGAGTTCGGGTCCAAACCCAGGCCCCCGCTCTCGTAAGCCCCGGTT  
 TAACTCTCCGTTACGAGTCAT**AAAAAACCTTCGGGAAGGTTGAC**ACTGTGAGGCGCC

#### ErmE\* insert containing promoter

-35 and -10 sequences shown in bold

AGCTTGCATGCCGGTCGACTCTAGAGGATCCTACCAACCGGCACGATTGTGCCACAACAGCATCGGGTGCCACGTGTGGACCGC  
 GTCGGTTCAG**ATCCTC**CCCCGACCTCTCGCCAGCC**GTCAAG**ATCGACCGCGTGCACCTGCGATCGCGGATCAACCGCGACTAGCATC  
 GGGCGCAAGCCACCCTCGAACGGACACTCGCACTTAAGACGGATCTTTTCCGCTGCATAACCCTGCTTCGGGGTCATTATAGCGA  
 TTTTTCGGTATATCCATCCTTTTTCGCACGATATACAGGATTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTATCCAACGG  
 CGTCAGC

#### GbnB synthetic analogues

BamHI | RBS | ATG | TAT signal | 6 Histidine | MunI | GbnB analogue | NotI | NdeI |

CAC | **GGATCC** | **AAAGGAGG** | ATG | **ACGAAGCCGGTCTGCCCTCGGGGGTCTCCCGCGCGGGTTCCTCGGGGGTCCCTGGGCGT**  
**CGGGGGCGGGTCTGCTCGCCGC** | **CACCATCACCATCACCAC** | **CAATTG** | GbnB analogue | **GCGGCGGC** | **CATATG** | CAC |

#### Streptococcus mutans

CAC**GGATCC****AAAGGAGG**ATG**ACGAAGCCGGTCTGCCCTCGGGGGTCTCCCGCGCGGGTTCCTCGGGGGTCCCTGGGCGT****CGCG**  
**GGCGGGTCTGCTCGCCGCC****CACCATCACCATCACCAC****CAATTG**CCTGGCCCTGGAGAACCCTGATCCAGATCCGCAACCGCAACCC  
 GGACAAGCTGATCCTGATCTCCGACGAGAAGTCTTCTCCTGGAAGGAGTACACCAACCTGGTCATCAACAACCTGCGCAACACCA  
 CCTGACGTCCGCTCTGAACAAGACCGACCGGCCATCATCATCTCCGAGAACACCTGGAAGGTCTTACCATCTACTCTGCCTG  
 TCCACCCAGAAGATCCCGTACTCCGGCATCGACTACTCCATGGAGGACGACAAGAAGGTGCGCCGATCAACAAGTCCGGCGCCAA  
 CACCGTCTTCTACTCCAAGGACCAGAAGCCGTCCAGAACCTGCGCAACTCCCTGAAGGGCGTCTCCTTATCTCCTTGGACATCC  
 TGCACGACGACATCGAGGGCTCCGACCTGTCCGACTTCAACATCAAGAAGCACTCCGACTCCATCGTCTCCTTCGGCTTCACTCC  
 GGCACCACCGCCCTGCCGAAGTGCATCTACCGCGACTACTCCTTCGCCACCGAGCGCATGAAGGAGTGACCAAGCTGTACAACCT  
 CAACGCCACCGACGCTTCTCCTGGTCACCATGCGGTTCTACCACGCTCCGTCACCGGCTGGGTCAAGCTGACCTGAACAACGGCG  
 GCTCCGTCGTCCTGGGCGACTTCAACAACCCGATCGACCTGTCTCCAAGATCAAGCAGTACGACATCACCAACATGCTGATCACC  
 CCGCGGTCTGAACTCCTGAACTTCTGCTGTAACCAGAGGGCTTCATCAACTCCACCGTCCGCTTTCATCATGGTCCGGGGCAA  
 GAACTTCCCGCGAAGCTGAAGGAGGAGACCCAGAACCCTGTTCCGCTCCGCTCCTGCACGAGTACTACGGCTCCTCCGAGACCGGCA  
 TCAACGTCCTGGCAACTCCTCCGACATGATGCTGTACCCGTCCTCCTCCGGCCGCGTCATGAAGGGTCCGACGTCATCATCGTC  
 GACTCCGACAACCGCAAGATCCGGAACAACACATCGGCCGATCGCCATCTACTCCTACCAGAACGCCACCGGCTACATCAACCA  
 GCGCTGGAGAAGTTCAACTACCGCCAGAAGGAGTACATCCTGACCTCCGACTACGGCTACGTCACAACAACGAGGGCTACATCTTCG  
 TCGTCCAGCGCATCCTGAACCACGAGAACAACAAGATCATCAACGTCCTCCAGATCGAGAACCAGCTGCGCTGATCAAGGACATC  
 GACGACGTCGCCATCGTCCAGAAGAACAACCTGCTGCTGGTCAACATCAAGCTGAAGAAGATCTCCGAGATGAAGCGCTCCCTGGT  
 CAACGACCTGGTCTGCTGGATCTTAGAGAAGACCAAGATCCCGTACGACCTGAAGTACACCGACGAGATCCACTACTCCATGTCCG  
 CAAAGGTCAAAGTACACCGAGGTCATCAACTCCGAGGGCCGCG**GCGGCGGC****CATATG**CAC

**Salmonella enterica**

CACGGATCCAAAGGAGGATGACGAAGCCGGTCCGTCCTCCGCGGGTCTCCGCGCGGGTTCCTCGGGGGTCCCTGGCGCTCGCGGGCGGTCTGCTCGCCGCCACCATCACCACCAATTGATCACCCCTGCGCCGCTGAACGAGATCGCCATCACC CGCGGCAACGACATCTGCATCATCGACAAGGAGCGCCAGTACACCTGGTACGACATCATCCGCCGACCAGTCCCGCATCGTCTTCCCTGC GCCCGCCTTCAACCCGGAGCAGCTGCGCTCCGTCTGTACTCTGTTCAAGAACTCCGTGACCTGATCTGCTGGCTGGCCGCTTC GCCACCTTGGGATCCCGGCCAACGGCTTGGACTACTCCCTGCCGATCGAGACCTTGGCGGCTGCTGATCAAGATCAACCCGGG CCTGATGCTGGTCTCCTTCCCTGTACTCCCGGACGAGCTGAACAAGCTGCACGTCCGCACCATCACCATGCTGGCCGTCGACG CCCCACCGACCCGGTTCATCGGCTCCATCGGCGAGTTCCACCACCCGGAGCTGGAGTCCCTGCTGGCCACCCACATCCCGCCCG TTCCGCTCCGTCTCCTGACCTCCGGCACCTCCTCCGCCCGAAGATCGTCTTCCGCTACAACCTCTTCGACGCCCGCGCTTCGA CTGGTTACCCAGCGCTTCAACTTCAACCACACGACGGCTTCCGTGATCCTGCCGTGTACCACGCCGCCGGCAACGGCTGGG CCCCATGTTATGGGCTGGGCGCTCCCTGCACCTGGTTCGACAGGACGACGAGTCCGCCCTGATCCAGGCCCTGTCCCTGAAC TCCGTCGAAGGCCACCGTTCATGACCCCGAACCTGGTCTCCCGCTGACCAAGCTGGCTCCGAGACCGTCTGCACCACTACCTGGC CTGGTCTGGTTCGGCGGCTCCTACTTCCCGTCAAGTCCAAGATCGCCGCTACACCCACTGGGCCACTCTTCAACGAGTACT ACGGCTGCACCGAGACCGGCTAACGTCTGTCCGAGTCTCCGACATGTTTCGAGTGCCTGGGCTCCGTCGGCCCGCTTCGAC GGCAACAAGATCCGCATCCTGGACGAGGACAACGTCCCGCTGAAGGCCGCAACCGCGCCGTATCGCCGTCGCCCTCTACATGCT GATGGACGAGTACGGCGACGGCTCCCGCCGTTTCATCGAGATCGACTCCGAGCGTACTTCCCTGATGGCCGACTACGGCTACCTGG ACACAACGGCCGCTGTTCTGTGATGAACCGCAACTCCGAGATCAAGTGCAGCAGGACATCTACCACATCGAGGAGCACCTGCGC GCCCTGCCGTGCATCACCGACGTCGCCCTGATCCCGATCCGCCAGCAGAACAAGGACCACATCCGCTGCATCTTCTCCGCCAAGTA CATCAACGAGGACGAGCTCCTTTCATCATGGACGAGATCAAGAACAAGATCAACCACATCGGCGTCCCGACTTCCCGCCACA TGGTCCGACAAGATCCCGTACTCCCGTCCGGCAAGTCCGCTTCTCCGAGATCGTCCAGACCTGACCGCCGCCGGCGCCGCGCCATA TGCAC

**Appendix B. Amino Acid Sequences**> Mmfr [*Streptomyces coelicolor* A3(2)]

MTSAQQPTPFVAVSNVPRGPHPQQERSIKTRAQILEAASEIFASRGYRGASVKDVAERVGMTKGAVYFHFPSKESLAIIVVEEHYA RWPAAEMEEIRIQFTPLETVEEMLHRAAQAFRDDPVMQAGARLOSERAFIDAELPLPYVDWTHLLEVPLODAREAGQLRAGVDPAA AARSLVAFFFGMQHVSDNLHQRADIMERWQELRELMFFALRA

> MmyR [*Streptomyces coelicolor* A3(2)]

MKQARAMRTRDQVLDAAAEFALHGYAGTNLATVAVRTGMTKGALYGHFPSKKALADELVSQSTETWNTIGRSIAETACAPETALR ALVLAVSRQMKHDIRFRAALRLAADCTMPAGGAPDLLDRIRREMAAARDTQQQAPYSPLATQPPDVVVHLLLTVAAYGLSFAAER GAPGRSPATTDKVVWELLLTALQLEDISTCHN

> GbnB analogue [*Streptococcus mutans* synthetic sequence]

MTKPVVPSGVSRGFLGGSLGVAGAVLLAAHHHHHQLLALLENLIQIRNRNPKLILISDEKFSWKEYTNLVINLRNTLQSVL NKTDRAIIISENTWKVFTIYSCLSTQKIPYSGIDYSMEDDKVAAINKSGANTVFPYSKDQKPSQNLRLNSLKGVSFISLDLHDDIE GSDLSDFNKIKHSDSIVSFGFTSGTGLPKCIYRDYSFATERMKELTKLYNFNATDVFLVTMPFYHVSVNGWVKLTLLNNGGSSVVLG DFNNPIDLSKIKQYDITMLITPPVLKSLNFVNLQOGFINSTVRFIMVGGKNFPKKEETQNLFGSVLHEYYGSSSETGINVLN SDDMLLYPSSSGRVMKGSVDIIVDSNRKIPNNHIGRIAIYSYQATGNQPLEKFNYRQKEYILTSYGYVNNEGYIFVVQRILNH ENNKIINVFQIENRLRIKDIDDAIVQKNNLLLVNIKLLKISEMKRSLVNDLVCWILEKTKIPYDLKYTDEIHYSMSGVKYKYTEV INSEGR

> GbnB analogue [*Salmonella enterica* GbnB synthetic sequence]

MTKPVVPSGVSRGFLGGSLGVAGAVLLAAHHHHHQLITLRLNEIATRNDICIIDKERQYTWYDIIRRTESRIVFLRRFNP EQLRSVCYLSKNSVDLICWLAFAFATLGI PANGLDYSLP IETLRGLLIKINPGLMLVSFSLYSPDELNKLHVRTITMLAVDAPTDPV IGSIGEFHHPLESLATHIPAPFRSVSLTSGTSSAPKIVLRYNSFDARRFDWFTQRFNFTHHDGFLILPLYHAAGNGWARMFMG LGASLHLVDQDDESALIQALSLSNVKATVMTPNLVSRLTKLASETVLHHYLRWLVGGSYFPVKSKIAAYTHLGHIFNEYGCTET GVNVLSESSDMFECPSVGRAFDGNKIRILDEDNVPLKAGNRGRIAVASYMLMDEYGDGSRPFIEIDSERYFLMADYGLDDNGRL FLMNRNSEIKCEQDIYHIEHLRALPCITDVALIPRQNKDHIRCIFSAKYNEDDVSFIMDEIKNKINHIGVTDFTAHMVDKIP YSPSGKVRFSEIVQTLTAA

## Appendix C. Sequence Analysis of MmfR and Homologues

**Table C.i Percentage similarity between MmfR and MmyR and their homologues from *S. venezuelae*, *S. avermitilis* and *S. albus***

*Details on percentage query coverage shown in following table*

*Amino acid sequences used: same as Table 7.3*

|                   | MmfR | MmyR | SgnR | GbnR | SAV_<br>2270 | SAV_<br>2268 | SSHG_<br>01258 |
|-------------------|------|------|------|------|--------------|--------------|----------------|
| <b>MmfR</b>       | 100  | 56   | 71   | 48   | 64           | 54           | 56             |
| <b>MmyR</b>       | 56   | 100  | 47   | 54   | 67           | 54           | 53             |
| <b>SgnR</b>       | 71   | 47   | 100  | 49   | 65           | 49           | 60             |
| <b>GbnR</b>       | 48   | 54   | 49   | 100  | 50           | 47           | 75             |
| <b>SAV_2270</b>   | 64   | 67   | 65   | 50   | 100          | 49           | 48             |
| <b>SAV_2268</b>   | 54   | 54   | 49   | 47   | 49           | 100          | 44             |
| <b>SSHG_01258</b> | 56   | 53   | 60   | 75   | 48           | 44           | 100            |

## References

1. Gerber N, Lechevalier H. Geosmin, an Earthy-Smelling Substance Isolated from Actinomycetes. *Applied Microbiology*. 1965;13(6):4.
2. Flürdh K, Buttner MJ. Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat Rev Microbiol*. 2009;7(1):36-49.
3. Jakimowicz D, van Wezel GP. Cell division and DNA segregation in Streptomyces: how to build a septum in the middle of nowhere? *Mol Microbiol*. 2012;85(3):393-404.
4. Webster's Revised Unabridged Dictionary. 1913 ed: C. & G. Merriam Co; 1913. natural product.
5. Kinghorn AD, Chin YW, Swanson SM. Discovery of natural product anticancer agents from biodiverse organisms. *Curr Opin Drug Discov Devel*. 2009;12(2):189-96.
6. Bentley SD, Chater KF, Cerdeño-Tárraga AM, Challis GL, Thomson NR, James KD, et al. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature*. 2002;417(6885):141-7.
7. Seipke RF, Kaltenpoth M, Hutchings MI. Streptomyces as symbionts: an emerging and widespread theme? *FEMS Microbiol Rev*. 2012;36(4):862-76.
8. Procópio RE, Silva IR, Martins MK, Azevedo JL, Araújo JM. Antibiotics produced by Streptomyces. *Braz J Infect Dis*. 2012;16(5):466-71.
9. Distler J, Ebert A, Mansouri K, Pissowotzki K, Stockmann M, Piepersberg W. Gene cluster for streptomycin biosynthesis in Streptomyces griseus: nucleotide sequence of three genes and analysis of transcriptional activity. *Nucleic Acids Res*. 1987;15(19):8041-56.
10. Dulmage HT. The production of neomycin by *Streptomyces fradiae* in synthetic media. *Appl Microbiol*. 1953;1(2):103-6.
11. Akagawa H, Okanishi M, Umezawa H. A plasmid involved in chloramphenicol production in Streptomyces venezuelae: evidence from genetic mapping. *J Gen Microbiol*. 1975;90(2):336-46.
12. Zipperer A, Konnerth MC, Laux C, Berscheid A, Janek D, Weidenmaier C, et al. Human commensals producing a novel antibiotic impair pathogen colonization. *Nature*. 2016;535(7613):511-6.
13. Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, et al. A new antibiotic kills pathogens without detectable resistance. *Nature*. 2015;517(7535):455-9.
14. Haneishi T, Terahara A, Hamano K, Arai M. New antibiotics, methylenomycins A and B. 3. Chemical modifications of methylenomycin A and structure-activity correlations in methylenomycins. *J Antibiot (Tokyo)*. 1974;27(6):400-7.
15. Ingroff E, Victoria A. *Medical Mycology in the United States: A Historical Analysis (1894–1996)*. Netherlands DS, editor 2013.
16. Fisher MH, Mrozik H. The chemistry and pharmacology of avermectins. *Annu Rev Pharmacol Toxicol*. 1992;32:537-53.
17. Demain AL, Vaishnav P. Natural products for cancer chemotherapy. *Microb Biotechnol*. 2011;4(6):687-99.
18. Pahl A, Keller U. FK-506-binding proteins from streptomycetes producing immunosuppressive macrolactones of the FK-506 type. *J Bacteriol*. 1992;174(18):5888-94.
19. Roberts T. *Metabolic Pathways of Agrochemicals: Insecticides and fungicides: Part II*. Hutson D, Roberts T, editors. Cambridge: The Royal Society of Chemistry; 1999. 975 p.
20. Linares JF, Gustafsson I, Baquero F, Martinez JL. Antibiotics as intermicrobial signaling agents instead of weapons. *Proc Natl Acad Sci U S A*. 2006;103(51):19484-9.
21. Ventola CL. The antibiotic resistance crisis: part I: causes and threats. *P T*. 2015;40(4):277-83.
22. Fleming A. Penicillin, Nobel Lecture. In: Fleming A, editor. *The Nobel Prize in Physiology or Medicine 1945*. Stockholm: Nobel Prize; 1945. p. 11.
23. Koehn FE, Carter GT. The evolving role of natural products in drug discovery. *Nat Rev Drug Discov*. 2005;4(3):206-20.
24. Koehn FE, Carter GT. Rediscovering natural products as a source of new drugs. *Discov Med*. 2005;5(26):159-64.
25. Lewis K. Platforms for antibiotic discovery. *Nat Rev Drug Discov*. 2013;12(5):371-87.
26. Challis GL. Genome mining for novel natural product discovery. *J Med Chem*. 2008;51(9):2618-28.
27. Harrison J, Studholme DJ. Recently published Streptomyces genome sequences. *Microb Biotechnol*. 2014;7(5):373-80.

28. Gomez-Escribano JP, Bibb MJ. Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. *J Ind Microbiol Biotechnol*. 2014;41(2):425-31.
29. Challis GL. Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology*. 2008;154(Pt 6):1555-69.
30. Lautru S, Deeth RJ, Bailey LM, Challis GL. Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nat Chem Biol*. 2005;1(5):265-9.
31. Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2009;37(Database issue):D5-15.
32. Ramos JL, Martinez-Bueno M, Molina-Henares AJ, Teran W, Watanabe K, Zhang XD, et al. The TetR family of transcriptional repressors. *Microbiology and Molecular Biology Reviews*. 2005;69(2):326-+.
33. Collins English Dictionary – Complete and Unabridged. 12 ed: HarperCollins; 2014. hormone.
34. Horinouchi S. A microbial hormone, A-factor, as a master switch for morphological differentiation and secondary metabolism in *Streptomyces griseus*. *Front Biosci*. 2002;7:d2045-57.
35. Niu G, Chater KF, Tian Y, Zhang J, Tan H. Specialised metabolites regulating antibiotic biosynthesis in *Streptomyces* spp. *FEMS Microbiol Rev*. 2016;40(4):554-73.
36. Miller MB, Bassler BL. Quorum sensing in bacteria. *Annu Rev Microbiol*. 2001;55:165-99.
37. Du YL, Shen XL, Yu P, Bai LQ, Li YQ. Gamma-butyrolactone regulatory system of *Streptomyces chattanoogensis* links nutrient utilization, metabolism, and development. *Appl Environ Microbiol*. 2011;77(23):8415-26.
38. Takano E, Chakraborty R, Nihira T, Yamada Y, Bibb MJ. A complex role for the gamma-butyrolactone SCBI in regulating antibiotic production in *Streptomyces coelicolor* A3(2). *Mol Microbiol*. 2001;41(5):1015-28.
39. Lee YJ, Kitani S, Nihira T. Null mutation analysis of an afsA-family gene, barX, that is involved in biosynthesis of the {gamma}-butyrolactone autoregulator in *Streptomyces virginiae*. *Microbiology*. 2010;156(Pt 1):206-10.
40. Kinoshita H, Ipposhi H, Okamoto S, Nakano H, Nihira T, Yamada Y. Butyrolactone autoregulator receptor protein (BarA) as a transcriptional regulator in *Streptomyces virginiae*. *J Bacteriol*. 1997;179(22):6986-93.
41. Kitani S, Yamada Y, Nihira T. Gene replacement analysis of the butyrolactone autoregulator receptor (FarA) reveals that FarA acts as a Novel regulator in secondary metabolism of *Streptomyces lavendulae* FRI-5. *J Bacteriol*. 2001;183(14):4357-63.
42. Kitani S, Doi M, Shimizu T, Maeda A, Nihira T. Control of secondary metabolism by farX, which is involved in the gamma-butyrolactone biosynthesis of *Streptomyces lavendulae* FRI-5. *Arch Microbiol*. 2010;192(3):211-20.
43. Cuthbertson L, Nodwell JR. The TetR Family of Regulators. *Microbiology and Molecular Biology Reviews*. 2013;77(3):440-75.
44. Orth P, Schnappinger D, Hillen W, Saenger W, Hinrichs W. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat Struct Biol*. 2000;7(3):215-9.
45. Le TBK, Schumacher MA, Lawson DM, Brennan RG, Buttner MJ. The crystal structure of the TetR family transcriptional repressor SimR bound to DNA and the role of a flexible N-terminal extension in minor groove binding *Nucleic Acids Research*. 2011;39(21):9433-47.
46. Dover LG, Corsino PE, Daniels IR, Cocklin SL, Tatituri V, Besra GS, et al. Crystal structure of the TetR/CamR family repressor *Mycobacterium tuberculosis* EthR implicated in ethionamide resistance. *J Mol Biol*. 2004;340(5):1095-105.
47. Saenger W, Orth P, Kisker C, Hillen W, Hinrichs W. The Tetracycline Repressor-A Paradigm for a Biological Switch. *Angew Chem Int Ed Engl*. 2000;39(12):2042-52.
48. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*. 2011;7.
49. Xu G, Wang J, Wang L, Tian X, Yang H, Fan K, et al. "Pseudo" gamma-butyrolactone receptors respond to antibiotic signals to coordinate antibiotic biosynthesis. *J Biol Chem*. 2010;285(35):27440-8.
50. Li X, Wang J, Li S, Ji J, Wang W, Yang K. ScbR- and ScbR2-mediated signal transduction networks coordinate complex physiological responses in *Streptomyces coelicolor*. *Sci Rep*. 2015;5:14831.

51. Hopwood DA, Kieser TB, M. J. Buttner, M. J. Chater, K. F. Practical *Streptomyces* Genetics. Norwich, UK: John Innes Foundation; 2000.
52. Wang W, Ji J, Li X, Wang J, Li S, Pan G, et al. Angucyclines as signals modulate the behaviors of *Streptomyces coelicolor*. *Proc Natl Acad Sci U S A*. 2014;111(15):5688-93.
53. Kitani S, Miyamoto KT, Takamatsu S, Herawati E, Iguchi H, Nishitomi K, et al. Avenolide, a *Streptomyces* hormone controlling antibiotic production in *Streptomyces avermitilis*. *Proc Natl Acad Sci U S A*. 2011;108(39):16410-5.
54. Corre C, Song L, O'Rourke S, Chater KF, Challis GL. 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by *Streptomyces coelicolor* genome mining. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(45):17510-5.
55. Malet N. Structure-activity relationship and biosynthesis of the methylenomycin furans Coventry: University of Warwick; 2012.
56. Muller R. Eine Diptheridee und eine Streptothrix mit gleichem blauen Farbstoff, sowie Untersuchungen über Streptothrix-Arten im Allgemeinen. *Zentralblatt für Bakteriologie und Parasitologie*. 1908;46:195-212.
57. Brockmann H, Hieronimus E. Über Actinomycetenfarbstoffe, V. Mitteil.: Zur Konstitution des Actinorhodins, III. *Mitteil Chem Ber*. 1955;88:1379-90.
58. Brockmann H, Pini H, Plotho O. Actinorhodin. *Chem Ber*. 1950;83:161.
59. Pridham TG. New Names and New Combinations in the Order Actinomycetales Buchanan 1917 - Issue 1424 of Technical bulletin (United States. Dept. of Agriculture). University of Virginia: U.S. Agricultural Research Service, 1970; 1970. 55 p.
60. Bentley SD, Brown S, Murphy LD, Harris DE, Quail MA, Parkhill J, 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*. 2004;51(6):1615-28.
61. Haug I, Weissenborn A, Brolle D, Bentley S, Kieser T, Altenbuchner J. *Streptomyces coelicolor* A3(2) plasmid SCP2\*: deductions from the complete sequence. *Microbiology*. 2003;149(Pt 2):505-13.
62. Liu G, Chater KF, Chandra G, Niu G, Tan H. Molecular regulation of antibiotic biosynthesis in *streptomyces*. *Microbiol Mol Biol Rev*. 2013;77(1):112-43.
63. Hayes A, Hobbs G, Smith CP, Oliver GP, Butler PR. Environmental signals triggering methylenomycin production by *Streptomyces coelicolor* A3(2). *Journal of Bacteriology*. 1997;179(17):5511-5.
64. O'Hara CM, Brenner FW, Miller JM. Classification, identification, and clinical significance of *Proteus*, *Providencia*, and *Morganella*. *Clin Microbiol Rev*. 2000;13(4):534-46.
65. Chater KF, Bruton CJ. Resistance, regulatory and production genes for the antibiotic methylenomycin are clustered. *EMBO Journal*. 1985;4(7):1893-2.
66. Wright LF, Hopwood DA. Identification of antibiotic determined by SCP1 plasmid of *Streptomyces coelicolor* A3(2). *Journal of General Microbiology*. 1976;95(JUL):96-106.
67. Vivian A. Genetic Control of Fertility in *Streptomyces coelicolor* A3(2) Plasmid Involvement in the Interconversion of U F and I F Strains Genetic Control of Fertility in *Streptomyces coelicolor* A3(2) Plasmid Involvement in the Interconversion of U F and I F Strains. *Journal of General Microbiology*. 1971;69:353-64.
68. Kirby R, Hopwood DA. Genetic Determination of Methylenomycin Synthesis by the SCP1 Plasmid of *Streptomyces coelicolor* A3(2). *Journal of General Microbiology*. 1977;98:239-52.
69. Kinashi H. Giant linear plasmids in *Streptomyces*: a treasure trove of antibiotic biosynthetic clusters. *J Antibiot (Tokyo)*. 2011;64(1):19-25.
70. Hobbs G, Obanye AIC, Petty J, Mason JC, Barratt E, Gardner DCJ, et al. An integrated approach to studying regulation of production of the antibiotic methylenomycin by *Streptomyces coelicolor* A3(2). *Journal of Bacteriology*. 1992;174(5):1487-94.
71. O'Rourke S, Wietzorrek A, Fowler K, Corre C, Challis GL, Chater KF. Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of methylenomycin production in *Streptomyces coelicolor*. *Molecular Microbiology*. 2009;71(3):763-78.
72. Biotech G. SnapGene Viewer. In: Glick B, editor. 3.0.3 ed. Chicago, IL 2004-2016.
73. Rea D, Fulop V, Zhou S, Styles K, Corre C. Structure of MmfR. Personal communication. School of Life Sciences: University of Warwick; 2015.
74. Corre C, Haynes SW, Malet N, Song L, Challis GL. A butenolide intermediate in methylenomycin furan biosynthesis is implied by incorporation of stereospecifically 13C-labelled glycerols. *Chem Commun (Camb)*. 2010;46(23):4079-81.



75. Sidda JD, Song L, Poon V, Al-Bassam M, Lazos O, Buttner MJ, et al. Discovery of a family of  $\gamma$ -aminobutyrate ureas *via* rational derepression of a silent bacterial gene cluster. *Chemical Science*. 2014;5:4.
76. Choi SU, Lee CK, Hwang YI, Kinoshita H, Nihira T. Cloning and functional analysis by gene disruption of a gene encoding a gamma-butyrolactone autoregulator receptor from *Kitasatospora setae*. *J Bacteriol*. 2004;186(11):3423-30.
77. Gasteiger E. HC, Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JME, editor. *The Proteomics Protocols Handbook*: Humana Press; 2005. p. 571-607.
78. Fisher SH, Bruton CJ, Chater KF. The glucose kinase gene of *Streptomyces coelicolor* and its use in selecting spontaneous deletions for desired regions of the genome. *Mol Gen Genet*. 1987;206(1):35-44.
79. Bowyer JE, de los Santos EL, Styles KM, Corre C, Bates DG. Modeling the Architecture of the Regulatory System Controlling Methylenomycin Production in *Streptomyces coelicolor*. *Asia Pacific Bioinformatics Conference 2017; Shenzhen, China: APBC2017*; 2016. p. 17.
80. Pullan ST, Chandra G, Bibb MJ, Merrick M. Genome-wide analysis of the role of GlnR in *Streptomyces venezuelae* provides new insights into global nitrogen regulation in actinomycetes. *BMC Genomics*. 2011;12:175.
81. Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, et al. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol*. 2003;21(5):526-31.
82. Wu H, Qu S, Lu C, Zheng H, Zhou X, Bai L, et al. Genomic and transcriptomic insights into the thermo-regulated biosynthesis of validamycin in *Streptomyces hygroscopicus* 5008. *BMC Genomics*. 2012;13:337.
83. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*. 2009;37:W202-W8.
84. Poon V. *Analysis and exploitation of AHFCA-dependent signalling systems in Streptomyces bacteria*: University of Warwick; 2015.
85. Xu Q, van Wezel GP, Chiu HJ, Jaroszewski L, Klock HE, Knuth MW, 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*. 2012;7(7):e41359.
86. Durante-Rodríguez G, Valderrama JA, Mancheño JM, Rivas G, Alfonso C, Arias-Palomo E, et al. Biochemical characterization of the transcriptional regulator BzdR from *Azoarcus* sp. CIB. *J Biol Chem*. 2010;285(46):35694-705.
87. Chater KF, 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*. 2008;46(1):1-11.
88. Chater KF. *Streptomyces inside-out: a new perspective on the bacteria that provide us with antibiotics*. *Philos Trans R Soc Lond B Biol Sci*. 2006;361(1469):761-8.
89. Bibb MJ, Molle V, Buttner MJ. sigma(BldN), an extracytoplasmic function RNA polymerase sigma factor required for aerial mycelium formation in *Streptomyces coelicolor* A3(2). *J Bacteriol*. 2000;182(16):4606-16.
90. Paget MSB, Hong H-J, Bibb MJ, Buttner MJ. The ECF sigma factors of *Streptomyces coelicolor* A3(2). In: Hodgson DA, Thomas CM, editors. *SGM*: Cambridge University Press; 2002. p. 105-22.
91. Nudler E, Mironov AS. The riboswitch control of bacterial metabolism. *Trends Biochem Sci*. 2004;29(1):11-7.
92. Craney A, Hohenauer T, Xu Y, Navani NK, Li Y, Nodwell J. A synthetic luxCDABE gene cluster optimized for expression in high-GC bacteria. *Nucleic Acids Research*. 2007;35(6):e46-Article No.: e.
93. Zaburannyi N, Rabyk M, Ostash B, Fedorenko V, Luzhetskyy A. Insights into naturally minimised *Streptomyces albus* J1074 genome. *BMC Genomics*. 2014;15:97.
94. Jefferson RA, Kavanagh TA, Bevan MW. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J*. 1987;6(13):3901-7.
95. Shindo T, Takahashi T, Nihira T, Yamada Y, Kato K, Shinmyo A. *Streptomyces*-derived induction system for gene expression in cultured plant cells. *J Biosci Bioeng*. 2006;102(6):552-9.
96. Myronovskiy M, Welle E, Fedorenko V, Luzhetskyy A. beta-Glucuronidase as a Sensitive and Versatile Reporter in Actinomycetes. *Applied and Environmental Microbiology*. 2011;77(15):5370-83.
97. Danilov VS, Zavilgelsky GB, Zarubina AP, Mazhul MM. The role of *luxCDE* genes in bioluminescence of bacteria. *Vestnik Moskovskogo Universiteta Seriya 16 Biologiya*. 2008(2):11-5.

98. Gupta R, K., Patterson S, S., Ripp S, Simpson M, L., Saylor GS. Expression of the *Photobacterium luminescens* lux genes (luxA, B, C, D, and E) in *Saccharomyces cerevisiae*. *Federation of European Microbiological Societies Yeast Research*. 2003;4:305-13.
99. Lin LY-C, Meighen EA. *Photobiological Sciences Online* [Internet]. KC S, editor: American Society for Photobiology. 2009. [cited 2016].
100. Hansen LH, Knudsen S, Sørensen SJ. The effect of the lacY gene on the induction of IPTG inducible promoters, studied in *Escherichia coli* and *Pseudomonas fluorescens*. *Curr Microbiol*. 1998;36(6):341-7.
101. Anné J, Maldonado B, Van Impe J, Van Mellaert L, Bernaerts K. Recombinant protein production and streptomycetes. *J Biotechnol*. 2012;158(4):159-67.
102. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol*. 2014;5:172.
103. Hildebrand F, Meyer A, Eyre-Walker A. Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet*. 2010;6(9):e1001107.
104. Dalton AC, Barton WA. Over-expression of secreted proteins from mammalian cell lines. *Protein Sci*. 2014;23(5):517-25.
105. Vrancken K, Anné J. Secretory production of recombinant proteins by *Streptomyces*. *Future Microbiol*. 2009;4(2):181-8.
106. Pozidis C, Lammertyn E, Politou AS, Anné J, Tsiftoglou AS, Sianidis G, et al. Protein secretion biotechnology using *Streptomyces lividans*: large-scale production of functional trimeric tumor necrosis factor alpha. *Biotechnol Bioeng*. 2001;72(6):611-9.
107. Noda S, Matsumoto T, Tanaka T, Kondo A. Secretory production of tetrameric native full-length streptavidin with thermostability using *Streptomyces lividans* as a host. *Microb Cell Fact*. 2015;14:5.
108. Wilkinson CJ, Hughes-Thomas ZA, Martin CJ, Böhm I, Mironenko T, Deacon M, et al. Increasing the efficiency of heterologous promoters in actinomycetes. *J Mol Microbiol Biotechnol*. 2002;4(4):417-26.
109. Agilent. QuikChange Lightning Site-Directed Mutagenesis Kit - Instruction Manual 2013 9th December 2014. Available from: <http://www.chem.agilent.com/library/usermanuals/Public/210518.pdf>.
110. Invitrogen. Champion™ pET Directional TOPO® Expression Kit – Five minute, directional TOPO® Cloning of blunt-end PCR products into vectors for high-level, inducible expression of *E. coli*. Champion™ pET Directional TOPO® Expression Kit: Life Technologies Corporation; 2010 [cited 2013 11th December 2013]. MAN0000214:[
111. MPBiomedicals. Instruction Manual: FastDNA SPIN kit for soil 2007 [cited 2013 8th February 2013]. 6560-200-07DEC:[Available from: <http://images.www.mpbio.com/docs/fastprep/FastDNASpinKitforsoil.pdf>.
112. ThermoScientific. GeneJET Plasmid Miniprep Kit: Thermo Fisher Scientific, Inc.; 2014 [cited 2015 25.9.15]. 11:[Product Information]. Available from: [https://tools.thermofisher.com/content/sfs/manuals/MAN0012655\\_GeneJET\\_Plasmid\\_Miniprep\\_UG.pdf](https://tools.thermofisher.com/content/sfs/manuals/MAN0012655_GeneJET_Plasmid_Miniprep_UG.pdf).
113. ThermoScientific. GeneJET Gel Extraction Kit Carlsbad, California: Thermo Fisher Scientific, Inc.; 2015 [cited 2015 25.9.15]. 12:[Product Information]. Available from: [https://tools.thermofisher.com/content/sfs/manuals/MAN0012661\\_GeneJET\\_Gel\\_Extraction\\_UG.pdf](https://tools.thermofisher.com/content/sfs/manuals/MAN0012661_GeneJET_Gel_Extraction_UG.pdf).
114. ThermoScientific. T4 DNA Ligase: Thermo Fisher Scientific Inc.; 2014 [cited 2015 25.9.15]. 10:[Product Information]. Available from: [https://tools.thermofisher.com/content/sfs/manuals/MAN0011990\\_T4\\_DNA\\_Ligase\\_5\\_Weiss\\_200\\_Weiss\\_U\\_UG.pdf](https://tools.thermofisher.com/content/sfs/manuals/MAN0011990_T4_DNA_Ligase_5_Weiss_200_Weiss_U_UG.pdf).
115. Ivashchenko R, Chernov M, Dudchak V, Salamatov V, Shayna A. Prism 6. In: Radushev D, editor. 6.0c for Mac OS X ed: GraphPad Software; 2013.
116. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering*. 1995;8:127-34.
117. Schrödinger. The PyMOL Molecular Graphics System. In: DeLano WL, editor. 1.5.0.4 ed: Schrödinger, DeLano Scientific LLC; 2006.
118. Grosdidier A, Zoete V, Michielin O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research*. 2011;39:W270-W7.

119. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. 2004;25(13):1605-12.
120. Boccard F, Smokvina T, Pernodet JL, Friedmann A, Guérineau M. The integrated conjugative plasmid pSAM2 of *Streptomyces ambofaciens* is related to temperate bacteriophages. *EMBO J*. 1989;8(3):973-80.
121. Invitrogen. Champion™ pET Directional TOPO® Expression Kit: Life Technologies Corporation; 2010 [cited 2013 11th December 2013]. MAN0000214:[
122. Paget MS, Chamberlin L, Atrih A, Foster SJ, Buttner MJ. Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J Bacteriol*. 1999;181(1):204-11.
123. Gomez-Escribano JP, Bibb MJ. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb Biotechnol*. 2011;4(2):207-15.
124. Chater KF, Wilde LC. Restriction of a bacteriophage of *Streptomyces albus* G involving endonuclease Sall. *J Bacteriol*. 1976;128(2):644-50.
125. MacNeil DJ, Gewain KM, Ruby CL, Dezeny G, Gibbons PH, MacNeil T. Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene*. 1992;111(1):61-8.
126. Hanahan D. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol*. 1983;166(4):557-80.
127. Novoradovsky A, Zhang V, Ghosh M, Hogrefe H, Sorge JA, Gaasterland T, editors. *Computational Principles of Primer Design for Site Directed Mutagenesis*. NSTI Nanotechnology Conference and Trade Show; 2005; Anaheim.
128. Widdick DA, Dilks K, Chandra G, Bottrill A, Naldrett M, Pohlschröder M, et al. The twin-arginine translocation pathway is a major route of protein export in *Streptomyces coelicolor*. *Proc Natl Acad Sci U S A*. 2006;103(47):17927-32.
129. Francis K, P., Joh D, Bellinger-Kawahara C, Hawkinson M, J., Purchio T, F., Contag P, R. Monitoring Bioluminescent *Staphylococcus aureus* Infections in Living Mice Using a Novel *luxABCDE* Construct. *Infection and Immunity*. 2000;68(6):3594–600.
130. Zhang L, Zhu B, Dai R, Zhao G, Ding X. Control of directionality in *Streptomyces* phage  $\phi$  BT1 integrase-mediated site-specific recombination. *PLoS One*. 2013;8(11):e80434.
131. Bibb MJ, Janssen GR, Ward JM. Cloning and analysis of the promoter region of the erythromycin resistance gene (*ermE*) of *Streptomyces erythraeus*. *Gene*. 1985;38(1-3):215-26.
132. Sezonov G, Duchêne AM, Friedmann A, Guérineau M, Pernodet JL. Replicase, excisionase, and integrase genes of the *Streptomyces* element pSAM2 constitute an operon positively regulated by the *pra* gene. *J Bacteriol*. 1998;180(12):3056-61.
133. Kelly JR, Rubin AJ, Davis JH, Ajo-Franklin CM, Cumbers J, Czar MJ, et al. Measuring the activity of BioBrick promoters using an in vivo reference standard. *J Biol Eng*. 2009;3:4.
134. Bai C, Zhang Y, Zhao X, Hu Y, Xiang S, Miao J, 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*. 2015;112(39):12181-6.
135. Wang W, Li X, Wang J, Xiang S, Feng X, Yang K. An Engineered Strong Promoter for *Streptomyces*. *Applied and Environmental Microbiology*. 2013;79(14):4484–92.
136. Mintun MA, Raichle ME, Kilbourn MR, Wooten GF, Welch MJ. A quantitative model for the in vivo assessment of drug binding sites with positron emission tomography. *Ann Neurol*. 1984;15(3):217-27.
137. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*. 2000;302(1):205-17.
138. Wang J, Wang W, Wang L, Zhang G, Fan K, Tan H, et al. A novel role of 'pseudo'  $\gamma$ -butyrolactone receptors in controlling  $\gamma$ -butyrolactone biosynthesis in *Streptomyces*. *Mol Microbiol*. 2011;82(1):236-50.
139. Wang GY, Graziani E, Waters B, Pan W, Li X, McDermott J, et al. Novel natural products from soil DNA libraries in a streptomycete host. *Org Lett*. 2000;2(16):2401-4.
140. Vrancken K, Van Mellaert L, Anné J. Cloning and expression vectors for a Gram-positive host, *Streptomyces lividans*. *Methods Mol Biol*. 2010;668:97-107.
141. Seipke RF. Strain-level diversity of secondary metabolism in *Streptomyces albus*. *PLoS One*. 2015;10(1):e0116457.
142. Dereeper A, Audic S, Claverie J-M, Blanc G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *Bmc Evolutionary Biology*. 2010;10.

143. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*. 2010;38:W695-W9.
144. Boc A, Diallo AB, Makarenkov V. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res*. 2012;40(Web Server issue):W573-9.
145. Mould RM, Robinson C. A proton gradient is required for the transport of two luminal oxygen-evolving proteins across the thylakoid membrane. *J Biol Chem*. 1991;266(19):12189-93.
146. Palmer T, Berks BC. The twin-arginine translocation (Tat) protein export pathway. *Nat Rev Microbiol*. 2012;10(7):483-96.
147. Creighton AM, Hulford A, Mant A, Robinson D, Robinson C. A monomeric, tightly folded stromal intermediate on the delta pH-dependent thylakoidal protein transport pathway. *J Biol Chem*. 1995;270(4):1663-9.
148. Li H, Jacques PE, Ghinet MG, Brzezinski R, Morosoli R. Determining the functionality of putative Tat-dependent signal peptides in *Streptomyces coelicolor* A3(2) by using two different reporter proteins. *Microbiology*. 2005;151(Pt 7):2189-98.
149. Lee PA, Tullman-Ercek D, Georgiou G. The bacterial twin-arginine translocation pathway. *Annu Rev Microbiol*. 2006;60:373-95.
150. Gauthier C, Li H, Morosoli R. Increase in xylanase production by *Streptomyces lividans* through simultaneous use of the Sec- and Tat-dependent protein export systems. *Appl Environ Microbiol*. 2005;71(6):3085-92.
151. Anné J, Van Mellaert L. *Streptomyces lividans* as host for heterologous protein production. *FEMS Microbiol Lett*. 1993;114(2):121-8.
152. Chen W, Qin Z. Development of a gene cloning system in a fast-growing and moderately thermophilic *Streptomyces* species and heterologous expression of *Streptomyces* antibiotic biosynthetic gene clusters. *BMC Microbiol*. 2011;11:243.
153. Sezonov G, Joseleau-Petit D, D'Ari R. *Escherichia coli* physiology in Luria-Bertani broth. *J Bacteriol*. 2007;189(23):8746-9.
154. Nikaido H. Small things considered [Internet]. Schaechter M, editor. San Diego, California: American Society For Microbiology. 2009. [cited 2016]. Available from: <http://schaechter.asmblog.org/schaechter/2009/11/the-limitations-of-lb-medium.html>.
155. Sidda JD, Corre C. Overproduction of *Streptomyces coelicolor*  $\gamma$ -butyrolactone signalling molecules in the superhost strain *Streptomyces* M1152. 2015.
156. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science*. 2010;327(5962):167-70.
157. Netzker T, Schroeckh V, Gregory MA, Flak M, Krespach MK, Leadlay PF, et al. An Efficient Method To Generate Gene Deletion Mutants of the Rapamycin-Producing Bacterium *Streptomyces iranensis* HM 35. *Appl Environ Microbiol*. 2016;82(12):3481-92.
158. Gust B, Kieser T, Chater K. REDIRECT technology: PCR-targeting system in *Streptomyces coelicolor* In: Centre JI, editor. 1 ed. Norwich: John Innes Centre; 2002.
159. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A*. 2000;97(12):6640-5.
160. Urban A, Eckermann S, Fast B, Metzger S, Gehling M, Ziegelbauer K, et al. Novel whole-cell antibiotic biosensors for compound discovery. *Appl Environ Microbiol*. 2007;73(20):6436-43.