



The Use of Heterologous Expression Platforms for the Production of Novel Actinomycete Antibiotics

Abigail Alford

John Innes Centre

Department of Molecular Microbiology

A thesis submitted to the University of East Anglia for the
degree of Doctor of Philosophy

September 2021

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

Antimicrobial resistance (AMR) poses a considerable threat to modern living. To combat this threat methods for the discovery and diversification of antibiotics are of great importance. This project focuses on the use of heterologous expression for the rapid production and structural diversification of natural products (NPs). Firstly, I exemplify a synthetic heterologous platform for class II type B lantipeptide production. Further to this, I expand on the use of *Saccharopolyspora erythraea* ISOM as a heterologous expression strain for NPs.

Lantipeptides are Ribosomally-synthesised and Post-translationally modified Peptides (RiPPs) with clinical utility as antibacterials and treatments for cystic fibrosis. Rapid peptide diversification was achieved by Gibson assembly using a library of oligonucleotides encoding lantipeptide core peptides and an engineered precursor peptide gene (*lanA*) housed within a synthetic cassette. The production of 83 novel lantipeptides was achieved, with 61 showing potential as antibacterials. Seven lantipeptides were purified and their structure and bioactivity characterised.

Production of actinomycetes cryptic NPs in *S. erythraea* ISOM was mostly unsuccessful. Only production of fasamycin C was achieved using a de-repressed formicamycin biosynthetic gene cluster (BGC). The de-repressed formicamycin BGC was used to compare NP expression profiles within the parental strain and within different heterologous hosts. The heterologous hosts had significantly different expression profiles of the NPs compared to the parental strain. A further six new glycosylated fasamycin-derived compounds were produced from heterologous host *S. erythraea* ISOM, exemplifying how heterologous production may be used for the diversification of NPs.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Acknowledgements

I would like to thank my primary supervisor Prof. Barrie Wilkinson for his support and guidance during my PhD. The start of my PhD was difficult, with my primary project coming to an end after a year with no results, but his advice and expertise allowed me to swiftly start new projects and to have faith in my ability to write this thesis within the difficult times of the COVID-19 pandemic. I would also like to thank Eleni Vikeli whose own PhD led to the creation of the kyamicin lantipeptide expression platform, upon which much of my work is based.

My thesis would not be possible without the support of colleagues past and present within the Molecular Microbiology department. I would like to give special thanks to Dr Silke Alt, who assisted me when I first began my PhD, Dr Edward Hems for his help with chemistry, Dr Martin Rejzek for his teachings and advice on HPLC purification and his work in optimising Tandem MS for lantipeptides, and Dr Sergey Nepogodiev, for his assistance with lantipeptide NMR analysis.

Special thanks also to Dr David Widdick for all his encouragement and humour, and Dr Carlo de-Oliveira-Martins and Dr Lionel Hill for their help with LCMS analysis of my samples, without which this thesis would not be possible.

I am grateful for all the friends I have made within and outside the department, who have made living in Norwich these last 4 years such an enjoyable experience, and with who I commiserate greatly with the pains of PhD life.

Finally, I would like to thank my mum Fiona Scriven for trying her best to understand my project and assist with correcting my many spelling errors. I would like to thank my late dad Peter Alford, for always believing in me, sparking my interest in science as a young child and never once letting me think I could not accomplish anything I set out to do. Lastly, thank you to my late Grandma and the rest of my family, both living and past, for loving me all these years.

Author's Declaration

The research described in this thesis was conducted at the John Innes Centre between October 2017 and September 2021. All data described is original and was obtained by the author, except where specific acknowledgement has been made.

Abbreviations

ABC	ATP-Binding Cassette
AMR	Antimicrobial Resistance
ATP	Adenosine Triphosphate
BGC	Biosynthetic Gene Cluster
BLAST	Basic Local Alignment Search Tool
BPC	Base Peak Chromatogram
CF	Cystic Fibrosis
CLF	Chain Length Factor
CoA	Coenzyme A
CV	Column Volume
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
EIC	Extracted Ion Chromatograms
ELSD	Evaporative Light Scattering Detection
HIV	Human Immunodeficiency Virus
HPLC	High Performance Liquid Chromatography
LCMS	Liquid Chromatography Mass Spectrometry
MALDI	Matrix-Assisted Laser Desorption Ionization
MIC	Minimum Inhibitory Concentration
MRSA	Methicillin-Resistant <i>Staphylococcus aureus</i>
MS	Mass Spectrometry
MSSA	Methicillin-Sensitive <i>Staphylococcus aureus</i>
NMR	Nuclear Magnetic Resonance
NCBI	National Centre of Biotechnology Information
NP	Natural Product
NRP	Non-Ribosomal Peptide
NRPS	Non-Ribosomal Peptide Synthetase
PAC	Phage-1 Artificial Chromosome
PCR	Polymerase Chain Reaction
PE	Phosphatidylethanolamine

PKS	Polyketide Synthase
PMF	Proton Motive Force
PPI	Proton Pump Inhibitor
PPM	Parts Per Million
PTM	Polycyclic Tetramate Macrolactam
RBS	Ribosome Binding Site
RiPP	Ribosomally-synthesised and Post-translationally modified Peptide
SARP	<i>Streptomyces</i> Antibiotic Regulatory Proteins
SNP	Single Nucleotide Polymorphism
TFD	Transcription Factor Decoy
VRE	Vancomycin-Resistant <i>Enterococci</i>
VRSA	Vancomycin-Resistant <i>Staphylococcus aureus</i>
WT	Wild-Type
XDR-TB	Extensively Drug-Resistant Tuberculosis

NRPS/PKS domains

A	Adenylation
ACP	Acyl Carrier Protein
AT	Acyltransferase
C	Condensation
CYC	Cyclases
DH	Dehydratase
ER	Enoylreductase
KR	Ketoreductase
KS	Ketosynthase
MT	Methyltransferase
PCP	Peptidyl Carrier Protein
TE	Thioesterase

Table of Contents

Abstract.....	ii
Acknowledgements.....	iii
Author’s Declaration	iv
Abbreviations	v
Table of Contents	vii
Figure Index.....	xii
Table Index.....	xviii
Chapter 1: Introduction.....	1
1. Introduction.....	2
1.1 Antimicrobial Natural Products.....	2
1.1.1 Antimicrobial Resistance – A Rising Global Crisis	3
1.1.2 A History of Antibiotic Discovery	6
1.2 Methods of Natural Product Discovery and Production.....	10
1.2.1 Bioassay-Guided Discovery	10
1.2.2 Genome Mining.....	13
1.2.3 Chemical Elicitors	15
1.2.4 Genetic Engineering	17
1.2.5 Heterologous Hosts.....	22
1.3 Actinomycetes as Heterologous Hosts and Producers of NPs.....	25
1.4 The Different Classes of Natural Products	27
1.4.1 Ribosomally-Synthesised and Post-Translationally Modified Peptides (RiPPs)....	27
1.4.2 Non-Ribosomal Peptides.....	36
1.4.3 Polyketides.....	39
1.5 Project Objectives	45
Chapter 2: Materials and Methods	46
2. Materials and Methods	47
2.1 Lab Supplies.....	47
2.2 PACs, Plasmids, Primers and Strains	48
2.2.1 Vectors	48

2.2.2 Primers	51
2.2.1 Bacterial Strains and Growth Conditions	51
2.2.3 Antibiotics	60
2.3 Culture and Production Media.....	61
2.3.1 Culture Media	61
2.3.2 Cell Culture and Stock Preparation	66
2.4 DNA, Cloning and Transformation	68
2.4.1 DNA	68
2.4.2 Cloning	71
2.4.3 Transformation	72
2.5 Bioassays	74
2.5.1 Overlay Bioassays.....	74
2.5.2 Minimum Inhibitory Concentration (MIC) Bioassays.....	74
2.6 Extraction of Antibiotics.....	75
2.6.1 Liquid.....	75
2.6.2 Solid.....	76
2.7 Chemical Analysis and Characterisation	78
2.7.1 Analytical LCMS (Liquid Chromatography Mass Spectrometry)	78
2.7.2 HPLC (High Performance Liquid Chromatography).....	79
2.7.3 Tandem Mass Spectrometry (MS)	80
2.7.4 Nuclear Magnetic Resonance (NMR).....	82
2.7.5 Carbohydrate Analysis	82
2.8 Phylogenetic Trees	83
Chapter 3: Diversification of Class II Type B Lantipeptides.....	84
3. Production and Diversification of Class II Type B Lantibiotics using a Heterologous Synthetic Expression System.....	85
3.1 Introduction	85
3.2 Objectives.....	88
3.3 A Synthetic Cassette System for the Expression of Class II Type B Lantipeptides based on the Kyamicin BGC.....	89
3.3.1 Optimization of the Kyamicin Synthetic Cassette System and Heterologous Host	95
3.4 Genome mining for BGCs Encoding Class II Type B Lantibiotics	99
3.5 Heterologous Production of Lantipeptides using the Kyamicin Synthetic Cassette System.....	103

3.5.1 Bioassays of Hybrid <i>lanA</i> Lantipeptide Construct Strains in the Kyamicin Cassette System against <i>B. subtilis</i>	103
3.5.2 Chemical Analysis of Lantipeptides Produced by Heterologous Expression	105
3.5.3 Comparison of Produced and Not Produced Lantipeptide Core Peptides	107
3.6 Design of an Alternative Expression System using the <i>Streptomyces roseoverticillatus</i> lantipeptide BGC	113
3.6.1 <i>Streptomyces roseoverticillatus</i> BGC.....	113
3.6.2 Rosiermycin Synthetic Cassette System.....	113
3.6.3 Attempted production of Lantipeptides using the Rosiermycin Synthetic Cassette System.....	114
3.6.4 Identification of a Truncated <i>lanL</i> Immunity Gene within pAMA3.....	115
3.7 Production of Rosiermycin from <i>Streptomyces roseoverticillatus</i>	119
3.7.1 Rosiermycin Expression and Synthetic Activation Cassettes	119
3.8 Purification of Lantipeptides from the Kyamicin Expression Platform	122
3.8.1 Optimization of Lantipeptide Production Media	122
3.8.2 Optimization of Lantipeptide Extraction from Liquid Media	124
3.8.2 Optimization of Lantipeptide HPLC Purification	126
3.9 Characterization of Purified Lantipeptides	129
3.9.1 Bioactivity of Lantipeptides	129
3.9.2 Chemical Characterization of Lantipeptides	135
3.10 Discussion.....	142
Chapter 4: A Synthetic Library of Lantipeptides	150
4. Design of a Synthetic <i>lanA</i> Core Peptide Library for Incorporation into the Kyamicin Synthetic Cassette System.....	151
4.1 Introduction	151
4.2 Objectives.....	155
4.3 Design of a Four-Part Synthetic Lantipeptide Oligonucleotide Library	156
4.4 Heterologous Production of a Four-Part Synthetic Lantipeptide Library	157
4.4.1 Bioassay against <i>B. subtilis</i> of the Four-Part Synthetic Lantipeptide Libraries ..	159
4.4.2 Sanger Sequencing of Synthetic Lantipeptide Library Colonies and Identification of Missing Mutants	161
4.4.3 Chemical Analysis of Synthetic Lantipeptides Produced by Heterologous Expression of the Four-Part Synthetic Lantipeptide Library	163
4.5 Discussion.....	175
Chapter 5: Production of Actinomycetes NPs in <i>S. erythraea</i> ISOM	182

5. Expression of Actinomycete BGCs in <i>Saccharopolyspora erythraea</i> ISOM	183
5.1 Introduction	183
5.2 Objectives.....	185
5.3 Attempted Production of Putative Antifungal Polyenes in <i>S. erythraea</i> ISOM	186
5.3.1 Identification of Putative Antifungal Polyene BGCs in Ant Symbiotes by Genome Sequencing.....	186
5.3.2 <i>S. erythraea</i> ISOM as a Potential Heterologous Host for <i>Pseudonocardia</i> BGCs.....	189
5.3.3 Insertion and Production of Nystatin Ps1 and Putative Polyene Ps2 in <i>S. erythraea</i> ISOM.....	190
5.3.4 Identification of a Frame Shift in the Nystatin Ps1 BGC within PAC 6_19B using Illumina Sequencing.....	194
5.4 Insertion and Expression of Actinomycetes NP BGCs in <i>S. erythraea</i> ISOM.....	195
5.4.1 PACs containing Actinomycetes BGCs	195
5.4.2 Production Media Assays of NP BGCs Inserted into <i>S. erythraea</i> ISOM.....	196
5.4.3 Mass Spectrometry Analysis of <i>S. erythraea</i> ISOM PAC Extracts Generated from Production Media Assays.....	198
5.5 Overexpression of Transcriptional Activators for Induction of Antascomycin Production from PAC 21A and PAC 21B.....	200
5.5.1 Antascomycin Transcriptional Activators	200
5.5.2 Conjugation of Antascomycin Transcriptional Activators into Heterologous Hosts <i>S. erythraea</i> ISOM and <i>S. coelicolor</i> M1146.....	205
5.5.3 Antascomycin Production Media Assays with Overexpression of Transcriptional Activators	206
5.5.4 Mass Spectra Analysis of Extracts Generated from Overexpression of Antascomycin Transcriptional Activators in Heterologous Hosts.....	206
5.6 Discussion.....	210
Chapter 6: Heterologous Production of Fasamycins and Formicamycins	213
6. Comparison of Heterologous Hosts for Production of Formicamycins and Fasamycins	214
6.1 Introduction	214
6.2 Objectives.....	218
6.3 Deletion of <i>forJ</i> in the <i>for</i> BGC of <i>S. formicae</i> KY5 and in PAC 215G containing the <i>for</i> BGC.....	219
6.4 Fasamycins and Formicamycins Production Profiles from Heterologous Hosts Expression Strains in Comparison to Parental Strain <i>S. formicae</i> KY5.....	221
6.4.2 HPLC Titre of Fasamycins and Formicamycins from the Parental Strain and Heterologous Host Strains	221

6.5 Identification of New Fasamycins Produced by <i>S. erythraea</i> ISOM/215G Δ forJ.....	224
6.5.1 Observation of New early Eluting Fasamycins by HPLC from <i>S. erythraea</i> ISOM/215G Δ forJ Solid Media Extracts.....	224
6.5.2 LCMS analysis of New Fasamycins	226
6.5.3 Carbohydrate Analysis of New Glycosylated Fasamycins	230
6.6 Discussion.....	232
Chapter 7: Discussion.....	236
7. Discussion.....	237
7.1 Diversification of Class II Type B Lantipeptides.....	238
7.2 Rosiermycin.....	243
7.3 Investigation of <i>S. erythraea</i> ISOM as a heterologous host for BGC expression	245
7.4 New Glycosylated Fasamycins	247
7.5 Conclusions	249
Chapter 8: References.....	252
Chapter 9: Appendix.....	284
9. Appendix.....	285
9.1 Chapter 2 Supplementary Material	285
9.1.1 Sequencing Primers	285
9.1.2 Cloning Primers	288
9.1.3 Oligonucleotide Templates	289
9.2 Chapter 3 Supplementary Material	290
9.2.1 UHPLC-HRMS of Produced Lantipeptides	290
9.2.2 Phylogenetic Trees of Identified Lantipeptides	297
9.2.3 Rosiermycin Synthetic Cassette Bioassay	298
9.2.4 Tandem MS	299
9.2.5 NMR	303
9.3 Chapter 5 Supplementary Material	310
9.3.1 Production Media Assays Mass Spectra of <i>S. erythraea</i> ISOM/215G Δ forJ.....	310
9.4 Chapter 6 Supplementary Material	312
9.4.1 Production of Fasamycin C and Fasamycin D in M1146/215G Δ forJ.....	312
9.4.2 Comparison of Mass Spectra of <i>S. formicae</i> KY5 WT and <i>S. formicae</i> KY5 Δ forJ in Liquid Media	313
9.4.3 Concentration Curves used to Determine Fasamycin and Formicamycin Titre from HPLC Chromatograms	314
9.5 Plasmid Maps	315

Figure Index

Figure 1.1: Chemical Structures of Select Clinical Antimicrobials Produced by <i>Streptomyces</i> Species	007
Figure 1.2: Timeline of Antibiotic Discovery, Resistance, and Clinical Use	008
Figure 1.3: Gene Cluster Approaches for Activating Silent BGCs	018
Figure 1.4: Compounds Induced via Transcriptional Activator Overexpression or De-repression of Transcriptional Repressors	021
Figure 1.5: General RiPP Biosynthetic Pathway	028
Figure 1.6: Overview of the Four Classes of Lantipeptide Biosynthesis	030
Figure 1.7: General Biosynthetic Scheme of Thioether Linkage in Lantipeptides ...	031
Figure 1.8: Biosynthesis and Structure of Cinnamycin	033
Figure 1.9: Non Ribosomal Peptides (NRPs) that are Used Clinically	036
Figure 1.10: Illustration of NRPS Biosynthesis	038
Figure 1.11: Structures and Mechanisms of Bacterial Type I, II and III PKSs	041
Figure 1.12: Formicamycin BGC and Gene Functions	043
Figure 1.13: Formicamycin I	044
Figure 3.1: Kyamicin Peptide Sequence, BGC, Biosynthesis and Structure	086
Figure 3.2: Activation of Kyamicin Biosynthesis in <i>Saccharopolyspora</i> KY21, and Expression in Heterologous Host <i>S. coelicolor</i> M1152	091
Figure 3.3: Duramycin Biosynthesis Induced in Heterologous Host <i>S. coelicolor</i> M1152 by Kyamicin Activation Cassette pEVK6	092
Figure 3.4: Kyamicin BGC and Two-Part Synthetic Cassette	094
Figure 3.5: Model of Regulation of Cinnamycin	095
Figure 3.6: Kyamicin Activation and Immunity Cassettes	096
Figure 3.7: Comparison of Kyamicin production Induced by Activation Cassettes pEVK6 and pAMA1	098
Figure 3.8: Comparison of Heterologous Hosts <i>S. coelicolor</i> M1152 and <i>S. coelicolor</i> M1146 for Kyamicin Production	098
Figure 3.9: Zone of Inhibition caused by Lantipeptides produced through the Heterologous Kyamicin Synthetic Cassette System	104

Figure 3.10: Extracted Ion Chromatograms (XICs) of Bioassay Plate Extracts from Strains L_16, L_24 and L_25	106
Figure 1.11: Sequence Alignment of Lantipeptides Produced and Not Produced using the Kyamicin Heterologous Cassette Platform	111
Figure 3.12: Rosiermycin Synthetic Cassettes	114
Figure 3.13: Bioassay of Strain Lr_00 (Kyamicin) and Lr_14 (Rosiermycin) with <i>kyoL</i> Immunity Cassette pAMA4	117
Figure 3.14: Bioassay of Strain Lr_00 (Kyamicin) and Lr_14 (Rosiermycin) with Activation Cassette pAMA5	118
Figure 3.15: Bioassay plates of <i>S. roseoverticillatus</i> with Synthetic Cassettes pAMA1 and pAMA5	120
Figure 3.16: Production of Rosiermycin from <i>S. roseoverticillatus</i> Wild-Type in Comparison to Activator Cassette pAMA1 and pAMA5 Mutants	121
Figure 3.17: Lantipeptide Production Media Optimization	123
Figure 3.18: Lantipeptide Extraction Solvent Optimization	125
Figure 3.19: Chromatogram of L_19 preparative HPLC	127
Figure 3.20: Antibacterial Comparison of L_19 Deoxy from Chromatographic Fractions 1-3	130
Figure 3.21: Comparative Bioassay of Novel Lantipeptides	132
Figure 3.22: Schematic Representation of L_19 Hydroxy F3 Reduction and Corresponding LCMS Peaks	136
Figure 3.23: Fragmentation of Fully Reduced L_19 Hydroxy F3 (zero sulphur)	137
Figure 3.24: L_19 Hydroxy F3 Chemical Structure with Corresponding Atom Assignment and ¹ H NMR Spectrum	140
Figure 3.25: L_19 Hydroxy F3 and L_19 Deoxy F3 HSQC Overlay and Identification of Asp15	141
Figure 4.1: Schematic of Kyamicin with PE Binding Pocket	153
Figure 4.2: Kyamicin Core Peptide Amino Acid Sequence	156
Figure 4.3: Methodology for Producing Synthetic Lantipeptide Libraries with Reduced Mutant Bias	159
Figure 4.4: Bioassays of the Four-Part Synthetic Lantipeptide Library	160

Figure 4.5 Bioassay plates of synthetic lantipeptide core peptide sequences within pWDW70	162
Figure 4.6: Box Plot Representation of Synthetic Lantipeptide Production	164
Figure 4.7: Tables Showing Effect of Single Amino Acid Mutation at Position Ala2 of Kyamycin Core Peptide	166
Figure 4.8: Tables Showing Effect of Single Amino Acid Mutation at Position Ala7 of Kyamycin Core Peptide	168
Figure 4.9: Tables Showing Effect of Single Amino Acid Mutation at Position Phe12 of Kyamycin Core Peptide	170
Figure 4.10: Tables Showing Effect of Single Amino Acid Mutation at Position Ala13 of Kyamycin Core Peptide	172
Figure 5.1: Comparison of Bacterial Gene Clusters and Chemical Structures of Nystatin A1, NPP and the Predicted Structure of Nystatin P1 and Putative Polyene Ps2.....	188
Figure 5.2: Colony PCR of <i>S. erythraea</i> SH ex-conjugants after Insertion of PACs 6_19B and 2_5K	191
Figure 5.3: Extracted Ion Chromatogram (EIC) and UV Absorbance of Nystatin A1 standard	193
Figure 5.4: Genetic Organization and Model Biosynthesis of the Nystatin Ps1	194
Figure 5.5: Extracted Ion Chromatograms (EIC) of <i>S. erythraea</i> SH/215G and <i>S. erythraea</i> SH/215G Δ forJ SV2 Liquid Media Extracts	199
Figure 5.6: Chemical Structure of Antascomicins	201
Figure 5.7: Comparative BGC Analysis of Rapamycin and Antascomicin	202
Figure 5.8: Extracted Ion Chromatograms (EIC) of <i>S. coelicolor</i> M1146/21A with the Antascomicin BGC and Transcriptional Activators	207
Figure 5.9: Extracted Ion Chromatograms (EIC) of <i>S. erythraea</i> SH with Antascomicin PAC 21A and Transcriptional Activators	209
Figure 6.1: Formicamycin BGC, Gene Annotation and Predicted Post-PKS Tailoring Steps	215
Figure 6.2: Chemical Structures of Fasamycins and Formicamycins isolated from <i>Streptomyces formicae</i>	216

Figure 6.3: Principle of λ Red-Mediated Replacement of a Target Gene by Homologous Recombination	220
Figure 6.4: Titre ($\mu\text{g}/\text{mL}$) of Fasamycins (Blue) and Formicamycins (Orange) Produced from <i>S. formicae</i> and Heterologous Host Strains	223
Figure 6.5: Six New Fasamycin Elution Peaks Identified in <i>S. erythraea</i> ISOM/215G Δ forJ HPLC Chromatogram	225
Figure 6.6: Extracted Ion Chromatogram (EIC) and LC-MS/MS Fragmentation of New Glycosylated Fasamycin Congeners	227
Figure 6.7: Hypothetical Structure of New Fasamycins 3, 4, 5 and 6	229
Figure 6.8: Carbohydrate Analysis (HPAE-PAD) of Carb2 and Carb3 obtained from Fasamycin-like Congeners 2 (m/z 649.16) and 3 (m/z 635.21)	231
Figure 7.1: Biosynthesis of Duramycin including Lanthionine (Lan) and Methyllanthionine ((Me)Lan) bond formation and Lysinoalanine bridge formation	240
Figure S1: Extracted Ion Chromatogram (EIC) $[\text{M}+3\text{H}]^{3+}$ M1146/pEVK6/L_01 (Cinnamycin)	290
Figure S2: EIC $[\text{M}+3\text{H}]^{3+}$ of M1146/pEV6/L_02 (Duramycin)	290
Figure S3: EIC $[\text{M}+3\text{H}]^{3+}$ of M1146/pEVK6/L_03 (Duramycin B)	291
Figure S4: EIC $[\text{M}+2\text{H}]^{2+}$ of M1146/pEVK6/L_07 (Mathermycin)	291
Figure S5: EIC $[\text{M}+2\text{H}]^{2+}$ of M1146/pEVK6/L_10	292
Figure S6: EIC $[\text{M}+3\text{H}]^{3+}$ of M1146/pEVK6/L_11	292
Figure S7: EIC $[\text{M}+2\text{H}]^{2+}$ of M1146/pEVK6/L_13	293
Figure S8: EIC $[\text{M}+2\text{H}]^{2+}$ of M1146/pEVK6/L_15	293
Figure S9: EIC $[\text{M}+2\text{H}]^{2+}$ of M1146/pEVK6/L_18	294
Figure S10: EIC $[\text{M}+2\text{H}]^{2+}$ of M1146/pEVK6/L_19	294
Figure S11: EIC $[\text{M}+2\text{H}]^{2+}$ of M1146/pEVK6/L_22	295
Figure S12: EIC $[\text{M}+2\text{H}]^{2+}$ of M1146/pEVK6/L_23	295
Figure S13: EIC $[\text{M}+2\text{H}]^{2+}$ of M1146/pEVK6/L_26	296
Figure S14: Phylogenetic Trees of Core Peptide Regions and LanA peptides from Identified Lantipeptides L_00 to L_35	297
Figure S15: Rosiermycin Heterologous Synthetic Cassette System Strains Overlaid with <i>B. subtilis</i>	298

Figure S16: Fragmentation of Fully Reduced L_19 Deoxy F3 (zero sulphur)	300
Figure S17: Fragmentation of Fully Reduced L_18 Hydroxy F3 (zero sulphur)	301
Figure S18: Fragmentation of Fully Reduced L_26 Deoxy F3 (zero sulphur)	302
Figure S19: L_19 Hydroxy F3 2D COSY NMR	306
Figure S20: L_19 Hydroxy F3 2D TOCSY NMR	307
Figure S21: L_19 Hydroxy F3 2D NOESY NMR	308
Figure S22: L_19 Hydroxy F3 2D HSQC NMR	309
Figure S23: Extracted Ion Chromatograms (EIC) of <i>S. erythraea</i> SH/215G and <i>S. erythraea</i> SH/215G Δ forJ	310
Figure S24: Extracted Ion Chromatograms (EIC) of <i>S. erythraea</i> SH/215G and <i>S. erythraea</i> SH/215G Δ forJ	311
Figure S25: Extracted Ion Chromatograms (EIC) of <i>S. coelicolor</i> M1146/215G Δ forJ	312
Figure S26: Extracted Ion Chromatograms (EIC) of <i>S. formicae</i> KY5 WT and <i>S. formicae</i> KY5 Δ forJ	313
Figure S27: Concentration Curves of Fasamycin E and Formicamycin E used to determine Titre from HPLC Chromatograms	314
Figure S28: pBluescript_II_K Plasmid Map	315
Figure S29: pEVK4 Plasmid Map	315
Figure S30: pEVK6 Plasmid Map	316
Figure S31: pWDW63 Plasmid Map	316
Figure S32: pWDW68 Plasmid Map	317
Figure S33: pWDW69 Plasmid Map	317
Figure S34: pWDW70 Plasmid Map	318
Figure S35: pAMA1 Plasmid Map	318
Figure S36: pAMA2 Plasmid Map	319
Figure S37: pAMA3 Plasmid Map	319
Figure S38: pAMA4 Plasmid Map	320
Figure S39: pAMA5 Plasmid Map	320
Figure S40: pADW11 Plasmid Map	321
Figure S41: pBF3 Plasmid Map	321
Figure S42: pESAC13 Map	322

Figure S43: pGP9 Plasmid Map	322
Figure S44: pIJ102567 Plasmid Map	323
Figure S45: pRF10 Plasmid Map	323

Table Index

Table 1.1: Class II Type B Lantipeptide Biosynthesis Genes	034
Table 2.1: Lab Supplies and Kits	047
Table 2.2: PACs used during this work	048
Table 2.3: Plasmids used during this work	049
Table 2.4: <i>E. coli</i> strains used in this work	052
Table 2.5: <i>Streptomyces</i> strains used in this work	054
Table 2.6: <i>Saccharopolyspora</i> strains used in this work	058
Table 2.7: Antibacterial activity indicator strains used in this work	059
Table 2.8: Antibiotics used in this work	060
Table 2.9: Protocol for Q5 PCR reaction	069
Table 2.10: Protocol for GoTaq PCR reaction	069
Table 2.11: Protocol for MyTaq PCR reaction	070
Table 2.12: HPLC Steps and Solvent for Purification of Lantipeptides	080
Table 3.1: Lantipeptide Core Peptides, Designation, GeneBank Accession Number	100
Table 3.2: Length, Proline Inclusion and No. of Amino Acid (aa) Changes compared to Kyamicin of Hybrid Lantipeptide Core Peptides. Table ranked by core peptide length	108
Table 3.3: Length, Proline Inclusion and No. of Amino Acid (aa) Changes compared to Kyamicin of Hybrid Lantipeptide Core Peptides. Table ranked by no. of amino acid changes compared to the kyamicin core peptide	109
Table 3.4: Lantipeptides Purified	128
Table 3.5: Minimum Inhibitory Concentrations of novel Lantipeptides ($\mu\text{g}/\text{mL}$)	134
Table 5.1: PACs, Associated NPs and the Parental Strains of BGCs Inserted into Heterologous Host <i>S. erythraea</i> SH	196
Table 5.2: Liquid production media assays for PACs in <i>S. erythraea</i> SH	197
Table S1: Primers used for Sangar Sequencing during this work	285
Table S2: Primers used for Cloning during this work	288
Table S3: Nucleotide Templated for Lantipeptide Core Peptide Oligonucleotides used during this work	289

Table S4: Calculated and Observed Mass Fragmentation of Reduced L_19 Hydroxy F3 (zero sulphur)	299
Table S5: Calculated and Observed Mass Fragmentation of Reduced L_19 Deoxy F3 (zero sulphur)	300
Table S6: Calculated and Observed Mass Fragmentation of Reduced L_18 Hydroxy F3 (zero sulphur)	301
Table S7: Calculated ad Observed Mass Fragmentation of Reduced L_26 Deoxy F3 (zero sulphur)	302
Table S8: Chemical Shifts in NMR Spectra of L_19 Hydroxy F3 (600 MHz, DMSO-d6, 298K)	303

Chapter 1:

Introduction

1. Introduction

1.1 Antimicrobial Natural Products

Antimicrobials have been widely used since the 1930s to treat infection, and refer to compounds that kill or inhibit the growth of bacteria, fungi or viruses (Zaman *et al*, 2017). They are an essential part of modern society, with medical use of antimicrobials increasing by 65% between 2000 and 2015, from 21.1 to 34.8 billion doses per year (Klein *et al*, 2018). Antimicrobials are also widely used in the agricultural sector, with over 80% of the antimicrobials consumed in the USA per year used as growth promoters to increase the efficiency of livestock production. This is achieved by preventing the growth of bacteria in the gastrointestinal tract, increasing feed-to-animal product efficiency, and reducing the rate of mortality by infection in close contact, high intensity livestock farming (Chattopadhyay, 2014; Gaskins *et al*, 2002; Hollis & Ahmed, 2013; Kirchhelle, 2018).

During cellular growth metabolites can be broadly separated into two categories, primary and secondary metabolites. Primary metabolites are directly involved in the normal growth, development, and reproduction of the cell. Secondary metabolites, also referred to as specialised metabolites, are not directly involved in these functions but usually play a secondary role in metabolism or carry out an important ecological function, such as cell-to-cell signalling, nutrient acquisition, or defence against competitors. Natural products (NPs) refer to primary and secondary metabolites originating from bacteria, fungi, plants and animals (Sorokina & Steinbeck, 2020; Zeng *et al*, 2018). The term can also refer to chemically synthesised compounds based upon natural metabolites (Hong, 2014; Maier, 2015).

Antimicrobial NPs are predominantly secondary metabolites produced to provide a competitive environmental advantage by defending against and inhibiting the growth of other microorganisms (Clardy *et al*, 2009; Katz & Baltz, 2016; Pishchany & Kolter, 2020; Van Arnem *et al*, 2017). For the course of this thesis the term NPs will be used to refer to antimicrobial, or potentially antimicrobial, NPs only.

1.1.1 Antimicrobial Resistance – A Rising Global Crisis

Without antimicrobials, standard medical treatments such as hip replacements, organ transplants and chemotherapy could not safely take place. However, an increasing number of deaths are attributed to antimicrobial resistant (AMR) infections, with over 35,000 deaths in the USA, and 700,000 deaths worldwide estimated to be caused per year by drug resistant diseases ((CDC) Centers for Disease Control & Prevention, 2019; (IACG) Interagency Coordination Group on Antimicrobial Resistance, 2019). The root cause of these deaths are strains of bacteria that are resistant to multiple antibiotics, including antibiotics of last resort. These types of bacteria are often called ‘superbugs’; examples of which include methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant *Enterococci* (VRE) and carbapenem-resistant tuberculosis. Carbapenem-resistant tuberculosis is also known as extensively drug-resistant tuberculosis (XDR-TB) and is incurable with current anti-TB drugs (Prasanna & Niranjana, 2019).

Resistance to antimicrobials can arise in multiple ways, including; through mutation of the antimicrobial target that impairs binding, changes to the cell wall preventing antibiotics entering the cell, an increase in efflux pumps expelling antibiotics from the cell, and the horizontal transfer of genes encoding antimicrobial resistance mechanisms from one cell to another (Munita & Arias, 2016; Reygaert, 2018). Horizontal gene transfer can result in the rapid spread of antibiotic resistance through a microbial population, transferring resistance from would-be benevolent or unharmed bacterial strains to virulent ones, and bestowing multi-drug resistance to infectious strains. For example, MRSA, one of the most common causes of hospital acquired infections, has acquired the vancomycin resistance operon *vanA* from VRE via a *Tn1546* transposon, creating multidrug resistant *S. aureus* (VRSA). Since 2002, 52 strains of VRSA have been isolated worldwide (Cong *et al.*, 2020; Périchon & Courvalin, 2009b).

While antimicrobial resistance is a natural phenomenon, and has been documented since the 1930s, it has seen an alarming rise in recent years (Zaman *et al.*, 2017). The reason for this rise may in part originate from the overuse and misuse of

antibiotics around the globe. Due to social pressure it has become commonplace in many areas to prescribe antibiotics for even mild maladies, including viral infections such as the common cold, which cannot be treated with antibiotics. While many people think it is better to be safe than sorry when it comes to their health this over-prescription can do more harm than good, by killing essential microflora without alleviating any symptoms and increasing the chance of AMR bacteria arising through selective pressure (Fair & Tor, 2014; Shallcross & Davies, 2014; Ventola, 2015). Furthermore, around the globe antibiotics are available for purchase without prescription (Mainous *et al*, 2009), leading many people to self-diagnose and self-administer, leading to ingestion of sub-inhibitory levels of antibiotics and selective pressure for bacteria with antibiotic resistance (Ajibola *et al*, 2018; Nepal & Bhatta, 2018).

This misuse of antibiotics has led to the repeated administration of antibiotics at subinhibitory concentrations. Low concentrations of antibiotics allow some bacteria to survive and selects for those with adaptations conferring resistance to low levels of antibiotics. Premature termination of antibiotic administration then leads to these bacteria multiplying, causing a return of disease symptoms. Subsequently, a stronger dose of antibiotics will be needed to control a second round of infection. Should this cycle of subinhibitory antibiotic administration continue microbes will adapt under the selective pressure to withstand extremely high doses of the antibiotic, rendering them ineffective (Sandegren, 2014; Ventola, 2015; Viswanathan, 2014; Westhoff *et al*, 2017).

Possibly the greatest misuse of antimicrobials happens in the agriculture and animal husbandry industries. The rise of high throughput factory farming creates breeding grounds for disease, with antimicrobials being routinely administered to healthy animals as preventatives to infection and to increase growth, leading to a great selective pressure on microbes to develop resistance (Fair & Tor, 2014; McEwen & Fedorka-Cray, 2002; Ventola, 2015). The widespread use of antifungals in monoculture crop farming also generates an environmental selective pressure to develop antifungal resistance (Fisher *et al*, 2018). Once it has arisen, antimicrobial

resistance can be passed from the agricultural microbiome to human microbial flora through food consumption and horizontal gene transfer (Ventola, 2015).

The problem of antimicrobial resistance is further compounded by the phenomenon of cross resistance. Cross resistance occurs when mutations arising due to pressure of a single antimicrobial also provide resistance to multiple other antimicrobials (Barker, 1999; Périchon & Courvalin, 2009a; Ray *et al.*, 1980). This happens because many antimicrobials share binding targets in just a few key systems within the cell. For example, bacteria which develop resistance against the quinolone antibiotic nalidixic acid also acquire resistance against the quinolone antibiotic ciprofloxacin, as both antibiotics work by inhibiting the action of the enzyme topoisomerase, which plays a key role in DNA replication (Périchon & Courvalin, 2009a). Thus, misuse of antimicrobials can induce resistance to multiple different antimicrobials with similar modes of action (Norby, 2009; Ray *et al.*, 1980).

As AMR continues to rise it is vital that new antibiotics with unique modes of action are discovered to replace those already widely used. However, without a change in the restrictions of use of antibiotics, and a change in global attitudes to antibiotic administration, it is assured that antibiotic resistance will continue to arise.

1.1.2 A History of Antibiotic Discovery

The concept of antimicrobial agents originated from Paul Ehrlich in the early 20th century. He proposed a theory that there could be a ‘magic bullet’, a compound capable of killing specific microbes without harming the host (Ehrlich, 1913; Witkop, 1999). In 1909 he developed the first of these ‘magic bullets’, an antimicrobial organoarsenic compound called Salvarsan which could treat syphilis. The use of this compound led to the discovery that microorganisms could develop resistance to antimicrobials (Kasten, 1996; Silberstein, 1924).

In 1928 Sir Alexander Fleming discovered the first NP antibiotic penicillin, produced by a *Penicillium* fungi (Fleming, 1929). However, it wasn’t until the 1940s that penicillin was produced on an industrial scale and its structure elucidated (Chain *et al*, 1940; Hodgkin, 1949). At around the same time, other microorganisms became the topic of intense examination in the hope of discovering further antimicrobial NPs. One such investigation focused on actinomycete soil bacteria, in particular *Streptomyces* species, which were shown to produce many different antimicrobials (Waksman *et al*, 1946), including; streptomycin, the first antibiotic used to treat tuberculosis (Schatz *et al*, 2005); cycloserine, a current second line antimicrobial used for the treatment of multidrug resistant tuberculosis (Offe, 1988; Yu *et al*, 2018); nystatin A, an antifungal used to treat fungal skin infections and oral thrush (Harman & Masterson, 2008; Lyu *et al*, 2016), and chloramphenicol, used to treat infections such as cholera, meningitis and typhoid fever (Butler *et al*, 1999; Duke *et al*, 2003; Gharagozloo *et al*, 1970; Lindenbaum *et al*, 1967; Rosove *et al*, 1950) (**Fig 1.1**).

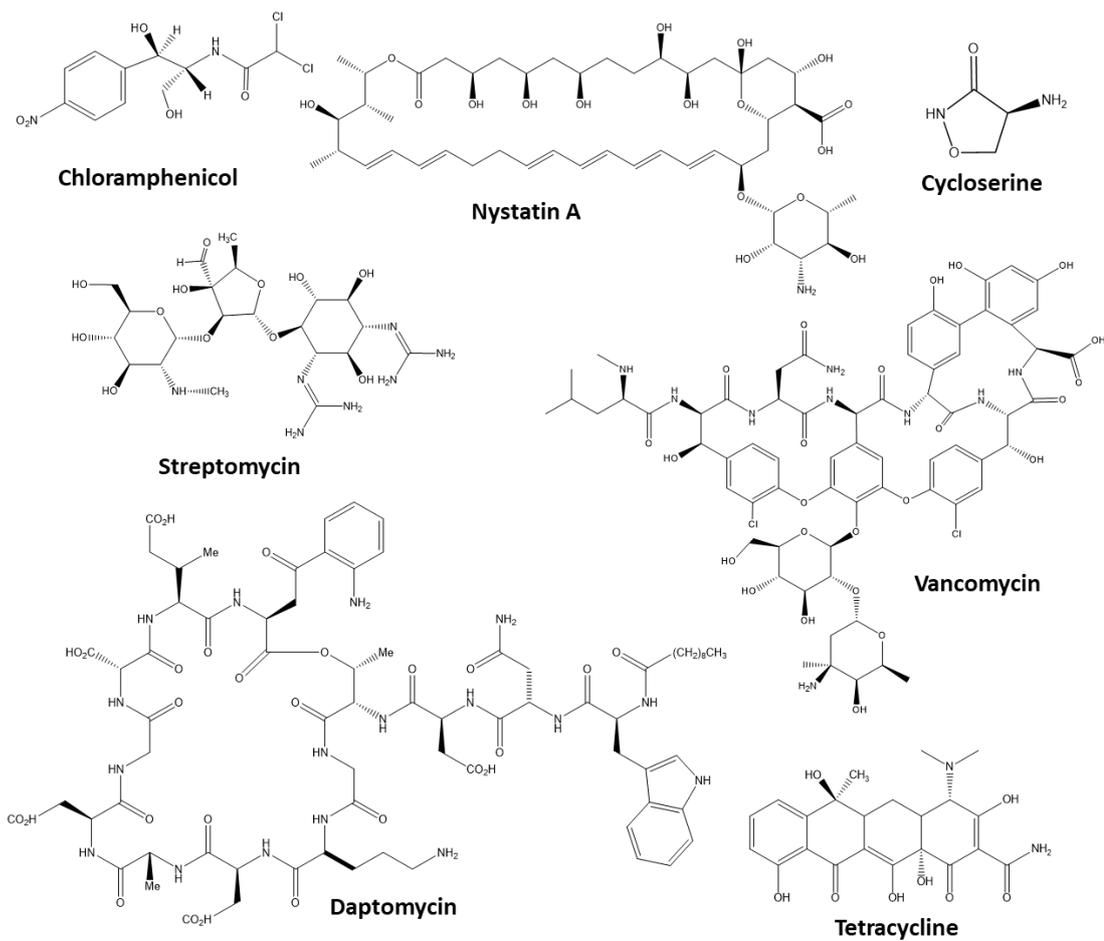


Figure 1.1: Chemical Structures of Select Clinical Antimicrobials Produced by *Streptomyces* Species. In order; Chloramphenicol from *Streptomyces venezuelae*, Nystatin A from *Streptomyces noursei*, Cycloserine from *Streptomyces garyphalus*, Streptomycin from *Streptomyces griseus*, Vancomycin from *Amycolatopsis orientalis* (previously called *Streptomyces orientalis*), Daptomycin from *Streptomyces roseosporus*, and Tetracycline from *Streptomyces aureofaciens*.

From this point on large scale fermentation of microorganisms to produce antimicrobials was widely exploited, resulting in what is known as the ‘Golden Age’ of antimicrobial discovery that encompassed the 1940s-1960s (Aminov, 2010; Jones *et al*, 2017). The widespread introduction of antibiotics resulted in a huge drop in infection-associated mortality, as well as a significant overall increase in life expectancy (Adedeji, 2016). However, since the 1960s only a few new classes of antibiotics have been discovered and approved for clinical use. Not only has the rate of discovery slowed significantly but resistance to novel antibiotics has arisen almost simultaneously (**Fig 1.2**).

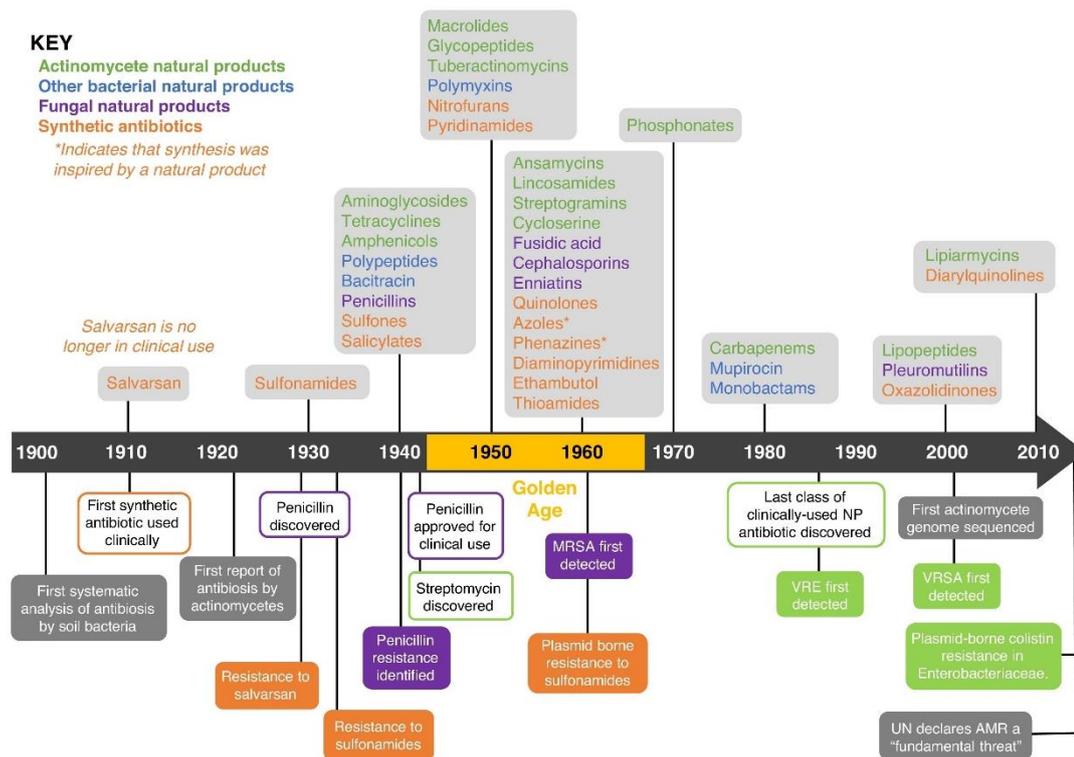


Figure 1.2 Timeline of Antibiotic Discovery, Resistance, and Clinical Use. Antibiotics are coloured per their source: green = actinomycetes, blue = other bacteria, purple = fungi and orange = synthetic. Drug-resistant strains: methicillin-resistant *S. aureus* (MRSA), vancomycin-resistant *enterococci* (VRE), vancomycin-resistant *S. aureus* (VRSA). Adapted from (Hutchings *et al*, 2019), Current Opinion in Microbiology, 53, p. 73, licenced under CC BY 4.0.

One of the reasons for the drop in the discovery rate of novel antimicrobials was the problem of re-discovery. Traditional screening programs, such as the Waksman platform, isolated actinomycetes bacterial strains from soil samples and screened them systematically for antibacterial activity by growing them under various conditions, challenging them with target bacteria and observing the appearance of zones of inhibition (Lewis, 2013). This process was time consuming and while it had at first yielded new antibiotics consistently it soon began to result in the re-discovery of known antibiotics as organisms were being isolated from very similar environments and then screened in very similar ways (Baltz, 2006). This led to a need to rapidly establish the novelty of a compound using dereplication protocols and this has driven many technological advances in methods such as mass spectrometry (MS) and nuclear magnetic resonance (NMR). In recent years improvements in automation and computer processing power has allowed us to

perform these techniques and others, such as genome mining and metagenomic sequencing, much faster at a reduced cost, opening up their potential for widespread use in drug discovery (Koehn & Carter, 2005). Nonetheless, re-discovery still poses a major challenge to modern antibiotic discovery, with many compounds being re-discovered consistently (Aminov, 2010; Gaudencio & Pereira, 2015; Peláez, 2006).

1.2 Methods of Natural Product Discovery and Production

1.2.1 Bioassay-Guided Discovery

Possibly the oldest method for the discovery of antimicrobials is bioassay-guided discovery, such as used in the previously described Waksman platform. This involves systematic screening of microbial cultures isolated from environmental samples against indicator strains to observe their antimicrobial activity. Conditions such as temperature, pH and growth media are varied to induce NP production. If activity is observed, chemical separation and purification techniques are then used to isolate and identify the NPs responsible.

As already stated, in more recent times these methods often results in the re-discovery of known compounds, wasting time, effort and money (Baltz, 2008). Despite this, bioassay guided discovery is still widely used and novel compounds continue to be discovered, for example two NPs from actinomycete isolated in Ihansia, Egypt were identified through bioassay-guided metabolite isolation and showed activity against *Bacillus subtilis* and MRSA (Sebak *et al*, 2020). Recently bioassay-guided approaches are increasingly run in conjunction with more advanced high throughput 'omic' techniques; such as genomics, transcriptomics, proteomics and metabolomics (Horgan & Kenny, 2011). These techniques can serve as quick and simple ways to: determine if observed bioactivity is caused by previously discovered compounds by identifying biosynthetic gene clusters (BGCs) within a strain, determine what BGCs are upregulated under specific environmental conditions by monitoring the change in mRNA profiles, and pre-emptively assign metabolites as potential antimicrobials, allowing for more targeted chemical separation and purification, thus saving time.

1.2.1.1 Underexploited Environmental Niches

To increase the chance of discovering novel antimicrobials, microbes are being sampled from extreme and under-explored environmental niches, including marine sediments (Jang *et al*, 2013; Mincer *et al*, 2002), deserts (Goodfellow *et al*, 2018) and arctic environments (Wietz *et al*, 2012), to name a few. The rationale behind this is that the biological diversity observed from these vastly different environments will lead to chemical diversity (Challinor & Bode, 2015; Lyddiard *et al*, 2016). Within such challenging environments resources are often scarce, leading to intense competition between organisms for survival. To secure these resources organisms are locked in an 'evolutionary arms race' to develop adaptations, such as antimicrobial NPs, to outcompete co-evolving species within the same environmental niche.

Within such highly competitive environments' cooperative relationships between organisms, called symbioses, can be an important factor for survival.

Symbionts are organisms that live in close physical association with one another, typically for mutual advantage (Akbar *et al*, 2018; Donia *et al*, 2014). Extended coevolution between symbionts can select for adaptations, including antimicrobial NP production, to provide an advantage against other competing organisms, such as by providing defence against infection and increasing nutrient acquisition (Adnani *et al*, 2017; Flórez *et al*, 2017; Koehler *et al*, 2013).

One such symbiotic relationship that selects for antimicrobial NP production is the tripartite insect-fungus-bacterium mutualism (Crawford & Clardy, 2011; Van Arnam *et al*, 2017). This is seen across multiple insect species that conduct 'fungus farming' as a method to secure food, including attine 'leaf-cutter' ants (Caldera *et al*, 2009; Currie, 2001; Mueller *et al*, 2001) and termites (Aanen *et al*, 2002; Benndorf *et al*, 2018). Within this system the insects are reliant on a specific fungal species for food, and sometimes nest building material. The insects 'farm' this fungus by providing it with food, in the form of vegetative material, in a clear two-way mutualistic relationship. However, pathogenic fungal species have co-evolved to invade the specific environmental niche of the nest. To defend against these

fungal invaders the insects have a second symbiotic relationship with mutualistic bacteria, including actinomycetes species. These bacteria produce antifungal NPs which can kill or inhibit the growth of the pathogenic fungus but have no effect against the symbiotic farmed fungus (Beemelmanns *et al*, 2016). Studies across many different systems have made it clear that the relationships between the cultivator fungi, pathogenic fungi, and insect carried mutualistic actinomycetes species have co-evolved over many millions of years (Currie *et al*, 2003; Fukuda *et al*, 2021; Holmes *et al*, 2016; Little & Currie, 2009).

Investigation of the bacterial species within these symbiotic systems has already identified several antimicrobial compounds, including dentigerumycin, a cyclic depsipeptide with highly modified amino acids produced by *Pseudonocardia* spp. associated with the attine ants species *Apterostigma dentigerum*, which selectively kills the parasitic fungus *Escovopsis* (Oh *et al*, 2009), and bacillaene A from *Bacillus* spp. associated with the termite species *Macrotermes natalensis* which selectively inhibits fungi antagonistic to the mutualistic cultivar fungi *Termitomyces* (Um *et al*, 2013). Further to this, the novel class II type B lantibiotic kyamicin, discussed further in Chapter 3 and upon which much of this thesis work is based, was identified from *Saccharopolyspora* spp. found on the cuticle of fungus farming *Tetraponera penzigi* ants (Vikeli *et al*, 2020).

However, many bacteria from such specific environmental niches, cannot be cultured under laboratory conditions, making assessing their biosynthetic potential challenging. This is thought to be due to the necessity of specific environmental factors, including inter-species relationships, needed for the survival of these microorganisms. For bacteria from niche environments these environmental conditions are often extreme, such as ultra-high or low temperatures, high or low pressure etc, and are difficult to replicate within a laboratory setting. It is estimated that over 99% of bacteria are unculturable under laboratory conditions (Baral *et al*, 2018). Therefore, other methodologies, such as genome mining and heterologous production, are used to investigate these microorganisms for novel NPs.

1.2.2 Genome Mining

Over the last 20 years many advances have occurred in DNA sequencing techniques and bioinformatics tools for the identification of NP BGCs, with huge impacts on NP discovery (Bachmann *et al*, 2014; Katz & Baltz, 2016). Genome sequencing has made it clear that even the most well studied strains contain high numbers of cryptic BGCs (Gram, 2015; Reen *et al*, 2015; Rutledge & Challis, 2015). For example, publication of the genome sequence of the model actinomycete strain *Streptomyces coelicolor* A3(2) in 2002 (Bentley *et al*, 2002), led to the surprising observation that it encoded over 20 BGCs indicating that this organism contained much greater biosynthetic potential than previously realised. These BGCs were termed 'silent' as they were not expressed under the laboratory conditions tested. Through the rise of metagenomic sequencing, the increase in computer processing power and the automation of genome assembly, it is now possible to extract the genetic material of unculturable microbes by harvesting mixed DNA from environmental samples, sequencing the DNA using next generation sequencing methods, and then re-assemble individual species genomes from the collective sequence in order to identify the BGCs present.

Computational tools available for the identification of BGCs and prediction of the compounds encoded include AntiSMASH (Blin *et al*, 2017), NaPDoS (Ziemert *et al*, 2012), GNP/PRISM (Johnston *et al*, 2015; Skinnider *et al*, 2015) and BAGEL (van Heel *et al*, 2018). AntiSMASH, NaPDoS and GNP/PRISM work through the rapid alignment of potential BGCs to an extensive database of known BGCs. Further to this, AntiSMASH breaks down each BGC into individual genes and annotates them with an assigned function based on homology to known genes within the National Centre of Biotechnology Information (NCBI) database. For modular assembly line pathways utilising non-ribosomal peptide synthetases (NRPS) and type I polyketide synthases (PKS), which are discussed further in Section 1.4, it also predicts the sequence of domains and modules present. NaPDoS uses a similar system identifying NRPS and PKS derived NPs through identification of characteristic elongation C- and KS- domains, creating a predicted structure of the final NP. GNP/PRISM can go further, linking the predicted putative biosynthetic products to

corresponding peaks in liquid chromatography mass spectrometry (LC-MS) data from strains within the GNP/iSNAP database. BAGEL is used to identify bacteriocins, a diverse group of Ribosomally-synthesised Post-translationally modified Peptides (RiPPs) with potential antimicrobial activity that are usually encoded for by small poorly conserved BGCs with little sequence homology.

To overcome a possible lack of homology to known genes for BGC identification these computational tools also consider the genomic context in which antimicrobial BGCs, such as bacteriocins, are found. For example, for each potential antimicrobial encoding BGC, the surrounding genome is analysed for genes encoding proteins involved in biosynthesis, transport, regulation and/or immunity, as antimicrobial BGCs are often found in association with these genes.

These tools have led to the discovery of many novel families of compounds but do have several limitations (Bachmann *et al.*, 2014; Ziemert *et al.*, 2016). Most obviously the extensiveness of the database used for homology comparison can limit the identification of BGCs. If a BGC present in the genome has no homology to any known BGCs and does not align to any BGC within the database, it may not be identified. As such, these tools often struggle to identify small BGCs with weak homology and BGCs for new NP classes. Secondly, there is a limitation to the accuracy of the annotation produced by these applications, as gene annotation and predicted NP structure is based upon the annotation of previous BGCs so continuous mis-annotation can occur.

Nonetheless, these rapid BGC identification and prediction tools have revolutionised the NP discovery pipeline. Previously, a bioactive compound would be discovered, and a 'back-to-front' approach then used to identify the BGC responsible for its biosynthesis. Now a 'front-to-back' approach can be utilised, in which a BGC of interest is identified and the novelty of the potential compound elucidated using database comparison and genetics before attempts to induce the production of the predicted NP begins. Production of novel NPs may then be induced through media experiments and the feeding of putative intermediates.

As many silent BGCs encode for NPs which impose a high energy cost on an organism they are only naturally expressed in response to specific stimuli and are under stringent regulation (Bibb, 2005). Common techniques used to induce NP production in bacterial strains are further discussed below, and can look to replicate this stress response, remove regulatory elements that limit expression, or impose new regulatory mechanisms altogether.

1.2.3 Chemical Elicitors

Supplementing growth media with chemical elicitors is a relatively simple way of inducing silent BGC expression without the need to use more challenging genetic techniques. This is especially relevant for talented species such as *Streptomyces* which can contain dozens of cryptic BGCs and where the identification and manipulation of all BGCs encoded in a genome would be extremely time consuming (Bibb, 2005). Various chemical elicitors have been identified in recent years that can induce the expression of silent BGCs. Examples of elicitors include a number of exogenous small molecules including antibiotics, which can act as signalling molecules when at subinhibitory concentrations but are toxic at high concentrations, a phenomenon known as hormesis (Okada & Seyedsayamdost, 2016; Seyedsayamdost, 2014). For example, sublethal concentrations of trimethoprim induced a change in the production of >100 metabolites in *Burkholderia thailandensis* fermentations, resulting in a metabolome very different to the uninduced *B. thailandensis*. These metabolites included two new analogues of capistrain, an antibacterial lasso peptide previously identified from this strain (Knappe *et al*, 2008; Okada *et al*, 2016).

Other chemical elicitors include inhibitors of histone deacetylases, which usually modify chromatin structure by removing acetyl groups from lysine residues on histones, allowing DNA to be wrapped more tightly and thus sequestering promoter regions of BGCs from the transcriptional machinery (Milazzo *et al*, 2020). Inhibiting histone deacetylation leaves promoter regions of genes free to be bound by transcriptional activators, thus inducing NP biosynthesis. One such histone

deacetylase inhibitor is sodium butyrate, the addition of which has been seen to alter the production of actinorhodin in *Streptomyces coelicolor* A3(2), causing upregulation or downregulation of actinorhodin on minimal and rich growth media, respectively (Moore *et al*, 2012). It has also been seen to elicit novel metabolites, inducing five cryptic pathways from *S. coelicolor* A3(2) as measured by quantitative PCR (Moore *et al.*, 2012) and 53 novel metabolites from a single strain (LAMA585) isolated from a deep-sea bacteria collection under laboratory conditions, as seen by liquid chromatography mass spectrometry (LC-MS) (de Felício *et al*, 2021).

Recently the advances in automation have led to high throughput screening of small molecule libraries to identify potential chemical elicitors through monitoring subsequent genomic and metabolomic changes of the target organism(s). This can be assessed through real-time PCR of the mRNA of a given gene or through whole cell metabolomics, monitoring the change in metabolic state of a strain through whole cell metabolite extraction. For example, using sonication and metabolite extraction through solvents, followed by mass spectrometry and gas chromatography (Christians *et al*, 2011; Xu *et al*, 2017). A successful example of this is the identification of canucin, RiPP with a β -hydroxylated amino acid, from a previously cryptic BGC within *Streptomyces canus* NRRL B3980 (Xu *et al*, 2019). The availability of high-throughput techniques means that hundreds of potential small molecule elicitors can be screened relatively quickly (Moore *et al.*, 2012).

While the application of chemical elicitors is simple, the technique has several constraints, the largest of which is the need to culture the bacterial strains under laboratory conditions. As mentioned before, most bacterial strains cannot be cultured within a laboratory imposing a restriction on what strains can be investigated. Secondly, as chemical elicitation is not a targeted approach the problem of re-discovery of known NPs can arise, as elicitation can induce the biosynthesis of multiple BGCs, including those already characterised. Thirdly, while chemical elicitors can lead to the induction of cryptic BGCs this can often be in minute amounts which are not sufficient for further purification and characterisation. Higher yields could be acquired through combination of chemical

elicitation and other techniques, such as genetic engineering, which is further explained below.

1.2.4 Genetic Engineering

With advances in DNA sequencing in addition to improved gene annotation software it is possible to predict the elements of many regions within cryptic BGCs, including promoters, transcriptional activators, and repressors. Manipulation of these regulatory elements, through the deletion of repressors, the constitutive or inducible expression of transcriptional activators and the refactoring of promoters, can potentially activate silent BGCs (**Fig 1.3**) (Baral *et al.*, 2018).

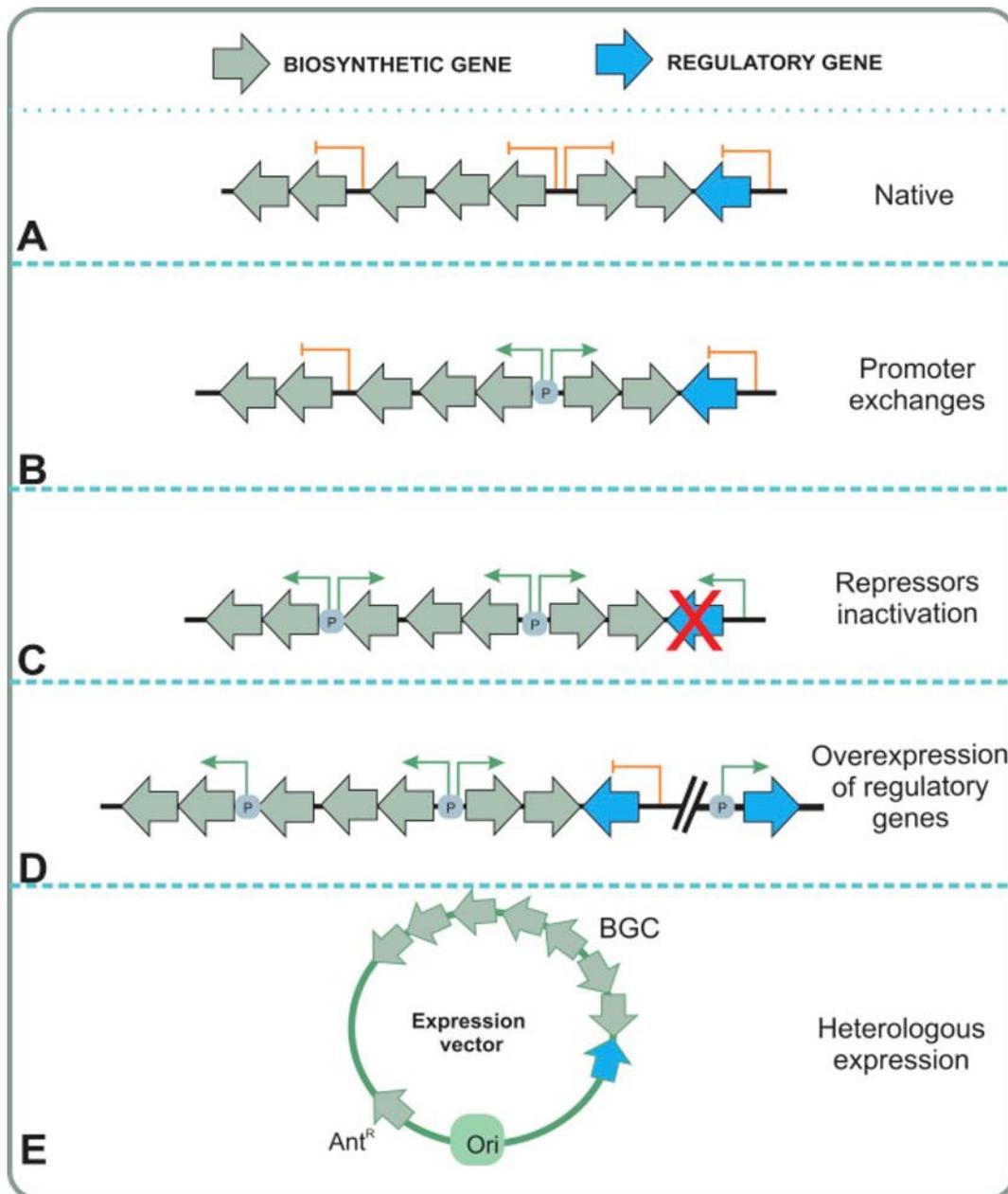


Figure 1.3: Gene Cluster Approaches for Activating Silent BGCs. Filled arrows indicate single genes within the operon including direction of transcription. Line arrows indicate start points and direction of transcription (green=transcription active, yellow=transcription inactive). (A) Native BGC silent under laboratory conditions, with no transcription observed. (B) Parts of the BGCs are activated by exchanging the native promoters with constitutively active promoters. (C) If present, the repressor is inactivated or removed from the BGC, leading to gene expression. (D) Overexpression of activator regulatory genes may also lead to activation of biosynthesis. (E) Cloning of the whole biosynthetic gene cluster and heterologous expression in a surrogate host can be used for artificial activation of the BGC. Abbreviations: Ant^R, Antibiotic selection marker; BGC, Biosynthetic gene cluster; Ori, Origin of replication; P, Promoters. Adapted from (Baral *et al.*, 2018), *Synthetic and Systems Biotechnology*, Vol 3, 3, p. 163-178, licensed under CC BY 4.0.

Within BGCs, operons consisting of many different genes, and sometimes the entirety of the BGC, can be under the control of a single promoter (**Fig 1.3A**). Regulation of this promoter can be complex, with activation induced by specific environmental stimuli. To negate this reliance on little understood complex regulatory networks for NP production native BGC promoters can be exchanged for well understood constitutive or inducible promoters (**Fig 1.3B**). For example, production of the blue pigment 'indigoidine' was induced in *Streptomyces albus* through placing the expression of the NRPS gene *bpsA* under control of the strong constitutive promoter *ermEp** (Olano *et al*, 2014). Furthermore, replacement of the native BGCs promoters of actinorhodin, undecylprodigiosin and polycyclic tetramate macrolactam (PTM) with the constitutive promoter *kasOp** using CRISPR/Cas9 in *S. albus*, *Streptomyces lividans* and *S. roseosporus*, respectively, led to production of the expected NPs (Zhang *et al*, 2017b).

Within BGCs pathway specific transcriptional regulatory genes can either initiate or repress the production of genes through binding to the promoter sequence. For transcriptional activators binding to the promoter region induces the binding of transcription factors, initiating transcription. For transcriptional repressors, binding to the promoter region prevents the binding of transcription factors, thus preventing transcription. Common transcriptional activators and repressors include *Streptomyces* Antibiotic Regulatory Proteins (SARPs) which are transcriptional activators found within actinobacteria (Romero-Rodríguez *et al*, 2015); LuxR transcriptional activators, in which ligand binding causes homodimerization allowing for binding to DNA in a *tra* box sequence which then initiates transcription (White & Winans, 2005), with a large sub-family of these transcriptional activators called LAL proteins (Large ATP-binding regulators of the LuxR family) are found throughout actinobacteria (Schrijver & Mot, 1999); and Tet repressor proteins (TetR), consisting of a helix-turn-helix DNA binding domain (Orth *et al*, 2000). A large subfamily of LuxR proteins, called the. Through the manipulation and control of transcriptional regulators NP production can also be induced (Baral *et al.*, 2018; Reen *et al.*, 2015).

Transcriptional activators are often overexpressed under strong constitutive promoters to induce NP production (**Fig 1.3D**). Potential transcriptional activators of BGCs are identified through their sequence homology to other known transcriptional regulators, cloned and placed under the control of a constitutive promoter within a vector, which is inserted into the bacterial strain of interest. Most commonly these transcriptional activators are pathway specific, but they can potentially affect multiple BGCs, causing induction of multiple different products (Krause *et al*, 2020; McLean *et al*, 2019). For example, the PAS-LuxR regulator PimM in *Streptomyces natalensis* controls pimarin biosynthesis, an antifungal polyketide used to treat eye infections, but its heterologous expression in *S. albus*, *Streptomyces nodosus* and *Streptomyces avermitilis* has led to increased production of the antifungals candicidin, amphotericin and filipin, respectively (**Fig 1.4A**) (Santos-Aberturas *et al*, 2011). In each case, a homologue of *pimM* is encoded in the native BGC.

In a similar fashion, deletion and redirection of transcriptional repressors, either by gene knock-out or by inducing the production of 'decoy' binding sites, can be used to activate multiple BGCs by relieving repression and thus permitting transcription (**Fig 1.3C**). By employing a transcription factor decoy strategy (TFD) eight large cryptic PKS and NRPS encoding BGCs were targeted for de-repression, leading to the characterization of a novel oxazole compound oxazolepoxidomycin A (**Fig 1.4B**) (Wang *et al*, 2019), a family of compounds including derivatives with antibacterial activity (Kakkar & Narasimhan, 2019). Furthermore, the knock-out of two TetR transcriptional repressors *totR5* and *totR3*, in conjugation with overexpression of the LAL transcriptional activator *totR1*, induced the production of three totopotensamides from *Streptomyces pactum* SCSIO 02999, one of which was novel (**Fig1.4B**) (Chen *et al*, 2017b).

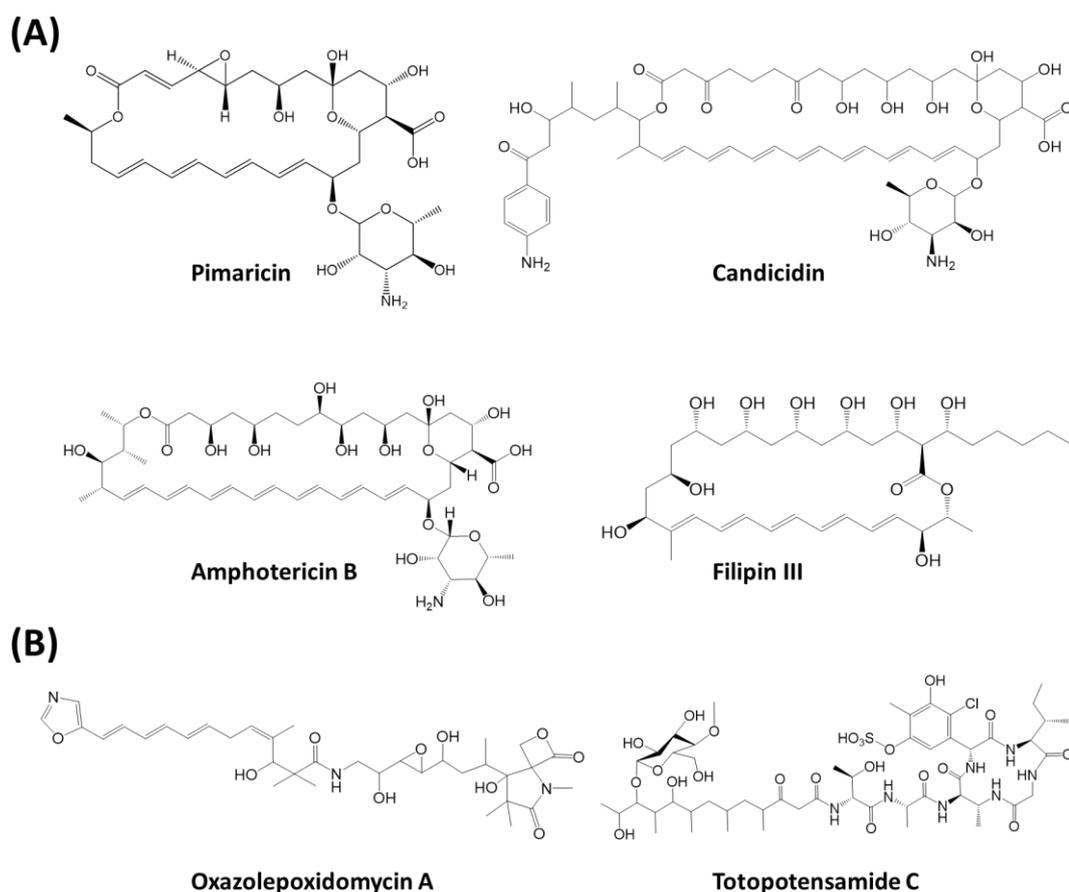


Figure 1.4: Compounds Induced via Transcriptional Activator Overexpression or De-repression of Transcriptional Repressors. (A) Pimaricin and homologs; candicidin, amphotericin B and filipin III, produced from *S. albus*, *Streptomyces nodosus* and *Streptomyces avermitilis*, respectively, by overexpression of PAS-LuxR transcriptional activator PimM (Santos-Aberturas *et al.*, 2011). (B) Novel compounds Oxazolepoxidomycin A (Wang *et al.*, 2019) and Totopotensamide C (Chen *et al.*, 2017b) produced via de-repression of transcriptional repressors by decoy and knock-out strategies, respectively.

Another way to overcome the difficulty with culturing microbes under laboratory conditions is to express cryptic or silent BGCs in a heterologous host (**Fig 1.3E**). These hosts are generally well-studied strains that grow well under laboratory conditions such as *Streptomyces coelicolor*, *Streptomyces lividans* and *Escherichia coli* etc. Heterologous expression also allows for the production of novel NPs without the need to isolate the parental strain by incorporating DNA from environmental samples into a host, as explained further below.

1.2.5 Heterologous Hosts

To facilitate the scale up of NP production the relevant BGC is cloned from the parental strain into a heterologous host. This can be done through an untargeted approach by the incorporation of random fragments of genomic DNA into a cosmid backbone. A cosmid is a hybrid plasmid and phage vector allowing for the cloning in of large fragments of genomic DNA, with the cosmid DNA being replicated as a plasmid when in bacterial cells or as a virus when in animal cells (Gupta & Yel, 2014). The inserted genomic DNA fragments can be isolated from an environmental sample without knowledge of the parental strain, from strains of interest that do not grow well under laboratory conditions, or from 'talented' strains containing many silent BGCs. Once the genomic DNA has been randomly fragmented and inserted the resulting cosmid library is then screened for BGC incorporation.

Alternatively, a targeted approach can be used by first identifying a potential antimicrobial NP BGC by sequencing. This BGC can then be synthesised synthetically one nucleotide at a time or cloned from the parental strain into a plasmid or cosmid backbone before being inserted into the heterologous host of choice. Plasmids are selected based upon their antibiotic selection marker, their suitability for scaling up of the plasmid for further assays, and their ability to either replicate within and express BGCs within the heterologous host or insert the BGC into the chromosome of the heterologous host. This later method is commonly used for inserting BGCs into actinobacteria hosts, as within this thesis, where the BGCs were incorporated into the heterologous hosts chromosome using phage-1 integration sites.

As heterologous host production involves the movement and manipulation of genetic material it is often paired with other genetic engineering techniques, such as the incorporation of constitutive promoters and BGC refactoring, as previously mentioned in Section 1.2.4. To date over one hundred actinomycete NPs have been produced using *Streptomyces* heterologous hosts (Gomez-Escribano & Bibb, 2011; Nah *et al*, 2017; Rutledge & Challis, 2015), including NPs such as actinorhodin (Malpartida & Hopwood, 1984), ansaseomycin (Liu *et al*, 2019), kinamycin (Liu *et al*,

2018), oxytetracycline (Yin *et al*, 2016) and kyamicin, the methodology for which will be used in this thesis (Vikeli *et al.*, 2020).

While the concept behind BGC expression in a heterologous host appears straightforward there are several things to consider. The first of these is the extent to which a host has been modified for ease of NP production. Many commonly used heterologous host strains have been genetically-modified to increase their ability to produce NPs. One way in which this has been accomplished is by the deletion of unwanted secondary metabolite BGCs to reduce the metabolic cost on the organism and increase the 'flux' of primary metabolites to the production of the inserted NP BGCs, as well eliminating production of native antimicrobial NPs (Baltz, 2010). Another way is by inserting integration sites to give orthogonal sites for BGC incorporation that will not affect the host chromosomes stability or primary metabolic pathways, and increase the efficiency of integration of the BGC encoding plasmid (Baltz, 2010). Additionally, increase in NP production within heterologous hosts can be accomplished by modification of ribosomal proteins through single point mutations to pleiotropically increase secondary metabolite production (Gomez-Escribano & Bibb, 2011; Palazzotto *et al*, 2019), as the translational machinery has been previously shown to significantly control the onset and extent of secondary metabolite production in some bacterial strains (Hu *et al*, 2002; Okamoto-Hosoya *et al*, 2000; Shima *et al*, 1996).

A second consideration with heterologous host use is the relatedness of the host to the parental strain. It is generally observed that greater success in heterologous NP production occurs if the host is closely related to the parental strain, as these hosts have greater biosynthetic and regulatory compatibility (Ongley *et al*, 2013; Zhang *et al*, 2018). For example, actinomycete DNA is extremely G+C rich and thus is unsuitable to be translated by *E. coli* hosts as deletions and single nucleotide polymorphisms (SNPs) would regularly occur. More closely related organisms are also more likely to produce the precursor molecules needed for production of the NP of interest. The metabolome and precursor availability can vary widely across different bacterial strains and it is important that a heterologous host is not only

able to produce precursor molecules in abundance, but that the metabolic 'flux' does not sequester them for primary metabolite production. Should a host not produce any of these precursor molecules, or their metabolic 'flux' restrains the precursors from being used in secondary metabolite synthesis, the target NP are not likely to be produced. Similarly, the regulatory background of the organism is of importance, as specific chaperone and transport proteins not encoded by the BGC but present ubiquitously within the cell could be needed for successful NP construction, with more closely related organisms being more likely to have conserved or homologous chaperone and transport proteins.

Lastly, the rate and ease of growth of a heterologous host should be considered. A good heterologous host should grow quickly and with relative ease under laboratory conditions to avoid chances of contamination, aid in high turnaround of assays and to be of any significant use where NP production is to be scaled up to large volumes.

Within this project the use of heterologous hosts and the systems that can be engineered for rapid production and diversification of NPs were investigated. Our focus was on actinomycetes, a large order of soil dwelling bacteria from which many well-known NPs are derived, including those produced from *Streptomyces* species (**Fig 1.1**).

1.3 Actinomycetes as Heterologous Hosts and Producers of NPs

Actinomycetes are prolific producers of NPs and are the source of around 45% of known bioactive bacterial metabolites (Abdel-Razek *et al*, 2020). Around 80% of these antimicrobials are produced from the genus *Streptomyces* (Berdy, 2012; Butler, 2008). Actinomycetes are Gram-positive bacteria with a high G+C content and diverse morphology. Many, including *Streptomyces* spp., are characterised by hyphal growth, the development of both substrate and aerial mycelium, and for undergoing spore formation (Chaudhary *et al*, 2013; Flardh & Buttner, 2009). While *Streptomyces* are found in a range of environments, such as the ocean (Zotchev, 2012) and desert (Gomez-Escribano *et al*, 2015), they are most common in the soil. Actinomycetes are a very diverse order of bacteria incorporating many sub orders, including Streptomycineae, Pseudonocardineae and Micromonosporineae, with over 150 genera of bacteria (Sowani *et al*, 2017), including *Streptomyces* and *Saccharopolyspora*, the two strains of heterologous hosts used within this thesis.

The *Streptomyces* genus contains a higher than average number of NP BGCs, making them a 'talented' genus of bacteria. The *Streptomyces* genus has been well-studied and there are over 550 species, with new ones being discovered regularly. Amongst them, genetically modified strains of *Streptomyces coelicolor* are widely used as heterologous hosts. The modified strains have had secondary metabolite pathways removed and their ribosomal proteins mutated to increase the success of NP production (Gomez-Escribano & Bibb, 2011). *S. coelicolor* is also easy to grow under laboratory conditions, making it an ideal heterologous host.

The *Saccharopolyspora* genus has not been as widely studied as other actinobacteria but does show great potential as a producer of chemically diverse NPs. *Saccharopolyspora* spp. produce the medically important antibiotic erythromycin (Oliylyk *et al*, 2007), the environmentally benign insecticide spinosyn (Kirst, 2010; Pan *et al*, 2011) and macrolide antibiotic sporeamicin (Yaginuma *et al*, 1992). *Saccharopolyspora* sp. KY3 produces kyamicin, a class II type B lantibiotic

which forms the basis for much of this thesis and will be discussed in more detail subsequently (Vikeli *et al.*, 2020).

As part of this thesis, we investigated a modified version of *Saccharopolyspora erythraea* that had been used as a heterologous host to previously produce erythromycin at titres up to 4 g/L. This strain is termed *S. erythraea* ISOM and has had the erythromycin BGC removed. Several of the NPs we sought to produce during this study were derived from ant associated bacteria of the genera of *Saccharopolyspora* and *Pseudonocardia* of which *S. erythraea* ISOM is more closely related to than *Streptomyces*. Thus, following the logic outlined in Section 1.2.5 that more closely related strains are more likely to succeed in producing NPs by heterologous production, it was hoped *S. erythraea* ISOM could be used to successfully produce the many different types of NPs originating from these strains.

1.4 The Different Classes of Natural Products

Actinomycete bacteria produce a wide range of NPs derived from a variety of biosynthetic pathways. Below I will outline some of the main classes of NPs produced, giving further detail on compounds discussed extensively within this thesis.

1.4.1 Ribosomally-Synthesised and Post-Translationally Modified Peptides (RiPPs)

Ribosomally-synthesised and post-translationally modified peptides (RiPPs) are derived from peptides that are ribosomally-produced and undergo subsequent post-translational modifications, such as halogenation, cyclisation and hydroxylation. These extensive modifications result in great structural diversity and over 20 subclasses of RiPPs are known (Montalbán-López *et al*, 2021). At the core of all RiPP BGCs are small structural genes that encode for a precursor peptide, a portion of which becomes the final natural product after modification. This gene and the subsequent precursor peptide are often designated as 'A' (Arnison *et al*, 2013). The precursor peptide is generally made up of two parts: a small core region, which constitutes the final product upon which post-translational modifications are performed, and a preceding leader peptide. The leader peptide is essential, containing recognition domains to which tailoring enzymes will bind (**Fig 1.5**). In rare cases this leader peptide can be attached to the C-terminus of the core peptide and is thus termed a 'follower' peptide. After maturation of the core peptide by post-translational modifications the leader peptide is then cleaved to leave the mature RiPP (Arnison *et al.*, 2013). Some RiPPs also contain a third region, termed the C-terminal recognition sequence. This is thought to be involved with excision and cyclisation (Arnison *et al.*, 2013).

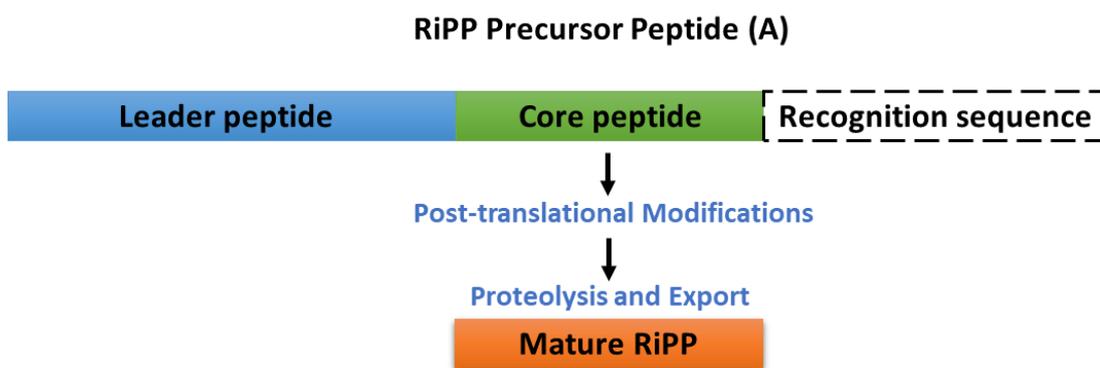


Figure 1.5: General RiPP Biosynthetic Pathway. A diagrammatic representation of the progress of a RiPP from precursor to mature peptide.

1.4.1.1 Class II Type B Lantipeptides

Lantipeptides, upon which much of this thesis is based, are RiPP NPs characterised by the presence of lanthionine (Lan) and 3-methylanthionine (MeLan) thioester bonds, formed by the attack of nucleophilic sulphur within cysteine residues to the electrophilic dehydroalanine and dehydrobutyrine moieties derived from the dehydration of serine and threonine residues (**Fig 1.7**). Lantibiotics are produced by Gram-positive bacteria and require extensive post-translational modifications. Based upon the enzymes responsible for formation of the Lan and MeLan bonds, lantipeptides can be separated into four classes (I, II, III, IV) (**Fig 1.6**) (Arnison *et al.*, 2013). Lantipeptides are also separated into two types, A or B, depending on their structure and mechanism of action.

Type A lantipeptides contain cyclic elements but have elongated structures and generally act through the formation of pores in the cytoplasmic membrane of bacterial cells, disrupting membrane integrity and causing collapse of the proton motive force (PMF) (McAuliffe *et al.*, 2001b; Moll *et al.*, 1996). The PMF is a gradient of protons maintained across the cytoplasmic membrane of bacteria that drives ATP synthesis, metabolite and ion accumulation through PMF-driven systems in the membrane, the collapse of which leads to cell death through cessation of energy-requiring reactions. Nisin, a Type A lantibiotic, has been used as a food preservative for many decades, preventing the growth of pathogenic and food spoilage Gram-positive bacteria (Gharsallaoui *et al.*, 2016).

In contrast Type B lantipeptides are small globular peptides which form complexes with membrane bound substrates, permeabilising the bacterial membrane (e.g cinnamycin) or inhibiting peptidoglycan synthesis (e.g mersacidin) (Brötz *et al*, 1998; Hosoda *et al*, 1996; McAuliffe *et al.*, 2001b). Within this study we focused on class II type B lantibiotics, exemplified by the antibiotic cinnamycin (**Fig 1.8B**). Cinnamycin selectively binds to amino phospholipid phosphatidylethanolamine (PE), a major component of many bacterial cell membranes (Vestergaard *et al*, 2019). The binding of cinnamycin has been theorised to alter the physical organisation of PE within the membrane causing membrane permeabilisation, reduced ATP-dependent calcium uptake and reduced ATP-dependent protein transport, leading to cell death (Choung *et al*, 1988; McAuliffe *et al.*, 2001b).

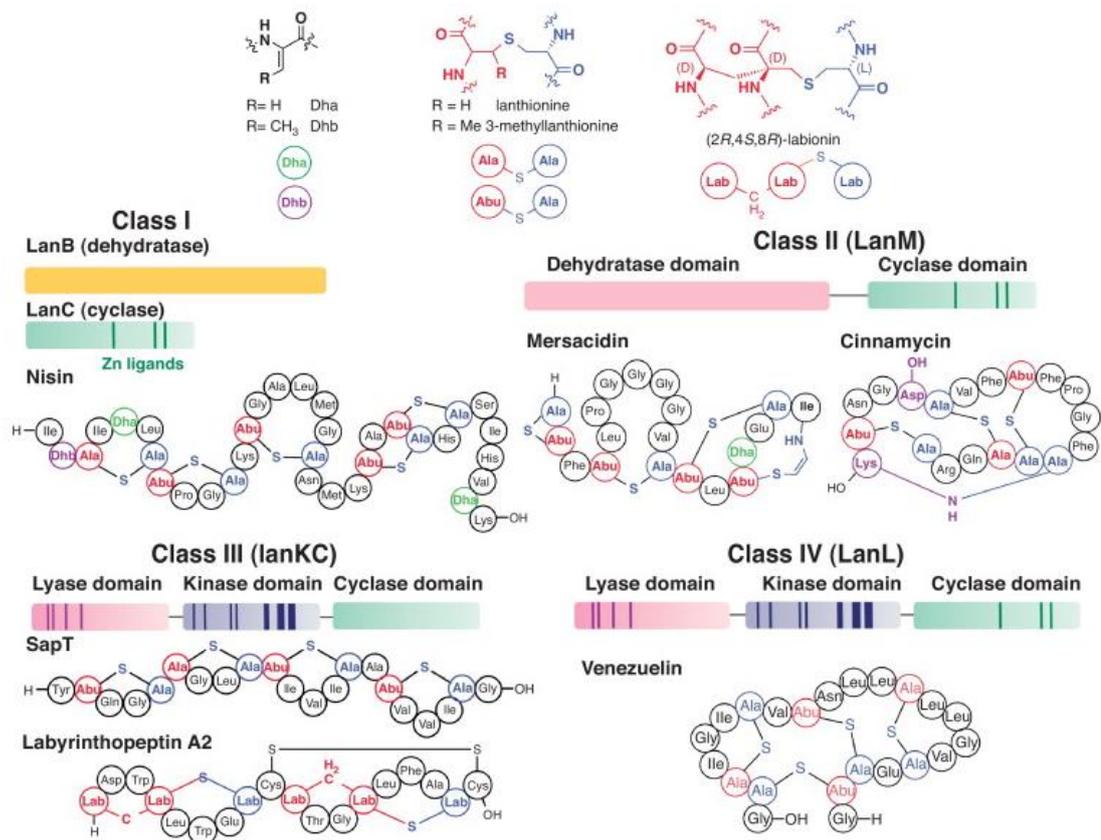


Figure 1.6: Overview of the Four Classes of Lantipeptide Biosynthesis. An overview demonstrating the differences in structure of the post-translational modification domains and structures of representative members. Class I lantibiotics (e.g. nisin) are generated by two independent dehydratase and cyclase enzymes (LanB and LanC, respectively). Class II lantipeptides (e.g. mersacidin and cinnamycin) are generated by a bifunctional enzyme with both dehydration and cyclization activities (LanM). Class III lantipeptides (e.g. SapT or labyrinthopeptin A2) are generated by a single protein that has lyase, kinase, and a cyclase domains which lacks zinc-ligands (lanKC). Some class III lantipeptides, such as labyrinthopeptin A2, also contain labionin moieties. Class IV lantibiotics (e.g. venezuelin) are generated by an enzyme that also contains lyase and kinase domains but with a LanC-like cyclase domain that contains zinc-ligands. Figure reproduced with permission from (van der Donk & Nair, 2014), *Current Opinion in Structural Biology*, 29, p. 58-66, licenced under Elsevier Ltd.

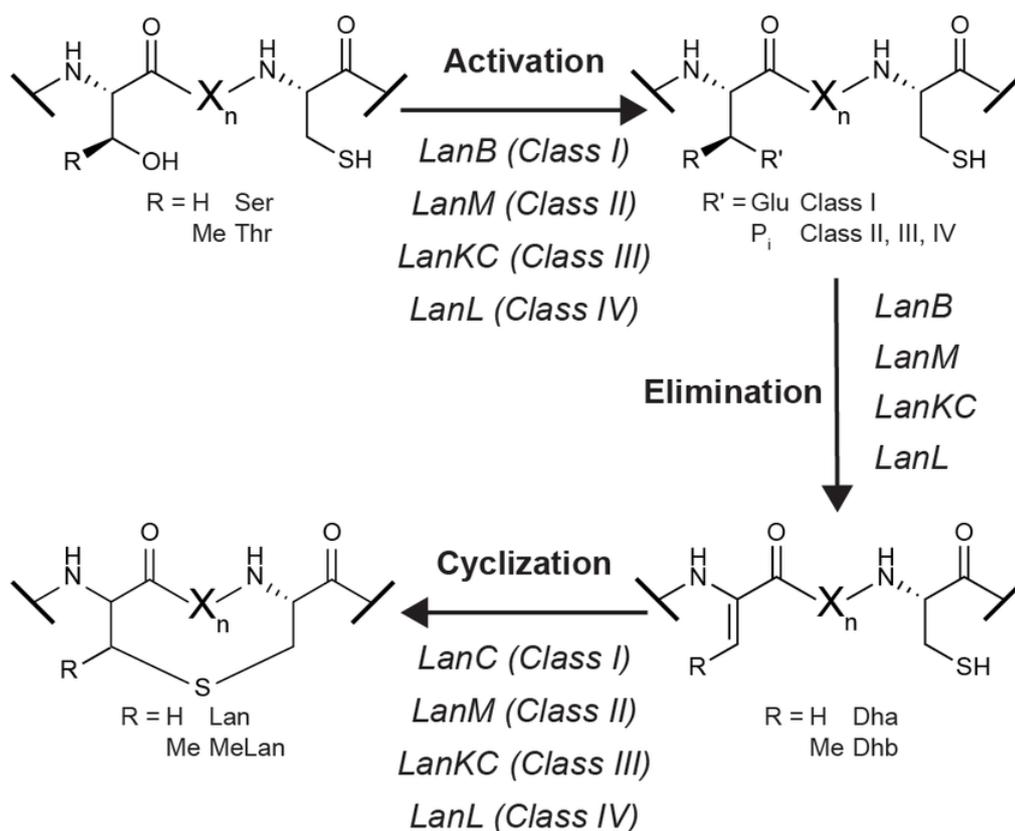


Figure 1.7: General Biosynthetic Scheme of Thioether Linkage in Lantipeptides.

Installation of lanthionine or methyllanthionine thioether crosslinks in the four different classes of lantipeptides (Class I, II, III, IV) and their respective genes. Dha: dehydroalanine, Dhb: dehydrobutyrate, Lan: lanthionine, MeLan: methyllanthionine. Adapted from, (Walker *et al*, 2020), *BMC Genomics*, 21, licenced under creative commons attribution licence 4.0 (CC BY 4.0).

Class II Type B lantipeptides also include molecules such as duramycin (Huo *et al*, 2017; Märki *et al*, 1991), ancovenin (Kido *et al*, 1983), and the recently characterised kyamicin, the BGC of which most of this work is based on (Vikeli *et al*, 2020). The genes involved in class II type B lantipeptide formation, proposed for cinnamycin biosynthesis, are included in **Table 1.1**, **Fig 1.8** (Widdick *et al*, 2003).

Firstly, the lanthionine synthetase CinM, catalyses the elimination of water from two serine residues, at position four and six (S4, S6), and two threonine residues at position eleven and eighteen (T11, T18) of the 19 amino acid core peptide region, forming dehydroalanine (Dha) and dehydrobutyrate (Dhb) residues, respectively. CinM then also catalyses crosslinking of the electrophilic Dha and Dhb β -carbon atoms with the nucleophilic sulphur atoms of cysteine residues at position one, five,

and fourteen (C1, C5, C14) forming a thioether link. This means Dha becomes an S-linked alanine (Ala) residue and Dhb becomes an S-linked L-alpha-aminobutyric acid (Abu) residue, forming the Lan and MeLan bridges characteristic of lantipeptides. Following this the hydroxylase CinX acts upon the aspartate residue at position fifteen (D15), introducing a β -hydroxy group. The lysinoalanine synthetase CinN then catalyses the formation of a single lysinoalanine bridge between Dha6 and the side chain amine moiety of the lysine residue at position nineteen (K19) (**Fig 1.8**) (Ökesli *et al*, 2011). Once modification to the core peptide is complete the leader peptide is cleaved proteolytically by type 1 signal peptidases of the general secretory (sec) pathway. Following this the mature peptide is transported across the cytoplasmic membrane by a two-component ATP-binding cassette (ABC) transporter system, encoded for by CinH and CinT (McAuliffe *et al.*, 2001b).

Further to this the cinnamycin BGC also encodes a two-component membrane spanning regulator (CinR and CinK) which is capable of sensing exogenous cinnamycin. This is thought to be involved in a positive feedback loop, discussed further in Chapter 3, where cinnamycin biosynthesis is induced at a low level by a stress response. This cinnamycin then activates a signalling pathway, resulting in the transcription of *cinL*, which encodes a phosphatidylethanolamine (PE) methyltransferase responsible for providing immunity to cinnamycin, and *cinR1*, which encodes a SARP transcriptional regulator, expression of which activates high levels of cinnamycin biosynthesis (O'Rourke *et al*, 2017).

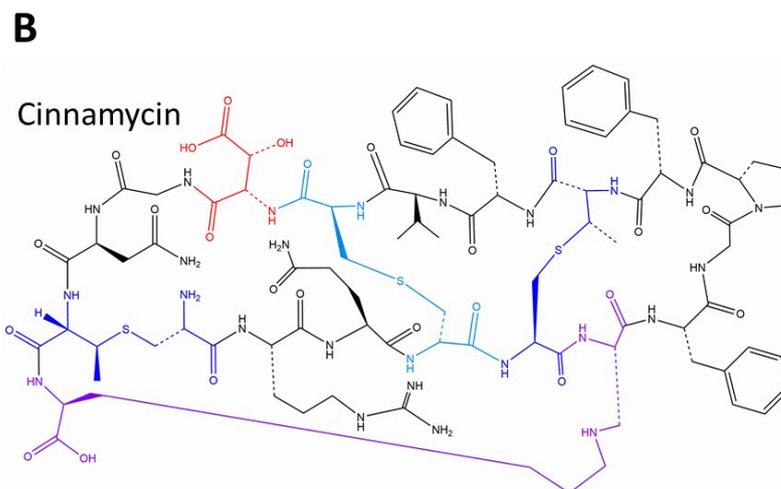
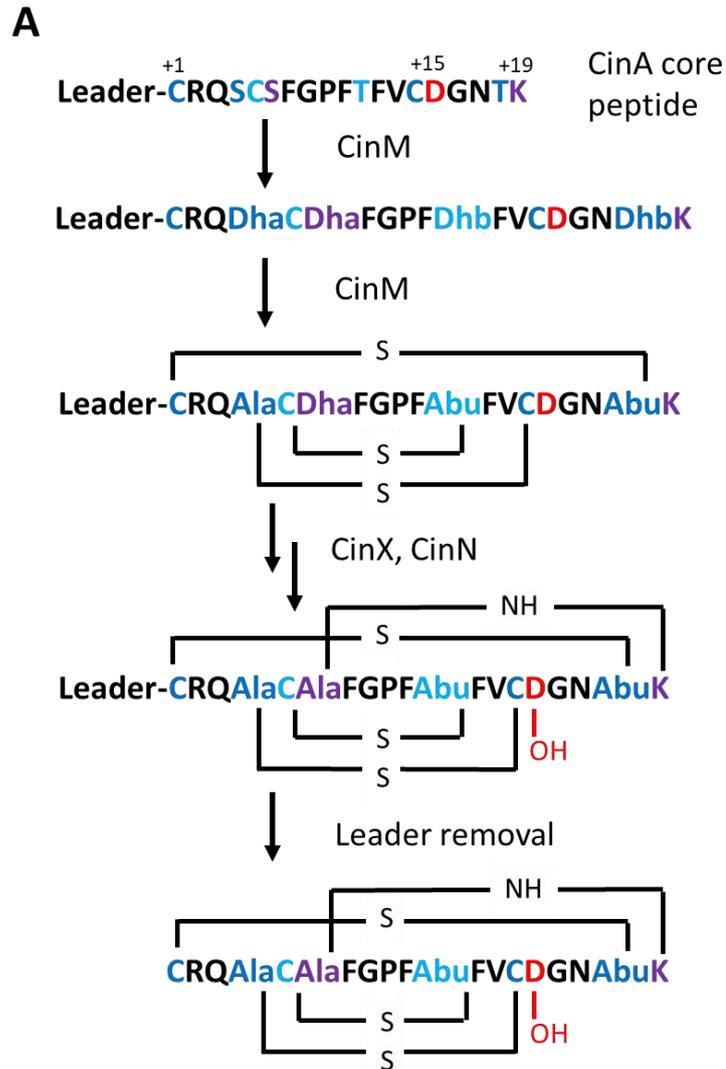


Figure 1.8: Biosynthesis and Structure of Cinnamycin. (A) Biosynthesis of cinnamycin, including formation of lanthionine, methyl-lanthionine and lysinoalanine bonds, and hydroxylation of aspartate. (B) Chemical structure of cinnamycin. Lanthionine residues = light blue, methyl-lanthionine residues = dark blue, lysinoalanine residues = purple, hydroxylated aspartate = red.

Table 1.1: Class II Type B Lantipeptide Biosynthesis Genes

Peptide	Cinnamycin Equivalent	Function	Description
LanA	CinA	Biosynthetic	Precursor peptide; including core and leader peptide regions
LanN	CinN	Biosynthetic	Lysinoalanine synthetase; formation of lysinoalanine bridge
LanM	CinM	Biosynthetic	Lanthionine synthetases; formation of lanthionine and methyl-lanthionine bridges
LanX	CinX	Biosynthetic	2-oxoglutarate (2OG, also known as α -ketoglutarate) dependent di-oxygenase (Gao et al, 2018; Loenarz & Schofield, 2011); addition of OH to D15
LanT	CinT	Export	Two-component ABC transporter; exports lantibiotic across cell membrane
LanH	CinH	Export	Two-component ABC transporter; exports lantibiotic across cell membrane
LanR	CinR	Regulation	Two-component regulator; response regulator of lanR1 and lanorf11
LanK	CinK	Regulation	Two-component regulator; membrane associated sensor kinase
LanR1	CinR1	Regulation	Streptomyces antibiotic regulatory protein (SARP); activator of putative biosynthetic operon
LanL	CinL	Immunity	Methyltransferase; methylates cell membrane providing immunity (O'Rourke <i>et al.</i> , 2017)
Lanorf11	Cinorf11	Unknown	Conserved, functionally uncharacterised protein with a domain similar to pfam family DedA (Widdick <i>et al.</i> , 2003)

The first reported class II type B lantibiotic was cinnamycin from *Streptomyces cinnamoneus* DSM 40646. It is active against Gram-positive bacteria and acts by binding the membrane amino phospholipid PE. This induces trans-bilayer lipid movement, compromising the structural integrity of the membrane and leading to cell lysis (Hullin-Matsuda *et al*, 2016). Further to this the related lantibiotic duramycin, which differs from cinnamycin by a single amino acid change (R2 to K2) and also has antibacterial activity, has also been shown to stimulate chloride secretion in human epithelial cells from cystic fibrosis (CF) patients through a number of mechanisms, such as increasing membrane permeability and activating calcium-activated chloride conductance (CaCC) channels. Duramycin has reached phase II clinical trials for the treatment of CF (Oliynyk *et al*, 2010). Further to this, it has been reported to block TIM1, a cofactor associated with Zika virus (ZIKV), and thus could have potential as an antiviral (Tabata *et al*, 2016).

The availability of whole genome sequencing data has led to the discovery of many class II type B lantipeptide BGCs. However, despite the clinical potential of these lantipeptides, very few have been isolated and their structure and/or antibacterial activity characterised to any level. Within this study we produced, diversified, validated the structure and assessed the antibacterial activity of novel class II type B lantipeptides using a synthetic cassette system based upon the BGC of the lantipeptide kyamicin (Vikeli *et al.*, 2020), wherein we replaced the core peptide region of *kyaA* with homologs, and used *S. coelicolor* as a heterologous host.

1.4.2 Non-Ribosomal Peptides

Non-ribosomal peptides (NRPs), as the name suggests, do not derive from peptides directly translated by the ribosome but are instead synthesised by non-ribosomal synthetases (NRPSs). NRPSs operate independently of the ribosome and are large multimodular assembly line enzymes that are not restricted to proteinogenic amino acid substrates and can utilise several hundred structurally varied monomers, producing a diverse family of NPs (Fischbach & Walsh, 2006; Marahiel, 2016; Payne *et al*, 2017). NRPS derived pharmaceuticals in clinical use include antibiotics such as vancomycin and daptomycin, and immunosuppressants such as cyclosporine (**Fig 1.9**).

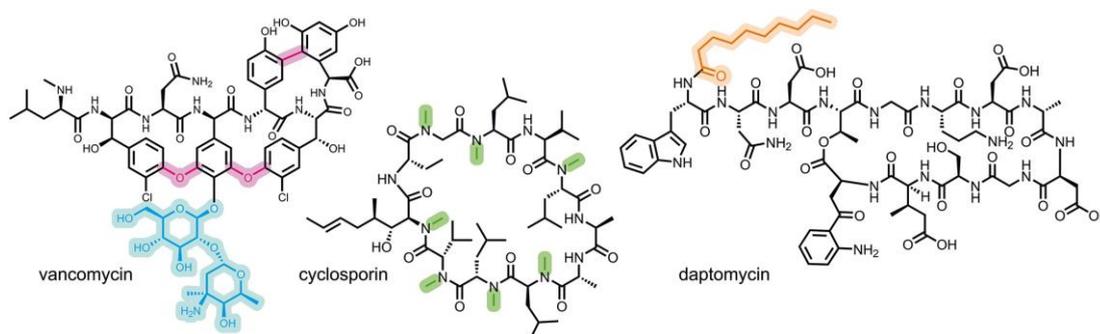


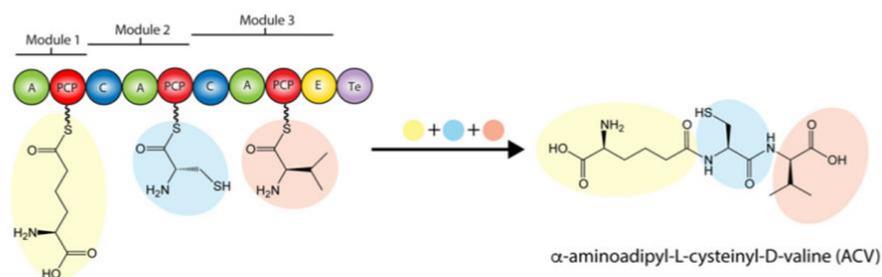
Figure 1.9 Non Ribosomal Peptides (NRPs) that are Used Clinically. Modifications like glycosylation (blue), side-chain cross-linking (purple), *N*-methylation (green), and lipidation (orange) increase the structural diversity of NRPs. Figure reproduced with permission, (Kries, 2016), *Journal of Peptide Science*, 22, p. 564-570, licenced to European Peptide Society and John Wiley & Sons, Ltd.

NRPSs have a modular architecture with biosynthesis proceeding in an assembly line fashion where each module is responsible for the selection and activation of an amino acid substrate followed by its incorporation into a growing peptide chain. Depending on the architecture of these modules NRPSs are broadly separated into three types: Type A linear NRPS, where each module is connected to the last to form one large multi-enzymatic cluster, Type B iterative NRPS, where modules do not form multi-enzymatic clusters and remain separated but in which peptide chain extension still occurs in a linear fashion, and Type C non-linear NRPS, in which modules are also iterative but peptide chain elongation does not occur in a linear fashion (**Fig 1.10**). Each module contains a core set of individual catalytic domains,

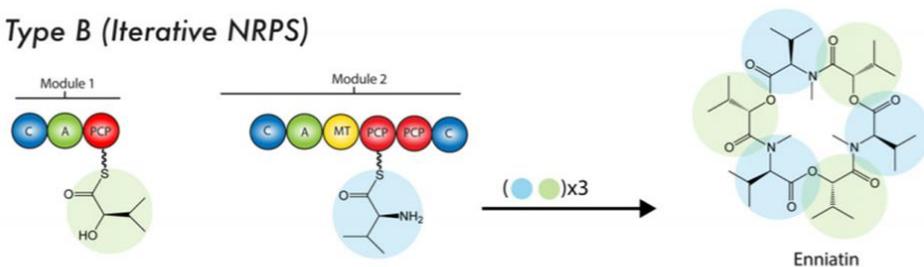
with the minimum set of domains being an adenylation (A), thiolation or peptidyl carrier protein (T/PCP) and condensation (C) domain.

To begin with an initiation module activates the first amino acid via the A domain using a two-part ATP-dependent reaction to give an acyl-phosphoric acid anhydride. This is then loaded onto the serine-attached 4'-phospho-pantethine (4'PP) sidechain of the PCP domain by a PCP catalysed reaction, creating a thioester bond. Following this, each elongation module loads a specific amino acid onto its PCP domain in an analogous way to the starting module. The C domains then catalyse a reaction between the thioester group of the peptide chain from the previous/upstream module and the amino group of the current/downstream module's amino acid, forming a peptide bond. The extended peptide is left attached to the PCP domain of the downstream module and extension of the peptide chain continues until reaching the termination module (Martínez-Núñez & López, 2016). The termination module contains either a thioesterase (TE) or modified C domain (the R domain) which hydrolyses the peptide chain from the PCP domain, causing cyclisation into lactams or lactones, or reducing the thioester bond to an aldehyde or alcohol group, respectively (Fischbach & Walsh, 2006). Lastly, in addition to the essential A, PCP and C domains, modules can contain a number of tailoring domains performing halogenation, cyclisation, methylation, epimerisation etc. The specificity of the amino acid sequence, the variability in the elongation modules and the presence of the tailoring domains result in a diverse range of NRPS NPs.

Type A (Linear NRPS)



Type B (Iterative NRPS)



Type C (Non-linear NRPS)

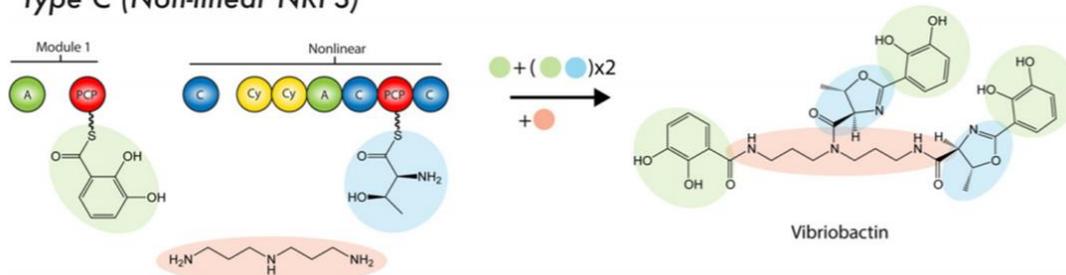


Figure 1.10: Illustration of NRPS Biosynthesis. A: adenylation domain; PCP: peptidyl carrier protein; C: condensation domain; Te: thioesterase domain; E: epimerase; MT: methyltransferase; Cy: cyclisation domain. Adapted from (Guzmán-Chávez *et al*, 2018), *Frontiers in Microbiology*, 9, licenced under CC BY 4.0

1.4.3 Polyketides

Polyketides are a diverse class of molecules generated by the successive condensation of short chain carboxylic chains. Several classes of polyketide synthases (PKSs) are responsible for their biosynthesis and these are separated into three main groups according to their organisation (Type I, II, III) (**Fig 1.11**). Type I PKSs comprise of large, multimodular enzymes similar in architecture to NRPSs, in which each module adds a new extender unit to a growing polyketide chain (**Fig 1.11A**) (Weissman, 2015).

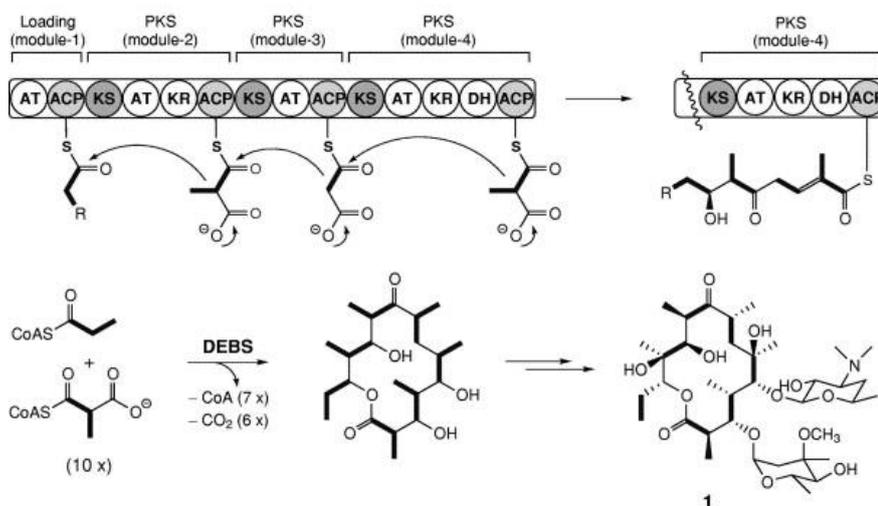
A module of a type I PKS must minimally contain an acyltransferase (AT), an acyl carrier protein (ACP) and a KS domain for polyketide chain elongation to occur. To begin with a starter module, comprising of a single AT and ACP domain only, loads a short coenzyme A (CoA) activated carboxylic acid chain substrate which is selected for by the AT domain. This is then transferred onto the sulphur atom of the phosphopantetheine moiety attached to the ACP domain, resulting in a thioester linkage. The KS in the downstream extender module then receives the polyketide chain from the upstream module and a decarboxylative condensation reaction occurs with an acyl extender unit recruited by the adjacent AT domain and bound to the module's ACP domain with a thioester linkage, as before (Weissman, 2015). The number of extension modules varies and each module can contain a number of tailoring domains, such as ketoreductases (KR), dehydratases (DH) or enoylreductases (ER), which turn the β -ketone group of the substrate to an alcohol, eliminate the β -hydroxyl group to form an α - β double bond, or create a fully reduced β -methylene group, respectively. Lastly the polyketide chain is released by a termination module containing a TE, after which further tailoring steps, such as cyclisation, are carried out to create the fully matured compound. The presence of tailoring domains, the specificity of the extender unit, the variability in the number of extension modules and the post-tailoring enzymes result in a great diversity of polyketide NPs.

Type II PKSs are comprised of several monodomain proteins which act iteratively to cause several rounds of polyketide chain extension, followed by removal of the

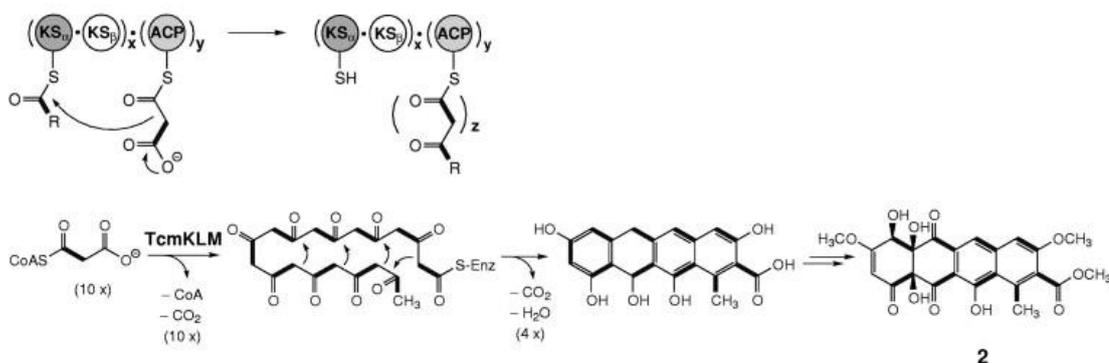
polyketide chain from the PKS often coupled with cyclisation and subsequent tailoring steps to generate the final molecule (Hertweck *et al*, 2007). Typically, a 'minimal type II PKS' consisting of a ketosynthase (KS) α/β heterodimer, in which only the KS- α is active, and an ACP iteratively elongates a poly- β -ketone chain linked to the ACP from malonyl-CoA units. Once the appropriate chain length has been reached, as determined by the inactive KS- β , also called the chain length factor (CLF), the full-length poly- β -ketone chain is then subjected to modification by other PKS component enzymes including KR, DH and cyclases (CYCs), to create the initial polyketide molecule which can then be modified by a range of tailoring enzymes (**Fig 1.11B**) (Dreier & Khosla, 2000; Ogasawara *et al*, 2015).

Type III PKSs are comprised of KS homodimers which can use acyl-CoA rather than acyl-ACP to anchor the polyketide chain. Each homodimer can carry out various reactions in a single catalytic centre, including priming, extension and cyclisation iteratively of one or several extender units to form the polyketide product which then undergoes tailoring steps to produce a mature polyketide (**Fig 1.11C**) (Katsuyama & Ohnishi, 2012; Yu *et al*, 2012).

(a) Type I PKS (noniterative)



(b) Type II PKS (iterative)



(c) Type III PKS (ACP-independent & iterative)

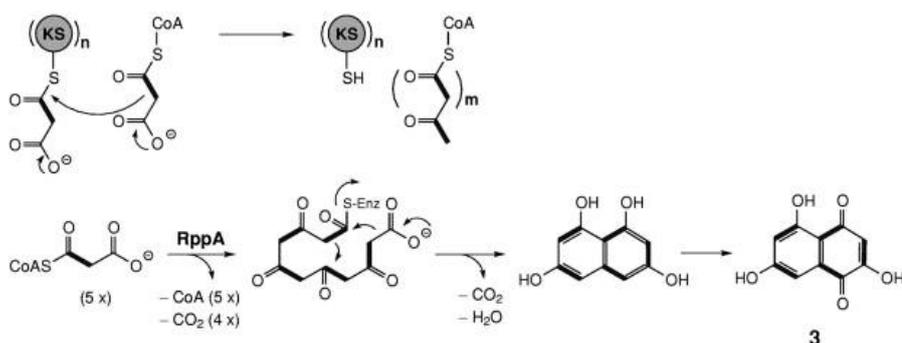


Figure 1.11: Structures and Mechanisms of Bacterial Type I, II and III PKSs. (a) Type I non-iterative PKS exemplified by the erythronolide synthase (6-Deoxyerythronolide B Synthase, DEBS) for erythronolide (1) biosynthesis. (b) Type II iterative PKS exemplified by the tetracenomycin C polyketide synthase genes (TcmKLM) for tetracenomycin (2) biosynthesis. (c) Type III ACP-independent iterative PKS exemplified by the type III PKS gene (RppA) for flavolin (3) biosynthesis. AT: acetyltransferase domain; ACP: acyl carrier protein domain; KS: ketosynthase domain; DH: dehydratase domain; KR: keto reductase domain. Atoms incorporated intact from the acyl CoA precursors into the resultant polyketides are shown in bold. Reproduced with permission, (Shen, 2003), *Current Opinion in Chemical Biology*, 7(2), p. 285-295, licenced to Elsevier Science Ltd.

1.4.3.1 Formicamycin Biosynthesis: an Exemplar Type II PKS Pathway

The formicamycins are pentacyclic polyketides produced by a type II PKS pathway in *Streptomyces formicae* KY5 that was isolated from the cuticle of an African fungus growing plant-ant *Tetraponera penzigi* (Qin *et al*, 2017). Fasamycin polyketides are precursors of the formicamycins and differ from them due to an aromatic C-ring structure and lack of any formal chiral centres (Feng *et al*, 2012). The formicamycins are highly modified compounds derived from these fasamycin precursors via a two-step ring expansion-contraction mechanism catalysed by gene product ForX and ForY, resulting in the formation of a non-aromatic C-ring and chiral centres at C10 and C19 (Qin *et al*, 2020). All formicamycins discovered so far are halogenated, as are the majority of fasamycin congeners.

The formicamycin BGC consists of 24 genes expressed on nine transcripts controlled by three cluster situated regulators: two MarR family transcriptional regulators, one of which represses the expression of the majority of biosynthetic genes (ForJ), the other activating the expression of transporter gene *forAA* (ForZ); and a two component transcriptional activator required for formicamycin biosynthesis (ForGF) (Devine *et al*, 2021). The genes ForABC comprise the minimal PKS, which produce a tridecaketide intermediate. This then goes through several additional tailoring steps, including cyclisation and dehydration (ForD, ForL, ForR) and several rounds of halogenation (ForV) and O-methylation to create multiple different formicamycins. Further speculative tailoring steps include hydration (ForN), decarboxylation (ForQ) and methylation (ForM/W) (**Fig 1.12**) (Devine *et al.*, 2021; Qin *et al.*, 2017).

Gene	Annotation
<i>forN</i>	Acyl hydrolase
<i>forM</i>	Methyltransferase
<i>forL</i>	PKS cyclase
<i>forK</i>	Na ⁺ /H ⁺ exchanger
<i>forJ</i>	MarR family transcriptional regulator
<i>forI</i>	ACC biotin carboxylase
<i>forH</i>	ACC carboxyl transferase
<i>forG</i>	Sensor histidine kinase
<i>forF</i>	LuxR family response regulator
<i>forE</i>	ACC biotin carboxy carrier protein
<i>forD</i>	PKS cyclase/dehydratase
<i>forC</i>	PKS ACP
<i>forB</i>	KS _β
<i>forA</i>	KS _α
<i>forR</i>	Putative PKS cyclase/chaperone
<i>forS</i>	PKS chaperone
<i>forT</i>	Methyltransferase
<i>forU</i>	Putative PKS cyclase/chaperone
<i>forV</i>	Halogenase
<i>forW</i>	Methyltransferase
<i>forX</i>	Monoxygenase
<i>forY</i>	Oxidoreductase
<i>forZ</i>	MarR family transcriptional regulator
<i>forAA</i>	Multidrug resistance protein

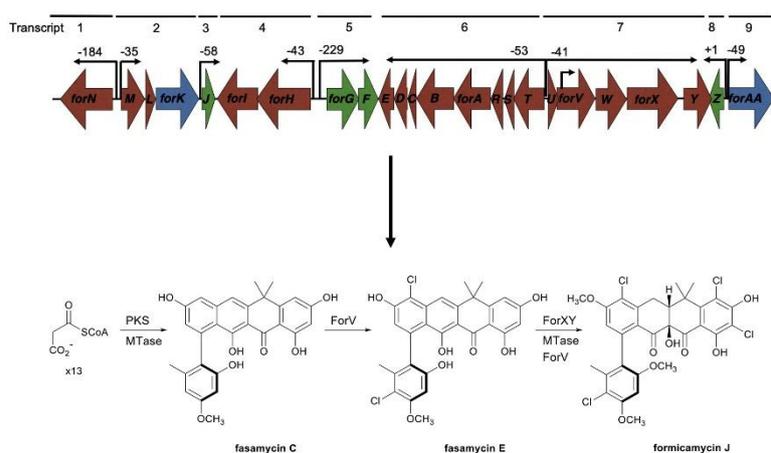


Figure 1.12: Formicamycin BGC and Gene Functions. Twenty-four genes on nine transcripts; red=biosynthetic genes, blue=transporters, green=regulatory genes. Ten transcription start sites were identified through cappable RNA sequencing. The table lists all genes with known/speculative function. Formicamycins are formed from fasamycins by ForX catalysed hydroxylation and ring expansion leading to a lactone intermediate. Reductive ring contraction is then catalysed by the flavin-dependant oxidoreductase ForY to form the formicamycin backbone. Adapted from (Devine *et al.*, 2021), *Cell Chemical Biology*, 28, 4, p. 515-523, licensed under CC BY 4.0.

Both formicamycins and fasamycins are active against MRSA and VRE but display no effect against Gram-negative bacteria or fungi. The minimum inhibitory concentration (MIC) for several formicamycins was determined and it was observed that the test strains did not acquire spontaneous resistance. Formicamycin I (**Fig 1.13**), which had a low MIC for MRSA and VRE (<2.5 and 1.25 μ M respectively), was applied to MRSA over 20 generations at sub-inhibitory concentrations (half MIC) to promote the development of spontaneous resistant mutants. There was no observed increase in MIC of formicamycin against the MRSA over the course of the experiment, indicating a high barrier for the selection of resistant mutants (Qin *et al.*, 2017).

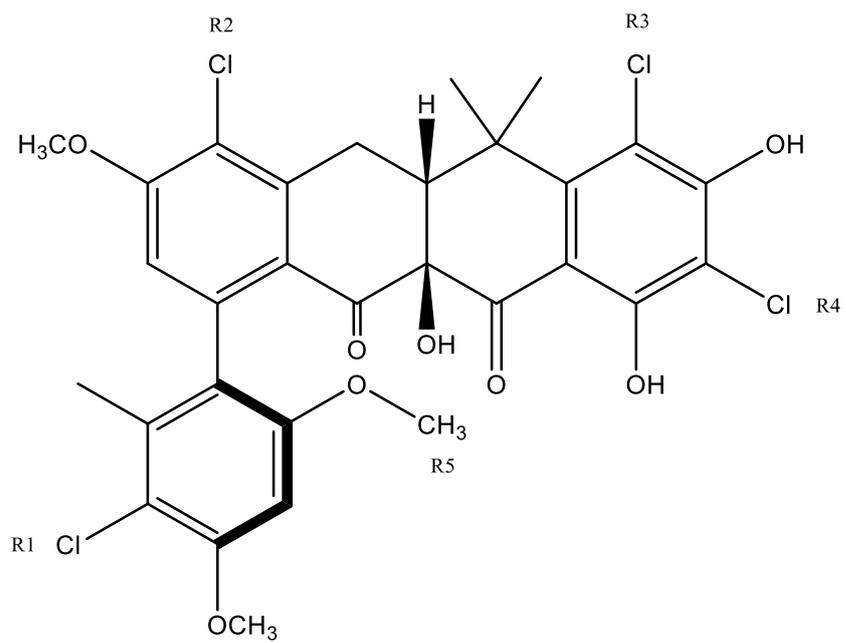


Figure 1.13: Formicamycin I, functional groups R1 to R5 are indicated.

1.5 Project Objectives

This study explores the production and diversification of different types of NPs using heterologous expression in both *Streptomyces* and *Saccharopolyspora* host systems.

Firstly, by using a two-part synthetic cassette platform created from the BGC of the class II type B lantipeptide kyamicin, in conjunction with heterologous host *S. coelicolor* M1146, several lantipeptides identified within the publicly available databases were produced, showcasing the ease of use of the system. Rapid diversification of lantipeptides was then achieved through the creation of a four-part synthetic library, further showcasing the flexibility of the platform. Select novel lantipeptides were then purified, their bioactivity characterised against several indicator strains and the structure of one novel lantipeptide elucidated. Lastly, a comparison of lantipeptide expression by the heterologous system and parental strains is presented, with close attention paid to post-translational cyclisation, lysinoalanine bond formation and hydroxylation.

Secondly, the expression of several unrelated BGCs that have been cloned into phage-artificial chromosomes (PACs) was undertaken, with a focus on inducing their expression in the host *S. erythraea* ISOM. *S. erythraea* ISOM was evaluated for its potential as a heterologous host for actinomycete NPs, comparing it to the commonly used actinomycete heterologous host *S. coelicolor*.

Finally, the effect that heterologous host choice can have on polyketide congener production profiles was compared, using formicamycin and fasamycin production as a model system. Changes in metabolic profile and titre between the heterologous hosts were observed, and a number of novel glycosylated fasamycin compounds unique to the heterologous system were identified.

Chapter 2:

Materials and

Methods

2. Materials and Methods

2.1 Lab Supplies

Reagents were obtained from the following commercial suppliers; Sigma-Aldrich, Merck, and Thermo Fisher Scientific. Solvents were obtained from Fisher Scientific in HPLC grade. Further commercially available kits, reagents, enzymes and DNA markers are shown in **Table 2.1**.

Table 2.1: Lab Supplies and Kits

Supplier	Enzyme and Kits
New England Biolabs (NEB)	1 kb and 100 bp DNA ladders Gibson Assembly® Master Mix Q5® High-Fidelity DNA Polymerase T4 DNA Ligase CutSmart® Buffer and appropriate Restriction Endonuclease Enzymes dNTPS
MP Biomedicals	FastDNA™ SPIN KIT
Promega	GoTaq® G2 Green Master Mix
Qiagen	QIAquick® Gel Extraction Kit QIAquick® PCR Purification Kit QIAprep® Spin Miniprep Kit
Bioline, Meridian Bioscience	MyTaq™ Red Mix
Sigma-Aldrich	PhasePrep™ BAC DNA Kit
Eurofins Genomics	Mix2Seq overnight Kit

2.2 PACs, Plasmids, Primers and Strains

2.2.1 Vectors

2.2.1.1 PACs

Phage-1 Artificial Chromosomes (PACs) were created by Bio S&T Inc (Montreal, Canada) using the pESAC13 vector. All PACs used in this work are listed in **Table 2.2**.

Table 2.2: PACs used during this work

PAC	Description/Genotype	Reference
6_19B	pESAC13 containing nystatin Ps1 BGC	(Holmes <i>et al.</i> , 2016)
2_5K	pESAC13 containing nystatin Ps2 BGC	(Holmes <i>et al.</i> , 2016)
215G	pESAC13 containing the formicamycin BGC from <i>Streptomyces formicae</i>	(Qin <i>et al.</i> , 2017)
215G Δ forJ	PAC 215G with the <i>forJ</i> repressors removed from the <i>for</i> BGC	(Devine <i>et al.</i> , 2021)
21A	pESAC13 containing the antascomycin BGC from <i>Micromonospora sp.</i>	Qin Zhiwei, JIC
21B	pESAC13 containing the antascomycin BGC from <i>Micromonospora sp.</i>	Qin Zhiwei, JIC
14M	pESAC13 containing the anthracimycin BGC from <i>Streptomyces</i> T676	(Alt & Wilkinson, 2015)
7G	pESAC13 containing the thiazolpeptide BGC from <i>Streptomyces</i> T676	Elena Stoyanova, JIC
24J	pESAC13 containing a lantipeptide BGC from <i>Streptomyces</i> T676	Charlie Owen, JIC
3D	pESAC13 containing a NRPS BGC from <i>Streptomyces</i> T676	Natalia Miguel Vior, JIC
TB73	pESAC13 containing a type 1 PKS macrolactam from <i>Streptomyces</i> T676	Thomas Booth, JIC

2.2.1.2 Plasmids

Plasmids used and constructed during this work are listed in **Table 2.3**. The synthetic operons pAMA2, pAMA3 and pAMA5 were ordered from GenScript® (Netherlands).

Table 2.3: Plasmids used during this work

Plasmid	Description/Genotype	Reference
pGP9	pSET152 derived ϕ BT-based integrative expression vector	(Gregory <i>et al</i> , 2003)
pSET152	ϕ C31 attP-conjugative vector	(Gregory <i>et al.</i> , 2003)
pIJ102567	oriT, ϕ BT1, attB-int, HygR, ermEp*	(Hong <i>et al</i> , 2005)
pESAC13	Derivative of pPAC-S1, ϕ C31-based large fragment integrative expression vector.	(Jones <i>et al</i> , 2013)
pBF3	A SV1-attP based integration expression vector	(Fayed <i>et al</i> , 2014)
pRF10	A TG1-attP based integration expression vector	(Fayed <i>et al</i> , 2015)
pTS1	pME3087 suicide vector; <i>ColE1</i> replicon, <i>IncP-1</i> , Mob; Tc ^R , modified with a <i>sacB</i> gene for counter-selection and an expanded multiple cloning site	Thomas Scott, JIC
pIJ773	PCR-targeting vector for mutagenesis, - <i>aac(3)IV oriT bla</i> AprR, flippase recognition target sites (FRT)	(Gust <i>et al</i> , 2003)
pIJ790	Vector for Redirect® mutagenesis, λ -RED (<i>gam</i> , <i>bet</i> , <i>exo</i>), <i>cat</i> , <i>araC</i> , <i>rep101ts</i>	(Gust <i>et al.</i> , 2003)
pADW11	An integration vector with ermE* and pSAM2 integration site.	David Widdick, JIC

pBF3TetR	pBF3 with tetracycline resistance genes from pTS1 inserted	This work
pGP9_SARP	pGP9 with antascomycin SARP transcriptional activator	Qin Zhiwei, JIC
pGP9_LuxR	pGP9 with antascomycin LuxR transcriptional activator	Qin Zhiwei, JIC
pGP9_AraC	pGP9 with antascomycin AraC transcriptional activator	Qin Zhiwei, JIC
pRF10_LuxR	pRF10 with antascomycin LuxR transcriptional activator	This work
pBF3TetR_AraC	pBF3TetR with antascomycin AraC transcriptional activator	This work
pWDW63	pSET152 containing the biosynthetic genes from the kyamycin BGC	(Vikeli <i>et al.</i> , 2020)
pWDW70	pWDW63 with the <i>kyaA</i> core peptide region removed	(Vikeli <i>et al.</i> , 2020)
pEVK6	pIJ10257 containing the <i>kyaR1</i> and <i>kyaL</i> genes	(Vikeli <i>et al.</i> , 2020)
pAMA1	pIJ10257 containing the entire activator and regulatory region of the kyamycin BGC	This work
pAMA2	pSET152 containing the <i>Streptomyces roseoverticillatus</i> lantibiotic BGC biosynthetic genes with the <i>lanA</i> core peptide removed	This work
pAMA3	pIJ10257 containing the <i>Streptomyces roseoverticillatus</i> lantibiotic BGC <i>lanR1</i> gene and a truncated <i>lanL</i> gene	This work
pAMA4	pADW11 carrying the <i>kyaL</i> gene	This work
pAMA5	pIJ10257 containing the <i>Streptomyces</i>	This work

	<i>roseovorticillatus</i> lantibiotic BGC <i>lanR1</i> gene and <i>lanL</i> gene	
pAMA6	pAMA2 with a stem loop attenuation between the <i>lanA</i> and <i>lanM</i> genes	This work
L_01 to L_35	pWDW70 with lantibiotic core peptides L_01 to L_35 inserted	This work
Lr_00 to Lr_35	pAMA2 with lantibiotic core peptides L_00 to L_35 inserted	This work
p6Lr_00	pAMA6 with the core peptide for kyamicin inserted	This work
p6Lr_14	pAMA6 with the core peptide from <i>Streptomyces</i> <i>roseovorticillatus</i> lantipeptide BGC inserted	This work

2.2.1.3 Synthetic Library

Synthetic libraries of lantibiotic core peptides, based on the kyamicin core peptide sequence of 19 amino acids, were ordered from GenScript® (Netherlands). Four libraries were created, each with a single amino acid substitution. In equal ratio the amino acids listed were replaced with all other 19 amino acids: A2, A7, F12, A13.

2.2.2 Primers

Primers used in this study are listed in Appendix 9.1.

2.2.1 Bacterial Strains and Growth Conditions

Saccharopolyspora and *Streptomyces* strains were grown on SF+M media with appropriate antibiotics at 30 °C unless stated otherwise. *E. coli* strains were grown on LB medium with appropriate antibiotics at 37 °C unless stated otherwise. *E. coli* strains used within this work are listed in **Table 2.4**. *Streptomyces* strains used in this work are listed in **Table 2.5**. *Saccharopolyspora* strains used in this work are listed in **Table 2.6**. Bioassay indicator strains are listed in **Table 2.7**.

Table 2.4: *E. coli* strains used in this work

Strain	Description/Genotype	Reference
<i>E. coli</i> DH5 α	Cloning strain; F ⁻ , <i>endA1</i> , <i>ginV44</i> , <i>thi-1</i> , <i>recA1</i> , <i>relA1</i> , <i>gyrA96</i> , <i>deoR</i> , <i>nupG</i> , $\phi 80lac\Delta(lacZ)M15$, $\Delta(lacIZYA-argF)U169$, <i>hsdR17(r_K⁻m_K⁺)</i> , λ -	(Hanahan, 1983)
<i>E. coli</i> ET12567/pUZ8002	<i>E. coli</i> ET12567 a methyltransferase deficient strain containing plasmid pUZ8002, a not self-transmissible plasmid which can mobilize other plasmids.	(Flett <i>et al</i> , 1997; MacNeil <i>et al</i> , 1992)
<i>E. coli</i> TOP10/pR9604	Cloning strain; F- <i>mcrA</i> $\Delta(mrr-hsdRMS-mcrBC)$ $\phi 80lacZ\Delta M15$ $\Delta lacX74$ <i>recA1</i> <i>araD139</i> $\Delta(araleu)7697$ <i>galU</i> <i>galk</i> <i>rpsL</i> (Str ^R) <i>endA1</i> <i>nupG</i> , containing a derivitave of pUB307 a not-self-transmissible plasmid which can mobilize other plasmids	Thermo Fisher Scientific, (Piffaretti <i>et al</i> , 1988)
<i>E. coli</i> DH10B	Cloning strain; F- <i>mcrA</i> $\Delta(mrr-hsdRMS-mcrBC)$ $\phi 80lacZ\Delta M15$ $\Delta lacX74$ <i>recA1</i> <i>endA1</i> <i>araD139</i> $\Delta(araleu)7697$ <i>galU</i> <i>galk</i> λ - <i>rpsL</i> (Str ^R) <i>nupG</i>	Thermo Fisher Scientific, (Durfee <i>et al</i> , 2008)
<i>E. coli</i> NEB [®] 5-alpha Competent (High Efficiency)	DH5 α derived T1 phage resistance and <i>endA</i> deficient cloning strain. Competent for use with Gibson assembly mixture.	NEB
<i>E. coli</i> C2925	<i>dam</i> and <i>dcm</i> methyltransferase deficient <i>E. coli</i> .	NEB
<i>E. coli</i> BW25113/pIJ790	<i>Mutagenesis strain</i> containg plasmid <i>pIJ790</i> ; λ -, $\Delta(arad-araB)567$,	(Datsenko & Wanner, 2000; Gust <i>et al.</i> , 2003)

	<i>ΔlacZ4787(::rrnB-4), lacI_{p4000}(lacIQ), rpoS369(Am), rph-1, Δ(rhaD-rhaB)568, hsdR514, Cm^R</i>	
--	---	--

Table 2.5: *Streptomyces* strains used in this work

Strain	Description/Genotype	Reference
<i>Streptomyces coelicolor</i> M1152	Antibiotic producing superhost [$\Delta act \Delta red \Delta cpk \Delta cda$, rpoB(C1298T)]	(Gomez-Escribano & Bibb, 2011)
<i>Streptomyces coelicolor</i> M1152/pEVK6	M1152 carrying a pIJ10257 kyamycin <i>lanR1</i> and <i>lanL</i> activator cassette	(Vikeli <i>et al.</i> , 2020)
<i>Streptomyces coelicolor</i> M1152/pEVK6/pWDW63	M1152 carrying pEVK6 and a pSET152 kyamycin biosynthetic cassette	(Vikeli <i>et al.</i> , 2020)
<i>Streptomyces coelicolor</i> M1152/pAMA1	M1152 with pIJ10257 carrying the entire activator and regulatory region of kyamycin	This work
<i>Streptomyces coelicolor</i> M1152/pAMA1/pWDW63	M1152 carrying pAMA1 and pWDW63	This work
<i>Streptomyces coelicolor</i> M1146	Antibiotic producing superhost [$\Delta act \Delta red \Delta cpk \Delta cda$]	(Gomez-Escribano & Bibb, 2011)
<i>Streptomyces coelicolor</i> M1146/pEVK6	M1146 carrying pEVK6	This work
<i>Streptomyces coelicolor</i> M1146/pEVK6/pWDW63	M1146 carrying pEVK6 and pWDW63	This work
<i>Streptomyces coelicolor</i> M1146/pEVK6/pWDW70	M1146 carrying pEVK6 and a pSET152 kyamycin biosynthetic cassette lacking the <i>kyaA</i> core peptide	This work
<i>Streptomyces coelicolor</i> M1146/pAMA1	M1146 carrying pAMA1	This work
<i>Streptomyces coelicolor</i> M1146/pAMA1/pWDW63	M1146 carrying pAMA1 and pWDW63	This work
<i>Streptomyces coelicolor</i> M1146/pEVK6/L_01 to L_35	M1146 carrying pEVK6 and pWDW70 with lantibiotic core peptides L_01 to L_35 inserted	This work
<i>Streptomyces coelicolor</i> M1146/pAMA3	M1146 carrying pIJ10257 containing rosiermycin <i>lanR1</i> and a truncated <i>lanL</i>	This work
<i>Streptomyces coelicolor</i> M1146/pAMA3/pAMA2	M1146 carrying pAMA3 and pSET157 containing the biosynthetic genes	This work

	of rosiermycin with the <i>rosA</i> core peptide removed	
<i>Streptomyces coelicolor</i> M1146/pAMA5	M1146 carrying pIJ10257 containing the <i>lanL</i> and <i>lanR1</i> genes for rosiermycin	This work
<i>Streptomyces coelicolor</i> M1146/pAMA5/pAMA2	M1146 carrying pAMA5 and pAMA2	This work
<i>Streptomyces coelicolor</i> M1146/pAMA3/pAMA6	M1146 carrying pAMA3 and pAMA2 with a stem loop attenuation	This work
<i>Streptomyces coelicolor</i> M1146/pAMA5/pAMA6	M1146 carrying pAMA5 and pAMA6	This work
<i>Streptomyces coelicolor</i> M1146/pEVK6/Lr_00	M1146 carrying pEVK6 and pAMA2 with the kyamicin core peptide inserted	This work
<i>Streptomyces coelicolor</i> M1146/pEVK6/Lr_14	M1146 carrying pEVK6 and pAMA2 with the rosiermycin core peptide inserted	This work
<i>Streptomyces coelicolor</i> M1146/pEVK6/p6Lr_00	M1146 carrying pEVK6 and pAMA6 with the kyamicin core peptide inserted	This work
<i>Streptomyces coelicolor</i> M1146/pEVK6/p6Lr_14	M1146 carrying pEVK6 and pAMA6 with the rosiermycin core peptide inserted	This work
<i>Streptomyces coelicolor</i> M1146/pAMA3/Lr_00 to Lr_35	M1146 carrying pAMA3 and pAMA2 with the core peptides from L_00 to L_35 inserted	This work
<i>Streptomyces coelicolor</i> M1146/pAMA3/Lr_00/pAMA4	M1146 carrying pAMA3, pAMA2 with the kyamicin core peptide and the <i>kyaL</i> pAMA4 cassette	This work
<i>Streptomyces coelicolor</i> M1146/pAMA3/Lr_14/pAMA4	M1146 carrying pAMA3, pAMA2 with the rosiermycin core peptide and the <i>kyaL</i> pAMA4 cassette	This work
<i>Streptomyces coelicolor</i> M1146/pAMA3/p6Lr_00	M1146 carrying pAMA3 and pAMA6 with the	This work

	kyamycin core peptide inserted	
<i>Streptomyces coelicolor</i> M1146/pAMA3/p6Lr_14	M1146 carrying pAMA3 and pAMA6 with the rosiermycin core peptide inserted	This work
<i>Streptomyces coelicolor</i> M1146/pAMA5/Lr_00	M1146 carrying pAMA5 and pAMA2 with the kyamycin core peptide inserted	This work
<i>Streptomyces coelicolor</i> M1146/pAMA5/Lr_14	M1146 carrying pAMA5 and pAMA2 with the rosiermycin core peptide inserted	This work
<i>Streptomyces coelicolor</i> M1146/pAMA5/p6Lr_00	M1146 carrying pAMA5 and pAMA6 with the kyamycin core peptide inserted	This work
<i>Streptomyces coelicolor</i> M1146/pAMA5/p6Ls_14	M1146 carrying pAMA5 and pAMA6 with the rosiermycin core peptide inserted	This work
<i>Streptomyces roseoverticillatus</i>	Potential lantibiotic producing strain, DSM 40845	DSMZ – Leibniz Institute, German collection of Microorganisms and Cell Cultures
<i>Streptomyces roseoverticillatus</i> /pAMA1	<i>S. roseoverticillatus</i> carrying pAMA1 large kyamycin activator cassette	This work
<i>Streptomyces roseoverticillatus</i> /pAMA5	<i>S. roseoverticillatus</i> carrying pAMA5 rosiermycin activator cassette	This work
<i>Streptomyces coelicolor</i> M1146/21A	M1146 carrying pESAC13 with antascomycin BGC from <i>Micromonospora sp.</i>	This work
<i>Streptomyces coelicolor</i> M1146/21A/SARP	M1146 carrying PAC 21A and the SARP activator	This work
<i>Streptomyces coelicolor</i> M1146/21A/LuxR	M1146 carrying PAC 21A and the LuxR activator	This work

<i>Streptomyces coelicolor</i> M1146/21A/AraC	M1146 carrying PAC 21A and the AraC activator	This work
<i>Streptomyces coelicolor</i> M1146/21A/LuxR/SARP	M1146 carrying the PAC 21A and SARP and LuxR	This work
<i>Streptomyces coelicolor</i> M1146/21A/AraC/SARP	M1146 carrying the PAC 21A and SARP and AraC	This work
<i>Streptomyces coelicolor</i> M1146/21A/AraC/LuxR	M1146 carrying the PAC 21A and LuxR and AraC	This work
<i>Streptomyces coelicolor</i> M1146/21A/AraC/LuxR/SARP	M1146 carrying the PAC 21A and SARP, LuxR and AraC	This work
<i>Streptomyces coelicolor</i> M1146/21B	M1146 carrying pESAC13 with antascomycin BGC from <i>Micromonospora sp.</i>	This work
<i>Streptomyces coelicolor</i> M1146/21B/SARP	M1146 carrying PAC 21B and the SARP activator	This work
<i>Streptomyces coelicolor</i> M1146/21B/LuxR	M1146 carrying PAC 21B and the LuxR activator	This work
<i>Streptomyces coelicolor</i> M1146/21B/AraC	M1146 carrying PAC 21B and the AraC activator	This work
<i>Streptomyces coelicolor</i> M1146/21B/LuxR/SARP	M1146 carrying the PAC 21B and SARP and LuxR	This work
<i>Streptomyces coelicolor</i> M1146/21B/AraC/SARP	M1146 carrying the PAC 21B and SARP and AraC	This work
<i>Streptomyces coelicolor</i> M1146/21B/AraC/LuxR	M1146 carrying the PAC 21B and LuxR and AraC	This work
<i>Streptomyces coelicolor</i> M1146/21B/AraC/LuxR/SARP	M1146 carrying the PAC 21B and SARP, LuxR and AraC	This work
<i>Streptomyces coelicolor</i> M1146/215G	M1146 carrying pESAC13 containing the formicamycin BGC	This work
<i>Streptomyces coelicolor</i> M1146/215G Δ forJ	M1146 carrying PAC 215G Δ forJ	This work
<i>Streptomyces formicea</i> KY3	Formicamycin producing strain	(Qin <i>et al.</i> , 2017)
<i>Streptomyces formicea</i> KY3 Δ forJ	Formicamycin producing strain with the <i>forJ</i> repressor deleted	(Devine <i>et al.</i> , 2021)

Table 2.6: *Saccharopolyspora* strains used in this work

Strain	Description/Genotype	Reference
<i>Saccharopolyspora erythraea</i> ISOM	Antibiotic producing host	Isomerase Therapeutics (Cambridge, UK)
<i>Saccharopolyspora erythraea</i> ISOM/215G	Heterologous host carrying formicamycin PAC 215G	This work
<i>Saccharopolyspora erythraea</i> ISOM/215G Δ forJ	Heterologous host carrying PAC 215G Δ forJ	This work
<i>Saccharopolyspora erythraea</i> ISOM/6_19B	Heterologous host carrying pESAC13 with nystatin Ps1 BGC	This work
<i>Saccharopolyspora erythraea</i> ISOM/2_5K	Heterologous host carrying pESAC13 with nystatin Ps2 BGC	This work
<i>Saccharopolyspora erythraea</i> ISOM/21A	Heterologous host carrying PAC 21A	This work
<i>Saccharopolyspora erythraea</i> ISOM/21A/SARP	Heterologous host carrying PAC 21A and the SARP activator	This work
<i>Saccharopolyspora erythraea</i> ISOM/21A/LuxR	Heterologous host carrying PAC 21A and the LuxR activator	This work
<i>Saccharopolyspora erythraea</i> ISOM/21A/AraC	Heterologous host carrying PAC 21A and the AraC activator	This work
<i>Saccharopolyspora erythraea</i> ISOM/21B	Heterologous host carrying PAC 21B	This work
<i>Saccharopolyspora erythraea</i> ISOM/21B/SARP	Heterologous host carrying PAC 21B and the SARP activator	This work
<i>Saccharopolyspora erythraea</i> ISOM/21B/LuxR	Heterologous host carrying PAC 21B and the LuxR activator	This work
<i>Saccharopolyspora erythraea</i> ISOM/21B/AraC	Heterologous host carrying PAC 21B and the AraC activator	This work
<i>Saccharopolyspora erythraea</i> ISOM/3D	Heterologous host carrying pESAC13 with NRPS from <i>Streptomyces</i> T676	This work
<i>Saccharopolyspora erythraea</i> ISOM/24J	Heterologous host carrying pESAC13 with lantipeptide BGC from <i>Streptomyces</i> T676	This work

<i>Saccharopolyspora erythraea</i> ISOM/TB 73	Heterologous host carrying pESAC13 with Type 1 PKS macrolactam from <i>Streptomyces</i> T676	This work
<i>Saccharopolyspora erythraea</i> ISOM/7G	Heterologous host carrying pESAC13 with thiazolpeptide BGC from <i>Streptomyces</i> T676	This work
<i>Saccharopolyspora erythraea</i> ISOM/14M	Heterologous host carrying pESAC13 with anthracimycin BGC from <i>Streptomyces</i> T676	This work
<i>Sacchropolyspora</i> KY3	<i>Saccharopolyspora</i> species encoding the kyamycin BGC	(Vikeli <i>et al.</i> , 2020)

Table 2.7: Antibacterial activity indicator strains used in this work

Strain	Description/Genotype	Reference
<i>Bacillus subtilis</i> EC1524	Bioassay strain; <i>trpC2</i> , subtilin BGC deleted	(Widdick <i>et al.</i> , 2003)
<i>E. coli</i> ATCC 25922	Indicator strain for <i>in vivo</i> tests, commonly used as a quality control strain	(Minogue <i>et al.</i> , 2014)
<i>E. coli</i> NR698	Bioassay strain with deletion in <i>imp4213</i> for enhanced uptake of small molecules	(Ruiz <i>et al.</i> , 2005)
Methicillin-resistant <i>Staphylococcus aureus</i> (MRSA)	Clinical isolate of MRSA isolated from a patient sample at the Norfolk and Norwich University Hospital (UK) and validated using a PCP2a test	(Qin <i>et al.</i> , 2017)
Methicillin-sensitive <i>Staphylococcus aureus</i> (MSSA)	Pure culture of <i>Staphylococcus aureus</i> subsp. <i>aureus</i> Rosenbach (ATCC 6538P)	(Cotter & Adley, 2001; Mowatt <i>et al.</i> , 1995)
<i>Streptomyces venezuelae</i>	Pure culture of <i>Streptomyces venezuelae</i> ATCC 10712	(Bradley & Ritzi, 1968)

2.2.3 Antibiotics

Unless stated otherwise all antibiotics were used at concentrations listed below in

Table 2.8.

Table 2.8: Antibiotics used in this work

Antibiotic	Culture Concentration ($\mu\text{g}/\text{mL}$)
Apramycin	50
Hygromycin	50
Thiostrepton	30
Nalidixic acid	25
Tetracycline	50
Kanamycin	50
Carbenicillin	50
Chloramphenicol	25

2.3 Culture and Production Media

2.3.1 Culture Media

2xYT: suspension of spores for conjugation

Tryptone 16 g

Yeast Extract 10 g

NaCl 5 g

Distilled water to 1 L

Adjust to pH 7.4

ABBA13: for sporulation of *Saccharopolyspora*

Bacto Peptone from Soybean 5 g

Starch 5 g

CaCO₃ 3 g

MOPS 2.1 g

Agar 20 g

R0 water to 1 L

After autoclaving

1 M MgSO₄ 10 mL

1% Thiamine 1 mL

1.2% FeSO₄·7H₂O 1 mL

Anthracimycin Production Media (A1): for production of anthracimycin

Starch 10 g

Yeast Extract 4 g

Peptone 2 g

CaCO₃ 1 g

Fe₂(SO₄)₃·4H₂O 40 mg

KBr 100 mg

Sea Salts 40 g

Distilled water to 1 L

Difco™ Nutrient Agar (Becton Dickinson) (DNA): for solid culture of *E. coli* with HygR

Difco™ Nutrient Agar 23 g

Distilled water to 1 L

GPP: for growth of *Saccharopolyspora*

Glucose 5 g

Glucidex-12 maltodextrin 80 g

Pharmamedium 60 g

CaCO₃ 8 g

Distilled water to 1 L

Adjust pH to 7

GYM: recovery of *Streptomyces roseoverticillatus*

Glucose 4 g

Yeast Extract 4 g

Malt Extract 10 g

CaCO₃ (solid medium) 2 g

Agar (solid medium) 12 g

Distilled water to 1 L

Adjust to pH 7.2

Instant March Agar (IMA): for sporulation of *Streptomyces*

SMAISOM 20 g

Agar 20 g

Tap water to 1 L

Lennox (L): for liquid culture of *E. coli* with HygR

Bacto-tryptone 10 g

Yeast Extract 5 g

NaCl 5 g

Glucose 1 g

Distilled water to 1 L

Add 2% Agar for solid media

Lysogeny Broth (LB): for *E. coli* culture

Bacto-tryptone 10 g

Yeast Extract 5 g

NaCl 10 g

Distilled water to 1 L

Adjust pH to 7.5

Add 2% Agar for solid media

***Micromonospora* Production Media (MM):** for production of antascomicin

Mannitol 20 g

Yeast Extract 1 g

Peptone casein (Tryptone) 20 g

Trace Elements

KH₂PO₄ 0.1 g

MgSO₄·7H₂O 0.005 g

CaCl₂·6H₂O 0.02 g

Agar 1 g

Distilled water to 1 L

Adjust pH to 6.5

R5: for overlay bioassays and production assays

Sucrose 103 g

K₂SO₄ 0.25 g

MgCl₂·6H₂O 10.12 g

Glucose 10 g

Casamino acid 0.1 g

Yeast Extract 5 g

TES 5.73 g

Trace element solution 2 mL

Distilled water to 1 L

Adjusted to pH 7.0 with 5M NaOH

For solid medium this was dispensed in 95 mL aliquots and 22 g of Difco Bacto agar was added

Trace elements for R5

ZnCl ₂	0.04 g
FeCl ₂ .6H ₂ O	0.2 g
CuCl ₂ .2H ₂ O	0.01 g
MnCl ₂ .4H ₂ O	0.01 g
Na ₂ B ₄ O ₇ .10H ₂ O	0.01 g
(NH ₄) ₆ Mo ₇ O ₂₄ .4H ₂ O	0.01 g

SF+M: for growth and conjugation of actinomycetes

Soya flour	20 g
Mannitol	20 g
Lab M N _o 1 agar	20 g

Tap water to 1 L

For liquid no Agar was added

SM7: for growth of *Streptomyces*

MOPS	20.9 g
L-Proline	15 g
Glycerol Solution	20 g
Sucrose	2.5 g
L-Glutamate	1.5 g
NaCl	0.5 g
K ₂ HPO ₄	2.0 g
0.2 M MgSO ₄	10 mL
0.02 M CaCl ₂	10 mL
Trace salts N ^o 1	5 mL

Distilled water to 1 L

Adjust to pH 6.5 with KOH (unless stated otherwise)

SOC: growth of *E. coli* competent cells

Peptone from casein	20 g
Yeast Extract	5 g
NaCl	0.58 g
KCl	0.186 g
MgCl ₂ .6H ₂ O	2.03 g
MgSO ₄ .7H ₂ O	2.46 g
Glucose	3.6 g
Distilled water to 1 L	

Soft Nutrient Agar (SNA): for overlay bioassays

Difco Nutrient Broth	8 g
Formedium	7 g
Distilled water to 1 L	

SV2: for growth of actinomycetes

Glucose	15 g
100% Glycerol solution	15 g
Soy Peptone	15 g
NaCl	3 g
CaCO ₃	1 g
Distilled water to 1 L	
For solid 2% Agar was added	

Tryptone Soya Broth (Oxoid) (TSB): for culture of actinomycetes

Tryptone Soya Broth	30 g
Distilled water to 1 L	

Yeast Extract-Malt Extract (YEME): for growth of *Saccharopolyspora*

Yeast Extract	3 g
Bacto-Peptone	5 g
Malt Extract	3 g

Glucose	10 g
Sucrose	340 g

Distilled water to 1 L

For solid agar addition of 15 g Difco Bacto agar

After autoclaving addition of $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ to a 5 mM final concentration

YP: production media for *S. erythraea* ISOM

Peptone from casein	20 g
---------------------	------

Yeast Extract	10 g
---------------	------

Agar (Solid media)	20 g
--------------------	------

Distilled water to 1 L

Adjust pH to 5.8

YPD: production media for *S. erythraea* ISOM

Peptone from casein	20 g
---------------------	------

Yeast Extract	10 g
---------------	------

40 % Glucose	100 mL
--------------	--------

Agar (Solid media)	20 g
--------------------	------

Distilled water to 1 L

Adjust pH to 5.8

2.3.2 Cell Culture and Stock Preparation

2.3.2.1 *E. coli* Glycerol Stocks

E. coli cultures were grown in 10 mL LB overnight at 37°C, 250 rpm. In a 2 mL cryotube 1 mL of culture was spun down and supernatant discarded. This was repeated before 750 μL of culture was mixed with 750 μL 50 % glycerol and vortexed to mixed. Glycerol stocks were stored at -80°C.

2.3.2.2 Actinomycetes Spore Stocks

Actinomycetes were spread on appropriate sporulation media plates and grown at 30 °C until a confluent lawn was present. Plates were flooded with 5 mL 20 % glycerol and spores suspended using a sterile Q-tip. Spores were extracted through a sterile cotton disc using a syringe. A further 5 mL 20 % glycerol was used to flood the plate and extraction repeated. The spore suspension was then centrifuged and all but 1-2 mL of the supernatant removed. Spores were vortexed into solution and transferred to cryotubes before being stored at -80 °C.

2.3.2.3 Preparation of Electrocompetent *E. coli*

A 10 mL overnight culture of single colonies of *E. coli* ET12567/pUZ8002 were grown in LB with appropriate antibiotics. Approximately 3 mL of these 10 mL cultures were then used to inoculate 100 mL LB with appropriate antibiotics. Cells were incubated at 37 °C, 250 rpm until the OD₆₀₀ was approximately equal to 0.4. The culture was then separated into 50 mL falcon and palleted by centrifugation at 4000 rpm for 5 min at 4 °C. Supernatant was discarded and the pellets were washed in 50 mL sterile ice-cold MilliQ water. Centrifugation was repeated and supernatant discarded. Cells were combined together and washed once more with 50 mL ice-cold sterile MilliQ water before being washed again in sterile ice-cold 10% glycerol solution. Cells were then re-suspended in 1-5 mL of sterile ice-cold 10% glycerol solution and centrifuged once more. Finally, the supernatant was discarded, and cells re-suspended in 0.5-1 mL 10% glycerol. Cells were then aliquoted in 50 µL volume and flash frozen using liquid nitrogen before being stored at -80 °C.

2.4 DNA, Cloning and Transformation

2.4.1 DNA

2.4.1.1 *Saccharopolyspora* Genomic DNA Extraction

Saccharopolyspora KY3 spores were used to inoculate 50 mL SV2 media and incubated for 4 days at 30 °C, 250 rpm. Mycelium was separated and genomic DNA extracted using the FastDNA™ SPIN KIT (MP Biomedicals). DNA was stored at -20 °C.

2.4.1.2 PAC Extraction

PAC DNA was extracted from two 500 mL LB *E. coli* cultures inoculated from a 10 mL starter culture and grown overnight at 37°C, 250 rpm. PAC DNA was extracted using the PhasePrep™ BAC DNA Kit (Sigma-Aldrich). PAC DNA inserts were visualised using PCR and gel electrophoresis and sequenced using PACbio Illumina sequencing. DNA was stored at -20 °C.

2.4.1.3 Plasmid Extraction

A single colony of appropriate *E. coli* strain was used to inoculate 10 mL LB containing appropriate antibiotic. Cultures were grown overnight at 37 °C, 250 rpm. Plasmid DNA was extracted using the QIAprep® Spin Miniprep Kit (Qiagen) and constructs sequenced by Sanger sequencing. DNA was stored at -20 °C.

2.4.1.4 PCR

PCR reactions were performed in a T100™ Thermal Cycler (Bio-Rad). The general protocol for PCR DNA amplification using Q5® High-Fidelity DNA Polymerase (NEB) for subsequent cloning is present in **Table 2.9**. The general protocol for PCR DNA amplification using GoTaq® G2 Green Master Mix (Promega) for subsequent DNA diagnostics is present in **Table 2.10**. The general protocol for colony PCR using MyTaq™ Red Mix (Bioline) for subsequent DNA diagnostics is present in **Table 2.11**. For colony PCR of *Saccharopolyspora* a sterile tip was used to crush mycelia growth in 5 µL of DMSO and 1 µL of this mix would be used as template DNA. For colony PCR of *Streptomyces* a sterile tip was used to transfer mycelial growth into 50 µL of

sterile water, vortexed until homogenous and incubated at 65 °C for 25 min, 1 µL of this mix would then be used as template DNA.

Table 2.9: Protocol for Q5 PCR reaction

Component	50 µL Reaction	Final Concentration
5X Q5 Reaction Buffer	10 µL	1X
10 mM dNTPs	1 µL	200 µM
10 µM Forward Primer	2.5 µL	0.5 µM
10 µM Reverse Primer	2.5 µL	0.5 µM
Template DNA	Variable	< 1000 ng
Q5 High-Fidelity DNA Polymerase	0.5 µL	0.02 U/µL
5X Q5 High GC Enhancer	10 µL	1X
Nuclease Free Water	To 50 µL	

Step	Temperature (°C)	Time	Repeat (x)
Initial Denaturation	98	1 min	1
Denaturation	98	10 sec	35
Annealing	50-72	1 min	
Elongation	72	3 min	
Final Elongation	72	10 min	1
Hold	10	∞	1

Table 2.10: Protocol for GoTaq PCR reaction

Component	25 µL Reaction	Final Concentration
2X GoTaq Green PCR Mix	12.5 µL	1X
10 µM Forward Primer	0.5 µL	0.5 µM
10 µM Reverse Primer	0.5 µL	0.5 µM
Template DNA	Variable	< 250 ng
Nuclease Free Water	To 25 µL	

Step	Temperature (°C)	Time	Repeat (x)
Initial Denaturation	98	5 min	1
Denaturation	98	30 sec	30
Annealing	50-72	30 min	
Elongation	72	45 s/kb	
Final Elongation	72	7 min	1
Hold	10	∞	1

Table 2.11: Protocol for MyTaq PCR reaction

Component	25 μ L Reaction	Final Concentration
2X MyTaq Red PCR Mix	12.5 μ L	1X
10 μ M Forward Primer	0.5 μ L	0.5 μ M
10 μ M Reverse Primer	0.5 μ L	0.5 μ M
Template DNA	Variable	< 250 ng
Nuclease Free Water	To 25 μ L	

Step	Temperature ($^{\circ}$ C)	Time	Repeat (x)
Initial Denaturation	98	3 min	1
Denaturation	98	15 sec	30
Annealing	50-72	15 min	
Elongation	72	15 s/kb	
Final Elongation	72	2 min	1
Hold	10	∞	1

2.4.1.5 Agarose Gel Electrophoresis

Agarose gels were prepared using agarose in Tris Borate EDTA (TBE, 90 mM Tris HCl, 90 mL boric acid, 2 mM EDTA) or Tris Acetate EDTA (TAE, 40 mM Tris HCl, 20 mM acetic acid, 1 mM EDTA) buffer to a final concentration between 0.6 and 1.0 %, depending on application and size of DNA fragments. About 10 μ g/mL ethidium bromide was added. DNA fragments were mixed with purple loading dye (NEB) and run alongside appropriately sized DNA ladder (NEB). Gel electrophoresis was performed in TBE or TAE buffer using a PowerPac™ Universal Power Supply (Bio-Rad) and run at 80-120 V until DNA fragments separated. DNA was then visualised using a UV Transilluminator.

2.4.1.6 Sequencing

Sanger

Sanger sequencing was carried out by Eurofins Genomics using their Mix2Seq overnight Kit, or by GeneWiz overnight sequencing service.

Illumina

PACbio Illumina sequencing was carried out by the DNA Sequencing Facility, Department of Biochemistry, University of Cambridge.

2.4.2 Cloning

2.4.2.1 Ligation

DNA fragments and vectors were cut with appropriate restriction endonucleases in a 50 μL reaction at 37 °C for at least 1 h (following manufacturer's instructions). Digested DNA fragments were purified on a 0.6 % agarose gel by electrophoresis and extracted using the QIAquick® Gel Extraction Kit (Qiagen) or via the QIAquick® PCR Clean Purification Kit (Qiagen) as appropriate. Ligation was carried out using T4 ligase (NEB) and fragments combined in a 3:1 ratio of insert to vector. Fragments were incubated together at 50 °C for 2 min before T4 ligase and 10x T4 ligase buffer added to a final volume of 20 μL . Ligations were cooled in water bath at room temperature before being incubated at 4 °C overnight. Constructs were transformed into DH5- α *E. coli* and analysed using Sanger sequencing.

2.4.2.1 Gibson Assembly

Lantibiotic core peptide fragments (\approx 128 bp) (Eurofins Genomics and IDT) were made up to concentration 0.25 pmol/ μL before assembly into appropriate vectors. Gibson assembly was carried out using 10 x Gibson Assembly Master Mix (NEB) according to instruction in a 10 μL volume. Samples were incubated at 50 °C for 15 min in a T100™ Thermal Cycler (Bio-Rad) before transformation into NEB® 5-alpha Competent (High Efficiency) *E. coli*. Constructs were analysed using enzymatic digest or colony PCR and Sanger sequencing.

2.4.3 Transformation

2.4.3.1 Transformation of Chemically Competent *E. coli*

Cells were aliquoted into 50 μ L volumes and mixed with either 5 μ L ligation mix, 2 μ L Gibson Assembly mix or 1 μ L of plasmid DNA. Cells were then incubated on ice for 30 min before being heat shocked at 42°C for 30-45 s and returned to ice for a further 2 min. The transformed cells were diluted in SOC medium to make the volume up to 1 mL and recovered at 37 °C, 250 rpm for 1 h. Cells were then plated onto LB plates with appropriate antibiotics or DNA plates if testing for hygromycin resistance to obtain single colonies. Plates were incubated overnight at 37 °C.

2.4.3.2 Transformation of Electrocompetent *E. coli*

Electrocompetent *E. coli* Et12567/pUZ8002 cells (50 μ L) were thawed and transferred to iced-cold cuvettes. Approximately 1-2 μ L of plasmid DNA was added and gently mixed. Samples were then electroporated using an Eppendorf Eporator® set to 2.5 kV. The electroporated cells were immediately diluted with 950 μ L of SOC media and transferred to a 1.5 mL microcentrifuge tube before being incubated at 37 °C, 250 rpm for 1 h. Cells were then plated onto LB agar or DNA agar containing appropriate antibiotics for single colony selection. Plates were incubated at 37 °C overnight.

2.4.3.3 Tri-Parental Mating for PACs

E. coli DH10B containing the PAC, *E. coli* TOPO10/pR9604 and either *E. coli* ET12567 or *E. coli* C2925 were grown in 10 mL overnight cultures, 2 mL was then used to inoculate 50 mL LB with appropriate antibiotics. This was incubated at 37 °C, 250 rpm until the OD₆₀₀ was between 0.4-0.6. Cells were then centrifuged at 4 °C, 3000 rpm for 5 min and supernatant discarded. Cell pellets were re-suspended and washed twice by centrifugation in fresh chilled LB media. The supernatant was discarded and the cells re-suspended in the flow back. In the middle of a LB agar plate with no antibiotics 20 μ L of each cell type was spotted and left to dry. Plates were incubated overnight at 37 °C. The cell spot was then streaked for single colonies on LB agar plates containing appropriate antibiotics and incubated

overnight at 37 °C. PCR was used to determine whether PACs had transferred to ET12567 *E. coli* successfully.

2.4.3.4 Actinomycete Conjugation

E. coli ET12567 or *E. coli* C2925 containing plasmid inserts were grown overnight in a 10 mL LB culture. This was then used to inoculate 50 mL LB media containing appropriate antibiotics. The culture was incubated at 37 °C, 250 rpm until OD₆₀₀ was approximately 0.4-0.6. Cells were then centrifuged at 4 °C, 3000 rpm for 7 min and supernatant discarded. Cells were washed twice by centrifugation using fresh chilled LB media and pellet resuspended in flow back. Spore stocks of appropriate actinomycetes were thawed on ice and 50 µL spun and glycerol removed. Spores were then resuspended in 0.5 mL 2YT media and heat shocked at 50 °C for 10 min. Spores were mixed with *E. coli* and spread on SF+M plates with 10 mM MgCl₂ and incubated at 30 °C overnight. For *Streptomyces* species only 100 µL of the sample mix was spread, for *Saccharopolyspora* species the sample was centrifuged for 1 min, the supernatant discarded and the pellet re-suspended in flow back before the entire mix was plated. The next day a 1 mL overlay with the appropriate antibiotics was spread and the plates returned to 30 °C until ex-conjugants appeared. Ex-conjugants were then spread on SF+M plates with appropriate antibiotics and incubated at 30 °C. PCR was used to determine whether inserts were incorporated successfully as appropriate.

2.5 Bioassays

2.5.1 Overlay Bioassays

Lantibiotic *Streptomyces* constructs were streaked onto the centre of R5 plates from spore stocks and left at 30 °C for 7 days. Bioindicator strains were plated from glycerol stock onto LB plate for single colonies and left overnight at 37 °C. A single colony was used to inoculate 10 mL LB, incubated at 37 °C, 250 rpm until OD₆₀₀ was about 0.4. SNA agar was melted and 5 mL agar with 0.5 mL indicator strain mixed in used to overlay plates. Bioassay plates were then incubated at 30 °C overnight and the zone of inhibition observed.

2.5.2 Minimum Inhibitory Concentration (MIC) Bioassays

A spot lawn method was used to determine the lantipeptide MICs. A 1 mg/mL stock solution was prepared, along with serial dilutions from 256 µg/mL to 8 µg/mL. Bioindicator strains other than *Streptomyces venezuelae* were grown as above until OD₆₀₀ was about 0.4. SNA plates were obtained with 1 mL indicator strain per 10 mL agar. For *S. venezuelae* spores from a high concentration spore stock were added directly to the liquid SNA agar in a 1000 times dilution. Once the plates were dry, 5 µL of each lantipeptide dilution was applied directly to the agar and incubated overnight at 30 °C. The MIC was defined as the lowest concentration for which a clear zone of inhibition was observed.

2.6 Extraction of Antibiotics

2.6.1 Liquid

2.6.1.1 Analytical

Lantibiotic Extraction

Unless stated otherwise spore stocks ($\approx 20 \mu\text{L}$) were used to inoculate 10 mL TSB incubated at 30 °C, 250 rpm for 5 days. Afterwards 1 mL of culture was mixed with 1 mL of methanol and shaken at 1200 rpm for 1-2 h on an IKA®VIBRAX VXR bench top shaker. Samples were then spun for 10 min and 1 mL of supernatant was transferred into HPLC vials for analysis. Extracts were stored at -20 °C.

***Saccharopolyspora* Fermentations**

Unless stated otherwise spore stocks were used to inoculate 10 mL TSB or SV2 cultures and incubated at 30 °C, 250 rpm for 2-4 days until good growth was observed. The culture was then used to further inoculate 40 mL of TSB or SV2 and incubated at 30 °C, 250 rpm for 2-3 days. Once strong growth is seen 0.5 mL of culture was used to inoculate 10 mL production media in falcon tubes with sterile bungs. Samples were then incubated at 30 °C, 250 rpm for 8 days unless otherwise stated. Extracts of 1 mL were taken at day 6 and day 8 unless otherwise stated and mixed with 1 mL methanol. Samples were then shaken at 1200 rpm for 1-2 h on a IKA®VIBRAX VXR bench top shaker. Samples were then spun for 10 min and 1 mL supernatant transferred into a HPLC vial for analysis. Extracts were stored at -20 °C.

2.6.1.2 Purification

Lantibiotics

Spore stocks were used to inoculate 10 mL TSB and incubated at 30 °C, 250 rpm for 2-3 days. Once strong growth was seen 5 mL was used to inoculate 500 mL liquid SFM media in 2.5 L conical flasks. Cultures were incubated at 30 °C, 250 rpm for 7 to 10 days as stated. Cultures were then transferred to 1 L ultracentrifuge tubes. If stated samples were then additionally sonicated using a MSE Soniprep 150 probe sonicator at 15 amp, cycle; 30 s on, 10 s off, for 5 min. Culture was then centrifuged at 5000 rpm for 30 min. Supernatant was filtered using a 0.22 μm filter (Sartorius,

Starlab) or multiple grade 292 filter discs (Sartorius). Up to 3 L of this filtered extract was then applied under pressure using inert air to a Biotage® SNAP KP-C18_HS 120g column and run on a Isolera One Biotage. Solvent A; R0 water, solvent B; methanol (MeOH), flowrate 50 mL/min. Wavelengths monitored; 254 nm, 210 nm. The first four column volumes were discarded to waste and 48 mL fractions were collected. The following gradient was used (% solvent B, number of column volume (CV)): 5 % 3.0 CV, 5 % 1 CV, 5-100% 10 CV, 100 % 2 CV.

Fractions were transferred into 50 mL falcon tubes and the flow back transferred to HPLC vials for UHPLC-HRMS analysis. Solvent was evaporated using a SP Scientific GenVac EZ-2 series. Fractions containing the wanted compound were then re-suspended in DMSO, pooled, solvent evaporated as before and weighed. The extracts were then run on preparative HPLC for further purification.

2.6.2 Solid

2.6.2.1 Agar Plates

***Streptomyces* Growth**

Unless stated otherwise 20 µL of spore stock was used to inoculate SF+M plates incubated at 30 °C for 9 days. The middle of the plate was extracted and the agar transferred to a 2 mL microcentrifuge tube and 1 mL ethyl acetate added. Samples were then shaken at 1200 rpm, for 1-2 h on a IKA®VIBRAX VXR bench top shaker. Samples were centrifuged for 15 min, separating out the ethyl acetate layer. 300 µL of the ethyl acetate layer was then transferred to a HPLC vial and the solvent evaporated using a SP Scientific GenVac EZ-2 series and the residue resuspended in 200 µL methanol for analysis. Extracts were stored at -20 °C.

***Saccharopolyspora* Growth**

Unless stated otherwise 20 µL of spore stock were used to inoculate solid agar plates incubated at 30 °C for 12 days. A quarter of the plate showing good growth was then removed using sterile razor blade and extracted with 5 mL methanol. Samples were shaken at 250 rpm, 30 °C for 1 h. A 2 mL aliquot was removed,

centrifuged and pellet discarded. Solvent was evaporated using a SP Scientific GenVac EZ-2 series and the residue resuspended in 200 μ L methanol, centrifuged and 150 μ L transferred to HPLC vials for analysis. Extracts were stored at -20 °C.

2.6.2.2 Bioassay Plates

The back of a sterile p1000 tip was used to take a core from the zone of inhibition if present, or from directly besides the bacterial growth if a zone of inhibition was not present. The core was transferred to a 2 mL microcentrifuge tube and freeze thawed at -80 °C for at least 10 min. Once thawed 300 μ L of MeOH was added and samples were shaken at 1200 rpm for 20 min on an IKA®VIBRAX VXR bench top shaker. Samples were then centrifuged for 10 min and the supernatant transferred to Thomson 0.2 μ m Filter Vials™ (HTS Labs) for analysis. Extracts were stored at -20 °C.

2.7 Chemical Analysis and Characterisation

2.7.1 Analytical LCMS (Liquid Chromatography Mass Spectrometry)

2.7.1.1 UHPLC-MS (Ultra High-Performance Liquid Chromatography-Mass Spectrometry)

Unless stated otherwise samples were injected (5 μ L) into a Shimadzu single quadrupole LCMS-2010A mass spectrometer equipped with Prominence HPLC system. Compounds were separated on a Kinetex[®] XB-C18 00B-4496-AN (50 x 2.1 mm, 2.6 μ m, 100 Å, Phenomenex). Solvent A; 0.1 % formic acid in water, solvent B; methanol. The following gradient was used: 0 min 2% B, 7 min 100% B, 9 min 100% B, 12 min 2% B. The UHPLC System was connected to a LCMS-IT-TOF liquid chromatograph mass spectrometer (Shimadzu) and mass spectra acquired in positive ion mode, scan range 100-1000 m/z . Data was analysed using LCMSsolutions[™] (Shimadzu).

2.7.1.2 UHPLC-HRMS (Ultra High-Performance Liquid Chromatography-High Resolution Mass Spectrometry)

Measurements of lantibiotics were acquired with a Waters Acquity UHPLC system equipped with an ACQUITY UHPLC HSS T3 1.8 μ m, 2.1 x 100 column (Waters) connected to a Synapt G2-Si high resolution mass spectrometer (Waters). Solvent A; 0.1% formic acid in water, solvent B; 0.1% formic acid in acetonitrile. Unless stated otherwise a flow rate of 200 μ L/min was used with the following gradient: 1 min 1 % B, 10 min 60 % B, 11 min 99% B, 13 min 99 % B, 13.10 min 1% B, 18 min 1 % B.

MS spectra were acquired with a scan time of 1 s in the range of m/z 50 to 1200 in positive MSE resolution mode. The following parameters were used: capillary voltage, 3 kV; cone voltage, 40 V; source offset, 80 V; source temperature, 120°C; desolvation temperature, 350 °C; desolvation gas flow, 800 L/h. A solution of sodium formate was used for calibration. A solution of leucine enkephalin (H₂O/MeOH/formic acid: 49.95/49.95/0.1) was used as a lock mass (556.2766 m/z) and was injected every 20 s during each run. The lock mass correction was applied during data analysis and data was analysed using Masslynx[™] (Waters) software.

2.7.2 HPLC (High Performance Liquid Chromatography)

2.7.2.1 Analytical HPLC

Unless stated otherwise extracts of formicamycins and fasamycins were run on an Agilent 1290 UHPLC system using a Gemini-NX C18 00F0-4453-E0 column (150 × 4.6 mm, 3 μm, 100 Å, Phenomenex). Solvent A; 0.1% formic acid in water, solvent B; 100% methanol. The following gradient was used; 0 min 50% B, 2 min 50% B, 16 min 100% B, 18 min 50% B, 20 min 50% B.

Unless stated otherwise lantibiotics were run on an Agilent 1260 Infinity II UHPLC-MS/ELSD system using a Kinetex® XB-C18 00D-4605-E0 column (100 x 4.6 mm, 5 μm, 100 Å, Phenomenex). Solvent A; 0.1% formic acid in water, solvent B; 0.1% formic acid in acetonitrile. The following gradient was used; 0 min 10% B, 1 min 10%, 11 min 98% B, 13 min 98% B, 13.10 10% B, 15 min 10% B. Mass spectra was acquired in positive ion mode.

All data was analysed using ChemStation™ (Agilent) software.

2.7.2.2 Preparative HPLC

Preparative HPLC was carried out using a Thermo Scientific UltiMate 3000 system with a Gemini-NX C18 00F-4454-P0-AX column (150 x 21.2 mm, 5 μm, 110 Å, Phenomenex). Solvent A; 0.5% formic acid in water, solvent B; 0.5% formic acid in methanol or acetonitrile, flow rate; 20 mL/min. The following gradient was used: 0 min 5% B, 2 min 5% B, 22 min 100% B, 24 min 100% B, 25 min 5% B, 29 min 5% B. Wavelength observed was 210nm and fractions were collected manually into 50 mL glass test tubes on observation of peaks. Data was analysed using Chromeleon™ (Thermofisher) software.

Semi-Preparative HPLC

Semi-preparative HPLC was carried out using a Thermo Scientific UltiMate 3000 system with a Luna C18(2) 00G-4252-N0 column (250 x 10 mm, 5 μm, 100 Å,

Phenomenex) Solvent A; 0.5% formic acid in water, solvent B; 0.5% formic acid in methanol or acetonitrile, flow rate; 4 mL/min. The following gradient was used: 0 min 5% B, 4 min 5% B, 34 min 100% B, 38 min 100% B, 39 min 5% B, 49 min 5% B. Wavelength observed was 210nm and fractions were collected manually into 4 mL glass test tubes on observation of peaks. Data was analysed using Chromeleon™ (ThermoFisher) software.

2.7.2.3 Lantibiotic Purification

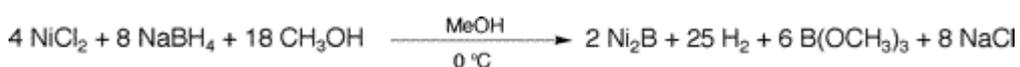
Lantibiotics were purified using the following steps after extraction on the Isolera One Biotage (2.6.1.2). Samples were dried down using a SP Scientific GenVac EZ-2 series under reduced pressure and weighed after each step. An aliquot of 100 µL was taken and analysed using UHPLC-HRMS or analytical UHPLC-MS/ELSD on the Agilent Infinity II system.

Table 2.12: HPLC Steps and Solvent for Purification of Lantipeptides

Step	Method	Solvent B
1	Preparative HPLC	Methanol+0.5% formic acid
2	Preparative HPLC	Acetonitrile+0.5% formic acid
3	Semi-Preparative HPLC	Methanol+0.5% formic acid
4	Semi-Preparative HPLC	Acetonitrile+0.5% formic acid

2.7.3 Tandem Mass Spectrometry (MS)

The organic thioether bonds in the lantibiotics were reduced to hydrocarbons (desulphurisation) by a nickel boride reaction (Ni₂B), prepared in situ from NiCl₂ and NaBH₄:



1 mg of lantipeptide was dissolved in methanol/water (1:1, 1 mL) before an aqueous solution of NiCl₂ was added (0.5 mL, 20 mg/mL, 42.1 µmol, Mw hexahydrate 237.71). The solution was mixed with 5 mg of solid NaBH₄ (132.2 µmol, Mw 37.83) in a 10 mL glass vial and sealed immediately. Hydrogen gas and a black Ni₂B precipitate were formed. The sample was heated for 4 hours at 55 °C with

vigorous stirring using a magnetic stirring plate (IKA®C-MAG HS 7) and water bath regulated with a thermostat (IKA®ETS-D5). The samples were then left to reach room temperature before a 20 µL aliquot was taken and diluted with 80 µL of a 1:1 mixture of 0.1% formic acid H₂O/methanol. The aliquot was centrifuged at 10,000 rpm for 5 min and the supernatant analysed by UHPLC-HRMS for the predicted fully reduced ion.

When the sample was fully reduced or 4 hours had passed the particulate was collected by centrifugation at 742 g for 10 min and the supernatant collected in a 4 mL vial. The precipitate was washed a second time with 1 mL methanol, subjected to ultrasonication for 30 min and centrifuged and supernatant collected as before. Solvent was then evaporated to dryness using a SP Scientific GenVac EZ-2 series before samples were redissolved in 10% acetonitrile (1 mL). A SPE C18 cartridge (Sep-Pak Plus, WAT020515) was washed with 10% acetonitrile (1 mL), sample applied, and the cartridge was washed with 10% acetonitrile and 100% acetonitrile, with flowthrough collected each time. The samples were evaporated to dryness as before and reconstituted in a 1:1 mixture of sterile water/methanol (1:1, 0.1 mL) before 1 µL was applied to a 2,5-dihydroxybenzoic acid (DHB) MALDI matrix and 1 µL of 10% TFA and 2 µL of 8mg/mL DHB in a 30% acetonitrile/0.1% TFA solution were added. The volatiles were evaporated at room temperature.

Tandem MS data was generated by Matrix Assisted Laser Desorption/Ionisation - Time of Flight (MALDI-TOF) LIFT using a Autoflex™ Speed MALDI-TOF/TOF mass spectrometer (Bruker Daltonics™ GmbH). The instrument was controlled by a FlexControl™ 3.4 (Bruker) method optimised for peptide detection and calibrated using peptide standards (Bruker). For sequence analysis fragments produced by Post Source Decay (PSD) were measured using the LIFT method (Bruker) (Suckau *et al*, 2003). All spectra were processed using FlexAnalysis™ 3.4 (Bruker) software.

2.7.4 Nuclear Magnetic Resonance (NMR)

NMR spectra were acquired on a Bruker Avance Neo 600 MHz spectrometer equipped with a 5 mm TCI cryoprobe. Experiments were carried out at 298K or 310K in DMSO- d_6 and chemical shifts reported in parts per million (ppm) relative to the residual deuterated solvent signals, which were used as an internal standard (δ_H/δ_C 2.50/39.5). 1H and ^{13}C resonances were assigned through 2D 1H - 1H COSY, 1H - ^{13}C -HSQCed, HMBC, TOCSY and NOESY experiments. Spectra processing and signal assignment was performed using Mnova14 software.

2.7.5 Carbohydrate Analysis

Purified fasamycin-like congeners were mixed with 1 mL of 1.0 M trifluoroacetic acid (TFA) and heated to 105 °C overnight. Samples were then diluted with 20 mL of Milli-Q water and the TFA removed by freeze drying. The residue was then suspended in 1 mL 5 % methanol and passed through a C18 solid phase extraction cartridge (Waters, Sep-Pak® Plus Short 360 mg C18), carbohydrates were then eluted with 5 % methanol (2 mL). The solvent was then dried using a SP Scientific GenVac EZ-2 series under reduced pressure, resuspended in water to a concentration of 100 μ M and carbohydrates analysed.

Carbohydrate analysis was performed using a Dionex ICS-5000 High performance anion exchange chromatography – pulsed amperometric detection (HPAE-PAD) instrument with a ThermoScientific CarboPac™ PA20 3 \times 150 mm analytical column. Solvent A; 7.8 mM NaOH, solvent B; 156 mM NaOH + 100 mM AcONa, flow rate; 0.25 mL/min, detection; pulsed amperometric detector, AgCl electrode, waveform *carbohydrates (standard method supplied with instrument), 0.25 mL/min. The following gradient was used: 0% B for 30 min, up to 100% B over 3 minutes, hold at 100% B for 20 min, down to 0% B over 3 min, hold 0% B for 14 min.

Sample injection volume was 5 μ L, compared to carbohydrate standards (Thermo Fisher Scientific) and verification performed with co-injection.

2.8 Phylogenetic Trees

Phylogenetic trees were created using the EMBL-EBI Simple Phylogeny tool. Firstly, amino acid sequences were aligned using the EMBL-EBI Clustal Omega multiple sequence alignment software using default settings. The alignment output was used to make a phylogenetic tree using the Simple Phylogeny tools default settings (Madeira *et al*, 2019). The generated phylogenetic tree was then visualised using Trex-Online-UQAM Newick viewer (Boc *et al*, 2012).

Chapter 3:
Diversification of
Class II Type B
Lantipeptides

3. Production and Diversification of Class II Type B Lantibiotics using a Heterologous Synthetic Expression System

3.1 Introduction

The widespread development of whole genome sequencing has resulted in a rise in published genomes featuring putative class II type B lantipeptide BGCs. Amongst this antibiotic family, duramycin has made it to clinical trials as a potential cystic fibrosis treatment (Oliynyk *et al.*, 2007) and cinnamycin has been used as a probe to further the understanding of cell lysis and division (Zhao, 2011), as well as an alternative treatment for atherosclerosis, due to its ability to induce trans-bilayer lipid movement (Märki *et al.*, 1991). Considering class II type B lantipeptides therapeutic potential the characterization of new members of this compound family is of great value. However, to date relatively few have been isolated, and only eight have been characterised at any level (Benndorf *et al.*, 2018; Chen *et al.*, 2017a; Kodani *et al.*, 2016; Märki *et al.*, 1991; Vikeli *et al.*, 2020).

Previously, a two-part synthetic cassette system which included a biosynthetic cassette (pWDW63) and activation cassette (pEVK6), described further in section 3.3, was created from the kyamicin BGC found in *Saccharopolyspora* KY21, a symbiont bacteria isolated from the cuticle of Kenyan *Tetraponera penzigi* fungus farming ants. Kyamicin is a class II type B lantipeptide comparable to cinnamycin, with homologous biosynthetic, transport and regulatory genes, which carry out the same post-translational modifications to the core peptide as previously described for the cinnamycin system in Chapter 1.4.4.1 (**Fig 3.1**). However, kyamicin has six amino acid residue differences in the core peptide compared to cinnamycin. The kyamicin expression platform was successfully used to produce kyamicin within the heterologous host *Streptomyces coelicolor* M1152 (Vikeli *et al.*, 2020). Through replacement of the *kyaA* core peptide with other *lanA* core peptides, the closely related lantibiotics cinnamycin B and duramycin were produced from the kyamicin biosynthetic machinery (Eleni Vikeli, JIC). Throughout this project *lan* shall be used

to refer to lantipeptide genes when discussed outside the context of a specific lantipeptide or used as the plural for lantipeptide genes from different BGCs.

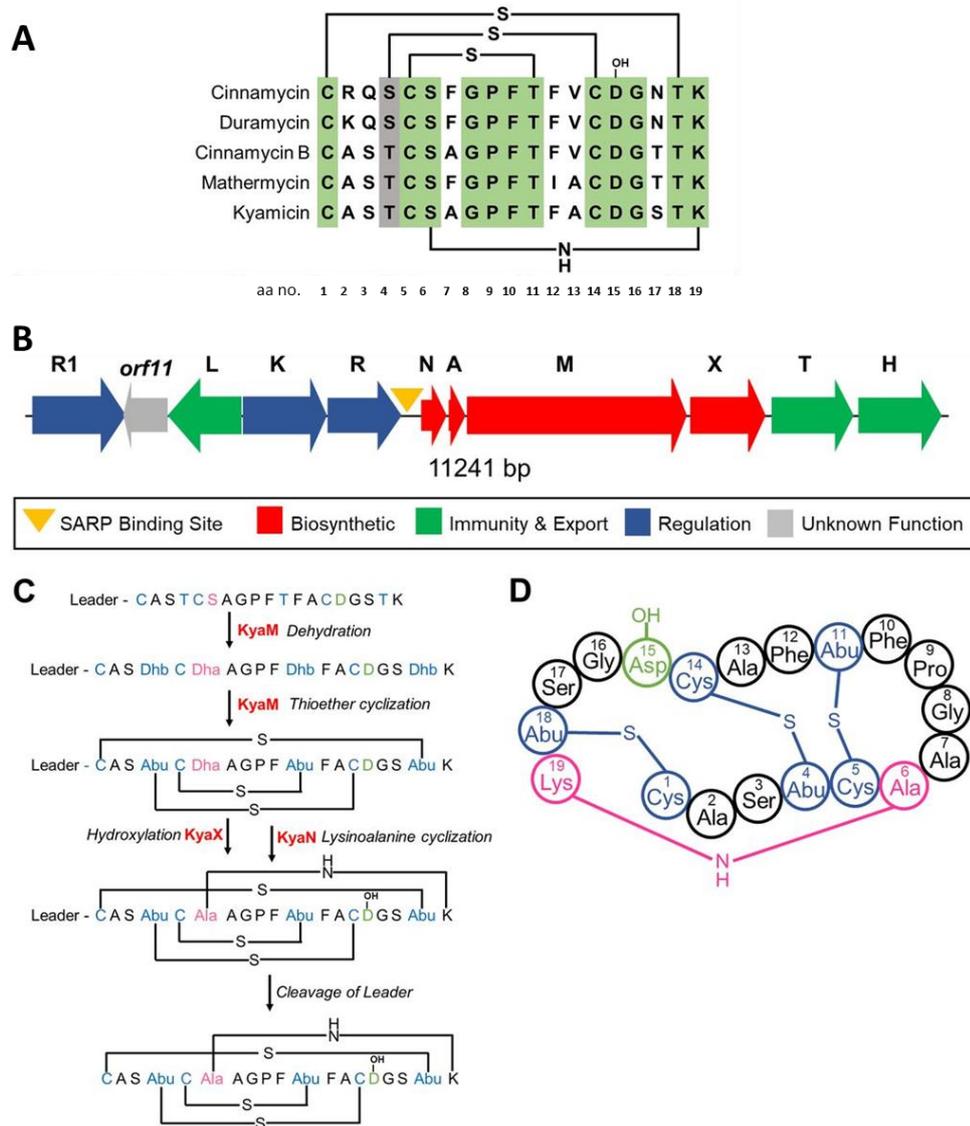


Figure 3.1: Kyamicin Peptide Sequence, BGC, Biosynthesis and Structure. (A) Core peptide alignment of kyamicin and known class II type B cinnamycin-like lantipeptides, with the positions of the characteristic thioether and lysinoalanine bridges of the mature peptide shown. Conserved residues (Green), similar residue (Grey). (B) The kyamicin BGC. (C) Schematic of kyamicin biosynthesis. Thioether bridges are formed first by KyaM causing dehydration of Thr4, Thr11, Thr18, and Ser6 to form dehydrobutyryne (Dhb) and dehydroalanine (Dha) residues, respectively. Dhb becomes S-linked aminobutyric acid (Abu) and Dha becomes S-linked Ala after thioester cyclization by KyaM. KyaX hydroxylates Asp15, and KyaN then forms the lysinoalanine bridge between Dha6 and Lys19. Lastly, following modification of the core peptide the leader peptide is proteolytically cleaved. (D) The 19 amino acid core peptide structure of kyamicin, with post-translational modifications. Blue: methyllanthionine bridges. Green: hydroxylated aspartate. Pink: lysinoalanine bridge. Adapted from (Vikeli *et al.*, 2020), *Applied and Environmental Microbiology*, 86(3), e01876-19, licensed under CC BY 4.0.

Using the kyamicin synthetic cassette system we first looked to produce novel lantipeptides identified within the public databases, with the aim to test their antimicrobial activity. The system was successful in enabling the production of 12 novel lantipeptides, including nine that displayed antibacterial activity against *Bacillus subtilis* EC1524.

In parallel, we attempted to build a second synthetic cassette system based upon the lantipeptide BGC found in *Streptomyces roseoverticillatus*. The new cassette system was devised following the same logic as that employed with the kyamicin BGC. However, this second example failed to produce any lantipeptides, including the putative product of the BGC. Modifications to the activation cassette and inclusion of a *kyaL* immunity gene on a separate expression vector failed to induce production of the lantipeptides from the rosierymycin biosynthetic cassette. However, *S. roseoverticillatus* grown under laboratory conditions showed antibacterial activity against *B. subtilis* EC1524 and produced rosierymycin, as identified through Ultra High Performance Liquid Chromatography Mass Spectrometry (UHPLC-MS). Attempts to overproduce rosierymycin through overexpression of transcriptional activators via incorporation of the kyamicin activator cassette (pAMA1) and *S. roseoverticillatus* activator cassette (pAMA5) led to a reduction in rosierymycin production, along with a distinct phylogenetic change.

Isolation of new lantipeptides was achieved through extensive use of HPLC. This work suggested the formation of potential epimers of the lysinoalanine bridge formed through both enzymatic and spontaneous chemical reaction (as reported for cinnamycin (Ökesli *et al.*, 2011)), and an uncyclized version lacking the lysinoalanine bridge which was biologically inactive. The bioactivities of the purified lantipeptides were assessed against several indicator strains, and their predicted structures validated through tandem mass spectrometry (Tandem MS) and nuclear magnetic resonance (NMR) of the compounds acquired.

3.2 Objectives

The objectives of the project were:

- Production of cryptic class II type B lantipeptides predicted by genome mining using the kyamicin synthetic cassette system
- Design of a further synthetic cassette system based on the *S. roseovorticillatus* BGC and investigation of its suitability for production of class II type B lantipeptides
- Isolation of novel class II type B lantipeptides for structural characterization and determination of their antibacterial activity

3.3 A Synthetic Cassette System for the Expression of Class II Type B Lantipeptides based on the Kyamicin BGC

A two-part cassette system was created from the kyamicin BGC to produce class II type B lantipeptides by David Widdick and Eleni Vikeli (at, JIC) before this project commenced. The system comprises of an activator (pEVK6) and biosynthetic cassette (pWDW70), the construction of which is outlined below.

Previous work investigating the regulatory mechanism of the homologous cinnamycin system, described further below in 3.3.1, had shown that CinR1 acted as a direct transcriptional activator for the *cinNAMX* biosynthetic operon and CinL was essential for immunity. While deletion of the two-component regulator CinKR only abolished cinnamycin production in the absence of *cinL* transcription. When *cinL* was transcribed under control of a constitutive promoter within an expression vector cinnamycin production was re-established, showing that CinKR is only involved in regulating CinL transcription (O'Rourke *et al.*, 2017). Based upon this work the kyamicin activator cassette comprises of only the *kyaR1* and *kyaL* genes under the control of a constitutive *ermE** promoter with the *kyaN* Ribosome Binding Site (RBS) as an intergenic link. The *kyaR1* gene encodes a *Streptomyces* Antibiotic Regulatory Protein (SARP), a pathway specific regulator (transcription factor), which goes on to activate the biosynthetic cassette. The *kyaL* gene encodes a PE-methyl transferase, which methylates the phospholipids of the cell membrane to prevent binding of lantipeptides, thus providing immunity. The RBS from *kyaN* was used as an intergenic region as it is an almost perfect example of an RBS.

The fragment *kyaR1L* was cloned into the expression vector pIJ102567 (Appendix 9.5) which utilises the ϕ BT1 integrase to site specifically integrate into the chromosome (Gregory *et al.*, 2003; Hong *et al.*, 2005); the final plasmid was named pEVK6 (Appendix 9.5) (**Fig 3.4**). Previously, the *kyaR1L* fragment had been cloned into expression vector pGP9 (which also utilises the ϕ BT1 integrase to site specifically integrate into the chromosome (Andexer *et al.*, 2011)) (Appendix 9.5), creating the plasmid pEVK4 (Appendix 9.5). Incorporation of pEVK4 into

Saccharopolyspora KY21 was then seen to induce production of the cryptic lantipeptide kyamicin (**Fig 3.2**) (Vikeli *et al.*, 2020).

The biosynthetic cassette contains the biosynthetic region from *kyaN* to *kyaH* of the kyamicin BGC and the intergenic region between *kyaN* and *kyaR1*, so as to include the SARP binding site. The biosynthetic region was cloned into expression vector pSET152 (Appendix 9.5), which utilises a ϕ C31 integrase to catalyse site specific integration into the chromosome (Bierman *et al.*, 1992); the final plasmid was named pWDW63 (Appendix 9.5). Introduction of pEVK6 and pWDW63 sequentially into the heterologous host *S. coelicolor* M1152 resulted in a zone of clearing against a *B. subtilis* EC1524 overlay. This was caused by production of kyamicin as verified by mass spectrometry. Small quantities of deoxykyamicin were also detected (**Fig 3.2**) (Vikeli *et al.*, 2020).

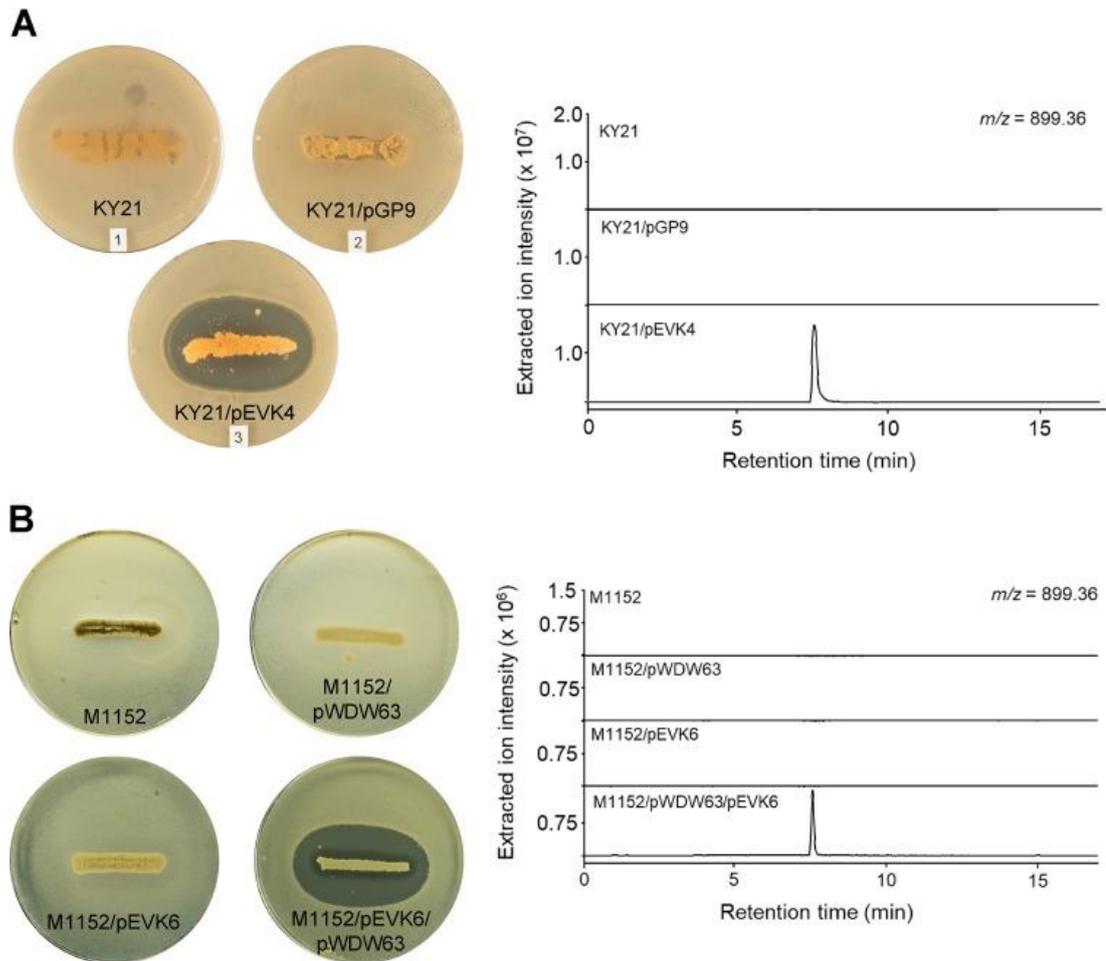


Figure 3.2: Activation of Kyamicin Biosynthesis in *Saccharopolyspora* KY21, and Expression in Heterologous Host *S. coelicolor* M1152. Bioassays were overlaid with indicator strain *B. subtilis* EC1524 and performed with biological triplicates. Extracts were analysed using UPLC-MS and extracted ion chromatograms are shown for calculated kyamicin m/z 899.36 ($[M + 2H]^{2+}$). (A) Kyamicin expression in *Saccharopolyspora* KY21. Activation cassette pEVK4 containing *kyaR1* and *kyaL* under control of a constitutive promoter induced production of kyamicin in KY21, resulting in a zone of inhibition. Wild-type strain and empty vector pGP9 are used as a negative controls. (B) Kyamicin expression in heterologous host *S. coelicolor* M1152. Activation cassette pEVK6, containing the *kyaR1L* fragment homologous to pEVK4, and kyamicin biosynthetic cassette pWDW63 induce production of kyamicin in M1152, resulting in a zone of inhibition. No zone of inhibition is seen for controls M1152/pEVK6 and M1152/pWDW63 showing the cassettes work in combination to produce kyamicin. Adapted from (Vikeli *et al.*, 2020), *Applied and Environmental Microbiology*, 86(3), e01876-19, licensed under CC BY 4.0.

The plasmid pEVK6 was also seen to induce the production of duramycin, a *Streptomyces* derived lantipeptide, in heterologous host *S. coelicolor* M1152. The plasmid pOJKKH was incorporated into *S. coelicolor* M1152 and contained the entirety of the duramycin biosynthetic region, including the SARP binding site upstream of *durN*, but lacking the regulatory and immunity genes (O'Rourke *et al.*, 2017); creating the strain *S. coelicolor* M1152/pOJKKH. When the plasmid pEVK6 was also incorporated into the chromosome duramycin was readily produced. This showed that the SARP and immunity genes of the kyamicin BGC, a *Saccharopolyspora* derived class II type B lantipeptide, can be used to activate *Streptomyces* derived class II type B lantipeptide BGCs, in a cross-genus activation (Fig 3.3) (Vikeli *et al.*, 2020).

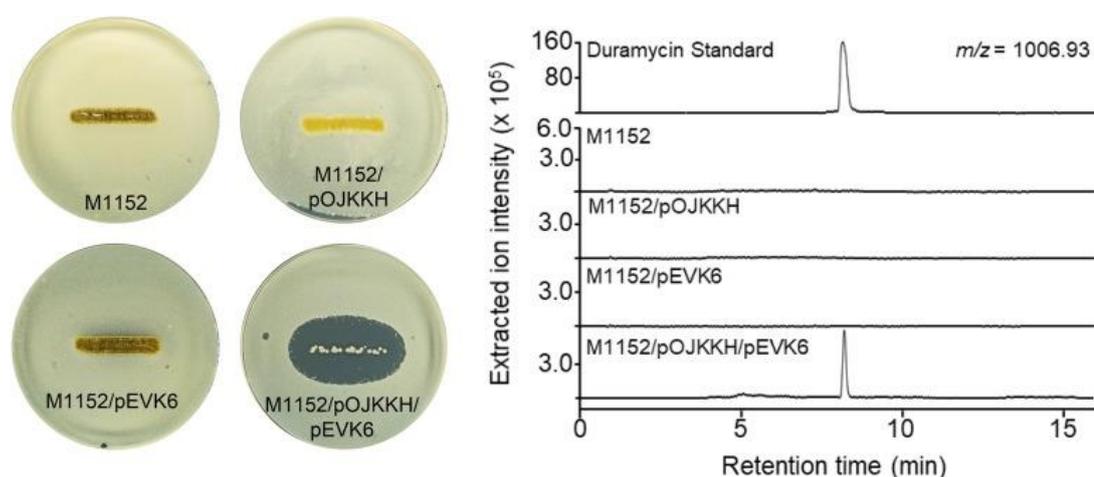


Figure 3.3: Duramycin Biosynthesis Induced in Heterologous Host *S. coelicolor* M1152 by Kyamicin Activation Cassette pEVK6. Bioassays were overlaid with *B. subtilis* EC1524 and performed with biological triplicates. Extracts were analyzed using UPLC-MS and extracted ion chromatograms are shown for the duramycin m/z 1,006.93 ($[M + 2H]^{2+}$). Duramycin is produced only from the M1152 strain carrying both pOJKKH and pEVK6, as shown by zone of inhibition. The duramycin peak detected aligns with an authentic standard of duramycin (0.1 mg/mL in 5% formic acid), displayed on a separate scale. No zone of inhibition is seen for controls M1152/pOJKKH and M1152/pEVK6, showing the cassettes work in combination to produce duramycin. Adapted from (Vikeli *et al.*, 2020), *Applied and Environmental Microbiology*, 86(3), e01876-19, licensed under CC BY 4.0.

To utilise the kyamicin biosynthetic cassette for the production of cryptic lantipeptides, the 19 amino acid core peptide encoding region of *kyaA* was removed from pWDW63 and replaced with a unique *Stu1* endonuclease restriction cut site that could be used to insert lantipeptide core peptide sequences identified through genome mining (Dave Widdick, JIC); the resulting plasmid was called pWDW70 (**Fig 3.4**). Thus, the expression of hybrid *kyaA* genes would lead to production of new KyaA variants, which, if used as substrates by the kyamicin biosynthetic genes, would lead to the production of new lantipeptide products.

To generate pWDW70 (Appendix 9.5) a PCR product covering a region encompassing a unique *EcoRI* site before *kyaN* to the beginning of the *kyaA* core peptide region was created using the biosynthetic cassette pWDW63 as a template. Primer design was used to engineer in a unique *StuI* site (AGG[^]CTT) upstream of the core peptide encoding region by utilising the last two nucleotides of codon E58 and the first nucleotide of codon V59 within *kyaA*, which formed the sequence AGG, after which a subsequent CTT sequence was added. Six random nucleotides then separated this *StuI* site from a *KpnI* (GGTAC[^]C) site at the end of the PCR fragment (GCG GGTACC TTAAA AGGCCT CGTTGGCCGCGATCCCCTTGG (KY3smStuI-R, Appendix 9.1.2). This PCR fragment was inserted by ligation into the vector pBlueScriptIIKS (Appendix 9.5) using the *EcoRI/KpnI* sites, to yield plasmid pWDW68 (Appendix 9.5).

Again, using pWDW63 as a template, a second PCR fragment was created encompassing from the stop codon of *kyaA* to a *KpnI* site located within *kyaX*, with a *StuI* site added upstream of the *kyaA* stop codon via primer design (GCG AGGCCT AGCATCGACGCGGTGAGCCTCC (KY3lgStuI-F, Appendix 9.1.2). This second PCR fragment was then cloned into pWDW68 via ligation at the *StuI/KpnI* sites. This gave plasmid pWDW69 (Appendix 9.5), encoding a fragment covering from *kyaN* through to *kyaX* but with the *kyaA* core peptide region removed and a unique *StuI* site engineered in its place. A PCR fragment was created using pWDW69 as a template encompassing from the unique *EcoRI* site before *kyaN* through to the *KpnI* site located within *kyaX*. This same region was removed from pWDW63 and the PCR

fragment inserted via ligation using the *EcoRI/KpnI* sites to give the biosynthetic cassette pWDW70, containing the whole kyamicin biosynthetic machinery but lacking the *kyaA* core peptide region (**Fig 3.4**). The biosynthetic cassette pWDW63, containing the kyamicin biosynthetic fragment with no modifications, was also used during this project as a positive control.

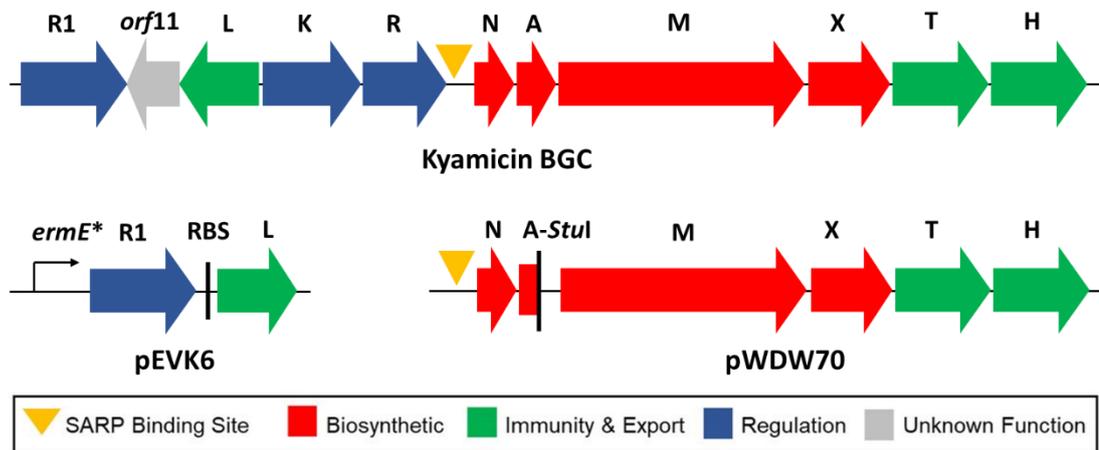


Figure 3.4: Kyamicin BGC and Two-Part Synthetic Cassette. pEVK6: pIJ102567 activation and immunity cassette; *ermE** constitutive promoter, *kyaR1* SARP, *kyaN* RBS, *kyaL* methyltransferase. pWDW70: pSET152 biosynthetic cassette; SARP binding site, *kyaN*, *kyaA* with core peptide removed and replaced with unique *Stu1* endonuclease restriction cut site, *kyaM*, *kyaX*, *kyaT*, *kyaH*. Adapted from (Vikeli *et al.*, 2020), *Applied and Environmental Microbiology*, 86(3), e01876-19, licensed under CC BY 4.0.

Considering the above proposed mechanism of regulation and the unknown function of *cinOrf11* we created a larger activator cassette containing the full regulatory and immunity region of the kyamicin BGC, from *kyaR1* to *kyaR*. Using *Saccharopolyspora* KY21 genomic DNA this region was generated using PCR and the fragment was inserted into expression vector pIJ102567 (Gregory *et al.*, 2003; Hong *et al.*, 2005) using the *NdeI/PacI* sites, to give pAMA1 (Appendix 9.5) (Fig 3.6). This plasmid, along with the smaller activator cassette pEVK6 described previously, were used to generate production of kyamicin in conjunction with the biosynthetic cassette pWDW63 (Fig 3.6).

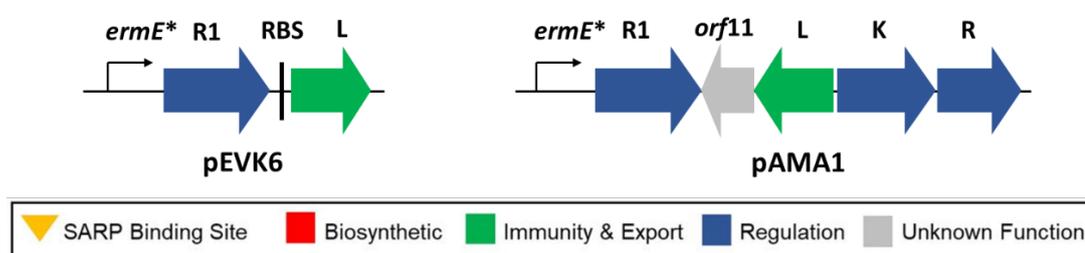


Figure 3.6: Kyamicin Activation and Immunity Cassettes. pEVK6: pIJ102567 activation cassette; *ermE** constitutive promoter, *kyaR1* SARP, *kyaN* RBS, *kyaL* methyltransferase. pAMA1: pIJ102567 activation cassette; *ermE** constitutive promoter, *kyaR1* SARP, *kyaOrf11*, *KyaL* methyltransferase, *KyaK* regulator, *kyaR* regulator.

Further to optimizing the activation cassette, we wished to compare heterologous hosts *S. coelicolor* M1152 (M1152) and *Streptomyces coelicolor* M1146 (M1146) (Gomez-Escribano & Bibb, 2011). These strains only differ by a single C1298T point mutation in the *rpoB* ribosome gene within M1152, which was previously shown to increase transcription levels of secondary metabolite BGCs (Gomez-Escribano & Bibb, 2011). It was reported that this mutation resulted in a marked decrease in growth rate and sporulation of M1152 when compared to M1146. This decrease in growth rate results in experiments taking longer to complete, and the decrease in sporulation causes greater difficulty in collecting viable spores, necessary for efficient conjugation of the kyamicin cassettes. Thus, we wanted to determine whether M1146, which has better growth characteristics overall, would be a suitable heterologous host for this work.

Following the introduction of activator cassettes pEVK6 or pAMA1 to *S. coelicolor* M1146/pWDW63 and *S. coelicolor* M1152/pWDW63, the resulting strains were tested for their levels of kyamicin production using bioassays against *B. subtilis* EC1524. We observed that the zone of inhibition in the *B. subtilis* EC1524 overlay was consistently smaller for the larger activator cassette pAMA1 compared to pEVK6, consistent with a lower level of kyamicin production as determined by UHPLC-MS (**Fig 3.7**). No observable difference could be seen between expression of pEVK6/pWDW63 in the two strains M1146 and M1152 (**Fig 3.8**). Thus, the pEVK6 cassette was used as the standard activator cassette for the kyamicin cassette system and heterologous host M1146 was used for all work going forward.

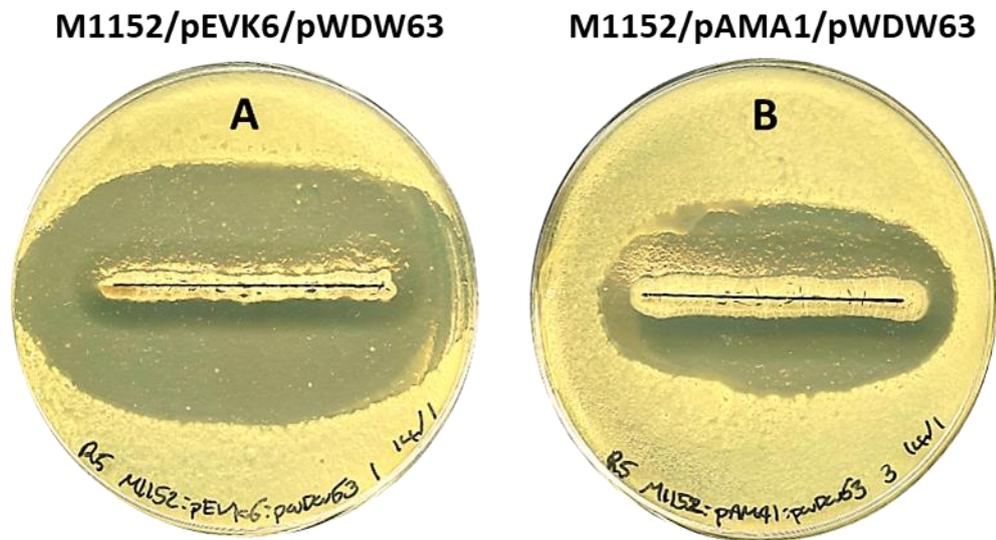


Figure 3.7: Comparison of Kyamicin production Induced by Activation Cassettes pEVK6 and pAMA1. Kyamicin biosynthetic cassette pWDW63, heterologous host *S. coelicolor* M1152, bioassays overlaid with *B. subtilis* EC1524, performed with biological triplicates. Kyamicin production shown by size of zone of inhibition. pEVK6 activator cassette contains *kyaR1* and *kyaL*. pAMA1 activator cassettes contains *kyaR1*, *kyaOrf11*, *kyaL*, *kyaK* and *kyaR*. (A) M1152/pEVK6/pWDW63 (B) M1152/pAMA1/pWDW63. Activator cassette pAMA1 induced less kyamicin production than activator cassette pEVK6, as seen by the zone of inhibition.

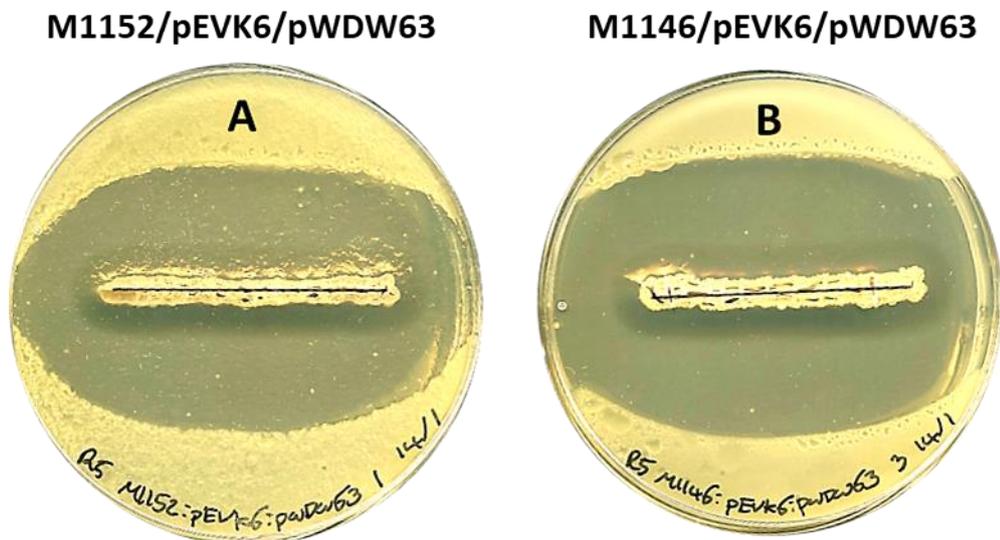


Figure 3.8: Comparison of Heterologous Hosts *S. coelicolor* M1152 and *S. coelicolor* M1146 for Kyamicin Production. Kyamicin activation cassette pEVK6, kyamicin biosynthetic cassette pWDW63, bioassays overlaid with *B. subtilis* EC1524, performed with biological triplicates. Kyamicin production shown by size of zone of inhibition. (A) M1152/pEVK6/pWDW63, (B) M1146/pEVK6/pWDW63. Kyamicin production in both heterologous hosts was comparable, as seen by the zone of inhibition.

3.4 Genome mining for BGCs Encoding Class II Type B Lantibiotics

Potential class II type B lantipeptide BGCs were identified through protein-protein comparison using NCBI BLAST (Pertsemlidis & Fondon, 2001) of KyaA, KyaN and KyaX against the free online databases. The *kyaA* gene encodes the lantipeptide core peptide region which is conserved across class II type B lantipeptides, as seen by previous comparison of kyamicin, cinnamycin, duramycin and other known lantipeptides *lanA* core peptide sequences (**Fig 3.1A**) (Vikeli *et al.*, 2020). However, the short length of the core peptides amino acid sequence, being only 19 amino acids long, means the probability of identifying homologous sequences in non-lantipeptide BGCs is high, and occurred on occasion during this work. Thus, the genes for KyaN and KyaX, previously discussed in Chapter 1.4.1.1 and Chapter 3.1, which are responsible for lysinoalanine bridge formation and Asp15 hydroxylation respectively, were also used for protein-protein comparison to identify true lantipeptide BGCs. These two post-translational modification are also characteristic of class II type B lantipeptides specifically, and are not shared amongst other classes of lantipeptides, nor class II type A lantipeptides as discussed in Chapter 1.4.1.1 (van der Donk & Nair, 2014).

In addition to the lantipeptide BGCs identified through genome mining, four unpublished class II type B lantipeptide *lanA* sequences (NAI698, NAI711, NAI716, NAI113) from strains related to the *Actinomadura* genus were donated by collaborators NAICONS, Milan, Italy. The lantipeptide core peptide regions used in this thesis were numbered using a L_XX system, with kyamicin designated L_00 (**Table 3.1**).

Table 3.1: Lantipeptide Core Peptides, Designation, GeneBank Accession Number

Origin	Core Peptide sequence (amino acid)	Designation	GeneBank Accession No.
<i>Saccharopolyspora</i> KY21(Kyamycin) (Vikeli <i>et al.</i> , 2020)	CASTCSAGPFTFACDGSTK	L_00	MK251553.1
<i>Streptomyces</i> <i>cinnamoneus</i> (Cinnamycin) (Lindenfelser <i>et al.</i> , 1959)	CRQSCSFGPFTFVCDGNTK	L_01	WP_190107705.1
<i>Streptomyces</i> <i>griseovercillatus</i> (Duramycin) (Hayashi <i>et al.</i> , 1990)	CKQSCSFGPFTFVCDGNTK	L_02	DD308185.1
<i>Streptoverticillium</i> R2075 (Duramycin B) (Fredenhagen <i>et</i> <i>al.</i> , 1990)	CRQSCSFGPLTFVCDGNTK	L_03	P36502.1
<i>Streptomyces</i> <i>griseoluteus</i> (R2107) (Duramycin C) (Fredenhagen <i>et</i> <i>al.</i> , 1990)	CANSCSYGPLTWSCDGNTK	L_04	P36503.1
<i>Actinomadura</i> spp. 5-7 (Rubrominin A) (Benndorf <i>et al.</i> , 2018)	CSSTCTSGPFTFACDGTTKG	L_05	WP_103565569.1
<i>Actinomadura</i> spp. 5-7 (Rubrominin B) (Benndorf <i>et al.</i> , 2018)	ACSSTCTSGPFTFACDGTTKG	L_06	WP_103565569.1

<i>Marinactinospora thermotolerans</i> (Mathermycin) (Chen <i>et al.</i> , 2017a)	CASTCSFGPFTIACDGTTK	L_07	WP_078762901.1
<i>Streptomyces africans</i>	CVQSCSFGPITAVCDGNTK	L_08	WP_086559519.1
<i>Thermosporothrix hazaken</i>	CVQSCSFGPITAICDGNTK	L_09	WP_111319848.1
<i>NAI698_Chromcin</i>	CQTSCSWGPIAVCDGTTK	L_10	N/A
<i>NAI711_Draftcin</i>	CRTSCSWGPIAVCDGTTK	L_11	N/A
<i>Nocardiopsis potens</i>	CRTSCSWGPFITACDGGTKP	L_12	WP_017591584.1
<i>NAI113cinA_59-7</i>	CGTSCSWGPFITIVCDGQTK	L_13	N/A
<i>Streptomyces roseovercillatus</i> (Rosiermycin) (This work)	CTSSCSSGRFTIICDGGTK	L_14	WP_030367889.1
<i>Actinomadura rubrobrunea</i>	CATSCSAGPFTIICDGGTK	L_15	WP_067909139.1
<i>NAI16cinA_60-8</i>	CATTCSAGPFTIICDGATKA	L_16	N/A
<i>Actinomadura macra</i>	CASTCSSGPFTFACDGGTKA	L_17	WP_067451807.1
<i>Actinomadura mexicana</i>	CASTCSSGPFTFACDGGTK	L_18	WP_089310450.1
<i>Nocardiopsis trehalosi</i>	CESTCSFGPFTFVCDGTSK	L_19	WP_076442089.1
<i>Actinomadura oligospora</i>	CISTCSYGPTTIICDGATKVG	L_20	WP_026413721.1
<i>Actinomadura oligospora</i>	CSSTCSFGPFTIVCDGTTKGQ	L_21	WP_026411382.1
<i>Oscillatoria spp.</i> <i>PCC 10802</i>	CRATCSFGPFTIVCDGTTK	L_22	WP_017718744.1
<i>Actinomadura spp.</i> <i>NEAU-G17</i>	CASTCTSGPLTFICDGTTK	L_23	WP_117357562.1

<i>Scytonema millei</i>	CKSTCTQGPYTIICDGTTK	L_24	WP_069351382.1
<i>Prochloron didemnid</i>	CASTCSSGPITAICDGTTK	L_25	ARD09210.1
<i>Prochloron didemnid</i>	CASTCSFGIVTIVCDGTTK	L_26	ARD09202.1
<i>Moorea producens</i> PAL	CESTCSFGIVTLICDGTTKS	L_27	WP_044491894.1
<i>Frankia</i> spp. EUN1F	CKTTCSSGPITIVCDGNTK	L_28	WP_026413721.1
<i>Moorea producens</i>	CQPSCDLGALTIVCDGVTK	L_29	WP_008178013.1
<i>Cylindrospermum stagnale</i>	CQTTCSFGWTVYCDGNTL	L_30	WP_015207184.1
<i>Cylindrospermum stagnale</i>	CRRTC VSGYTIRCDGVTV	L_31	WP_015208675.1
<i>Oscillatoria</i> spp. PCC 10802	CRRTC VSGWTIRCDGATV	L_32	WP_017720753.1
<i>Nostoc</i> spp. MBR_98	CDSTCVSGWTILCDGSTF	L_33	OCQ94429.1
<i>Tychonema bourrellyi</i>	CRCTCVTGFTLRCDGTSM	L_34	WP_096832221.1
<i>Microbispora rosea</i>	CPATHQCIHTHVCRKTY	L_35	WP_076442089.1

3.5 Heterologous Production of Lantipeptides using the Kyamicin Synthetic Cassette System

Synthetic DNA fragments encoding the lantipeptide core peptide sequence, with end regions homologous to pWDW70 extending 30 bp either side of the unique *Stu1* site, were ordered from Eurofins or Integrated DNA Technologies (IDT). The nucleotide sequences of the lantipeptide core peptides were not codon optimized and translated from amino acid sequences using Bioinformatics Sequence Manipulation Suite Reverse Translation tool (Stothard, 2000). Gibson assembly was used to clone these fragments into pWDW70 that had been linearised with *StuI*. This led to hybrid *lanA* genes comprised of the *kyaA* leader sequence and the target core peptide region. Constructs were confirmed by Sanger sequencing and transformed into *E. coli* ET12567/pUZ8002 cells before conjugation into *S. coelicolor* M1146/pEVK6. Strains were designated M1146/pEVK6/L_XX respectively and verified through antibiotic selection and colony PCR followed by Sanger sequencing.

3.5.1 Bioassays of Hybrid *lanA* Lantipeptide Construct Strains in the Kyamicin Cassette System against *B. subtilis*

All M1146/pEVK6/L_XX strains were streaked onto R5 plates, incubated for 7 days at 30°C and then overlaid with *B. subtilis* EC1524 in SNA media and incubated further overnight. M1146 and M1146/pEVK6/pWDW70 (empty vector) were used as negative controls. M1146/pEVK6/pWDW63 was used as a kyamicin producing positive control which causes a clear zone of inhibition against *B. subtilis* EC1524, as seen previously (Vikeli *et al.*, 2020).

Strains containing plasmids expressing genes containing the known cinnamycin, duramycin, duramycin B and mathermycin core peptides (L_01, L_02, L_03, L_07 respectively) showed clear bioactivity, indicating their successful expression and processing by the kyamicin biosynthetic genes. Furthermore, strains expressing hybrid *lanA* genes encoding for the cryptic lantipeptides L_10, L_11, L_13, L_15, L_18, L_19, L_22, L_23 and L_26 also showed antibacterial activity against *B. subtilis* EC1524 (**Fig 3.9**). Other than the strains expressing the genes designated L_22 and

L_26 which originate from marine dwelling Cyanobacteria species, all other producing species belonged to the Actinobacteria phylum.

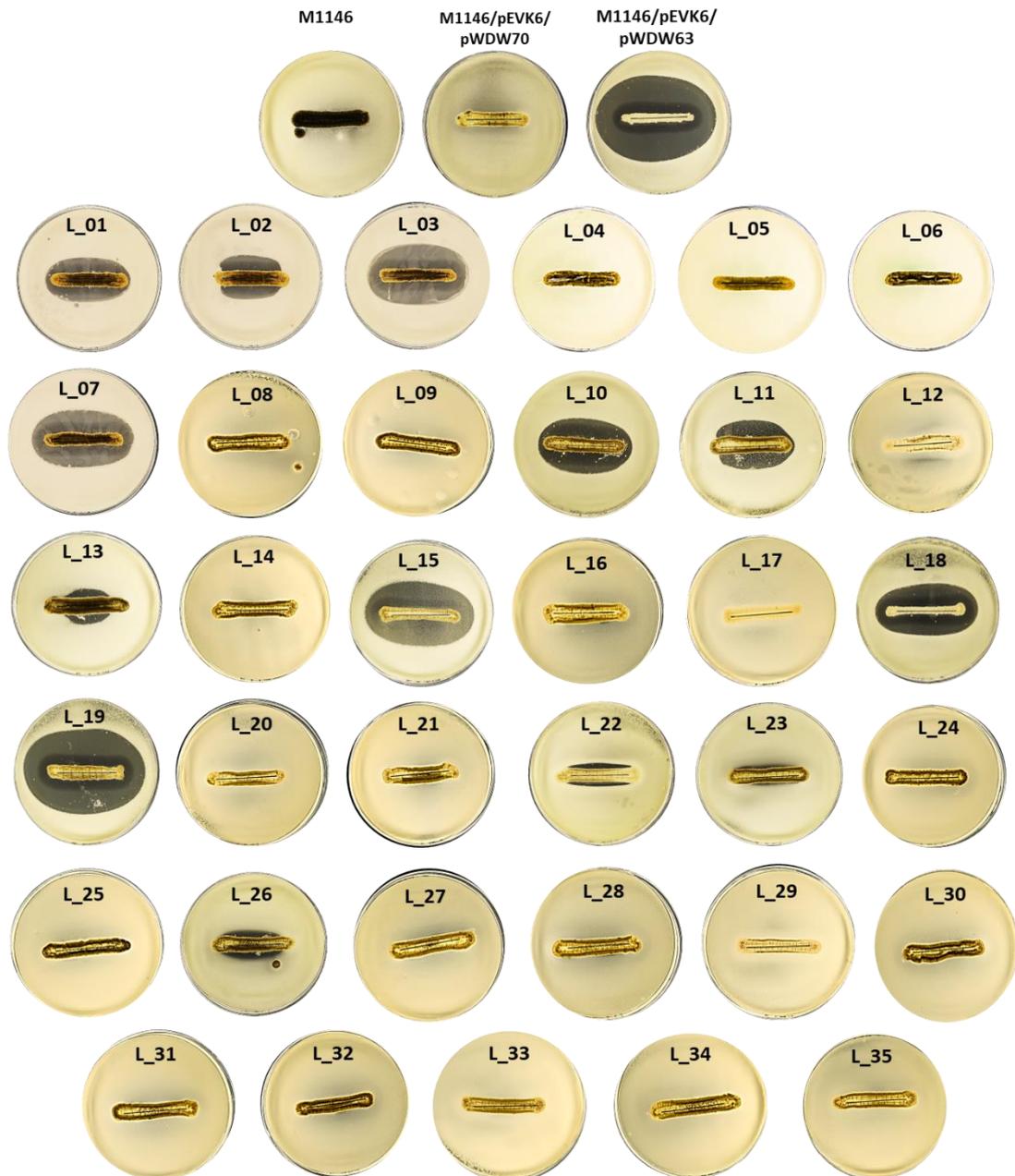


Figure 3.9: Zone of Inhibition caused by Lantipeptides produced through the Heterologous Kyamicin Synthetic Cassette System. Overlay *Bacillus subtilis*. Heterologous host *Streptomyces coelicolor* M1146. Plates were produced in triplicate. Strains consist of M1146/pEVK6/L_XX. Strains M1146 and M1146/pEVK6/pWDW70 were used as negative controls. Strain M1146/pEVK6/pWDW63 was used as a kyamicin producing positive control, with a clear zone of inhibition seen. Strains with zone of inhibition; L_01, Cinnamycin; L_02, Duramycin; L_03, Duramycin B; L_07, Mathermycin, L_10, L_11, L_13, L_15, L_18, L_19, L_22, L_23, L_26

3.5.2 Chemical Analysis of Lantipeptides Produced by Heterologous Expression

Agar plugs were taken from all bioassay plates described above, freeze-thawed and lantipeptides extracted by shaking in methanol for 20 min. All strains were also grown for 5 days in liquid TSB media at 37 °C, 250 rpm, after which aliquots were taken and lantipeptides extracted by shaking with methanol for 1 hour. Extracts were analysed using UHPLC-HRMS on a Synapt G2-Si high resolution mass spectrometer using electrospray ionization. For each predicted product molecule a m/z value was calculated based on the core peptide amino acid formula and anticipated post translational modifications corresponding to a loss of four water molecules and the addition of a hydroxyl group at the Asp15 residue. We also searched for congeners lacking the Asp15 hydroxylation.

All strains showing zones of inhibition gave LCMS chromatograms on a Synapt G2-Si high resolution mass spectrometer showing molecules consistent with the expected masses for the mature lantipeptides and/or a deoxy form; the high-resolution MS data for these was within ± 5 ppm of the calculated values. The compound produced by the L_22 strain was only observed in the deoxy form, and only the product of the L_03 (duramycin B) strain was produced exclusively in its hydroxylated form; all other strains produced mixtures of hydroxylated and deoxy forms (Appendix 9.2.1). Analysis of extracts made from strains that did not inhibit *B. subtilis* EC1524, indicated that lantipeptides were also produced by strains L_16 (hydroxylated and deoxy form), L_24 (deoxy form only) and L_25 (deoxy form only) (**Fig 3.10**).

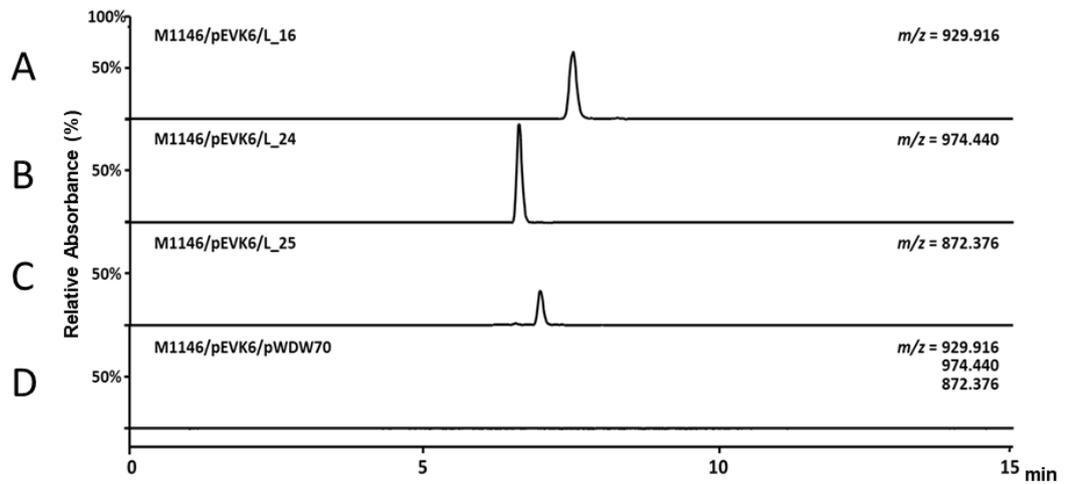


Figure 3.10: Extracted Ion Chromatograms (XICs) of Bioassay Plate Extracts from Strains L_16, L_24 and L_25. Agar plugs extracts taken from *B. subtilis* EC1524 overlay plates, analyses by UHPLC-MS. Masses described correspond to calculated $[M + 2H]^{2+}$ ions of predicted lantipeptide compound. (A) L_16 deoxy; m/z observed = 929.9160, m/z calculated = 929.9179, $\Delta = -2.0432$ ppm. (B) L_24 deoxy; m/z observed = 974.4417, m/z calculated = 974.4399, $\Delta = -1.8472$ ppm. (C) L_25 deoxy; m/z observed = 872.3754, m/z calculated = 872.3786, $\Delta = -3.668132$ ppm. (D) pWDW70 empty vector negative control.

3.5.3 Comparison of Produced and Not Produced Lantipeptide Core Peptides

In an attempt to identify features of the lantipeptide core peptide sequences that determined successful lantipeptide production in the kyamicin cassette system the core peptide sequences of the lantipeptides that were produced were compared to those not produced. The lantipeptides were then ranked by the length of the core peptide sequence (**Table 3.2**) and by the number of amino acid changes in the core peptide sequence when compared to the core peptide sequence of kyamicin (**Table 3.3**). Several trends were seen; firstly, successfully produced lantipeptides generally had a core peptide of 19 amino acids long, equal in length to that of kyamicin. With only a single produced lantipeptide, L_16, having a core peptide sequence 20 amino acids long. Secondly, produced lantipeptides generally retained the conserved GPFT region containing the central proline amino acid, which plays an important part in determining the structural conformation of the lantipeptide product and thus formation of the PE binding pocket (Elvas *et al*, 2017; Zhao, 2011). Thirdly, when compared to kyamicin, the core peptides of the produced lantipeptides had on average a lower number of amino acid residue changes than the core peptides of lantipeptides not produced. With the core peptides of produced lantipeptides having an average of 5.6 amino acid residue differences compared to the kyamicin core peptide, while core peptides of lantipeptides that were not produced had an average of 8.2 amino acid residue differences. However, some non-produced lantipeptide core peptide sequences were highly similar to kyamicin. Such as L_17, which only differed by two amino acid residues. So, similarity to the kyamicin core peptide sequence could not guarantee production of lantipeptides using the kyamicin heterologous cassette system.

Table 3.2: Length, Proline Inclusion and No. of Amino Acid (aa) Changes compared to Kyamicin of Hybrid Lantipeptide Core Peptides. Table ranked by core peptide length

Kyamicin core peptide	CASTCSAGPFTFACDGSTK								
Produced	Length (aa)	Proline (Y/N)	No. aa changes from kyamicin core peptide	Core Peptide Amino Acid Sequence	Not Produced	Length (aa)	Proline (Y/N)	No. aa changes from kyamicin core peptide	Core Peptide Amino Acid Sequence
L_01	19	Y	6	CRQSCSFGPFTFVCDGNTK	L_35	17	N	15	CPATHQCIHTHVCRKTY
L_02	19	Y	6	CKQSCSFGPFTFVCDGNTK	L_30	18	N	9	CQTTCSEFGWTVYCDGNTL
L_03	19	Y	7	CRQSCSFGPLTFVCDGNTK	L_31	18	N	10	CRRTCIVSGYTIRCDGVTV
L_07	19	Y	3	CASTCSFGPFTIACDGTTK	L_32	18	N	10	CRRTCIVSGWTIRCDGATV
L_10	19	Y	8	CQTSCSWGPITAVCDGTTK	L_33	18	N	9	CDSTCVSGWTILCDGSTF
L_11	19	Y	8	CRTSCSWGPITAVCDGTTK	L_34	18	N	11	CRCTCVTGFTLRCDGTSM
L_13	19	Y	7	CGTSCSWGPFTIVCDGQTK	L_04	19	Y	8	CANSCSYGPLTWSCDGNTK
L_15	19	Y	5	CATSCSAGPFTIICDGGTK	L_08	19	Y	8	CVQSCSFGPITAVCDGNTK
L_18	19	Y	1	CASTCSSGPFTFACDGSTK	L_09	19	Y	8	CVQSCSFGPITAVCDGNTK
L_19	19	Y	4	CESTCSFGPFTFVCDGTSK	L_14	19	N	7	CTSSCSSGRFTIICDGGTK
L_22	19	Y	6	CRATCSFGPFTIVCDGTTK	L_28	19	Y	7	CKTTCSSGPITIVCDGNTK
L_23	19	Y	5	CASTCTSGPLTFICDGTTK	L_29	19	N	10	CQPSCDLGALTIVCDGVTK
L_24	19	Y	7	CKSTCTQGPTYIICDGTTK	L_05	20	Y	5	CSSRCTSGPFTFACGTTKG
L_25	19	Y	5	CASTCSSGPITAVCDGTTK	L_12	20	Y	8	CRTSCSWGPFTIACDGSTKP
L_26	19	N	6	CASTCSFGIVTIVCDGTTK	L_17	20	Y	2	CASTCSSGPFTFACDGSTKA
L_16	20	Y	6	CATTCAGPFTIICDGATKA	L_27	20	N	8	CESTCSFGIVTLICDGTTKS
					L_06	21	Y	6	ACSSTCTSGPFTFACDGTTKG
					L_20	21	Y	8	CISTCSYGPTTIICDGATKVG
					L_21	21	Y	7	CSSTCSFGPFTIVCDGTTKGQ

Table 3.3: Length, Proline Inclusion and No. of Amino Acid (aa) Changes compared to Kyamicin of Hybrid Lantipeptide Core Peptides. Table ranked by no. of amino acid changes compared to the kyamicin core peptide

Kyamicin core peptide	CASTCSAGPFTFACDGSTK								
Produced	Length (aa)	Proline (Y/N)	No. aa changes from kyamicin core peptide	Core Peptide Amino Acid Sequence	Not Produced	Length (aa)	Proline (Y/N)	No. aa changes from kyamicin core peptide	Core Peptide Amino Acid Sequence
L_18	19	Y	1	CASTCSSGPFTFACDGSTK	L_17	20	Y	2	CASTCSSGPFTFACDGSTKA
L_07	19	Y	3	CASTCSFGPFTIACDGTTK	L_05	20	Y	5	CSSRCTSGPFTFACGTTKG
L_19	19	Y	4	CESTCSFGPFTFVCDGTSK	L_06	21	Y	6	ACSSTCTSGPFTFACDGTTKG
L_15	19	Y	5	CATSCSAGPFTIICDGGTK	L_14	19	N	7	CTSSCSSGRFTIICDGGTK
L_23	19	Y	5	CASTCTSGPLTFICDGTTK	L_21	21	Y	7	CSSTCSFGPFTIVCDGTTKGQ
L_25	19	Y	5	CASTCSSGPITAICDGTTK	L_28	19	Y	7	CKTTCSSGPITIVCDGNTK
L_01	19	Y	6	CRQSCSFGPFTFVCDGNTK	L_04	19	Y	8	CANSCSYGPLTWSCDGNTK
L_02	19	Y	6	CKQSCSFGPFTFVCDGNTK	L_08	19	Y	8	CVQSCSFGPITAVCDGNTK
L_16	20	Y	6	CATTCAGPFTIICDGATKA	L_09	19	Y	8	CVQSCSFGPITAICDGNTK
L_22	19	Y	6	CRATCSFGPFTIVCDGTTK	L_12	20	Y	8	CRTSCSWGPFITACDGSTKP
L_26	19	N	6	CASTCSFGIVTIVCDGTTK	L_20	21	Y	8	CISTCSYGPTTICDGATKVG
L_03	19	Y	7	CRQSCSFGPLTFVCDGNTK	L_27	20	N	8	CESTCSFGIVTLICDGTTKS
L_13	19	Y	7	CGTSCSWGPFITIVCDGQTK	L_30	18	N	9	CQTTCSFGWTVYCDGNTL
L_24	19	Y	7	CKSTCTQGPTYIICDGTTK	L_33	18	N	9	CDSTCVSGWTILCDGSTF
L_10	19	Y	8	CQTCSCSWGPIAVCDGTTK	L_29	19	N	10	CQPSCDLGALTIVCDGVTK
L_11	19	Y	8	CRTSCSWGPIAVCDGTTK	L_31	18	N	10	CRRTCIVSGYTIRCDGVTV
					L_32	18	N	10	CRRTCIVSGWTIRCDGATV
					L_34	18	N	11	CRCTCVTGFTLRCDGSTM
					L_35	17	N	15	CPATHQCIHTHVCRKTY

To further compare the produced and not produced lantipeptides the core peptides for each were aligned with the core peptide sequence of kyamicin (**Fig 3.1**). The produced lantipeptides were seen to be strongly conserved across most residues, including all residues involved in the formation of the conserved lanthionine, methyllanthionine and lysinoalanine bonds of class II type B lantipeptides. In comparison for several of the not produced lantipeptides these core residues are at spatially different positions compared to the kyamicin core peptide residues due to difference in the length of the core peptide region (L_05, L_30, L_31, L_32, L_33, L_34 and L_35). This could affect the ability of the kyamicin biosynthetic enzymes to form the lanthionine, methyllanthionine and lysinoalanine bonds, and thus effect the production of the lantipeptides. However, several of the not produced lantipeptides do contain all conserved residues in the same position as kyamicin, so changes to other amino acid residues is enough to affect the production of the lantipeptides.

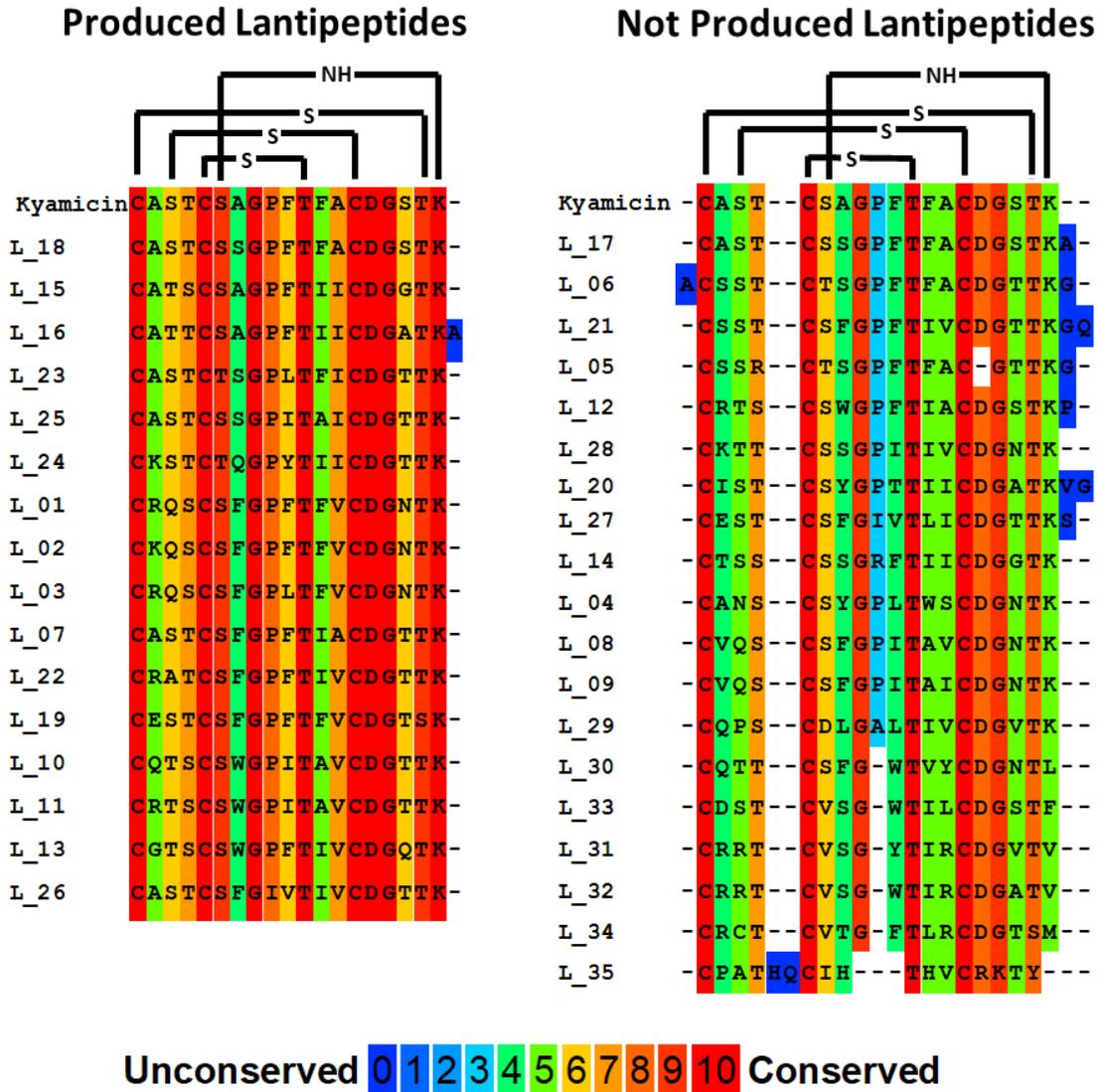


Figure 2.11: Sequence Alignment of Lantipeptides Produced and Not Produced using the Kyamicin Heterologous Cassette Platform. Conserved lanthionine, methylanthionine and lysinoalanine bonds shown. Core peptide sequences aligned against the kyamicin core peptide sequence. Uncovered residues shaded blue, strongly conserved residues shaded red.

To try and further distinguish if there was any relationship between the lantipeptides produced by the kyamicin heterologous production platform, phylogenetic trees were created from alignments of all the lantipeptides conserved core peptide sequences and *lanA* peptides. The phylogenetic trees were generated using the EMBL-EBI Clustal Omega alignment and Simple Phylogeny tool, as detailed in Chapter 2.8. However, no clustering was observed between the produced and non-produced lantipeptides *lanA* or core peptide sequences, and thus no

relationship between evolutionary clade and lantipeptide production could be distinguished (Appendix 9.2.2).

3.6 Design of an Alternative Expression System using the *Streptomyces roseovercillatus* lantipeptide BGC

Another synthetic cassette system based upon the lantibiotic BGC of *Streptomyces roseovercillatus* was designed following the same logic as the kyamicin synthetic cassette system. The reason for this was twofold; firstly, to try and produce the remaining naturally occurring lantipeptides not produced via the kyamicin synthetic cassette system; and secondly, to compare the two synthetic cassettes and provide insight into potential factors affecting the success of lantipeptide production, as well as constraints upon the systems.

3.6.1 *Streptomyces roseovercillatus* BGC

S. roseovercillatus is genetically distinct from *S. erythraea* KY3, the originator strain of the kyamicin BGC, and its lantipeptide BGC was selected as the basis for the secondary expression system for several reasons. Firstly, its core peptide was not expressed within the kyamicin synthetic cassette system so its corresponding lantipeptide remained cryptic. Secondly, this core peptide did not contain the conserved proline, a trait shared with many other lantipeptides not produced using the kyamicin system. Lastly, the organisation of the lantipeptide BGC was identical to that of kyamicin, with the genes clearly separated into a regulatory and immunity region and a biosynthetic and transport region, and homologs for all major genes were readily identified through antiSMASH software, allowing for the use of the same design principles as used for the kyamicin cassette system. The lantipeptide expected to be produced from the *S. roseovercillatus* BGC corresponded to L_14 and was putatively named rosiermycin

3.6.2 Rosiermycin Synthetic Cassette System

The biosynthetic cassette from *rosN* to *rosH*, including the intergenic region from *rosN* to *rosR1* and thus the SARP binding site, was inserted into expression vector pSET152 (Bierman *et al.*, 1992), using *EcoRI/XbaI* sites. The *rosA* core peptide region was replaced during design with a unique *EcoRV* site, giving plasmid pAMA2 (Appendix 9.5) (**Fig 3.11**). The activation cassette was designed for the expression of

rosR1 and *rosL* using the same principle as for pEVK6, with the *kyaN* RBS as an intergenic region. This fragment was then inserted into integration vector pIJ102567 (Gregory *et al.*, 2003; Hong *et al.*, 2005), using *NdeI/HindIII* sites, and puts the gene under the control of an *ermE** promoter; the resulting plasmid was named pAMA3 (Appendix 9.5) (**Fig 3.12**). Creation and insertion of fragments into the appropriate expression vectors was carried out by GenScript (Netherlands).

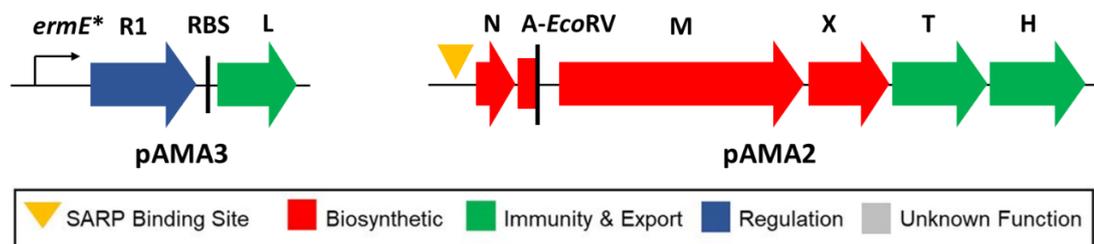


Figure 3.12: Rosiermycin Synthetic Cassettes. pAMA3; pIJ102567 activation cassette: *ermE** constitutive promoter, *rosR1* SARP, *kyaN* RBS, *rosL* methyltransferase. pAMA2; pSET152 biosynthetic cassette: SARP binding site, *rosN*, *rosA* with core peptide region removed and replaced with a unique *EcoRV* site, *rosM*, *rosX*, *rosT*, *rosH*.

3.6.3 Attempted production of Lantipeptides using the Rosiermycin Synthetic Cassette System

Synthetic DNA fragments encoding the lantipeptide core peptide sequences including ends homologous to 30 bp either side of the unique *EcoRV* site within *rosA* were ordered from Eurofins or IDT and inserted into *EcoRV* linearised pAMA2 via Gibson assembly. Constructs were confirmed by Sanger sequencing and transformed into *E. coli* ET12567/pUZ8002 cells before conjugation into *S. coelicolor* M1146/pAMA3. Strains were designated M1146/pAMA3/Lr_XX respectively. With Lr_14 being equal to the *rosiermycin* biosynthetic region without modification.

3.6.3.1 Bioassays of Hybrid Lantipeptide *lanA* Constructs Strains in the Rosiermycin Cassette System against *B. subtilis*

All strains were streaked on R5 plates and incubated for 7 days at 30°C before being overlaid with *B. subtilis* EC1524 in SNA agar and incubated further overnight. Strains M1146 and M1146/pAMA3/pAMA2 were used as negative controls and

M1146/pEVK6/pWDW63 was used as a kyamicin producing positive control, which caused a zone of inhibition against *B. subtilis* EC1524 as previously described (Vikeli *et al.*, 2020). Despite the rosiermycin synthetic cassette system following the same logic as the successful kyamicin synthetic cassette system no zones of inhibition were observed (data not shown, Appendix 9.2.3). This was also true for M1146/pAMA3/Lr_14 which comprised of the native rosiermycin *lanA* core peptide region within the rosiermycin biosynthetic cassette under control of its native promoter.

As described previously agar plugs were taken of all bioassay plates and lantipeptides extracted with methanol after freeze-thawing. As before extracts were analysed using UHPLC-HRMS on a Synapt G2-Si high resolution mass spectrometer using electrospray ionization. For each predicted lantipeptide product a m/z value was calculated based on the core peptide amino acid formula and anticipated post translational modifications corresponding to a loss of four water molecules and the addition of a hydroxyl group at the Asp15 residue. We also searched for congeners lacking the Asp15 hydroxylation. No lantipeptides were identified as being produced from any strain with the rosiermycin cassette system, including rosiermycin itself from strain M1146/pAMA3/Lr_14.

3.6.4 Identification of a Truncated *lanL* Immunity Gene within pAMA3

Considering the similarity of the kyamicin and rosiermycin BGCs and the success of the kyamicin synthetic cassette system we did not think the logic of the rosiermycin synthetic cassette design was responsible for the lack of lantipeptide production. Instead, we investigated each rosiermycin cassette and gene in turn, comparing them to their kyamicin homologs to identify any substantial differences.

In doing so, and from communication with David Widdick (Department of Molecular Microbiology, John Innes Centre, 2019), we discovered that the *rosL* immunity PE-methyl transferase gene was truncated by 46 amino acids. This occurred through designation of the start site based upon an erroneous annotation of the cinnamycin

lanL gene within the public databases. This truncation could result in a non-functioning PE-methyl transferase and thus no immunity to the produced lantipeptides. As all strains still grew it is possible that an unknown form of regulation within the host prevented the production of lantipeptides when not immune.

Previously, it had been shown that *kyaL* worked to provide immunity against duramycin within heterologous host *S. coelicolor* M1152 (**Fig 3.3**). Thus, to provide a working *lanL* immunity gene, and induce lantipeptide production, a PCR fragment containing *kyaL* was cloned from pEVK6 and inserted using *NdeI/HindIII* sites into integration vector pADW11 (Appendix 9.5). This vector was selected for its use of integrate pSAM2, which allowed for specific site integration into the chromosome (Boccard *et al*, 1989; Pernodet *et al*, 1984), allowing the third plasmid to be integrated into the heterologous host chromosome along with plasmids pAMA2 and pAMA3, which utilised the site specific integrases ϕ BT1 and ϕ C31, respectively. The *kyaL* gene was under control of a constitutive promoter ermE*, giving plasmid pAMA4 (Appendix 9.5). pAMA4 was then conjugated into M1146/pAMA3/Lr_00 (kyamicin) and M1146/pAMA3/Lr_14 (rosiermycin), to give M1146/pAMA3/Lr_00/pAMA4 and M1146/pAMA3/Lr_14/pAMA4, respectively. Bioassays were carried out against *B. subtilis* EC1524 as before in 3.5.1 but no zones of inhibition were observed, and no lantipeptides were identified from agar plug extracts of the bioassay plates using UHPLC-MS (**Fig 3.13**). M1146 and empty vector pAMA2 were used as negative controls, M1146/pEVK6/pWDW63 was used as a kyamicin positive control, producing a zone of inhibition against *B. subtilis* EC1524.

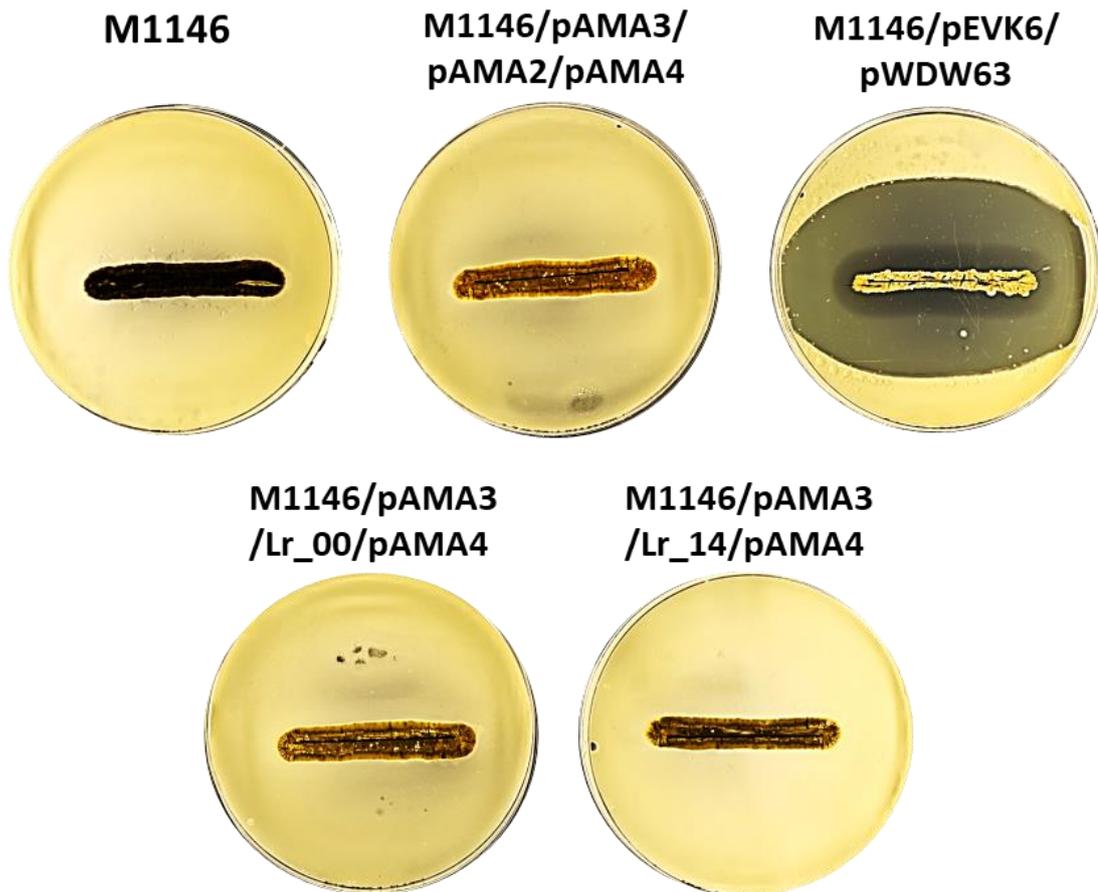


Figure 3.13: Bioassay of Strain Lr_00 (Kyamicin) and Lr_14 (Rosiermycin) with *kyaL* Immunity Cassette pAMA4. Heterologous host *S. coelicolor* M1146, overlaid with *B. subtilis* EC1524, biosynthetic cassette pAMA2, activation cassette pAMA3, performed with biological triplicates. M1146 and M1146/pAMA3/pAMA2/pAMA4 were used as negative controls. M1146/pEVK6/pWDW63 was used as a kyamicin producing positive control, with a clear zone of inhibition seen. No zones of inhibition were seen for strains Lr_00 and Lr_14.

To determine whether production of lantipeptides from the rosiermycin cassette system could be induced through incorporation of a non-truncated *rosL* immunity gene a synthetic DNA fragment was designed containing the *rosR1* and the non-truncated *rosL* genes, connected via the *kyaN* RBS. The fragment was created and inserted into integration vector pIJ102567 using *NdeI/HindIII* sites by GenScript (Netherlands); giving plasmid pAMA5 (Appendix 9.5). The activation cassette pAMA5 was then conjugated into *S. coelicolor* M1146, followed subsequently by biosynthetic cassettes Lr_00 (kyamicin) and Lr_14 (rosiermycin), to create strains M1146/pAMA5/Lr_00 and M1146/pAMA5/Lr_14, respectively. Bioassays were carried out against *B. subtilis* EC1524 as before in 3.5.1, but no zones of inhibition were observed and no lantipeptides were identified from agar plug extracts of the

bioassay plates using UHPLC-MS (**Fig3.14**). M1146 and empty vector pAMA2 were used as negative controls, M1146/pEVK6/pWDW63 was used as a kyamycin positive control, producing a zone of inhibition against *B. subtilis* EC1524.

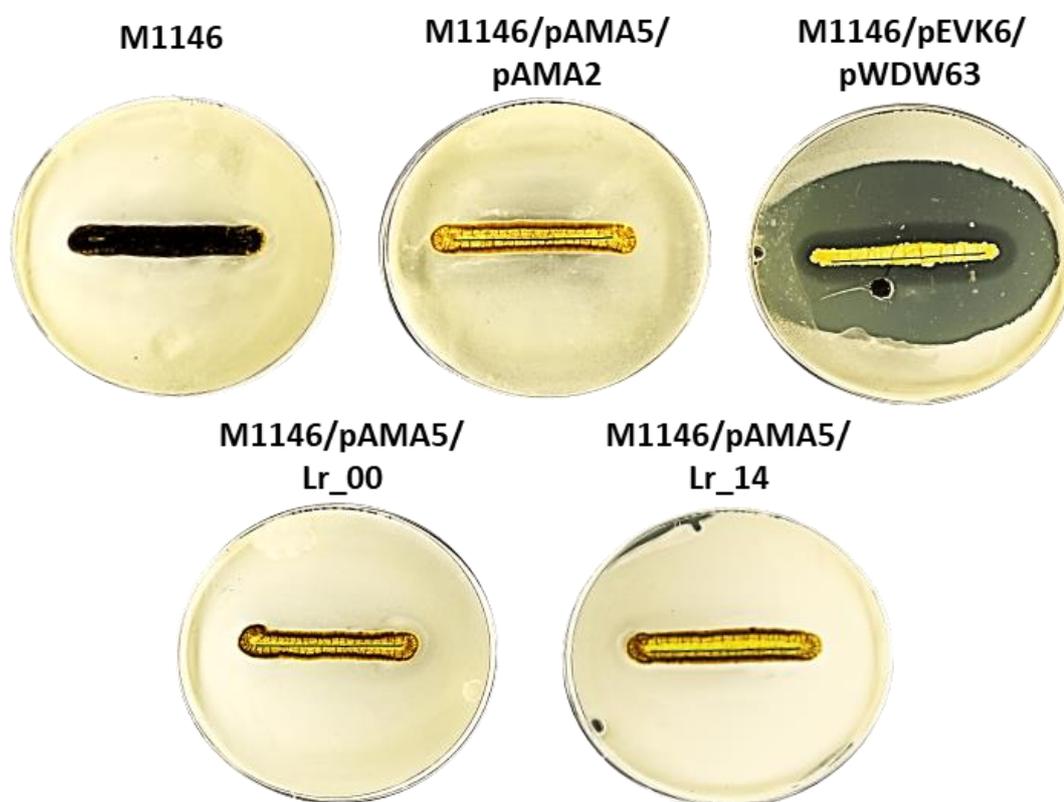


Figure 3.14: Bioassay of Strain Lr_00 (Kyamycin) and Lr_14 (Rosiermycin) with Activation Cassette pAMA5. Heterologous host *S. coelicolor* M1146, overlaid with *B. subtilis* EC1524, biosynthetic cassette pAMA2, performed with biological triplicates. M1146 and M1146/pAMA5/pAMA2 were used as negative controls. M1146/pEVK6/pWDW63 was used a kyamycin producing positive control, with a clear zone of inhibition seen. No zones of inhibition were seen for strains Lr_00 and Lr_14.

3.7 Production of Rosiermycin from *Streptomyces*

roseoverticillatus

As attempts to induce expression of lantipeptides using the rosiermycin synthetic cassette system had been met with failure we were interested to see if we could induce production of rosiermycin from the parental strain. *S. roseoverticillatus* was ordered from DSMZ (Leibniz Institute, German collection of Microorganisms and Cell Culture) as a dried mycelium stock, re-hydrated and grown on GYM media before being transferred to SF+M media. The strain grew quickly and sporulated well at 30°C with a bright pink pigmentation under these conditions.

3.7.1 Rosiermycin Expression and Synthetic Activation Cassettes

Preliminary experiments revealed that *S. roseoverticillatus* produced a compound capable of inhibiting growth of *B. subtilis* EC1524 under laboratory conditions. The production of a molecule consistent with the characteristics expected for the product of the cryptic lantipeptide BGC was confirmed by LCMS analysis and this lantipeptide was putatively called rosiermycin. To investigate the effect overexpression of the SARP encoding gene from the *ros* BGC promoters had on rosiermycin production, the activation cassettes pAMA1 and pAMA5 were conjugated into *S. roseoverticillatus*. This resulted in a phylogenetic change, with a loss of pigmentation and reduced growth in comparison to the wild type. Conjugation with activation cassette pEVK6 was also attempted to no success and an empty vector negative control was created using pIJ102567.

Strains were plated on R5 media and incubated for 7 days before being overlaid with *B. subtilis* EC1524 in SNA media and incubated further overnight.

Unexpectedly, wild type *S. roseoverticillatus* had a larger zone of inhibition compared to strains with integrated activator cassettes pAMA1 and pAMA5 (**Fig 3.15**). *S. coelicolor* M1146 was used as a negative control and M1146/pEVK6/pWDW63 used as a kyamicin producing positive control, producing a zone of inhibition against *B. subtilis* EC1524.

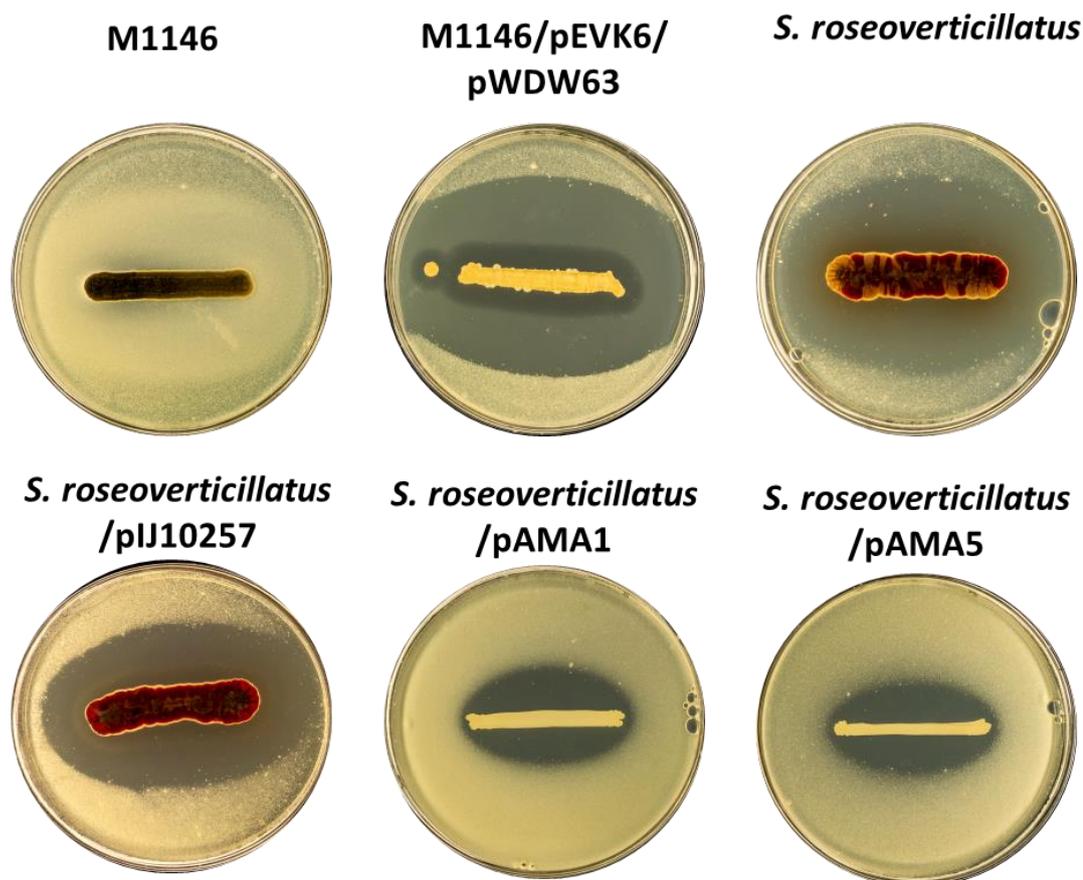


Figure 3.15: Bioassay plates of *S. roseovorticillatus* with Synthetic Cassettes pAMA1 and pAMA5. Overlay with *B. subtilis* EC1524, performed in triplicate. Rosiermycin production shown by size of zone of inhibition. *S. roseovorticillatus*/pAMA1 and *S. roseovorticillatus*/pAMA5 have reduced rosiermycin production in comparison to wild-type strain *S. roseovorticillatus*. *S. coelicolor* M1146 used as a negative control. M1146/pEVK6/pWDW63 used as a positive kyamycin producing positive control, with clear zone of inhibition seen. *S. roseovorticillatus* wild-type strain also used as a positive control for comparison. pIJ10257 used as an empty vector control.

Agar plug extracts were acquired as before in 3.5.1, and analysed by UHPLC-MS using the calculated masses for hydroxylated and deoxy $[M+2H]^{2+}$ m/z ions (933.911 and 925.911, respectively) and $[M+3H]^{3+}$ m/z ions (622.940 and 617.607, respectively), as before in 3.5.2. Mass data for all peaks observed as corresponding to lantipeptides was within ± 5 ppm of the calculated values, and the area under the peak was averaged for the triplicates (**Fig 3.16**). Wild type *S. roseovorticillatus* produced the most rosiermycin, producing almost four times more deoxy compound as compared to the activator cassette mutants. This discrepancy in production could partially be caused by the reduced growth of the mutants.

However, most notably, production of hydroxylated rosierymycin was absent in the activator cassette mutants.

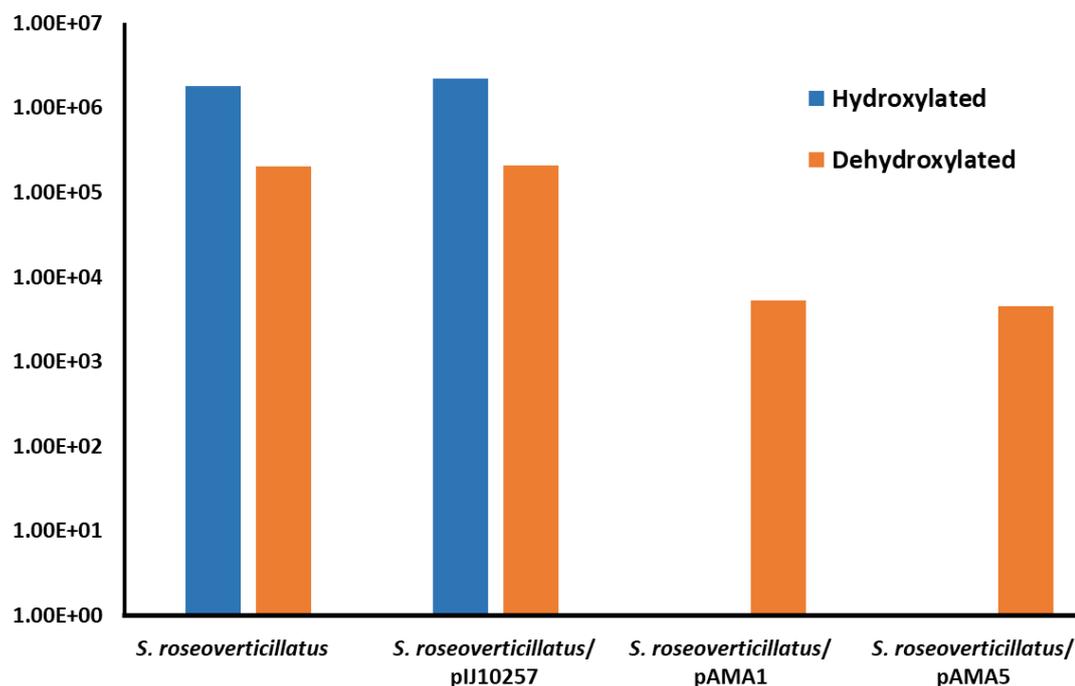


Figure 3.16: Production of Rosierymycin from *S. roseoverticillatus* Wild-Type in Comparison to Activator Cassette pAMA1 and pAMA5 Mutants. Agar plug Bioassays were performed in triplicate and plate extracts analysed using UHPLC-MS. The area under the peak was averaged. Inclusion of activator cassettes pAMA1 and pAMA5 to *S. roseoverticillatus* reduced deoxy rosierymycin production and abolished hydroxylated rosierymycin production. Calculated hydroxylated m/z (Blue); $[M+2H]^{2+}$ 933.911, $[M+3H]^{3+}$ 622.940. Calculated deoxy m/z (Orange); $[M+2H]^{2+}$ 925.911, $[M+3H]^{3+}$ 617.607. Error bars = 1 standard deviation.

3.8 Purification of Lantipeptides from the Kyamicin Expression Platform

3.8.1 Optimization of Lantipeptide Production Media

Before large scale purification of novel lantipeptides from the kyamicin expression platform began, growth conditions were optimised through the assessment of several different liquid and solid production medias (**Fig 3.17**). Strain M1146/pEVK6/L_19 was selected to test these conditions as it produced a novel lantipeptide in large amounts, as seen through previous bioassays and LCMS analysis in 3.5.

All extracts were generated using methanol and shaking. For liquid fermentation samples the cell pellet and supernatant were first separated by centrifugation and extracts generated from each separately at day 5 and day 7 of incubation at 30°C. Solid media extracts were generated from agar plugs taken at day 7 of incubation at 30 °C. All extracts were analysed using UHPLC-MS on a Synapt G2-Si high resolution mass spectrometer and the lantipeptide product identified through mass fragment search (m/z hydroxylated = 980.389, m/z deoxy = 972.389), and the area under the peak quantified (**Fig 3.17**). Liquid SF+M media was identified as the best liquid media producer of hydroxylated L_19 with solid R5 media also producing large amounts of hydroxylated L_19 and liquid R5 media being the best producer of deoxy L_19. When taking the supernatant and pellet fractions as one and considering the ease of producing and extracting compounds from liquid cultures in comparison to solid plate extraction, liquid SF+M media was selected for further production and purification of the novel lantipeptides.

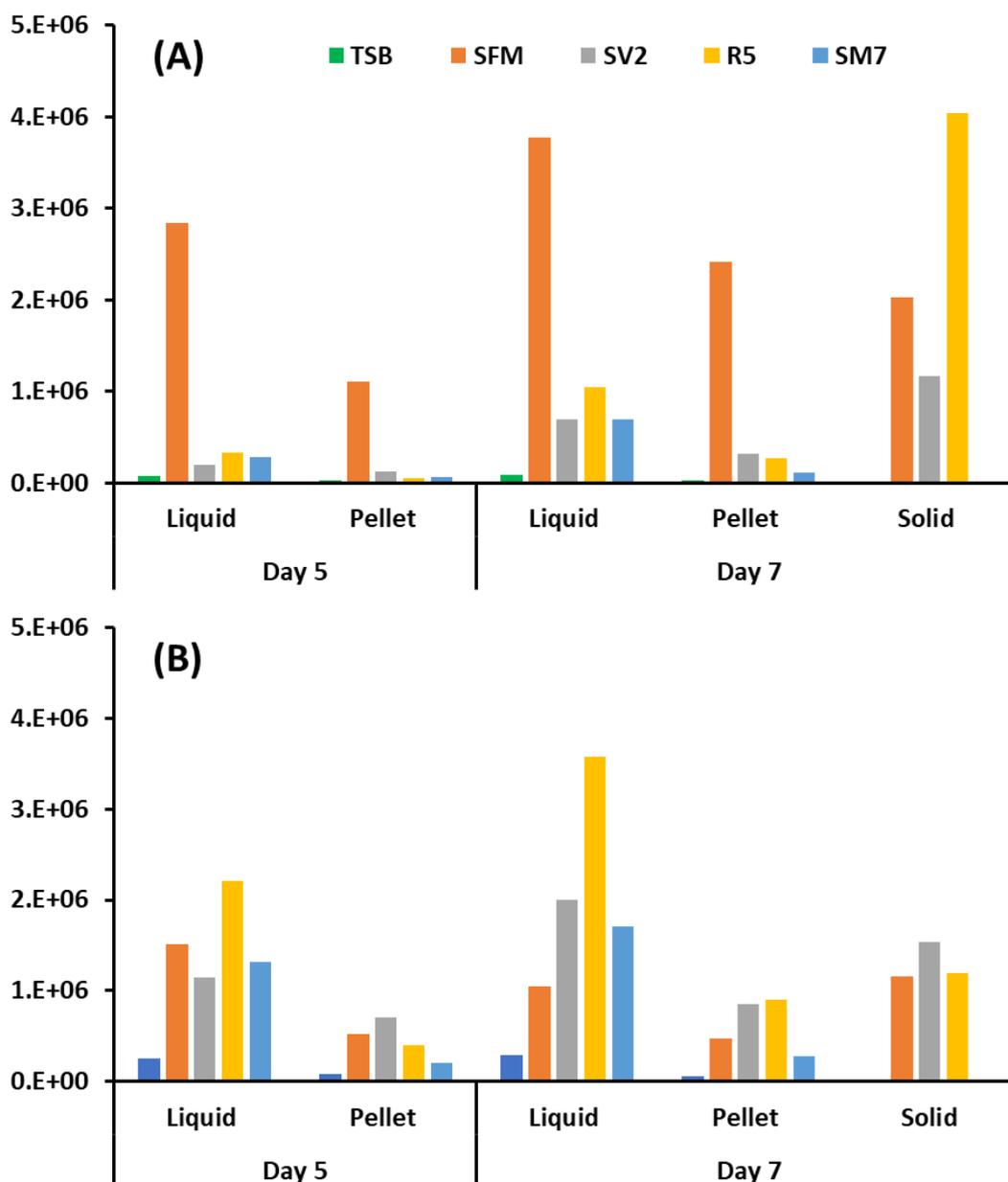


Figure 3.17: Lantipeptide Production Media Optimization. Strain grown, M1146/pEVK6/L₁₉. Liquid media grown at 30 °C, 250 rpm, for 5 or 7 days, supernatant and pellet separated by centrifugation, extraction was achieved with 50 % MeOH. Solid media grown at 30 °C for 7 days, agar plug extraction was achieved with 50 % MeOH following freeze-thawing. Extracts analysed using UHPLC-MS, performed in duplicate and the average area under the peak given. (A) Hydroxylated L₁₉; *m/z* = 980.389. (B) Deoxy L₁₉; *m/z* = 972.389.

3.8.2 Optimization of Lantipeptide Extraction from Liquid Media

Before large scale production of lantipeptides began the extraction method using methanol was assessed for its optimality by comparing several different extraction solvents and conditions (for the extraction of M1146/pEVK6/L_19 from liquid SF+M media) (**Fig.3.18**).

Strain M1146/pEVK6/L_19 (10 mL cultures) was grown for 7 days at 30 °C before 1 mL aliquots were taken and the pellet and supernatant separated by centrifugation before compounds were extracted by either methanol or ethyl acetate in a 1:1 ratio. The samples were vortexed and sonicated in an ultrasonic bath before being shaken and then centrifuged. The ethyl acetate organic layer was separated, the solvent removed by evaporation and the residue re-suspended in 1 mL 50 % methanol. Additionally, HP-20 resin was trialed as a potential way to extract lantipeptides from solution. Firstly, 0.5 g of HP-20 resin was added to the supernatant of a M1146/pEVK6/L_19 10 mL culture, which was then incubated with shaking for a further 3 h. The resin was then separated out by filtering with filtration paper and this was considered as the pellet fraction of the sample. The supernatant was extracted as before, with a 1 mL aliquot extracted with methanol in a 1:1 ratio. To separate any compounds bound to the HP-20 resin 10 mL of methanol was added and the sample was homogenised through vortexing before being sonicated in an ultrasonic bath and shaken for a further hour. The resin was then separated out through a filter and 0.5 mL of the supernatant mixed with distilled water to get a 1 mL 50 % MeOH extract.

All extracts were analysed using UHPLC-HRMS using a Synapt G2-Si high resolution mass spectrometer and the L_19 lantipeptide identified through mass fragment search (m/z hydroxylated = 980.389, m/z deoxy = 972.389), and the area under the peak quantified. This was treated as an equivalent representation of the true amount of lantipeptide present within the sample (**Fig 3.18**).

As shown below neither ethyl acetate nor HP-20 resin worked to extract lantipeptide L₁₉. Whereas the lantipeptide was present in large quantities in both the supernatant and pellet fraction of the methanol extract. Thus, going forward methanol was still considered the optimum extraction solvent.

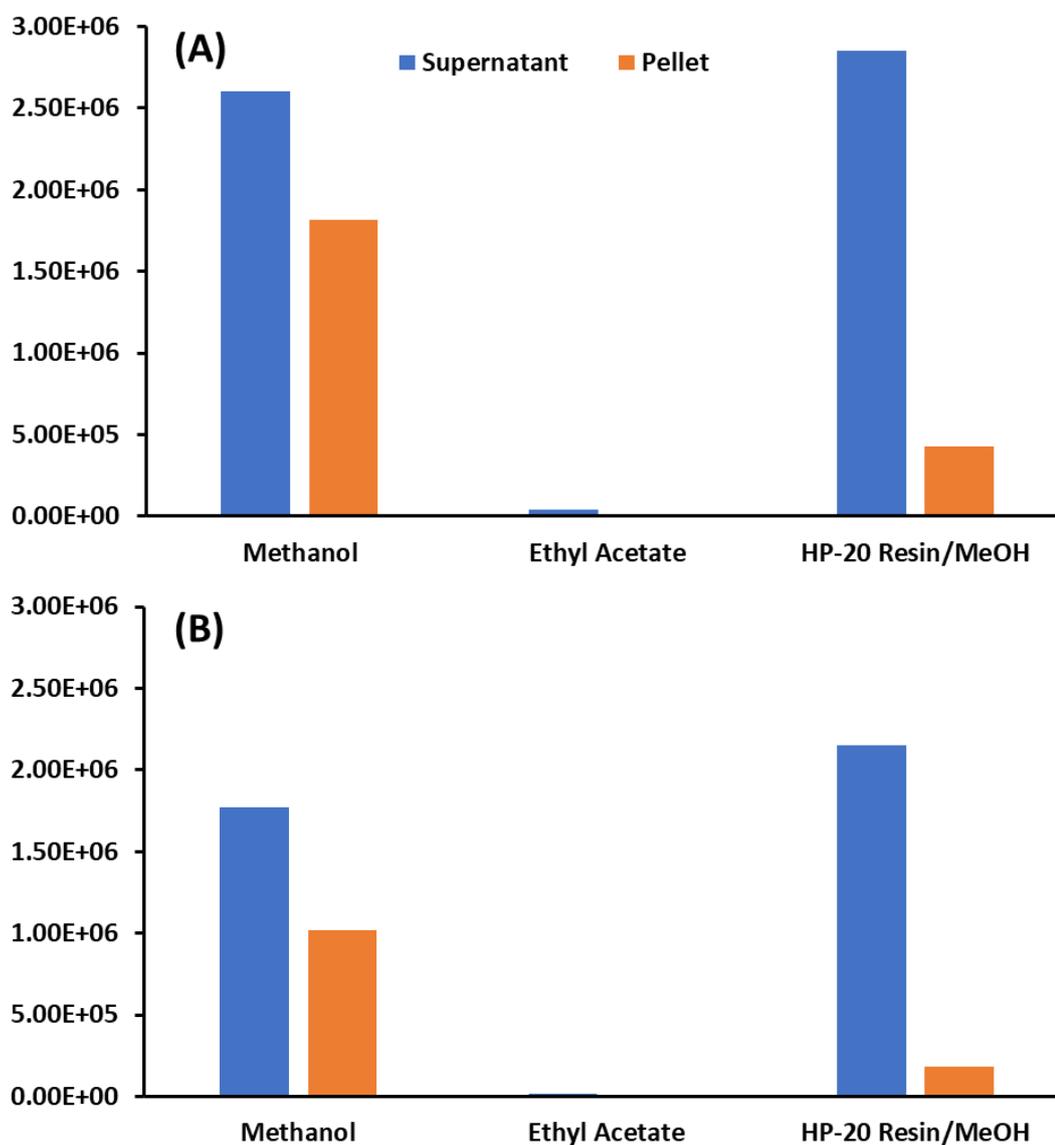


Figure 3.18: Lantipeptide Extraction Solvent Optimization. Culture; M1146/pEVK6/L₁₉, 10 mL liquid SF+M, extracts taken after 7 days, performed in duplicate, analysed on UHPLC-MS and the average area under the peak given. Solvents; Methanol, Ethyl Acetate, HP-20 resin supernatant incubation followed by MeOH extraction. (A) Hydroxylated L₁₉; $m/z = 980.389$, (B) Deoxy L₁₉; $m/z = 972.389$.

3.8.2 Optimization of Lantipeptide HPLC Purification

Following the results obtained above in Sections 3.8.1 and 3.8.2 all lantipeptides producing strains were grown in SF+M liquid media at 30 °C for 7 days and compounds extracted from the supernatant using methanol. Purification of lantipeptides was achieved through extensive use of HPLC. To obtain enough compound for further assays 3 to 9 litres of lantipeptide producing culture were grown in 500 mL volumes and the supernatant separated and filtered. Initial extraction was carried out by loading the supernatant of the sample onto a 120 g SNAP-Ultra Biotage C18 column and running a 5-100% water/methanol gradient, with the lantipeptides typically eluting at 55-75 % methanol as observed by LCMS (Materials & Methods 2.6.1.2). Fractions containing lantipeptide were then pooled and purified further using preparative HPLC, again with a C18 column and a 5-100 % water/methanol gradient and the resulting fractions analysed by LCMS (Materials & Methods 2.7.2.2). At this stage it was observed that lantipeptides of the same m/z produced using the kyamycin cassette platform eluted in three distinct chromatographic peaks, as shown for lantipeptide L_19 purified from a culture of M1146/pEVK6/L_19 (**Fig 3.19**). LCMS showed that each of the three chromatographic fractions contained an m/z matching that of the predicted lantipeptide to within ± 5 ppm. This pattern of purification was not observed for rosierymycin produced from the parental strain *S. roseoverticillatus* in liquid SFM, which eluted in a singular chromatographic peak and was produced almost exclusively as a hydroxylated compound (data not shown).

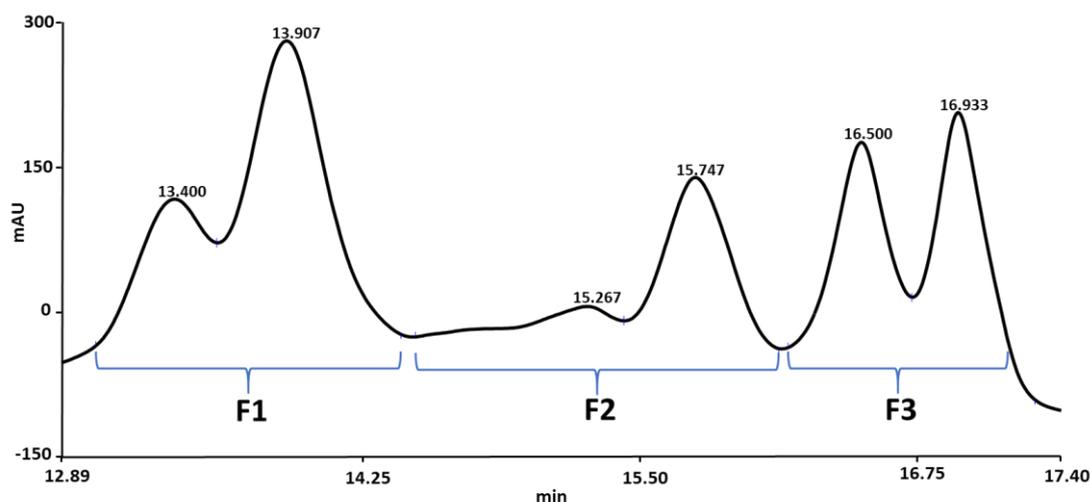


Figure 3.19: Chromatogram of L_19 preparative HPLC. Gradient; 5-100% MeOH. Column; reverse phase C18. Peaks were collected as three fractions (F1-3). UHPLC-MS identified L_19 Hydroxy and L_19 Deoxy within each fraction ($m/z \leq \pm 5$ ppm), with L_19 Hydroxy being the latter peak of each fraction.

As purification of the lantipeptides proceeded the different chromatographic fractions were kept separate. Full purification of the lantipeptide fractions was achieved through a further preparative HPLC step using an 5-100 % acetonitrile gradient, followed by a semi-prep HPLC step using a C18 column with a 5-100 % methanol gradient and then a last semi-prep HPLC step again using a C18 column with a 5-100 % acetonitrile gradient. Purity of the samples was assessed at each stage using analytical HPLC with evaporative light scattering detection (ELSD) and mass spectrometry. The lantipeptides were considered pure when they comprised ≥ 90 % of the sample, as assessed through ELSD. Purification of lantipeptide L_19 hydroxy and deoxy was achieved across all chromatographic fractions, while purification of lantipeptide L_18 hydroxy and deoxy, lantipeptide L_22 and lantipeptide L_26 deoxy was achieved from singular chromatographic fractions. The lantipeptide compound, the preparative-HPLC chromatographic fraction it was purified from (1-3) and the amount purified lantipeptide are shown in **Table 3.4**. Rosiermycin, which only eluted in a single chromatographic fraction as previously stated, was also purified (**Table 3.4**).

Table 3.4: Lantipeptides Purified during the course of this work, with the preparative HPLC chromatographic fraction (1-3) they were purified from and sample weight (mg)

Lantipeptide	HPLC Fraction	Weight (mg)
L_19 Deoxy	1	13.40
L_19 Deoxy	2	1.5
L_19 Deoxy	3	6.7
L_19 Hydroxy	3	3.5

Lantipeptide	HPLC Fraction	Weight (mg)
L_18 Deoxy	1	1.4
L_18 Hydroxy	3	5.6

Lantipeptide	HPLC Fraction	Weight (mg)
L_22 Deoxy	3	0.9

Lantipeptide	HPLC Fraction	Weight (mg)
L_26 Deoxy	2	5.7

Lantipeptide	HPLC Fraction	Weight (mg)
Rosiermycin Hydroxy	1	2.6

3.9 Characterization of Purified Lantipeptides

3.9.1 Bioactivity of Lantipeptides

3.9.1.1 Difference in Bioactivity between Purified Chromatographic Fractions of Lantipeptide L_19 Deoxy

To assess whether the preparative-HPLC chromatographic fraction the lantipeptides were purified from affected their antimicrobial activity lantipeptide L_19 Deoxy was purified from all three preparative-HPLC fractions (F1-3), which were previously observed to contain the same m/z by LCMS (Section 3.8.2, **Fig 3.19**). Previous experiments (Section 3.10.1.2, **Fig 3.21**, see below) had shown that lantipeptide L_19 Deoxy F3 had antimicrobial activity against *B. subtilis* EC1524 at 1 mg/mL. Thus, lantipeptide L_19 Deoxy from all three chromatographic fractions was spotted onto R5 plates impregnated with *B. subtilis* EC1524 at 1 mg/mL and left to grow overnight (**Fig 3.20**). Apramycin (5 $\mu\text{g/mL}$) and DMSO were used as positive and negative controls, respectively.

Only lantipeptide L_19 Deoxy F3 had antibacterial activity against *B. subtilis* EC1524. Thus, in this case, the early eluting L_19 lantipeptide compounds do not have the same antibacterial activity as the latter eluting L_19 lantipeptide compounds. Therefore, the preparative-HPLC chromatographic fraction that a novel lantipeptide compound was purified from could affect the antimicrobial activity observed. As all lantipeptides produced from the kyamycin cassette platform had a similar preparative-HPLC traces to lantipeptide L_19, with three distinct chromatographic peaks (F1-3) each containing m/z equivalent lantipeptide compound (in both hydroxy and deoxy form if both were produced), we expected that the later purified fractions (F3) of the lantipeptides to also show antibacterial activity above that of the earlier purified fractions (F1, F2), similarly to that seen for lantipeptide L_19 Deoxy fractions.

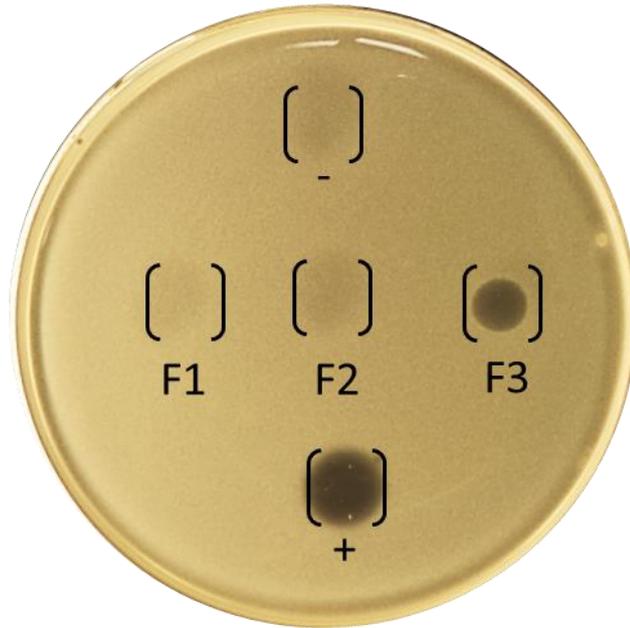


Figure 3.20: Antibacterial Comparison of lantipeptide L₁₉ Deoxy from Chromatographic Fractions 1-3. Plate impregnated with *B. subtilis* EC1524. (-) DMSO negative control, (+) Apramycin 5 µg/mL positive control, (F1) L₁₉ Deoxy from preparative HPLC chromatographic fraction 1, (F2) L₁₉ Deoxy from preparative HPLC chromatographic fraction 2, (F3) L₁₉ Deoxy from preparative HPLC chromatographic fraction 3.

3.9.1.2 Minimum Inhibitory Concentration (MIC) of Purified Lantipeptides

All lantipeptides isolated during this work were assessed for their antibacterial activity and MICs calculated. Several indicator strains were used, including clinically relevant strains Methicillin Resistant *Staphylococcus aureus* (MRSA) (Qin *et al.*, 2017) and Methicillin-Susceptible *Staphylococcus aureus* (MSSA); the soil dwelling actinomycete *Streptomyces venezuelae*; *E. coli* ATCC 25922 a gram-negative strain; and *E. coli* NR698, a gram-negative strain with a deficient outer membrane which mimics a 'gram-positive cell envelope' state (Krüger *et al.*, 2019; Ruiz *et al.*, 2005), were also tested. The lantipeptide duramycin, shown to have bioactivity against *B. subtilis* EC1524 (Huo *et al.*, 2017; Vikeli *et al.*, 2020), was used as a direct comparator and apramycin (5 µg/mL) and DMSO were used as positive and negative controls, respectively (**Fig 3.21**).

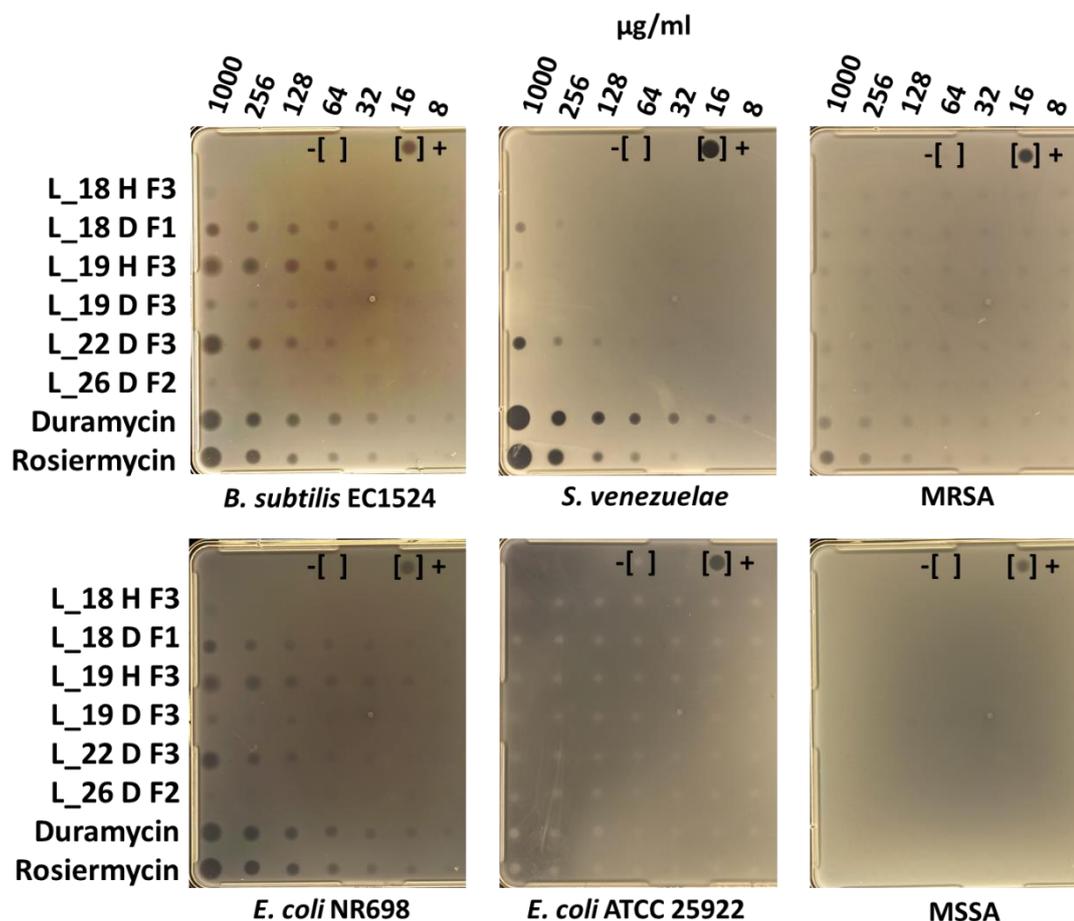


Figure 3.21: Comparative Bioassay of Novel Lantipeptides. Lantipeptides were spotted (5 µL) from 1000 to 8 µg/mL, plated in triplicate. (-) DMSO negative control (5 µL), (+) Apramycin 5 µg/mL positive control (5 µL). Lantipeptides; L_18 Hydroxy F3, L_18 Deoxy F1, L_19 Hydroxy F3, L_19 Deoxy F3, L_22 Deoxy F3, L_26 Deoxy F3, Duramycin, Rosiermycin. MRSA; methicillin resistance *Staphylococcus aureus*, MSSA; methicillin susceptible *Staphylococcus aureus*.

The hydroxylation of the Asp15 residue has been reported to be important for the binding to target membrane lipid PE and thus the antimicrobial activity of lantipeptides, with the deoxy version of lantipeptides reported as being inactive or having greatly reduce activity in comparison to the hydroxylated version (Lopatniuk *et al*, 2017; Vestergaard *et al.*, 2019; Vikeli *et al.*, 2020). Thus, we would expect active deoxy compounds to have reduced activity compared to *m/z* equivalent active hydroxylated compounds.

For indicator strains against which antimicrobial activity was seen (*B. subtilis* EC1524, *E. coli* NR698 and *S. venezuelae*) lantipeptide L_19 Hydroxy F3 showed

greater activity than lantipeptide L_19 Deoxy F3, as was expected from studies previously reported (Lopatniuk *et al.*, 2017; Vestergaard *et al.*, 2019; Vikeli *et al.*, 2020). L_26 Deoxy F2 showed no antibacterial activity, including against *B. subtilis* EC1524, against which a zone of inhibition was previously observed for the producing strain M1146/pEVK6/L_26 (**Fig 3.9**). As no L_26 Hydroxy was seen to be produced during purification of the compound it is likely that the F2 fraction of L_26 Deoxy does not contain the active form of the lantipeptide compound. L_22 Deoxy F3 showed antibacterial activity against strains *B. subtilis* EC1524, *E. coli* NR698 and *S. venezuelae* but as no hydroxylated form of the compound could be purified a comparison could not be made for the effect of Asp15 hydroxylation on activity. L_18 Hydroxy F3, which we would expect to be the most active form of the compound, showed no activity against any of the indicator strains. In comparison, we would have expected to see no antibacterial activity from L_18 Deoxy F1, considering purification was achieved from the fastest eluting chromatographic peak (F1) and the compound is not hydroxylated. However, lantipeptide L_18 Deoxy F1 did show activity against *B. subtilis* EC1524, *E. coli* NR698 and *S. venezuelae*, potentially indicating that the order of elution of mass equivalent intermediates and/or epimers of the lantipeptide products is variable, and not all active compounds elute in the later chromatographic fractions.

No lantipeptides had bioactivity against the *E. coli* ATCC 25922 despite having bioactivity against *E. coli* NR698, suggesting that lantipeptides are incapable of passing through the outer membrane of Gram-negative bacteria to bind to the PE of the inner membrane. L_18 Deoxy F1 L_22 Deoxy F3, duramycin and rosiermycin had bioactivity against *S. venezuelae*, with MICs of 1000, 256, 16 and 64 µg/mL, respectively. Only duramycin and rosiermycin showed bioactivity against MSSA, with MICs of 1000 and 256 µg/mL, respectively, and no lantipeptides had bioactivity against MSSA. Rosiermycin had the greatest bioactivity of the novel lantipeptides, with bioactivity against all strains except *E. coli* ATCC 24922 and MSSA. However, excepting MRSA, duramycin had lower MICs than rosiermycin across all other indicator strains (**Table 3.5**).

Table 3.5: Minimum Inhibitory Concentrations of novel Lantipeptides ($\mu\text{g}/\text{mL}$).

MIC was calculated from agar plate assays, performed in triplicate. >1000 was used when no zone of inhibition was seen.

Sample	<i>B. subtilis</i> EC1524	<i>E. coli</i> NR698	<i>E. coli</i> ATCC 25922	<i>S. venezuelae</i>	MRSA	MSSA
L_18 H F3	>1000	>1000	>1000	>1000	>1000	>1000
L_18 D F1	1000	256	>1000	1000	>1000	>1000
L_19 H F3	64	128	>1000	>1000	>1000	>1000
L_19 D F3	1000	>1000	>1000	>1000	>1000	>1000
L_22 D F3	128	256	>1000	256	>1000	>1000
L_26 D F2	>1000	>1000	>1000	>1000	>1000	>1000
Duramycin	8	32	>1000	16	1000	>1000
Rosiermycin	64	64	>1000	64	256	>1000

3.9.2 Chemical Characterization of Lantipeptides

3.9.2.1 Structural Validation of Novel Lantipeptides by Tandem Mass Spectrometry (MS)

To help in validating the predicted structures of the new lantipeptides isolated here, tandem mass spectrometry (MS) was performed to determine the peptide sequence of each compound. However, due to the highly crosslinked nature of these compounds, they do not readily fragment within the mass spectrometer. To address this, we utilised a published procedure for chemical reduction and desulphurisation of the lanthionine linkages by treatment with nickel chloride (NiCl_2) and sodium borohydride (NaBH_4) (see Chapter 2, section 2.7.3). This method was previously described during analysis of the lantibiotic cinnamycin B (Kodani *et al.*, 2016) and subsequently by the Wilkinson group during analysis of kyamicin (Vikeli *et al.*, 2020). During the chemical reduction of lantipeptide L_19 Hydroxy F3, UHPLC-MS was used to track the degree to which reduction of the lantipeptide had occurred; this showed that although complete reduction of the compound was not achieved within the 4 hour time frame of the assay, a fully reduced (zero sulphur) species was produced (**Fig 3.22**). Subsequently MS/MS fragmentation was then carried out on the partly reduced lantipeptide mixture using Matrix-Assisted Laser Desorption Ionization (MALDI) LIFT methods, and fragmentation of the fully reduced zero sulphur parental ion determined.

After reduction of lantipeptide L_19 Hydroxy_F3, a $[\text{M}+2\text{H}]^{2+}$ ion at 935.4569 m/z was identified, matching the predicted m/z of 935.4545 ($\Delta = 2.6$ ppm). This corresponded to the expected loss of three sulphur atoms and the gain of six hydrogen atoms, to give the molecular formula $\text{C}_{86}\text{H}_{124}\text{N}_{20}\text{O}_{27}$ (**Fig 3.22**).

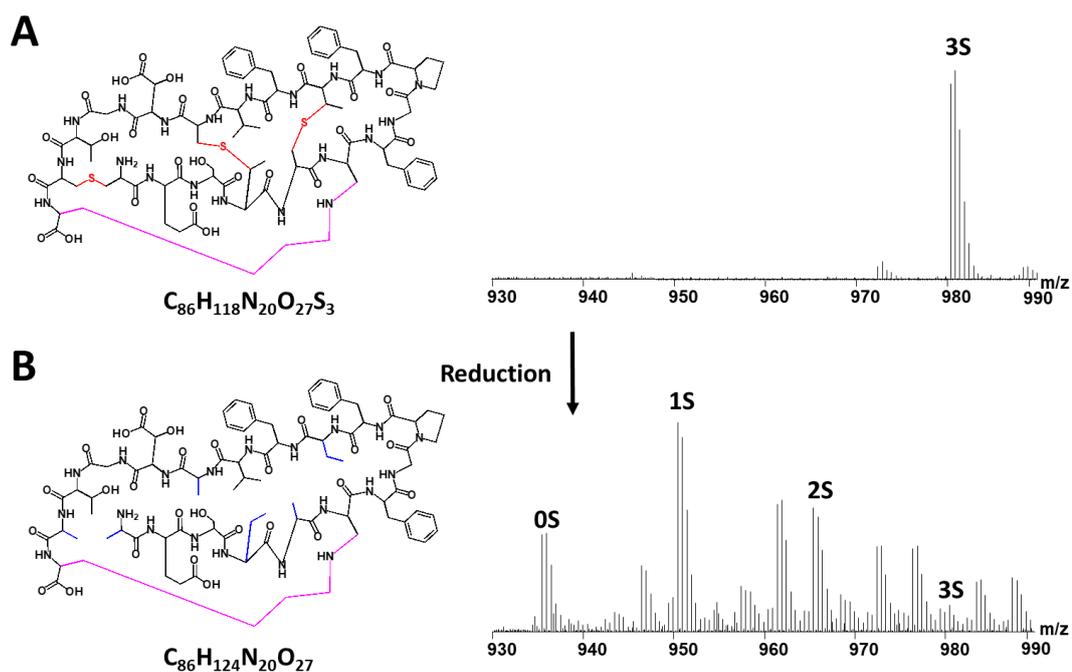


Figure 3.22: Schematic Representation of L_19 Hydroxy F3 Reduction and Corresponding LCMS Peaks. Mass spectra obtained using UHPLC-MS (A) Chemical structure of L_19 Hydroxy F3 corresponding to molecular formula $C_{86}H_{118}N_{20}O_{27}S_3$ (m/z calculated = 980.3891, m/z observed = 980.3886, Δ = -0.5 ppm). (B) Chemical structure of L_19 Hydroxy F3 after reduction of the three thioether bonds corresponding to the loss of three sulphur atoms and gain of six hydrogen atoms. Mass spectra observed for each reduction; 2S corresponding to molecular formula $C_{86}H_{120}N_{20}O_{27}S_2$ (m/z calculated = 965.4109, m/z observed = 965.4109, Δ = 0.0 ppm), 1S corresponding to molecular formula $C_{86}H_{122}N_{20}O_{27}S_1$ (m/z calculated = 950.4327, m/z observed = 950.4291, Δ = -3.7878), 0S corresponding to molecular formula $C_{86}H_{124}N_{20}O_{27}$ (m/z calculated = 935.4545, m/z observed = 935.4569, Δ = 2.7 ppm). Red; intact thioether bonds, blue; reduce thioether bonds, pink; lysinoalanine bridge.

While the lantipeptide L_19 Hydroxy F3 sample could not be fully reduced fragmentation was still carried out on the identified lantipeptide L_19 Hydroxy F3 (zero sulphur) ion using MALDI LIFT MS/MS fragmentation, and the complete ion series for the lantipeptide L_19 Hydroxy peptide seen (**Fig 3.23**). Fragmentation of the lysinoalanine bridge resulted in a glycine at residue six and a N=C double bond at the end of the K19 side chain, as had been shown previously (Kodani *et al.*, 2016; Vikeli *et al.*, 2020).

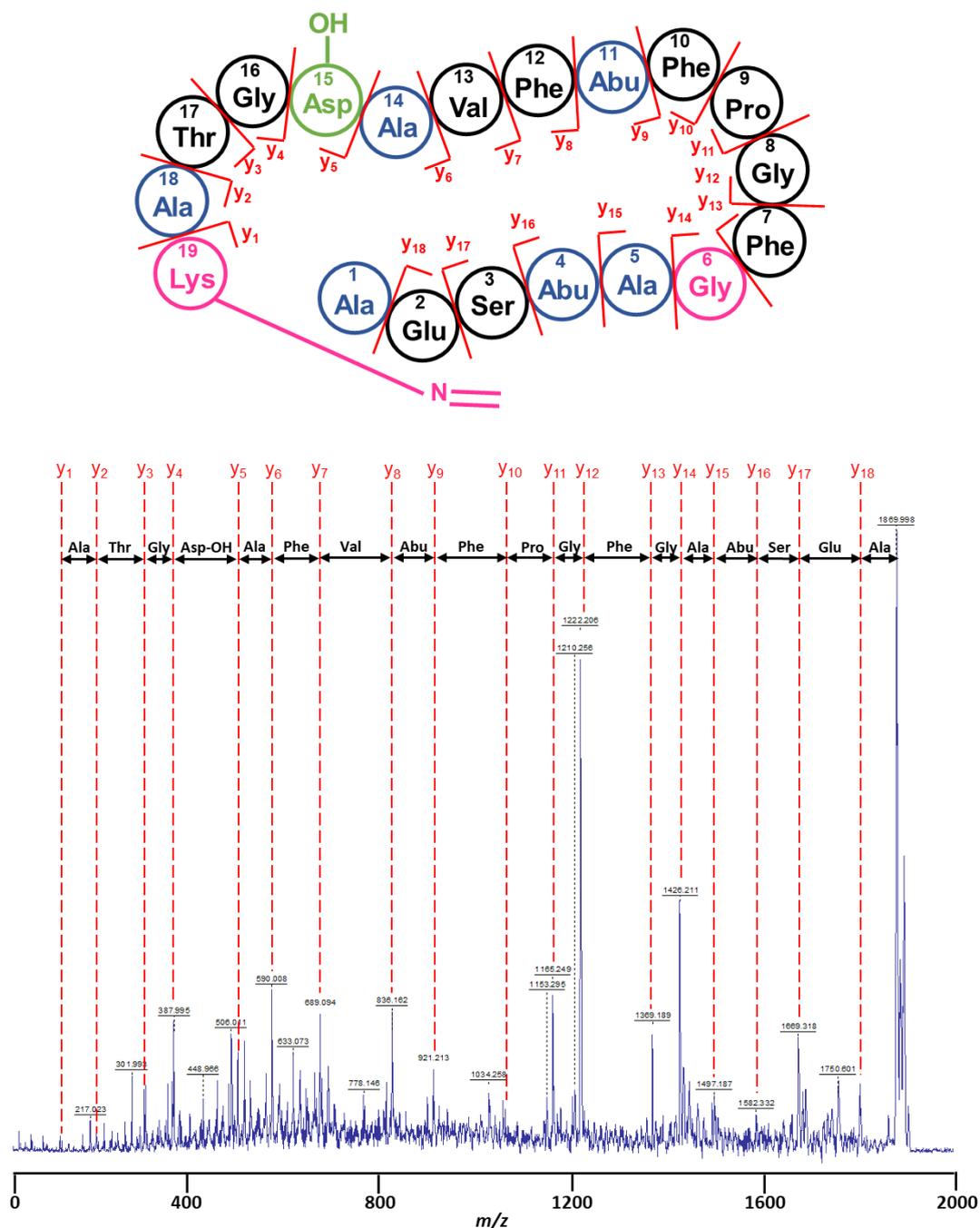


Figure 3.23: Fragmentation of Fully Reduced L₁₉ Hydroxy F3 (zero sulphur). MALDI LIFT MS/MS fragmentation showed the y ion series of the reduced peptide (zero sulphur). The expected connectivity of the complete peptide was observed, with the expected mass loss per amino acid residue fragmented. The fragmentation of the lysinoalanine bridge resulted in a glycine at residue six and a N=C double bond at the end of the K19 side chain.

Lantipeptides L_19 Deoxy F3, L_18 Hydroxy F3 and L_26 Deoxy F3 were also successfully chemically reduced and desulphurisation of the lanthionine linkages carried out as above, with the fully reduced ion (zero sulphur) m/z matching those calculated to within ± 5 ppm. Tandem MS using MALDI LIFT MS/MS fragmentation of the fully reduced compounds (zero sulphur) showed the complete ion series of the expected peptides, validating the core peptide amino acid sequence of the lantipeptides (Appendix 9.2.4), and lending confidence to the predicted structures of these lantipeptides.

Reduction and MALDI LIFT MS/MS fragmentation was attempted for purified rosiermycin and L_18 Deoxy F1 but the fully reduced ion could not be detected by UHPLC-MS at the time of this thesis being completed.

3.9.2.2 Structural Validation of Novel Lantipeptide L_19 Hydroxy F3 and L_19 Deoxy F3 by Nuclear Magnetic Resonance (NMR)

NMR analysis was carried out to gain further insight into the structure of the lantipeptide L_19. Pure lantipeptide L_19 Hydroxy F3 and L_19 Deoxy F3 were dissolved in deuterated DMSO and a ^1H NMR spectrum recorded at 600 MHz, as detailed in Chapter 2.7.4. Further to this, two-dimensional NMR experiments COESY, HSQC, TOCSY and NOESY were carried out to enable identification of proton-carbon correlations. Assignment was facilitated by comparison of chemical shifts in the ^1H and ^{13}C NMR spectra to corresponding atoms in similar lantipeptides, i.e cinnamycin (Kessler *et al.*, 1988; Wakamatsu *et al.*, 1990), cinnamycin B (Kodani *et al.*, 2016) and kyamicin (Vikeli *et al.*, 2020). All NMR spectra acquisition and assignment were carried out by Dr Sergey Nopogodiev, manager of the NMR facility at JIC, on my behalf.

Lantipeptide L_19 Hydroxy F3 was used to optimise NMR data acquisition and the assigned spectra allowed assignment of many of the expected residue resonances, including the conserved proline (Pro9) thought to play a role in the conformation of the lantipeptide and formation of the PE binding pocket (**Fig 3.24**) (Elvas *et al.*, 2017; Zhao, 2011). However, resonances for residues Lys19 and Ala6, between

which the characteristic lysinoalanine bridge of class II type B lantipeptides is formed, could not be assigned. The Lys19 ^1H and ^{13}C NMR signal resonances had been previously reported for cinnamycin, cinnamycin B, and duramycin B and C (Kessler *et al.*, 1988; Kodani *et al.*, 2016; Wakamatsu *et al.*, 1990; Zimmermann *et al.*, 1993), but was not found for other cinnamycin-like compounds, such as kyamicin (Vikeli *et al.*, 2020). Comparison of the HSQC of lantipeptide L_19 Hydroxy F3 to lantipeptide L_19 Deoxy F3 identified the Asp15 residue, distinguishing between the hydroxylated and deoxy state (**Fig 3.25**).

NMR data was also acquired by Dr Sergey Nopogodiev for purified lantipeptides L_19 Deoxy F3, L_18 Hydroxy F3, L_26 Deoxy F2 and rosiermycin, but no distinguishing residues or features of the molecules could be readily identified and the spectra were unassigned at the point of publishing this work, so the data was not included in this work. Assignment of these spectra will be carried out in the future for further investigation and structural validation of the novel lantipeptides.

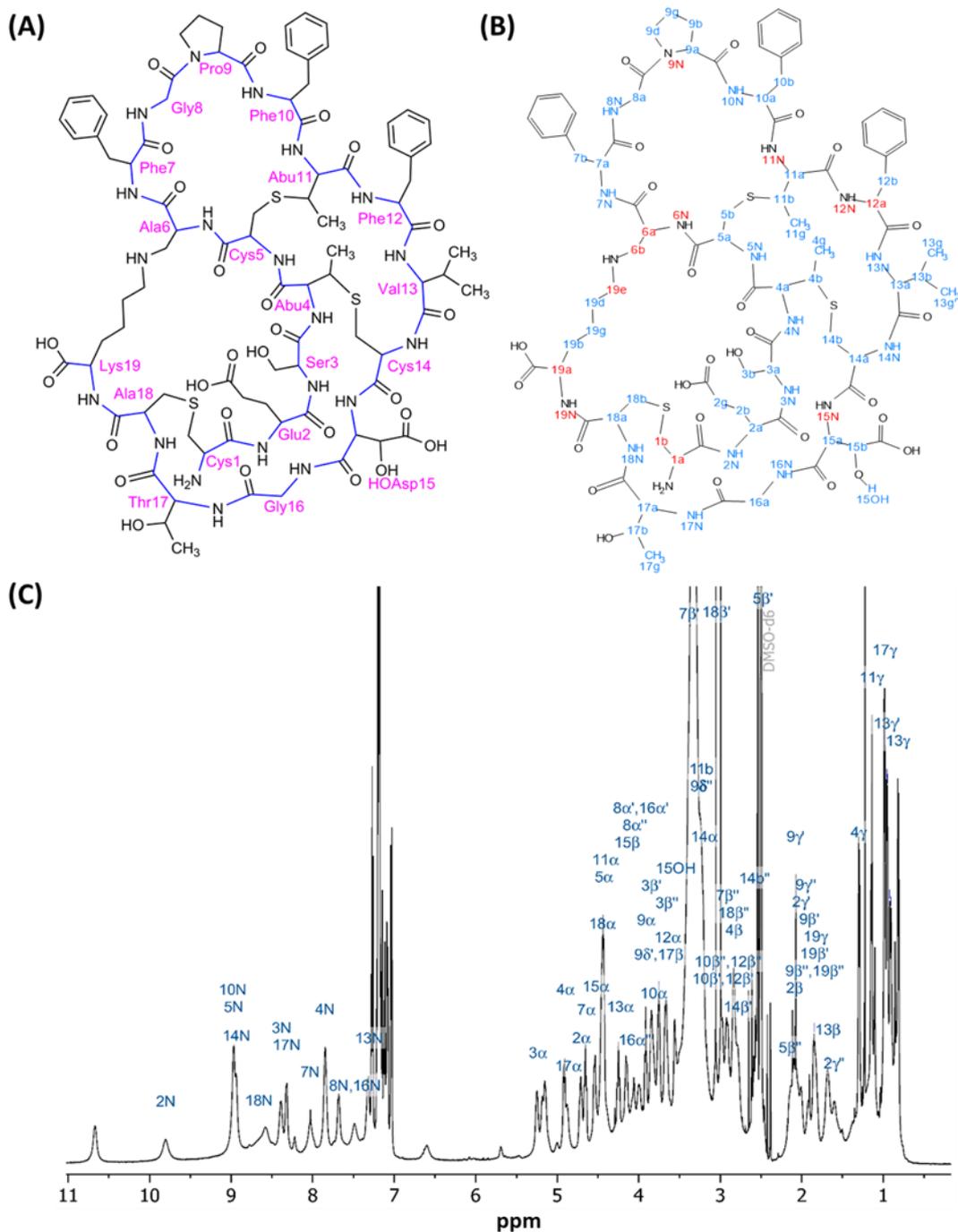


Figure 3.24: L₁₉ Hydroxy F3 Chemical Structure with Corresponding Atom Assignment and ¹H NMR Spectrum. (A) L₁₉ Hydroxy schematic representation with amino acid residues numbered (pink). (B) Corresponding structure schematic of L₁₉ Hydroxy with atom numbering used in ¹H NMR spectra. a,b,g,d,e were used instead of α, β, γ, δ, ε respectively, to assign ¹H resonance when attached to carbon atoms and ¹H resonance attached to nitrogen atoms was termed N. Atoms corresponding to assigned peaks (blue), atoms not assigned (red). (C) ¹H NMR spectrum (600 MHz, d₆-DMSO, 298K) of L₁₉ Hydroxy F3, with peaks assigned to expected ¹H resonance points within the compound. Data acquired and spectrum assignment by Sergey Nopogodiev, JIC.

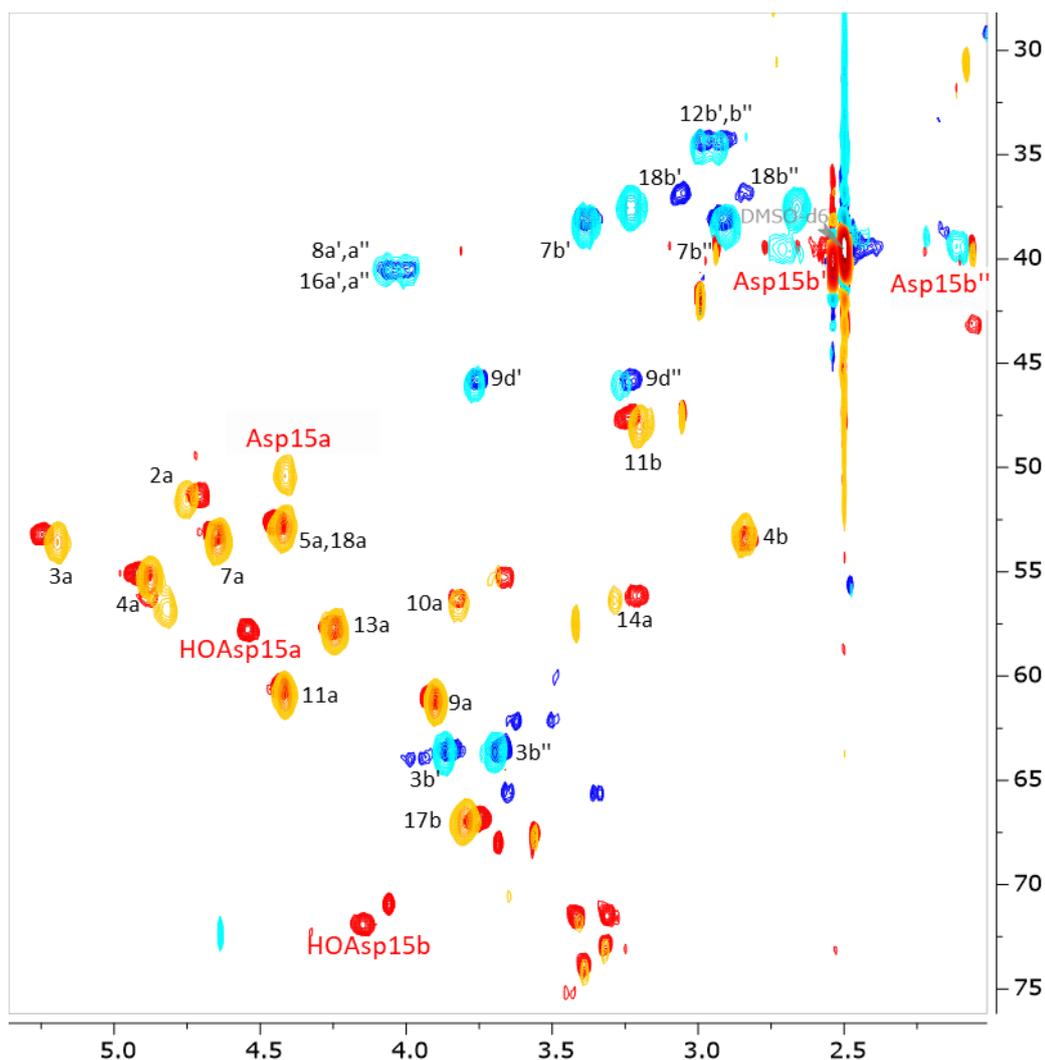


Figure 3.25: L_19 Hydroxy F3 and L_19 Deoxy F3 HSQC Overlay and Identification of Asp15. Fragment of an overlay of HSQC-EDITED 2D spectra of L_19 Hydroxy F3 (CH red, CH2 dark blue) with L_19 Deoxy F3 (CH yellow, CH2 light blue). Numbers correspond to expected atom cross-peaks within L_19 (Fig 3.24B). Comparison of 2D spectra revealed positions of hydroxy-Asp15 cross-peaks (HOAsp15a and HOAsp15b). Hydroxy Asp15 cross-peaks are not present in L_19 Deoxy F3, instead a cross-peak for deoxy-Asp15 (Asp15a, Asp15b and Asp15b'') is present, though location of cross-peak Asp15b could not be unambiguously defined. Other resonances for the amino acid residues of L_19 Hydroxy F3 and L_19 Deoxy F3 coincide or are very close, showing no distinct difference. HSQC spectra was acquired and assigned by Sergey Nopogodiev, JIC.

3.10 Discussion

Using the previously reported kyamicin heterologous expression system a version was engineered which allowed the core peptide sequence of the *lanA* gene to be rapidly exchanged with any other using synthetic DNA fragments. Genome mining identified 31 *lanA* core peptide sequences within the public databases, and a further 4 *lanA* core peptide sequences were supplied by NAICONS, Milan, Italy. Of the 35 potential lantipeptides, 16 were produced using the kyamicin synthetic cassette system as identified by LCMS, 12 of which were novel lantipeptides. Lantipeptides cinnamycin, duramycin, duramycin B, mathermycin and nine novel lantipeptides exhibited antibacterial activity against *B. subtilis* EC1524.

When optimizing the kyamicin synthetic cassette system there was a distinct difference in the zone of inhibition observed between strains M1146/pEVK6/pWDW63 and M1146/pAMA1/pWDW63, with the latter having a much larger zone in inhibition than the former. This was assumed to correlate with increased production of kyamicin from the pWDW63 biosynthetic cassette. Both activation cassettes contained the essential *kyaL* ad *kyaR1* genes, and for both activation cassettes the *kyaR1* gene is directly downstream of the *ermE** promoter. Thus, it is not expected that the amount of *KyaR1* produced from both activation cassettes to be distinctly different, but this could be measured through protein pull down and quantification.

Within the pEVK6 activation cassette *kyaL* is directly downstream of *kyaR1* and has been modified to read in the same direction, thus putting directly downstream of the strong *ermE** promoter. However, in the pAMA1 activation cassette the *kyaL* gene has not been modified and reads in the opposite direction to the *kyaR1* gene. Thus, it is not directly downstream of the *ermE** promoter and instead must lie downstream of a native streptomyces promoter within the sequence (not identified). This promoter may not produce *kyaL* to the levels produced by the *ermE** promoter in the pEVK6 activation cassette. As it is known that the methylation of the membrane by *KyaL* is essential for the initiation of production of

lantipeptides in the native system, a lack of KyaL, and thus the lack of membrane methylation, could cause repression of the kyamycin biosynthetic genes, leading to less kyamycin production and a small zone of inhibition. Again, this could be investigated through protein pull down and quantification of KyaL, or quantification of *kyaL* mRNA, comparing the KyaL protein or mRNA levels between the M1146/pEVK6/pWDW63 and M1146/pAMA1/pWDW63 strains over time. Alternatively, the level of membrane methylation could be monitored overtime and compared to overall kyamycin production.

Comparison of the kyamycin core peptide sequence to the core peptide sequences of the lantipeptides produced and not produced from the kyamycin heterologous cassette system revealed several key factors that influenced the successful production of the lantipeptides. One influencing factor is the length of the core peptide, with all but one of the produced lantipeptides having a core peptide 19 amino acids long, the same length as kyamycin. Thus, all the key amino acid residues used to produce the lanthionine, methyllanthionine and lysinoalanine bonds are spatially conserved. Another factor is the presence of the conserved proline, which is important for the conformation of the lantipeptide and formation of the PE binding pocket. Lastly, the number of amino acid changes as compared to the kyamycin core peptide sequence effects the production of the lantipeptides, with the produced lantipeptides having on average fewer amino acid changes compared to the kyamycin core peptide sequence when compared to the lantipeptides not produced.

However, closer comparison of the lantipeptide core peptide sequences also reveals that several more factors must influence the production of the lantipeptides. When comparing the core peptide sequences of the produced lantipeptide L_16 and the not produced lantipeptide L_17, both were 20 amino acids long, containing an additional alanine after K19 of the core peptide sequence. For both lantipeptides the key amino acid residues involved in the formation of the lanthionine, methyllanthionine and lysinoalanine bonds were conserved and the conserved proline was present. When comparing the amino acid residues changes compared

to the kyamicin core peptide L_16 differed by 6 residues and L_17 differed by only two residues, with one of these residues in each lantipeptide being the extra Ala20. As such, it would be expected that between the two L_17 would be the lantipeptide more likely to be produced using the kyamicin heterologous synthetic cassette system. Thus, the single amino acid change of Ala7 to Ser7 in L_17 effected the production of the lantipeptide more than all other amino acid residue changes in L_16. This could be due to the nature of the amino acid change, as alanine is non-polar residue while serine is a polar residue, so the residue will behave very differently. In Chapter 4 a lantipeptide containing the Ala7 to Ser7 mutation but lacking the extra Ala20 was successfully produced from the kyamicin heterologous synthetic cassette system. However, it was produced to only very low levels with almost no hydroxylation of the compound. So, for L_17 this amino acid change along with the extra Ala20 was enough to abolish production of L_17 entirely. This shows that single amino acid changes can have a large impact on the ability of the kyamicin biosynthetic machinery to produce a lantipeptide depending on the position of the amino acid and the nature of the modification. This is further investigated in Chapter 4 of this thesis.

Using LCMS it was determined that most lantipeptide products were produced as a mixture of congeners, including the hydroxylated product (at the Asp15 residue) and the deoxy biosynthetic precursor. In several cases these deoxy forms also displayed antibacterial activity (as seen for strains M1146/pEVK6/L_10, L_11, L_13, L_22, L_23, L_26 which only produced deoxygenated lantipeptides). As post translational modification enzymes often bind to the leader peptide before acting upon the core peptide (Oman & van der Donk, 2010) and the kyamicin leader peptide was still maintained within the biosynthetic cassette it was anticipated that the post translational modification enzymes would work to high efficacy. The lack of hydroxylation for some of the molecules indicates that *kyoX* does not work at high efficiency in every case, and appears to display some degree of specificity for the core peptide sequence. It has been previously hypothesised that the abundance of deoxy lantipeptide derivatives produced in heterologous host expression systems occurs to reduce the strain on the host, as hydroxylation of Asp15 is reported to

play a part in binding to the target PE, and thus the antibacterial activity of the compound (Ahmed *et al.*, 2020; Lopatniuk *et al.*, 2017; Vestergaard *et al.*, 2019; Vikeli *et al.*, 2020; Zhao *et al.*, 2008). Deoxy lantipeptides are reported to have reduced activity compared to the hydroxylated variants, which could lessen the strain on heterologous hosts that do not naturally produce lantipeptides and thus have systems to establish immunity. However, it was previously seen that the deoxy form of kyamicin was produced by the parental strain *Saccharopolyspora* KY3, and the deoxy form of rosiermycin was produced by parental strain *Streptomyces roseoverticillatus*. This shows that deoxy intermediates are relatively common even within strains that naturally produce lantipeptides and so should have systems to establish immunity to their antibacterial activity.

The lack of hydroxylation by LanX enzymes could potentially be attributed to a lack of *lanX* transcription in comparison to the lantipeptide core peptide, as it is known that short intergenic repeats and stem loops, structures associated with a decreased level of translation and ribosome 'stalling', occur in the intergenic region between *lanA* and *lanM* in class II type B lantipeptide BGCs (Boakes *et al.*, 2010; McAuliffe *et al.*, 2001a; McAuliffe *et al.*, 2001b; Widdick *et al.*, 2003). However, as the lantipeptide biosynthetic machinery is shared across all strains, including the intergenic region between *lanA* and *lanM*, it would be unexpected for one strain to have a significantly different production of KyaX compared to another within the same heterologous host expression platform. Alternatively, KyaX could have a variable catalytic rate for different lantipeptide core peptide substrates, with the differences between the lantipeptide core peptide and the kyamicin core peptide reducing the catalytic rate of Asp15 hydroxylation. To determine whether the amount of KyaX produced or the catalytic rate of the enzyme for different lantipeptide core peptide substrates have a greater impact on Asp15 hydroxylation the amount of KyaX could be quantified and compared to the amount of lantipeptide core peptide produced, and the prevalence of Asp15 hydroxylation. Should the amount of KyaX not differ between strains with high prevalence of Asp15 hydroxylation compared to strains with high prevalence of deoxy lantipeptide then it is likely that the lack of binding efficiency, and thus catalytic activity, of KyaX is

responsible for the high-level production of deoxy lantipeptide intermediates produced using the kyamicin heterologous expression system.

To try and further produce novel lantipeptides and increase our understanding of the constraints of the kyamicin synthetic cassette system a second cassette system was created based upon the putatively named rosierymycin BGC of *S. roseoverticillatus*, following the same design principles of the kyamicin synthetic cassette system. The putatively named rosierymycin heterologous synthetic cassette system contained genes homologous to the kyamicin BGC with homologous gene organisation. Synthetic DNA fragments containing the 35 potential *lanA* core peptides previously inserted into the kyamicin synthetic cassette system, as well as a fragment containing the kyamicin core peptide, were inserted into the rosierymycin synthetic cassette system. However, the system failed to produce any of the 36 potential lantipeptides, including rosierymycin. Firstly, it was believed this was due to the incorporation of a truncated version of the *rosL* immunity gene on the pAMA3 activator cassette. However, inclusion of a *kyal* immunity gene on a further cassette pAMA4 and substitution of pAMA3 with cassette pAMA5, containing a non-truncated *rosL* immunity gene along with the *rosR1* SARP transcriptional activator, failed to induce lantipeptide production from the rosierymycin synthetic cassette system.

Thus, a three-pronged investigation into the rosierymycin synthetic cassette system needs to take place. Firstly, the production of the SARP promoter RosR1 within the host should be assessed; should this not be produced it is likely that it is being targeted for destruction and thus is not capable of inducing activation of the biosynthetic pathway. To circumnavigate this the biosynthetic cassette could also be placed under the control of a constitutive promoter. Secondly, the methylation of PE within the cell membrane should be assessed; lack of methylation would be the result of an inactive *rosL* immunity methyltransferase gene. It has already been postulated that secondary regulatory mechanisms are in place to prevent the cell from producing large amounts of lantipeptides before full methylation of the cell wall (O'Rourke *et al.*, 2017). However, inclusion of the *kyal* PE-methyl transferase,

known to work within the heterologous host *S. coelicolor* M1146 from the success of the kyamicin synthetic expression system, did not induce the production of lantipeptides from the rosiermycin synthetic cassette system, so lack of PE-methylation may not be a deciding factor in the lack of lantipeptide production. Thirdly, the production of post translational modification enzymes *rosM* and *rosN* should be assessed; if *rosM* and *rosN* are not being produced lantipeptides will not form the characteristic lanthionine and methylanthionine bonds and will not cyclise through the formation of the lysinoalanine bridge. These bonds, characteristic of class II type B lantipeptides, make the compounds very stable, without them it is possible that the pro-peptide would break down within the cell.

Due to the failure of the rosiermycin synthetic cassette system to produce any lantipeptides, including rosiermycin itself, *S. roseoverticillatus* was investigated for its potential to produce rosiermycin. When grown under laboratory conditions *S. roseoverticillatus* did produce a compound capable of inhibiting the growth of *B. subtilis* EC1524, confirmed by LCMS analysis to share the predicted characteristic of rosiermycin. To try and enhance production of rosiermycin the activation cassettes pAMA1 (kyamicin *kyaR1* to *kyaR*) and pAMA5 (*rosR1L* and *rosL*) were inserted with the intention that overexpression of a lantipeptide SARP transcriptional activator under control of a constitutive *ermE** promoter would result in an increase in rosiermycin production. However, while insertion of cassettes pAMA1 and pAMA5 caused a phenotypic change in *S. roseoverticillatus*, resulting in a loss of pink pigmentation and reduced growth, rosiermycin production was reduced compared to the of the wild type strain, with deoxy rosiermycin production decreased 10-fold and hydroxylated rosiermycin production abolished. This unexpected effect on growth and reduction in rosiermycin production indicates that overexpression of either the *lanR1* SARP transcriptional activator or the *lanL* immunity PE-methyltransferase gene caused a negative feedback loop on lantipeptide production and affected further metabolic processes within the cell, resulting in reduced growth and loss of pigmentation. Thus, further investigation into the regulatory mechanisms of lantipeptides may reveal potential cross-regulation and secondary functions within the cell.

To validate our heterologous expression platform approach, we attempted purification of several new lantipeptides from the kyamicin synthetic cassette system, using HPLC systems. During this work it was noted that in some cases up to three isomers of each lantipeptides hydroxylated and deoxy forms (if both were produced) were present in the scaled-up fermentation products, except for during kyamicin production where only one isomer was seen (Eleni Vikeli, JIC). It was also observed that rosiermycin was also produced as a single isomer when produced from the parental strain *S. roseoverticillatus*. We hypothesised that these apparent isomers might correspond to variations in the status of the lysinoalanine bridge region of the lantipeptides. Previous work showed that although formation of the lysinoalanine bridge is catalysed by LanN enzymes it can occur spontaneously, leading to formation of two epimeric forms of the compound (An *et al*, 2018). We thus hypothesised that if KyaN does not function equally efficiently for all of the altered lantipeptide products, due to greater specificity to the kyamicin core peptide than previously thought, then the three chromatographic peaks observed might correspond to the uncyclized and two epimeric forms of the compounds. Given that in the case of DurN, during duramycin biosynthesis, the hydroxyl group at Asp15 provides substrate assistance for the cyclization reaction (An *et al.*, 2018), we reasoned that for the deoxy lantipeptide forms observed the major product may be non-cyclised, but this was difficult to determine purely based on peak areas, as these varied considerably.

While several lantipeptides were purified none showed greater antibacterial activity than the commercially available duramycin, and rosiermycin had the greatest antibacterial activity of the novel lantipeptides. Lantipeptide L_26 Deoxy F2 showed no bioactivity against *B. subtilis* EC1524, although no hydroxylated L_26 was seen during purification and previously a zone of inhibition was observed from the strain M1146/pEVK6/L_26. This was expected, as it was thought that only the later eluting fraction of the lantipeptide contained an active compound, as was seen previously with lantipeptide L_19 Deoxy F3, with the earlier eluting lantipeptide fractions predicted to contain an non-cyclised or incorrectly cyclised epimer of the compound

which were inactive, due to the lysinoalanine bridge being essential for the formation of the PE binding pocket (An *et al.*, 2018; Huo *et al.*, 2017). However, lantipeptide L_18 Deoxy F1, purified from the fastest eluting chromatographic fraction and thus expected to contain a non-cyclised or incorrectly cyclised epimer of the lantipeptide compound, showed antibacterial activity against *B. subtilis* EC1524. Thus, the elution sequence of the cyclised, incorrectly cyclised epimer and correctly cyclised active form of the lantipeptides could be unexpectedly variable. To investigate this the compounds should be re-purified, this time with the aim to purify all potential epimers and compare their antibacterial activity. To try and increase the relative amount of the correctly cyclised isomer for further assessment, *kyaN* could be overproduced by insertion into a separate cassette under control of a constitutive promoter.

Overall, the heterologous expression platform described here displayed great potential for the production and rapid diversification of class II type B lantibiotics. While there are still many questions to be answered about why the kyamicin synthetic cassette system has shown such success in producing lantipeptides when a system created following the same design principle but derived from *S. roseoverticillatus* failed to produce any lantipeptides at all, the expression platform nonetheless provides a foundation upon which future lantipeptide research and production can be based. As other classes of lantipeptides BGCs also have core peptides upon which post translational modifications occur these too could be produced and diversified using a similar model of synthetic cassettes in a heterologous expression platform.

Chapter 4:

A Synthetic Library of Lantipeptides

4. Design of a Synthetic *lanA* Core Peptide Library for Incorporation into the Kyamicin Synthetic Cassette System

4.1 Introduction

In Chapter 3 it was shown that the kyamicin biosynthetic machinery was capable of performing post-translational modification of hybrid *KyaA* proteins by modifying the core peptide region. Using this principle and a kyamicin heterologous expression system the core peptides of naturally occurring lantipeptides, identified through genome mining, were inserted into the kyamicin *kyaA* gene and 16 lantipeptides produced. Given this success, and the relative ease with which changes could be introduced into the core peptide encoding region of the *kyaA* gene, we sought to further diversify class II type B lantipeptides and investigate the constraints of the kyamicin expression platform, by constructing an exemplar lantipeptide library using synthetic biology principles.

Furthermore, extensive libraries of lantipeptide RiPPs based upon ProcA2.8 core peptides, created using the promiscuous substrate tolerant enzyme ProcM, which dehydrates and cyclizes up to 30 different linear precursor peptides encoded *Prochlorococcus* genome, have been developed in *E. coli* (Yang *et al*, 2018). The ProcA2.8 core peptide sequence was randomised at 10 amino acid positions using the N_WY codon, which can introduce one of eight potential amino acids at each position, creating a mutagenesis library of 10⁶ core peptides. The library of compounds was transformed into a plasmid containing the ProcM enzyme and the produced compounds screened for their ability to bind to a single target of interest, in this case a proton pump inhibitor (PPI) seen to be essential for the budding of human immunodeficiency virus (HIV) from infected cells. Binding to this target was linked to the de-repression of kyamicin resistance gene and *HIS3* gene, which act as selection markers. Thus, only cells containing a compound binding to the target will survive upon selection medium containing kyamicin and histidine. The oligonucleotide sequence for compounds showing strong binding to the target

molecule was determined through plasmid extraction and sequencing, identifying compound XY3-3 which bound strongly to the target (Yang *et al.*, 2018).

While this approach to library creation and screening allows for the rapid production and testing of huge libraries it is only possible within an *E. coli* or similar host system. In which the compounds are produced from plasmids, without the need from chromosomal insertion. It also relies on a highly engineered selection system, with the binding target linked to a factor capable of causing de-repression of a selectable marker upon binding of the target to a NP. As it was not within the scope of this project to design such a system for the binding target of the class II type B lantipeptides PE, and as the synthetic heterologous cassette system could only be utilised for the production of lantipeptides when chromosomally inserted into *Streptomyces*, a bacterium which does not allow for the easy extraction of DNA, the creation of such extensive oligonucleotide libraries was beyond the reach of this project. Instead, a four-part single site mutagenesis library was created.

Previously, a single saturation mutagenesis library of type B lantipeptide actagardine A has been created. Wherein each amino acid of the actagardine A core peptide, except for those involved in bond formation, was substituted with every other potential amino acid, creating a library of 228 mutants (Boakes *et al.*, 2012). Each mutant was individually inserted into host strain actinomycete *Actinoplanes garbadinenis* ATCC 31049 and screened for NP production and antimicrobial activity. A similar single saturation mutagenesis library for kyamicin could be accomplished using the two-part heterologous synthetic cassette previously described in Chapter 3. If residues Cys1, Thr4, Cys5, Ser6, Thr11, Cys14, Thr18, Lys19, which are involved in the formation of the conserved lanthionine, methyllanthionine and lysinoalanine bonds, as well as the conserved Asp15, which is hydroxylated by *lanX* and is important for formation of the PE binding pocket and thus antimicrobial activity, are excluded then single saturation mutagenesis library of 190 mutants could be created based upon the kyamicin core peptide. However, due to the time constraints of this project, it was deemed infeasible to screen such an extensive library. Instead, the four amino acids seen to have the greatest

variability in the lantipeptides identified in Chapter 3.4 were substituted with every other potential amino acid, to create a four-part library of 78 potential mutants. These mutants were screened for antimicrobial activity against *B. subtilis* EC1524 and for lantipeptide production by LCMS.

It was assumed that all antibacterial activity exhibited by lantipeptides produced from the mutagenesis of kyamicin would occur due to the binding of the target PE, and thus destabilisation of the membrane, as previously discussed in Chapter 1.4.1. Thus, it is expected that substitution of amino acid residues not involved in the binding to PE would have less of an effect on the activity of the lantipeptides compared to those that make up the PE binding pocket. It has been shown for cinnamycin and duramycin that the PE binding pocket is comprised of amino acid residues Phe7 to Cys14, and it is expected that the same amino acids make up the binding pocket in kyamicin (**Fig 4.1**) (Zhao, 2011).

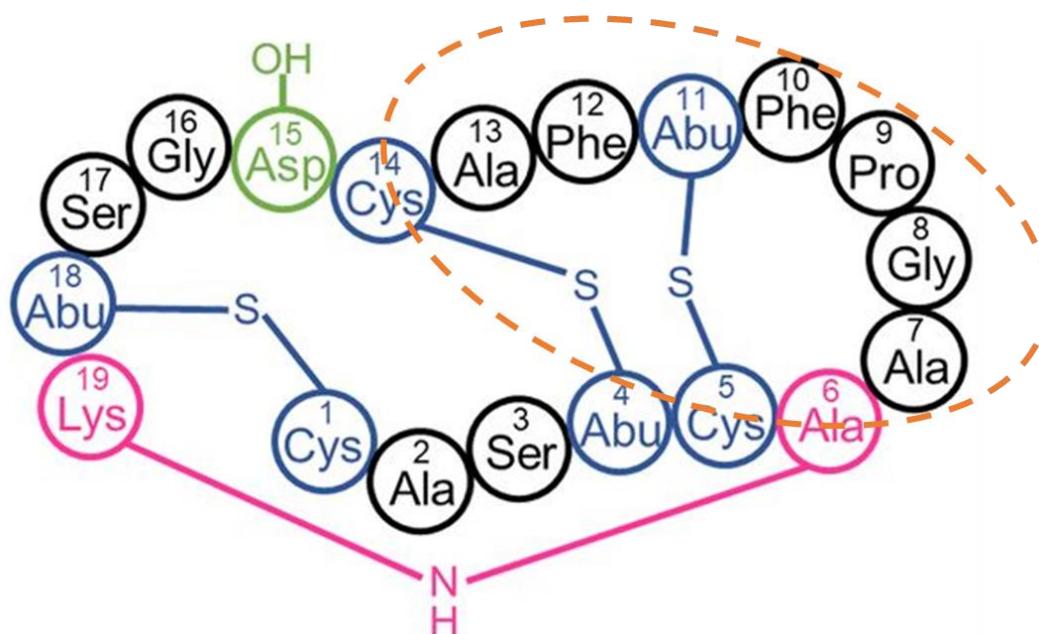


Figure 4.1: Schematic of Kyamicin with PE Binding Pocket. 19 amino acid structure of kyamicin with conserved lantionine and methylanthionine bonds (blue), lysinoalanine bond (pink) and hydroxylated Asp15 (green). PE binding pocket formed by Cys5 to Cys14 circled (orange dashed line).

The PE binding pocket resembles a hydrophobic pocket that fits around PE, with an extensive hydrogen-bonding network forming between all three hydrogen atoms of

the primary ammonium group of the PE head (Vestergaard *et al.*, 2019). Furthermore, cinnamycin reacts with the phosphate group of PE through a phosphate-binding site formed by the Phe10-Abu11-Phe12-Val13 moieties (Vestergaard *et al.*, 2019). Additionally, the carboxylate of hydroxylated Asp15 forms a tight interaction with the charged amine of PE, adding to the selectivity of cinnamycin for the lipid. This binding is also stabilised by an ionic interaction pair between the PE ammonium group and the carboxylate of the hydroxylated Asp15 (Vestergaard *et al.*, 2019; Zhao, 2011). Thus, should the substitution of amino acids in the kyamicin core peptide effect the ability of enzyme LanX to hydroxylate Asp15 it could be expected that the PE binding pocket would not form correctly and the lantipeptide would lose antibacterial activity. Overall, it is expected that the mutagenesis of amino acids within the binding pocket will have a greater impact on the lantipeptides antibacterial activity compared to changes in amino acids outside the binding pocket.

4.2 Objectives

The objectives of this project were:

- Design of a synthetic library of class II type B lantipeptides based on the kyamicin core peptide sequence
- Production of synthetic class II type B lantipeptides using the kyamicin heterologous expression platform
- Assessment and relative quantification of production and antibacterial activity of synthetic lantipeptides produced from the kyamicin heterologous expression platform

4.3 Design of a Four-Part Synthetic Lantipeptide Oligonucleotide

Library

To achieve diversification of class II type B lantipeptides a four-part synthetic library of oligonucleotides was designed. Within each oligonucleotide library a single amino acid residue of the kyamycin core peptide was changed and replaced with all other possible residues. The residues Ala2, Ala7, Phe12 and Ala13 were changed (**Fig 4.2**). These residues were selected for modification as they were not associated with the lanthionine and methyllanthionine bond formation or the lysinoalanine bridge and showed high variability within the naturally occurring lantipeptide core peptides, identified through genome mining in Chapter 3.4.

Each oligonucleotide library contained an equal mixture of all possible amino acid variants and were ordered from GenScript (Netherlands). The libraries were called A2, A7, F12 and A13 respectively based upon the amino acid residue changed.

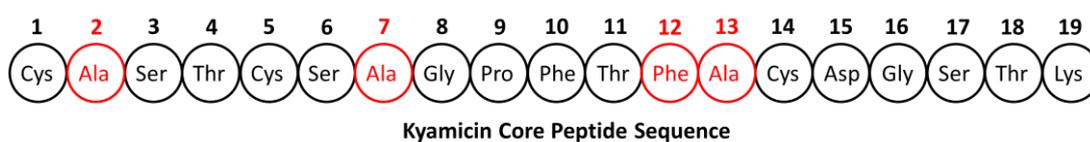


Figure 4.2: Kyamicin Core Peptide Amino Acid Sequence. Red residues were selected for synthetic library variation. Numbers correspond to amino acid number.

4.4 Heterologous Production of a Four-Part Synthetic

Lantipeptide Library

To produce each library, 76 synthetic DNA fragments (19 for each library) containing the nucleotide sequence of the modified kyamycin core peptides were designed with regions homologous to 30 bp either side of the unique *StuI* site in the kyamycin biosynthetic cassette pWDW70. The synthetic DNA fragment libraries were then cloned using Gibson assembly into *StuI* linearised pWDW70, and the reaction was transformed into *E. coli* NEB® 5-alpha Competent (High Efficiency) cells, as described in Chapter 3.5. It was assumed that the synthetic lantipeptide fragments had all assembled into the linearised vector and were present in a normal distribution throughout the plasmids.

To begin with the transformed *E. coli* were then used to inoculate an overnight culture of LB media and incubated overnight to produce a mixed cell population containing plasmids covering the library of synthetic *lanA* core peptide oligonucleotides. The plasmid libraries were not sequenced and it was assumed there was a normal distribution of the different synthetic lantipeptide core peptides within the plasmid libraries. This was not sequenced and it was assumed that the synthetic *lanA* core peptides were then extracted to create a plasmid library. The plasmid library was transformed by electroporation into *E. coli* ET12567/pUZ8002 to create a mixed cell population. This mixed cell population was transferred to LB media and incubated overnight. An aliquot of the cell culture was used to inoculate fresh LB medium, and this mixed cell population was then used to conjugate the heterologous host *S. coelicolor* M1146/pEVK6, resulting in a mixed population of *S. coelicolor* M1146 spores containing a single synthetic lantipeptide construct from the library. The mixed spore population was plated onto SF+M media with added MgCl₂ with antibiotic selection for single colonies.

The open access GLUE programme (Firth & Patrick, 2005; Patrick *et al*, 2003) was used to determine the number of resulting *S. coelicolor* M1146 colonies needed for a 0.95 probability that 100 % of the possible *lanA* variants were represented. This

assumes a normal distribution of all lantipeptide core peptides across the colonies. The calculated number of colonies needed was 112.4, therefore 120 ex-conjugant colonies were selected and tested as individual patches per library.

Each of the 120 colonies per library was grown on SF+M agar at 30 °C for a minimum of five days, mycelia from each ex-conjugant patch was then subjected to colony PCR and Sanger sequencing. Analysis of the resulting sequences indicated that the initial *S. coelicolor* M1146 libraries exhibited strong bias towards only one or two mutants. For example, initial sequencing of library A2 identified that mutants A2_W and A2_Y represented 96 % of the colonies. Similar bias was seen for the other libraries, with 19 of 20 random colonies selected from library A13 identified as mutant A13_M.

When repeating this procedure, the constructed plasmids were analysed at each stage by colony PCR followed by Sanger sequencing, this revealed that the mutant bias was stochastic and was introduced during the first Gibson assembly step during construction of the plasmids, due to inefficiency in the incorporating the core peptides into the linearised vector. Attempts to increase the efficiency of Gibson assembly through modifications of the insert to vector ratio, de-salting of the Gibson mixture for detoxification, and the use of different competent cells, proved futile, with each reaction leading to a small but random number of plasmids dominating.

As such, to compensate for this bias we chose to carry out 10 Gibson assembly reactions in parallel for each library **(1)**. Each of the 10 Gibson reactions was then transformed into *E. coli* DH5 α separately **(2)** and the plasmid DNA extracted **(3)**. Each of these plasmid DNA extractions contained a selection of mutants from the oligonucleotide library, heavily skewed towards one or two mutants. Each of the mixed plasmid populations were then transformed into *E. coli* ET12567/pUZ8002 **(4)**, all 10 of the transformed *E. coli* cultures were then pooled **(5)** before conjugation into M1146/pEVK6 **(6)**. The conjugation was then plated for single colony selection **(7)** and 120 individual colonies picked and plated for each of the

four oligonucleotide libraries (8) (Fig 4.3). Although this approach significantly improved the distribution of plasmids produced not all mutants were covered, with six mutants missing across all libraries, as seen below in 4.4.2.

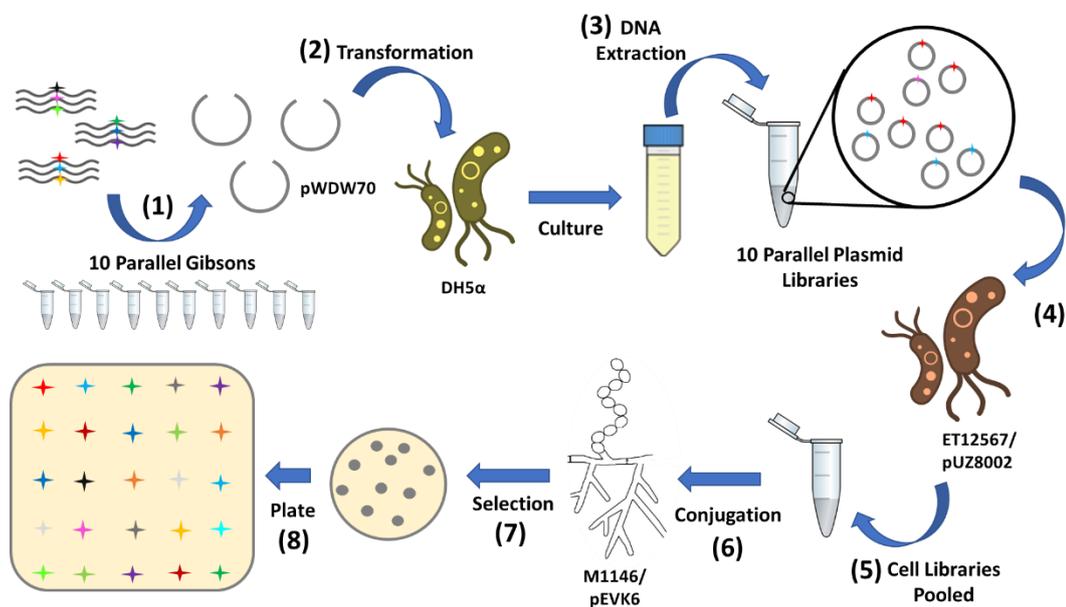


Figure 4.3: Methodology for Producing Synthetic Lantipeptide Libraries with Reduced Mutant Bias. (1) 10 Parallel Gibson assembly reactions were carried out between *Stu*I linearised biosynthetic cassette pWDW70 and one of the normally distributed lantipeptide core peptide oligonucleotide libraries. (2) All 10 Gibson assembly reactions were transformed in parallel into *E. coli* DH5 α electrocompetent cells. (3) All 10 samples were then cultured, and mixed plasmid populations extracted. (4) The mixed plasmid populations were transformed into *E. coli* ET12567/pUZ8002 and cultured. (5) All 10 *E. coli* cultures were then pooled into a single cell culture. (6) The single cell culture was then conjugated into M1146/pEVK6. (7) Transformed M1146/pEVK6 was then plated for single colony selection. (8) 120 colonies were selected and plated for each of the four synthetic lantipeptide libraries.

4.4.1 Bioassay against *B. subtilis* of the Four-Part Synthetic Lantipeptide Libraries

Following construction using the modified procedure to minimise bias, 120 colonies were spotted onto R5 plates for all four libraries (A2, A7, F12, A13), and the plates were incubated for 7 days at 30°C, overlaid with *B. subtilis* EC1524 in SNA media and incubated further overnight. The strains M1146 and M1146/pEVK6/pWDW70 were used as negative controls and strain M1146/pEVK6/pWDW63 used as a kyamicin producing positive control, which exhibited a zone of inhibition against *B. subtilis*

EC1524, as had been seen previously (Vikeli *et al.*, 2020). Clear zones of inhibition were observed for multiple colonies from each library (Fig 4.4).

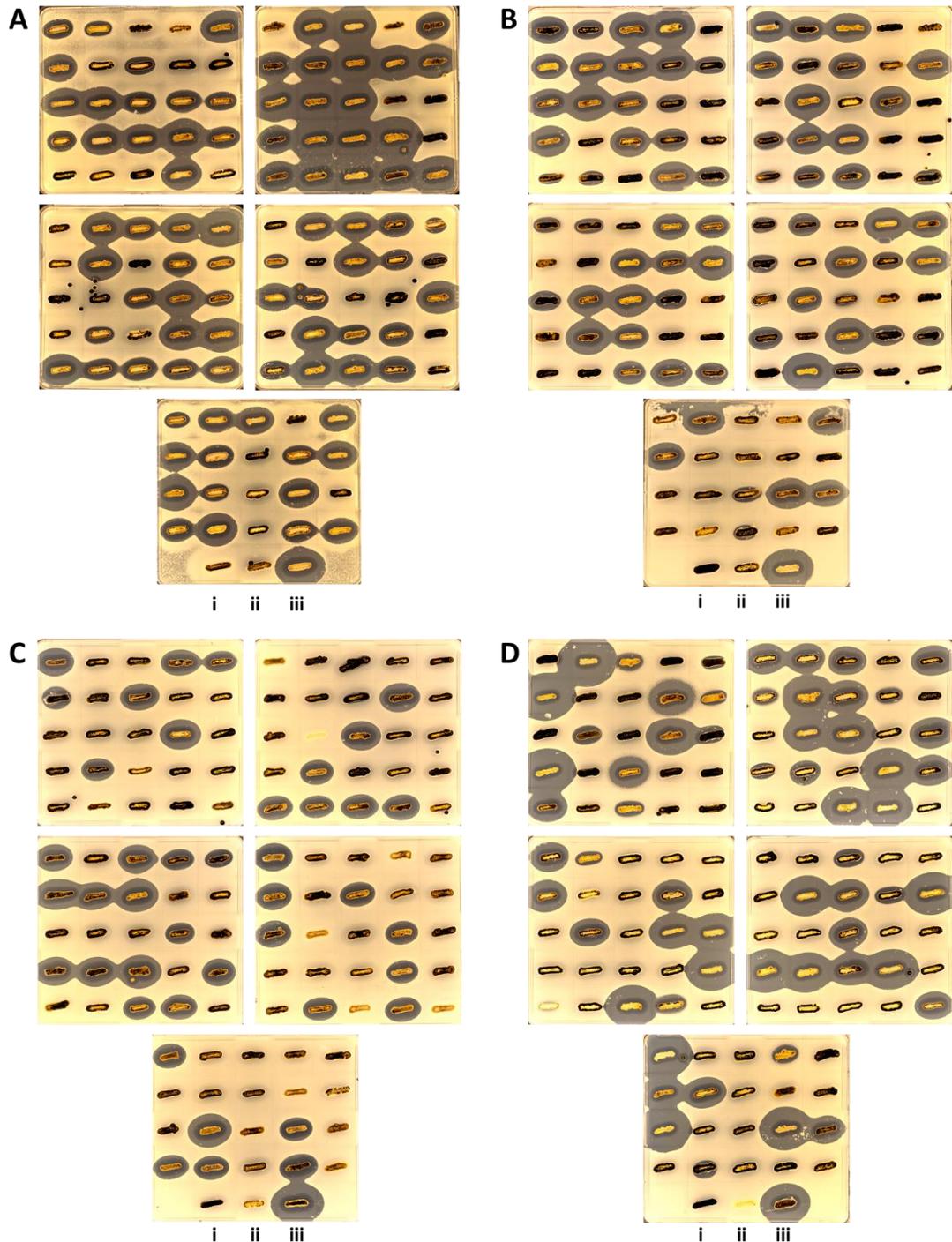


Figure 4.4: Bioassays of the Four-Part Synthetic Lantipeptide Library. Heterologous host *S. coelicolor* M1146/pEVK6, overlaid with *B. subtilis* EC1524. M1146 and M1146/pEVK6/pWDD70 used as negative controls, M1146/pEVK6/pWDD63 used as a kyamicin producing positive control. (A) Synthetic Library Ala2 (A2) (B) Synthetic Library

Ala7 (A7) (C) Synthetic Library Phe12 (F12) (D) Synthetic Library Ala13 (A13). (i) M1146, (ii) M1146/pEVK6/pWDW70, (iii) M1146/pEVK6/pWDW63.

4.4.2 Sanger Sequencing of Synthetic Lantipeptide Library Colonies and Identification of Missing Mutants

Before overlaying with *B. subtilis* master plates of each synthetic lantipeptide *lanA* library were created and used to identify the *kyaA* core peptide mutation present in each colony via colony PCR and Sanger sequencing. This showed that 70 of the potential 76 synthetic lantipeptide encoding plasmids were present. The missing mutants from the libraries were identified as A2_D, A2_Q, A2_S, A2_V, A13_R and A13_W. To complete the library synthetic DNA fragments encoding each of these core peptides were designed as before and ordered from Eurofins or IDT, cloned into the *StuI* site of pWDW70 as before, and finally conjugated into *S. coelicolor* M1146/pEVK6. Each strain was then verified by colony PCR and Sanger sequencing and a *B. subtilis* bioassay carried out as above (**Fig 4.5**). The strains M1146 and M1146/pEVK6/pWDW70 were used as negative controls and strain M1146/pEVK6/pWDW63 was used as a kyamycin producing positive control as before. Mutant A2_D produced a small zone of inhibition, with mutants A2_Q, A2_S and A2_V producing substantial zones of inhibition. Neither mutants A13_R or A13_W produced a zone of inhibition.

Calculated m/z values for all new synthetic lantipeptides were generated as before in Chapter 3, section 3.5.2, based on the core peptide amino acid formula and anticipated post translational modifications corresponding to a loss of four water molecules and the addition of a hydroxyl group at the Asp15 residue. The m/z value for congeners lacking the Asp15 hydroxylation was also calculated and used in the following MS analysis in section 4.4.3.

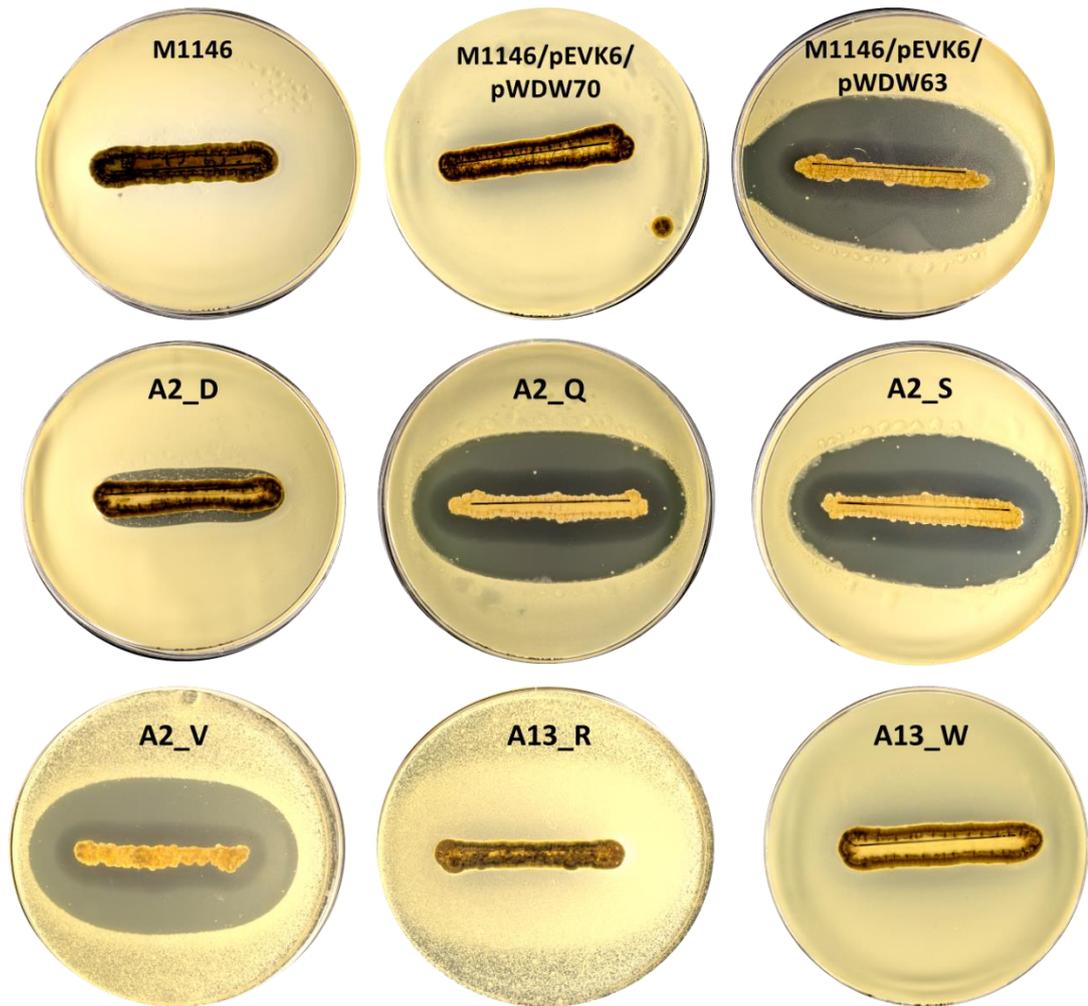


Figure 4.5 Bioassay plates of synthetic lantipeptide core peptide sequences within pWDW70. Heterologous host *S. coelicolor* M1146/pEVK6, overlaid with *B. subtilis* EC1524. M1146 and M1146/pEVK6/pWDW70 used as negative controls, M1146/pEVK6/pWDW63 used as a kyamycin producing positive control. Performed with biological triplicates. A2 and A13 represent at what residue of the kyamycin core peptide sequence the mutations occur. Mutations represented by the single code amino acid notation.

4.4.3 Chemical Analysis of Synthetic Lantipeptides Produced by Heterologous Expression of the Four-Part Synthetic Lantipeptide Library

Agar plugs were taken for all colonies in each library and from the overlay plates in Section 4.4.2, compounds were then extracted using freeze-thaw and methanol extraction with shaking for 20 min. The resulting extracts were analysed using UHPLC-MS on a Synapt G2-Si high resolution mass spectrometer using electrospray ionization. Where a lantipeptide product was seen, the observed m/z values were examined and found to be in agreement with the calculated m/z values to within ± 5 ppm for the predicted lantipeptide based upon the present core peptide sequence (data not shown).

In total, based on the combined LCMS and bioassay data, 73 synthetic lantipeptides were confirmed to be produced, with 72 of these being new lantipeptide compounds. Overall, 96% of the libraries were produced, with 100% of library A2, 100% of A7, 84% of F12 and 100% of A13 libraries produced, respectively. Bias was still seen within all libraries, with some mutants seen in abundance while others were only seen once. Of the produced lantipeptides 52 showed antibacterial activity against *B. subtilis* EC1524. For the majority of these, the lantipeptide products were produced as a mixture of the deoxy and hydroxylated forms, as seen before with lantipeptides identified through genome mining produced using the heterologous expression platform in Chapter 3.

To gain a qualitative understanding of the production level for each of the new synthetic lantipeptides, a single colony from every mutant within the libraries was selected from each LCMS dataset and the lantipeptide production analysed by integration of the area under the peak corresponding to the calculated m/z value for each predicted lantipeptide. The overall production of the lantipeptide was then ranked from 0-4, with 0 being no production and the values for 1-4 assigned based on the quartile separation of the peak area for the lantipeptides produced; this was visualised using a box plot (**Fig 4.6**). The percentage of hydroxylated versus deoxy

congener was then determined. The strength of the zones of inhibition were also ranked from 0-4, with 0 being no zone of inhibition and 4 indicating a zone of inhibition equal or larger to that produced by the kyamicin producing positive control M1146/pEVK6/pWDW63. Comparison of these different data points allowed the relative activity of each new compound to be assessed, at least in a qualitative manner

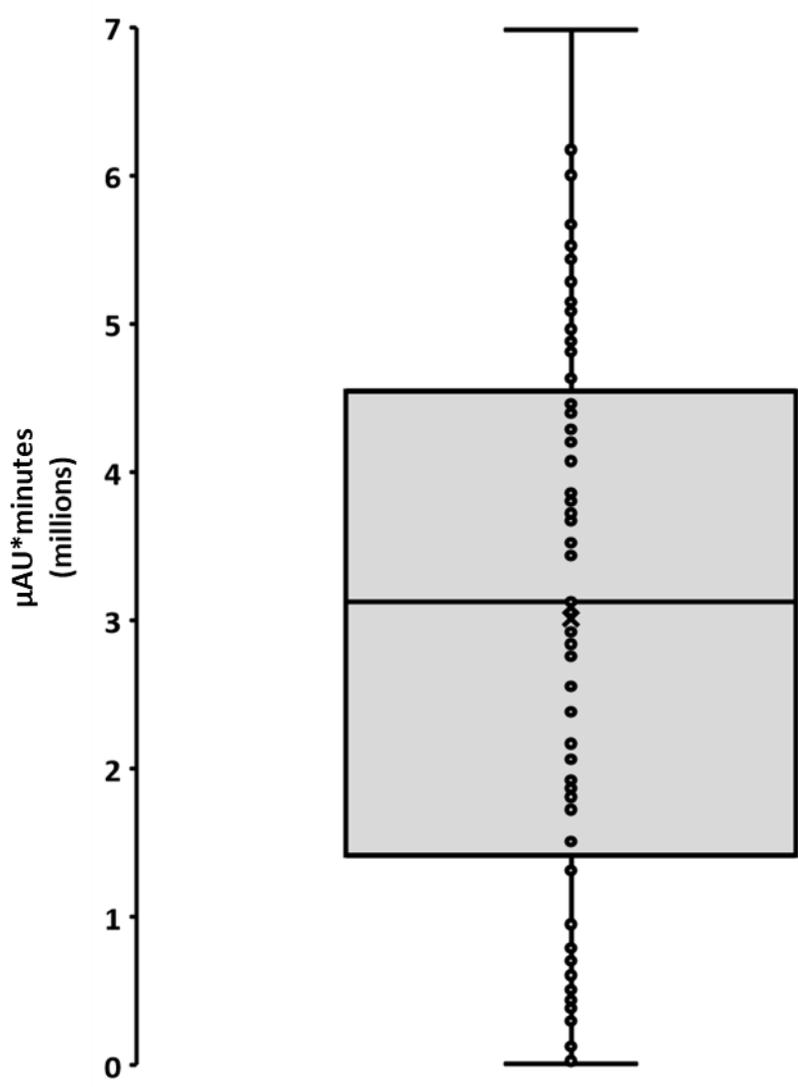
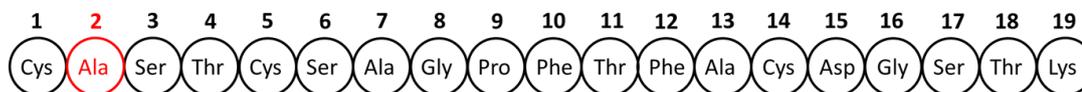


Figure 4.6: Box Plot Representation of Synthetic Lantipeptide Production. Production of lantipeptides quantified using UHPLC-MS area under a peak. 0 represent no production of the lantipeptide. End point and closed circles represent a single synthetic mutant. Quantification of lantipeptide production 0-4; 0 = 0, no production of lantipeptide; 1 = $\leq 1,400,000$; 2 = $>1,400,000$ to $\leq 3,100,000$; 3 = $>3,100,000$ to $\leq 4,500,000$; 4 = $>4,500,000$. A single replicate was used.

Each library was looked at individually and the data organised by amino acid type, percentage of hydroxylation, lantipeptide production rank and zone of inhibition rank. This was done to identify effects that specific types of amino acid changes had on the lantipeptide. It was assumed that all lantipeptides mode of action was the same as kyamicin, with the lantipeptides binding to the target PE within the membrane of the indicator strain *B. subtilis* EC1524, causing membrane destabilisation and death of the bacteria.

Library A2

Kyamicin Core Peptide Sequence



(A) Sorted by Amino Acids

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Alanine (A)	59	4	4
Isoleucine (I)	73	4	4
Leucine (L)	55	4	3
Methionine (M)	83	3	4
Phenylalanine (F)	58	3	2
Tryptophan (W)	65	3	3
Valine (V)	79	2	4
Glycine (G)	69	3	3
Cysteine (C)	53	2	1
Proline (P)	65	3	1
Serine (S)	63	2	3
Asparagine (N)	95	2	2
Threonine (T)	60	4	3
Glutamine (Q)	72	2	2
Tyrosine (Y)	83	2	4
Arginine (R)	88	2	4
Lysine (K)	80	3	3
Histidine (H)	63	2	2
Aspartic Acid (D)	67	3	1
Glutamic Acid (E)	76	4	3

(B) Sorted by % Hydroxylation

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Asparagine (N)	95	2	2
Arginine (R)	88	2	4
Methionine (M)	83	3	4
Tyrosine (Y)	83	2	4
Lysine (K)	80	3	3
Valine (V)	79	2	4
Glutamic Acid (E)	76	4	3
Isoleucine (I)	73	4	4
Glutamine (Q)	72	2	2
Glycine (G)	69	3	3
Aspartic Acid (D)	67	3	1
Proline (P)	65	3	1
Tryptophan (W)	65	3	3
Serine (S)	63	2	3
Histidine (H)	63	2	2
Threonine (T)	60	4	3
Alanine (A)	59	4	4
Phenylalanine (F)	58	3	2
Leucine (L)	55	4	3
Cysteine (C)	53	2	1

(C) Sorted by Production Rank

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Alanine (A)	59	4	4
Isoleucine (I)	73	4	4
Leucine (L)	55	4	3
Threonine (T)	60	4	3
Glutamic Acid (E)	76	4	3
Methionine (M)	83	3	4
Tryptophan (W)	65	3	3
Phenylalanine (F)	58	3	2
Glycine (G)	69	3	3
Proline (P)	65	3	1
Lysine (K)	80	3	3
Aspartic Acid (D)	67	3	1
Valine (V)	79	2	4
Cysteine (C)	53	2	1
Glutamine (Q)	72	2	2
Serine (S)	63	2	3
Tyrosine (Y)	83	2	4
Asparagine (N)	95	2	2
Arginine (R)	88	2	4
Histidine (H)	63	2	2

(D) Sorted by Zone of Inhibition Rank

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Alanine (A)	59	4	4
Isoleucine (I)	73	4	4
Methionine (M)	83	3	4
Valine (V)	79	2	4
Tyrosine (Y)	83	2	4
Arginine (R)	88	2	4
Leucine (L)	55	4	3
Tryptophan (W)	65	3	3
Glycine (G)	69	3	3
Threonine (T)	60	4	3
Serine (S)	63	2	3
Lysine (K)	80	3	3
Glutamic Acid (E)	76	4	3
Phenylalanine (F)	58	3	2
Glutamine (Q)	72	2	2
Asparagine (N)	95	2	2
Histidine (H)	63	2	2
Proline (P)	65	3	1
Cysteine (C)	53	2	1
Aspartic Acid (D)	67	3	1

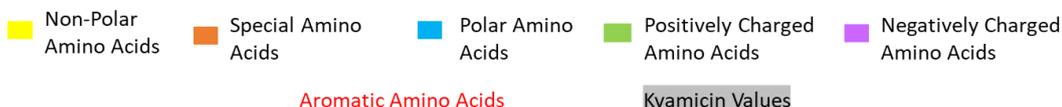
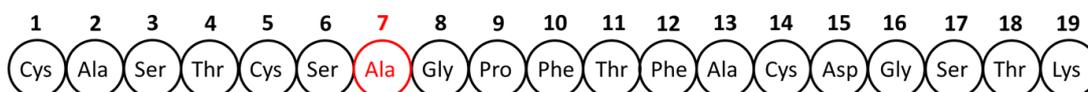


Figure 4.7: Tables Showing Effect of Single Amino Acid Mutation at Position Ala2 of Kyamicin Core Peptide. (A) A2 library sorted by amino acid type. (B) A2 library ranked by % of hydroxylated lantipeptide in descending order. (C) A2 library ranked by lantipeptide production value in descending order. (D) A2 library ranked by zone of inhibition value in descending order. Yellow = Non-Polar Amino Acids, Orange = Special Amino Acids, Blue = Polar Amino Acids, Green=Positively Charged Amino Acids, Pink = Negatively Charged Amino Acids, Red Type = Aromatic Amino Acids, Grey Shading = Kyamicin WT core peptide

Position Ala2 of the kyamicin core peptide showed the greatest variability in the natural lantipeptides identified in Chapter 3.4 (**Table 3.2**). This amino acid does not make up part of the PE binding pocket, so changes to the amino acid was not expected to affect the ability of the lantipeptides to bind PE and work as an antibacterial. While changes to amino acid did affect the percentage of hydroxylation, lantipeptide production and the size of the zone of inhibition all lantipeptides were produced and all showed antibacterial activity as expected (**Fig 4.7**). Non-polar amino acid mutants had the least effect on the production of the lantipeptides, being produced in large quantities. Similarly, these mutants also had large zone of inhibition, likely correlating to the high production value of the lantipeptides. Hydroxylation of all mutants was over 50% and mutants A2_V, A2_Y and A2_R, which had large zones of inhibition (rank 4) but lower levels of production (rank 2), had higher levels of hydroxylation (79, 83 and 88 % respectively), so had potentially higher levels of active compound compared to mutants produced with lower levels of hydroxylation, as the hydroxylation of the Asp15 residue has been shown to be important in the formation of the PE binding pocket, as previously discussed in section 4.4 (Vestergaard *et al.*, 2019; Zhao, 2011)

Library A7

Kyamycin Core Peptide Sequence



(A) Sorted by Amino Acids

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Alanine (A)	59	4	4
Isoleucine (I)	91	3	1
Leucine (L)	81	4	0
Methionine (M)	84	4	3
Phenylalanine (F)	77	3	3
Tryptophan (W)	91	2	4
Valine (V)	87	4	3
Glycine (G)	95	3	3
Cysteine (C)	99	1	2
Proline (P)	56	1	0
Serine (S)	3	1	4
Asparagine (N)	93	2	3
Threonine (T)	39	2	2
Glutamine (Q)	91	4	4
Tyrosine (Y)	86	4	4
Arginine (R)	98	1	3
Lysine (K)	94	1	0
Histidine (H)	90	4	4
Aspartic Acid (D)	100	2	0
Glutamic Acid (E)	82	3	0

(B) Sorted by % Hydroxylation

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Aspartic Acid (D)	100	2	0
Cysteine (C)	99	1	2
Arginine (R)	98	1	3
Glycine (G)	95	3	3
Lysine (K)	94	1	0
Asparagine (N)	93	2	3
Isoleucine (I)	91	3	1
Tryptophan (W)	91	2	4
Glutamine (Q)	91	4	4
Histidine (H)	90	4	4
Valine (V)	87	4	3
Tyrosine (Y)	86	4	4
Methionine (M)	84	4	3
Glutamic Acid (E)	82	3	0
Leucine (L)	81	4	0
Phenylalanine (F)	77	3	3
Alanine (A)	59	4	4
Proline (P)	56	1	0
Threonine (T)	39	2	2
Serine (S)	3	1	4

(C) Sorted by Production Rank

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Alanine (A)	59	4	4
Methionine (M)	84	4	3
Leucine (L)	81	4	0
Valine (V)	87	4	3
Glutamine (Q)	91	4	4
Tyrosine (Y)	86	4	4
Histidine (H)	90	4	4
Phenylalanine (F)	77	3	3
Isoleucine (I)	91	3	1
Glycine (G)	95	3	3
Glutamic Acid (E)	82	3	0
Tryptophan (W)	91	2	4
Asparagine (N)	93	2	3
Threonine (T)	39	2	2
Aspartic Acid (D)	100	2	0
Cysteine (C)	99	1	2
Proline (P)	56	1	0
Arginine (R)	98	1	3
Lysine (K)	94	1	0
Serine (S)	3	1	4

(D) Sorted by Zone of Inhibition Rank

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Alanine (A)	59	4	4
Tryptophan (W)	91	2	4
Glutamine (Q)	91	4	4
Tyrosine (Y)	86	4	4
Serine (S)	3	1	4
Histidine (H)	90	4	4
Methionine (M)	84	4	3
Valine (V)	87	4	3
Phenylalanine (F)	77	3	3
Glycine (G)	95	3	3
Asparagine (N)	93	2	3
Arginine (R)	98	1	3
Cysteine (C)	99	1	2
Threonine (T)	39	2	2
Isoleucine (I)	91	3	1
Leucine (L)	81	4	0
Proline (P)	56	1	0
Lysine (K)	94	1	0
Glutamic Acid (E)	82	3	0
Aspartic Acid (D)	100	2	0

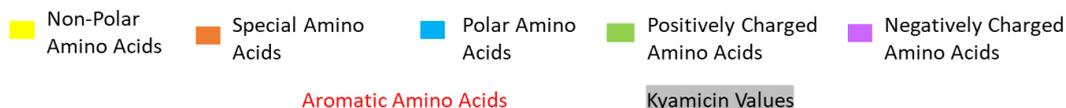
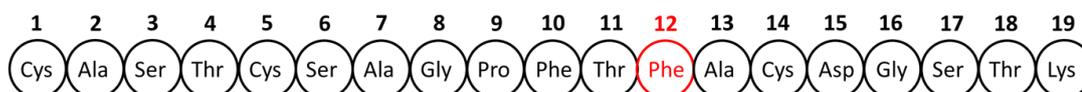


Figure 4.8: Tables Showing Effect of Single Amino Acid Mutation at Position Ala7 of Kyamycin Core Peptide. (A) A7 library sorted by amino acid type. (B) A7 library ranked by % of hydroxylated lantipeptide in descending order. (C) A7 library ranked by lantipeptide production value in descending order. (D) A7 library ranked by zone of inhibition value in descending order. Yellow = Non-Polar Amino Acids, Orange = Special Amino Acids, Blue = Polar Amino Acids, Green=Positively Charged Amino Acids, Pink = Negatively Charged Amino Acids, Red Type = Aromatic Amino Acids, Grey Shading = Kyamycin WT core peptide

Position A7 of the kyamicin core peptide is just within the PE binding pocket of the lantipeptides but is not thought to be involved in formation of bonds to the PE. It also showed a high level of variability in the natural lantipeptides identified in Chapter 3.4 (**Table 3.2**). All lantipeptide mutants were produced but not all lantipeptides showed antibacterial activity (**Fig 4.8**). Most mutants were produced with high levels of hydroxylation, above that of kyamicin (59%), with the exception of A7_P, A7_T and A7_S, of which only 56, 39 and 3% was hydroxylated, respectively. Similar to library A2, non-polar amino acids mutants had the highest level of production from all the mutants and also had largest zones of inhibition. However, most polar amino acid mutations also had large zones of inhibition, indicating that the hydrophobicity of this amino acid, and its ability to form hydrogen bonds, has little impact on the activity of the lantipeptide. This is consistent with the amino acids lack of involvement in binding directly to the target PE, despite being part of the PE binding pocket formed by the lantipeptide. However, A7_S, which showed a high level of antibacterial activity (rank 4), was produced to very low levels (rank 1) and almost exclusively in its deoxy state, with only 3 % of the lantipeptide product produced in its hydroxylated form. This suggests that the deoxy version of some lantipeptides have the potential to be highly active and form the PE binding pocket without need of the hydroxylated Asp15.

Library F12

Kyamycin Core Peptide Sequence



(A) Sorted by Amino Acids

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Alanine (A)	28	4	0
Isoleucine (I)	62	4	3
Leucine (L)	37	4	3
Methionine (M)	69	4	3
Phenylalanine (F)	59	4	4
Tryptophan (W)	--	0	0
Valine (V)	78	4	3
Glycine (G)	--	0	0
Cysteine (C)	96	3	0
Proline (P)	--	0	0
Serine (S)	45	1	0
Asparagine (N)	55	1	0
Threonine (T)	93	2	0
Glutamine (Q)	5	1	0
Tyrosine (Y)	62	3	3
Arginine (R)	0	0	0
Lysine (K)	27	1	0
Histidine (H)	33	1	0
Aspartic Acid (D)	0	1	0
Glutamic Acid (E)	0	2	0

(B) Sorted by % Hydroxylation

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Cysteine (C)	96	3	0
Threonine (T)	93	2	0
Valine (V)	78	4	3
Methionine (M)	69	4	3
Isoleucine (I)	62	4	3
Tyrosine (Y)	62	3	3
Phenylalanine (F)	59	4	4
Asparagine (N)	55	1	0
Serine (S)	45	1	0
Leucine (L)	37	4	3
Histidine (H)	33	1	0
Alanine (A)	28	4	0
Lysine (K)	27	1	0
Glutamine (Q)	5	1	0
Aspartic Acid (D)	0	1	0
Glutamic Acid (E)	0	2	0
Arginine (R)	0	1	0
Proline (P)	--	0	0
Glycine (G)	--	0	0
Tryptophan (W)	--	0	0

(C) Sorted by Production Rank

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Phenylalanine (F)	59	4	4
Valine (V)	78	4	3
Methionine (M)	69	4	3
Isoleucine (I)	62	4	3
Leucine (L)	37	4	3
Alanine (A)	28	4	0
Cysteine (C)	96	3	0
Tyrosine (Y)	62	3	3
Threonine (T)	93	2	0
Glutamic Acid (E)	0	2	0
Glutamine (Q)	5	1	0
Asparagine (N)	55	1	0
Serine (S)	45	1	0
Histidine (H)	33	1	0
Lysine (K)	27	1	0
Arginine (R)	0	1	0
Aspartic Acid (D)	0	1	0
Proline (P)	--	0	0
Glycine (G)	--	0	0
Tryptophan (W)	--	0	0

(D) Sorted by Zone of Inhibition Rank

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Phenylalanine (F)	59	4	4
Valine (V)	78	4	3
Methionine (M)	69	4	3
Isoleucine (I)	62	4	3
Leucine (L)	37	4	3
Tyrosine (Y)	62	3	3
Alanine (A)	28	4	0
Cysteine (C)	96	3	0
Threonine (T)	93	2	0
Glutamine (Q)	5	1	0
Asparagine (N)	55	1	0
Serine (S)	45	1	0
Histidine (H)	33	1	0
Lysine (K)	27	1	0
Arginine (R)	0	1	0
Aspartic Acid (D)	0	1	0
Glutamic Acid (E)	0	2	0
Tryptophan (W)	--	0	0
Proline (P)	--	0	0
Glycine (G)	--	0	0

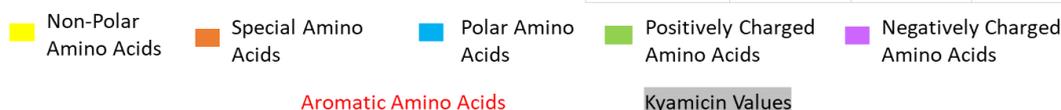


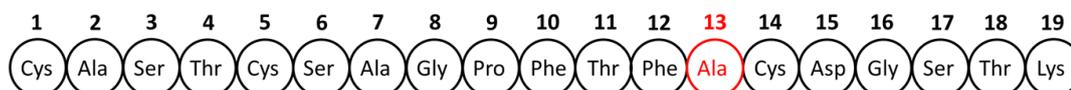
Figure 4.9: Tables Showing Effect of Single Amino Acid Mutation at Position Phe12 of Kyamycin Core Peptide. (A) F12 library sorted by amino acid type. (B) F12 library ranked by % of hydroxylated lantipeptide in descending order. (C) F12 library ranked by lantipeptide production value in descending order. (D) F12 library ranked by zone of inhibition value in descending order. Yellow = Non-Polar Amino Acids, Orange = Special Amino Acids, Blue = Polar Amino Acids, Green=Positively Charged Amino Acids, Pink = Negatively Charged Amino Acids, Red Type = Aromatic Amino Acids, Grey Shading = Kyamycin WT core peptide

Position F12 is within the centre of the PE binding pocket and showed the lowest level of variability in the natural lantipeptides identified in Chapter 3 XXX of all 4

amino acid residues varied across the synthetic lantipeptide libraries. Thus, change to this amino acid was expected to have the greatest effect on the activity of the lantipeptide (**Fig 4.9**). Indeed, three mutants, F12_P, F12_G and F12_W were not produced at all, and seven other mutants (35% of the library) were only produced in very low quantities (rank 1 production). Most mutants that were produced showed no zone of inhibition, showing that the lantipeptide had lost the ability to bind PE. Non-polar amino acids mutants, excepting F12_W which was not produced, were produced in large quantities. All produced non-polar amino acid mutants, excepting F12_A, also had had large zones of inhibition (rank 4 and 3). Only mutant F12_Y, containing the polar amino acid tyrosine, also had a zone of inhibition. Most mutants also had lower levels of hydroxylation compared to the A2 and A7 synthetic libraries, with F12_D, F12_E and F12_R only produced in a deoxygenated form. Thus, changes to the F12 amino acid can have a large effect on the hydroxylation of the Asp15 residue, previously shown to be important for the correct formation of the PE binding pocket, potentially contributing to the lack of antibacterial activity of many of the mutants.

Library A13

Kyamycin Core Peptide Sequence



(A) Sorted by Amino Acids

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Alanine (A)	59	4	4
Isoleucine (I)	74	4	4
Leucine (L)	76	4	4
Methionine (M)	89	4	4
Phenylalanine (F)	70	3	3
Tryptophan (W)	91	3	4
Valine (V)	70	3	3
Glycine (G)	78	3	1
Cysteine (C)	92	1	1
Proline (P)	51	1	0
Serine (S)	93	3	4
Asparagine (N)	88	2	0
Threonine (T)	82	3	3
Glutamine (Q)	90	2	0
Tyrosine (Y)	85	3	1
Arginine (R)	97	1	0
Lysine (K)	84	1	0
Histidine (H)	94	2	1
Aspartic Acid (D)	29	2	0
Glutamic Acid (E)	28	2	0

(B) Sorted by % Hydroxylation

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Arginine (R)	97	1	0
Histidine (H)	94	2	1
Serine (S)	93	3	4
Cysteine (C)	92	1	1
Tryptophan (W)	91	3	4
Glutamine (Q)	90	2	0
Methionine (M)	89	4	4
Asparagine (N)	88	2	0
Tyrosine (Y)	85	3	1
Lysine (K)	84	1	0
Threonine (T)	82	3	3
Glycine (G)	78	3	1
Leucine (L)	76	4	4
Isoleucine (I)	74	4	4
Phenylalanine (F)	70	3	3
Valine (V)	70	3	3
Alanine (A)	59	4	4
Proline (P)	51	1	0
Aspartic Acid (D)	29	2	0
Glutamic Acid (E)	28	2	0

(C) Sorted by Production Rank

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Alanine (A)	59	4	4
Methionine (M)	89	4	4
Leucine (L)	76	4	4
Isoleucine (I)	74	4	4
Phenylalanine (F)	70	3	3
Valine (V)	70	3	3
Tryptophan (W)	91	3	4
Glycine (G)	78	3	1
Threonine (T)	82	3	3
Serine (S)	93	3	4
Tyrosine (Y)	85	3	1
Glutamine (Q)	90	2	0
Asparagine (N)	88	2	0
Histidine (H)	94	2	1
Aspartic Acid (D)	29	2	0
Glutamic Acid (E)	28	2	0
Cysteine (C)	92	1	1
Proline (P)	51	1	0
Arginine (R)	97	1	0
Lysine (K)	84	1	0

(D) Sorted by Zone of Inhibition Rank

Amino Acid	Percentage Hydroxylated (%)	Lantipeptide Production (0-4)	Zone of Inhibition (0-4)
Alanine (A)	59	4	4
Methionine (M)	89	4	4
Leucine (L)	76	4	4
Isoleucine (I)	74	4	4
Tryptophan (W)	91	3	4
Serine (S)	93	3	4
Phenylalanine (F)	70	3	3
Valine (V)	70	3	3
Threonine (T)	82	3	3
Glycine (G)	78	3	1
Tyrosine (Y)	85	3	1
Cysteine (C)	92	1	1
Histidine (H)	94	2	1
Proline (P)	51	1	0
Glutamine (Q)	90	2	0
Asparagine (N)	88	2	0
Arginine (R)	97	1	0
Lysine (K)	84	1	0
Aspartic Acid (D)	29	2	0
Glutamic Acid (E)	28	2	0



Figure 4.10: Tables Showing Effect of Single Amino Acid Mutation at Position Ala13 of Kyamycin Core Peptide. (A) A13 library sorted by amino acid type. (B) A13 library ranked by % of hydroxylated lantipeptide in descending order. (C) A13 library ranked by lantipeptide production value in descending order. (D) A13 library ranked by zone of inhibition value in descending order. Yellow = Non-Polar Amino Acids, Orange = Special Amino Acids, Blue = Polar Amino Acids, Green=Positively Charged Amino Acids, Pink = Negatively Charged Amino Acids, Red Type = Aromatic Amino Acids, Grey Shading = Kyamycin WT core peptide

Amino acid A13 is also within the PE binding pocket of the lantipeptide, but like A7 is not thought to be involved directly in forming bonds with PE. Thus, changes to this amino acid were not thought to have as much effect on the activity of the lantipeptide compared to changes to the F12 amino acid. Indeed, once again all mutants within the library were produced, however, only two thirds of them showed antibacterial activity (**Fig 4.10**). Indicating that changes to the A13 residue has a greater effect on the antibacterial activity of the lantipeptides than changes to A7 and A2. Once again, the majority of mutants had high levels of hydroxylation, above that of kyamicin (59%), except negatively charged residues aspartic acid and glutamic acid, which had only 29 and 28 % hydroxylated compound, respectively. This shows that negatively charged residues at position A13 can have a large impact on the hydroxylation of Asp15, which can affect the proper formation of the PE binding pocket. Indeed, these mutants had no zone of inhibition and low levels of production. Non-polar amino acids were produced in the greatest quantities and had large zones of inhibition (rank 4 and 3). Some polar amino acid residues (serine and threonine) were also produced in large quantities (rank 3) and showed strong zone of inhibition (rank 3), while others showed no zone of inhibition at all (glutamine and asparagine). Overall, the hydrophobicity of the amino acid did not seem to affect the production and antibacterial activity of the lantipeptides as much as the residue having a charge, with only a single positively charged residue, histidine, showing a very minor zone of inhibition (rank 1).

General trends were also observed from the comparison of lantipeptide production to zone of inhibition, showing that there was a general trend, as seen previously with the naturally occurring lantipeptides, that the greater the production of a lantipeptide the larger its zone of inhibition. However, there were a few exceptions to this within the synthetic libraries A2 and A7. With A2_R, A2_V and A2_Y all having rank 2 production but rank 4 zones of inhibition. Similarly, A7_R, A7_S and A7_W also had a low level of lantipeptide production for the size of their zone of inhibition, with A7_S having an extremely low production rank of 1 and a zone of inhibition rank of 4. In general, mutants with a large zone of inhibition of rank 4 were produced predominantly in their hydroxylated state, at or above 75%. This is

consistent with the currently reported data that hydroxylation of the Asp15 residue is important for formation of the PE binding pocket and thus activity of the lantipeptide, as previously discussed in Chapter 3.10.1.2 (Lopatniuk *et al.*, 2017; Vestergaard *et al.*, 2019; Vikeli *et al.*, 2020).

4.5 Discussion

Within this chapter single site mutagenesis was used to produce several synthetic lantipeptides. Saturation mutagenesis libraires have been used extensively in for the production of novel biological compounds (Molloy *et al*, 2013), for directed evolution for the optimization of therapeutic proteins (Siloto & Weselake, 2012), enzymes (van der Meer *et al*, 2016), promoters (Kircher *et al*, 2019), transcription enhances (Smith *et al*, 2013) and ribosome binding sites (Wang *et al*, 2009). Saturation mutagenesis has also been used to investigate protein-sequence function relationships (Acevedo-Rocha *et al*, 2018; Hecht *et al*, 2013) and allow for the generation of biomolecular fitness landscapes (Hietpas *et al*, 2011; Klesmith *et al*, 2015; Melnikov *et al*, 2014). Using the kyamicin heterologous expression system previously discussed in Chapter 3 a four-part synthetic library of lantipeptides was engineered, based upon amino acid residue substitution of the *kyaA* core peptide sequence. Four residues, which were not part of the lanthionine, methyllanthionine bonds or lysinoalanine bridge characteristic of class II type B lantipeptides were selected for mutations. These residues were Ala2, Ala7, Phe12 and Ala13 of the *kyaA* core peptide.

Initial attempts to create a synthetic library with a normal distribution of mutants were unsuccessful, due to random bias introduced by inefficiency of the Gibson assembly of the synthetic DNA fragments into the *Stul* linearised pWDW70 kyamicin biosynthetic cassette. Attempts to increase the efficiency of the Gibson assembly step by optimization of the insert to vector ratio, de-salting of the Gibson assembly mix and use of different competent cells were unsuccessful, bringing into question the suitability for the system for further expression of synthetic libraries of lantipeptides going forward. To reduce the effect of random bias on the four-part synthetic library 10 samples were processed in parallel for each residue library. While this did not eliminate the bias, as certain mutants remained more prevalent than others beyond what would be expected to occur through random chance in a normal distribution profile, it did improve the coverage of mutants in the synthetic library to a workable standard, with 90 of the potential 96 mutants generated at

least once. Going forward, modifications will need to be made to the synthetic cassette system to allow for more efficient incorporation of synthetic DNA fragments, either by improvement and optimization of the Gibson assembly or by employing an alternative, more efficient, method for DNA assembly.

Upon the creation of a workable four-part synthetic library with minimum bias, bioassays against *B. subtilis* EC1524 using 120 ex-conjugant colonies per residue library were carried out, and extracts analysed through UHPLC-MS. In total, 71 new synthetic lantipeptides were produced, with 52 having antibacterial activity against *B. subtilis* EC1524. Of these, 14 strains displayed zones of inhibition equivalent to that of the kyamicin producing positive control M1146/pEVK6/pWDW63.

For relative quantification of the production of the synthetic lantipeptides and their activity against *B. subtilis* EC1524 a ranking system was established. Lantipeptide production was ranked from 0-4, with 0 representing very low levels of observable lantipeptide production by LCMS, and 4 being comparable production to that seen for kyamicin produced using the kyamicin cassette system. Kyamicin was considered to have a high level of production from the cassette system when relative to the production of other lantipeptides identified through genome mining and produced using the synthetic cassette system, seen in Chapter 3 (data not shown). The size of the zone of inhibition in the *B. subtilis* EC1524 overlay was also ranked from 0-4, with 0 representing no observable zone of inhibition and 4 representing a zone of inhibition again equivalent to that seen for the kyamicin producing strain.

Several variants from the libraries were seen to have relatively high levels of antibacterial activity, as seen by large zones of inhibition against *B. subtilis* EC1524, coupled with comparatively low levels of production. Mutants A2_R, A2_V, A2_Y, A7_W all have a production value of 2 and a zone of inhibition value of 4, while A7_R has a production value of 1 and a zone of inhibition value of 3 and A7_S has a production value of 1 and a zone of inhibition value of 4. Thus, these compounds are of potential interest for pre-clinical assessment by purification and MIC assays, as performed in Chapter 3. Following this, synthetic lantipeptide core peptides could be rationally designed to try and improve other clinically relevant factors as

well as antibacterial activity, such as peptide solubility by increasing the number of polar residues in the core peptide.

The pattern of antibacterial activity of the synthetic lantipeptides produced offered some insight into the importance of specific amino acids for the binding of the lantipeptides to the target PE. Amino acids Phe7 to Cys14 were seen in cinnamycin and duramycin to be within the PE binding pocket (Zhao, 2011), so it would be expected that amino acid residue substitution at A7, F12 and A13 would have a greater impact on the antibacterial activity of the produced lantipeptides compared to A2, and indeed this was seen to be true. Changes to amino acid residues that effect the ability of LanX to hydroxylate Asp15 are also expected to have an impact on the antibacterial activity of the lantipeptide, as the carboxylate of hydroxylated Asp15 forms a tight interaction with the charged amine of PE (Vestergaard *et al.*, 2019; Zhao, 2011) and is needed for correct PE binding pocket formation (Lopatniuk *et al.*, 2017; Vestergaard *et al.*, 2019; Vikeli *et al.*, 2020). So, it is expected that compounds produced with little hydroxylation would have smaller zones of inhibition as they exhibit lesser antibacterial activity.

No changes to residue A2 resulted in a loss of antibacterial activity, though reductions in the zones of inhibition within the *B. subtilis* EC1524 overlay were seen. As A2 is not involved in the PE binding pocket changes to this amino acid were not expected to cause loss of antibacterial activity. Mutants with small zones of inhibition but high levels of lantipeptide production tended to have lower levels of hydroxylation. This indicates that the dehydroxylated compounds were potentially not active, so that high level of lantipeptide production did not equal a large amount of active compound with antibacterial activity. However, overall, the variations to amino acid residue A2 did not have a large impact on the antibacterial activity, hydroxylation and production of the lantipeptides.

Changes to the A7 and A13 residues resulted in no loss of production but a few changes (D, E, K, L, P and D, E, K, P, N, Q, R, respectively) did cause a loss of antibacterial activity. These amino acid residues are within the PE binding pocket

but are not involved directly with binding to the PE. As expected, changes to these amino acids had a greater impact on the lantipeptide activity and production compared to changes to A2, but less of an impact compared to changes to F12, which is directly involved in binding to PE. Changes to the polarity of amino acids at position A7 did not affect the antibacterial activity of the lantipeptides, showing that this amino acid is not involved in the hydrogen bond formation with PE. Changes to the A13 residue had a larger impact on the antibacterial activity of the lantipeptides compared to changes in A7, with only two thirds of the lantipeptides having any antibacterial activity at all. It was observed that lantipeptides with mutations to amino acids with a charge had no zone of inhibition and low levels of production. Showing that while the A13 amino acid is not directly involved in the binding to PE having charged residues at this position may interfere with the formation of other bonds to the PE, possibly interfering with hydrogen bond formation or the formation of the phosphate bond with the nearby hydroxylated Asp15 residue. Indeed, negatively charged residues resulted in low levels of Asp15 hydroxylation, possibly due to the negative charge interfering with the binding of the LanX enzyme. While positively charged amino acids generally had high levels of hydroxylation, and so did not interfere with the LanX enzyme, these mutants still had no antibacterial activity, excepting histidine, which had a very small zone of inhibition (rank 1). Showing that a charge of any kind could impact PE bond formation.

Lastly, changes to the F12 residue almost entirely resulted in lantipeptides showing no antibacterial activity against *B. subtilis* EC1524, with no zones of inhibition observed. Only lantipeptides with mutations to other non-polar amino acids, excepting alanine, still retained antibacterial activity, the polar tyrosine (Y) being the only exception. It was expected that changes to amino acid F12 would have the greatest impact on the antibacterial activity of the lantipeptides as the phosphate group of the Phe10-Abu11-Phe12-Val13 moieties directly binds PE (Vestergaard *et al.*, 2019). Changing F12 not only resulted in loss of PE binding and antibacterial activity for most mutants but also abolished the production of three lantipeptides entirely (F12_W, F12_P and F12_G). This could be due to the changes caused to the

structure of the lantipeptide at this position causing a lack of formation of the conserved lanthionine, methyllanthionine and lysinoalanine bonds, without which the lantipeptide would maintain a linear form, have no activity and likely be targeted for degradation in the cytoplasm. Proline (P) at position 11 of the lantipeptide is essential for conformation of the peptide and formation of the PE binding pocket (Elvas *et al.*, 2017; Zhao, 2011). It is also possible that this proline effects the spatial positioning of the conserved amino acids needed for the lanthionine, methyllanthionine and lysinoalanine bond formation by the LanM and LanN enzymes. The addition of another proline (F12_P) at adjacent amino acid 12 could completely change the structure of the immature peptide and thus the spatial positioning of these conserved amino acid residues, thus effecting the ability of the enzymes to act upon the core peptide. Tryptophan (F12_W) is also the largest aromatic amino acid, so could also spatially interfere with the formation of the PE binding pocket and structure of the lantipeptide core peptide. In comparison to this glycine (F12_G) is a very small amino acid with no carbon side chains, making it a very flexible amino acid. Possibly the addition of glycine adjacent to the conserved proline introduced too much flexibility to the structure, causing a lack of formation of the PE binding pocket and again interfering in the spatial positioning of the conserved amino acids, preventing the the biosynthetic post-translational modification enzymes from acting.

Further synthetic *lanA* core peptide libraries modifying all other 15 amino acid residues in turn could be created to monitor the effect changes to each residue has on the binding efficiency to PE, as well as production of the lantipeptide. With residues Gly8, Pro9, Phe10, Thr11 and Cys14, in the binding pocket of PE, expected to have a large effect on the antibacterial activity of the compound. Substitution to conserved residues Cys1, Ser3, Cys5, Ser6, Thr11, Cys14, Thr 18 and Lys19, involved in the formation of the lanthionine, methyllanthionine and lysinoalanine bonds would also be expected to have an large impact on the antibacterial activity of the compounds, or more likely abolish production of the lantipeptides altogether, as the mature peptide would not form, and the immature pre-peptide would likely degrade within the cell. Abolishment of lantipeptide production by changes to

amino acid residues not directly involved in the formation of these conserved characteristics could indicate a repression effect on the post-translational modification enzymes.

Potentially, amino acid changes may impede the enzymes from acting upon the core peptide due to physical obstruction. Alternatively, amino acids in the core peptide could provide some assistance in specially positioning the post-translational modification enzymes, with changes to these residues impeding this effect. While it is known that the *lanA* leader peptide contains recognition domains to which the post-translational modification enzymes bind (Oman & van der Donk, 2010), and that this binding is important for efficient post-translational modification of the core peptide, little is known about the effect, if any, of residues in within the core peptide sequence on the special positioning and activity of the post-translational modification enzymes.

Changes to the conserved residue Asp15 could result in the loss of hydroxylation of the lantipeptides. This hydroxylation has been previously seen to be important for the formation of the PE binding pocket (Lopatniuk *et al.*, 2017; Vestergaard *et al.*, 2019; Vikeli *et al.*, 2020), but results from this study have shown that it could be possible for the PE binding pockets to form correctly without the need for hydroxylation. As seen with the mutant A7_S, of which only 3% of the compound is hydroxylated, but which exhibited a zone of inhibition rank 4, despite only having a production rank 1. Thus, it is expected that the deoxy version of the compound is active, and thus has a correctly formed PE binding pocket despite the lack of hydroxylation at the Asp15 residue. As such, substitution of Asp15 might not necessarily result in the loss of antibacterial activity for the lantipeptides.

Overall, we rapidly diversified the lantipeptides produced using the kyamicin heterologous expression system, with 87 lantipeptides produced, including those assessed in Chapter 3. Most of these lantipeptides were produced in a combination of their hydroxylated and deoxy forms. Of these lantipeptides 83 were novel, with 61 having antibacterial activity against *B. subtilis* EC1524. The kyamicin

heterologous expression system shows great promise as a system to produce synthetically designed lantipeptides, allowing for rational design and pre-clinical assessment after purification of the lantipeptides. It also provides the opportunity to investigate the relationship between amino acid residues within the lantipeptide core peptide with the binding target PE. However, should larger synthetic libraries be created, optimization of the incorporation of mixed populations of synthetic DNA fragments into the biosynthetic cassette is needed to prevent the introduction of random bias and assess all synthetic lantipeptide variants.

Chapter 5:
Production of
Actinomycetes NPs
in *S. erythraea* ISOM

5. Expression of Actinomycete BGCs in

Saccharopolyspora erythraea ISOM

5.1 Introduction

The recent increase in accessibility of whole genome sequencing has led to the discovery of many cryptic BGCs with no known associated product produced under laboratory conditions, including many BGCs which may encode for antimicrobial NPs (Foulston, 2019; Sekurova *et al*, 2019). The need for novel antimicrobials to tackle the rising threat of antimicrobial resistant pathogens, which has been reported to cause more than 35,000 deaths per year in the USA alone ((CDC) Centers for Disease Control & Prevention, 2019; Prestinaci *et al*, 2015), has led to increased focus on investigating cryptic BGCs to elucidate their products. One current approach to uncovering the products of cryptic BGCs is to use well characterised heterologous host strains as platforms for their expression (Chen *et al*, 2019). Advantages of these heterologous hosts over parental strains are that they have been well studied, can be easily genetically modified and grow well under laboratory conditions. However, finding a suitable heterologous host within which to successfully produce NPs of cryptic BGCs can be difficult, with a several factors affecting the success of the host, such as the similarity of regulatory elements to those of the parental strain and availability of metabolites in the cell, as previously discussed in Chapter 1, section 1.2.5. Due to these difficulties there is a need to expand the range of heterologous hosts available in order to achieve production of the many different cryptic BGCs identified within the public database (Fernández & Vega, 2016; Zhang *et al*, 2016).

The focus of this project was the assessment of *Saccharopolyspora erythraea* ISOM as a heterologous host. *S. erythraea* ISOM was provided by Isomerase Therapeutics (Cambridge, UK) and is derived from an industrial, medium titre erythromycin producer from which the erythromycin BGC has been deleted. Plasmids and BACs can be introduced at specific integration sites on the chromosome without compromising the host genome and the strain behaves well using standard

microbiological techniques (personal communication, Dr Matthew Gregory, Isomerase Therapeutics).

In this chapter a series of phage P1 artificial chromosomes (PACs) encoding BGCs of interest, including BGCs for known compounds anthracimycin, antascomycin (**Fig 5.6**) and formicamycin (**Fig 6.2**), in addition to BGCs for cryptic NPs were introduced into *S. erythraea* ISOM to assess its potential for heterologous expression. PACs were created by Bio S&T Inc, Montreal, Canada by isolation of genomic DNA from actinomycetes strains of interests, followed by the generation and insertion of DNA fragments containing BGCs of interest into a pESAC13 vector (Appendix 9.5) (Sosio *et al*, 2000; Tocchetti *et al*, 2018). pESAC13 is derived from *E. coli*-*Streptomyces* Artificial Chromosome pPAC-S1 (Sosio *et al.*, 2000) and contains a phage P1 origin of replication, a ϕ C31 integration system for stable site-specific insertion into actinomycetes genomes and an *oriT* replication site allowing for the transfer of the PAC into actinomycetes by conjugation. Previously, PAC integration into heterologous hosts has been used to successfully produce FK506 (tacrolimus) from heterologous host *S. coelicolor* (Jones *et al.*, 2013).

After insertion of the PACs into *S. erythraea* ISOM the resulting strains were grown under a range of media under different conditions and tested for compound production using LCMS methods. Of the 11 PACs (containing nine different BGCs) tested only *S. erythraea* ISOM/215G Δ *forJ*, encoding the formicamycin/fasamycin BGC, led to compound production, and in this case only when the gene *forJ* encoding for a biosynthetic transcriptional repressor (Devine *et al.*, 2021) was deleted from the BGC.

5.2 Objectives

The main objectives of the project were:

- Production of putative antifungal polyenes related to the known molecule nystatin by expression of the silent BGCs identified in two *Pseudonocardia* strains isolated from fungus farming ants
- Expression of cryptic and known BGCs in *S. erythraea* ISOM; these BGCs were cloned from a range of actinomycete strains under investigation in our lab
- Overexpression of putative activators of the antascomicin BGC to induce antascomicin production from the cryptic BGC in heterologous hosts *S. erythraea* ISOM and *S. coelicolor* M1146

5.3 Attempted Production of Putative Antifungal Polyenes in *S. erythraea* ISOM

5.3.1 Identification of Putative Antifungal Polyene BGCs in Ant Symbiotes by Genome Sequencing

Pseudonocardia symbionts of *Acromyrmex* ants, commonly called attine or leaf cutter ants, found in Gamboa, Panama were identified as belonging to two species, *Pseudonocardia octospinosus* (Ps1) and *Pseudonocardia echinator* (Ps2) (Holmes *et al.*, 2016). Ants from the same colony would have only one of either the Ps1 or Ps2 strains on their cuticle. Full genome sequencing of five Ps1 and Ps2 isolates each from 10 different ant colonies revealed that Ps1 and Ps2 have distinct BGC profiles, with each containing a different polyene-like BGC (Holmes *et al.*, 2016). This was of interest as polyenes are well known polyketide natural products with antifungal activity (Aldholmi *et al.*, 2019; Vicente *et al.*, 2003). A well know clinical example is nystatin which is used to treat *Candida* skin infections (**Fig 5.1**) (Falah-Tafti *et al.*, 2010; Lyu *et al.*, 2016).

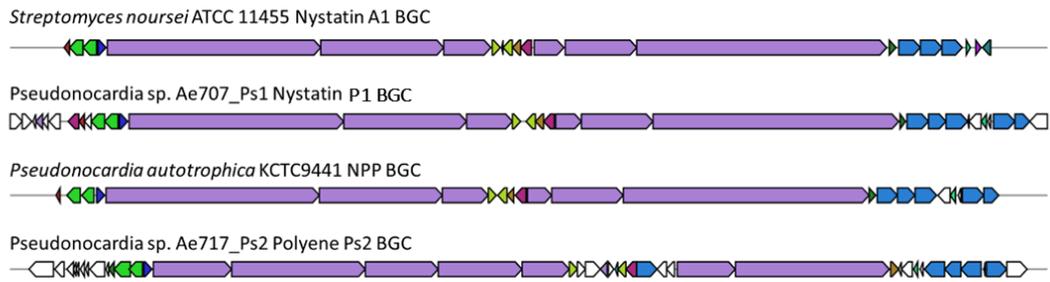
The polyene-like BGCs found in Ps1 strains are highly similar to the BGC from *Pseudonocardia octospinosus* isolated from another attine ant species and encoding for a molecule called nystatin P1 (Barke *et al.*, 2010), making it likely the BGCs encode for the same compound (Holmes *et al.*, 2016). Although the absolute chemical structure of nystatin P1 was not fully determined, LCMS/MS data suggested it consisted of a nystatin derivative with an additional deoxyhexose attached to the mycosamine sugar moiety. Based on the very high similarity of the two BGCs and the published LCMS data, it is highly likely that the products of nystatin P1 and the Ps1 BGCs are identical (**Fig 5.1**).

The polyketide synthase (PKS) genes encoded in the nystatin P1 and nystatin Ps1 BGCs show a high level of amino acid similarity with the nystatin-like polyene compound NPP produced by *Pseudonocardia autotrophica* KCTC9441, which is not an attine ant symbiont (Kim *et al.*, 2009) (**Fig 5.1**). The chemical structure of nystatin NPP has been fully determined and is identical to nystatin but with a D-glucosamine

sugar attached to the mycosamine moiety (**Fig 5.1**). The additional sugars make nystatin NPP and P1 more water soluble than nystatin itself. Consistent with these observations, the nystatin P1, nystatin Ps1 and NPP BGCs all contain additional glycosyl transferases compared to that of the original nystatin producer *Streptomyces noursei* ATCC (Fjaervik & Zotchev, 2005).

The putative polyene BGCs from Ps2 stains showed lower sequence conservation with the NPP and other nystatin BGCs (Holmes *et al.*, 2016) and lacks several modules compared to the nystatin Ps1 PKS, including the loading module, first two elongation modules and a module responsible for the incorporation of a fully reduced acetate unit. Analysis of the alternative loading module and other precursor supply genes in the BGC suggest that biosynthesis of the putative polyene product may be initiated by a 3-amino-5-hydroxybenzoate starter unit (**Fig 5.1**).

(A)



(B)

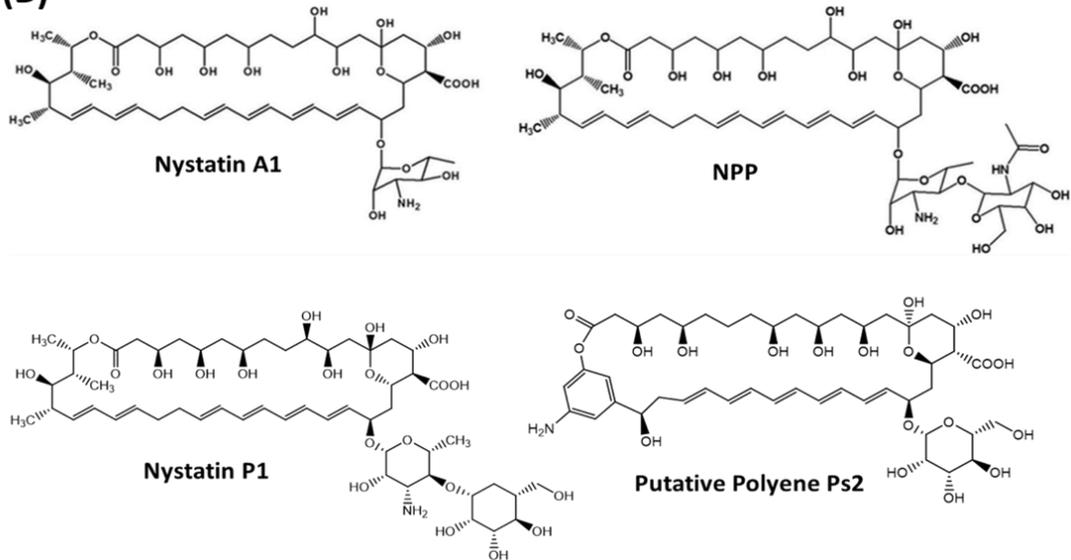


Figure 5.1: Comparison of Bacterial Gene Clusters and Chemical Structures of Nystatin A1, NPP and the Predicted Structure of Nystatin P1 and Putative Polyene Ps2. (A) BGC Alignment and comparison of Nystatin A1 from *Streptomyces noursei* ATCC 11455, Nystatin P1 from *Pseudonocardia* sp. Ae707_Ps1, NPP from *Pseudonocardia autotrophica* KCTC9441 and putative polyene Ps2 from *Pseudonocardia* sp. Ae717_Ps2. (B) Chemical structure of Nystatin A1, Nystatin P1, NPP and predicted chemical structure deduced from PKS assembly line and post-PKS tailoring steps (Holmes *et al.*, 2016) of putative polyene Ps2. Adapted from (Kim *et al.*, 2015), *PLOS ONE*, 10(4), e0123270, licensed under CC BY 4.0

5.3.2 *S. erythraea* ISOM as a Potential Heterologous Host for *Pseudonocardia* BGCs

Pseudonocardia Ps1 and Ps2 strains were difficult to cultivate under laboratory conditions and did not display antifungal activity against *Escovopsis weberi*, the pathogenic invader of leaf cutter ant colonies, as would be expected should the strains be producing nystatin-like antifungal compounds (Holmes, N., personal communication, 2018). PACs containing the polyene BGCs from Ps1 and Ps2 had previously been generated for us by Bio S&T Inc (Montreal, Canada). The PAC 6_19B contained the Ps1 lineage BGC and was cloned from strain Ae707_Ps1; and PAC 2_5K contained the BGC from Ps2 lineage and was cloned from strain Ae717_Ps2 (Holmes *et al.*, 2016).

S. erythraea is a bacterium in the Pseudonocardiaceae family along with *Pseudonocardia* and is therefore phylogenetically closer to *Pseudonocardia* than the more commonly used actinomycete heterologous host *S. coelicolor*. Consequently, we hypothesised that *S. erythraea* ISOM might successfully produce the putative nystatin-like compounds from the *Pseudonocardia* spp. BGCs where expression in *S. coelicolor* had previously failed (Holmes, N., personal communication, 2018).

5.3.3 Insertion and Production of Nystatin Ps1 and Putative Polyene Ps2 in *S. erythraea* ISOM

5.3.3.1 Conjugation of the Nystatin Ps1 and Putative Polyene Ps2 BGCs into *S. erythraea* ISOM

The PACs 6_19B and 2_5K, containing the polyene BGCs from Ps1 and Ps2 lineage respectively, were tri-parentally mated into methylation deficient *E. coli* ET12567 cells with use of a pR9604 plasmid, readying the PACs for intergenic transfer into *S. erythraea* ISOM (Flett *et al.*, 1997). Once the PACs had been conjugated into *S. erythraea* ISOM and ex-conjugates selected for by apramycin resistance, colony PCR was used to determine successful integration of the polyene BGCs of Ps1 and Ps2, with primer pairs annealing to the left (L), within the centre (C) and to the right (R) of the Ps1 BGC, and to the left (L), within the centre (C), on the right (R), and to the far-right (fR) of the Ps2 polyene BGC. Colony PCR was used instead of Sanger sequencing to assess conjugation success as the size of the BGCs made Sanger sequencing of the whole BGCs inefficient. Conjugation into *S. erythraea* ISOM was not very efficient, resulting in only two ex-conjugants for strain *S. erythraea* ISOM/6_19B and one ex-conjugant for strain *S. erythraea* ISOM/2_5K (**Fig 5.2**). However, this latter clone was not brought forward as the right-hand PCR product band of the 2_5K ex-conjugant was 300 bp larger than expected and the far-right band could not be amplified (**Fig 5.2B**), indicating a possible re-combination event or misalignment during insertion of the PAC into the *S. erythraea* chromosome.

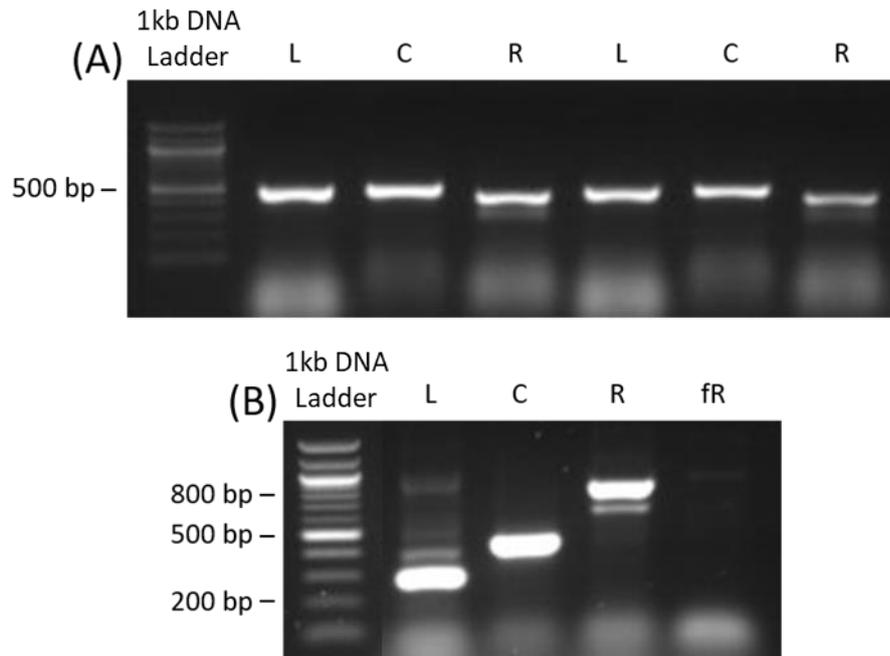


Figure 5.2: Colony PCR of *S. erythraea* ISOM ex-conjugants after Insertion of PACs 6_19B and 2_5K. PCR fragments were run on a 1% agarose gel with a 100 bp to 1 kb ladder. PAC 6_19B contains the nystatin Ps1 BGC, PAC 2_5K contains the putative polyene Ps2 BGC. Heterologous host was *S. erythraea* ISOM. (A) Two ex-conjugants of 6_19B, with bands of expect size to the left (L), within the centre (C) and to the right (R) of the nystatin Ps1 BGC. (B) One ex-conjugant of 2_5K, with the PCR fragment within the right of the putative polyene Ps2 BGC (R) 300 bp larger than expected, and the PCR fragment to the right of the putative polyene Ps2 BGC (fR) no longer visible. L = Left, C = Centre, R = Right, fR = Far Right of insert.

5.3.3.2 Production Media Assay to Induce Production of Nystatin Ps1 Polyene Product from *S. erythraea* ISOM/6_19B

A single ex-conjugant of *S. erythraea* ISOM/6_19B was brought forward and multiple fermentation media assessed to try and induce production of nystatin Ps1. Production assays were carried out with liquid media (SFM, ABBA13, TWM, YP, YPD, SV2) and on solid media plates (SFM, ABBA13, TWM, YP, YPD), with *S. erythraea* ISOM used as a negative control. Liquid media cultures were grown for 8 days at 30°C with shaking and extracts taken at day 6 and day 8. Solid media cultures were grown for 12 days at 30°C and extracts generated from approximately 5 mL of the plate. Samples were grown in duplicate and metabolites extracted using methanol.

Extracts were analysed using UHPLC-MS on a Shimadzu single quadrupole LCMS-2010A mass spectrometer and compared to a nystatin A1 standard (Merk) positive control showing the expected m/z peak of 926.5, with a standard polyene UV profile at 292, 305 and 320 nm (**Fig 5.3**). The putative product of the nystatin Ps1 BGC was expected to have a m/z of 1088.6 (the same as nystatin P1) with a UV absorbance similar to that of the nystatin A1 standard (Barke *et al.*, 2010). However, no peak of this m/z was observed for any of the fermentation conditions, nor were any other polyene UV absorption profiles present (data not shown)

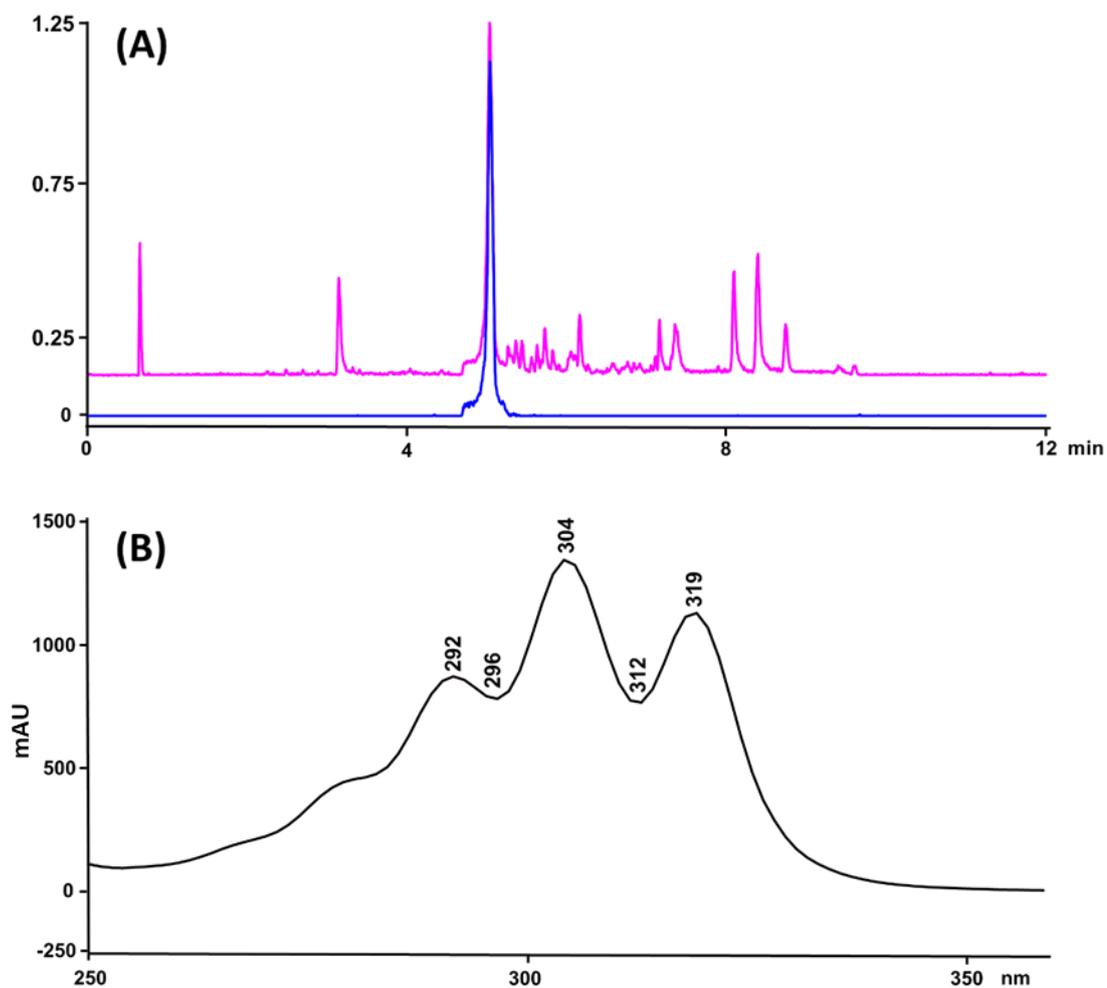


Figure 5.3: Extracted Ion Chromatogram (EIC) and UV Absorbance of Nystatin A1 standard. Nystatin A1 5 mg/mL 1 μ L analysed UHPLC-MS with mass fragment search. (A) Mass spectra of nystatin A1, Pink; Base Peak Chromatogram, Blue: nystatinA1 $m/z = 926.5$. EIC scaled 0 to 1.25, 10,000,000 intensity. Run time = 12 min. (B) UV absorbance of nystatin A1, peaks seen at 292, 304 and 319 nm. Equivalent to polyene UV profile of 292, 305, 320 nm. Scale -250 to 1500 milli-absorbance units (mAU).

5.3.4 Identification of a Frame Shift in the Nystatin Ps1 BGC within PAC 6_19B using Illumina Sequencing

As nystatin Ps1 was not produced within the heterologous host *S. erythraea* ISOM the PAC 6_19B was sent for Illumina sequencing to check for detrimental mutations. This identified two nucleotide changes within the nystatin Ps1 BGC: a cytosine (C) to thymine (T) point mutation and a double guanosine (G) deletion, present in a keto reductase and a keto synthase domain of the nystatin Ps1 PKS, respectively. Sanger sequencing of these regions from PAC 6_19B, genomic DNA of the parental strain Ae707_Ps1 and four other Ps1 strains (Ae150A_Ps1, Ae168_Ps1, Ae263_Ps1, Ae356_Ps1) (Holmes *et al.*, 2016), revealed that the C-to-T point mutation was a sequencing error but that the double guanosine deletion was present only within PAC 6_19B. The guanosine deletion was located within elongation module 1 of the nystatin Ps1 PKS (**Fig 5.4**) and led to a frame shift of all downstream codons. This would result in all further elongation modules of the PKS within the same open reading frame being non-functioning, likely causing premature termination of translation of the nystatin Ps1 PKS. As such production of nystatin Ps1 from PAC 6_19B was considered unobtainable.

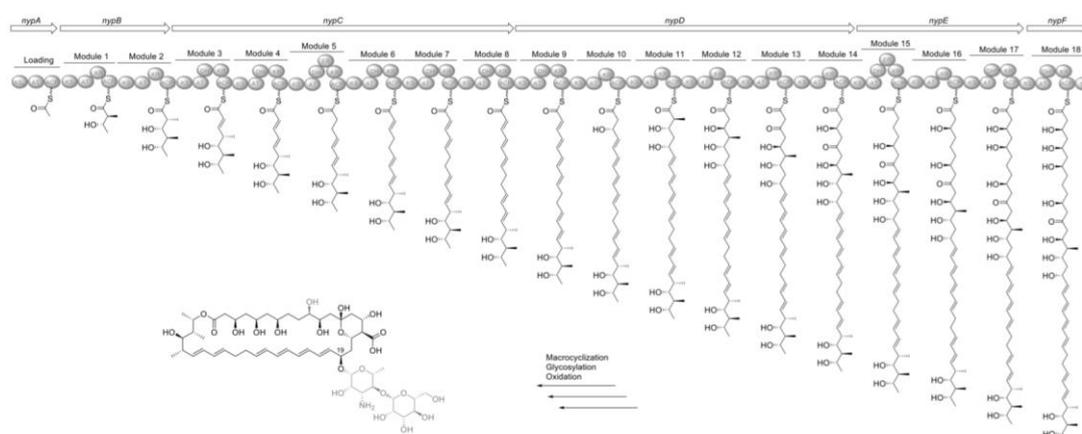


Figure 5.4: Genetic Organization and Model Biosynthesis of the Nystatin P1 from Ae707_Ps1. Biosynthesis genes *nysA* to *nysF* encoding for the loading module and all 18 elongations modules. ACP, acyl carrier protein; AT, acyl transferase; DH, dehydratase; ER, enoylreductase; KR, ketoreductase, KS, ketosynthase; KSS, *nys*-like loading KS; TE, thioesterase. Proposed structures of the mature compound pictured in gray are due to post-PKS tailoring steps. Used with permission from (Holmes *et al.*, 2016), *Frontiers in Microbiology*, 7, p. 2073, licensed under CC BY 4.0

5.4 Insertion and Expression of Actinomycetes NP BGCs in *S. erythraea* ISOM

To further explore *S. erythraea* ISOM as a potential heterologous host several other PACs containing BGCs originating from actinomycete strains deriving from other projects in our lab were investigated (**Table 5.1**). As in section 5.3.3.2 above a range of different liquid media were used in fermentation of the resulting strains to try and induce NP production.

5.4.1 PACs containing Actinomycetes BGCs

PACs containing BGCs of interest from a range of actinomycete species were obtained from colleagues in the Molecular Microbiology Department at the John Innes Centre (**Table 5.1**). Previously, anthracimycin, encoded by PAC 14M, was successfully produced in heterologous host *S. coelicolor* M1146, M1152 and M1154 (Alt & Wilkinson, 2015); antascomycin was produced from parental strain *Micromonospora* sp. A92-306401, but only in minute amounts not suitable for further study (Fehr *et al.*, 1996); and fasamycins and formicamycins were produced from parental strain *Streptomyces formica* KY5, and in greater quantities from *S. formicae* KY5 Δ forJ (Devine *et al.*, 2021; Qin *et al.*, 2017). All other BGCs investigated were identified through genetic sequencing efforts and their respective NP were not identified from the parental strains after fermentation under laboratory conditions or after heterologous expression of the relevant PACs in *S. coelicolor* (apart from the formicamycin BGC which did express in this strain, see section 5.4.3).

Table 5.1: PACs, Associated NPs and the Parental Strains of BGCs Inserted into Heterologous Host *S. erythraea* ISOM

PAC	BGC/Natural Product	Parental Strain	Reference
14M	Polyketide/ Anthracimycin	<i>Streptomyces</i> sp. T676	(Alt & Wilkinson, 2015)
7G	Thiopeptide/ unknown	<i>Streptomyces</i> sp. T676	Elena Stoyanova, JIC
24J	Lantipeptide/unknown	<i>Streptomyces</i> sp. T676	Charles Owen, JIC
TB_73	Macrolactam/unknown	<i>Streptomyces</i> sp. T676	Thomas Booth, JIC
3D	NRPS/unknown	<i>Streptomyces</i> sp. T676	Natalia Miguel Vior, JIC
21A 21B	Macrolide/ Antascomycin	<i>Micromonospora</i> sp. A92-306401	(Fehr <i>et al.</i> , 1996), Qin Zhiwei, JIC
215G 215G Δ forJ	Polyketide/ Fasamycins and Formicamycins	<i>Streptomyces</i> <i>formicae</i> KY3	(Devine <i>et al.</i> , 2021; Qin <i>et al.</i> , 2017)

5.4.2 Production Media Assays of NP BGCs Inserted into *S. erythraea* ISOM

Ex-conjugants of each PAC integrated into *S. erythraea* ISOM were grown in three different liquid production medias to try and induce NP production (**Table 5.2**). SV2 was used as a standard production media and GPP was a rich production media that *S. erythraea* ISOM had previously shown strong growth in. The third production media was selected based on NP production previously seen in parental strains or other heterologous host strains. ABBA13 media, which *S. erythraea* ISOM could readily sporulate on (media recipes, Chapter 2, section 2.3), was used if no NP production had been previously seen. Samples were incubated for 8 days at 30°C with shaking and metabolite extracts generated using methanol at day 6 and day 8, the assays were carried out in duplicate and *S. erythraea* ISOM used as a negative control.

Table 5.2: Liquid production media assays for PACs in *S. erythraea* ISOM

PAC	Production Media
14M	SV2, GPP, A1
7G	SV2, GPP, ABBA13
24J	SV2, GPP, ABBA13
TB_73	SV2, GPP, ABBA13
3D	SV2, GPP, ABBA13
21A 21B	SV2, GPP, MM
215G 215G Δ forJ	SV2, GPP, SFM

5.4.3 Mass Spectrometry Analysis of *S. erythraea* ISOM PAC Extracts Generated from Production Media Assays

Extracts from the *S. erythraea* ISOM PAC production media assays were analysed using UHPLC-MS using a Shimadzu single quadrupole LCMS-2010A mass spectrometer to identify NPs produced. Mass spectra of the extracts were compared to extracts generated from the *S. erythraea* ISOM negative control. If a m/z value was known for the NP associated with the BGC of interest within the PAC, this was also searched for. Apart from strain *S. erythraea* ISOM/215G Δ forJ which produced several fasamycins, no other PAC expression strains produced any identifiable NPs (data not shown). *S. erythraea* ISOM/215G Δ forJ produced fasamycin C and small amounts of fasamycin D in all three production medias assessed (fasamycin C; m/z known = 473.1595, m/z observed = 473.160, Δ = 1.1 ppm; fasamycin D; m/z = 507.1205, observed m/z = 507.1230, Δ = 4.9 ppm.) (**Fig. 5.5**) (Appendix 9.3). The PAC 215G Δ forJ is a derivative of PAC 215G and contains the *S. formicae* KY5 formicamycin BGC, but with the transcriptional repressor gene *forJ* replaced with an apramycin resistance marker (Devine *et al.*, 2021). Deletion of *ForJ* in *S. formicae* KY5 was previously seen to cause over production of formicamycins and fasamycins on solid media and induce production in liquid media, where no production was previously observed (Devine *et al.*, 2021). Along with previous results the successful production of fasamycins from strain *S. erythraea* ISOM/215G Δ forJ, but not from strain *S. erythraea* ISOM/215G shows the impact the *ForJ* transcriptional repressor has on the production profile of the BGC. It is notable that no formicamycins were produced in these experiments, and this result will be examined further in Chapter 6 of this thesis

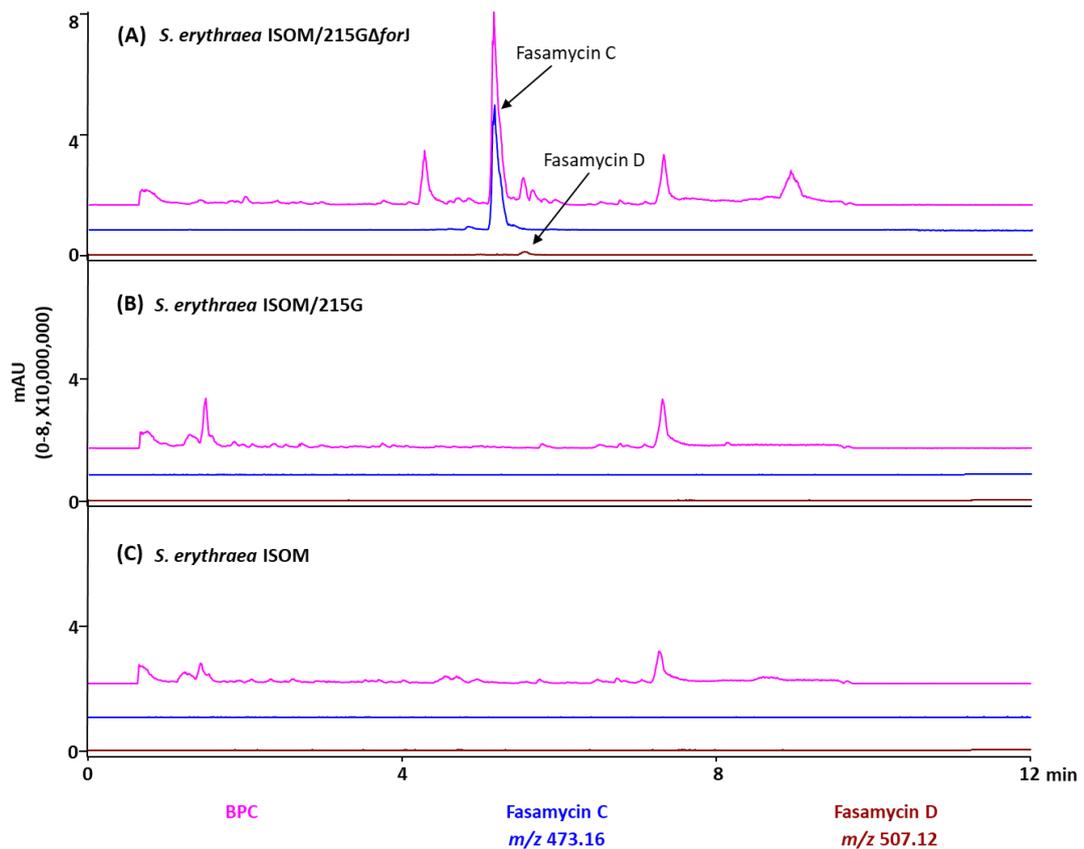


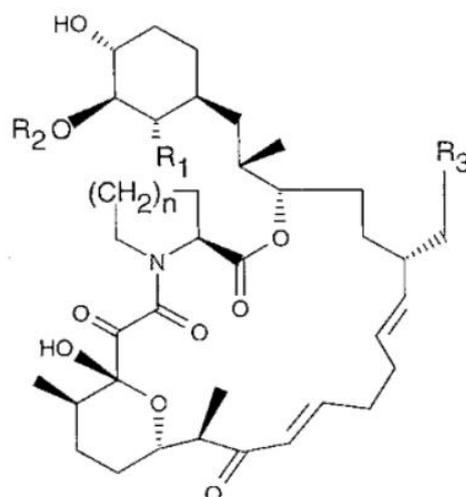
Figure 5.5: Extracted Ion Chromatograms (EIC) of *S. erythraea* ISOM/215G and *S. erythraea* ISOM/215G Δ forJ SV2 Liquid Media Extracts. Extracts taken at day 8, analysed by UHPLC-MS, *S. erythraea* ISOM extracts used as a negative control. Formicamycins and fasamycin compounds searched for using known m/z values (Qin *et al.*, 2017). Pink = Base Peak Chromatogram (BPC), Blue = Fasamycin C m/z 473.16, Brown = Fasamycin D m/z 507.12. (A) *S. erythraea* ISOM/215G Δ forJ, producing fasamycin C (m/z known = 473.1595, m/z observed = 473.160, Δ = 1.0567 ppm) and fasamycin D (m/z = 507.1205, observed m/z = 507.123, Δ = 4.9298 ppm). (B) *S. erythraea* ISOM/215G, EIC equivalent to negative control with no observable fasamycins. (C) *S. erythraea* ISOM negative control. All EIC scaled from 0 to 8, x10,000,000 intensity. Run = 12 min.

5.5 Overexpression of Transcriptional Activators for Induction of Antascomycin Production from PAC 21A and PAC 21B

As seen by the results above in 5.4.3 the removal of pathway specific repressors from BGCs can induce NP production within the parental strain and heterologous hosts. Similarly, the expression of pathway specific transcriptional activators under control of constitutive promoters can also lead to over production of NPs (Aigle & Corre, 2012; Guo *et al*, 2015; Reen *et al.*, 2015; Ren *et al*, 2017), such as when the overexpression of the LuxR transcriptional activator gene *nam1* in *Streptomyces* sp. LZ35 led to production of three novel naphthalenic octaketide ansamycins from the putative cryptic anasamycin BGC *nam* (Li *et al*, 2015). Based on these previous results, we decided to overexpress three potential transcriptional activators identified in the antascomycin BGC in an attempt to induce production of antascomycin in heterologous hosts *S. erythraea* ISOM and *S coelicolor* M1146.

5.5.1 Antascomycin Transcriptional Activators

Antascomycin (**Fig 5.6**), produced in minute amounts from *Micromonospora* sp. A92-306401 (Fehr *et al.*, 1996), is a macrolide polyketide antibiotic capable of binding to the abundant intracellular binding protein FKBP12, which is known to act as the receptor for the immunosuppressant drug FK506 (Tong & Jiang, 2015). However, unlike other FKBP12 binding macrolides, such as rapamycin and ascomycin, antascomycin does not show immunosuppressive activity (Fehr *et al.*, 1996), making it of great interest for its potential clinical use as an antimicrobial without immunosuppressive side effects. Three potential transcriptional activators of the antascomycin BGC were previously identified by genetic sequencing and annotation, these were a SARP, a LuxR and an AraC gene (**Fig 5.7**) (data not published, Qin Zhiwei, JIC).



Antascomycin A	R ₁ =H	R ₂ =H	R ₃ =H	n=2
Antascomycin B	R ₁ =OH	R ₂ =H	R ₃ =H	n=2
Antascomycin C	R ₁ =OH	R ₂ =CH ₃	R ₃ =H	n=2
Antascomycin D	R ₁ =H	R ₂ =H	R ₃ =H	n=1
Antascomycin E	R ₁ =H	R ₂ =H	R ₃ =OH	n=2

Figure 5.6: Chemical Structure of Antascomicins, Reproduced from (Fehr *et al.*, 1996), *Journal of Antibiotics*, 49(3), p. 230-233, licensed under CC BY 4.0.

Streptomyces Antibiotic Regulatory Proteins (SARPs), are a family of transcriptional activator proteins characterized by an N-terminal OmpR-type winged helix-turn-helix DNA-binding domain (Wietzorrek & Bibb, 1997). This domain recognises and binds to direct heptameric repeats (TCGAGXX) spaced by 4 or 15 nt apart and located 8 nt upstream of a promoter (Arias *et al.*, 1999; Sheldon *et al.*, 2002). LuxR transcriptional regulators work in a similar manner with their own characteristic helix-turn-helix DNA binding domain located in the C-terminal. They can be activated through several different mechanisms, the most common of which is being phosphorylated by a transmembrane kinase in a two-component sensory transduction system. LuxR transcriptional regulators most often act as transcriptional activators but can also act as repressors or have a dual role, depending on conditions and the BGC (Birck *et al.*, 2002; Maris *et al.*, 2002). AraC transcriptional regulators have a 99 amino acid conserved region, usually within the

coelicolor, and to account for the possibility of multiple transcriptional activators being needed to successfully activate the antascomycin BGC, I attempted to insert the transcriptional activators in combination in heterologous hosts *S. erythraea* ISOM/21A, *S. erythraea* ISOM/21B, *S. coelicolor* M1146/21A and *S. coelicolor* M1146/21B.

5.5.1.1 Cloning of Antascomycin Transcriptional Activators into Integration Plasmids for Pair-Wise Conjugation in *S. erythraea* ISOM

To begin with, PCR fragments for the genes encoding the three potential antascomycin transcriptional activators (SARP, LuxR and AraC) were generated from the genomic DNA of *Micromonospora sp.* A92-306401 and inserted via ligation using *NdeI/HindIII* sites into expression plasmid pGP9, which utilises a ϕ BT1 integrase site to specifically integrate into the chromosome (Andexer *et al.*, 2011), generating plasmids; pGP9_SARP, pGP9_LuxR and pGP9_AraC (Qin Zhiwei, JIC).

S. erythraea ISOM contains several plasmid integration sites and it is possible to carry out pair-wise conjugation of integration plasmids into the chromosome. Personal communications with Isomerase Therapeutics informed us that vector candidates pBF3 (Appendix 9.5) (Fayed *et al.*, 2014) and pRF10 (Appendix 9.5) (Fayed *et al.*, 2015), which contained the phage-1 integrate sites SV1 and TG1, respectively, were effective for use with this strain. However, both vectors had the same antibiotic resistance marker (hygromycin), making it impossible to identify through antibiotic resistance assays whether both plasmids had integrated into the chromosome successfully. Thus, an antibiotic resistance marker for tetracycline was cloned into the integration vector pBF3 and the hygromycin resistance gene removed. This was accomplished by generating a PCR fragment of the tetracycline resistance genes *tetA* and *tetR* from the *Pseudomonas* suicide vector pTS1 (Thomas Scott, JIC), with 30 bp either side of a unique *SacI* site in the hygromycin resistance gene of pBF3. Using Gibson assembly, the tetracycline PCR DNA fragment was then inserted into the linearized pBF3, conferring tetracycline resistance and disrupting the hygromycin resistance gene, generating integration vector pBF3TetR.

A PCR fragment for the LuxR encoding gene was created using pGP9_LuxR as a template and inserted into vector pRF10 by Gibson assembly using a unique *Bgl*II site; giving plasmid pRF10_LuxR. A PCR fragment for the AraC encoding gene was created using pGP9_AraC as a template and inserted into pBF3TetR by Gibson assembly using a unique *Xba*I site; giving plasmid pBF3TetR_AraC. Transformation of empty vector controls revealed that methylation deficient *E. coli* ET12567/pUZ8002 cells, used for conjugation into *S. erythraea* ISOM, contained a tetracycline resistance gene, so were unsuitable for cloning of the pBF3TetR plasmid. Thus, *E. coli* C2925/pR9604 cells, another suitable methylation deficient *E. coli* strain were used for conjugation of pBF3TetR_AraC and the pBF3TetR empty vector control into *S. erythraea* ISOM.

5.5.2 Conjugation of Antascomicin Transcriptional Activators into Heterologous Hosts *S. erythraea* ISOM and *S. coelicolor* M1146

5.5.2.1 Sequential Conjugation of Antascomicin Transcriptional Activators into *S. coelicolor* M1146

The antascomicin transcriptional activators pGP9_SARP, pRF10_LuxR and pBF3TetR_AraC were conjugated into strains *S. coelicolor* M1146/21A and *S. coelicolor* M1146/21B, which were created through conjugation of the two antascomicin PACs, 21A and 21B, into heterologous host *S. coelicolor* M1146. The transcriptional activators were inserted into the strains consecutively and all possible combinations were achieved (Chapter 2.2.1, **Table 2.5**). The following strains were created from M1146/21A; M1146/21A/AraC, M1146/21A/LuxR, M1146/21A/SARP, M1146/21A/AraC/LuxR, M1146/21A/AraC/SARP, M1146/21A/LuxR/SARP, M1156/21A/AraC/LuxR/SARP. It was observed that insertion of pBF3TetR_AraC resulted in a phenotypic change, causing slower growth and smaller colonies.

5.5.2.2 Attempted Pair-Wise Conjugation of Antascomicin Transcriptional Activators into *S. erythraea* ISOM

We initially attempted to introduce the transcriptional activator containing plasmids into *S. erythraea* ISOM in a pair-wise fashion using conjugation; communication from the Isomerase suppliers suggested this should be possible. However, in our hands the efficiency of conjugation was not good enough for this to occur, and overall the conjugation efficiency was lower than that for *S. coelicolor* M1146. Therefore, conjugation of the antascomicin transcriptional activators was carried out individually. However, the insertion of any of the plasmid into the genome resulted in a phenotypic change abolishing sporulation and causing slower growth. No ex-conjugants could be generated from attempts to conjugate with *S. erythraea* ISOM mycelium stocks, thus the antascomicin transcriptional activators could only be conjugated singly into strains *S. erythraea* ISOM/21A and *S. erythraea* ISOM/21B (Chapter 2, section 2.2.1, **Table 2.5**). The following strains were created

from *S. erythraea* ISOM/21A; *S. erythraea* ISOM/21A/AraC, *S. erythraea* ISOM/21A/LuxR, *S. erythraea* ISOM/21A/SARP.

5.5.3 Antascomycin Production Media Assays with Overexpression of Transcriptional Activators

The *S. coelicolor* M1146 strains containing the antascomycin PACs and transcriptional activators in all possible combinations and the *S. erythraea* ISOM strains containing the antascomycin PACs and a single transcriptional activator were grown in SV2, GPP and MM liquid production media for 8 days at 30°C with shaking to try and induce production of antascomycin. Metabolic extracts were generated with methanol at days 6 and day 8 and extracts from the parent strains *S. erythraea* ISOM and *S. coelicolor* M1146 used as negative controls.

5.5.4 Mass Spectra Analysis of Extracts Generated from Overexpression of Antascomycin Transcriptional Activators in Heterologous Hosts

The extracts were analysed using UHPLC-MS on a Shimadzu single quadrupole LCMS-2010A mass spectrometer and the mass spectra compared to the *S. coelicolor* M1146 and *S. erythraea* ISOM negative controls. The known m/z values for the antascomicins were also searched for; antascomycin A $m/z = 659.4028$; antascomycin B+E $m/z = 675.3977$; antascomycin C $m/z = 689.4133$; antascomycin D $m/z = 645.3871$ (Fehr *et al.*, 1996). No antascomicins, potential pathway intermediates, or variant products were identified, though differences in mass spectra between extracts from strains containing the antascomycin BGC, strains overexpressing the transcriptional activators and the heterologous host negative controls were observed for both the *S. coelicolor* M1146 (**Fig 5.8**) and *S. erythraea* ISOM strains (**Fig 5.9**).

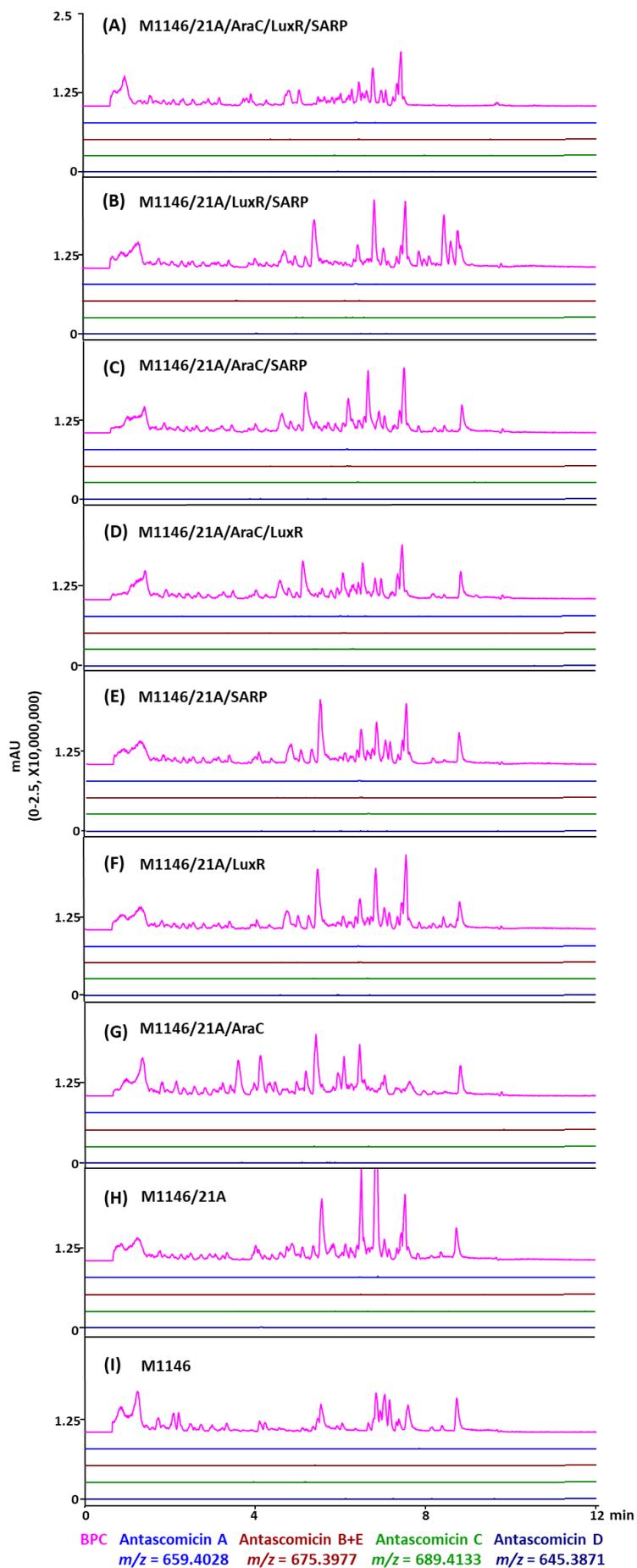


Figure 5.8: Extracted Ion Chromatograms (EIC) of *S. coelicolor* M1146/21A with the Antascomicin BGC and Transcriptional Activators. Samples grown in SV2 media, extracts generated at day 8, analysed using UHPLC-MS and mass fragment search, *S. coelicolor* M1146 was used as a negative control. No antascomicins, biosynthetic intermediates or variant products were identified. BPC (Pink) = Base Peak Chromatogram. Extracted Ion chromatograms are shown for: Antascomicin A (Light Blue) $m/z = 659.4028$; Antascomicin B+E (Brown) $m/z = 675.3977$; Antascomicin C (Green) $m/z = 689.4133$; Antascomicin D (Dark Blue) $m/z = 645.3871$. (A) Strain M1146/21A/AraC/LuxR/SARP, (B) Strain M1146/21A/LuxR/SARP, (C) Strain M1146/21A/AraC/SARP, (D) Strain M1146/21A/LuxR/SARP, (E) Strain M1146/21A/SARP, (F) Strain M1146/21A/LuxR, (G) Strain M1146/21A/AraC, (H) Strain M1146/21A, (I) Strain M1146 negative control. All EIC scaled from 0 to 2.5, $\times 10,000,000$ intensity. Run = 12 min.

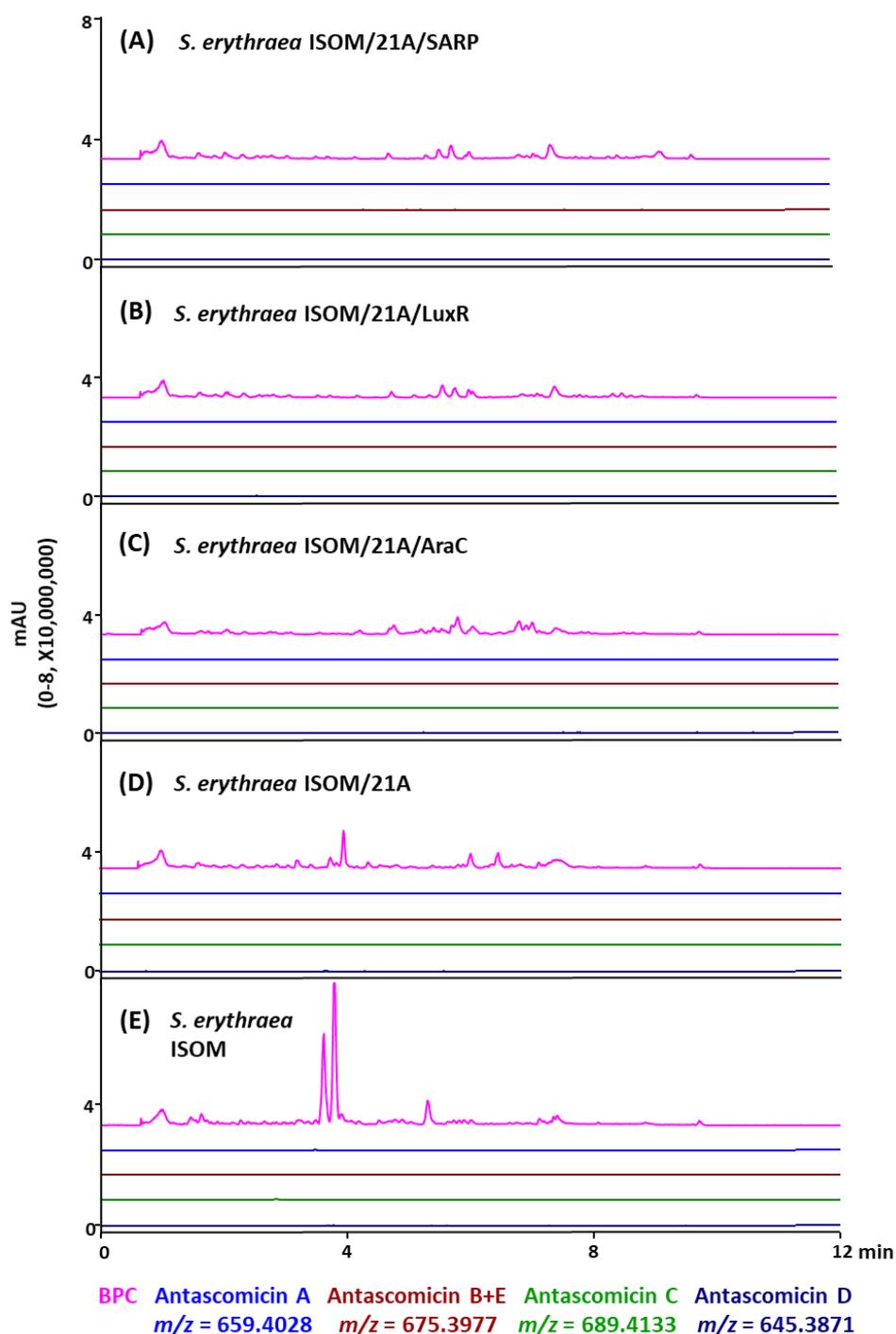


Figure 5.9: Extracted Ion Chromatograms (EIC) of *S. erythraea* ISOM with Antascomycin PAC 21A and Transcriptional Activators. Samples grown in SV2 media, extracts generated at day 8, analysed using UHPLC-MS and mass fragment search, *S. erythraea* ISOM was used as a negative control. No antascomicins, biosynthetic intermediates or variant products were identified. BPC (Pink) = Base Peak Chromatogram. Extracted ion chromatograms are shown for: Antascomycin A (Light Blue) $m/z = 659.4028$; Antascomycin B+E (Brown) $m/z = 675.3977$; Antascomycin C (Green) $m/z = 689.4133$; Antascomycin D (Dark Blue) $m/z = 645.3871$. (A) Strain *S. erythraea* ISOM/21A/SARP, (B) Strain *S. erythraea* ISOM/21A/LuxR, (C) Strain *S. erythraea* ISOM/21A/AraC, (D) Strain *S. erythraea* ISOM/21A, (E) Strain *S. erythraea* ISOM negative control. All EIC scaled from 0 to 8, x10,000,000 intensity. Run = 12 min.

5.6 Discussion

Several PACs containing BGCs of interest cloned from a range of actinomycetes were conjugated into the alternative heterologous host *S. erythraea* ISOM. This strain is derived from a parent that has been selected in an industrial setting for high erythromycin production and good growth in fermenters. Further, *Saccharopolyspora* is a distant relative of *Streptomyces* and thus we hoped might provide an alternative regulatory background for the expression of heterologous BGCs. Thus, we hoped that BGCs which had remained silent in the well-studied host *S. coelicolor* might be expressed and produce new metabolites. In addition, we were interested to see how BGCs known to be expressed in their native context, or heterologously in *S. coelicolor* might behave in this strain.

Unfortunately, only strains of *S. erythraea* ISOM conjugated with a PAC containing the formicamycin BGC (*S. erythraea* ISOM/215G Δ forJ) produced the expected products. Unfortunately, PAC 6_19B, containing the nystatin Ps1 BGC that originated from *Pseudonocardia octospinosus* (Ps1), could not be assessed for NP production within *S. erythraea* ISOM due to a two-nucleotide deletion causing a frame shift in module 1 of the PKS. It was hoped that *S. erythraea* ISOM would be a suitable host for production of the polyene BGCs from Ps1 and Ps2 (which could also not be successfully conjugated into the host) as it is more closely related to the *Pseudonocardia* parental strains than *Streptomyces* sp., with both strains being part of the Pseudonocardiaceae family, a class of actinobacteria. Thus *S. erythraea* ISOM could not be assessed for its suitability as a heterologous host for BGCs originating from strains of the Pseudonocardiaceae family. Moreover, heterologous expression of the anthracimycin BGC from *S. erythraea* ISOM was not successful, whereas it had been from *S. coelicolor*. However, heterologous host *S. erythraea* ISOM but it was useful as a host for heterologous expression of fasamycin/formicamycins, with several new fasamycin like compounds produced, discussed further in Chapter 6. Like *S. coelicolor*, *S. erythraea* ISOM was not an effective host for heterologous expression of the other BGCs tested.

The PAC, 215G Δ *forJ*, that led to successful production of NPs within *S. erythraea* ISOM will be investigated in further detail in Chapter 6 of this study. However, it should be noted that the BGC contained on this PAC also produced NP within heterologous host *S. coelicolor* M1146 and within parental strain *S. formicae* KY5 (**Fig 6.4**) (Devine *et al.*, 2021). Considering that insertion of PAC 215G, from which PAC 215G Δ *forJ* is derived, into *S. erythraea* ISOM and *S. coelicolor* M1146 did not induce production of the fasamycin and/or formicamycin NPs it can be assumed that in this case the critical factor affecting compound production is not the host species, but the removal of the gene encoding the ForJ repressor.

Furthermore, it was observed that fasamycin and formicamycins from *S. erythraea* ISOM/215G Δ *forJ* could only be produced consistently on solid media, with liquid media production being inconsistent (data not shown). It has been generally observed that changes to growth conditions, such as between liquid and solid media, can result in different secondary metabolite profiles of actinomycetes (Gebreyohannes *et al.*, 2013). In the case of the formicamycins and fasamycins, the parent organism only produces these NPs on solid agar unless the repressor gene *forJ* is deleted (Devine *et al.*, 2021). Considering this, further studies using the *S. erythraea* ISOM for NP production should also include trialling different solid production media as well as different liquid production media.

When investigating the possibility of inducing antascomycin production from the PACs 21A and 21B, overexpression of all three identified transcriptional activators (SARP, LuxR and AraC) singularly, in pairs and in triplicate, provided no benefit when tested in *S. coelicolor*. Overexpression of all three transcriptional activators could not be achieved in *S. erythraea* due to the inability to conjugate in multiple expression plasmids containing the transcriptional activators, due to insertion of a single transcriptional activator causing a phenotypic change abolishing sporulation. To overcome this limitation all three transcriptional activators could be included on a single expression plasmid under control of a constitutive promoter.

Over the course of this investigation several observations were made on the suitability of *S. erythraea* ISOM as a heterologous host of actinomycetes BGCs. This suitability was judged in comparison to the common heterologous host *S. coelicolor*, which has also been extensively used in this study. Firstly, the efficiency of the conjugation into *S. erythraea* was low, with only one or two ex-conjugants being produced in some instances from 100 % volume of the spore/cell conjugation mixture. When compared to *S. coelicolor*, from which we would expect a minimum of 20+ colonies from only 10 % volume of a comparative spore/cell conjugation mixture, this difference is notable. Secondly, the extent of false positive ex-conjugants was high, with many ex-conjugants growing well under antibacterial selection but not containing the BGC insert, as assessed through colony PCR. Lastly, the growing period of *S. erythraea* ISOM after conjugation was much longer than expected, taking up to two weeks for ex-conjugants to appear, and longer for sporulation to occur.

For future work trialling heterologous strains for expression of BGCs more varied production media should be examined, in both liquid and solid form. More varied growth conditions should be assessed, such as temperature and pH. Genetic manipulation of the BGCs via deletion of path-specific repressors and/or the overexpression of transcriptional activators should also be examined, as this has previously been shown to induce production of several different BGCs (Aigle & Corre, 2012; Guo *et al.*, 2015; Reen *et al.*, 2015; Ren *et al.*, 2017). This includes the *for* BGC examined in this project, with fasamycins and formicamycins being produced from PAC 215G Δ *for*J in heterologous hosts but not from PAC 215G; this is discussed further and in greater detail in Chapter 6 of this work. Finally, codon optimisation, promoter and terminator manipulation and RBS adjustments could be utilised to make the BGCs more suitable for heterologous host production. For example it was reported that RBS adjustment could lead to a 2-3 fold increase in *gusA* reporter gene expression in the common heterologous host *Streptomyces lividans* (Elena *et al.*, 2014; Horbal *et al.*, 2018; Myronovskyi & Luzhetskyy, 2019; Zhou *et al.*, 2019).

Chapter 6:

Heterologous

Production of

Fasamycins and

Formicamycins

6. Comparison of Heterologous Hosts for Production of Formicamycins and Fasamycins

6.1 Introduction

As reported in Chapter 5, *S. erythraea* ISOM/215G Δ forJ produced predominantly fasamycin C in SV2, GPP and SFM liquid media. Previously, fasamycins and formicamycins were also produced in liquid media from the strain *S. formicae* KY5 Δ forJ (Devine *et al.*, 2021) and inconsistently on solid media from the heterologous host strain *S. coelicolor* M1146/215G Δ forJ (this work, data not shown). In this chapter I investigated the impact on the titre and production profile of formicamycins and fasamycins when produced from these three different strains.

Fasamycins and formicamycins are pentacyclic type 2 polyketides produced from *Streptomyces formicae* KY5, isolated from the cuticle of Kenyan *Tetraponera* leafcutter ants. They display potent activity against methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin resistant *enterococci* (VRE), with formicamycin J having a minimum inhibitory concentration of 0.625 and 1.25 μ M, respectively. (Qin *et al.*, 2017). As discussed in Chapter 1.4.3.1, the *for* BGC contains 24 genes of known or speculative function, with two MarR family transcriptional regulators, one of which (ForJ) represses the majority of the biosynthetic genes. Deletion of *forJ* gene led to overproduction of fasamycins and formicamycins in strain *S. formicae* KY5 Δ forJ and induced production of the NPs to occur in liquid media, whereas previously production from *S. formicae* KY5 had only been observed on solid media (Devine *et al.*, 2021). It was also observed previously in Chapter 5 that deletion of *forJ* from the *for* BGC was essential to induce production of fasamycins and formicamycins in the heterologous hosts *S. erythraea* ISOM and *S. coelicolor* M1146.

The genes ForABC comprise the minimal PKS, of which the tridecaketide product goes through a series of tailoring steps, firstly producing fasamycins by cyclisation and dehydration (ForD, ForL, ForR). Fasamycins can then undergo ForV catalysed halogenation, this acts as a gatekeeper step for formicamycin formation, with

chlorination of the fasamycins intermediates being necessary for the catalytic activity of enzymes ForX and ForY, which convert the molecule to a formicamycin. Formicamycin formations occurs through hydroxylation and ring expansion of the fasamycin intermediate by ForX, leading to a lactone intermediate, this intermediate is then acted upon by the flavin dependant oxidoreductase ForY, which then converts it into a formicamycin by catalysing a reductive ring contraction (**Fig 6.1**) (Devine *et al.*, 2021). Once the formicamycin skeleton is formed the flavin dependant halogenase ForV can perform several rounds of halogenation on different carbon atoms to make a number of different formicamycins (**Fig 6.2**) (Devine *et al.*, 2021; Qin *et al.*, 2019b).

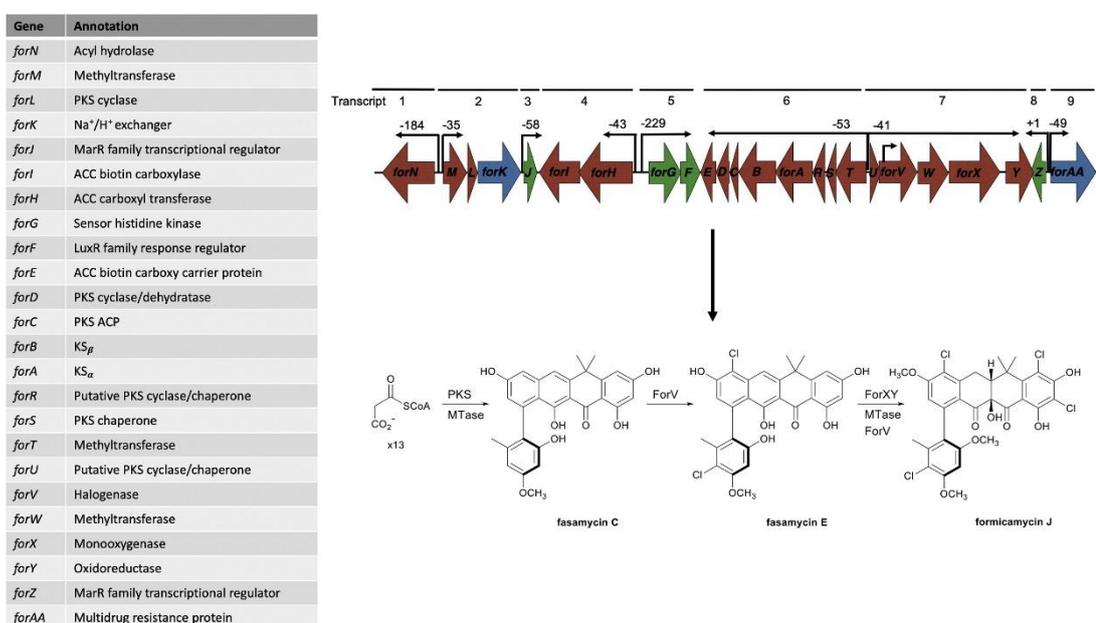
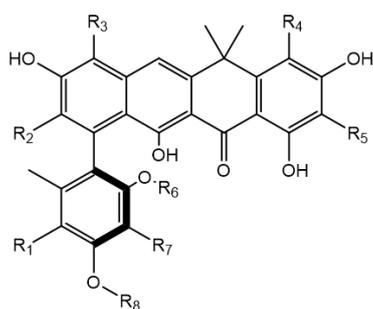
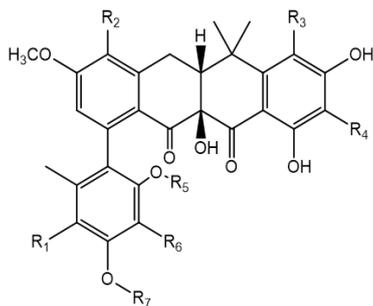


Figure 6.1: Formicamycin BGC, Gene Annotation and Predicted Post-PKS Tailoring Steps.

The *for* BGC comprises 24 genes on nine transcripts; red=biosynthetic genes, blue=transporters, green=regulatory genes. 10 transcription start sites were identified through cappable RNA sequencing. The table lists all genes with known/speculative function. Formicamycins are formed from fasamycins by ForX catalysed hydroxylation and ring expansion leading to a lactone intermediate. Reductive ring contraction is then catalysed by the flavin-dependant oxidoreductase ForY to form the formicamycin backbone. Reproduced from (Devine *et al.*, 2021), *Cell Chemical Biology*, 28(4), p. 515-523, licensed under CC BY 4.0.



Fasamycin C R₁ = H, R₂ = H, R₃ = H, R₄ = H, R₅ = H, R₆ = H, R₇ = H, R₈ = CH₃
Fasamycin D R₁ = H, R₂ = H, R₃ = Cl, R₄ = H, R₅ = H, R₆ = H, R₇ = H, R₈ = CH₃
Fasamycin E R₁ = Cl, R₂ = H, R₃ = Cl, R₄ = H, R₅ = H, R₆ = H, R₇ = H, R₈ = CH₃
Fasamycin F R₁ = H, R₂ = COOH, R₃ = H, R₄ = H, R₅ = H, R₆ = H, R₇ = H, R₈ = H
Fasamycin L R₁ = Cl, R₂ = H, R₃ = Cl, R₄ = H, R₅ = Cl, R₆ = H, R₇ = H, R₈ = CH₃
Fasamycin M R₁ = Cl, R₂ = H, R₃ = Cl, R₄ = H, R₅ = H, R₆ = CH₃, R₇ = Cl, R₈ = H
Fasamycin N R₁ = Cl, R₂ = H, R₃ = Cl, R₄ = Cl, R₅ = H, R₆ = H, R₇ = H, R₈ = CH₃
Fasamycin O R₁ = Cl, R₂ = H, R₃ = Cl, R₄ = H, R₅ = Cl, R₆ = CH₃, R₇ = H, R₈ = CH₃
Fasamycin P R₁ = Cl, R₂ = H, R₃ = Cl, R₄ = Cl, R₅ = H, R₆ = CH₃, R₇ = Cl, R₈ = H
Fasamycin Q R₁ = Cl, R₂ = H, R₃ = Cl, R₄ = Cl, R₅ = Cl, R₆ = H, R₇ = H, R₈ = CH₃



Formicamycin A R₁ = H, R₂ = Cl, R₃ = H, R₄ = H, R₅ = CH₃, R₆ = H, R₇ = CH₃
Formicamycin B R₁ = Cl, R₂ = Cl, R₃ = H, R₄ = H, R₅ = H, R₆ = H, R₇ = CH₃
Formicamycin C R₁ = H, R₂ = Cl, R₃ = Cl, R₄ = H, R₅ = CH₃, R₆ = H, R₇ = CH₃
Formicamycin D R₁ = Cl, R₂ = Cl, R₃ = Cl, R₄ = H, R₅ = H, R₆ = H, R₇ = CH₃
Formicamycin E R₁ = Cl, R₂ = Cl, R₃ = Cl, R₄ = H, R₅ = CH₃, R₆ = H, R₇ = CH₃
Formicamycin F R₁ = Cl, R₂ = Cl, R₃ = H, R₄ = Cl, R₅ = CH₃, R₆ = H, R₇ = CH₃
Formicamycin G R₁ = H, R₂ = Cl, R₃ = Cl, R₄ = Cl, R₅ = CH₃, R₆ = H, R₇ = CH₃
Formicamycin H R₁ = Cl, R₂ = H, R₃ = Cl, R₄ = Cl, R₅ = CH₃, R₆ = H, R₇ = CH₃
Formicamycin I R₁ = Cl, R₂ = Cl, R₃ = Cl, R₄ = Cl, R₅ = H, R₆ = H, R₇ = CH₃
Formicamycin J R₁ = Cl, R₂ = Cl, R₃ = Cl, R₄ = Cl, R₅ = CH₃, R₆ = H, R₇ = CH₃
Formicamycin K R₁ = H, R₂ = Cl, R₃ = Br, R₄ = Cl, R₅ = CH₃, R₆ = H, R₇ = CH₃
Formicamycin L R₁ = Cl, R₂ = Cl, R₃ = Br, R₄ = Cl, R₅ = CH₃, R₆ = H, R₇ = CH₃
Formicamycin M R₁ = H, R₂ = Br, R₃ = H, R₄ = H, R₅ = CH₃, R₆ = H, R₇ = CH₃
Formicamycin R R₁ = Cl, R₂ = Cl, R₃ = Cl, R₄ = Cl, R₅ = CH₃, R₆ = Cl, R₇ = H
Formicamycin S R₁ = Cl, R₂ = Cl, R₃ = Cl, R₄ = Cl, R₅ = CH₃, R₆ = Cl, R₇ = CH₃

Figure 6.2: Chemical Structures of Fasamycins and Formicamycins isolated from *Streptomyces formicae*. Fasamycin C-F, and Formicamycins A-M reported by (Qin *et al.*, 2019b; Qin *et al.*, 2017). Fasamycins L-Q and Formicamycins R-S reported by (Devine *et al.*, 2021). Adapted from (Qin *et al.*, 2019b), *Nature Communications*, 10, 3611, licensed under CC BY 4.0.

As discussed in Chapter 1.2.5, the choice of heterologous host can have an impact on the expression profiles of BGCs and the resulting products produced. This is due to various factors, such as; precursor availability, the compatibility of the hosts regulatory factors with the BGCs regulatory elements, and the specificity of chaperone and transport proteins (Ongley *et al.*, 2013; Zhang *et al.*, 2018). Thus, the profiles of NPs produced from a BGC can change depending on differences between heterologous hosts and be distinctly different from that seen in the parental strain. This difference in expression was seen for the NPs fasamycin and formicamycins when produced from parental strain *S. formicae* KY5 WT and *S. formicae* KY5 $\Delta forJ$, which is a mutant of the parental strain, and heterologous host strains *S. erythraea* ISOM/215G $\Delta forJ$ and *S. coelicolor* M1146/215G $\Delta forJ$. It was observed by HPLC and UHPLC-MS that *S. formicae* KY5 WT and *S. formicae* KY5 $\Delta forJ$ produced fasamycins C and E, and formicamycins A through J (**Fig 6.2**), and produced 3.7x and 3.2x more formicamycins than fasamycins, respectively. Comparatively *S. coelicolor*

M1146/215G Δ forJ showed highly variable levels of production of formicamycins A, B, C, D and H amongst biological triplicates, and trace amounts of fasamycins C and D.

S. erythraea ISOM/215G Δ forJ produced only trace amounts of formicamycins A and B, but produced comparatively large amounts of fasamycin C, as well as six new fasamycins identified by HPLC and UHPLC-MS. LC-MS/MS fragmentation of the new fasamycin-like compounds identified the addition of a sugar moiety to the fasamycin C skeleton. Subsequent carbohydrate analysis carried out by Dr Edward Hems, JIC identified a new fasamycin-like compound of m/z 635.21 with a glucose moiety, a new fasamycin-like compound with a m/z of 635.21 with a galactose moiety and a new fasamycin-like compound with a m/z of 649.16 with a glucuronic acid moiety. Purification of the new fasamycin-like compounds is ongoing to determine their true structure.

6.2 Objectives

The main objectives of the project were:

- Generation of *S. coelicolor* M1146 and *S. erythraea* ISOM heterologous host strains containing the *for* BGC PAC 215G and ForJ deletion PAC 215G Δ *forJ*
- Comparison of the titre and expression profile of fasamycins and formicamycins produced in the repressor deleted strains compared to parental strain *S. formicae* KY5 and *S. formicae* KY5/ Δ *forJ*
- Identification of potential new fasamycin congeners arising from expression of the *for* BGC in heterologous host *S. erythraea* ISOM

6.3 Deletion of *forJ* in the *for* BGC of *S. formicae* KY5 and in PAC 215G containing the *for* BGC

The gene *forJ* was identified as a repressor of the *for* BGC via deletion studies in *S. formicae* KY5, where *forJ* was removed from the *for* BGC via CRISPR/Cas9 mediated editing (Devine *et al.*, 2021). It was determined via 3x-FLAG-tagged fusion constructs and Chromatin Immunoprecipitation (ChIP) sequencing, which identifies the binding sites of DNA-associated proteins (Park, 2009), that ForJ represses the *for* biosynthetic machinery by binding to multiple sites on the BGC, acting as a roadblock to the transcriptional machinery and thus preventing transcription of the *for* biosynthesis genes (Devine *et al.*, 2021). Deletion of *forJ* also led to the production of fasamycins and formicamycins from *S. formicae* KY5 in liquid media, something not seen before, and a large increase in fasamycin and formicamycin production on solid media, with 28-fold and 5-fold more compound produced, respectively (Devine *et al.*, 2021; Qin *et al.*, 2017).

Following this success *forJ* was deleted from PAC 215G, which was created using *S. formicae* KY5 genomic DNA and contained the entire *for* BGC. This was used to complement a deletion of the *for* BGC in *S. formicae* (the strain was named *S. formicae* KY5 Δ *for*), restoring production of fasamycins and formicamycins (personal communication, Rebecca Devine, JIC). Deletion of *forJ* in the 215G PAC was carried out by Dr Rebecca Devine using Redirect[®] technology (**Fig 6.3**) (Gust *et al.*, 2004). Firstly, a linear PCR fragment was generated which contained an apramycin resistance gene flanked by regions homologous to either side of the *forJ* gene. This was accomplished by using primer design to incorporate the *forJ* flanking regions and the apramycin resistance gene in plasmid pIJ773 as a template (Gust *et al.*, 2003). The linear PCR fragment was then inserted into strain *E. coli* BW25113/pIJ790/215G, which had been previously created via tri-parental mating. Once the linear PCR fragment was within the cell, double homologous crossover then occurred between the *forJ* flanking regions on the PCR fragment and within PAC 215G, thus replacing the *forJ* gene with an apramycin resistance gene marker, creating PAC 215G Δ *forJ* (Rebecca Devine, JIC) (Gust *et al.*, 2002). This was greatly

facilitated by the expression of λ Red genes on the pIJ790 plasmid, which greatly increases the chance of homologous recombination occurring in *E. coli* (Donald L. Court et al, 2002; Poteete, 2001).

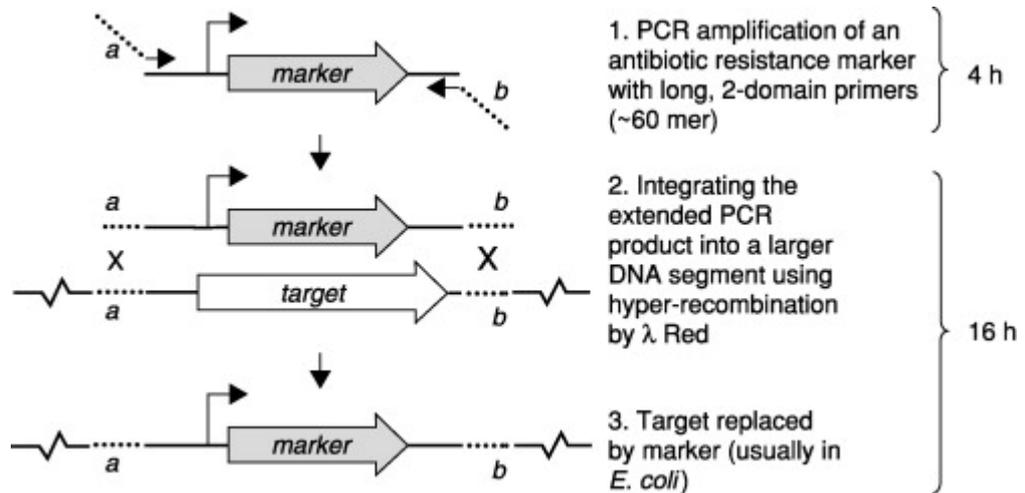


Figure 6.3: Principle of λ Red-Mediated Replacement of a Target Gene by Homologous Recombination. Dotted lines a and b represent 39 bp extension sequences of primers/PCR product homologous to the region adjacent to the target gene. A linear PCR fragment of the antibiotic resistance marker with adjacent regions homologous to regions adjacent to the target gene is created. λ Red mediated hyper homologous recombination results in double cross over of the adjacent regions, replacing the target gene with the antibiotic resistance marker. Reproduced with permission from (Gust et al., 2004), *Advances in Applied Microbiology*, 54, p. 107-128, licenced under Elsevier Ltd.

6.4 Fasamycins and Formicamycins Production Profiles from Heterologous Hosts Expression Strains in Comparison to Parental Strain *S. formicae* KY5

To create fasamycin and formicamycin heterologous production strains the PACs 215G and 215G Δ forJ were inserted into the heterologous expression strains *S. coelicolor* M1146 and *S. erythraea* ISOM via conjugation, to create strains *S. coelicolor* M1146/215G, *S. coelicolor* M1146/215G Δ forJ, *S. erythraea* ISOM/215G and *S. erythraea* ISOM/215G Δ forJ. *S. coelicolor* M1146 and *S. erythraea* ISOM were used as negative controls, the parental strain *S. formicae* KY5 wild-type (WT) was used to determine the base line production of fasamycins and formicamycins within the assay, and strain *S. formicae* KY5 Δ forJ used as a comparison for the effect of formicamycin and fasamycin production after deletion of ForJ.

Both liquid and solid media assays were carried out with biological triplicates, except in the case of *S. formicae* KY5 WT where production only occurs on solid media. Liquid media assays were incubated for 7 days at 30°C with shaking before extracts were generated. Solid media assays were incubated over 10 days at 30°C, three cores of agar were taken around the plate for extraction to account for uneven growth of the strains and uneven NP production. All metabolic extracts were generated with ethyl acetate, the solvent removed by evaporation, and the residue redissolved in methanol before analysis by HPLC and UHPLC-MS.

6.4.2 HPLC Titre of Fasamycins and Formicamycins from the Parental Strain and Heterologous Host Strains

HPLC analysis was carried out by Hannah McDonald, JIC on an Agilent 1290 UHPLC system with a Gemini-NX C18 column using a 50-100% water+0.1% formic acid/methanol gradient over 20 min. Production of fasamycins and formicamycins in liquid media from *S. erythraea* ISOM/215G Δ forJ (Chapter 5.4.3, **Fig 5.5**) and *S. coelicolor* M1146/215G Δ forJ (Appendix 9.4.1) could not be replicated consistently, with NP production only occurring about every 1 in 4 fermentations for every distinct biological sample of the two strains. When conducting the assays below no

fasamycins or formicamycins were produced in liquid media from the *S. erythraea* ISOM and *S. coelicolor* M1146 heterologous host strains. As reported previously *S. formicae* KY5 $\Delta forJ$ produced formicamycins and fasamycins in liquid media, with *S. formicae* KY5 WT showing no production (Appendix 9.4.2) (Devine *et al.*, 2021).

Fasamycin and formicamycin production on solid media was quantified using HPLC chromatogram spectra (**Fig 6.4**). Fasamycins were monitored at 418 nm and formicamycins were monitored at 285 nm, the area under the peak was averaged across all biological triplicates and titre calculated by comparison with standard curves of fasamycin E (10, 20, 50, 80, 200 μM) or formicamycin I (10, 20, 50, 100, 200, 400 μm), respectively (Appendix 9.4.3).

S. formicae KY5 $\Delta forJ$ produced the largest amount of fasamycins and formicamycins, with over 7.2 times more fasamycins and 6 times more formicamycins than the parental strain *S. formicea* KY5 WT, and is comparable with results seen previously (Devine *et al.*, 2021). Heterologous host strains *S. coelicolor* M1146/215G and *S. erythraea* ISOM/215G produced no fasamycins or formicamycins, while strains *S. coelicolor* M1146/215G $\Delta forJ$ and *S. erythraea* ISOM/215G $\Delta forJ$ did show production, indicating that the deletion of *forJ* was essential for production of fasamycins and formicamycins in heterologous hosts, as was discussed previously in Chapter 5. Strain *S. coelicolor* M1146/215G $\Delta forJ$ produced very little fasamycins but comparatively large amounts of formicamycins, about 2 times the amount produced by *S. formicae* KY5 WT. However, as seen by the standard deviation of the sample ($156.82 \pm 171.61 \mu\text{M}/\text{mL}$) the amount of formicamycins produced differed hugely across triplicates. This was consistent with the unpredictable production pattern of the NPs we had observed previously from the same strain in liquid media (data not shown). When excluding the outlying biological triplicate, the titre of formicamycins was reduced by 2.5 times in *S. coelicolor* M1146/215G $\Delta forJ$ compared to *S. formicae* KY5 WT. Strain *S. erythraea* ISOM/215G $\Delta forJ$ produced trace levels of formicamycins but produced 1.6 times more fasamycins than *S. formicae* KY5 WT. This was comprised predominantly of

the non-halogenated fasamycin C and a series of new fasamycin congeners as discussed further below in section 6.5.

UHPLC-MS was used to identify the specific fasamycins and formicamycins produced from each strain. *S. formicae* KY5 WT and *S. formicae* KY5 $\Delta forJ$ produced a wide range of multi halogenated fasamycins and formicamycins containing 1 to 4 halogenated groups (fasamycins E and formicamycin A to J, **Fig 6.2**) (data not shown). Comparatively, *S. coelicolor* M1146/215G $\Delta forJ$ produced only small amounts of fasamycins C and D, and formicamycins A, B, C, D, and H (**Fig 6.2**), containing up to 3 halogenated groups. Strain *S. erythraea* ISOM/215G $\Delta forJ$ had trace levels of halogenated formicamycins A and B but was predominantly seen to only produce non-halogenated fasamycin C and a series of new fasamycin congeners, discussed further below in section 6.5.

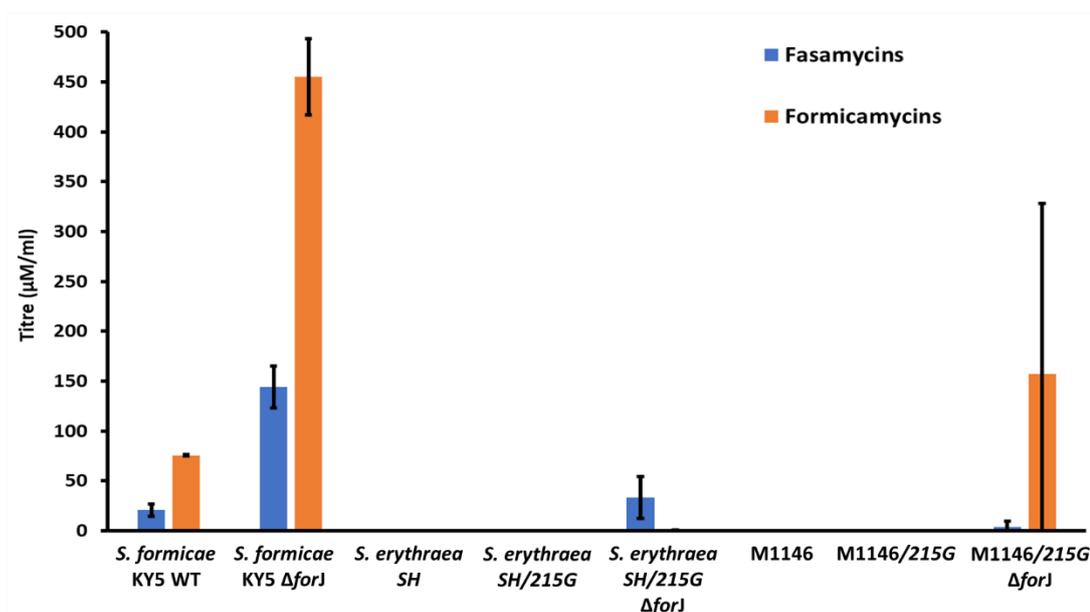


Figure 6.4: Titre ($\mu\text{g}/\text{mL}$) of Fasamycins (Blue) and Formicamycins (Orange) Produced from *S. formicae* and Heterologous Host Strains. Assay was carried out with biological triplicates. Strains were grown on solid SF+M media and extracts were generated using ethyl acetate; $n=3$; analysed using HPLC at 418nm (fasamycins) and 285 nm (formicamycins). *S. erythraea* ISOM and *S. coelicolor* M1146 were used as negative controls. Fasamycin and formicamycin production from *S. formicae* KY5 WT was taken as baseline production. No fasamycins or formicamycins were produced from strains *S. erythraea* ISOM/215G or *S. coelicolor* M1146/215G. *S. formicae* KY5 $\Delta forJ$ produced the most fasamycin and formicamycins. Error bars = 1 standard deviation, 215G = *for* BGC PAC, *forJ* = MarR family BGC repressor.

6.5 Identification of New Fasamycins Produced by *S. erythraea* ISOM/215G Δ forJ

6.5.1 Observation of New early Eluting Fasamycins by HPLC from *S. erythraea* ISOM/215G Δ forJ Solid Media Extracts

Previously, fasamycin C was consistently seen to elute before all other fasamycins and formicamycins (Devine *et al.*, 2021; Qin *et al.*, 2020; Qin *et al.*, 2017). However, for the *S. erythraea* ISOM/215G Δ forJ solid media extracts six new peaks were seen to elute before fasamycin C (**Fig 6.5**). The chromophores of all six peaks matched the distinctive UV profile of fasamycin polyketides as seen previously (280, 350, 420 nm) (Devine *et al.*, 2021; Qin *et al.*, 2020; Qin *et al.*, 2017). Thus, we identified six new fasamycins exclusively produced from heterologous host strain *S. erythraea* ISOM/215G Δ forJ.

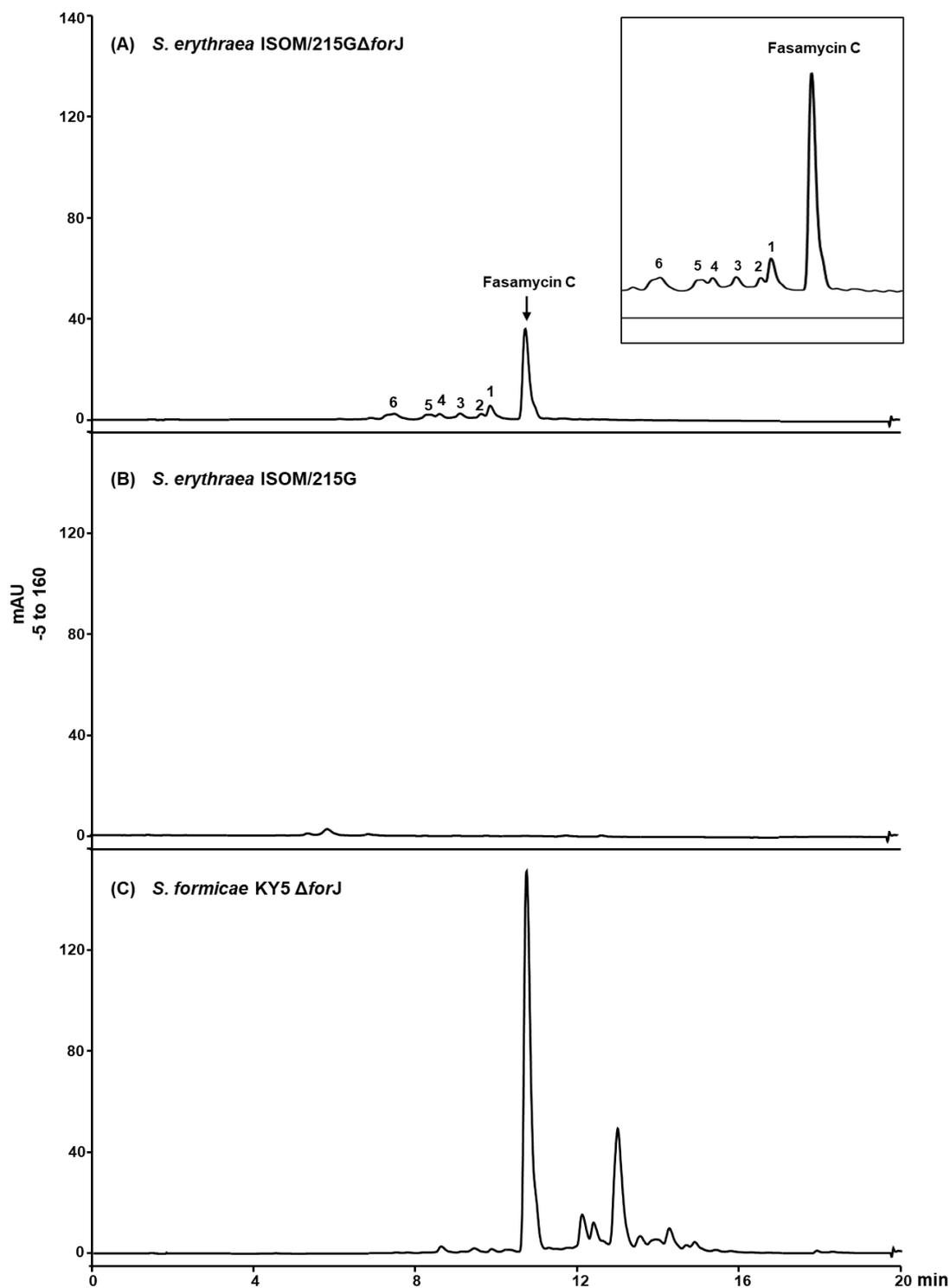
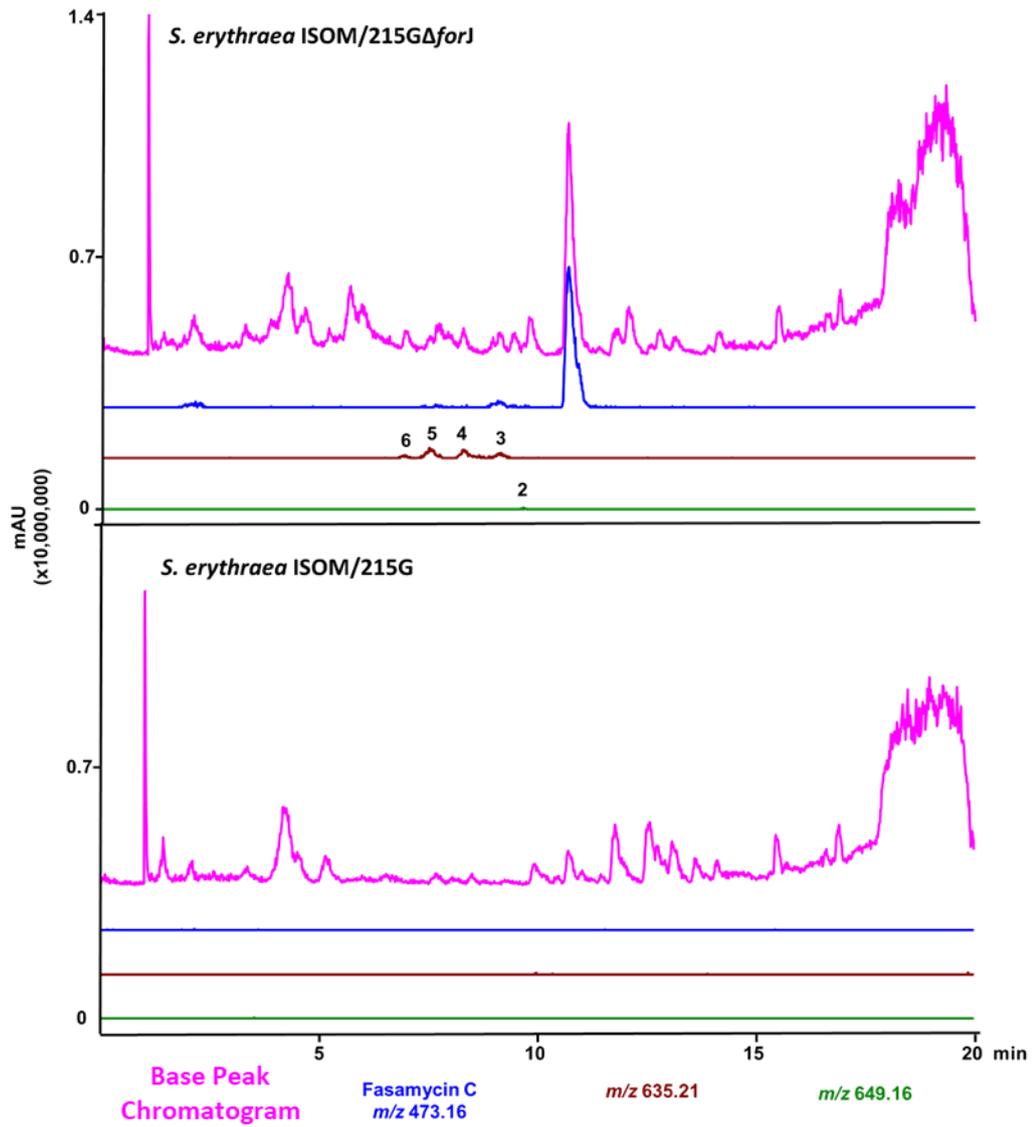


Figure 6.5: Six New Fasamycin Elution Peaks Identified in *S. erythraea* ISOM/215G Δ forJ HPLC Chromatogram. HPLC chromatograms (418 nm), water + 0.1% FA/methanol gradient 50-100 %, run time 20 min, peaks measured in milli-absorbance units (mAU), scaled -5 to 160. All HPLC run by Hannah McDonald, JIC. (A) *S. erythraea* ISOM/215G Δ forJ chromatogram showing fasamycin C ($m/z = 473.16$), the previously fastest eluted fasamycin compound, and new faster eluting fasamycin-like chromatographic peaks (UV profile matches fasamycin polyketides, 280, 350, 420 nm). (B) *S. erythraea* ISOM/215G chromatogram showing no peaks. (C) *S. formicae* KY5 Δ forJ showing fasamycin C, but no faster eluting fasamycin-like polyketide peaks (as determine by UV profile).

6.5.2 LCMS analysis of New Fasamycins

The six new fasamycin peaks were analysed by Hannah McDonald, JIC by UHPLC-MS on a Shimadzu single quadrupole LCMS-2010A mass spectrometer using a C18 reverse phase analytical column with a 50-100 % water+0.1 % formic acid/methanol gradient over 20 min. An m/z of 635.21 was assigned to peaks 3, 4, 5 and 6, peak 2 was assigned an m/z of 649.16 and peak 1 remained unassigned at this time (**Fig 6.6A**). LC-MS/MS fragmentation analysis of peaks 3, 4 and 5 showed a distinct fragment loss of 162, commonly associated with a loss of a hexose moiety. LC-MS/MS fragmentation analysis of peak 2 showed a distinct fragment loss of 176, commonly associated with loss of a uronic acid, a class of sugar acids with both a carboxylic acid functional group (**Fig 6.6B**).

(A) LC-MS



(B) MS-MS Fragmentation

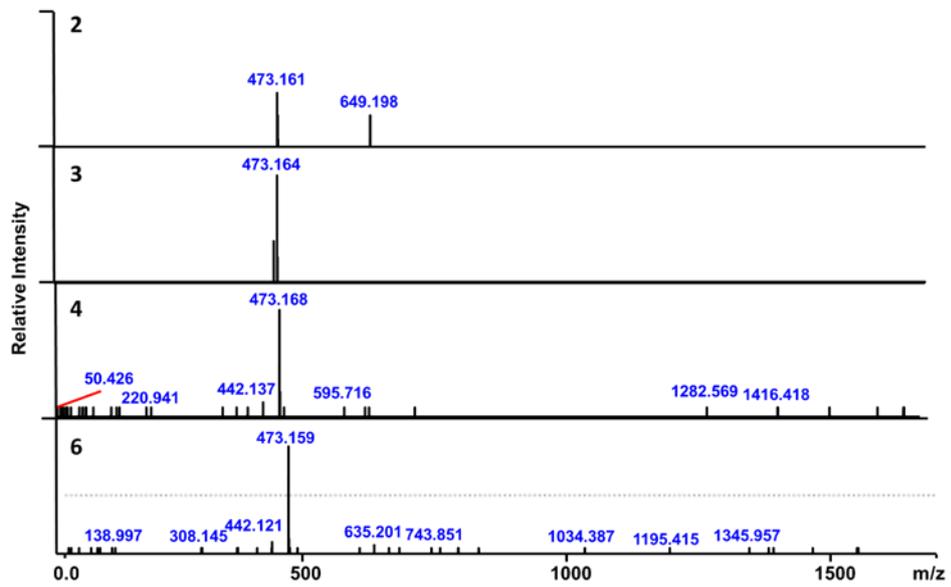


Figure 6.6: Extracted Ion Chromatograms (EIC) and MS/MS Fragmentation of New Glycosylated Fasamycin Congeners. UHPLC-MS analysed with mass fragment search. (A) Mass spectra of *S. erythraea* ISOM/215G and *S. erythraea* ISOM/215G Δ forJ, Pink: Base Peak Chromatogram, Blue: EIC fasamycin C $m/z = 473.16$, Brown: EIC $m/z = 635.21$, fasamycin congeners 3, 4, 5 and 6, Green: EIC $m/z = 649.16$, fasamycin congener 2. Fasamycin congener 1 was not assigned. EIC scaled 0 to 1.4, 10,000,000 intensity. Run time = 20 min. (B) LC-MS/MS for glycosylated fasamycin congeners 2, 3,4 and 5. Mass fragmentation was not obtained for congener 6. All MS/MS fragmentation resulted in an ion of $m/z = 473.16$, equivalent to fasamycin C.

The incorporation of a sugar moiety would account for the early elution of the new fasamycin congeners, as sugar moieties are polar and thus the compounds would be expected to be less hydrophobic than fasamycin C and elute at a lower percentage of organic solvent, as was seen in **Fig 6.5**. LC-MS/MS fragmentation of the peaks 2, 3, 4 and 5 resulted in a glycone ion with m/z of 473.16, equivalent to fasamycin C. Thus, it was proposed that fasamycin-like congeners 3, 4 and 5 were comprised of a fasamycin C skeleton with an additional hexose moiety, with fasamycin-like congener 2 comprised of a fasamycin C skeleton with an additional uronic acid moiety. As the fasamycin-like congener 6 has the same mass as congeners 3, 4 and 5 ($m/z = 635.21$) it is likely to also comprise of a hexose moiety attached to a fasamycin C skeleton, though LC-MS/MS fragmentation has not been obtained at this time. The difference in elution times of peaks 3, 4, 5 and 6, despite their equivalent mass, was hypothesised to result from the attachment of the hexose moiety to different carbon atoms on the fasamycin C skeleton (**Fig 6.7**).

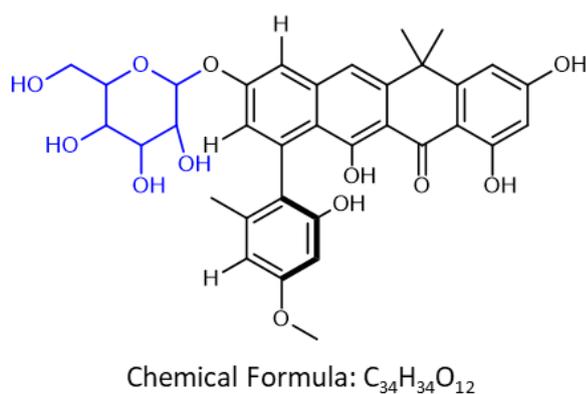


Figure 6.7: Hypothetical Structure of New Fasamycins 3, 4, 5 and 6. Fasamycin C skeleton (black) with additional hexose group (blue). The position of the additional hexose group has not been confirmed and is only theoretical.

6.5.3 Carbohydrate Analysis of New Glycosylated Fasamycins

Fasamycin-like congeners 2 and 3 were purified by Hannah McDonald, JIC and carbohydrate analysis carried out by Edward Hems, JIC by hydrolysing the sugar moieties from the fasamycin C skeleton in trifluoroacetic acid (TFA) at 105°C overnight. TFA was removed through the addition of water and freeze drying before the samples were loaded onto a C18 cartridge and the sugar moieties eluted in 5 % methanol, carbohydrate elutes of fasamycin-like congeners 2 and 3 were termed Carb2 and Carb3, respectively. The solvent was evaporated and the carbohydrate samples re-suspended in water to a concentration of 100 µM. The carbohydrate samples were then run on a Dionex ICS-5000 High performance anion exchange chromatography-pulsed amperometric detection (HPAE-PAD) instrument with a ThermoScientific CarboPac™ PA20 3×150 mm analytical column using an increasing concentration gradient of sodium hydroxide (NaOH). The samples were compared to standards of glucose, galactose and mannose (Carb3) and galacturonic acid and glucuronic acid (Carb2), and co-injection performed for verification. Carb2 was seen to contain only glucuronic acid, indicating that fasamycin-like congener 2 comprised of a single compound glycosylated with glucuronic acid (**Fig 6.8A**). Carb3 was seen to contain both galactose and glucose, indicating that fasamycin-like congener 3 was comprised of two compounds, one glycosylated with galactose and another glycosylated with glucose (**Fig 6.8B**). It is likely that both compounds are distinct new fasamycin-like compounds with similar elution times on HPLC, thus accounting for the single peak seen.

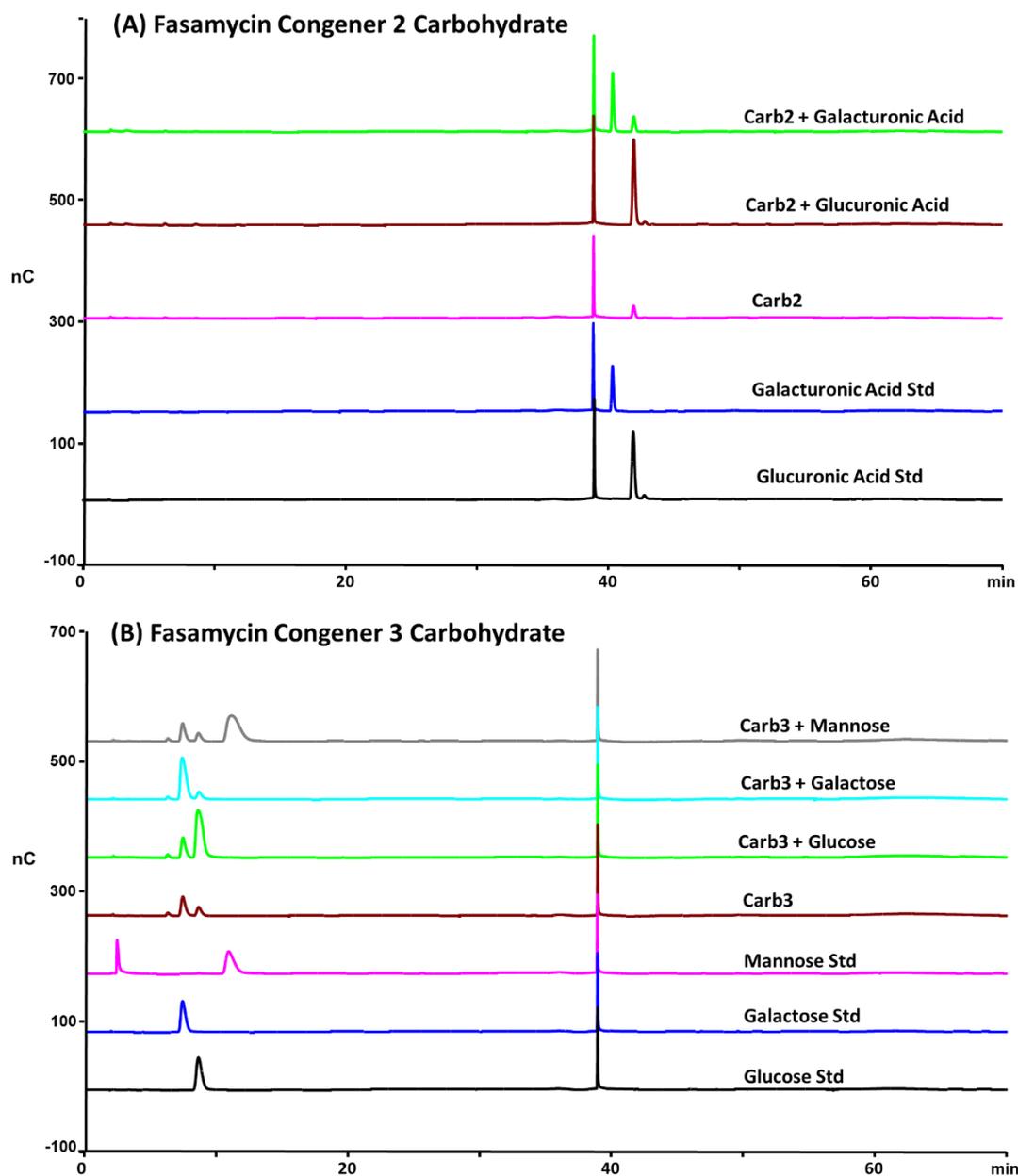


Figure 6.8: Carbohydrate Analysis (HPAE-PAD) of Carb2 and Carb3 obtained from Fasamycin-like Congeners 2 (m/z 649.16) and 3 (m/z 635.21). Congeners purified by HPLC from *S. erythraea* ISOM/215G Δ forJ grown on SF+M media. Carbohydrate hydrolysed from fasamycin skeleton (1.0 M TFA, 105°C, overnight), purified using C18 column and eluted in 5 % MeOH. (A) Carbohydrate from fasamycin-like congener 2 with m/z 649.16 (Carb2) predicted to contain a uronic acid moiety by MS/MS fragmentation. Sample seen to contain glucuronic acid. (B) Carbohydrate obtained from fasamycin-like congener 3 with m/z of 635.21 (Carb3) predicted to contain a hexose moiety by MS/MS fragmentation. Sample seen to contain galactose and glucose.

6.6 Discussion

There was a clear difference between the production profile of fasamycins and formicamycins from heterologous host strains in comparison to the *S. formicae* WT strain. *S. formicae* KY5 WT and *S. formicae* KY5 $\Delta forJ$ produced a wide range of multi halogenated formicamycins (1 Cl to 4 Cl, formicamycin A to J, respectively), with a comparatively small amount of fasamycins, suggesting that most fasamycins produced are successfully halogenated by ForV and converted into formicamycins by ForX and ForY, as discussed in 6.1. As was previously observed *S. formicae* KY5/ $\Delta forJ$ produced several times more fasamycins and formicamycins than *S. formicae* KY5 WT, with an increase of 7.2x fasamycins and 6x formicamycins, respectively (Devine *et al.*, 2021). Comparatively, the heterologous host strains *S. coelicolor* M1146/215G and *S. erythraea*/215G could not produce fasamycins or formicamycins without the $\Delta forJ$ knock out. Even with the ForJ biosynthesis repressor removed, significantly less fasamycins and formicamycins were produced compared to *S. formicae* KY5 $\Delta forJ$.

Strain *S. coelicolor* M1146/215G $\Delta forJ$ only produced formicamycins A to D, and H (1 Cl to 3 Cl), suggesting that halogenation was not as efficient as in *S. formicae* KY5 WT. Previously, halogenation had been identified as a crucial gatekeeping step for production of formicamycins from the fasamycin precursors (Qin *et al.*, 2020). If this was not occurring efficiently we would expect to see an accumulation of the non-halogenated fasamycin C precursor. However, *S. coelicolor* M1146/215G $\Delta forJ$ extracts had very little fasamycin C, suggesting that halogenation was occurring at a speed relative to fasamycin C production, but that only select formicamycins were produced. This lack of halogenation could be due to a reduction in the efficiency of the ForV halogenase, caused by the difference in the intercellular environment of the heterologous host *S. coelicolor* M1146 compared to *S. formicae* KY5.

The strain *S. erythraea* ISOM/215G $\Delta forJ$ produced trace levels of formicamycins, but instead there was accumulation of fasamycin C, as would be expected if halogenation was not occurring. Further to this, six new early eluting novel

fasamycin congeners were observed. LC-MS/MS analysis indicated these were derivatives of fasamycin C modified with a hexose. Glycosylation would make the compounds more hydrophilic, accounting for their early elution under the analytical HPLC conditions used. *S. erythraea* ISOM has previously been used as a heterologous system for the biotransformation of different NPs with deoxyhexose residues. This includes the production of engineered spinosyn and other macrolide analogues (Gaisser *et al*, 2009; Gaisser *et al*, 2002; Schell *et al*, 2008). In these experiments, the strain was shown to possess a background glycosylation activity capable of adding D-glucose to a number of these molecules. Currently it is unknown whether the glycosyltransferase responsible for attachment of the hexose moieties to the fasamycin C scaffold is derived from the host *S. erythraea* ISOM genome, or whether it is encoded within the 215GΔ*forJ* PAC. In the latter case, the *for* BGC does not encode a glycosyltransferase so it would most likely be encoded in the regions of *S. formicae* KY5 genomic DNA flanking the *for* BGC within the PAC.

Clearly heterologous expression has a detrimental effect on the activity of the ForV halogenase encoded within the *for* BGC. This could potentially arise from misfolding of the halogenase due to lack of specific folding or chaperone proteins within the heterologous host (Gasser *et al*, 2008; Hartl & Hayer-Hartl, 2002; Kolaj *et al*, 2009). Misfolding could cause the halogenase to be retained within inclusion bodies within the cell, rendering it unable to act against the fasamycin precursors, or cause the halogenase to be targeted for degradation (Jackson & Hewitt, 2016). Thus, further studies should assess the ForV halogenase for transcriptional expression, via methods such as real time PCR, and for correct folding, via methods such as protein-tag pull down, ELISA, NMR, and *in vitro* activity assays. Should the ForV halogenase not be folded correctly overexpression of ForV on a separate inducible plasmid may result in the production of more formicamycins, with the aim that producing a large amount of the ForV enzyme will compensate for any inefficiency potentially caused by misfolding (Amrein *et al*, 1995; Kolaj *et al.*, 2009). Previously, similar systems of overexpression have been used to prevent the aggregation of recombinant proteins in *E. coli* by using expression plasmids to over express trigger factors and chaperone proteins (Nishihara *et al*, 2000).

Alternatively, the lack of formicamycins could be due to toxicity of the compounds to the heterologous hosts leading to the introduction of mutations in the *for* BGC to suppress the production of formicamycins. Toxicity induced selective pressure has been known to cause an accumulation of mutations in *E. coli*. Previously, FtsZ, an essential protein for the formation of the cytoskeletal scaffold during cell division (Haney *et al*, 2001; Ma & Margolin, 1999), was deleted from *E. coli*. At first the insertion of foreign FtsZ proteins did not complement this mutation, but an accumulation of 11 single nucleotide polymorphisms in various metabolic pathway genes caused the foreign FtsZ proteins to start functioning for cell division (Gardner *et al*, 2017). Thus, toxicity induced selective pressure could lead to an accumulation of mutations within and outside the *for* BGC suppressing formicamycin production. To test this the entire genome of *S. erythraea* ISOM/215G Δ *forJ* could be sequenced and compared to the genome of *S. erythraea* ISOM, to identify any mutations outside the 215G Δ *forJ* PAC region, and the genome of *S. formicae* KY5, to identify any mutations within the 215G Δ *forJ* PAC region. As there is an overall lack of halogenation in strain *S. erythraea* ISOM/215G Δ *forJ*, with only very trace amounts of halogenated formicamycins A and B produced, and it is known that the halogenation activity of ForV acts as a gatekeeper step before conversion of the fasamycin skeleton to the formicamycin skeleton, mutations preventing the activity of ForV would result in a lack of formicamycin production.

Purification of the novel fasamycin congeners produced from *S. erythraea* ISOM/215G Δ *forJ* is still currently underway in the laboratory. This will allow for further structural and antibacterial activity characterisation, using such methods as NMR and MIC assays, respectively. As fasamycin C has antibacterial activity against clinically relevant strains, including against MRSA and VRE, and the addition of sugar moieties generally increases the solubility of compounds (Solá & Griebenow, 2009; Tams *et al*, 1999), an important factor for clinically viable antibiotics, the new congeners could be of great interest for further study should they retain the bioactivity of fasamycin C.

Overall, we have shown that heterologous host selection can have a large impact on the expression profile of a BGC, resulting in very different titre profiles for NPs and even producing new compound variants. Thus, the choice of heterologous host should be recognised for the impact it may have on the metabolite profile of cryptic BGCs, and the ease of production and purification of NPs in future studies.

Chapter 7:

Discussion

7. Discussion

The overarching theme of this thesis was to investigate the potential of heterologous host expression systems for the production and diversification of both known and new NPs. The production of novel antimicrobial NPs is of increasing importance due to the rise of AMR worldwide, which is leading to increasing deaths by what were once previously preventable infections ((CDC) Centers for Disease Control & Prevention, 2019; (IACG) Interagency Coordination Group on Antimicrobial Resistance, 2019). Throughout this project it was found that heterologous host expression was best used in combination with synthetic biology approaches to produce and investigate novel NPs. This was exemplified through the rapid diversification of class II type B lantipeptides by substitution of the core peptide region within a synthetic cassette system, and through the production of new fasamycin-like glycosylated compounds in *S. erythraea* ISOM when the *forJ* transcriptional repressor is deleted. It was also seen, through the identification and production of rosiermycin in *S. roseoverticillatus*, that by combining genetic sequencing and annotation with the culture of bacterial strains novel NPs could still be yielded from previously studied, laboratory grown strains.

Five major outcomes were achieved in this research:

- Diversification of class II type B lantipeptides using a synthetic heterologous expression platform
- Production of rosiermycin from *S. roseoverticillatus*
- Investigation of *S. erythraea* ISOM as a heterologous host for actinomycetes NP
- Observation of the impact of repressor deletion and heterologous host choice on the production of fasamycins and formicamycins from the *for* BGC
- Production of novel glycosylated fasamycin congeners from the heterologous host *S. erythraea* ISOM

7.1 Diversification of Class II Type B Lantipeptides

Before commencing this project work by Eleni Vikeli at the JIC showed that the kyamicin heterologous expression platform promised a method for the rapid production and diversification of class II type B lantipeptides, through easy modification of the core peptide sequence. The expansion of the class II type B lantipeptide family is potentially very valuable, as the known molecules cinnamycin and duramycin have demonstrated activities as antibiotics, inhibitors of viral entry and as potential therapeutics for cystic fibrosis (An *et al.*, 2018; Oliynyk *et al.*, 2007; Willey & van der Donk, 2007). Through the identification of 35 lantipeptide core peptide sequences by genome mining of the public databases, and the creation of a synthetic library of core peptide sequences based on the kyamicin core peptide, I reported the production of 83 novel class II type B lantipeptides using the heterologous expression platform, 61 of which had antibacterial activity against *B. subtilis*.

Purification of four novel lantipeptides from this platform was achieved and their antibacterial activity assessed against a range of indicator strains. The lantipeptides showed no antibacterial activity against gram-negative bacteria but showed activity against some gram-positive bacteria, including strains of *Streptomyces* bacteria. This was expected, as all but two (L_22 and L_26) of the lantipeptide core peptide sequences from the public databases originated from species in the Actinomycete order of Actinobacteria, of which the *Streptomyces* genus is a part. These bacteria are most commonly found in soil, which is a highly diverse, competitive environment in which bacteria are often competing for space, carbon and mineral resources. It is well documented that many bacteria within the soil produce antimicrobial NPs to inhibit or kill competing bacteria and gain a competitive advantage (Chandra & Kumar, 2017). Thus, lantipeptides are hypothesized to be antimicrobial NPs produced by gram-positive bacteria to kill related gram-positive bacteria found in the same environmental niche.

In most cases, based on LCMS analysis, the new lantipeptide molecules were produced in the expected Asp-15 hydroxylated forms along with the non-hydroxylated biosynthetic intermediate. However, it was also noted during HPLC purification that in most cases there appeared to be multiple forms of each molecule type, e.g., multiple separately eluting peaks with a mass corresponding to the hydroxylated or non-hydroxylated lantipeptide forms. Purification of some of the new lantipeptides was attempted and for one example (L_19) we were able to isolate three isomeric congeners for the non-hydroxylated form of the molecule. Only one of the free isomeric congeners (lantipeptide L_19 Deoxy Fraction 3) displayed antibacterial activity against *B. subtilis* EC1524.

It has been established that the final biosynthetic step of cinnamycin-like lanthipeptides involves cyclization to form the distinctive lysinoalanine bridge, as exemplified by duramycin synthesis below (**Fig 7.1**) (An *et al.*, 2018; Huo *et al.*, 2017; Ökesli *et al.*, 2011; Repka *et al.*, 2017). This step is catalysed by the CinN family of enzymes, and that this occurs in a substrate assisted manner requiring hydroxylation of the Asp15 residue (An *et al.*, 2018); but, it was also ascertained that this reaction can occur spontaneously leading to two epimeric forms of the lantipeptide at the lysinoalanine bridge (An *et al.*, 2018). As it is known that the stereochemistry of a NP can play a vital role in a molecule's ability to actively bind its target (Elder *et al.*, 2020), this spontaneously formed lantipeptide epimer is likely to have a distorted PE binding pocket. Thus, although we could not obtain direct structural evidence to verify the hypothesis, we believe that the three isolated forms of the molecules may comprise the two epimeric forms of the compounds, and potentially the uncyclized lantipeptide. If correct, and taking into the account the results of antibacterial assays against *B. subtilis* with the purified congeners, our data suggest that only the enzymatically formed epimer of the cyclised lantipeptide molecules exhibit antibacterial activity.

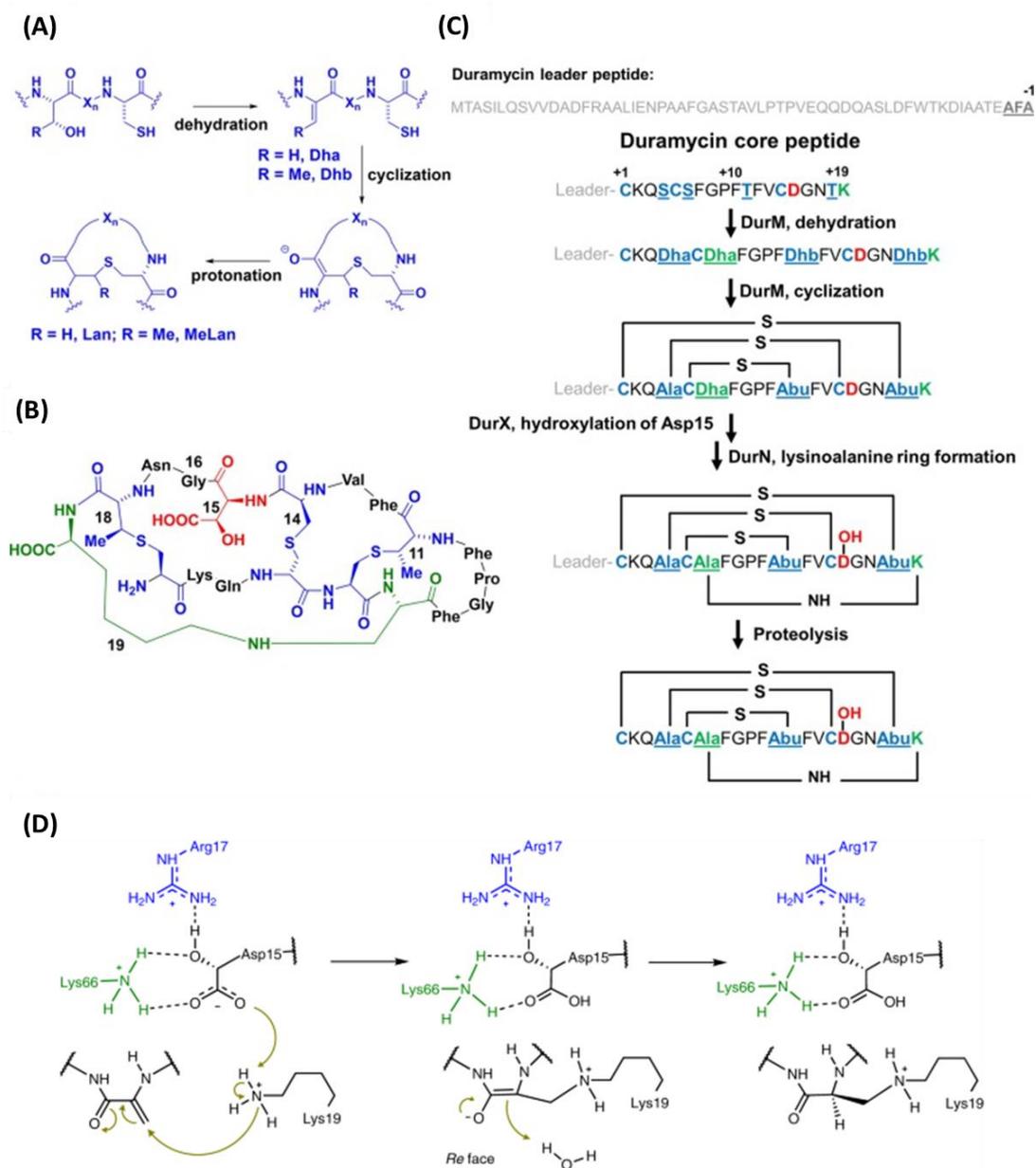


Figure 7.1: Biosynthesis of Duramycin including Lanthionine (Lan) and Methyllanthionine ((Me)Lan) bond formation and Lysinoalanine bridge formation. (A) General biosynthesis of Lan and (Me)Lan bonds. (B) Structure of Duramycin, Lan and (Me)Lan bonds (blue), lysinoalanine bond (red), hydroxylated aspartate (red). (C) Post translational maturation of duramycin. The core peptide sequence is shown, residues that will form the Lan and (Me)Lan cross-links are indicated (blue), the residues that will form the lysinoalanine bond are indicated (green), and the hydroxylated aspartate residue is indicated (red). (D) proposed hydroxylated aspartate assisted lysinoalanine bond formation catalysed by DurN. Features of the two monomers of DurN that catalyse the reaction are in blue and green. Features of the substrate/product are in black. Reproduced with permission from (Huo et al., 2017), *Applied and Environmental Microbiology*, 83(3), e02698-16, licensed under American Society of Microbiology. Reproduced with permission from (An et al., 2018), *Nature Chemical Biology*, 14, p. 928-933, licensed under Springer Nature America, Inc.

The creation of a second heterologous expression platform using the lantipeptide BGC from *S. roseovorticillatus*, following the same design principles as used for the kyamicin heterologous expression platform, did not result in the production of any class II type B lantipeptides, including the native product of this system (putatively named rosiermycin). Potentially, this lack of lantipeptide production arose because of regulatory elements associated with the system that we are unaware of. Alternatively, and while unlikely, the RosR1 and RosL could be different enough from KyaR1 and KyaL to cross interact with metabolic elements of the heterologous host and this could potentially lead to repression of the rosiermycin biosynthetic machinery. Future investigation should focus on the regulation of the rosiermycin BGC, and the interaction of lantipeptide regulatory elements with regulatory elements of the heterologous host.

The successful generation of synthetic lantipeptides now paves the way for the rational design of lantipeptides for the improvement of drug like factors, including antimicrobial activity and biophysical properties such as solubility. Several of the new synthetic lantipeptides we identified warrant further investigation, as the qualitative assessment of the synthetic lantipeptide library identified a number of strains producing comparatively small amounts of lantipeptide but which displayed a large zone of inhibition against *B. subtilis* when compared to the average lantipeptide production and zone of inhibition in the synthetic library. The MIC of antibiotics is highly important when considering their suitability for clinical use, as dosages must be kept to a minimum while still killing/inhibiting bacteria. Thus, future work should look to purify and assess the MIC of these synthetic lantipeptides with low production by a high level of antibacterial activity. Furthermore, the successful production of synthetic lantipeptides enables an investigation into the importance of single amino acid residues for the binding to the target PE, potentially improving understanding into the binding motif and mechanism of action of class II type B lantipeptides

The abundance of novel lantipeptides produced from the kyamicin heterologous expression platform showcases the potential for the diversification of NP through

synthetic biology approaches and heterologous host expression. As most families of RiPP NPs also contain core peptide sequences upon which post translational modifications occur, similar synthetic expression platforms could be created for other RiPP antimicrobials. Indeed, expression platforms seeking to modify the core peptide through saturation mutagenesis have already been carried out for a number of lantipeptides, including lacticin 3147 and nukacin-ISK-1 (Class I) (Field *et al*, 2013; Islam *et al*, 2009) and merscedin and actingardine (Class II) (Appleyard *et al*, 2009; Boakes *et al.*, 2012). The incorporation of non-canonical amino acids within the core peptide region has also been used to produce derivatives of lantipeptides nisin (Class I) (Bartholomae *et al*, 2018; Zambaldo *et al*, 2017) and lichenicidin (Class II) (Oldach *et al*, 2012).

Further to this expression platforms could be created for the diversification of III, IV and the newly discovered class V of lantipeptides (Hegemann & Süßmuth, 2020; Knerr & Donk, 2012; Ren *et al*, 2020; Xu *et al*, 2020). These other classes of lantipeptides produce NP which are used as food preservatives (class I nisin) (Lubelski *et al*, 2008) and antimicrobials (class I, subtilin; class III, andalusicin) (Grigoreva *et al*, 2021; Qin *et al*, 2019a). Thus, production and diversification of these classes of lantipeptides would also potentially increase the number of novel antimicrobials that could one day replace the currently used antibiotics to which AMR is seen to readily arise.

7.2 Rosiermycin

Production of rosiermycin was unsuccessful through both the kyamicin and rosiermycin heterologous expression platforms but was achieved through culturing of the parental strain *S. roseoverticillatus* under laboratory conditions. Addition of the activator cassettes created from the rosiermycin (pAMA5) and kyamicin (pAMA1) BGCs expressing the SARP transcriptional activator and PE-methyltransferase immunity genes resulted in a lack of growth and reduction in rosiermycin production. This result, along with the lack of production of lantipeptides from the rosiermycin heterologous expression platform discussed above, once again indicates the possibility of unknown regulatory elements associated with the system. Future work stemming from this project should focus on investigating the rosiermycin regulatory system and its wider effect on the metabolome of *S. roseoverticillatus*.

In recent years the identification and growth of bacterial strains under laboratory conditions for the discovery of novel antimicrobial NPs has been plagued by the problem of rediscovery (Aminov, 2010; Baltz, 2006; Gaudencio & Pereira, 2015; Peláez, 2006). However, the production of rosiermycin from the parental strain under laboratory conditions does show that there are any number of strains that are capable of being grown in the laboratory that have not been investigated in full for their potential to produce NPs. As *S. roseoverticillatus* had been previously investigated and reported to produce two different antimicrobial poly-amino acids, γ -poly(dl-glutamic acid) (gPGA) and ϵ -poly(l-lysine) (ePL) (Nishikawa & Kobayashi, 2009), and a further antibacterial compound carbazomycin B (Shi *et al*, 2021), it would be expected that these earlier studies would have already identified the readily produced rosiermycin, but this was not the case. This showcases the importance of genome sequencing and annotation for the discovery of NPs.

The knowledge that *S. roseoverticillatus* encoded for a class II type B lantipeptide allowed for easy identification by LCMS of the rosiermycin NP from growth extracts. With the increasing abundance of annotated genomes within the public database

available to help identify BGCs by homologous comparison, and the speed in which such comparisons can be carried out getting ever faster, a change in methodology for the identification of NP of interest from strains readily and previously grown within the laboratory should be adopted. The identified strains should first be genetically sequenced, then BGCs identified, and the corresponding NP predicated using tools such as antiSMASH. This would then be followed by growth on a range of different production medias under different conditions. This way the potential NPs of a strain will have already been identified, their structure predicated and thus their presence more readily searched for in crude growth extracts, allowing for identification of multiple NPs at once, including NPs that do not readily show antibacterial activity against common indicator strains. This methodology could go some way to improving the depth of investigation into the NP biosynthetic potential of laboratory grown strains. However, it does depend on similar compounds and associated BGCs having been identified and uploaded to the public databases, so would not help in identifying truly novel families of NPs.

Overall, while establishing synthetic approaches for the production of BGCs from unculturable bacterial strains or metagenomic samples in heterologous hosts is of great importance and shows great promise for the diversification of novel NPs, the NPs that can be identified by strains readily culturable in the laboratory has been far from exhausted. With the full biosynthetic potential of many already investigated strains still not understood to any great depth.

7.3 Investigation of *S. erythraea* ISOM as a heterologous host for BGC expression

The increase in whole genome sequencing, and subsequent discovery of cryptic BGC which may encoded for antimicrobial NPs (Foulston, 2019; Sekurova *et al.*, 2019), has highlighted the need to expand the range of heterologous hosts available for NP production (Chen *et al.*, 2019). The suitability of a host for NP production can be affected by a number of factors, such as the similarity of regulatory elements to those of the parental strain and availability of metabolites in the cell (Fernández & Vega, 2016; Zhang *et al.*, 2016). Thus, we investigated *S. erythraea* ISOM as a potential heterologous host for actinomycete NP production. During this study multiple actinomycetes polyketides, RiPPs and NRPS NP BGCs were inserted into *S. erythraea* ISOM, but only 1 of the 11 BGCs inserted resulted in NP production. When these BGCs were examined for heterologous expression in the well-studied host *S. coelicolor* similar results were seen, with the successful heterologous expression of only one additional BGC being observed.

Only strain *S. erythraea* ISOM/215G Δ forJ, containing the *for* BGC but lacking the biosynthetic repressor ForJ, showed production of the associated fasamycin NPs. The strain *S. erythraea* ISOM/215G, containing the *for* BGC with the ForJ repressor, did not produce NPs. Thus, the removal of the ForJ repressor has the greater impact on inducing fasamycin and formicamycin production than insertion into the heterologous host. Previously, the removal of the ForJ repressor was also seen to have a large impact on the production of fasamycin and formicamycin within the parental strain *S. formicae* KY5, inducing NP production in liquid media (Devine *et al.*, 2021). Accordingly, the other BGCs used in this project should be investigated for potential genetic modifications that could aid in inducing NP production, such as repressor knock out or promoter refactoring. However, it should be noted that attempts to induce antascomycin production in *S. erythraea* ISOM and *S. coelicolor* M1146 through the co-expression of the antascomycin transcriptional activators were not successful. As insertion of the transcriptional activators on separate integration plasmids resulted in a phenotypic change in both hosts it is likely that

the transcriptional activators were being produced within the cell and affecting other metabolic functions of the hosts but having no effect on antascomycin production.

The lack of NPs produced during this study shows the limit of heterologous expression as a technique, and how it is very often unsuccessful. While it is generally observed that the more closely related a heterologous host strain is to the strain from which BGC originates, the greater is the likelihood of NP production (Ongley *et al.*, 2013; Zhang *et al.*, 2018), this is in no way a guarantee for success (Zhang *et al.*, 2017a). *S. erythraea* ISOM could not be assessed in this work as a producer of NPs originating from the Pseudonocardiaceae family, of which it is also a part. This was due to a possible recombination event or misalignment during insertion of putative polyene Ps2 PAC into the *S. erythraea* ISOM chromosome, as seen by the presence of a PCR band 300 bp larger than expected, and due to the presence of a two-nucleotide deletion in the nystatin Ps1 PAC, causing a frameshift.

Moving forward, the nystatin Ps1 and putative polyene Ps2 PACs should be remade, ensuring that there are no genetic mutations, and reinserted into the host to assess their production. Should the associated compounds be produced this could indicate that the other NP BGCs inserted during this work originated from actinomycetes species too distantly related to *S. erythraea* ISOM for successful NP production to occur. It is likely this would be due to differences in regulatory elements or metabolic precursors between the heterologous host and parental strain. Should the associated nystatin-like polyene compounds not be produced we can consider that the relatedness of the parental strain to the host was not the limiting factor explaining the lack of NP production from the other actinomycetes BGCs inserted.

7.4 New Glycosylated Fasamycins

The deletion of the ForJ repressor in the *for* BGC leads to greater titres of NPs in the native strain *S. formicae* KY5 (Devine *et al.*, 2021). It also induced NP production from the *for* BGC in heterologous host *S. coelicolor* M1146, though this was not consistent. When expressed in *S. erythraea* ISOM the derepressed *for* BGC led to the production of known fasamycins in addition to at least six new glycosylated fasamycin congeners. LCMS analysis of the six new fasamycin-like compounds identified four congeners with a m/z of 635.21 (3, 4, 5 and 6) and one congener with an m/z of 649.16 (2), with congener 1 remaining unassigned at this time. LC-MS/MS fragmentation indicated that the congeners 3, 4 and 5 contained a hexose group bound to a fasamycin C skeleton and that congener 2 contained a uronic acid group, again bound to a fasamycin C skeleton. Purification and carbohydrate analysis on congener 3 suggested it was actually comprised of two fasamycin-like congeners, containing either a glucose or a galactose group, and that congener 2 was comprised of a single fasamycin-like compound containing a glucuronic acid group. Currently it is unknown how these sugar moieties are bound to the fasamycin C skeleton as the congeners are still in the process of being purified in sufficient quantities to allow their structures to be determined. Once purified, future work will also focus on assessing the potential antibacterial activity of the compounds.

The production of the glycosylated fasamycin congeners shows the effect that heterologous hosts regulatory and metabolic elements can have on NP expression profiles and titres. It is well documented that heterologous host expression can affect the titre of NPs (Ke & Yoshikuni, 2020; Wenzel *et al.*, 2005), however, production of derivatives of NPs is most often caused by the feeding of different precursor molecules, or by genetic refactoring of the BGC within the heterologous host (Alberti *et al.*, 2017; Baltz, 2010; Khaw *et al.*, 1998; Pickens *et al.*, 2011; Sandmann, 2002). Our work demonstrates that without these interferences it is still possible for heterologous host expression to have a far-reaching effect on NP expression profiles and the production of new congeners. Thus, heterologous host

production should be considered as a method of NP production even for compounds that can be readily produced from their native strains. Future work with actinomycetes NPs should look to compare a range of different heterologous hosts for the same BGC, assessing the differences in titres of the NPs and determining the possible production of new derivatives.

7.5 Conclusions

While insertion of BGCs into heterologous hosts does not guarantee production of NPs the ease of use of heterologous hosts and their genetic tractability makes them valuable assets for NP production when coupled with genetic engineering and synthetic biology approaches. Generally, for heterologous production a BGC of interest is first identified, commonly by genome mining, and compared to homologous BGCs available in the public databases. While this can offer insight into the role of the genes within the BGC and the possible NP they encodes for, it does not give a full overview of the regulation of the BGCs in the native strain, nor how the BGCs regulatory elements and NP biosynthesis may affect the wider metabolome within a heterologous host.

The effect inserted BGCs have on heterologous host gene expression and the wider metabolome has not been widely investigated, with efforts instead focused on increasing NP titre from heterologous hosts by deletion of secondary metabolite clusters (Zhang *et al.*, 2016), and improving the 'flux' of metabolites to create the desired NP (Bilyk *et al.*, 2017; Cook *et al.*, 2021; Pham *et al.*, 2021). While phenotypic changes, such as loss of pigmentation and lack of sporulation, are often seen upon expression of a BGC in a heterologous host the cause of these changes is rarely investigated, merely theorised. As reported, current thinking is that the more closely related a host is to the parental strain of a BGC the more likely successfully production of NP will arise through incorporation of the BGC into the host (Ongley *et al.*, 2013; Zhang *et al.*, 2018), and thus it is common to select heterologous hosts based on their relatedness to the parental strain of the BGC of interest. Considering this, if the regulatory elements found on the inserted BGC are similar to regulatory elements of the host it is possible they could potentially interact with cluster or global regulatory elements of the host, resulting in signalling cascades which radically change the host metabolome, preventing sporulation and abolishing production of pigmentation (McLean *et al.*, 2019; Teijaro *et al.*, 2019).

Investigating the interactions of BGC regulatory elements to targets within the heterologous host could provide insights into the regulation of the BGCs within the parental strain. This would be of great importance for understanding strains that cannot be cultured under laboratory conditions by facilitating the investigation of BGC regulation without the need to culture the parental strain. If the heterologous host used is sufficiently similar to that of the parental strain the interaction of BGC regulatory elements with the wider cell metabolome may also provide insight into the systems of regulation and interaction of metabolites in the parental strain. This could be particularly useful for investigating BGCs identified from strains found in extreme environmental niches, which are being investigated as potential sources of novel antimicrobials (Goodfellow *et al.*, 2018; Jang *et al.*, 2013; Mincer *et al.*, 2002; Wietz *et al.*, 2012). The understanding of host metabolome and BGC gene interaction could also potentially indicate the suitability of a strain to become a heterologous host for NP production. Should the common regulatory elements of BGCs, such as SARP and LuxR transcriptional activators in *Streptomyces* BGCs (Liu *et al.*, 2013; Romero-Rodríguez *et al.*, 2015; Wei *et al.*, 2018), be seen to cause metabolic changes in the host, affecting its wider metabolome and regulatory mechanisms, that strain may not be suitable as a heterologous host.

Genetic engineering and synthetic biology approaches have proven to be invaluable tools for NP production during this work. Promoter design, transcriptional activator expression and deletion of transcriptional repressors have long been known to effect NP production and are successful strategies to induce NP production within heterologous hosts (Baral *et al.*, 2018; Chen *et al.*, 2017b; Olano *et al.*, 2014; Reen *et al.*, 2015; Wang *et al.*, 2019; Zhang *et al.*, 2017b). As computational tools and databases continue to improve, more accurate annotation of BGCs should follow, allowing for more in-depth refactoring and genetic manipulation of BGCs of interest. Following this, more encompassing synthetic systems, such as the synthetic heterologous production platform used in this project to produce class II type B lantipeptides, can be created. This allows for not only the production of novel NPs but also the rapid diversification and rational synthetic design of NPs for improved clinical potential. This is of increasing importance as AMR bacteria

continue to arise worldwide, causing deaths from previously preventable infections ((CDC) Centers for Disease Control & Prevention, 2019; (IACG) Interagency Coordination Group on Antimicrobial Resistance, 2019), and making the need for novel antimicrobials a pressing concern.

To avoid a post-antibiotic era and develop novel therapies we must continue to expand the tool kit of synthetic biology, deepening our understanding of the methods already available, such as heterologous hosts, while innovating systems for the future.

Chapter 8:

References

8. References

(CDC) Centers for Disease Control, Prevention (2019). Antibiotic resistance threats in the United States, 2019, in: National Center for Emerging Z., Infectious Diseases . Division of Healthcare Quality Promotion. Antibiotic Resistance C., Strategy U. (Eds.). <http://dx.doi.org/10.15620/cdc:82532>, Atlanta, GA.

(IACG) Interagency Coordination Group on Antimicrobial Resistance, (2019). No Time To Wait: Securing the Future from Drug-Resistant Infections.

Aanen DK, Eggleton P, Rouland-Lefèvre C, Guldborg-Frøsvlev T, Rosendahl S, Boomsma JJ (2002) The evolution of fungus-growing termites and their mutualistic fungal symbionts. *Proceedings of the National Academy of Sciences* 99: 14887-14892

Abdel-Razek AS, El-Naggar ME, Allam A, Morsy OM, Othman SI (2020) Microbial Natural Products in Drug Discovery. *Processes* 8

Acevedo-Rocha CG, Ferla M, Reetz MT (2018) Directed Evolution of Proteins Based on Mutational Scanning. *Methods Mol Biol* 1685: 87-128

Adedeji WA (2016) THE TREASURE CALLED ANTIBIOTICS. *Ann Ib Postgrad Med* 14: 56-57

Adnani N, Rajski SR, Bugni TS (2017) Symbiosis-inspired approaches to antibiotic discovery. *Natural product reports* 34: 784-814

Ahmed Y, Rebets Y, Estévez MR, Zapp J, Myronovskiy M, Luzhetskyy A (2020) Engineering of *Streptomyces lividans* for heterologous expression of secondary metabolite gene clusters. *Microbial cell factories* 19: 5-5

Aigle B, Corre C (2012) Waking up *Streptomyces* secondary metabolism by constitutive expression of activators or genetic disruption of repressors. *Methods Enzymol* 517: 343-366

Ajibola O, Omisakin OA, Eze AA, Omoleke SA (2018) Self-Medication with Antibiotics, Attitude and Knowledge of Antibiotic Resistance among Community Residents and Undergraduate Students in Northwest Nigeria. *Diseases* 6: 32

Akbar N, Siddiqui R, Iqbal M, Sagathevan K, Khan NA (2018) Gut bacteria of cockroaches are a potential source of antibacterial compound(s). *Lett Appl Microbiol* 66: 416-426

Alberti F, Khairudin K, Venegas ER, Davies JA, Hayes PM, Willis CL, Bailey AM, Foster GD (2017) Heterologous expression reveals the biosynthesis of the antibiotic pleuromutilin and generates bioactive semi-synthetic derivatives. *Nat Commun* 8: 1831-1831

Aldholmi M, Marchand P, Ourliac-Garnier I, Le Pape P, Ganesan A (2019) A Decade of Antifungal Leads from Natural Products: 2010–2019. *Pharmaceuticals* 12: 182

Alt S, Wilkinson B (2015) Biosynthesis of the Novel Macrolide Antibiotic Anthracimycin. *Acs Chemical Biology* 10: 2468-2479

Aminov RI (2010) A brief history of the antibiotic era: lessons learned and challenges for the future. *Frontiers in microbiology* 1: 134-134

Amrein KE, Takacs B, Stieger M, Molnos J, Flint NA, Burn P (1995) Purification and characterization of recombinant human p50csk protein-tyrosine kinase from an Escherichia coli expression system overproducing the bacterial chaperones GroES and GroEL. *Proc Natl Acad Sci U S A* 92: 1048-1052

An L, Cogan DP, Navo CD, Jiménez-Osés G, Nair SK, van der Donk WA (2018) Substrate-assisted enzymatic formation of lysinoalanine in duramycin. *Nature Chemical Biology* 14: 928-933

Andexer JN, Kendrew SG, Nur-e-Alam M, Lazos O, Foster TA, Zimmermann A-S, Warneck TD, Suthar D, Coates NJ, Koehn FE et al (2011) Biosynthesis of the immunosuppressants FK506, FK520, and rapamycin involves a previously undescribed family of enzymes acting on chorismate. *Proceedings of the National Academy of Sciences* 108: 4776-4781

Appleyard AN, Choi S, Read DM, Lightfoot A, Boakes S, Hoffmann A, Chopra I, Bierbaum G, Rudd BAM, Dawson MJ et al (2009) Dissecting Structural and Functional Diversity of the Lantibiotic Mersacidin. *Chemistry & Biology* 16: 490-498

Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J et al (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* 30: 108-160

Bachmann BO, Van Lanen SG, Baltz RH (2014) Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J Ind Microbiol Biotechnol* 41: 175-184

- Baltz RH** (2006) Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *Journal of Industrial Microbiology and Biotechnology* 33: 507-513
- Baltz RH** (2008) Renaissance in antibacterial discovery from actinomycetes. *Current Opinion in Pharmacology* 8: 557-563
- Baltz RH** (2010) Streptomyces and Saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. *J Ind Microbiol Biotechnol* 37: 759-772
- Baral B, Akhgari A, Metsä-Ketelä M** (2018) Activation of microbial secondary metabolic pathways: Avenues and challenges. *Synth Syst Biotechnol* 3: 163-178
- Barke J, Seipke RF, Gruschow S, Heavens D, Drou N, Bibb MJ, Goss RJM, Yu DW, Hutchings MI** (2010) A mixed community of actinomycetes produce multiple antibiotics for the fungus farming ant Acromyrmex octospinosus. *Bmc Biology* 8
- Barker KF** (1999) Antibiotic resistance: a current perspective. *Br J Clin Pharmacol* 48: 109-124
- Bartholomae M, Baumann T, Nickling JH, Peterhoff D, Wagner R, Budisa N, Kuipers OP** (2018) Expanding the Genetic Code of Lactococcus lactis and Escherichia coli to Incorporate Non-canonical Amino Acids for Production of Modified Lantibiotics. *Frontiers in Microbiology* 9
- Beemelmans C, Guo H, Rischer M, Poulsen M** (2016) Natural products from microbes associated with insects. *Beilstein J Org Chem* 12: 314-327
- Benndorf R, Guo H, Sommerwerk E, Weigel C, Garcia-Altare M, Martin K, Hu H, Kufner M, de Beer ZW, Poulsen M et al** (2018) Natural Products from Actinobacteria Associated with Fungus-Growing Termites. *Antibiotics (Basel, Switzerland)* 7: 83
- Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D et al** (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* 417: 141-147
- Berdy J** (2012) Thoughts and facts about antibiotics: where we are now and where we are heading. *The Journal of antibiotics* 65: 385-395

Bibb MJ (2005) Regulation of secondary metabolism in streptomycetes. *Current Opinion in Microbiology* 8: 208-215

Bierman M, Logan R, O'Brien K, Seno ET, Rao RN, Schoner BE (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* 116: 43-49

Bilyk B, Horbal L, Luzhetskyy A (2017) Chromosomal position effect influences the heterologous expression of genes and biosynthetic gene clusters in *Streptomyces albus* J1074. *Microbial cell factories* 16: 5-5

Birck C, Malfois M, Svergun D, Samama J (2002) Insights into signal transduction revealed by the low resolution structure of the FixJ response regulator. *J Mol Biol* 321: 447-457

Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M *et al* (2017) antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45: W36-w41

Boakes S, Appleyard AN, Cortés J, Dawson MJ (2010) Organization of the biosynthetic genes encoding deoxyactagardine B (DAB), a new lantibiotic produced by *Actinoplanes liguriae* NCIMB41362. *The Journal of antibiotics* 63: 351-358

Boakes S, Ayala T, Herman M, Appleyard AN, Dawson MJ, Cortés J (2012) Generation of an actagardine A variant library through saturation mutagenesis. *Applied Microbiology and Biotechnology* 95: 1509-1517

Boc A, Diallo AB, Makarenkov V (2012) T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Research* 40: W573-W579

Boccard F, Smokvina T, Pernodet JL, Friedmann A, Guérineau M (1989) The integrated conjugative plasmid pSAM2 of *Streptomyces ambofaciens* is related to temperate bacteriophages. *EMBO J* 8: 973-980

Bradley SG, Ritzi D (1968) Composition and ultrastructure of *Streptomyces venezuelae*. *Journal of bacteriology* 95: 2358-2364

Brötz H, Bierbaum G, Leopold K, Reynolds PE, Sahl HG (1998) The lantibiotic mersacidin inhibits peptidoglycan synthesis by targeting lipid II. *Antimicrob Agents Chemother* 42: 154-160

Butler MS (2008) Natural products to drugs: natural product-derived compounds in clinical trials. *Nat Prod Rep* 25: 475-516

Butler T, Sridhar CB, Daga MK, Pathak K, Pandit RB, Khakhria R, Potkar CN, Zelasky MT, Johnson RB (1999) Treatment of typhoid fever with azithromycin versus chloramphenicol in a randomized multicentre trial in India. *Journal of Antimicrobial Chemotherapy* 44: 243-250

Caldera EJ, Poulsen M, Suen G, Currie CR (2009) Insect Symbioses: A Case Study of Past, Present, and Future Fungus-growing Ant Research. *Environmental Entomology* 38: 78-92

Chain E, Florey HW, Gardner AD, Heatley NG, Jennings MA, Orr-Ewing J, Sanders AG (1940) PENICILLIN AS A CHEMOTHERAPEUTIC AGENT. *The Lancet* 236: 226-228

Challinor VL, Bode HB (2015) Bioactive natural products from novel microbial sources. *Annals of the New York Academy of Sciences* 1354: 82-97

Chandra N, Kumar S (2017) Antibiotics Producing Soil Microorganisms. In: *Antibiotics and Antibiotics Resistance Genes in Soils: Monitoring, Toxicity, Risk Assessment and Management*, Hashmi M.Z., Strezov V., Varma A. (eds.) pp. 1-18. Springer International Publishing: Cham

Chattopadhyay MK (2014) Use of antibiotics as feed additives: a burning question. *Frontiers in Microbiology* 5

Chaudhary AK, Dhakal D, Sohng JK (2013) An insight into the "-omics" based engineering of streptomycetes for secondary metabolite overproduction. *Biomed Res Int* 2013: 968518

Chen E, Chen Q, Chen S, Xu B, Ju J, Wang H (2017a) Mathermycin, a Lantibiotic from the Marine Actinomycete *Marinactinospora thermotolerans* SCSIO 00652. *Appl Environ Microbiol* 83

Chen R, Wong HL, Burns BP (2019) New Approaches to Detect Biosynthetic Gene Clusters in the Environment. *Medicines (Basel)* 6: 32

Chen R, Zhang Q, Tan B, Zheng L, Li H, Zhu Y, Zhang C (2017b) Genome Mining and Activation of a Silent PKS/NRPS Gene Cluster Direct the Production of Totopotensamides. *Org Lett* 19: 5697-5700

Choung SY, Kobayashi T, Takemoto K, Ishitsuka H, Inoue K (1988) Interaction of a cyclic peptide, Ro09-0198, with phosphatidylethanolamine in liposomal membranes. *Biochim Biophys Acta* 940: 180-187

Christians U, Klawitter J, Hornberger A, Klawitter J (2011) How unbiased is non-targeted metabolomics and is targeted pathway screening the solution? *Curr Pharm Biotechnol* 12: 1053-1066

Clardy J, Fischbach MA, Currie CR (2009) The natural history of antibiotics. *Curr Biol* 19: R437-R441

Cong Y, Yang S, Rao X (2020) Vancomycin resistant Staphylococcus aureus infections: A review of case updating and clinical features. *Journal of Advanced Research* 21: 169-176

Cook TB, Jacobson TB, Venkataraman MV, Hofstetter H, Amador-Noguez D, Thomas MG, Pflieger BF (2021) Stepwise genetic engineering of Pseudomonas putida enables robust heterologous production of prodigiosin and glidobactin A. *Metabolic Engineering* 67: 112-124

Cotter G, Adley CC (2001) Comparison and evaluation of antimicrobial susceptibility testing of enterococci performed in accordance with six national committee standardized disk diffusion procedures. *J Clin Microbiol* 39: 3753-3756

Crawford JM, Clardy J (2011) Bacterial symbionts and natural products. *Chem Commun (Camb)* 47: 7559-7566

Currie CR (2001) A community of ants, fungi, and bacteria: a multilateral approach to studying symbiosis. *Annu Rev Microbiol* 55: 357-380

Currie CR, Wong B, Stuart AE, Schultz TR, Rehner SA, Mueller UG, Sung G-H, Spatafora JW, Straus NA (2003) Ancient Tripartite Coevolution in the Attine Ant-Microbe Symbiosis. *Science* 299: 386-388

Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* 97: 6640-6645

de Felício R, Ballone P, Bazzano CF, Alves LFG, Sigrist R, Infante GP, Niero H, Rodrigues-Costa F, Fernandes AZN, Tonon LAC et al (2021) Chemical Elicitors Induce Rare Bioactive Secondary Metabolites in Deep-Sea Bacteria under Laboratory Conditions. *Metabolites* 11

Devine R, McDonald HP, Qin Z, Arnold CJ, Noble K, Chandra G, Wilkinson B, Hutchings MI (2021) Re-wiring the regulation of the formicamycin biosynthetic gene cluster to enable the development of promising antibacterial compounds. *Cell Chemical Biology* 28: 515-523.e515

Donald L. Court, James A. Sawitzke a, Thomason LC (2002) Genetic Engineering Using Homologous Recombination. *Annual Review of Genetics* 36: 361-388

Donia MS, Cimermancic P, Schulze CJ, Wieland Brown LC, Martin J, Mitreva M, Clardy J, Linington RG, Fischbach MA (2014) A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* 158: 1402-1414

Dreier J, Khosla C (2000) Mechanistic analysis of a type II polyketide synthase. Role of conserved residues in the beta-ketoacyl synthase-chain length factor heterodimer. *Biochemistry* 39: 2088-2095

Duke T, Michael A, Mokela D, Wal T, Reeder J (2003) Chloramphenicol or ceftriaxone, or both, as treatment for meningitis in developing countries? *Archives of Disease in Childhood* 88: 536

Durfee T, Nelson R, Baldwin S, Plunkett G, Burland V, Mau B, Petrosino JF, Qin X, Muzny DM, Ayele M et al (2008) The Complete Genome Sequence of *Escherichia coli* DH10B: Insights into the Biology of a Laboratory Workhorse. *Journal of Bacteriology* 190: 2597-2606

Ehrlich P (1913) Address in Pathology, ON CHEMIOTHERAPY: Delivered before the Seventeenth International Congress of Medicine. *Br Med J* 2: 353-359

Elder FCT, Feil EJ, Snape J, Gaze WH, Kasprzyk-Hordern B (2020) The role of stereochemistry of antibiotic agents in the development of antibiotic resistance in the environment. *Environment International* 139: 105681

Elena C, Ravasi P, Castelli ME, Peirú S, Menzella HG (2014) Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. *Frontiers in microbiology* 5: 21-21

Elvas F, Stroobants S, Wyffels L (2017) Phosphatidylethanolamine targeting for cell death imaging in early treatment response evaluation and disease diagnosis. *Apoptosis : an international journal on programmed cell death* 22

Fair RJ, Tor Y (2014) Antibiotics and bacterial resistance in the 21st century. *Perspect Medicin Chem* 6: 25-64

Falah-Tafti A, Jafari AA, Lotfi-Kamran MH, Fallahzadeh H, Hayan RS (2010) A Comparison of the eFficacy of Nystatin and Fluconazole Incorporated into Tissue Conditioner on the In Vitro Attachment and Colonization of Candida Albicans. *Dent Res J (Isfahan)* 7: 18-22

Fayed B, Ashford DA, Hashem AM, Amin MA, El Gazayerly ON, Gregory MA, Smith MCM (2015) Multiplexed integrating plasmids for engineering of the erythromycin gene cluster for expression in Streptomyces spp. and combinatorial biosynthesis. *Applied and environmental microbiology* 81: 8402-8413

Fayed B, Younger E, Taylor G, Smith MC (2014) A novel Streptomyces spp. integration vector derived from the S. venezuelae phage, SV1. *BMC biotechnology* 14: 51

Fehr T, Sanglier JJ, Schuler W, Gschwind L, Ponelle M, Schilling W, Wioland C (1996) Antascomicins A, B, C, D and E - Novel FKBP12 binding compounds from a micromonospora a strain. *Journal of Antibiotics* 49: 230-233

Feng Z, Chakraborty D, Dewell SB, Reddy BVB, Brady SF (2012) Environmental DNA-Encoded Antibiotics Fasamycins A and B Inhibit FabF in Type II Fatty Acid Biosynthesis. *J Am Chem Soc* 134: 2981-2987

Fernández FJ, Vega MC (2016) Choose a Suitable Expression Host: A Survey of Available Protein Production Platforms. In: *Advanced Technologies for Protein Complex Production and Characterization*, Vega M.C. (ed.) pp. 15-24. Springer International Publishing: Cham

Field D, Molloy EM, Iancu C, Draper LA, O' Connor PM, Cotter PD, Hill C, Ross RP (2013) Saturation mutagenesis of selected residues of the α -peptide of the lantibiotic lacticin 3147 yields a derivative with enhanced antimicrobial activity. *Microbial biotechnology* 6: 564-575

Firth AE, Patrick WM (2005) Statistics of protein library construction. *Bioinformatics (Oxford, England)* 21: 3314-3315

Fischbach MA, Walsh CT (2006) Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* 106: 3468-3496

Fisher MC, Hawkins NJ, Sanglard D, Gurr SJ (2018) Worldwide emergence of resistance to antifungal drugs challenges human health and food security. *Science* 360: 739

Fjaervik E, Zotchev SB (2005) Biosynthesis of the polyene macrolide antibiotic nystatin in *Streptomyces noursei*. *Applied Microbiology and Biotechnology* 67: 436-443

Flardh K, Buttner MJ (2009) *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nature reviews Microbiology* 7: 36-49

Fleming A (1929) On the Antibacterial Action of Cultures of a *Penicillium*, with Special Reference to their Use in the Isolation of *B. influenzae*. *Br J Exp Pathol* 10: 226-236

Flett F, Mersinias V, Smith CP (1997) High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS Microbiol Lett* 155: 223-229

Flórez LV, Scherlach K, Gaube P, Ross C, Sitte E, Hermes C, Rodrigues A, Hertweck C, Kaltenpoth M (2017) Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect-defensive mutualism. *Nat Commun* 8: 15172-15172

Foulston L (2019) Genome mining and prospects for antibiotic discovery. *Current Opinion in Microbiology* 51: 1-8

Fredenhagen A, Fendrich G, Märki F, Märki W, Gruner J, Raschdorf F, Peter HH (1990) Duramycins B and C, two new lanthionine containing antibiotics as inhibitors of phospholipase A2. Structural revision of duramycin and cinnamycin. *The Journal of antibiotics* 43: 1403-1412

Fukuda TTH, Helfrich EJM, Mevers E, Melo WGP, Van Arnem EB, Andes DR, Currie CR, Pupo MT, Clardy J (2021) Specialized Metabolites Reveal Evolutionary History and Geographic Dispersion of a Multilateral Symbiosis. *ACS Central Science* 7: 292-299

Gaisser S, Carletti I, Schell U, Graupner PR, Sparks TC, Martin CJ, Wilkinson B (2009) Glycosylation engineering of spinosyn analogues containing an L-olivose moiety. *Org Biomol Chem* 7: 1705-1708

Gaisser S, Martin CJ, Wilkinson B, Sheridan RM, Lill RE, Weston AJ, Ready SJ, Waldron C, Crouse GD, Leadlay PF *et al* (2002) Engineered biosynthesis of novel spinosyns bearing altered deoxyhexose substituents. *Chem Commun (Camb)*: 618-619

Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL (1997) Arac/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev* 61: 393-410

Gao S-S, Naowarajna N, Cheng R, Liu X, Liu P (2018) Recent examples of α -ketoglutarate-dependent mononuclear non-haem iron enzymes in natural product biosyntheses. *Natural product reports* 35: 792-837

Gardner KAJA, Osawa M, Erickson HP (2017) Whole genome re-sequencing to identify suppressor mutations of mutant and foreign Escherichia coli FtsZ. *PLoS one* 12: e0176643-e0176643

Gaskins HR, Collier CT, Anderson DB (2002) ANTIBIOTICS AS GROWTH PROMOTANTS:MODE OF ACTION. *Animal Biotechnology* 13: 29-42

Gasser B, Saloheimo M, Rinas U, Dragosits M, Rodríguez-Carmona E, Baumann K, Giuliani M, Parrilli E, Branduardi P, Lang C et al (2008) Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microbial cell factories* 7: 11-11

Gaudencio SP, Pereira F (2015) Dereplication: racing to speed up the natural products discovery process. *Nat Prod Rep* 32: 779-810

Gebreyohannes G, Moges F, Sahile S, Raja N (2013) Isolation and characterization of potential antibiotic producing actinomycetes from water and sediments of Lake Tana, Ethiopia. *Asian Pac J Trop Biomed* 3: 426-435

Gharagozloo RA, Naficy K, Mouin M, Nassirzadeh MH, Yalda R (1970) Comparative trial of tetracycline, chloramphenicol, and trimethoprim-sulphamethoxazole in eradication of Vibrio cholerae El Tor. *Br Med J* 4: 281-282

Gharsallaoui A, Oulahal N, Joly C, Degraeve P (2016) Nisin as a Food Preservative: Part 1: Physicochemical Properties, Antimicrobial Activity, and Main Uses. *Crit Rev Food Sci Nutr* 56: 1262-1274

Gomez-Escribano JP, Bibb MJ (2011) Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters. *Microbial biotechnology* 4: 207-215

Gomez-Escribano JP, Castro JF, Razmilic V, Chandra G, Andrews B, Asenjo JA, Bibb MJ (2015) The Streptomyces leeuwenhoekii genome: de novo sequencing and assembly in single contigs of the chromosome, circular plasmid pSLE1 and linear plasmid pSLE2. *BMC Genomics* 16: 485

Goodfellow M, Nouioui I, Sanderson R, Xie F, Bull AT (2018) Rare taxa and dark microbial matter: novel bioactive actinobacteria abound in Atacama Desert soils. *Antonie van Leeuwenhoek* 111: 1315-1332

Gram L (2015) Silent clusters - speak up! *Microbial biotechnology* 8: 13-14

Gregory MA, Till R, Smith MCM (2003) Integration site for Streptomyces phage phiBT1 and development of site-specific integrating vectors. *Journal of bacteriology* 185: 5320-5323

Grigoreva A, Andreeva J, Bikmetov D, Rusanova A, Serebryakova M, Garcia AH, Slonova D, Nair SK, Lippens G, Severinov K et al (2021) Identification and characterization of andalusicin: N-terminally dimethylated class III lantibiotic from *Bacillus thuringiensis* sv. andalusiensis. *iScience* 24: 102480

Guo F, Xiang S, Li L, Wang B, Rajasarkka J, Grondahl-Yli-Hannuksela K, Ai G, Metsa-Ketela M, Yang K (2015) Targeted activation of silent natural product biosynthesis pathways by reporter-guided mutant selection. *Metab Eng* 28: 134-142

Gupta S, Yel L (2014) 10 - Molecular Biology and Genetic Engineering. In: *Middleton's Allergy (Eighth Edition)*, Adkinson N.F., Bochner B.S., Burks A.W., Busse W.W., Holgate S.T., Lemanske R.F., O'Hehir R.E. (eds.) pp. 162-183. W.B. Saunders: London

Gust B, Challis GL, Fowler K, Kieser T, Chater KF (2003) PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A* 100: 1541-1546

Gust B, Chandra G, Jakimowicz D, Yuqing T, Bruton CJ, Chater KF (2004) λ Red-Mediated Genetic Manipulation of Antibiotic-Producing Streptomyces. In: *Advances in Applied Microbiology*, pp. 107-128. Academic Press:

Gust B, Kieser T, Chater K (2002) PCR targeting system in *Streptomyces coelicolor* A3 (2). *John Innes Centre* 3: 1-39

Guzmán-Chávez F, Zwahlen RD, Bovenberg RAL, Driessen AJM (2018) Engineering of the Filamentous Fungus *Penicillium chrysogenum* as Cell Factory for Natural Products. *Frontiers in Microbiology* 9

Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166: 557-580

- Haney SA, Glasfeld E, Hale C, Keeney D, He Z, de Boer P** (2001) Genetic analysis of the Escherichia coli FtsZ.ZipA interaction in the yeast two-hybrid system. Characterization of FtsZ residues essential for the interactions with ZipA and with FtsA. *The Journal of biological chemistry* 276: 11980-11987
- Harman JW, Masterson JG** (2008) The mechanism of nystatin action. *Irish Journal of Medical Science (1926-1967)* 32: 249
- Hartl FU, Hayer-Hartl M** (2002) Molecular Chaperones in the Cytosol: from Nascent Chain to Folded Protein. *Science* 295: 1852-1858
- Hayashi F, Nagashima K, Terui Y, Kawamura Y, Matsumoto K, Itazaki H** (1990) The structure of PA48009: the revised structure of duramycin. *The Journal of antibiotics* 43: 1421-1430
- Hecht M, Bromberg Y, Rost B** (2013) News from the protein mutability landscape. *J Mol Biol* 425: 3937-3948
- Hegemann JD, Süßmuth RD** (2020) Matters of class: coming of age of class III and IV lanthipeptides. *RSC Chemical Biology* 1: 110-127
- Hertweck C, Luzhetskyy A, Rebets Y, Bechthold A** (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat Prod Rep* 24: 162-190
- Hietpas RT, Jensen JD, Bolon DN** (2011) Experimental illumination of a fitness landscape. *Proc Natl Acad Sci U S A* 108: 7896-7901
- Hodgkin DC** (1949) The X-ray analysis of the structure of penicillin. *Adv Sci* 6: 85-89
- Hollis A, Ahmed Z** (2013) Preserving antibiotics, rationally. *N Engl J Med* 369: 2474-2476
- Holmes NA, Innocent TM, Heine D, Al Bassam M, Worsley SF, Trottmann F, Patrick EH, Yu DW, Murrell JC, Schiott M *et al*** (2016) Genome Analysis of Two Pseudonocardia Phylotypes Associated with Acromyrmex Leafcutter Ants Reveals Their Biosynthetic Potential. *Frontiers in Microbiology* 7
- Hong HJ, Hutchings MI, Hill LM, Buttner MJ** (2005) The role of the novel Fem protein VanK in vancomycin resistance in Streptomyces coelicolor. *The Journal of biological chemistry* 280: 13055-13061

- Hong J** (2014) Natural product synthesis at the interface of chemistry and biology. *Chemistry* 20: 10204-10212
- Horbal L, Siegl T, Luzhetskyy A** (2018) A set of synthetic versatile genetic control elements for the efficient expression of genes in Actinobacteria. *Scientific reports* 8: 491-491
- Horgan RP, Kenny LC** (2011) 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist* 13: 189-195
- Hosoda K, Ohya M, Kohno T, Maeda T, Endo S, Wakamatsu K** (1996) Structure determination of an immunopotentiator peptide, cinnamycin, complexed with lysophosphatidylethanolamine by 1H-NMR1. *J Biochem* 119: 226-230
- Hu H, Zhang Q, Ochi K** (2002) Activation of antibiotic biosynthesis by specified mutations in the rpoB gene (encoding the RNA polymerase beta subunit) of *Streptomyces lividans*. *J Bacteriol* 184: 3984-3991
- Hullin-Matsuda F, Makino A, Murate M, Kobayashi T** (2016) Probing phosphoethanolamine-containing lipids in membranes with duramycin/cinnamycin and aegerolysin proteins. *Biochimie* 130: 81-90
- Huo L, Ökesli A, Zhao M, van der Donk WA** (2017) Insights into the Biosynthesis of Duramycin. *Applied and environmental microbiology* 83: e02698-02616
- Hutchings MI, Truman AW, Wilkinson B** (2019) Antibiotics: past, present and future. *Current Opinion in Microbiology* 51: 72-80
- Islam MR, Shioya K, Nagao J, Nishie M, Jikuya H, Zendo T, Nakayama J, Sonomoto K** (2009) Evaluation of essential and variable residues of nukacin ISK-1 by NNK scanning. *Molecular Microbiology* 72: 1438-1447
- Jackson MP, Hewitt EW** (2016) Cellular proteostasis: degradation of misfolded proteins by lysosomes. *Essays Biochem* 60: 173-180
- Jang KH, Nam S-J, Locke JB, Kauffman CA, Beatty DS, Paul LA, Fenical W** (2013) Anthracimycin, a potent anthrax antibiotic from a marine-derived actinomycete. *Angewandte Chemie (International ed in English)* 52: 7822-7824
- Johnston CW, Skinnider MA, Wyatt MA, Li X, Ranieri MRM, Yang L, Zechel DL, Ma B, Magarvey NA** (2015) An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. *Nat Commun* 6: 8421

Jones AC, Gust B, Kulik A, Heide L, Buttner MJ, Bibb MJ (2013) Phage P1-Derived Artificial Chromosomes Facilitate Heterologous Expression of the FK506 Gene Cluster. *Plos One* 8

Jones MB, Nierman WC, Shan Y, Frank BC, Spoering A, Ling L, Peoples A, Zullo A, Lewis K, Nelson KE (2017) Reducing the Bottleneck in Discovery of Novel Antibiotics. *Microbial Ecology* 73: 658-667

Kakkar S, Narasimhan B (2019) A comprehensive review on biological activities of oxazole derivatives. *BMC Chem* 13: 16-16

Kasten FH (1996) Paul Ehrlich: Pathfinder in Cell Biology 1. Chronicle of His Life and Accomplishments in Immunology, Cancer Research, and Chemotherapy1: Paul Ehrlich's Recipe for Success: "Patience, Ability, Money and Luck". *Biotechnic & Histochemistry* 71: 2-37

Katsuyama Y, Ohnishi Y (2012) Type III polyketide synthases in microorganisms. *Methods Enzymol* 515: 359-377

Katz L, Baltz RH (2016) Natural product discovery: past, present, and future. *J Ind Microbiol Biotechnol* 43: 155-176

Ke J, Yoshikuni Y (2020) Multi-chassis engineering for heterologous production of microbial natural products. *Current Opinion in Biotechnology* 62: 88-97

Kessler H, Steuernagel S, Will M, Jung G, Kellner R, Gillessen D, Kamiyama T (1988) The Structure of the Polycyclic Nonadecapeptide Ro 09-0198. *Helvetica Chimica Acta* 71: 1924-1929

Khaw LE, Böhm GA, Metcalfe S, Staunton J, Leadlay PF (1998) Mutational biosynthesis of novel rapamycins by a strain of *Streptomyces hygroscopicus* NRRL 5491 disrupted in *rapL*, encoding a putative lysine cyclodeaminase. *J Bacteriol* 180: 809-814

Kido Y, Hamakado T, Yoshida T, Anno M, Motoki Y, Wakamiya T, Shiba T (1983) Isolation and characterization of ancovenin, a new inhibitor of angiotensin I converting enzyme, produced by actinomycetes. *The Journal of antibiotics* 36: 1295-1299

Kim BG, Lee MJ, Seo J, Hwang YB, Lee MY, Han K, Sherman DH, Kim ES (2009) Identification of functionally clustered nystatin-like biosynthetic genes in a rare

actinomycetes, *Pseudonocardia autotrophica*. *Journal of Industrial Microbiology & Biotechnology* 36: 1425-1434

Kim H-J, Kim M-K, Lee M-J, Won H-J, Choi S-S, Kim E-S (2015) Post-PKS Tailoring Steps of a Disaccharide-Containing Polyene NPP in *Pseudonocardia autotrophica*. *PLOS ONE* 10: e0123270

Kircher M, Xiong C, Martin B, Schubach M, Inoue F, Bell RJA, Costello JF, Shendure J, Ahituv N (2019) Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. *Nat Commun* 10: 3583

Kirchhelle C (2018) Pharming animals: a global history of antibiotics in food production (1935–2017). *Palgrave Communications* 4: 96

Kirst HA (2010) The spinosyn family of insecticides: realizing the potential of natural products research. *The Journal of antibiotics* 63: 101-111

Klein EY, Van Boeckel TP, Martinez EM, Pant S, Gandra S, Levin SA, Goossens H, Laxminarayan R (2018) Global increase and geographic convergence in antibiotic consumption between 2000 and 2015. *Proceedings of the National Academy of Sciences* 115: E3463-E3470

Klesmith JR, Bacik JP, Michalczyk R, Whitehead TA (2015) Comprehensive Sequence-Flux Mapping of a Levoglucosan Utilization Pathway in *E. coli*. *ACS Synth Biol* 4: 1235-1243

Knappe TA, Linne U, Zirah S, Rebuffat S, Xie X, Marahiel MA (2008) Isolation and Structural Characterization of Capistruin, a Lasso Peptide Predicted from the Genome Sequence of *Burkholderia thailandensis* E264. *J Am Chem Soc* 130: 11446-11454

Knerr PJ, Donk WA (2012) Discovery, Biosynthesis, and Engineering of Lantipeptides. *Annual Review of Biochemistry* 81: 479-505

Kodani S, Komaki H, Ishimura S, Hemmi H, Ohnishi-Kameyama M (2016) Isolation and structure determination of a new lantibiotic cinnamycin B from *Actinomadura atramentaria* based on genome mining. *J Ind Microbiol Biotechnol* 43: 1159-1165

Koehler S, Doubský J, Kaltenpoth M (2013) Dynamics of symbiont-mediated antibiotic production reveal efficient long-term protection for beewolf offspring. *Frontiers in Zoology* 10: 3

Koehn FE, Carter GT (2005) The evolving role of natural products in drug discovery. *Nat Rev Drug Discov* 4: 206-220

Kolaj O, Spada S, Robin S, Wall JG (2009) Use of folding modulators to improve heterologous protein production in Escherichia coli. *Microbial cell factories* 8: 9-9

Krause J, Handayani I, Blin K, Kulik A, Mast Y (2020) Disclosing the Potential of the SARP-Type Regulator PapR2 for the Activation of Antibiotic Gene Clusters in Streptomyces. *Frontiers in Microbiology* 11

Kries H (2016) Biosynthetic engineering of nonribosomal peptide synthetases. *Journal of Peptide Science* 22: 564-570

Krüger M, Richter P, Strauch SM, Nasir A, Burkovski A, Antunes CA, Meißgeier T, Schlücker E, Schwab S, Lebert M (2019) What an Escherichia coli Mutant Can Teach Us About the Antibacterial Effect of Chlorophyllin. *Microorganisms* 7: 59

Lewis K (2013) Platforms for antibiotic discovery. *Nature Reviews Drug Discovery* 12: 371-387

Li S, Li Y, Lu C, Zhang J, Zhu J, Wang H, Shen Y (2015) Activating a Cryptic Ansamycin Biosynthetic Gene Cluster To Produce Three New Naphthalenic Octaketide Ansamycins with n-Pentyl and n-Butyl Side Chains. *Org Lett* 17: 3706-3709

Lindenbaum J, Greenough WB, Islam MR (1967) Antibiotic therapy of cholera. *Bull World Health Organ* 36: 871-883

Lindenfelser LA, Pridham TG, Kemp CE (1959) Antibiotics against plant disease. V. Activity of cinnamycin against selected microorganisms. *Antibiot Chemother (Northfield)* 9: 690-695

Little AEF, Currie CR (2009) Parasites may help stabilize cooperative relationships. *BMC Evolutionary Biology* 9: 124

Liu G, Chater KF, Chandra G, Niu G, Tan H (2013) Molecular regulation of antibiotic biosynthesis in streptomyces. *Microbiol Mol Biol Rev* 77: 112-143

Liu SH, Wang W, Wang KB, Zhang B, Li W, Shi J, Jiao RH, Tan RX, Ge HM (2019) Heterologous Expression of a Cryptic Giant Type I PKS Gene Cluster Leads to the Production of Ansaseomycin. *Organic Letters* 21: 3785-3788

Liu X, Liu D, Xu M, Tao M, Bai L, Deng Z, Pfeifer BA, Jiang M (2018) Reconstitution of Kinamycin Biosynthesis within the Heterologous Host *Streptomyces albus* J1074. *Journal of Natural Products* 81: 72-77

Loenarz C, Schofield CJ (2011) Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem Sci* 36: 7-18

Lopatniuk M, Myronovskyi M, Luzhetskyy A (2017) *Streptomyces albus*: A New Cell Factory for Non-Canonical Amino Acids Incorporation into Ribosomally Synthesized Natural Products. *ACS Chem Biol* 12: 2362-2370

Lubelski J, Rink R, Khusainov R, Moll GN, Kuipers OP (2008) Biosynthesis, immunity, regulation, mode of action and engineering of the model lantibiotic nisin. *Cellular and Molecular Life Sciences* 65: 455-476

Lyddiard D, Jones GL, Greatrex BW (2016) Keeping it simple: lessons from the golden era of antibiotic discovery. *FEMS Microbiol Lett* 363

Lyu X, Zhao C, Yan Z-M, Hua H (2016) Efficacy of nystatin for the treatment of oral candidiasis: a systematic review and meta-analysis. *Drug Des Devel Ther* 10: 1161-1171

Ma X, Margolin W (1999) Genetic and functional analyses of the conserved C-terminal core domain of *Escherichia coli* FtsZ. *J Bacteriol* 181: 7531-7544

MacNeil DJ, Gewain KM, Ruby CL, Dezeny G, Gibbons PH, MacNeil T (1992) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* 111: 61-68

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD et al (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 47: W636-w641

Maier ME (2015) Design and synthesis of analogues of natural products. *Organic & Biomolecular Chemistry* 13: 5302-5343

Mainous AG, 3rd, Everett CJ, Post RE, Diaz VA, Hueston WJ (2009) Availability of antibiotics for purchase without a prescription on the internet. *Ann Fam Med* 7: 431-435

Malpartida F, Hopwood DA (1984) Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. *Nature* 309: 462-464

Marahiel MA (2016) A structural model for multimodular NRPS assembly lines. *Natural Product Reports* 33: 136-140

Maris AE, Sawaya MR, Kaczor-Grzeskowiak M, Jarvis MR, Bearson SM, Kopka ML, Schröder I, Gunsalus RP, Dickerson RE (2002) Dimerization allows DNA target site recognition by the NarL response regulator. *Nat Struct Biol* 9: 771-778

Märki F, Hänni E, Fredenhagen A, van Oostrum J (1991) Mode of action of the lanthionine-containing peptide antibiotics duramycin, duramycin B and C, and cinnamycin as indirect inhibitors of phospholipase A2. *Biochemical Pharmacology* 42: 2027-2035

Martínez-Núñez MA, López VELY (2016) Nonribosomal peptides synthetases and their applications in industry. *Sustainable Chemical Processes* 4: 13

McAuliffe O, O'Keeffe T, Hill C, Ross RP (2001a) Regulation of immunity to the two-component lantibiotic, lactacin 3147, by the transcriptional repressor LtnR. *Mol Microbiol* 39: 982-993

McAuliffe O, Ross RP, Hill C (2001b) Lantibiotics: structure, biosynthesis and mode of action. *FEMS Microbiology Reviews* 25: 285-308

McEwen SA, Fedorka-Cray PJ (2002) Antimicrobial use and resistance in animals. *Clin Infect Dis* 34 Suppl 3: S93-s106

McLean TC, Wilkinson B, Hutchings MI, Devine R (2019) Dissolution of the Disparate: Co-ordinate Regulation in Antibiotic Biosynthesis. *Antibiotics (Basel)* 8

Melnikov A, Rogov P, Wang L, Gnirke A, Mikkelsen TS (2014) Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res* 42: e112

Milazzo G, Mercatelli D, Di Muzio G, Triboli L, De Rosa P, Perini G, Giorgi FM (2020) Histone Deacetylases (HDACs): Evolution, Specificity, Role in Transcriptional Complexes, and Pharmacological Actionability. *Genes* 11

- Mincer TJ, Jensen PR, Kauffman CA, Fenical W** (2002) Widespread and Persistent Populations of a Major New Marine Actinomycete Taxon in Ocean Sediments. *Applied and Environmental Microbiology* 68: 5005-5011
- Minogue TD, Daligault HA, Davenport KW, Bishop-Lilly KA, Broomall SM, Bruce DC, Chain PS, Chertkov O, Coyne SR, Freitas T et al** (2014) Complete Genome Assembly of Escherichia coli ATCC 25922, a Serotype O6 Reference Strain. *Genome announcements* 2: e00969-00914
- Moll GN, Roberts GCK, Konings WN, Driessen AJM** (1996) Mechanism of lantibiotic-induced pore-formation. *Antonie van Leeuwenhoek* 69: 185-191
- Molloy EM, Field D, O' Connor PM, Cotter PD, Hill C, Ross RP** (2013) Saturation mutagenesis of lysine 12 leads to the identification of derivatives of nisin A with enhanced antimicrobial activity. *PloS one* 8: e58530-e58530
- Montalbán-López M, Scott TA, Ramesh S, Rahman IR, van Heel AJ, Viel JH, Bandarian V, Dittmann E, Genilloud O, Goto Y et al** (2021) New developments in RiPP discovery, enzymology and engineering. *Natural Product Reports* 38: 130-239
- Moore JM, Bradshaw E, Seipke RF, Hutchings MI, McArthur M** (2012) Chapter Eighteen - Use and Discovery of Chemical Elicitors That Stimulate Biosynthetic Gene Clusters in Streptomyces Bacteria. In: *Methods in Enzymology*, Hopwood D.A. (ed.) pp. 367-385. Academic Press:
- Mowatt MR, Luján HD, Cotten DB, Bowers B, Yee J, Nash TE, Stibbs HH** (1995) Developmentally regulated expression of a Giardia lamblia cyst wall protein gene. *Mol Microbiol* 15: 955-963
- Mueller UG, Schultz TR, Currie CR, Adams RMM, Malloch D** (2001) The Origin of the Attine Ant-Fungus Mutualism. *The Quarterly Review of Biology* 76: 169-197
- Munita JM, Arias CA** (2016) Mechanisms of Antibiotic Resistance. *Microbiol Spectr* 4: 10.1128/microbiolspec.VMBF-0016-2015
- Myronovskiy M, Luzhetskyy A** (2019) Heterologous production of small molecules in the optimized Streptomyces hosts. *Natural Product Reports* 36: 1281-1294
- Nah HJ, Pyeon HR, Kang SH, Choi SS, Kim ES** (2017) Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in Streptomyces Species. *Front Microbiol* 8

- Nepal G, Bhatta S** (2018) Self-medication with Antibiotics in WHO Southeast Asian Region: A Systematic Review. *Cureus* 10: e2428-e2428
- Nishihara K, Kanemori M, Yanagi H, Yura T** (2000) Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. *Applied and environmental microbiology* 66: 884-889
- Nishikawa M, Kobayashi K** (2009) Streptomyces roseoverticillatus produces two different poly(amino acid)s: lariat-shaped γ -poly(L-glutamic acid) and ϵ -poly(L-lysine). *Microbiology* 155: 2988-2993
- Norby B** (2009) CHAPTER 98 - Antimicrobial Resistance in Human Pathogens and the Use of Antimicrobials in Food Animals: Challenges in Food Animal Veterinary Practice. In: *Food Animal Practice (Fifth Edition)*, Anderson D.E., Rings D.M. (eds.) pp. 479-488. W.B. Saunders: Saint Louis
- O'Rourke S, Widdick D, Bibb M** (2017) A novel mechanism of immunity controls the onset of cinnamycin biosynthesis in Streptomyces cinnamoneus DSM 40646. *Journal of Industrial Microbiology & Biotechnology* 44: 563-572
- Offe HA** (1988) Historical Introduction and Chemical Characteristics of Antituberculosis Drugs. In: *Antituberculosis Drugs*, Bartmann K. (ed.) pp. 1-30. Springer Berlin Heidelberg: Berlin, Heidelberg
- Ogasawara Y, Yackley BJ, Greenberg JA, Rogelj S, Melançon CE, III** (2015) Expanding our Understanding of Sequence-Function Relationships of Type II Polyketide Biosynthetic Gene Clusters: Bioinformatics-Guided Identification of Frankiamicin A from Frankia sp. EAN1pec. *PLOS ONE* 10: e0121505
- Oh DC, Poulsen M, Currie CR, Clardy J** (2009) Dentigerumycin: a bacterial mediator of an ant-fungus symbiosis. *Nature Chemical Biology* 5: 391-393
- Okada B, Seyedsayamdost M** (2016) Antibiotic dialogues: Induction of silent biosynthetic gene clusters by exogenous small molecules. *FEMS Microbiology Reviews* 41: fuw035
- Okada BK, Wu Y, Mao D, Bushin LB, Seyedsayamdost MR** (2016) Mapping the Trimethoprim-Induced Secondary Metabolome of Burkholderia thailandensis. *ACS Chemical Biology* 11: 2124-2130
- Okamoto-Hosoya Y, Sato TA, Ochi K** (2000) Resistance to paromomycin is conferred by rpsL mutations, accompanied by an enhanced antibiotic production in Streptomyces coelicolor A3(2). *The Journal of antibiotics* 53: 1424-1427

Ökesli A, Cooper LE, Fogle EJ, van der Donk WA (2011) Nine post-translational modifications during the biosynthesis of cinnamycin. *J Am Chem Soc* 133: 13753-13760

Olano C, García I, González A, Rodriguez M, Rozas D, Rubio J, Sánchez-Hidalgo M, Braña AF, Méndez C, Salas JA (2014) Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074. *Microbial biotechnology* 7: 242-256

Oldach F, Al Toma R, Kuthning A, Caetano T, Mendo S, Budisa N, Süßmuth RD (2012) Congeneric Lantibiotics from Ribosomal In Vivo Peptide Synthesis with Noncanonical Amino Acids. *Angewandte Chemie International Edition* 51: 415-418

Oliynyk I, Varelogianni G, Roomans GM, Johannesson M (2010) Effect of duramycin on chloride transport and intracellular calcium concentration in cystic fibrosis and non-cystic fibrosis epithelia. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica* 118: 982-990

Oliynyk M, Samborskyy M, Lester JB, Mironenko T, Scott N, Dickens S, Haydock SF, Leadlay PF (2007) Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nature Biotechnology* 25: 447-453

Oman TJ, van der Donk WA (2010) Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nature chemical biology* 6: 9-18

Ongley SE, Bian X, Neilan BA, Müller R (2013) Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Natural Product Reports* 30: 1121-1138

Orth P, Schnappinger D, Hillen W, Saenger W, Hinrichs W (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat Struct Biol* 7: 215-219

Palazzotto E, Tong Y, Lee SY, Weber T (2019) Synthetic biology and metabolic engineering of actinomycetes for natural product discovery. *Biotechnology Advances* 37: 107366

Pan YL, Yang X, Li J, Zhang RF, Hu YF, Zhou YG, Wang J, Zhu BL (2011) Genome Sequence of the Spinosyns-Producing Bacterium *Saccharopolyspora spinosa* NRRL 18395. *Journal of Bacteriology* 193: 3150-3151

Park PJ (2009) ChIP–seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics* 10: 669-680

Patrick WM, Firth AE, Blackburn JM (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries Wayne M. Patrick and Andrew E. Firth contributed equally to this work. *Protein Engineering, Design and Selection* 16: 451-457

Payne JAE, Schoppet M, Hansen MH, Cryle MJ (2017) Diversity of nature's assembly lines – recent discoveries in non-ribosomal peptide synthesis. *Molecular BioSystems* 13: 9-22

Peláez F (2006) The historical delivery of antibiotics from microbial natural products—Can history repeat? *Biochemical Pharmacology* 71: 981-990

Périchon B, Courvalin P (2009a) Antibiotic Resistance. In: *Encyclopedia of Microbiology (Third Edition)*, Schaechter M. (ed.) pp. 193-204. Academic Press: Oxford

Périchon B, Courvalin P (2009b) VanA-Type Vancomycin-Resistant *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy* 53: 4580-4587

Pernodet JL, Simonet JM, Guérineau M (1984) Plasmids in different strains of *Streptomyces ambofaciens*: free and integrated form of plasmid pSAM2. *Mol Gen Genet* 198: 35-41

Pertsemlidis A, Fondon JW (2001) Having a BLAST with bioinformatics (and avoiding BLASTphemy). *Genome Biology* 2: reviews2002.2001

Pham VTT, Nguyen CT, Dhakal D, Nguyen HT, Kim T-S, Sohng JK (2021) Recent Advances in the Heterologous Biosynthesis of Natural Products from *Streptomyces*. *Applied Sciences* 11: 1851

Pickens LB, Tang Y, Chooi Y-H (2011) Metabolic engineering for the production of natural products. *Annu Rev Chem Biomol Eng* 2: 211-236

Piffaretti J-C, Arini A, Frey J (1988) pUB307 mobilizes resistance plasmids from *Escherichia coli* into *Neisseria gonorrhoeae*. *Molecular and General Genetics MGG* 212: 215-218

Pishchany G, Kolter R (2020) On the possible ecological roles of antimicrobials. *Mol Microbiol* 113: 580-587

Poteete AR (2001) What makes the bacteriophage λ Red system useful for genetic engineering: molecular mechanism and biological function. *FEMS Microbiology Letters* 201: 9-14

Prasanna A, Niranjana V (2019) Classification of Mycobacterium tuberculosis DR, MDR, XDR Isolates and Identification of Signature Mutation Pattern of Drug Resistance. *Bioinformatics* 15: 261-268

Prestinaci F, Pezzotti P, Pantosti A (2015) Antimicrobial resistance: a global multifaceted phenomenon. *Pathog Glob Health* 109: 309-318

Qin Y, Wang Y, He Y, Zhang Y, She Q, Chai Y, Li P, Shang Q (2019a) Characterization of Subtilin L-Q11, a Novel Class I Bacteriocin Synthesized by Bacillus subtilis L-Q11 Isolated From Orchard Soil. *Frontiers in Microbiology* 10

Qin Z, Devine R, Booth TJ, Farrar EHE, Grayson MN, Hutchings MI, Wilkinson B (2020) Formicamycin biosynthesis involves a unique reductive ring contraction. *Chemical Science* 11: 8125-8131

Qin Z, Devine R, Hutchings MI, Wilkinson B (2019b) A role for antibiotic biosynthesis monooxygenase domain proteins in fidelity control during aromatic polyketide biosynthesis. *Nat Commun* 10: 3611

Qin Z, Munnoch JT, Devine R, Holmes NA, Seipke RF, Wilkinson KA, Wilkinson B, Hutchings MI (2017) Formicamycins, antibacterial polyketides produced by Streptomyces formicae isolated from African Tetraponera plant-ants. *Chemical Science* 8: 3218-3227

Ray S, Dastidar SG, Chakrabarty AN (1980) Antibiotic cross-resistance patterns of ambodryl and promazine resistant mutants. *Br J Exp Pathol* 61: 465-470

Reen FJ, Romano S, Dobson ADW, O'Gara F (2015) The Sound of Silence: Activating Silent Biosynthetic Gene Clusters in Marine Microorganisms. *Marine Drugs* 13: 4754-4783

Ren H, Shi C, Bothwell IR, van der Donk WA, Zhao H (2020) Discovery and Characterization of a Class IV Lanthipeptide with a Nonoverlapping Ring Pattern. *ACS Chemical Biology* 15: 1642-1649

Ren H, Wang B, Zhao H (2017) Breaking the silence: new strategies for discovering novel natural products. *Current opinion in biotechnology* 48: 21-27

Repka LM, Chekan JR, Nair SK, van der Donk WA (2017) Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes. *Chemical Reviews* 117: 5457-5520

Reygaert WC (2018) An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiol* 4: 482-501

Romero-Rodríguez A, Robledo-Casados I, Sánchez S (2015) An overview on transcriptional regulators in *Streptomyces*. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1849: 1017-1039

Rosove L, Chudnoff JS, Bower AG (1950) Chloramphenicol in the treatment of typhoid fever. *Calif Med* 72: 425-430

Ruiz N, Falcone B, Kahne D, Silhavy TJ (2005) Chemical conditionality: a genetic strategy to probe organelle assembly. *Cell* 121: 307-317

Rutledge PJ, Challis GL (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature reviews Microbiology* 13: 509-523

Sandegren L (2014) Selection of antibiotic resistance at very low antibiotic concentrations. *Ups J Med Sci* 119: 103-107

Sandmann G (2002) Combinatorial Biosynthesis of Carotenoids in a Heterologous Host: A Powerful Approach for the Biosynthesis of Novel Structures. *ChemBioChem* 3: 629-635

Santos-Aberturas J, Payero TD, Vicente CM, Guerra SM, Cañibano C, Martín JF, Aparicio JF (2011) Functional conservation of PAS-LuxR transcriptional regulators in polyene macrolide biosynthesis. *Metab Eng* 13: 756-767

Schatz A, Bugie E, Waksman SA (2005) Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. 1944. *Clin Orthop Relat Res*: 3-6

Schell U, Haydock SF, Kaja AL, Carletti I, Lill RE, Read E, Sheehan LS, Low L, Fernandez MJ, Grolle F et al (2008) Engineered biosynthesis of hybrid macrolide polyketides containing D-angolosamine and D-mycaminose moieties. *Org Biomol Chem* 6: 3315-3327

Schrijver A, Mot R (1999) A subfamily of MalT-related ATP-dependent regulators in the LuxR family. *Microbiology (Reading)* 145 (Pt 6): 1287-1288

Sebak M, Saafan AE, AbdelGhani S, Bakeer W, El-Gendy AO, Espriu LC, Duncan K, Edrada-Ebel R (2020) Bioassay- and metabolomics-guided screening of bioactive soil actinomycetes from the ancient city of Ichnasia, Egypt. *PLOS ONE* 14: e0226959

Sekurova ON, Schneider O, Zotchev SB (2019) Novel bioactive natural products from bacteria via bioprospecting, genome mining and metabolic engineering. *Microbial biotechnology* 12: 828-844

Seyedsayamdost MR (2014) High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proceedings of the National Academy of Sciences of the United States of America* 111: 7266-7271

Shallcross LJ, Davies DSC (2014) Antibiotic overuse: a key driver of antimicrobial resistance. *British Journal of General Practice* 64: 604-605

Shen B (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology* 7: 285-295

Shi T, Guo X, Zhu J, Hu L, He Z, Jiang D (2021) Inhibitory Effects of Carbazomycin B Produced by *Streptomyces roseovorticillatus* 63 Against *Xanthomonas oryzae* pv. *oryzae*. *Frontiers in Microbiology* 12

Shigdel UK, Lee S-J, Sowa ME, Bowman BR, Robison K, Zhou M, Pua KH, Stiles DT, Blodgett JAV, Udway DW et al (2020) Genomic discovery of an evolutionarily programmed modality for small-molecule targeting of an intractable protein surface. *Proceedings of the National Academy of Sciences* 117: 17195-17203

Shima J, Hesketh A, Okamoto S, Kawamoto S, Ochi K (1996) Induction of actinorhodin production by rpsL (encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans* and *Streptomyces coelicolor* A3(2). *J Bacteriol* 178: 7276-7284

Silberstein S (1924) Zur Frage der salvarsanresistenten Lues. *Archiv für Dermatologie und Syphilis* 147: 116-130

Siloto RMP, Weselake RJ (2012) Site saturation mutagenesis: Methods and applications in protein engineering. *Biocatalysis and Agricultural Biotechnology* 1: 181-189

Skinnider MA, Dejong CA, Rees PN, Johnston CW, Li H, Webster AL, Wyatt MA, Magarvey NA (2015) Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Res* 43: 9645-9662

Smith JD, McManus KF, Fraser HB (2013) A novel test for selection on cis-regulatory elements reveals positive and negative selection acting on mammalian transcriptional enhancers. *Mol Biol Evol* 30: 2509-2518

Solá RJ, Griebenow K (2009) Effects of glycosylation on the stability of protein pharmaceuticals. *J Pharm Sci* 98: 1223-1245

Sorokina M, Steinbeck C (2020) Review on natural products databases: where to find data in 2020. *J Cheminform* 12: 20

Sosio M, Giusino F, Cappellano C, Bossi E, Puglia AM, Donadio S (2000) Artificial chromosomes for antibiotic-producing actinomycetes. *Nat Biotechnol* 18: 343-345

Sowani H, Kulkarni M, Zinjarde S, Javdekar V (2017) Chapter 7 - *Gordonia* and Related Genera as Opportunistic Human Pathogens Causing Infections of Skin, Soft Tissues, and Bones. In: *The Microbiology of Skin, Soft Tissue, Bone and Joint Infections*, Kon K., Rai M. (eds.) pp. 105-121. Academic Press:

Stothard P (2000) The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28: 1102-1104

Suckau D, Resemann A, Schuerenberg M, Hufnagel P, Franzen J, Holle A (2003) A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Anal Bioanal Chem* 376: 952-965

Tabata T, Petitt M, Puerta-Guardo H, Michlmayr D, Wang C, Fang-Hoover J, Harris E, Pereira L (2016) Zika Virus Targets Different Primary Human Placental Cells, Suggesting Two Routes for Vertical Transmission. *Cell host & microbe* 20: 155-166

Tams JW, Vind J, Welinder KG (1999) Adapting protein solubility by glycosylation.: N-Glycosylation mutants of *Coprinus cinereus* peroxidase in salt and organic solutions. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* 1432: 214-221

Teijaro CN, Adhikari A, Shen B (2019) Challenges and opportunities for natural product discovery, production, and engineering in native producers versus heterologous hosts. *Journal of industrial microbiology & biotechnology* 46: 433-444

- Tocchetti A, Donadio S, Sosio M** (2018) Large inserts for big data: artificial chromosomes in the genomic era. *FEMS Microbiology Letters* 365
- Tong M, Jiang Y** (2015) FK506-Binding Proteins and Their Diverse Functions. *Curr Mol Pharmacol* 9: 48-65
- Um S, Fraimout A, Sapountzis P, Oh DC, Poulsen M** (2013) The fungus-growing termite *Macrotermes natalensis* harbors bacillaene-producing *Bacillus* sp. that inhibit potentially antagonistic fungi. *Sci Rep* 3: 3250
- Van Arnam EB, Currie CR, Clardy J** (2017) Defense contracts: molecular protection in insect-microbe symbioses. *Chemical Society reviews*
- van der Donk WA, Nair SK** (2014) Structure and mechanism of lanthipeptide biosynthetic enzymes. *Current Opinion in Structural Biology* 29: 58-66
- van der Meer JY, Biewenga L, Poelarends GJ** (2016) The Generation and Exploitation of Protein Mutability Landscapes for Enzyme Engineering. *Chembiochem* 17: 1792-1799
- van Heel AJ, de Jong A, Song C, Viel JH, Kok J, Kuipers OP** (2018) BAGEL4: a user-friendly web server to thoroughly mine RIPPs and bacteriocins. *Nucleic acids research* 46: W278-W281
- Ventola CL** (2015) The antibiotic resistance crisis: part 1: causes and threats. *P T* 40: 277-283
- Vestergaard M, Berglund NA, Hsu P-C, Song C, Koldsø H, Schiøtt B, Sansom MSP** (2019) Structure and Dynamics of Cinnamycin–Lipid Complexes: Mechanisms of Selectivity for Phosphatidylethanolamine Lipids. *ACS Omega* 4: 18889-18899
- Vicente MF, Basilio A, Cabello A, Peláez F** (2003) Microbial natural products as a source of antifungals. *Clinical Microbiology and Infection* 9: 15-32
- Vikeli E, Widdick DA, Batey SFD, Heine D, Holmes NA, Bibb MJ, Martins DJ, Pierce NE, Hutchings MI, Wilkinson B** (2020) *In Situ* Activation and Heterologous Production of a Cryptic Lantibiotic from an African Plant Ant-Derived *Saccharopolyspora* Species. *Applied and Environmental Microbiology* 86: e01876-01819

Viswanathan VK (2014) Off-label abuse of antibiotics by bacteria. *Gut Microbes* 5: 3-4

Wakamatsu K, Choung SY, Kobayashi T, Inoue K, Higashijima T, Miyazawa T (1990) Complex formation of peptide antibiotic Ro09-0198 with lysophosphatidylethanolamine: proton NMR analyses in dimethyl sulfoxide solution. *Biochemistry* 29: 113-118

Waksman SA, Schatz A, Reynolds DM (1946) PRODUCTION OF ANTIBIOTIC SUBSTANCES BY ACTINOMYCETES. *Annals of the New York Academy of Sciences* 48: 73-86

Walker M, Mitchell D, Donk W (2020) *Precursor peptide-targeted mining of more than one hundred thousand genomes expands the lanthipeptide natural product family*

Wang B, Guo F, Dong S-H, Zhao H (2019) Activation of silent biosynthetic gene clusters using transcription factor decoys. *Nature Chemical Biology* 15: 111-114

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460: 894-898

Wei J, He L, Niu G (2018) Regulation of antibiotic biosynthesis in actinomycetes: Perspectives and challenges. *Synth Syst Biotechnol* 3: 229-235

Weissman KJ (2015) Uncovering the structures of modular polyketide synthases. *Nat Prod Rep* 32: 436-453

Wenzel SC, Gross F, Zhang Y, Fu J, Stewart AF, Müller R (2005) Heterologous Expression of a Myxobacterial Natural Products Assembly Line in Pseudomonads via Red/ET Recombineering. *Chemistry & Biology* 12: 349-356

Westhoff S, van Leeuwe TM, Qachach O, Zhang Z, van Wezel GP, Rozen DE (2017) The evolution of no-cost resistance at sub-MIC concentrations of streptomycin in *Streptomyces coelicolor*. *ISME J* 11: 1168-1178

White CE, Winans SC (2005) Identification of amino acid residues of the *Agrobacterium tumefaciens* quorum-sensing regulator TraR that are critical for positive control of transcription. *Mol Microbiol* 55: 1473-1486

Widdick DA, Dodd HM, Barraille P, White J, Stein TH, Chater KF, Gasson MJ, Bibb MJ (2003) Cloning and engineering of the cinnamycin biosynthetic gene cluster from *Streptomyces cinnamoneus cinnamoneus* DSM 40005. *Proceedings of the National Academy of Sciences* 100: 4316-4321

Wietz M, Månsson M, Bowman JS, Blom N, Ng Y, Gram L (2012) Wide distribution of closely related, antibiotic-producing *Arthrobacter* strains throughout the Arctic Ocean. *Applied and environmental microbiology* 78: 2039-2042

Willey JM, van der Donk WA (2007) Lantibiotics: peptides of diverse structure and function. *Annu Rev Microbiol* 61: 477-501

Witkop B (1999) Paul Ehrlich and his Magic bullets--revisited. *Proc Am Philos Soc* 143: 540-557

Xu F, Nazari B, Moon K, Bushin LB, Seyedsayamdost MR (2017) Discovery of a Cryptic Antifungal Compound from *Streptomyces albus* J1074 Using High-Throughput Elicitor Screens. *J Am Chem Soc* 139: 9203-9212

Xu F, Wu Y, Zhang C, Davis KM, Moon K, Bushin LB, Seyedsayamdost MR (2019) A genetics-free method for high-throughput discovery of cryptic microbial metabolites. *Nature Chemical Biology* 15: 161-168

Xu M, Zhang F, Cheng Z, Bashiri G, Wang J, Hong J, Wang Y, Xu L, Chen X, Huang S-X et al (2020) Functional Genome Mining Reveals a Class V Lanthipeptide Containing a d-Amino Acid Introduced by an F420H2-Dependent Reductase. *Angewandte Chemie International Edition* 59: 18029-18035

Yaginuma S, Morishita A, Ishizawa K, Murofushi S, Hayashi M, Mutoh N (1992) Sporeamicin A, a new macrolide antibiotic. I. Taxonomy, fermentation, isolation and characterization. *The Journal of antibiotics* 45: 599-606

Yang X, Lennard KR, He C, Walker MC, Ball AT, Doigneaux C, Tavassoli A, van der Donk WA (2018) A lanthipeptide library used to identify a protein-protein interaction inhibitor. *Nature chemical biology* 14: 375-380

Yin S, Li Z, Wang X, Wang H, Jia X, Ai G, Bai Z, Shi M, Yuan F, Liu T et al (2016) Heterologous expression of oxytetracycline biosynthetic gene cluster in *Streptomyces venezuelae* WVR2006 to improve production level and to alter fermentation process. *Applied Microbiology and Biotechnology* 100: 10563-10572

Yu D, Xu F, Zeng J, Zhan J (2012) Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* 64: 285-295

Yu X, Zeng X, Shi W, Hu Y, Nie W, Chu N, Huang H (2018) Validation of Cycloserine Efficacy in Treatment of Multidrug-Resistant and Extensively Drug-Resistant Tuberculosis in Beijing, China. *Antimicrobial agents and chemotherapy* 62: e01824-01817

Zaman SB, Hussain MA, Nye R, Mehta V, Mamun KT, Hossain N (2017) A Review on Antibiotic Resistance: Alarm Bells are Ringing. *Cureus* 9: e1403-e1403

Zambaldo C, Luo X, Mehta AP, Schultz PG (2017) Recombinant Macrocyclic Lanthipeptides Incorporating Non-Canonical Amino Acids. *J Am Chem Soc* 139: 11646-11649

Zeng X, Zhang P, He W, Qin C, Chen S, Tao L, Wang Y, Tan Y, Gao D, Wang B et al (2018) NPASS: natural product activity and species source database for natural product research, discovery and tool development. *Nucleic Acids Res* 46: D1217-d1222

Zhang JJ, Tang X, Zhang M, Nguyen D, Moore BS (2017a) Broad-Host-Range Expression Reveals Native and Host Regulatory Elements That Influence Heterologous Antibiotic Production in Gram-Negative Bacteria. *mBio* 8: e01291-01217

Zhang MM, Wang Y, Ang EL, Zhao H (2016) Engineering microbial hosts for production of bacterial natural products. *Natural product reports* 33: 963-987

Zhang MM, Wong FT, Wang Y, Luo S, Lim YH, Heng E, Yeo WL, Cobb RE, Enghiad B, Ang EL et al (2017b) CRISPR-Cas9 strategy for activation of silent Streptomyces biosynthetic gene clusters. *Nature chemical biology*: 10.1038/nchembio.2341

Zhang Y, Chen M, Bruner SD, Ding Y (2018) Heterologous Production of Microbial Ribosomally Synthesized and Post-translationally Modified Peptides. *Frontiers in Microbiology* 9

Zhao M (2011) Lantibiotics as probes for phosphatidylethanolamine. *Amino Acids* 41: 1071-1079

Zhao M, Li Z, Bugenhagen S (2008) ^{99m}Tc-labeled duramycin as a novel phosphatidylethanolamine-binding molecular probe. *J Nucl Med* 49: 1345-1352

Zhou C, Ye B, Cheng S, Zhao L, Liu Y, Jiang J, Yan X (2019) Promoter engineering enables overproduction of foreign proteins from a single copy expression cassette in *Bacillus subtilis*. *Microbial Cell Factories* 18: 111

Ziemert N, Alanjary M, Weber T (2016) The evolution of genome mining in microbes – a review. *Natural Product Reports* 33: 988-1005

Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR (2012) The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* 7: e34064

Zimmermann N, Freund S, Fredenhagen A, Jung G (1993) Solution structures of the lantibiotics duramycin B and C. *Eur J Biochem* 216: 419-428

Zotchev SB (2012) Marine actinomycetes as an emerging resource for the drug development pipelines. *Journal of biotechnology* 158: 168-175

Chapter 9:

Appendix

9. Appendix

9.1 Chapter 2 Supplementary Material

9.1.1 Sequencing Primers

Table S1: Primers used for Sangar Sequencing during this work

Primer	Sequence 5'-3'
HNA171_Ps1Nys_LF	TCTCCAGACGTACATGTTCCGG
HNA172_Ps1Nys_LR	CTGATGATGATCCTCTTCCCTCGG
HNA173_Ps1Nys_CF	CATGTGACCGTCTTCCCTCC
HNA174_Ps1Nys_CR	TGCTCCTCTTCTCGGCCTAC
HNA175_Ps1Nys_RF	GTCGTGGGTGTCGTCGAAGT
HNA176_Ps1Nys_RR	GTCGGTACTTTCCATGGTCAGTT
Ps1 CtoT_F	GAGGAGAACAGCACGAAC
Ps1 CtoT_R	GCACCGTTCTGATCACC
Ps1 newCtoT_R	AGGTCACCGAGCTGGA
Ps1 GGGtoG_F	GCTGAACTCCACGAAGAC
Ps1 GGGtoG_R	GAATCGTTGTGGGATCTGGT
HNA183_Ps2Nys_NewLF	GTCAGCACGAGGTGGTCGAG
HNA184_Ps2Nys_NewLR	GTCGTCAGCTCGTCGTCCTCA
HNA185_Ps2Nys_NewCF	GACCTCGAAGACGCTGTCCG
HNA186_Ps2Nys_NewCR	CCCAGATACCGCTGCTGAAC
HNA187_Ps2Nys_NewRF	TGAAGGTACTGGCCCAGGTC
HNA188_Ps2Nys_NewRR	CAGTCCTGGCACAGCTGGT
HNA189_Ps2Nys_FarRF	ACACCTTCCGCCAGGACGT
HNA190_Ps2Nys_FarRR	GAGCAGCTCACGGGTCATCA
215G_Left_F	AAGGACATTCGCCTCGTCAGC
215G_Left_R	TTGCCCGTCTTGATCCGTTCC
215G_Centre_F	GCACGACGTAGAGGAACTC
215G_Centre_R	GTACGAGCACATACGCCATC
215G_Right_F	GGACGTAAGTACAGACAATTTCTC
215G_Right_R	TTTCGAGGACATGGAAGATCG
Z.Qin_left_F	GTTGCTGCCGACATGATGGTG
Z.Qin_left_R	CATCCGGCTGGTTCGAGGAC
Z.Qin_centre_F	GCAAGCAGCATGTGTTGCGCC
Z.Qin_centre_R	GATCCTGGATCAAAACCTCGTCC
Z.Qin_right_F	GTACCGGACCTCGATGTGGC
Z.Qin_right_R	CATCGAGGCCCGGCTGTAC
S.Alt_left_F	CTCACGCTTCGTAGTGTTAGTTTC
S.Alt_left_R	GACGTTGATCCAGAAGACCGATC
S.Alt_centre_F	GTCTACGTCGGCGTCATGTAC
S.Alt_centre_R	GTGGCCGATGTTTCTGACTTCA
S.Alt_right_F	CTTCTGTGTGATCGTCGGCATG
S.Alt_right_R	CTGTCTGCACCATGCTCATG

7G_Left_F	CACGACGGCCACTCGGACACGAAG
7G_Left_F	GGTGGTGAGTACGGCGATGGCCAC
7G_Centre_F	GACGCCAGTGACGGTGACC
7G_Centre_R	GACCTGACCGAGTCCGGTCTCCAC
7G_Right_F	GATGAAGGCCACGTGCCGCCAG
7G_Right_R	GCCCATCTCCACTACGTCCGCCTG
Lanti_left_F	GTCCTCGTCGACAGCGAGCCGATC
Lanti_left_R	CTCGACGACGACGCACCGCTC
Lanti_centre_F	GACGGTTCGGGCTGGAGGTCGTG
Lanti_centre_R	CTGCGCGCGGACAACCTCTGAG
Lanti_right_F	CAAGTCGACCCTGCTGCGGGTGAT
Lanti_right_R	GAGGAGAGGAACACCGCCTCCGAG
Natalia cluster1 left_F	GTCGTGCTCCAGCCGTCGAG
Natalia cluster1 left_R	GAGCACACGGTGATCTGCCCTGAC
Natalia cluster1 centre_F	GACGCCGTAGCCGAGGACCATGTC
Natalia cluster1 centre_R	CATCGCCCTGACCCGGGCAC
Natalia cluster1 right_F	GACACCGAGAGCGGCCTGGTCAC
Natalia cluster1 right_R	GTCGTGGGCGGCGGCGAGTTC
T.Booth Cluster3 left_F	CTCGCCGTCTCCTCCGCCTCCAG
T.Booth Cluster3 left_R	GTCGTGCGCCTTGCCGACTG
T.Booth Cluster3 centre_F	GCATCACGGCCATGGACCACAG
T.Booth Cluster3 centre_R	GCTGAACCACTTCTGGTCGGGCTC
T.Booth Cluster3 right_F	CTGCGAAGCGGTCCAGGACGACGTTG
T.Booth Cluster3 right_R	ACCGACCTCTACTACGGCGACCTG
AraC_Primer1_F	ACCCTGATCGAGAGGTA
AraC_Primer2_F	TACGAGAGGGAGATCCTGT
AraC_Primer3_R	GGCTCGGTCAGCGAATA
LuxR_Primer1_F	AAGACCGAACTGCTGGAC
LuxR_Primer2_F	ATCTGCATCGACGACGTG
LuxR_PrimerEuro2_F	ATCGCCATGATCCTCACCGAGTTG
LuxR_Primer3_F	TCGACGCCGAGACG
LuxR_PrimerEuro3_F	AGTACCTGCAACTGGCCTAC
LuxR_Primer4_F	TGGTCGTGCACTGGCT
LuxR_Primer5_F	TGACGGCCCTGCTCTA
LuxR_PrimerEuro5_F	TCTCCAGCAACGACCAGTA
LuxR_PrimerEuro1_R	GTGATGAATTCTCCTCGGC
SARP_Primer1_F	GACCCGAAGTTGATCGT
SARP_Primer2_F	TGCTGAGCTGCGTACCCT
SARP_PrimerEuro1_F	TGCTGGACAACGCGAACAC
SARP_Primer3_F	TGCTGAGCTGCGTACCCT
SARP_Primer4_F	ATCGCCAACGGCTTC
SARP_Primer5_F	GAGAGCCTGGTGGTCTA
SARP_Primer6_F	TGCACCTGACCACCT
SARP_Primer7_R	TTGAACCACTCGCCCA
deltaCinAseqF	GGCCGAGCGCTCGTGGAGTTG
deltaCinAseqR	GGGCGGCCAGATCGAGTGCTC

kyaseqF1	GACCTCCAGCGCCACCTGCAC
kyaseqF2	ATGCCGGTGCCGAGATCGTC
kyaseqF3	CGCACCGGTGCGTTCGCCGAC
kyaseqR1	GTCATGGTGATCTCCTTCGTC
kyaseqR2	CCCGTGCAGCAGGTCCAGGTG
kyaseqR3	GATGGACTCCAGGTGCTCCTG
kyaseqR4	GTAGGCGAGGTAGTCGTCCTC
deltaCinXseqF	CCCGCCGCGACCGTGAGCACC
deltaCinXseqR	CGATCGCGGGTGCCTCCTCGG
LanA_F_Primer	GAGGGACTACCGGAGCAGGAGAA
LanA_R_Primer	AGTGCTCGGTTCGGTCTGCAT
pAMA1_P1E_F	ATGAAACCGCTGTCGTTCCACG
pAMA1_P2_F	AACTGGAGCTGGAGCTGG
pAMA1_P3E_F	GAATCCGCACCTTCCGAATC
pAMA1_P4_F	GAACAGCTCGCTCGCGCGGTC
pAMA1_P5E_F	AGGTAGCCGAAGAAGGTGAG
pAMA1_P6_F	CGGGAGAAGGCGAACCGTC
pAMA1_P7_F	GGGAACCTGCTGCAGCCGCA
pAMA1_P8_F	CTGGTGCAGGAAGCGCTGA
pAMA1_P9E_F	GGATTCTGCTCAAGGACAG
pAMA1_P8_R	TCAGCGCTTCCTGCACCAG
pAMA1_P5E_R_KyaL	CTCACCTTCTTCGGCTACCT
pAMA1_P2_R_KyaR1	AACTGGAGCTGGAGCTGG
Roseo_P1_F	AAGTGTGTGGAGAGCAAT
Roseo_P2_F	TCCAGGATGCCGAGGA
Roseo_P3_F	AGTACGTGCCGCATGG
Roseo_P4_F	TGGTCGAGGACTTCCA
Roseo_P5_F	GGATACGGCTACGAGGCGCT
Roseo_P6_F	GGATATTCCGGTGCCTTT
Roseo_P7_F	CTACGAACTCGCGGACGGGT
Roseo_P8_F	CGTCGAACAGGACTGGTA
Roseo_P9_F	CCACCGGGGCGAGACAGAAAGAG
Roseo_P10_F	ATCAACCTGGACGAGCA
Roseo_P11_F	GACCAAGACCTACGGACACGT
Roseo_P12_F	ACTATCTCGAAGAGGCCGA
Roseo_P13_F	AGATGATGCACTTCTGGCG
Roseo_P14_F	ATGATGTTCCCCGTGAGCGCC
Roseo_P1_R	GCTGCAAGGCGATTAAGTTGGG
Roseo_P2_R	TCCTCGGCATCCTGGA

9.1.2 Cloning Primers

Table S2: Primers used for Cloning during this work

Primer	Sequence 5'-3'
pIJ10257_Nde1_f	CGATGCTGTTGTGGGCACAATC
pIJ10257_Pac1_R	CCTCCAACGTCATCTCGTTCTC
Arc_HindIII_R	CCAAGCTTTCACGGAAAGCTCGGCTC
ArC_NedI_F	CCCATATGGAGCTGGAACCCCTGATCG
AraC-pGP9Ligation_F	CTAGCATATGAGGAGGAATTCATCGTGGAGCTGGAACCCCTGATCG
AraC-pGP9Ligation_R	CTAGAAGCTTCTATCACGGAAAGCTCGGCTC
AraCpBF3_F	GGTACCGAATTCCTCGAGTAGGAGGAATTCATCGTGGAGCTGGAAACCCTGAT
AraCpBF3_R	GACGTCGCATGCTCCTCTAGATCTCACGGAAAGCTCGGCTCGC
AraCpBF3_xba1_F	CTAGTCTAGAAGGAGGAATTCATCGTGGAGCTGGAACCCCTGATCGAGA
AraCpBF3_xba1_R	CTAGTCTAGACTATCACGGAAAGCTCGGCTCGCTCGCC
AraCpBF3_xba1Gibson_F	CGAGAACCCTAGGTACCGAATTCCTCGAGTAGGAGGAATTCATCGTGGAGCTGGAACCCCTGATCGAGAGGTAC
AraCpBF3_xba1Gibson_R	AATTGGGCCCGACGTCGCATGCTCCTCTAGATCTCACGGAAAGCTCGGCTCGCTCGCCCGCAA
LuxRpRF10_F	GAACCTAGGATCCAAGCTTAAGGAGGAATTCATCGTGGAGTTTTACGACC
LuxRpRF10_R	CGCTCACTGGTACCATGCATAGATCTCAGCCCGGCGGCCGACGGTCG
TetA-R-Sac1_Primer_F	GGCCGCGAGCTCTCAATCGTCACCCTTTCTCG
TetA-R-Sac1_Primer_R	GCGGCCGAGCTCTCAGCGATCGGCTCGTTGCC
KyaBigAc_F	GGCCGCCATATGAAACCGCTGTCGTTCCACGTCCTC
KyaBigAc_R	GCGGCCTTAATTAATCAGCCGGACTGGTCGGCC
pAMA2stemloopattern_F	GAGGGCGCCGTGGCCATGAACGCCGTCGATATCCCCGTACGTCGCTCCGGTGAATGAACC GCATCCTGCGTGCATCCGACCGACCCGCCG
pAMA2stemloopattern_R	ACAGCACCACGACGAGTTCGTCCTTGGAGTACTGCGGCGGCACCGACCGGTGCGCCGCT
KY3smNdeI-F	GCGGAATTCGGAAGCCGCGGATCTCCTGCCGTGG
KY3smStuI-R	GCG GGTACC TTTAAA AGGCCT CGTTGGCCGCGATCCCCTTGG
KY3lgStuI-F	GCG AGGCCT AGCATCGACGCGGTGAGCCTCC
KY3lgNdeI-R2	CCGATAGCTGTTGCGCGGGTACCAC

9.1.3 Oligonucleotide Templates

Nucleotide templates used to create oligonucleotides for insertion of lantipeptide core peptide sequences into cassettes pWDW70 and pAMA2 via Gibson assembly, XXXXX denotes where the core peptide nucleotide sequence should be placed.

Table S3: Nucleotide Templated for Lantipeptide Core Peptide Oligonucleotides used during this work

Template	Sequence 5'-3'
pWDW70 core peptide insertion template	TCTGGACCAAGGGGATCGCGGCCAACGAGGTCGTCGCC XXXXX TAGCCTAGCATCGACGCGGTGAGCCTCCGGGCT
pAMA2 core peptide insertion template	TTCTGGACCGAGGGCGCCGTGGCCATGAACGCCGTCGAG XXXXX TAGATCCCCGTACGTCGCTCCGGTGAATGAACC

9.2 Chapter 3 Supplementary Material

9.2.1 UHPLC-HRMS of Produced Lantipeptides

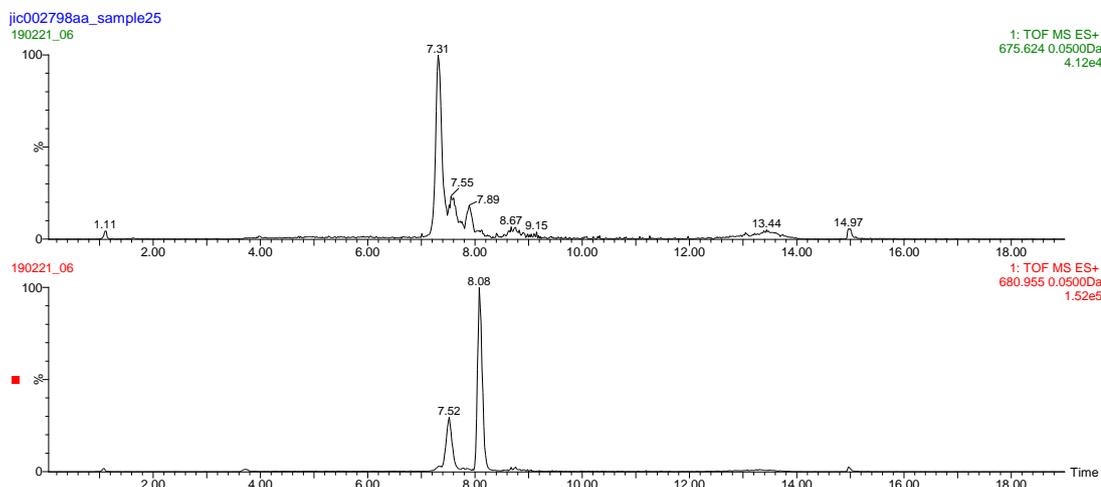


Figure S1: Extracted Ion Chromatogram (EIC) $[M+3H]^{3+}$ M1146/pEVK6/L_01 (Cinnamycin).

Hydroxylated: expected $m/z = 680.9553$, observed $m/z = 680.9536$, $\Delta = -2.5$ ppm.

Dehydroxylated: expected $m/z = 675.6236$, observed $m/z = 675.6216$, $\Delta = -3.0$ ppm.

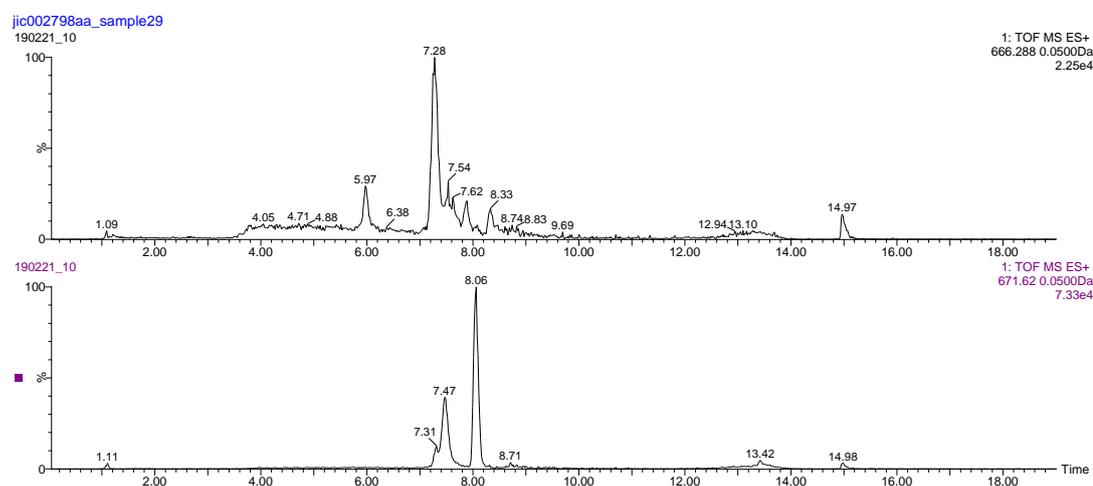


Figure S2: EIC $[M+3H]^{3+}$ of M1146/pEV6/L_02 (Duramycin).

Hydroxylated: expected $m/z = 671.6199$, observed $m/z = 671.6183$, $\Delta = -2.4$ ppm.

Dehydroxylated: expected $m/z = 666.2883$, observed $m/z = 666.2856$, $\Delta = -4.1$ ppm.

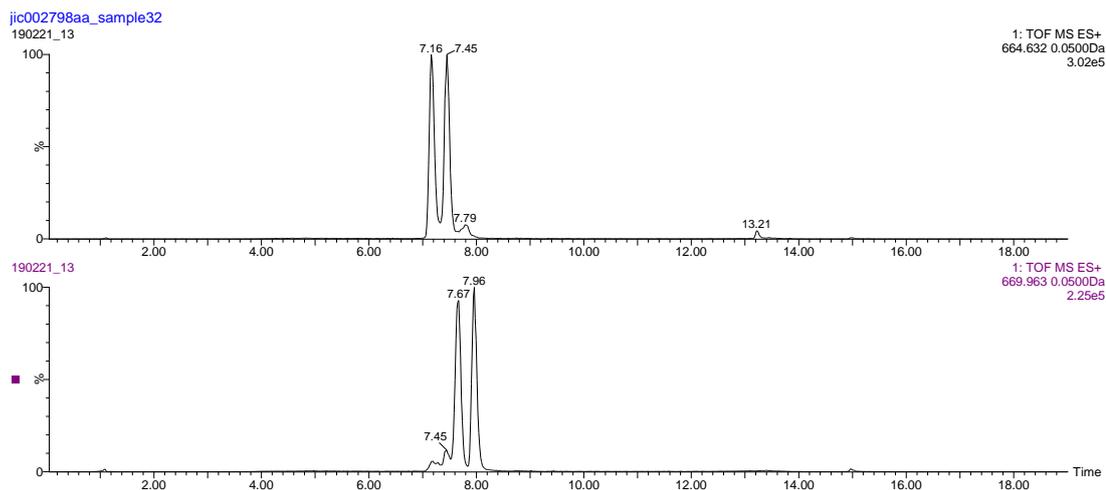


Figure S3: EIC $[M+3H]^{3+}$ of M1146/pEVK6/L_03 (Duramycin B). Hydroxylated: expected $m/z = 669.9631$, observed $m/z = 669.9598$, $\Delta = -4.9$ ppm. Dehydroxylated: expected $m/z = 664.6315$, observed $m/z = 664.6229$, $\Delta = -12.9$ ppm.

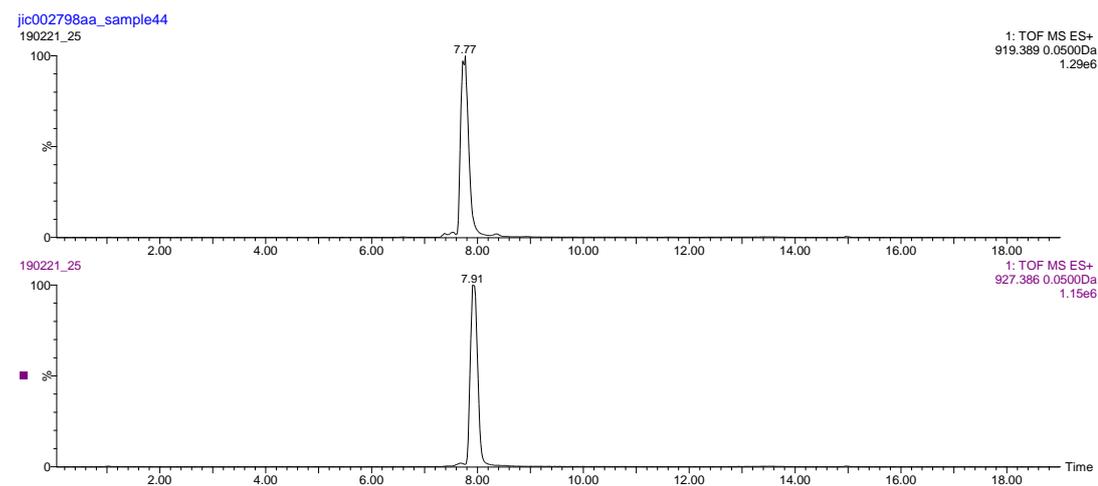


Figure S4: EIC $[M+2H]^{2+}$ of M1146/pEVK6/L_07 (Mathermycin). Hydroxylated: expected $m/z = 927.3864$, observed $m/z = 927.3862$, $\Delta = -0.2$ ppm. Dehydroxylated: expected $m/z = 919.3890$, observed $m/z = 919.3895$, $\Delta = 0.5$ ppm.

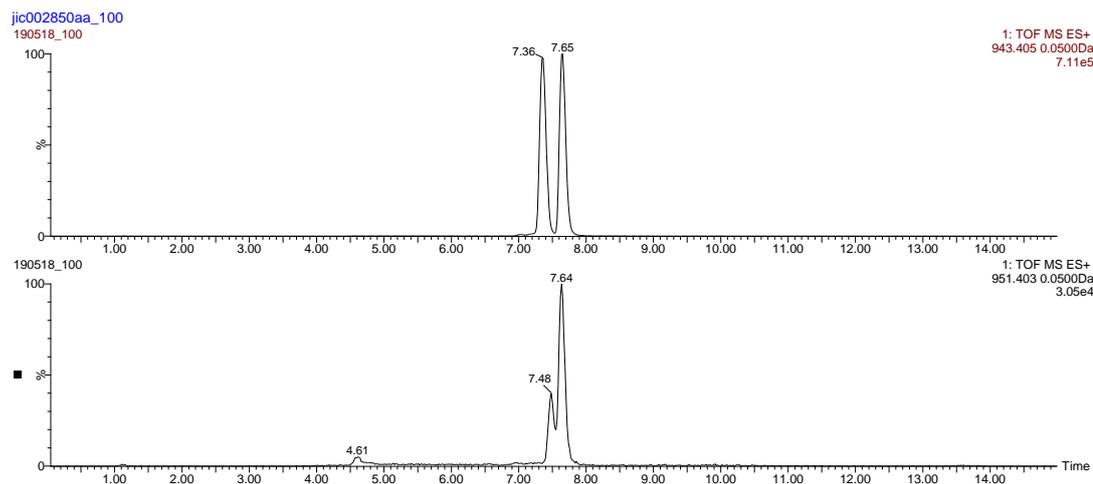


Figure S5: EIC $[M+2H]^{2+}$ of M1146/pEVK6/L_10. Hydroxylated: expected $m/z = 951.4026$, observed $m/z = 951.3993$, $\Delta = -3.5$. Dehydroxylated: expected $m/z = 943.4052$, observed $m/z = 943.4037$, $\Delta = -1.6$ ppm.

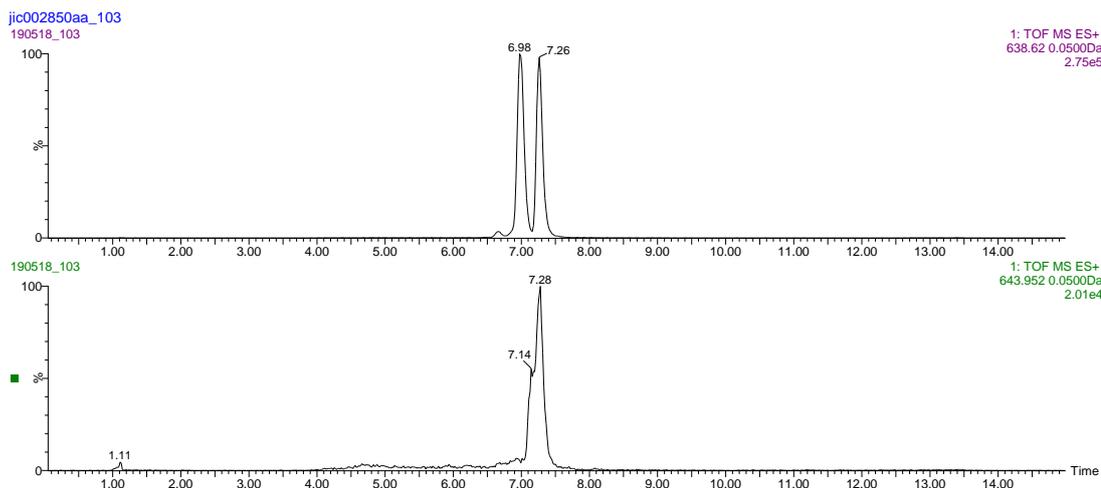


Figure S6: EIC $[M+3H]^{3+}$ of M1146/pEVK6/L_11. Hydroxylated: expected $m/z = 643.9517$, observed $m/z = 643.9510$, $\Delta = -1.1$ ppm. Dehydroxylated: expected $m/z = 638.6200$, observed $m/z = 638.6182$, $\Delta = -2.8$ ppm.

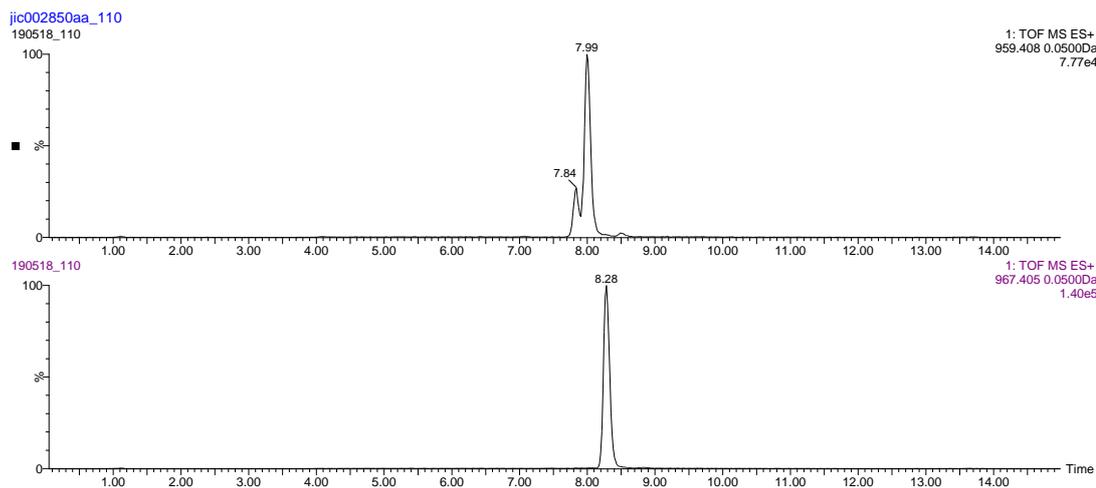


Figure S7: EIC $[M+2H]^{2+}$ of M1146/pEVK6/L_13. Hydroxylated: expected $m/z = 967.4052$, observed $m/z = 967.4020$, $\Delta = -3.3$ ppm. Dehydroxylated: expected $m/z = 959.4077$, observed $m/z = 959.4046$, $\Delta = -4.3$ ppm.

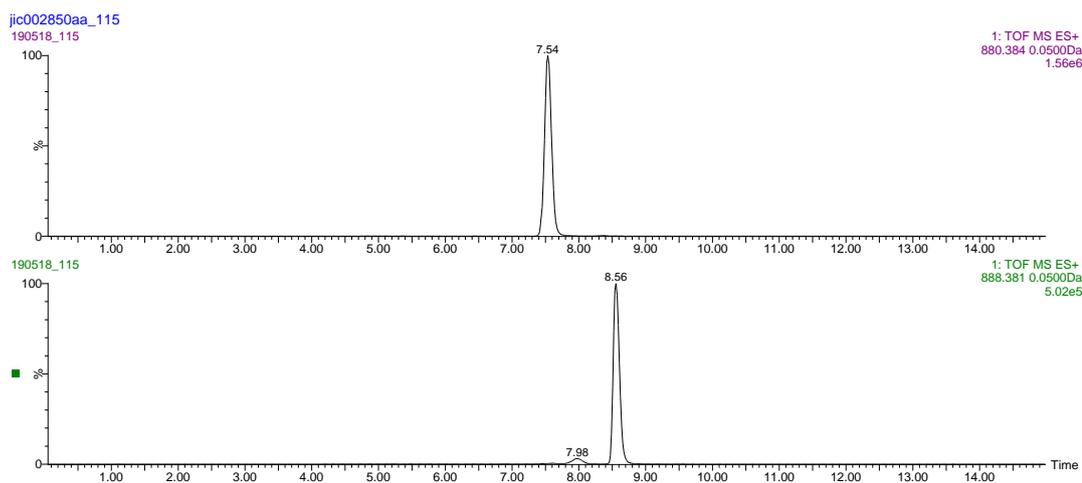


Figure S8: EIC $[M+2H]^{2+}$ of M1146/pEVK6/L_15. Hydroxylated: expected $m/z = 888.3811$, observed $m/z = 888.3790$, $\Delta = -2.4$ ppm. Dehydroxylated: expected $m/z = 880.3837$, observed $m/z = 880.3824$, $\Delta = -1.5$ ppm.

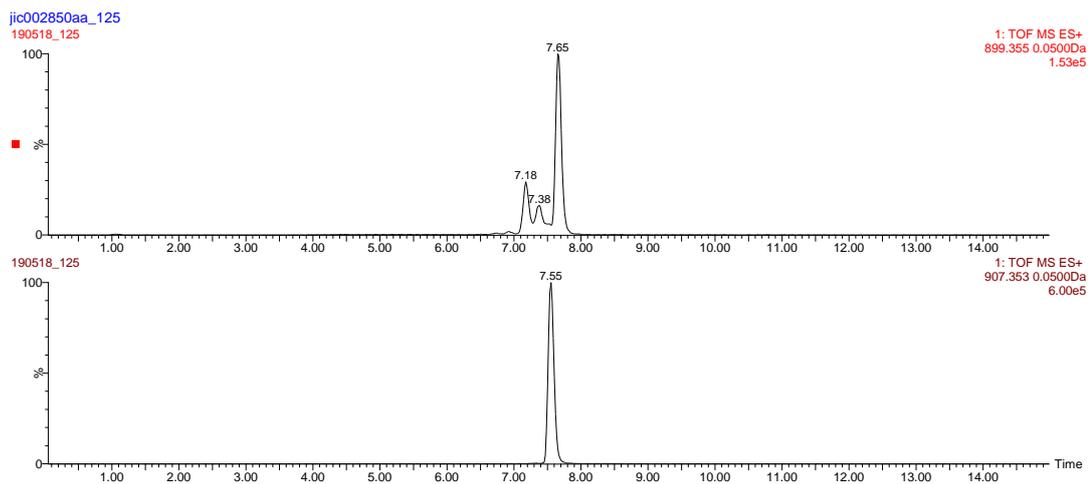


Figure S9: EIC [M+2H]²⁺ of M1146/pEVK6/L_18. Hydroxylated: expected m/z = 907.3526, observed m/z = 907.3494, Δ = -3.5 ppm. Dehydroxylated: expected m/z = 899.3551, observed m/z = 899.3508, Δ = -4.8 ppm.

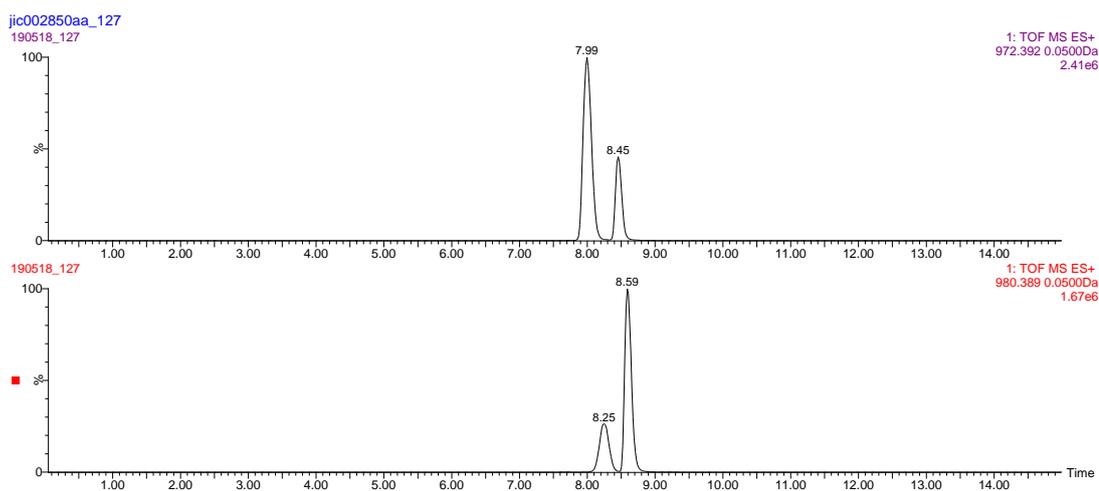


Figure S10: EIC [M+2H]²⁺ of M1146/pEVK6/L_19. Hydroxylated: expected m/z = 980.3892, observed m/z = 980.3882, Δ = -1.0 ppm. Dehydroxylated: expected m/z = 972.3917, observed m/z = 972.3892, Δ = -2.6 ppm.

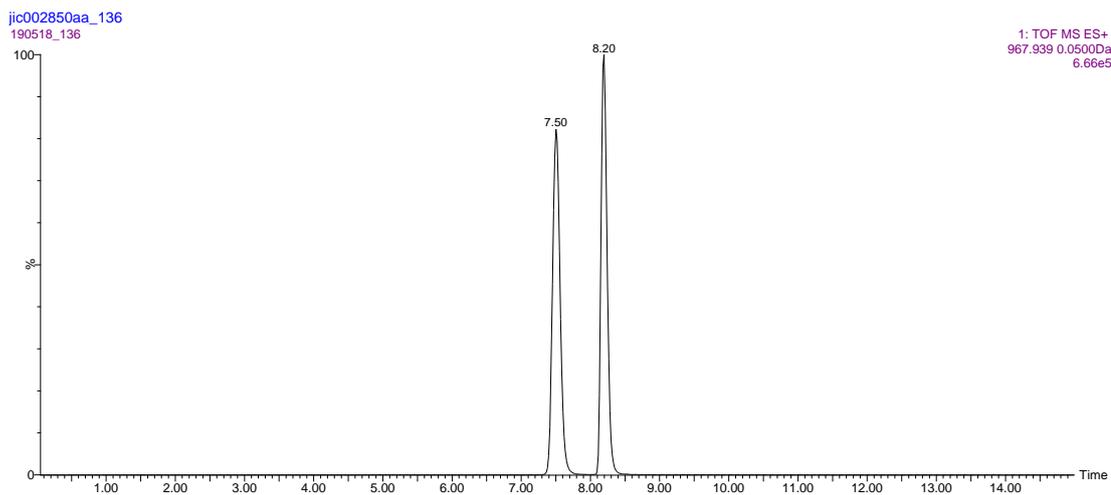


Figure S11: EIC $[M+2H]^{2+}$ of M1146/pEVK6/L_22. Dehydroxylated: expected $m/z = 967.9392$, observed $m/z = 967.9366$, $\Delta = -2.7$ ppm.

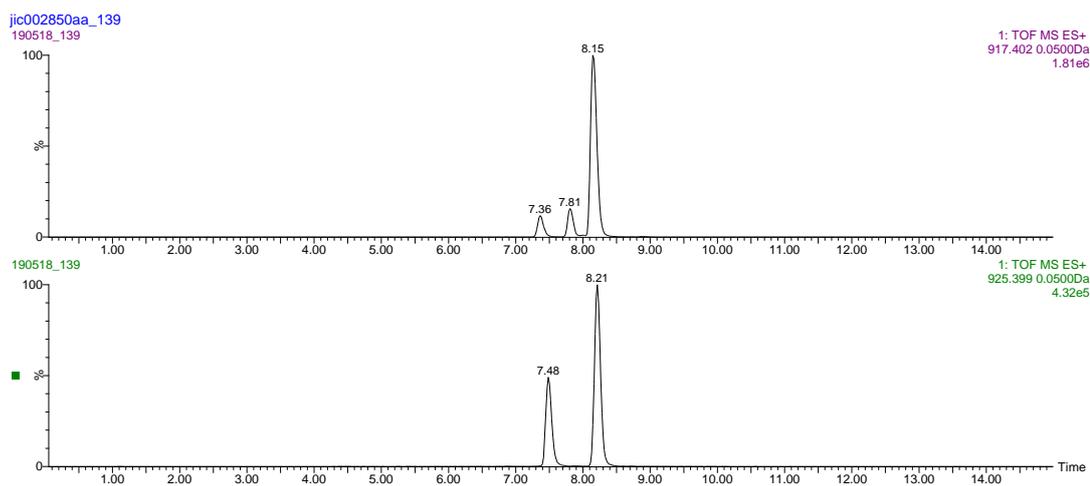


Figure S12: EIC $[M+2H]^{2+}$ of M1146/pEVK6/L_23. Hydroxylated: expected $m/z = 925.3995$, observed $m/z = 925.3964$, $\Delta = -3.3$ ppm. Dehydroxylated: expected $m/z = 917.4021$, observed $m/z = 917.3983$, $\Delta = -4.1$ ppm.

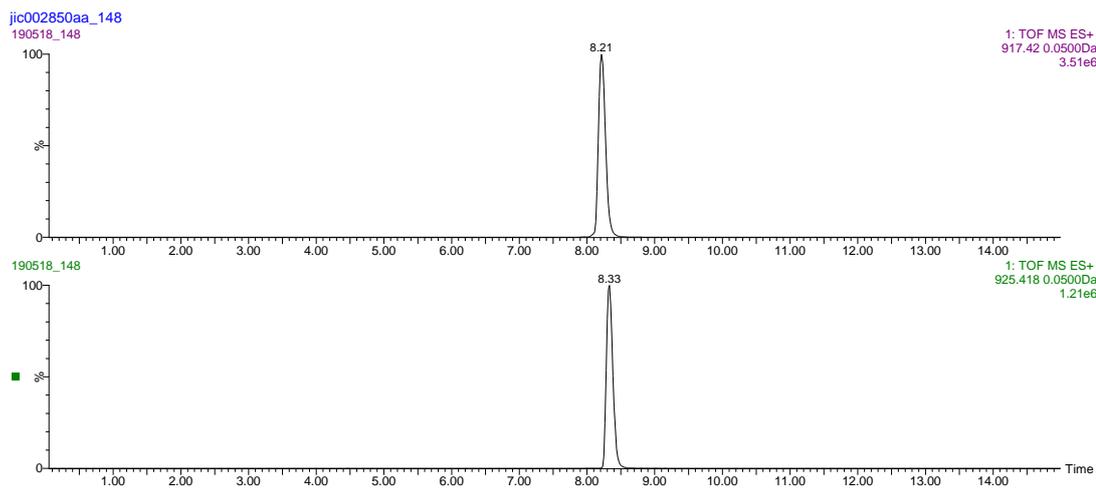


Figure S13: EIC $[M+2H]^{2+}$ of M1146/pEVK6/L_26. Hydroxylated: expected $m/z = 925.4177$, observed $m/z = 925.4160$, $\Delta = -1.8$ ppm. Dehydroxylated: expected $m/z = 917.4203$, observed $m/z = 917.4186$, $\Delta = -1.9$ ppm.

9.2.2 Phylogenetic Trees of Identified Lantipeptides

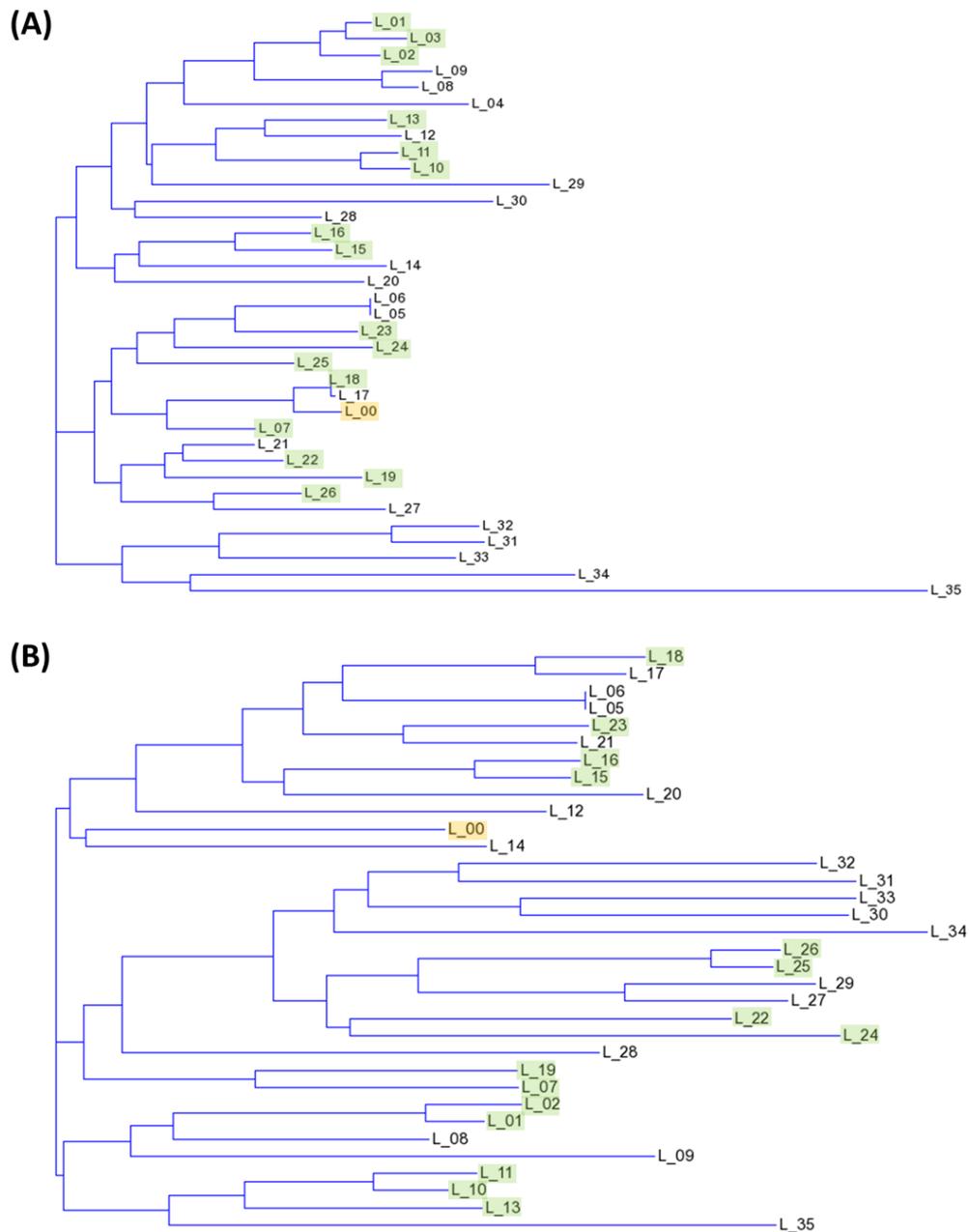


Figure S14: Phylogenetic Trees of Core Peptide Regions and LanA peptides from Identified Lantipeptides L_00 to L_35. (A) Phylogenetic tree of core peptide sequences from identified lantipeptides. (B) Phylogenetic tree of LanA amino acid sequences from identified lantipeptides. (Green) lantipeptides expressed using the kyamicin heterologous expression platform, (Yellow) Kyamicin, (Clear) Lantipeptides not expressed using the kyamicin heterologous expression platform.

9.2.3 Rosiermycin Synthetic Cassette Bioassay



Figure S15: Rosiermycin Heterologous Synthetic Cassette System Strains Overlaid with *B. subtilis*. Heterologous host *Streptomyces coelicolor* M1146, carried out with biological triplicates, no zones of inhibition seen. Lantipeptide core peptide sequences L_00 to L_35 tested, labelled Lr_XX respectively (Table 3.1) (i) M1146, negative control (ii) M1146/pAMA3/pAMA2, empty vector negative control (iii) M1146/pEVK6/pWDW63, kyamycin producing positive control.

9.2.4 Tandem MS

9.2.4.1 L_19

Table S4: Calculated and Observed Mass Fragmentation of Reduced L_19 Hydroxy F3 (zero sulphur)

	Chemical Formula [M+H] ⁺	Amino Acid Loss	Calculated Mass	Mass Loss	Observed Mass	Difference
Precursor	C ₈₆ H ₁₂₅ N ₂₀ O ₂₇		1869.902		1869.997	0.095
Y₁	C ₈₃ H ₁₂₀ N ₁₉ O ₂₆	Ala	1798.865	71.037	1798.509	-0.356
Y₂	C ₇₈ H ₁₁₃ N ₁₈ O ₂₃	Glu	1669.823	129.042	1669.318	-0.505
Y₃	C ₇₅ H ₁₀₈ N ₁₇ O ₂₁	Ser	1582.792	87.031	1582.332	-0.460
Y₄	C ₇₁ H ₁₀₁ N ₁₆ O ₂₀	Abu	1497.740	85.052	1497.184	-0.556
Y₅	C ₆₈ H ₉₆ N ₁₅ O ₁₉	Ala	1426.703	71.037	1426.194	-0.509
Y₆	C ₆₆ H ₉₃ N ₁₄ O ₁₈	Gly	1369.682	57.021	1369.189	-0.493
Y₇	C ₅₇ H ₈₄ N ₁₃ O ₁₇	Phe	1222.614	147.068	1222.206	-0.408
Y₈	C ₅₅ H ₈₁ N ₁₂ O ₁₆	Gly	1165.593	57.021	1165.249	-0.344
Y₉	C ₅₀ H ₇₄ N ₁₁ O ₁₅	Pro	1068.541	97.052	1068.261	-0.280
Y₁₀	C ₄₁ H ₆₅ N ₁₀ O ₁₄	Phe	921.473	147.068	921.213	-0.260
Y₁₁	C ₃₇ H ₅₈ N ₉ O ₁₃	Abu	836.421	85.052	836.162	-0.259
Y₁₂	C ₂₈ H ₄₉ N ₈ O ₁₂	Phe	689.353	147.068	689.094	-0.259
Y₁₃	C ₂₃ H ₄₀ N ₇ O ₁₁	Val	590.285	99.068	590.008	-0.277
Y₁₄	C ₂₀ H ₃₅ N ₆ O ₁₀	Ala	519.248	71.037	518.974	-0.274
Y₁₅	C ₁₆ H ₃₀ N ₅ O ₆	Asp-OH	388.227	131.021	387.995	-0.232
Y₁₆	C ₁₄ H ₂₇ N ₄ O ₅	Gly	331.206	57.021	331.005	-0.201
Y₁₇	C ₁₀ H ₂₀ N ₃ O ₃	Thr	230.159	101.047	230.029	-0.130
Y₁₈	C ₇ H ₁₅ N ₂ O ₂	Ala	159.122	71.037	159.091	-0.031
		Lys+CH2				

Table S5: Calculated and Observed Mass Fragmentation of Reduced L₁₉ Deoxy F3 (zero sulphur)

	Chemical Formula [M+H] ⁺	Amino Acid Loss	Calculated Mass	Mass Loss	Observed Mass	Difference
Precursor	C ₈₆ H ₁₂₅ N ₂₀ O ₂₆		1853.907		1853.819	-0.088
Y ₁	C ₈₃ H ₁₂₀ N ₁₉ O ₂₅	Ala	1782.870	71.037	1782.425	-0.455
Y ₂	C ₇₈ H ₁₁₃ N ₁₈ O ₂₂	Glu	1653.828	129.042	1653.318	-0.510
Y ₃	C ₇₅ H ₁₀₈ N ₁₇ O ₂₀	Ser	1566.797	87.031	1566.498	-0.299
Y ₄	C ₇₁ H ₁₀₁ N ₁₆ O ₁₉	Abu	1481.745	85.052	1481.166	-0.579
Y ₅	C ₆₈ H ₉₆ N ₁₅ O ₁₈	Ala	1410.708	71.037	1410.119	-0.589
Y ₆	C ₆₆ H ₉₃ N ₁₄ O ₁₇	Gly	1353.687	57.021	1353.123	-0.564
Y ₇	C ₅₇ H ₈₄ N ₁₃ O ₁₆	Phe	1206.619	147.068	1206.143	-0.476
Y ₈	C ₅₅ H ₈₁ N ₁₂ O ₁₅	Gly	1149.598	57.021	1149.194	-0.404
Y ₉	C ₅₀ H ₇₄ N ₁₁ O ₁₄	Pro	1052.546	97.052	1052.128	-0.418
Y ₁₀	C ₄₁ H ₆₅ N ₁₀ O ₁₃	Phe	905.478	147.068	905.083	-0.395
Y ₁₁	C ₃₇ H ₅₈ N ₉ O ₁₂	Abu	820.426	85.052	820.070	-0.356
Y ₁₂	C ₂₈ H ₄₉ N ₈ O ₁₁	Phe	673.358	147.068	672.963	-0.395
Y ₁₃	C ₂₃ H ₄₀ N ₇ O ₁₀	Val	574.290	99.068	573.875	-0.415
Y ₁₄	C ₂₀ H ₃₅ N ₆ O ₉	Ala	503.252	71.037	502.815	-0.438
Y ₁₅	C ₁₆ H ₃₀ N ₅ O ₆	Asp	388.227	115.026	387.833	-0.394
Y ₁₆	C ₁₄ H ₂₇ N ₄ O ₅	Gly	331.206	57.021	330.786	-0.420
Y ₁₇	C ₁₀ H ₂₀ N ₃ O ₃	Thr	230.159	101.047	229.807	-0.352
Y ₁₈	C ₇ H ₁₅ N ₂ O ₂	Ala	159.122	71.037	159.000	-0.122
		Lys+CH ₂				

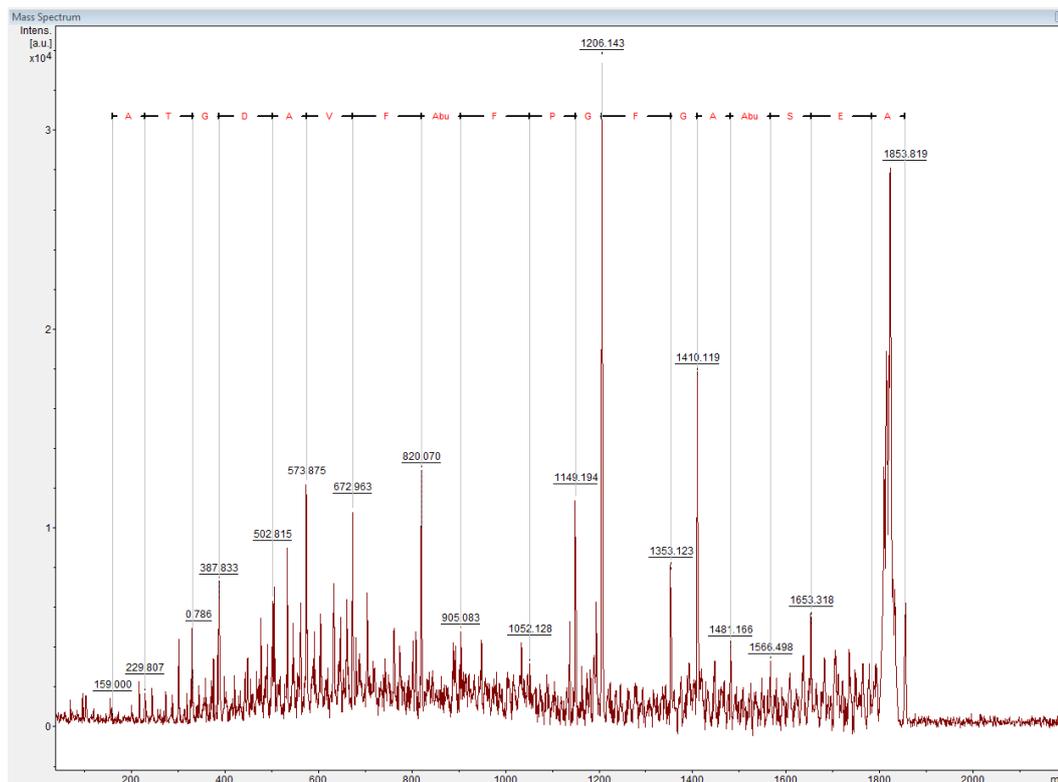


Figure S16: Fragmentation of Fully Reduced L₁₉ Deoxy F3 (zero sulphur). MALDI LIFT MS/MS fragmentation showed the y ion series of the reduced peptide (zero sulphur). The expected connectivity of the complete peptide was observed, with the expected mass loss per amino acid residue fragmented. The fragmentation of the lysinoalanine bridge resulted in a glycine at residue six and a N=C double bond at the end of the K19 side chain.

9.2.4.2 L_18

Table S6: Calculated and Observed Mass Fragmentation of Reduced L_18 Hydroxy F3 (zero sulphur)

	Chemical Formula [M+H] ⁺	Amino Acid Loss	Calculated Mass	Mass Loss	Observed Mass	Difference
Precursor	C ₇₆ H ₁₁₅ N ₂₀ O ₂₆		1723.82914		1723.925	0.09586
Y ₁	C ₇₃ H ₁₁₀ N ₁₉ O ₂₅	Ala	1652.79214	71.037	1652.713	-0.07914
Y ₂	C ₇₀ H ₁₀₅ N ₁₈ O ₂₄	Ala	1581.75514	71.037	1581.401	-0.35414
Y ₃	C ₆₇ H ₁₀₀ N ₁₇ O ₂₂	Ser	1494.72414	87.031	1494.35	-0.37414
Y ₄	C ₆₃ H ₉₃ N ₁₆ O ₂₁	Abu	1409.67214	85.052	1409.187	-0.48514
Y ₅	C ₆₀ H ₈₈ N ₁₅ O ₂₀	Ala	1338.63514	71.037	1338.211	-0.42414
Y ₆	C ₅₈ H ₈₅ N ₁₄ O ₁₉	Gly	1281.61414	57.021	1281.169	-0.44514
Y ₇	C ₅₅ H ₈₀ N ₁₃ O ₁₇	Ser	1194.58314	87.031	1193.876	-0.70714
Y ₈	C ₅₃ H ₇₇ N ₁₂ O ₁₆	Gly	1137.56214	57.021	1136.995	-0.56714
Y ₉	C ₄₈ H ₇₀ N ₁₁ O ₁₅	Pro	1040.51014	97.052	1040.212	-0.29814
Y ₁₀	C ₃₆ H ₆₁ N ₁₀ O ₁₄	Phe	893.44214	147.068	893.186	-0.25614
Y ₁₁	C ₃₅ H ₅₄ N ₉ O ₁₃	Abu	808.39014	85.052	808.152	-0.23814
Y ₁₂	C ₂₆ H ₄₅ N ₈ O ₁₂	Phe	661.32214	147.068	661.112	-0.21014
Y ₁₃	C ₂₃ H ₄₀ N ₇ O ₁₁	Ala	590.28514	71.037	590.05	-0.23514
Y ₁₄	C ₂₀ H ₃₅ N ₆ O ₁₀	Ala	519.24814	71.037	519.044	-0.20414
Y ₁₅	C ₁₆ H ₃₀ N ₅ O ₆	Asp-OH	388.22714	131.021	388.041	-0.18614
Y ₁₆	C ₁₄ H ₂₇ N ₄ O ₅	Gly	331.20614	57.021	331.092	-0.11414
Y ₁₇	C ₁₁ H ₂₂ N ₃ O ₃	Ser	244.17514	87.031	244.112	-0.06314
Y ₁₈	C ₇ H ₁₅ N ₂ O ₂	Abu	159.12314	85.052	159.21	0.08686
		Lys+CH2				

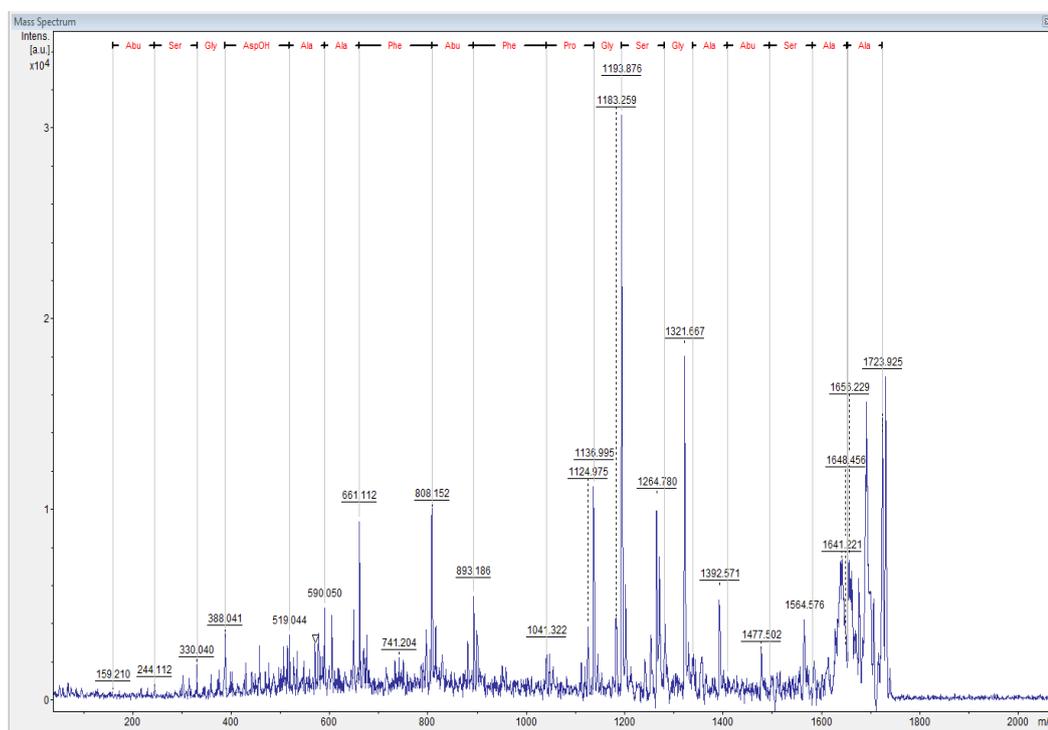


Figure S17: Fragmentation of Fully Reduced L_18 Hydroxy F3 (zero sulphur). MALDI LIFT MS/MS fragmentation showed the y ion series of the reduced peptide (zero sulphur). The expected connectivity of the complete peptide was observed, with the expected mass loss per amino acid residue fragmented. The fragmentation of the lysinoalanine bridge resulted in a glycine at residue six and a N=C double bond at the end of the K19 side chain.

9.2.4.3 L_26

Table S7: Calculated and Observed Mass Fragmentation of Reduced L_26 Deoxy F3 (zero sulphur)

	Chemical Formula [M+H] ⁺	Amino Acid Loss	Calculated Mass	Mass Loss	Observed Mass	Difference
Precursor	C ₇₉ H ₁₃₁ N ₂₀ O ₂₄		1743.96454		1744.018	0.05346
Y ₁	C ₇₆ H ₁₂₆ N ₁₉ O ₂₃	Ala	1672.92754	71.037	1672.189	-0.73854
Y ₂	C ₇₃ H ₁₂₁ N ₁₈ O ₂₂	Ala	1601.89054	71.037	1601.51	-0.38054
Y ₃	C ₇₀ H ₁₁₆ N ₁₇ O ₂₀	Ser	1514.85954	87.031	1514.749	-0.11054
Y ₄	C ₆₆ H ₁₀₉ N ₁₆ O ₁₉	Abu	1429.80754	85.052	1429.553	-0.25454
Y ₅	C ₆₃ H ₁₀₄ N ₁₅ O ₁₈	Ala	1358.77054	71.037	1358.305	-0.46554
Y ₆	C ₆₁ H ₁₀₁ N ₁₄ O ₁₇	Gly	1301.74954	57.021	1301.421	-0.32854
Y ₇	C ₅₂ H ₉₂ N ₁₃ O ₁₆	Phe	1154.68154	147.068	1154.449	-0.23254
Y ₈	C ₅₀ H ₈₉ N ₁₂ O ₁₅	Gly	1097.66054	57.021	1097.412	-0.24854
Y ₉	C ₄₄ H ₇₈ N ₁₁ O ₁₄	Ile	984.57594	113.0846	984.559	-0.01694
Y ₁₀	C ₃₉ H ₆₉ N ₁₀ O ₁₃	Val	885.50794	99.068	885.499	-0.00894
Y ₁₁	C ₃₅ H ₆₂ N ₉ O ₁₂	Abu	800.45594	85.052	800.448	-0.00794
Y ₁₂	C ₂₉ H ₅₁ N ₈ O ₁₁	Ile	687.37134	113.0846	687.378	0.00666
Y ₁₃	C ₂₄ H ₄₂ N ₇ O ₁₀	Val	588.30334	99.068	588.292	-0.01134
Y ₁₄	C ₂₁ H ₃₇ N ₆ O ₉	Ala	517.26634	71.037	517.251	-0.01534
Y ₁₅	C ₁₇ H ₃₂ N ₅ O ₆	Asp	402.23934	115.027	402.256	0.01666
Y ₁₆	C ₁₅ H ₂₉ N ₄ O ₅	Gly	345.21834	57.021	345.24	0.02166
Y ₁₇	C ₁₁ H ₂₂ N ₃ O ₃	Thr	244.17034	101.048	244.26	0.08966
Y ₁₈	C ₇ H ₁₅ N ₂ O ₂	Abu	159.11834	85.052	159.253	0.13466
		Lys-CH2				

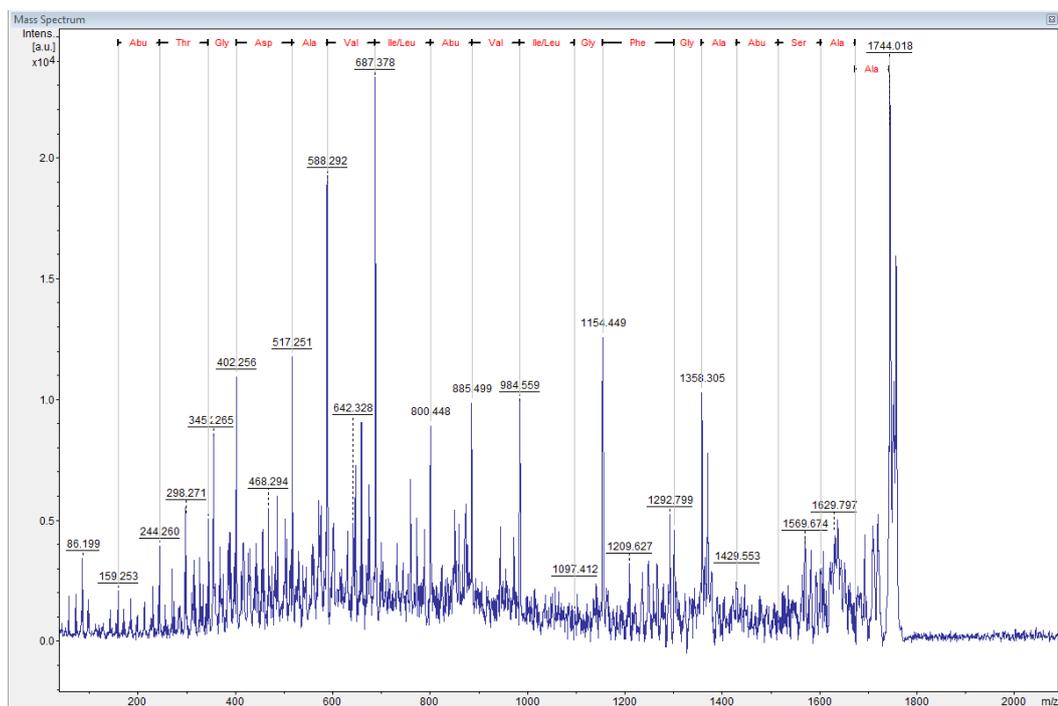


Figure S18: Fragmentation of Fully Reduced L_26 Deoxy F3 (zero sulphur). MALDI LIFT MS/MS fragmentation showed the y ion series of the reduced peptide (zero sulphur). The expected connectivity of the complete peptide was observed, with the expected mass loss per amino acid residue fragmented. The fragmentation of the lysinoalanine bridge resulted in a glycine at residue six and a N=C double bond at the end of the K19 side chain.

9.2.5 NMR

9.2.5.1 L_19 Hydroxy F3

Table S8: Chemical Shifts in NMR Spectra of L_19 Hydroxy F3 (600 MHz (H), 150 MHz (C), DMSO-d₆, 298K)

Residue	Amino Acid	Position	H	C
1	Cys	NH	ND	
1		α	ND	ND
1		β	ND	ND
2	Glu	NH	9.80	
2		α	4.72	51.4
2		β	2.09	24.2
2		γ		29.0
2		γ'	2.00	
2		γ''	1.59	
3	Ser	NH	8.38	
3		α	5.25	53.2
3		β		63.7
3		β'	3.85	
3		β''	3.66	
4	Thr	NH	7.85	
4		α	4.91	55.1
4		β	2.83	53.4
4		γ	1.30	22.7
5	Cys	NH	8.97	
5		α	4.44	52.7
5		β		38.7
5		β'	2.48	
5		β''	2.15	
6	Ser	NH	ND	
6		α	ND	ND
6		β	ND	ND
7	Phe	NH	8.03	
7		α	4.65	53.2
7		β		38.1
7		β'	3.38	
7		β''	2.91	
8	Gly	NH	7.48	
8		α		40.6
8		α'	4.00	
8		α''	4.06	
9	Pro			
9		α	3.92	61.0
9		β		ND

9		β'	1.91	
9		β''	1.85	
9		γ		ND
9		γ'	2.07	
9		γ''	1.94	
9		δ	ND	45.8
9		δ'	3.75	
9		δ''	3.23	
10	Phe	NH	8.98	
10		α	3.83	56.2
10		β		34.3
10		β'	2.98	
10		β''	2.92	
11	Thr	NH	ND	
11		α	4.43	60.6
11		β	3.23	47.6
11		γ	1.14	18.0
12	Phe	NH	ND	
12		α	3.66	55.3
12		β		34.3
12		β'	2.98	
12		β''	2.92	
13	Val	NH	7.31	
13		α	4.26	57.7
13		β	1.68	31.6
13		γ	0.82	18.3
13		γ'	0.95	18.4
14	Cys	NH	8.94	
14		α	3.21	56.2
14		β		34.3
14		β'	2.79	
14		β''	2.58	
15	Asp	NH	ND	
15		α	4.54	57.8
15		β	4.15	71.9
15		OH	3.55	
16	Gly	NH	7.48	
16		α		40.6
16		α'	4.00	
16		α''	4.06	
17	Thr	NH	8.32	
17		α	4.88	ND
17		β	3.75	66.9
17		γ	0.98	17.6
18	Ser	NH	8.66	

18		α	4.46	52.7
18		β		36.9
18		β'	3.05	
18		β''	2.84	
19	Lys	NH	ND	
19		α	ND	ND
19		β		28.2
19		β'	1.85	
19		β''	1.85	
19		γ	1.83	22.9
19		δ	ND	24.2
19		ϵ	ND	
19		NH'	ND	

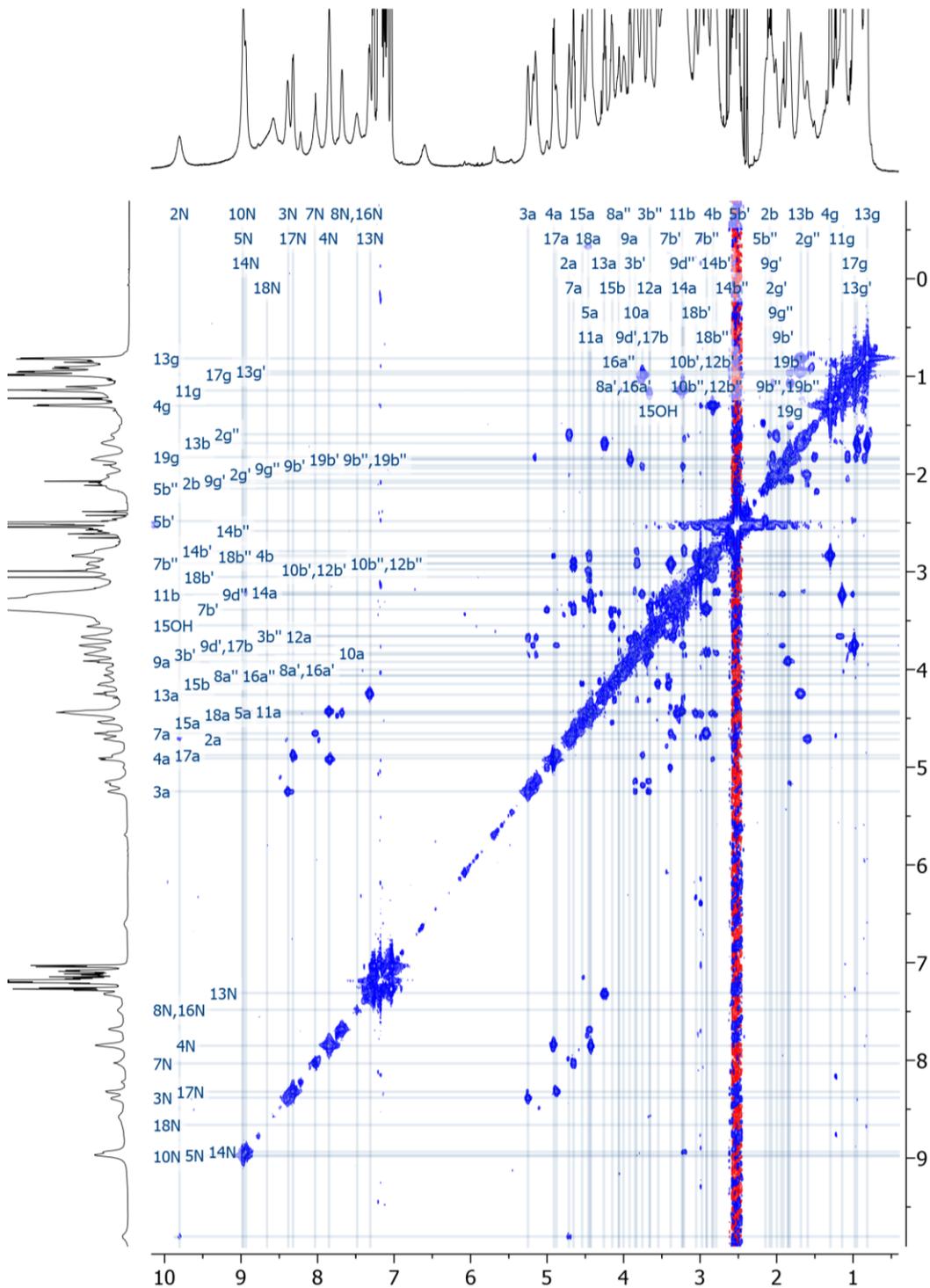


Figure S19: L_19 Hydroxy F3 2D COSY NMR (600 MHz, DMSO- d_6 , 298K). Numbers correspond to expected atom cross-peaks within L_19 Hydroxy (Fig 3.24B).

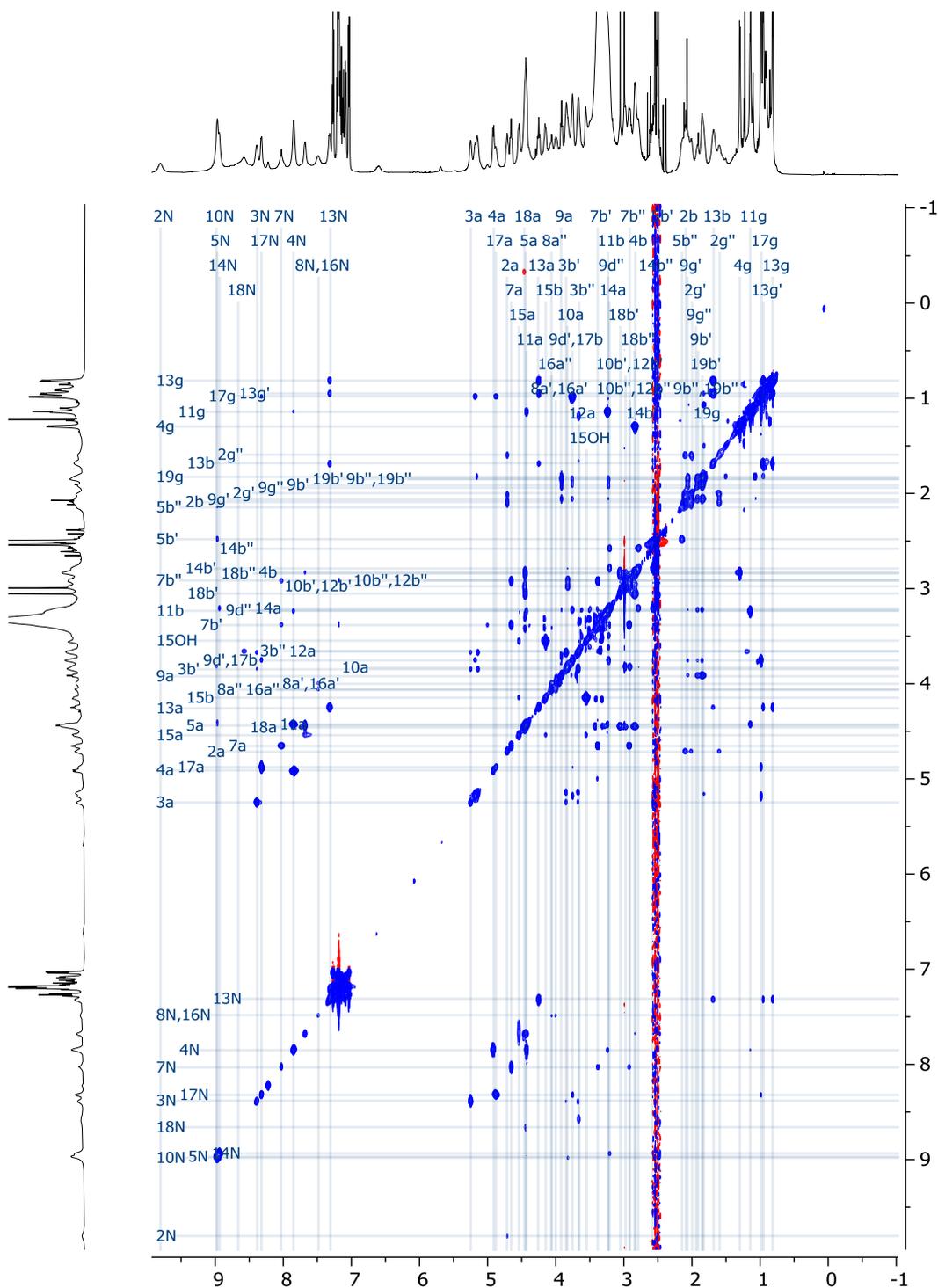


Figure S20: L_19 Hydroxy F3 2D TOCSY NMR (600 MHz, DMSO-d₆, 298K). Numbers correspond to expected atom cross-peaks within L_19 Hydroxy (Fig 3.24B).

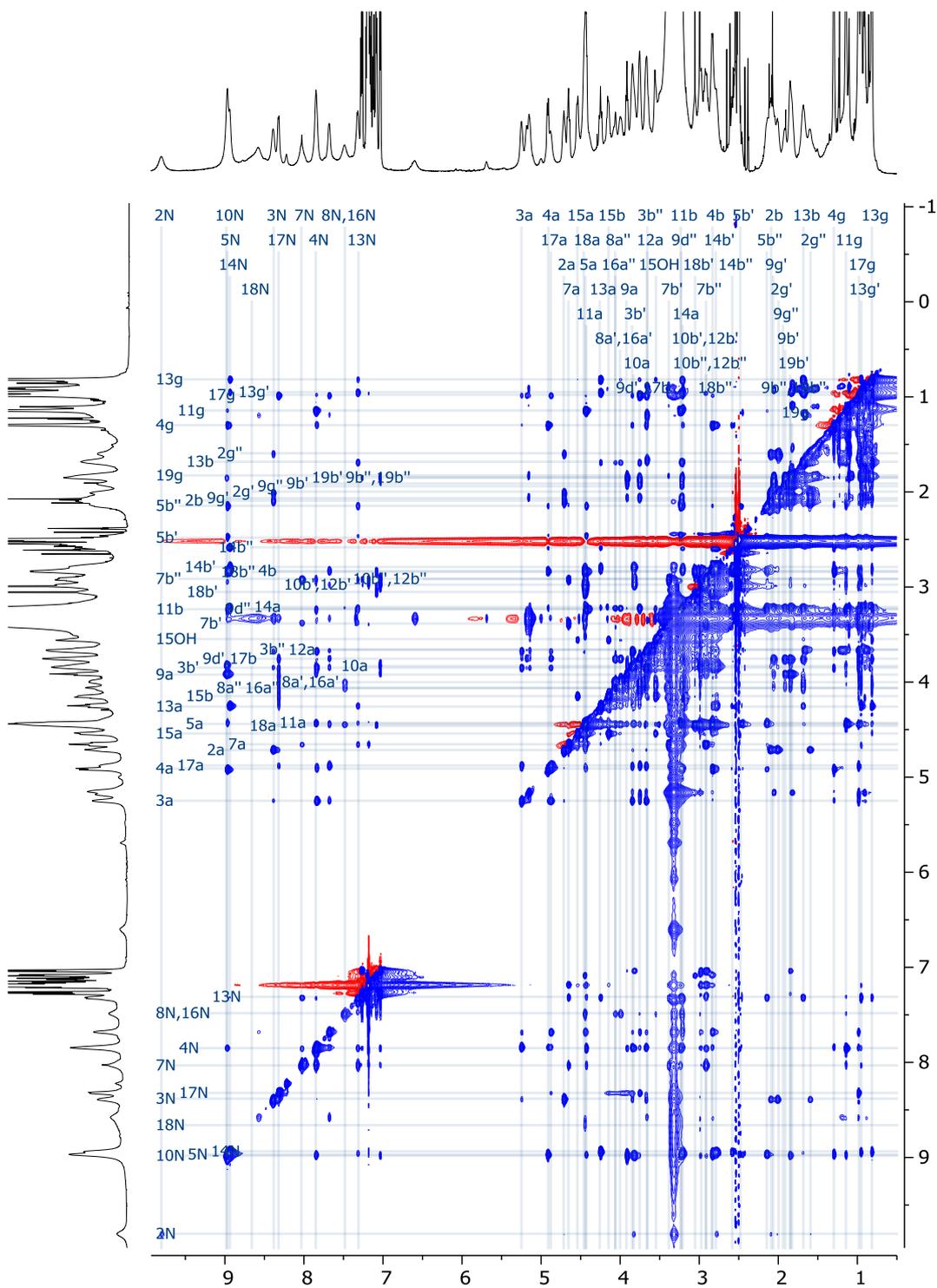


Figure S21: L_19 Hydroxy F3 2D NOESY NMR (600 MHz, DMSO-d₆, 298K). Numbers correspond to expected atom cross-peaks within L_19 Hydroxy (Fig 3.24B).

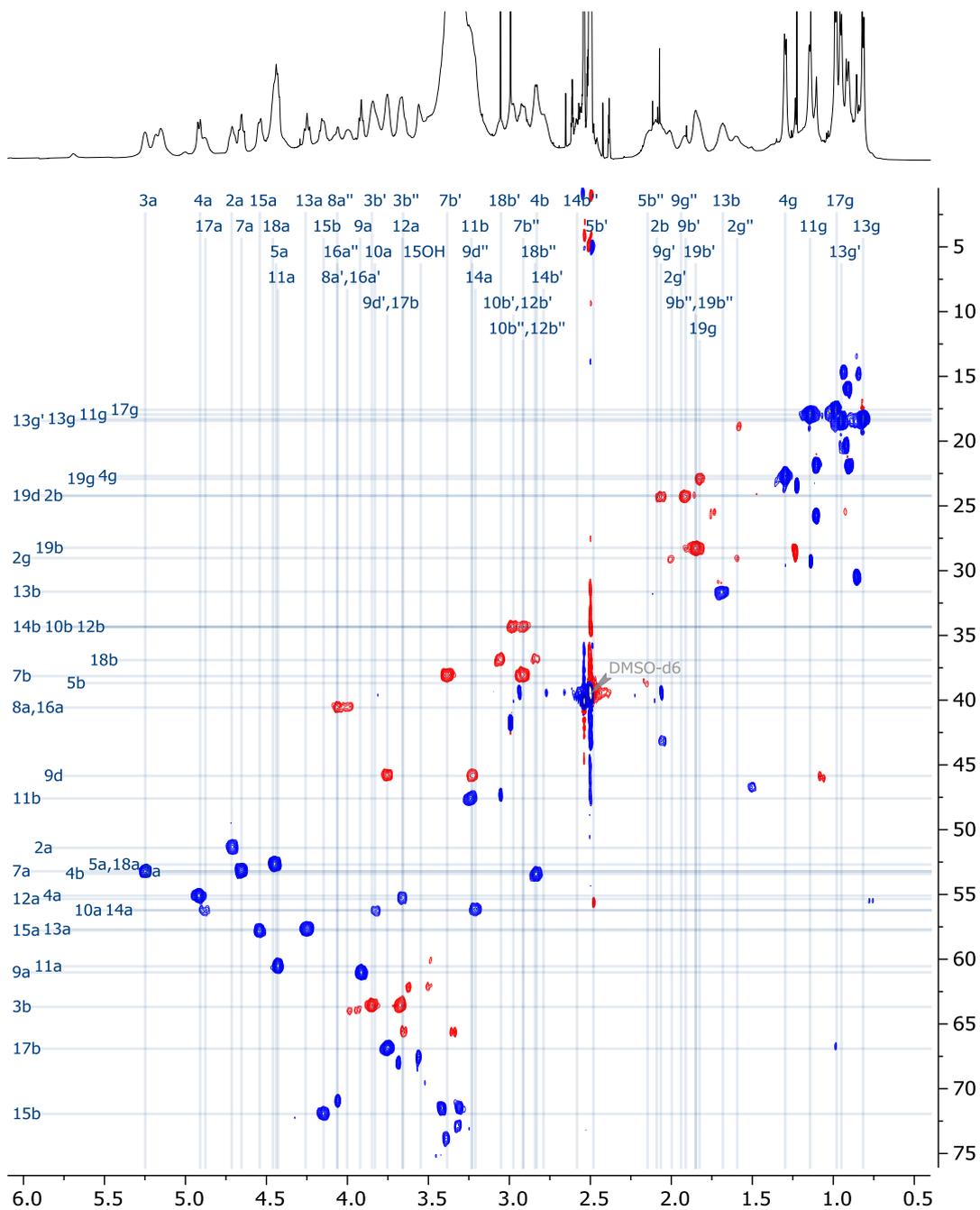


Figure S22: L_19 Hydroxy F3 2D HSQC NMR (600 MHz, DMSO-d₆, 298K). Numbers correspond to expected atom cross-peaks within L_19 Hydroxy (Fig 3.24B).

9.3 Chapter 5 Supplementary Material

9.3.1 Production Media Assays Mass Spectra of *S. erythraea*

ISOM/215G Δ forJ

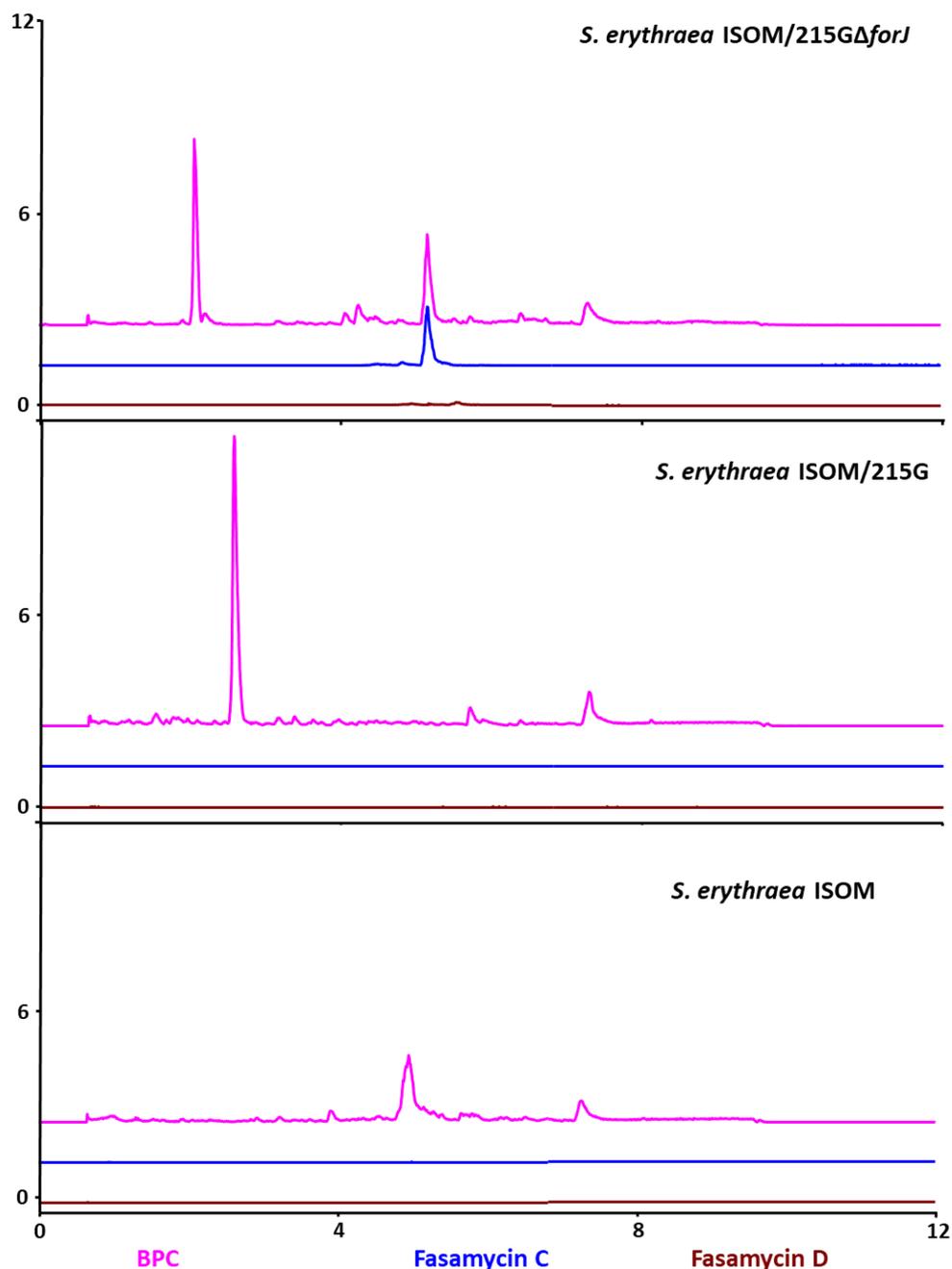


Figure S23: Extracted Ion Chromatograms (EIC) of *S. erythraea* ISOM/215G and *S. erythraea* ISOM/215G Δ forJ. Cultures grown in liquid GPP media, extracts generated day 8, analysed by UHPLC-MS. Formicamycins and fasamycin compounds searched for using known m/z (Qin *et al*, 2017). Fasamycin C; m/z known = 473.1595, m/z observed = 473.159, Δ = -1.1 ppm. Fasamycin D; m/z = 507.1205, observed m/z = 507.121, Δ = 1.0 ppm. All EIC scaled from 0 to 12, $\times 10,000,000$ intensity. Run = 12 min.

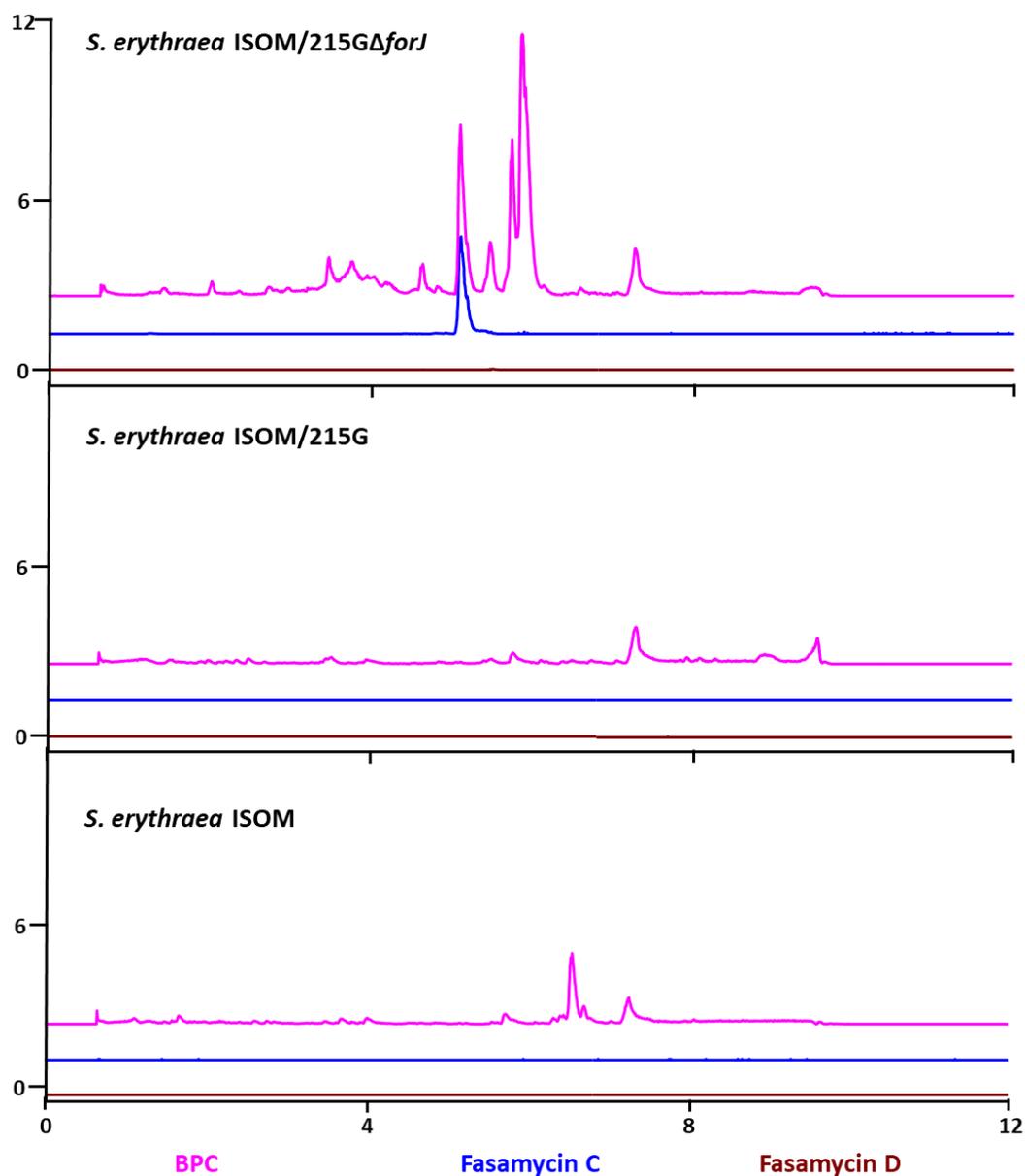


Figure S24: Extracted Ion Chromatograms (EIC) of *S. erythraea* ISOM/215G and *S. erythraea* ISOM/215G Δ forJ. Cultures grown in liquid SFM media, extracts taken at day 8, analysed by UHPLC-MS. Formicamycins and fasamycin compounds searched for using known m/z (Qin *et al.*, 2017). Fasamycin C; m/z known = 473.1595, m/z observed = 473.158, Δ = -3.2 ppm. All EIC scaled from 0 to 12, $\times 10,000,000$ intensity. Run = 12 min.

9.4 Chapter 6 Supplementary Material

9.4.1 Production of Fasamycin C and Fasamycin D in M1146/215GΔforJ

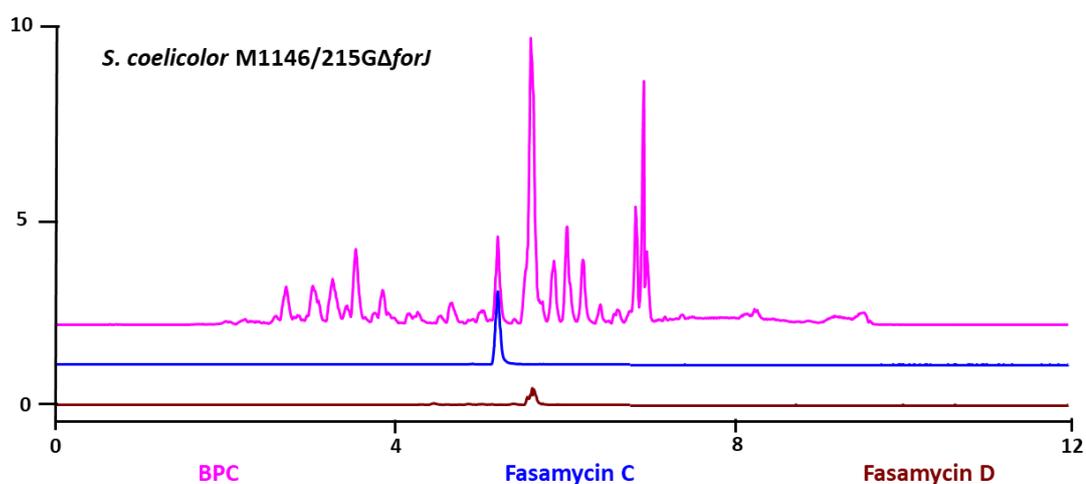


Figure S25: Extracted Ion Chromatograms (EIC) of *S. coelicolor* M1146/215GΔforJ. Cultures grown in liquid SV2 media, extracts taken at day 8, analysed by UHPLC-MS. Formicamycins and fasamycin compounds searched for using known m/z (Qin *et al.*, 2017). Fasamycin C; m/z known = 473.1595, m/z observed = 473.160, Δ = 1.1 ppm. Fasamycin D; m/z = 507.1205, observed m/z = 507.121, Δ = 1.0 ppm. All EIC scaled from 0 to 10, x10,000,000 intensity. Run = 12 min.

9.4.2 Comparison of Mass Spectra of *S. formicae* KY5 WT and *S. formicae* KY5 $\Delta forJ$ in Liquid Media

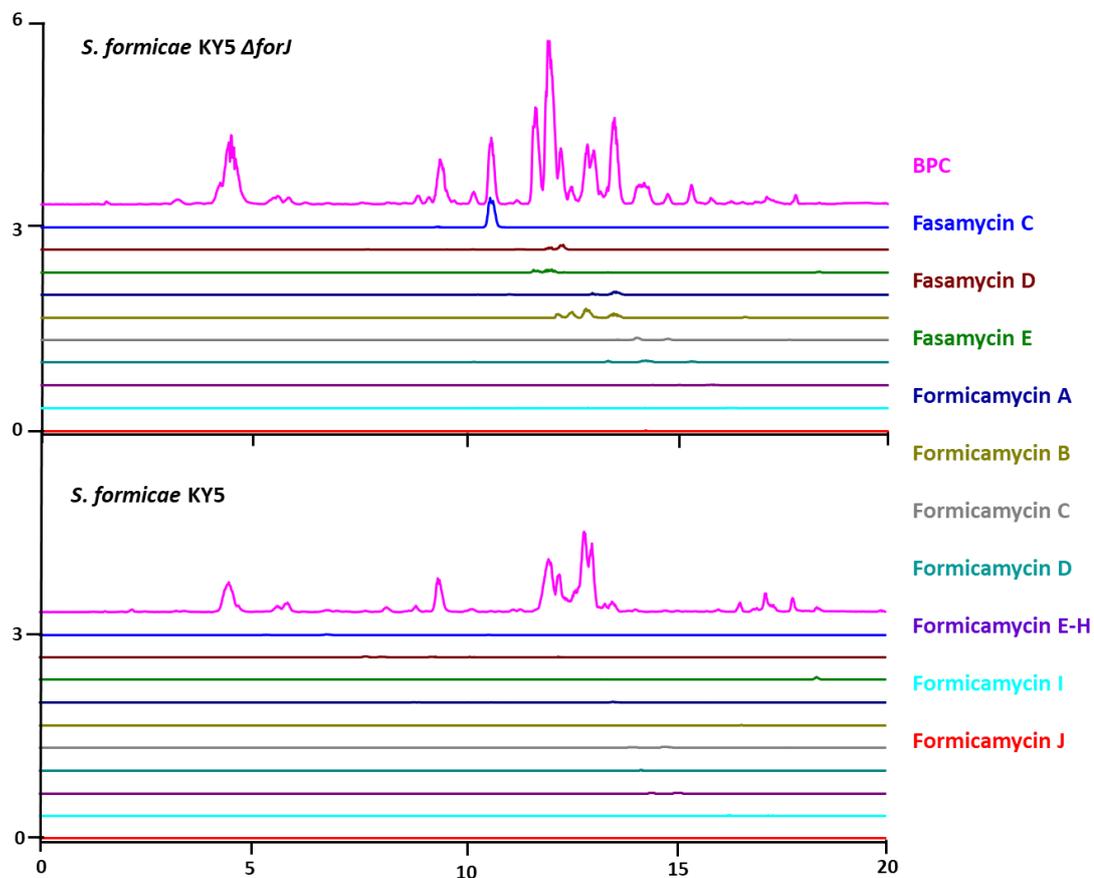


Figure S26: Extracted Ion Chromatograms (EIC) of *S. formicae* KY5 WT and *S. formicae* KY5 $\Delta forJ$. Cultures grown in liquid SV2 media, extracts taken at day 8, analysed by UHPLC-MS. Formicamycins and fasamycin compounds searched for using known m/z (Qin *et al.*, 2017). All EIC scaled from 0 to 6, x100,000,000 intensity. Run = 20 min.

9.4.3 Concentration Curves used to Determine Fasamycin and Formicamycin Titre from HPLC Chromatograms

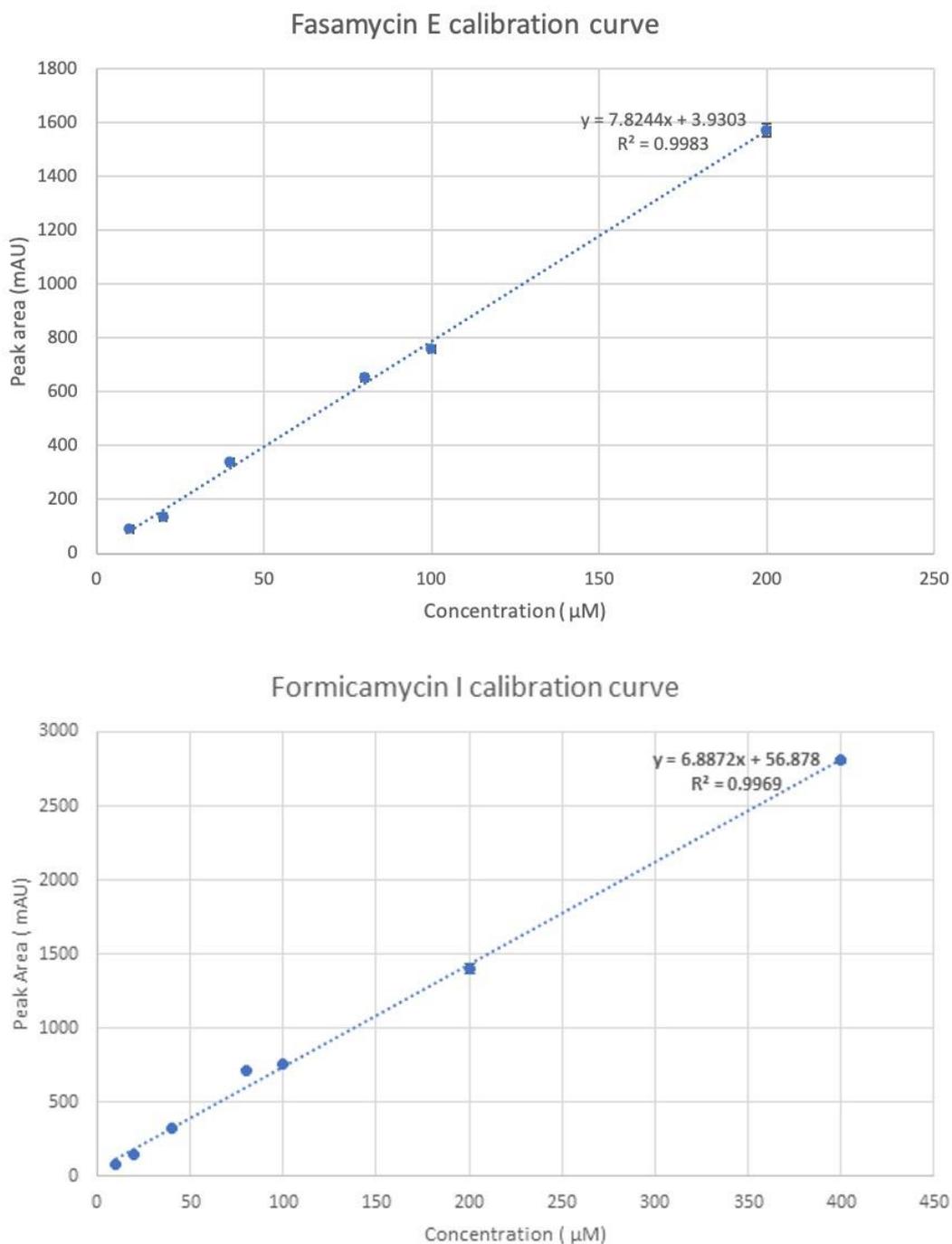


Figure S27: Concentration Curves of Fasamycin E and Formicamycin E used to determine Titre from HPLC Chromatograms. Area under peak in HPLC chromatogram plotted against known concentrations of Fasamycin E and Formicamycin I, Line of Best Fit R^2 and Y values used to determine titre of Fasamycins and Formicamycins produced during this work. Adapted from (Devine *et al.*, 2021), *Cell Chemical Biology*, 28(4), 515-523, licenced under CC BY-NC-ND.

9.5 Plasmid Maps

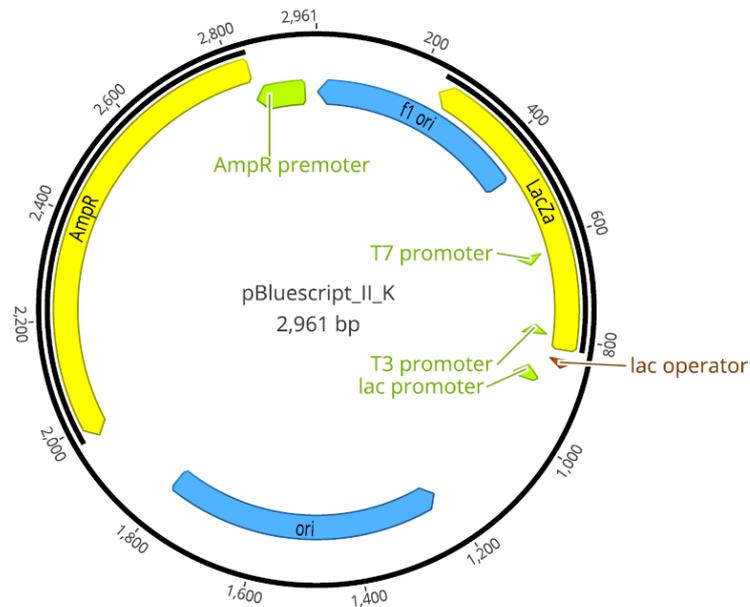


Figure S28: pBluescript_II_K Plasmid Map. Functional elements include antibiotic resistance markers (AmpR), origins of replication (ori, f1 ori) promoters (AmpR promoter, T7, T3, lac), lac operator and lac gene (LacZa).

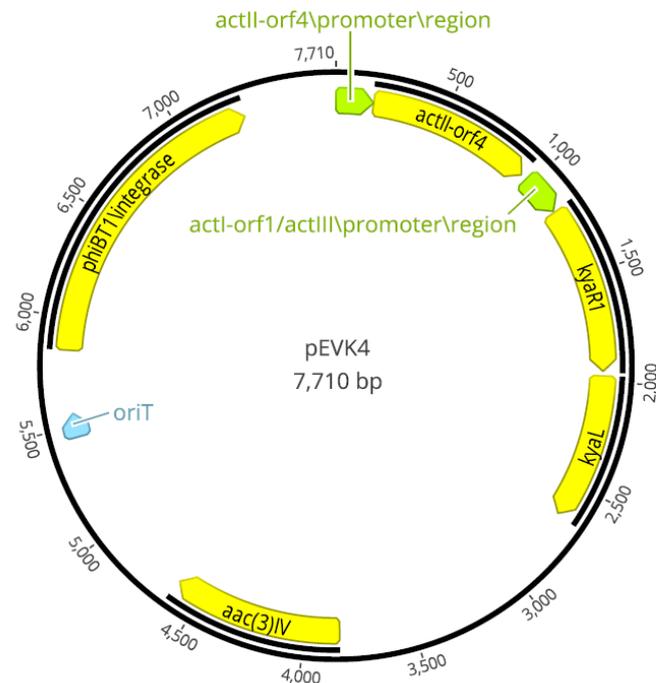


Figure S29: pEVK4 Plasmid Map. Functional elements include integrases (phiBT1), origins of replication (oriT), promoters (actII-orf4, actI-orf1), apramycin resistance genes (aac(3)IV), inserted cassette genes (kyaR1, kyaL).

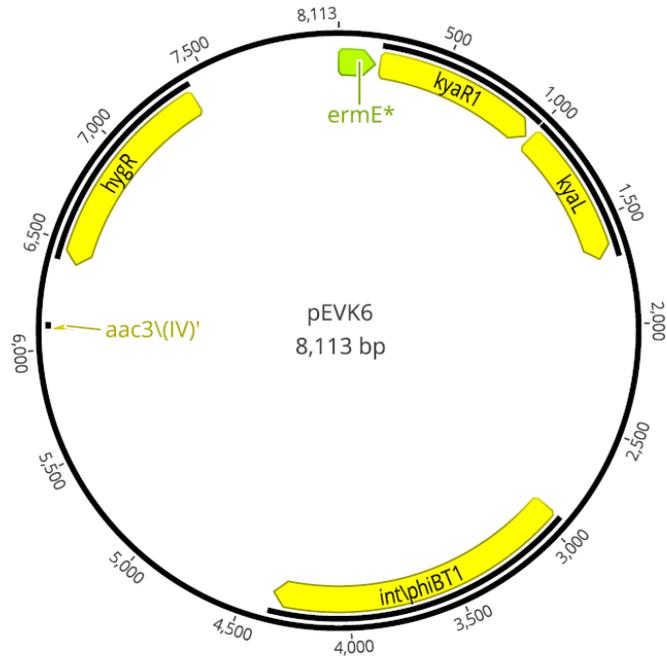


Figure S30: pEVK6 Plasmid Map. Functional genes include integrases (*phiBT1*), antibiotic resistance markers (*hygR*), promoters (*ermE**) origins of replication (*oriT*) and inserted cassette genes (*kyaR1*, *kyaL*).

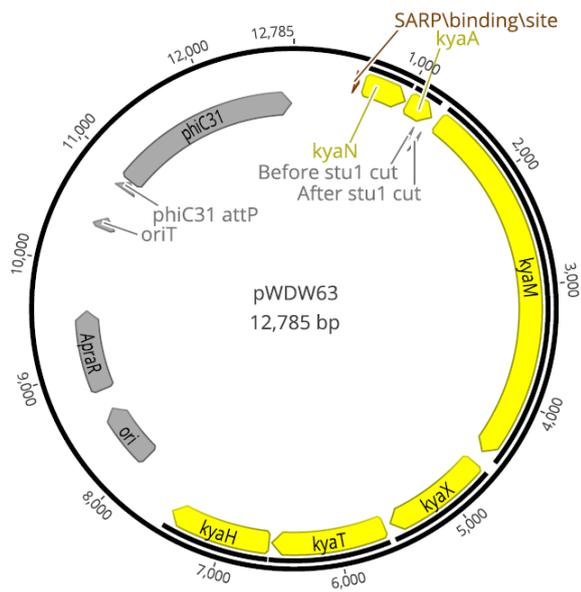


Figure S31: pWDW63 Plasmid Map. Functional genes include antibiotic resistance markers (*ApraR*), integrases (*phiC31*), origins of replication (*oriT*, *ori*), inserted cassette genes (*SARP* binding site, *kyaN*, *kyaM*, *kyaX*, *kyaT*, *kyaH*).

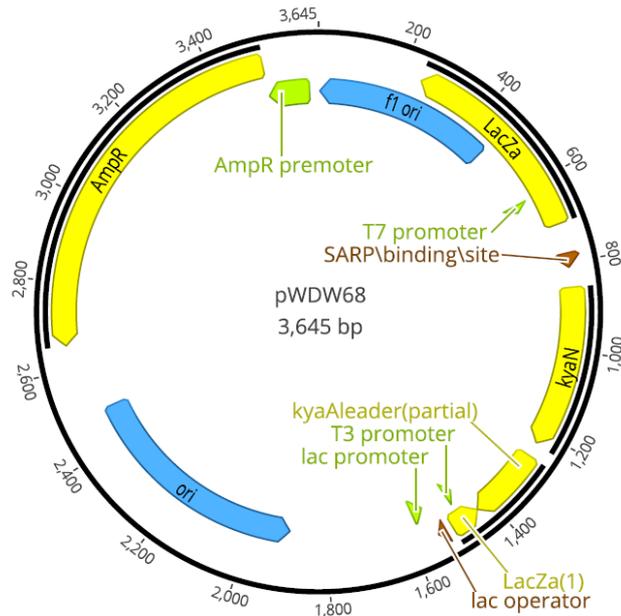


Figure S32: pWDW68 Plasmid Map. Functional genes include antibiotic resistance markers (AmpR), promoters (AmpR promoter, T7, T3, lac), origins of replication (ori, f1 ori) inserted cassette genes (SARP binding site, *kyaN*, *kyaA* partial) and LacZa genes.

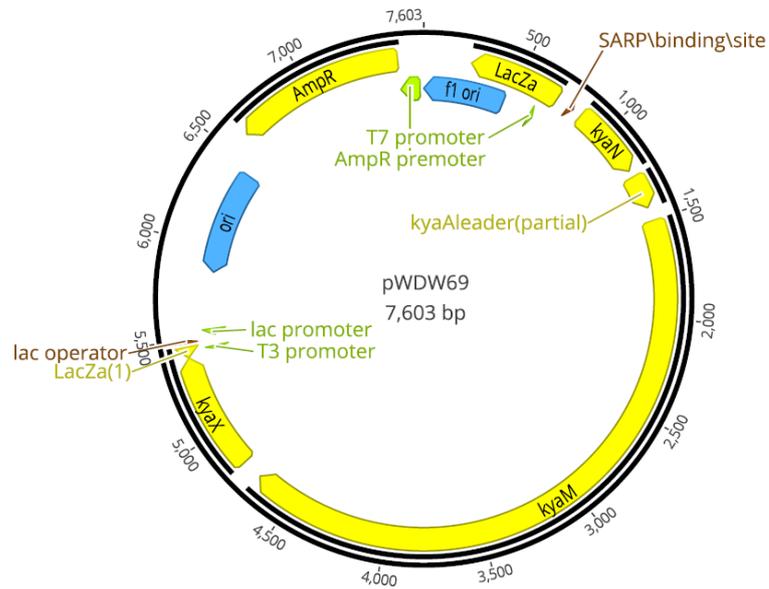


Figure S33: pWDW69 Plasmid Map. Functional genes include promoters (AmpR promoter, T7, T3, lac), antibiotic resistance markers (AmpR), inserted cassette genes (SARP binding site, *kyaN*, *kyaA* partial, *kyaM*, *kyaX*) and LacZa genes.

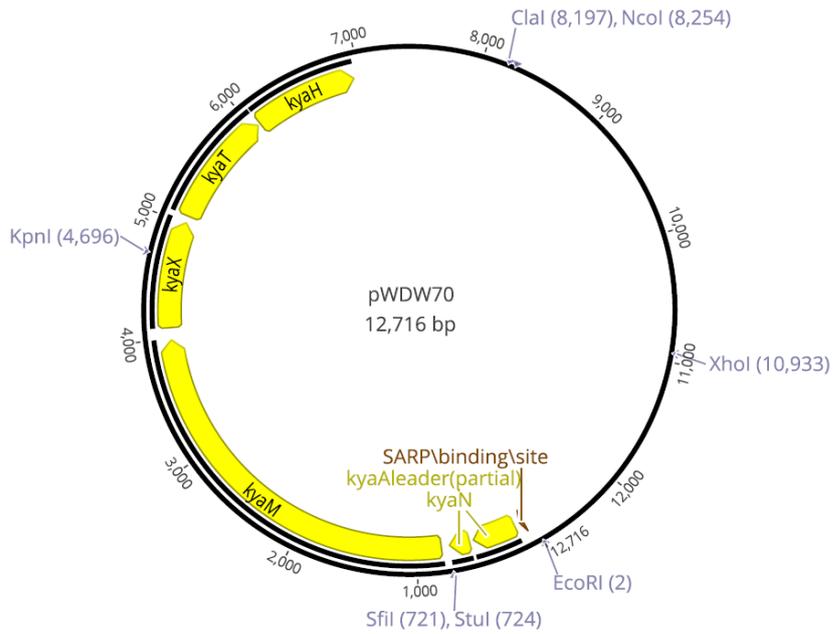


Figure S34: pWDW70 Plasmid Map. Functional genes include inserted cassette genes (SARP binding site, *kyaN*, *kyaA* partial, *kyaM*, *kyaX*, *kyaT*, *kyaH*).

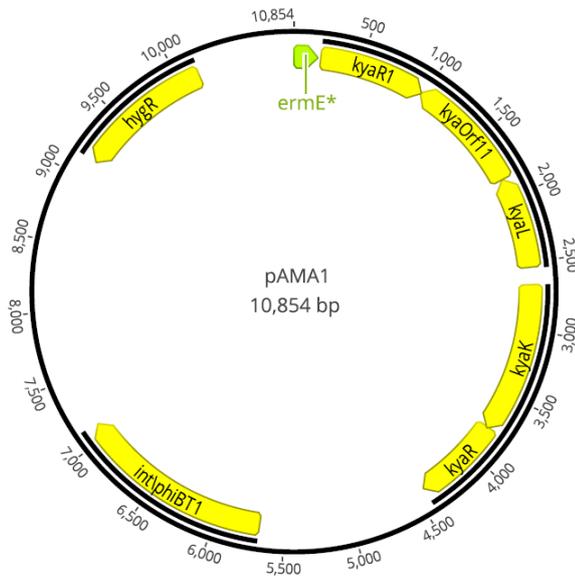


Figure S35: pAMA1 Plasmid Map. Functional genes include promoters (*ermE**), antibiotic resistance markers (*hygR*), origins of replication (*oriT*) and inserted cassette genes (*kyaR1*, *kyaOrf11*, *kyaL*, *kyaK*, *kyaR*).

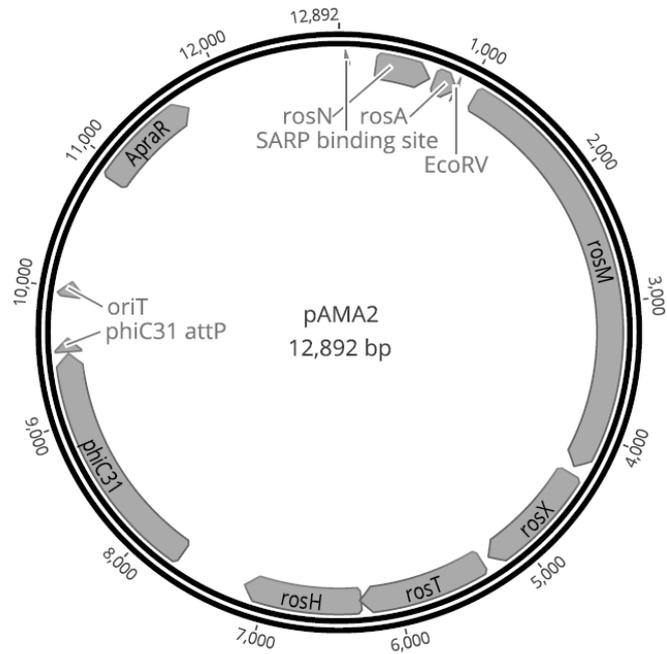


Figure S36: pAMA2 Plasmid Map. Functional genes include integrase (*phiC31*), antibiotic resistance markers (*ApraR*), origin of replication (*oriT*) and inserted cassette genes (SARP binding site, *rosN*, *rosA* partial, *rosM*, *rosX*, *rosT*, *rosH*).

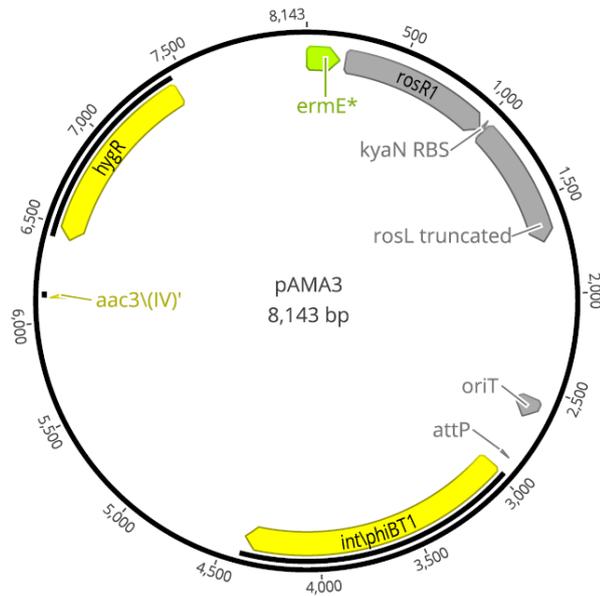


Figure S37: pAMA3 Plasmid Map. Functional genes include integrases (*phiBT1*), antibiotic resistance markers (*hygR*), promoters (*ermE**), origin of replication (*ori*) and inserted cassette genes (*rosR1*, truncated *rosL*).

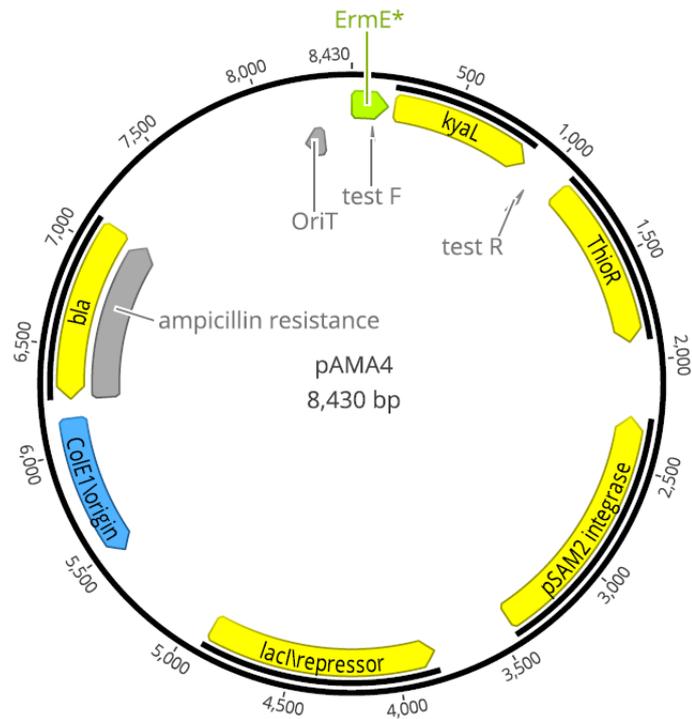


Figure S38: pAMA4 Plasmid Map. Functional genes include origins of replication (*oriT*, *ColE1*), integrase (*pSAM2*), antibiotic resistance markers (*ThioR*) and inserted cassette genes (*kyaL*).

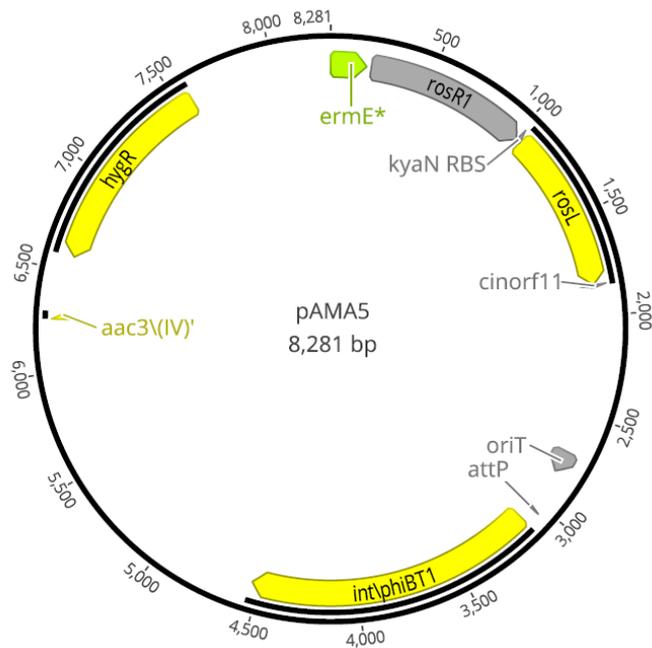


Figure S39: pAMA5 Plasmid Map. Functional genes include origins or replication (*oriT*), integrases (*phiBT1*), promoters (*ermE**), antibiotic resistance markers (*hygR*) and inserted cassette genes (*rosR1*, *rosL*).

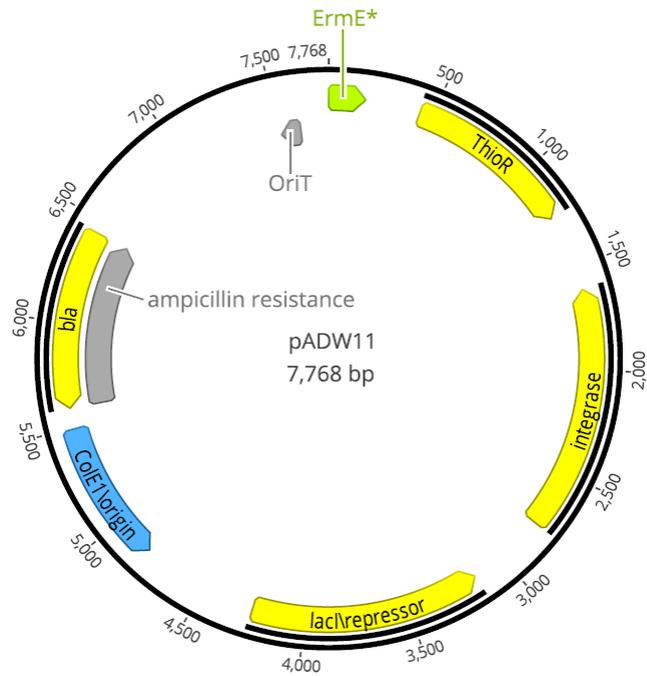


Figure S40: pADW11 Plasmid Map. Functional genes include antibiotic resistance genes (ThioR), integrase pSAM2, origins of replication (oriT, ColE), promoters (*ermE**).

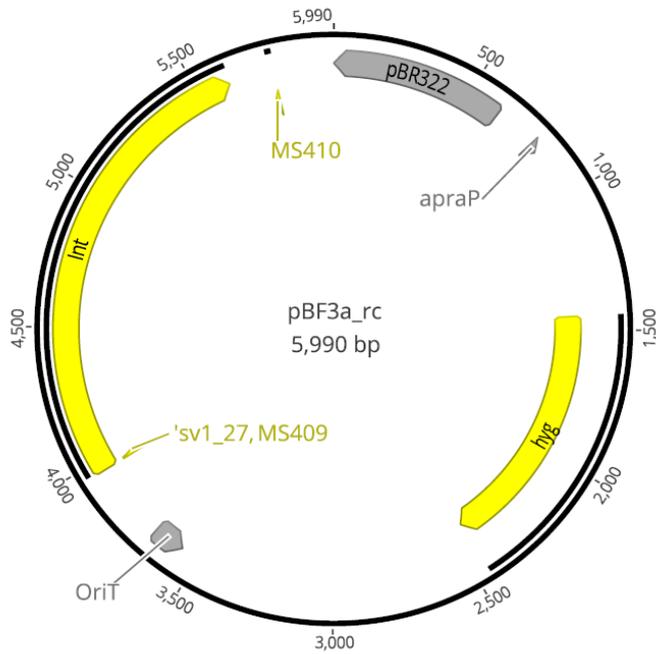


Figure S41: pBF3 Plasmid Map. Functional genes include antibiotic resistance markers (hygR), origins of replication (oriT) and integrase SV1.

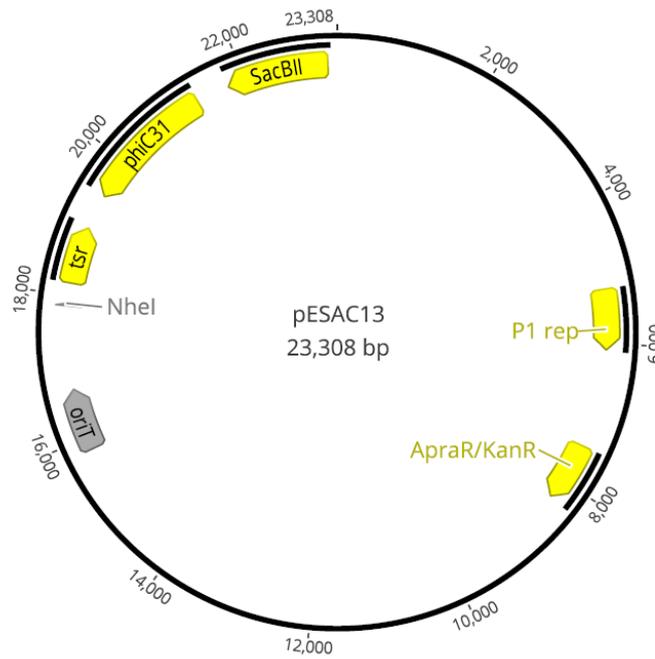


Figure S42: pESAC13 Map. Functional genes include antibiotic resistance markers (AprR, KanR), integrases (phiC31), origin of replication (oriT, phage P1 rep).

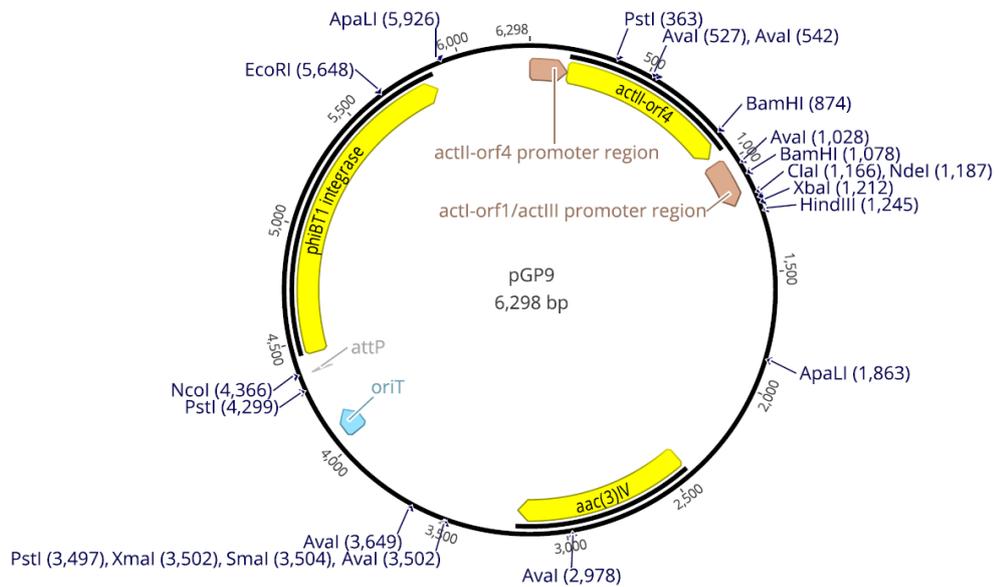


Figure S43: pGP9 Plasmid Map. Functional genes integrase (phiBT1), origins of replication (oriT), promoters (actII-orf4, actI-orf1), apramycin resistance marker (aac(3)IV).

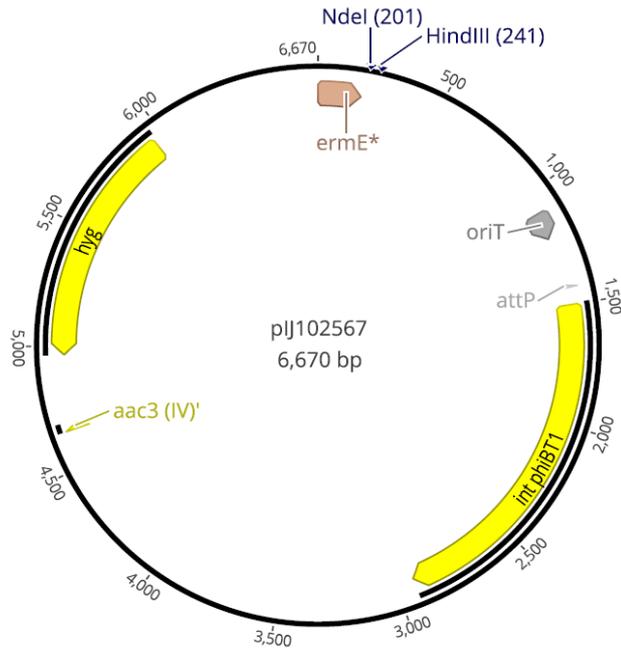


Figure S44: pIJ102567 Plasmid Map. Functional genes integrase (phiBT1), antibiotic resistance markers (hygR), promoter (*ermE**).

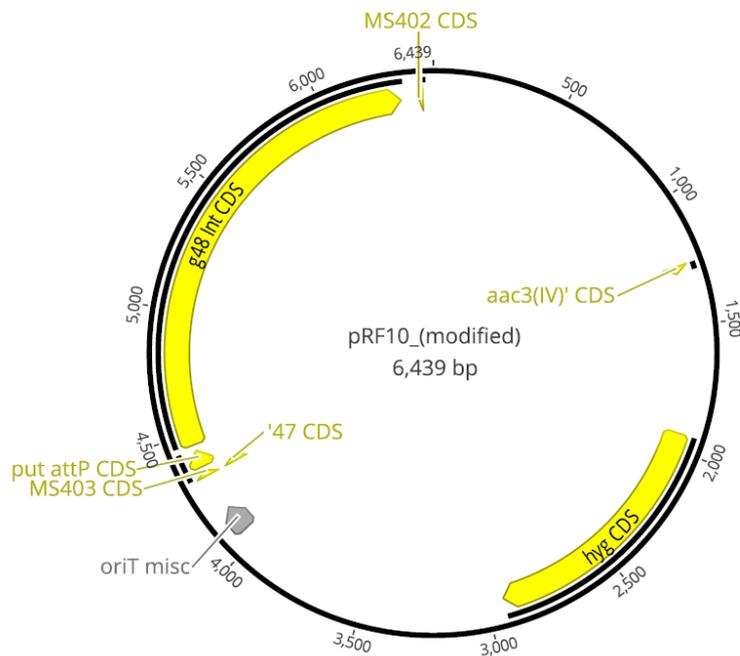


Figure S45: pRF10 Plasmid Map. Functional genes include antibiotic resistance marker (hyg CDS), origin of replication (oriT), integrase TG1.