

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

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/130107>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

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



Studies on the biosynthesis of gladiolin

by

Marianne Costa

Thesis

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

Department of Chemistry

September 2018

Contents

List of figures	I
List of tables.....	VII
Acknowledgements	IX
Declaration	X
Abbreviations	XI
Abstract	XII

Introduction.....	1
-------------------	---

1.1 Antibiotics.....	1
1.1.1 The discovery of antibiotics	1
1.1.2 Antibiotic modes of actions	2
1.1.3 Antibiotic resistance.....	5
1.2 Polyketide synthases	6
1.2.1 <i>Cis</i> -AT PKSs.....	7
1.2.1.1 The minimal catalytic module.....	7
1.2.1.2 Accessory domains.....	11
1.2.1.3 The principal of co-linearity	14
1.2.1.4 Post-PKS modifications	15
1.2.2 <i>Trans</i> -AT PKSs.....	16
1.2.2.1 Chain initiation mechanisms	18
1.2.2.2 Non-elongating KS domains	19
1.2.2.3 Unusual domain architectures.....	19
1.2.2.4 Recruitment of unusual domains.....	20
1.2.2.5 Introduction of methyl branches during chain assembly	21

1.3 Gladiolin	23
---------------------	----

O-methylation during chain assembly by *trans*-AT modular polyketide synthases30

2.1 Introduction.....	30
2.2 <i>In vitro</i> reconstitution of embedded OMT domain activity.....	32
2.2.1 Purification and characterisation of GldM11 OMTACP	32
2.2.2 Methylation assay design	33
2.2.2 Confirming the <i>in vitro</i> activity of GldM11 OMTACP	34
2.3 Identification of key catalytic residues	37
2.3.1 Sequence alignment.....	37
2.3.2 Investigating the activity of GldM11 OMT-ACP(H154A)	41
2.4 Structural studies	42
2.4.1 Crystallisation trials.....	42
2.4.2 Homology modelling	43
2.5 Phylogenetic analysis	45
2.6 Comparing the substrate tolerance of embedded OMT domains from two different clades.....	47
2.7 ACP domain specificity and the role of the KS ⁰ domain.....	50
2.7.1 The role of the KS ⁰ domain.....	50
2.7.2 ACP domain specificity in the gladiolin O-methylation module	53
2.7.3 ACP domain crosstalk	54
2.8 Summary	57

Investigating cryptic starter unit amidation59

3.1 Introduction.....	59
3.2 Identification of BGCs containing homologs of GbnM and GbnC ...	62
3.3 Investigating the role of <i>gbnM</i>	64
3.3.1 Creation of knock out mutants in <i>B. gladioli</i> BCC1622	64

3.3.2 <i>In vivo</i> studies.....	66
3.3.3 Isolation and structure elucidation of amide gladiolin	68
3.3.4 Stability of amide gladiolin in protic solvents.....	71
3.3.5 <i>In vitro</i> studies	73
3.4 Antimicrobial activity of amide gladiolin	74
3.5 Investigating the action of <i>gbnC</i>.....	79
3.5.1 <i>In vivo</i> studies.....	79
3.5.2 Complementation with the homolog SorQ	81
3.6 Exploration of polyketide chain initiation in the biosynthesis of gladiolin.....	82
3.6.1 Proposed mechanisms of chain initiation.....	82
3.6.2 Establishing the importance of <i>gbnA</i> and <i>gbnB</i> in the biosynthesis of gladiolin	85
3.6.3 Overproduction & purification of proteins involved in chain initiation ...	86
3.6.4 Investigating the substrate tolerance of GbnB.....	88
3.8 Summary	91

Conclusions and perspectives.....92

4.1 <i>O</i>-methylation during chain assembly by <i>trans</i>-AT modular polyketide synthases	92
4.1.1 Exploration of the substrate tolerance of embedded OMT domains....	92
4.1.2 Characterising the interactions between embedded OMT domains and ACP domains.....	93
4.1.3 Structural studies of embedded OMT domains.....	95
4.1.4 Bioengineering <i>O</i> -methylation	96
4.2 Investigating cryptic starter unit amidation	97
4.2.1 Amide gladiolin	97
4.2.2 The role of GbnC in the biosynthesis of gladiolin	98
4.2.3 Exploration of polyketide chain initiation in gladiolin biosynthesis	99
4.2.4 Revised gladiolin biosynthesis.....	100

Experimental97

5.1 Microbial strains and culture conditions.....	102
5.1.1 Microbial strains.....	102
5.1.2 Plasmids.....	104
5.1.3 Media.....	104
5.1.4 Culture conditions.....	106
5.2 Buffers	106
5.3 DNA manipulation and cloning procedures	108
5.3.1 DNA isolation and purification.....	108
5.3.2 PCR.....	108
5.3.3 DNA size and concentration analysis	110
5.3.4 Recombinant plasmid creation	110
5.3.4.1 Traditional cloning.....	110
5.3.4.1.1 Restriction digestion.....	110
5.3.4.1.2 Ligation.....	113
5.3.4.2 TOPO cloning	113
5.3.4.3 TA cloning.....	114
5.3.4.4 Transformation of <i>E. coli</i>	114
5.3.4.4.1 Chemical transformation	114
5.3.4.4.2 Electroporation	115
5.3.4.5 Sequencing.....	115
5.3.5 Mutagenesis	115
5.3.6 Gene deletion and complementation	116
5.3.6.1 Gene deletion.....	116
5.3.6.2 Gene complementation.....	119
5.3.6.3 Analysis of the metabolite profile of knockout mutants	120
5.3.6.4 High-resolution LC-MS.....	121
5.4 Isolation and characterisation of amide gladiolin.....	121
5.4.1 Production and purification of gladiolin and amide gladiolin	121
5.4.2 NMR spectroscopy	123
5.4.3 Stability analysis	123

5.5 Protein overproduction, purification and characterisation.....	123
5.5.1 Protein overproduction.....	123
5.5.2 Protein purification.....	126
5.5.3 SDS-PAGE analysis.....	127
5.5.4 Determining protein concentration.....	128
5.5.5 Determining protein oligomerisation state.....	128
5.5.6 Intact protein MS	129
5.6 Protein Crystallisation trials.....	129
5.6.1 Protein overproduction and purification	129
5.6.2 Crystallisation trials.....	130
5.6.3 Cleavage of the hexahistidine tag.....	131
5.7 Assays.....	131
5.7.1 Biochemical assays.....	131
5.7.1.1 Methylation assays	131
5.7.1.2 Transacylation assay	132
5.7.1.3 ACP-OMT interaction assays.....	133
5.7.1.4 GbnM activity assay.....	134
5.7.1.5 AT domain loading assay.....	135
5.7.2 Antimicrobial activity assays.....	136
5.7.2.1 Broth microdilution method	136
5.7.2.2 Resazurin microtiter assay.....	136
5.8 Computational methods.....	137
5.8.1 ClusterTools searches.....	137
5.8.2 Sequence alignments	138
5.8.3 Phylogenetic analysis	138
5.8.4 Homology modelling of GldM11 OMT.....	138
References	139
Appendices	149

List of figures

Figure 1.1: The discovery and structure of the first antibiotic, penicillin. The image (which has been modified) was taken from the paper on penicillin by A. Fleming	1
Figure 1.2: The structure of a range of natural product antibiotics from the different classes discovered during the golden era of antibiotics.....	3
Figure 1.3: Antibiotic targets. Examples of antibiotics that act upon each of the targets are listed.....	4
Figure 1.4: The biosynthesis of 6-dEB.....	8
Figure 1.5: Mechanism of the conversion of <i>apo</i> -ACP domains to <i>holo</i> -ACP domains catalysed by 4'-phosphopantetheinyl transferase.....	9
Figure 1.6: Examples of commonly used starter and extender units in polyketide biosynthesis by <i>cis</i> -AT PKSs.....	10
Figure 1.7: Mechanism of the loading of the extender unit (methylmalonyl-CoA for the 6-dEB PKS) or starter unit (propionyl-CoA for the 6-dEB PKS) onto the PKS catalysed by the AT domain.	10
Figure 1.8: Mechanism of intermodular polyketide translocation and chain elongation catalysed by the KS domain.....	11
Figure 1.9: Mechanism of the reduction of the β -keto group catalysed by the KR domain.....	12
Figure 1.10: Mechanism of the reduction of the β -hydroxyl group catalysed by the DH domain	13
Figure 1.11: Mechanism of the reduction of the α / β -double bond catalysed by the ER domain.....	13
Figure 1.12: The module structure of <i>trans</i> -AT PKSs compared to <i>cis</i> -AT PKSs.....	16

Figure 1.13: Pathway for bacillaene biosynthesis, showing the ACP-bound polyketide intermediates following α - and β -carbon processing	17
Figure 1.14: The most commonly utilised chain initiation mechanisms by <i>trans</i> - and <i>cis</i> -AT PKSs.	19
Figure 1.15: Mechanism of double bond migration catalysed by the DH* domain	21
Figure 1.16: Mechanism of <i>O</i> -methylation catalysed by embedded OMT domains.....	22
Figure 1.17: General mechanism for the formation of β -branches in PKSs..	23
Figure 1.18: The structure of gladiolin and etnangien.....	24
Figure 1.19: The organisation of the gladiolin biosynthetic gene cluster in <i>B. gladioli</i> BCC0238	25
Figure 1.20: Proposed pathway for gladiolin biosynthesis, showing the ACP-bound predicted polyketide intermediates following α - and β -carbon processing.....	28
Figure 2.1: A closer look at the <i>O</i> -methylation module from the gladiolin biosynthetic pathway.....	31
Figure 2.2: (A) 10 % SDS-PAGE gel showing purified GldM11 OMT-ACP. (B) A graph showing the results of size exclusion chromatography for GldM11 OMT-ACP.....	32
Figure 2.3: (A) Overview of the methylation assays conducted. (B) The enzymatic reactions that were used to convert the pantetheine substrates to CoA analogues.....	34
Figure 2.4: (A) Stacked deconvoluted ESI-MS spectra of GldM11 OMT-ACP with ACP bound simplified substrate mimics after incubation with SAM. (B) A closer look at the deconvoluted spectrum for the (<i>S</i>)-3-hydroxyoctanyl substrate mimic. (C) Spectrum showing the key phosphopantetheine ejection ions for the (<i>S</i>)-3-hydroxyoctanyl substrate mimic.	35
Figure 2.5: (A) Stacked deconvoluted ESI-MS spectra of GldM11 OMT-ACP with the 3,5-dihydroxyhexanoyl ACP bound substrate mimics after incubation	

with SAM. (B) Mechanism of lactonization of the 3,5-dihydroxyhexanoyl substrates, the transition states are shown..	37
Figure 2.6: The structure of secondary metabolites produced by <i>trans</i> -AT PKSs that utilise embedded <i>OMT</i> domains during chain assembly	39
Figure 2.7: Sequence logo, constructed using WebLogo, of a sequence alignment of all reported embedded <i>OMT</i> domains from <i>trans</i> -AT PKSs, showing the conserved histidine residue proposed to be the key catalytic base	40
Figure 2.8: The proposed mechanism for SAM dependent methylation in <i>trans</i> -AT embedded <i>OMTs</i> .	40
Figure 2.9: Stacked deconvoluted ESI-MS spectra of GldM11 <i>OMT</i> -(<i>S</i>)-3-hydroxyoctanoyl-ACP (A) and GldM11 <i>OMT</i> -(3 <i>S</i> ,5 <i>S</i>)-3,5-dihydroxyhexanoyl-ACP (B) after a 1 h 30 min incubation with SAM. GldM11 <i>OMT</i> -ACP WT is represented in purple and GldM11 <i>OMT</i> -ACP(H153A) is shown in grey.....	41
Figure 2.10: 8 % SDS-PAGE gel showing the purified proteins that were used for structural studies.....	43
Figure 2.11: Homology model of Gld <i>OMT</i> generated using the intensive mode of Phyre2. (A) Overall model with the defining class I MT fold shown in purple. (B) Close up showing the proximity of the proposed key catalytic histidine (purple) to SAH (teal).	44
Figure 2.12: Phylogenetic tree of <i>OMT</i> domains from modular type I PKSs. Embedded <i>OMT</i> domains are highlighted, with those found in <i>trans</i> -AT PKSs shown in purple and those from <i>cis</i> -AT PKSs shown in blue	45
Figure 2.13: The hypothesised substrates for the embedded <i>OMT</i> domains from <i>trans</i> -AT PKSs	47
Figure 2.14: 10 % SDS-PAGE gel showing the purified proteins, MisM9 <i>OMT</i> -ACP, MisM1 <i>OMT</i> -ACP and RizM18 <i>OMT</i> -ACP.....	48
Figure 2.15: A summary of the substrate tolerance determined for several embedded <i>OMT</i> domains from <i>trans</i> -AT PKSs.	50
Figure 2.16: 15 % SDS-PAGE gel showing the purified ACP domains used in this study.	51

Figure 2.17: Stacked deconvoluted ESI-MS spectra of GldM11-(<i>S</i>)-3-hydroxyoctanoyl-ACP(I) and <i>holo</i> -GldM11ACP(II) after incubation with and without the GldM11KS OMT didomain	52
Figure 2.18: Stacked deconvoluted ESI-MS spectra of (<i>S</i>)-3-hydroxyoctanoyl-ACP domains from the gladiolin methylation module, after incubation with Gld OMT	54
Figure 2.19: Stacked deconvoluted ESI-MS spectra of (<i>S</i>)-3-hydroxyoctanoyl-ACP(II) domains, after incubation with Gld OMT. Methylation (+ 14 Da) is only observed for GldM11ACP(II) and MisM9ACP(II).....	55
Figure 2.20: Stacked deconvoluted ESI-MS spectra of (<i>R</i>)-3-hydroxyoctanoyl-ACP(II) domains, after incubation with MisM1 OMT.....	57
 Figure 3.1: (A) The gladiolin biosynthetic gene cluster, the two genes of interest <i>gbnC</i> and <i>gbnM</i> are highlighted in purple. (B) The general activity of the proposed enzymes.....	60
Figure 3.2: (A) Examples of protecting groups used in the biosynthesis of desertomycin and vicienistatin. (B) Schematic showing how the intermediate proposed in the initiation of gladiolin biosynthesis would be prone to cyclisation.....	61
Figure 3.3: (A) The structure of secondary metabolites produced by BGCs containing homologs of both GbnC and GbnM (B) Percentage identity matrices showing the similarity between each gene and their identified homologs.....	63
Figure 3.4: Overview of the homing endonuclease I-SceI mutagenesis system used to generate deletion mutants in <i>B. gladioli</i> BCC1622.....	65
Figure 3.5: Agarose gel electrophoresis showing genetic evidence for the creation of a <i>gbnM</i> deletion mutant in <i>B. gladioli</i> BCC1622.....	66
Figure 3.6: Base peak chromatograms from LC-MS analyses of extracts from agar-grown cultures of <i>B. gladioli</i> BCC1622 WT, <i>B. gladioli</i> BCC1622Δ <i>gbnM</i> and <i>B. gladioli</i> BCC1622Δ <i>gbnM</i> following <i>in trans</i> expression of <i>gbnM</i>	67
Figure 3.7: Chemical structure of amide gladiolin showing the key HMBC and COSY correlation confirming the position of the amide moiety.	69

Figure 3.8: (A) The conversion of amide gladiolin to <i>iso</i> -amide gladiolin. (B) Base peak chromatograms from the LC-MS time-course analysis showing the stability of amide gladiolin and gladiolin in methanol.....	72
Figure 3.9: (A) Overview of the assay used to demonstrate activity of GbnM <i>in vitro</i> . (B) 10 % SDS-PAGE analysis showing the presence of GbnM in the cell lysate. (C) Base peak chromatograms from LC-MS analysis of the assay showing the conversion of amide gladiolin to gladiolin by GbnM.....	74
Figure 3.10: The rezasurin microtiter assay used to determine the MIC of amide gladiolin and gladiolin for <i>M. smegmatis</i>	77
Figure 3.11: Image of sorangicin (purple) in complex with <i>Thermus aquaticus</i> RNA polymerase (PDB: 1YNJ ¹)	78
Figure 3.12: Agarose gel electrophoresis showing genetic evidence for the creation of a <i>gbnC</i> deletion mutant in <i>B. gladioli</i> BCC1622.....	79
Figure 3.13: (A) Base peak chromatograms from LC-MS analyses of extracts from agar-grown cultures of <i>B. gladioli</i> BCC1622 WT, <i>B. gladioli</i> BCC1622Δ <i>gbnC</i> and <i>B. gladioli</i> BCC1622Δ <i>gbnC</i> following <i>in trans</i> expression of GbnC. The extracted ion chromatogram for the [M+Na] ⁺ ion is also shown. (B) Base peak chromatograms from LC-MS analyses of extracts from cultures of <i>B. gladioli</i> BCC1622Δ <i>gbnC</i> where the media was supplemented with increasing concentrations of arabinose.....	80
Figure 3.14: Base peak chromatograms from LC-MS analyses of extracts from agar-grown cultures of <i>B. gladioli</i> BCC1622 WT, <i>B. gladioli</i> BCC1622Δ <i>gbnC</i> and <i>B. gladioli</i> BCC1622Δ <i>gbnC</i> following <i>in trans</i> expression of <i>sorQ</i> . The extracted ion chromatogram for the [M+Na] ⁺ ion is also shown.	82
Figure 3.15: Proposed chain initiation mechanism in the biosynthesis of the glutarimide-containing polyketides	83
Figure 3.16: Proposed mechanisms of polyketide chain initiation in the biosynthesis of gladiolin.....	84
Figure 3.17: (A) Agarose gel electrophoresis showing genetic evidence for the creation of a <i>gbnA</i> and <i>gbnB</i> deletion mutant in <i>B. gladioli</i> BCC1622. (B) Base peak chromatograms from LC-MS analyses of extracts from agar-grown cultures of <i>B. gladioli</i> BCC1622 WT, and the generated mutants.....	86

Figure 3.18: (A) 12 % SDS-PAGE analysis of the purified AT domains GbnB and EtnK. (B) 15 % SDS-PAGE analysis of the purified ACP domains GbnA and GldM1ACP(I).	87
Figure 3.19: Stacked deconvoluted ESI-MS spectra of the standalone AT domains GbnB, EtnK and PksC showing whether they are capable of accepting succinyl-CoA, malonyl-CoA or succinamic-CoA as substrates.	89
Figure 3.20: (A) Overview of the VbxP catalysed reaction used to generate succinamic-CoA. (B) Base peak chromatograms from the LC-MS analysis of the CoA transferase, VbxP, catalysed reaction of acetyl-CoA with succinamic acid and the control reaction using boiled VbxP.	90
Figure 4.1: Proposed substrates for further exploring the substrate tolerance of embedded OMT domains from <i>trans</i> -AT PKSs.	93
Figure 4.2: (A) 8 % SDS-PAGE gel showing purified GldM11KS-OMT-ACP. (B) 2D class averages of the negative stain TEM images of the tri-domain protein obtained by classification of different conformations using EMAN 2	96
Figure 4.3: Proposed bioengineering of the gladiolin PKS to introduce methylation of the hydroxyl-group at position 23.	97
Figure 4.4: Revised pathway for gladiolin biosynthesis, showing the ACP-bound predicted polyketide intermediates following α - and β -carbon processing.	101

List of tables

Table 1.1: The putative functions of the non-PKS-encoding genes in the gladiolin BGC.	26
Table 3.1: NMR assignments for amide gladiolin (DMSO- d_6 , 600 Hz) compared to those published for gladiolin.....	69
Table 3.2: MIC values determined for amide gladiolin against <i>C. albicans</i> and a range of gram-negative and gram-positive bacteria. The MIC values determined for gladiolin in this study are also shown as a comparison.	75
Table 3.3: MIC values determined for amide gladiolin against <i>M. smegmatis</i> and <i>M. tuberculosis</i>	77
Table 5.1: Bacterial and fungal strains used within this project..	102
Table 5.2: Plasmids used within this project.	104
Table 5.3: Antibiotics used within this project.	106
Table 5.4: Volumes of reagents used for 25 μ L PCR reactions.....	109
Table 5.5: Oligonucleotide primers used for the creation of recombinant plasmids.....	111
Table 5.6: Oligonucleotide primers used for the creation of pET151 constructs.....	113
Table 5.7: Oligonucleotide primers used for the creation of the pETSUMO- <i>gbcC</i> construct.....	114

Table 5.8: Oligonucleotide primers used for the creation of the GldM11 OMTACP(H153A) mutant, the mutated residues are shown in purple.	116
Table 5.9: Oligonucleotide primers used for the creation of pGPI vectors	117
Table 5.10: Oligonucleotide primers used for colony PCR to screen for successful crossover mutants.	118
Table 5.11: Oligonucleotide primers used for the creation of pMLBAD vectors, the restriction sites used are underlined.	120
Table 5.12: Elution conditions for high-resolution LC-MS analysis.	121
Table 5.13: Elution conditions for HPLC purification of amide gladiolin and gladiolin.	122
Table 5.14: Overproduction conditions for recombinant proteins.	124
Table 5.15: Volumes of reagents used in the preparation of SDS-PAGE gels.	127
Table 5.16: Elution conditions for intact protein MS.	129
Table 5.17: The composition of the methylation assays.	132
Table 5.18: The composition of the ACP domain loading reactions.	133

Acknowledgements

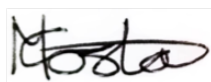
First and foremost I would like to thank Professor Greg Challis for giving me the opportunity to work on such a varied and exciting project. I would then like to thank Dr Lona Alkhalaf, alongside Greg, for her continued guidance and advice throughout. One of the things I have enjoyed most about doing this PhD, is having so many opportunities to learn new skills, so I must give special thanks to everyone who has taken time to share their expertise with me: Dr Matthew Jenner, Xinyun Jian, Dr Yousef Dashti, Dr Chris Fage, Dr Dani Zabala, Christian Hobson, Dr Emzo de los Santos, Dr Simone Kosol and Dr Joleen Masschelein. I would also like to give thanks to everyone who has collaborated with me on this project.

Finally, I would like to thank all members of the Challis group, past and present, for their friendship and the many good memories. I also need to thank Michael and my wonderful friends for keeping me smiling throughout this project. Thanks are also owed to BBSRC MIBTP for funding.

Declaration

The experimental work reported in this thesis is original research carried out by the author, unless stated otherwise, in the Department of Chemistry, University of Warwick, between September 2015 and September 2018. No material has been submitted in any previous application for any degree.

Contributions from collaborators are references throughout the text.



Marianne Costa

Date: 30/09/18

List of abbreviations

A	Adenylation (domain)
ACP	Acyl carrier protein (domain)
AH	Acyl hydrolase (domain)
AL	Acyl ligase (domain)
Amp	Ampicillin
ARG	Antibiotic resistance gene
AT	Acyltransferase (domain)
ATP	Adenosine triphosphate
BGC	Biosynthetic gene cluster
bp	Base pairs
C	Condensation (domain)
CMT	C-methyltransferase (domain)
CoA	Coenzyme A
6-dEB	6-deoxyerythronolide B
DH	Dehydratase (domain)
DH*	Enoyl isomerase (domain)
dH ₂ O	Deionized water
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
ECH	Enoyl-CoA dehydratase (enzyme)
EIC	Extracted ion chromatogram
ER	Enoylreductase (domain)
ESI	Electrospray ionization
FAS	Fatty acid synthase
GNAT	GCN5-related <i>N</i> -acetyltransferase (domain)
h	Hour(s)
HMGS	3-hydroxy-3-methyl glutaryl-CoA synthase

HMM	Hidden Markov model
HPLC	High performance liquid chromatography
IPTG	Isopropyl- β -D-thiogalactoside
Kan	Kanamycin
KR	Ketoreductase (domain)
KS	Ketosynthase (domain)
KS ⁰	Non-elongating ketosynthase (domain)
MCS	Multiple cloning site
MIC	Minimum inhibitory concentration
min	Minute(s)
MS	Mass spectrometry
NADPH	β -nicotinamide adenine dinucleotide 2'-phosphate
NMR	Nuclear magnetic resonance
NRPS	Non-ribosomal peptide synthase
NTA	Nitriloacetic acid
OD ₆₀₀	Optical density at 600 nm
OMT	O-methyltransferase (domain)
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PKS	Polyketide synthase
PMX	Polymyxin
RNA	Ribonucleic acid
RT	Room temperature
SAH	S-adenosyl homocysteine
SAM	S-adenosyl methionine
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
T _A	Annealing temperature
T _c	Tetracycline
TE	Thioesterase (domain)
Tp	Trimethoprim
WT	Wild type

Abstract

Gladiolin is a macrolide antibiotic with potent activity against *Mycobacterium tuberculosis*. It is biosynthesised by a *trans*-AT polyketide synthase (PKS) with many unique features including an *O*-methylation module, with a dedicated *O*-methyltransferase (OMT) domain. In this work, the OMT-ACP didomains from the gladiolin PKS, as well as three others from *trans*-AT PKSs, were overproduced and purified. Their activity and substrate tolerance were determined, providing the first biochemical proof of *O*-methylation on polyketide assembly lines and revealing useful substrate promiscuity. The OMT domains were shown to fall into two clades, which appear to correspond their ability to interact with specific ACP domains.

The gene cluster also contained two putative auxiliary genes of unknown function; *gbnC* that encodes for an asparagine synthetase, and *gbnM* which encodes for an amidase. Inactivation of *gbnM* generated a mutant producing a novel amide derivative of gladiolin, which maintained antimicrobial activity. Purified GbnM was shown to catalyse hydrolysis of the amide derivative to produce gladiolin. Although it was established, by *in vivo* gene deletion, that *gbnC*, proposed to be responsible for the initial amidation, plays an essential biosynthetic role, *in vitro* studies were not possible due to insolubility of the protein.

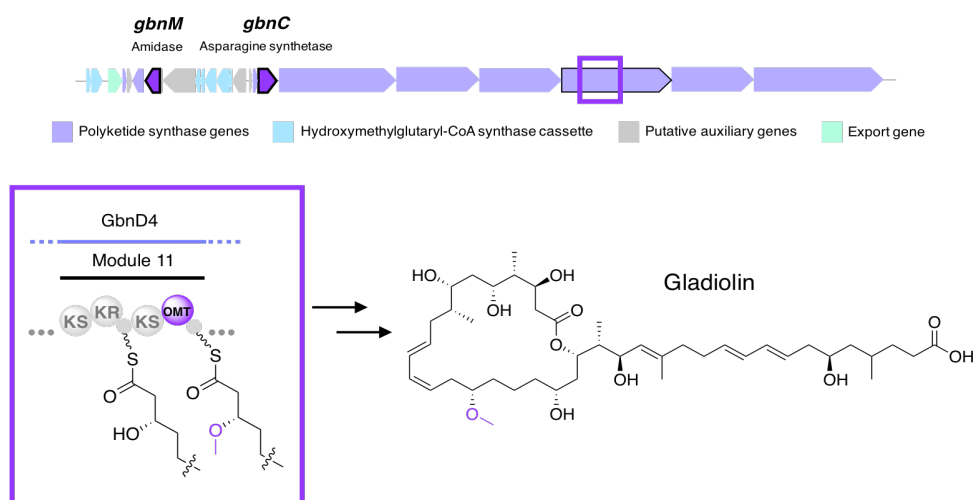


Figure 1: Gladiolin structure and biosynthetic gene cluster, highlighting the OMT domain containing region, and the genes *gbnM* and *gbnC*.

Chapter 1

Introduction

1.1 Antibiotics

1.1.1 The discovery of antibiotics

The first discovered antibiotic, penicillin (**1**), has now been in clinical use for over 70 years. It was discovered serendipitously by the microbiologist Alexander Fleming in 1928, when he observed a zone around a contaminating fungus, *Penicillium notatum*, on an agar plate in which bacteria did not grow, and decided to investigate further^{2,3} (Figure 1.1). It took over a decade, and contributions from scientists worldwide to establish a method to mass-produce penicillin for clinical use, but since then penicillin and other antibiotics have revolutionised the field of medicine.

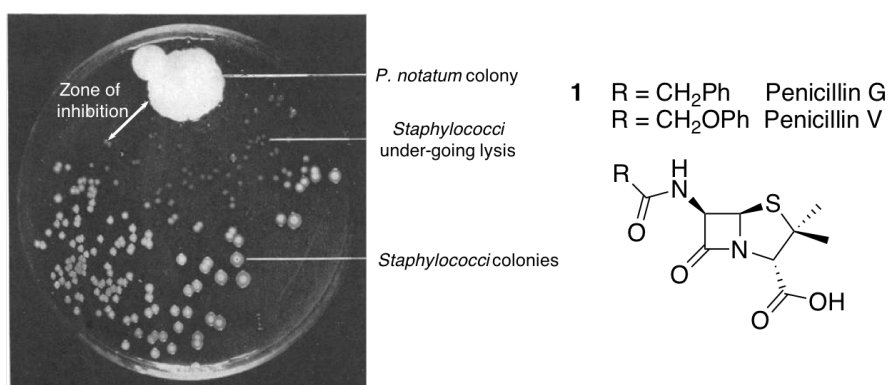


Figure 1.1: The discovery and structure of the first antibiotic, penicillin. The image (which has been modified) was taken from the paper on penicillin by A. Fleming³.

The discovery of penicillin opened the eyes of the scientific community to the potential of biologically active secondary metabolites, and between the 1940's and 1960's nearly all of the natural product antibiotic classes used today were identified (Figure 1.2)⁴. The success of the golden era of antibiotic discovery can be in part accredited to Selman Waksman, who pioneered the systematic screening of soil-derived *Actinomyces*. The Waksman platform was simple and based on the principals of penicillin discovery; the soil-derived *Actinomyces* were screened for antimicrobial activity against test microorganisms by detecting zones of growth inhibition on an overlay plate⁵. The capability of the platform was realised in 1943 with the discovery of the first treatment for *Mycobacterium tuberculosis*, streptomycin (**2**), which was isolated from *Streptomyces griseus*⁶. After twenty years this platform became significantly less utilised, however, owing to the increasing difficulty of identifying novel compounds against the background of known compounds, as well as the emergence of antibiotic resistance. The appearance of antibiotic resistance diverted efforts towards chemically modifying the existing antibiotic scaffolds to overcome this problem, and away from the discovery of new antibiotics.

1.1.2 Antibiotic modes of action

For discovered antibiotics to be clinically applicable they must be chemically stable, readily soluble, be able to easily enter bacterial cells and inhibit critical pathways in pathogens that are absent or significantly different in mammalian cells. Antibiotics can be either bactericidal and kill bacteria (penicillin (**1**) and rifampicin (**5**)), or bacteriostatic and prevent the growth of bacteria (erythromycin (**4**) and chlortetracycline (**6**)).

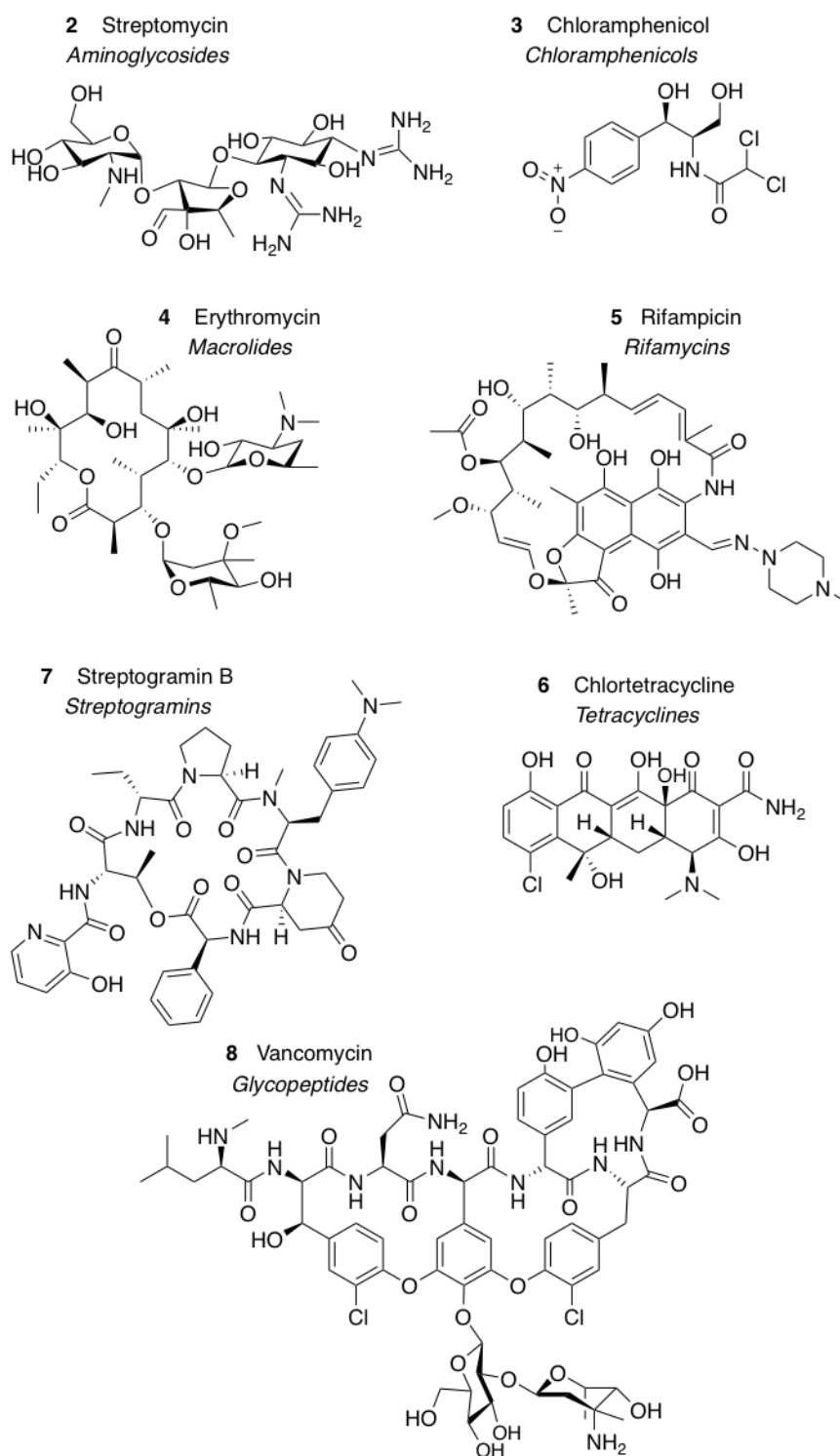


Figure 1.2: The structure of a range of natural product antibiotics from the different classes discovered during the golden era of antibiotics.

There are three main antibiotic targets; cell wall synthesis, DNA and RNA synthesis and protein synthesis (Figure 1.3). Penicillin (**1**) and vancomycin (**8**) are examples of antibiotics that act by inhibiting bacterial cell wall biosynthesis. The cell wall is made up of peptidoglycan, which is a covalently cross-linked polymer matrix made up of sugars and amino acids. Penicillin (**1**) works by preventing cross-linking of peptidoglycan whereas vancomycin (**8**) acts as a steric inhibitor of peptidoglycan maturation^{7,8}. Although these two antibiotics target different steps in cell wall biosynthesis, they both ultimately lead to the loss of structural integrity of the bacterial cell, resulting in cell death by lysis. Rifampicin (**5**) on the other hand induces bacterial cell death by binding to RNA polymerase, and sterically preventing RNA elongation when the transcript becomes 2 to 3 nucleotides in length⁹. Generally, antibiotics that inhibit protein synthesis by binding to the ribosome are bacteriostatic rather than bactericidal. The ribosome consists of two subunits (30S and 50S subunits) both of which are targets for antibiotics. Chlortetracycline (**6**) is an example of an antibiotic that acts on the 30S subunit and prevents attachment of aminoacyl-tRNAs to the ribosome that are carrying the amino acid building blocks required for protein assembly¹⁰.

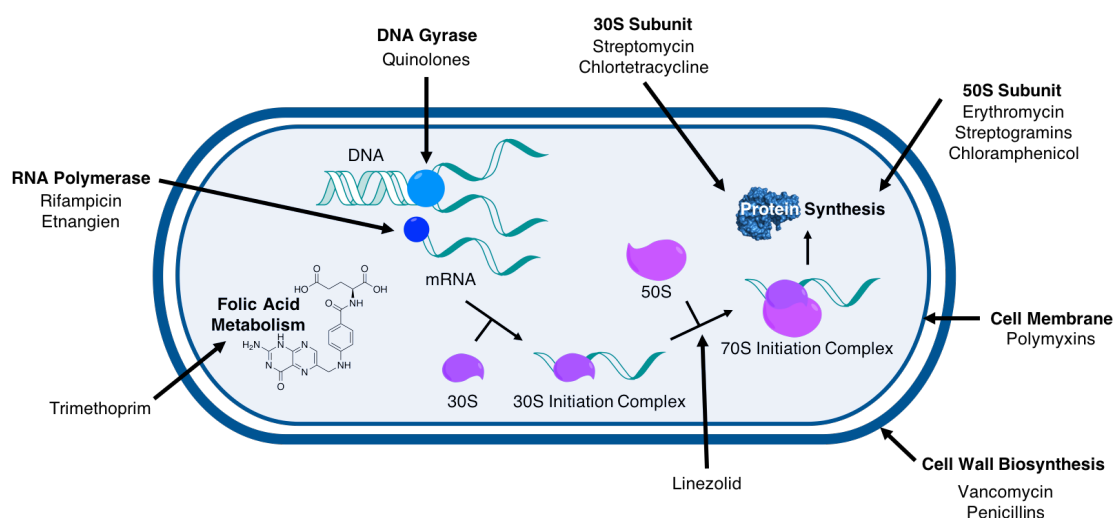


Figure 1.3: Antibiotic targets. Examples of antibiotics that act upon each of the targets are listed.

1.1.3 Antibiotic resistance

Antibiotic resistance was quickly observed in pathogenic bacteria, with resistance having been identified for all clinically used antibiotics⁵. Bacteria are able to acquire resistance *via* mutations in their chromosomal genes and by horizontal gene transfer¹¹. Antibiotic resistance genes (ARGs) pre-existed the clinical use of antibiotics owing to the fact that bacteria have encountered such compounds, which are produced by other microorganisms, in nature for millions of years, and evidence is emerging that some of the ARGs found in pathogenic bacteria actually originate from the producing microorganisms, which possess such genes to prevent suicide during the production of antibiotics¹².

There are three major resistance mechanisms employed by bacteria. The first, and one of the most successful mechanisms involves the use of enzymes to destroy or modify the antibiotic so that efficacy is lost¹³. Another mechanism involves the modification of the antibiotic target, this can be through genetic mutations or post-translational enzymatic modification of the target site. Vancomycin (**8**), an antibiotic that acts by inhibiting cell wall biosynthesis, is an interesting example where the antibiotic target is completely remodelled to avoid antibiotic binding. When resistant bacteria detect the presence of vancomycin they begin to change the cell wall composition so that efficacy is lost¹⁴. The third mechanism involves preventing the antibiotic from accessing the target. This is achieved by either reducing the permeability of the outer membrane to the antibiotic or by overexpressing efflux pumps to remove the antibiotic from the cell¹³. Often resistance to the same antibiotic is achieved through more than one of these mechanisms.

Antibiotic resistance is now becoming a worldwide problem, due to the increasing prevalence of multi-drug resistant bacteria, which are difficult or in some cases impossible to treat. The need to develop new therapies to overcome this problem has put antibiotic discovery back in the limelight. The

recent advent in genomics revealed that most bacterial species in fact remain uncultured, it is estimated that uncultured species outnumber cultured species by two or three orders of magnitude¹⁵. This highlights the huge, still untapped, potential of microorganisms in the search for new drugs, especially owing to the fact that developments in metagenomics enables bypassing the requirement of cultivation¹⁶. Much research is also being carried out to try and obtain a greater understanding of the complex biosynthetic machineries that are responsible for microbial natural products, in order to try and realise the potential of these systems for manipulation to create libraries of novel compounds with new activities.

1.2 Polyketide synthases

Polyketide synthases (PKSs) play an important role in microbial secondary metabolism and constitute one of the major mechanisms of antibiotics biosynthesis. Erythromycin (**4**), rifampicin (**5**) and chlortetracycline (**6**) are all polyketide antibiotics and despite their structural diversity they are all assembled from small acyl units by PKSs^{17–19}. PKSs are evolutionarily related to fatty acid synthases (FASs), which play a fundamental role in primary metabolism and can be found across all organisms^{20,21}. Many parallels can be drawn between these two biosynthetic machineries, thus logically PKSs were classified in the same way as FASs, according to the architecture and activity of their biosynthetic enzymes.

Briefly, PKSs can be divided into three major classes. Type I PKSs are large, multifunctional enzymes that are arranged into modules, and within each module there are a number of active sites that are referred to as domains. The covalently bound biosynthetic intermediates are passed, in an orderly manner, from one domain to the next to build the polyketide²². Type II PKSs typically produce aromatic compounds (for example chlortetracycline (**6**)) and consist of a set of discrete mono- or bi-functional enzymes that are used iteratively to

assemble the polyketide chain, that is tethered to the PKS throughout the biosynthesis²³. Type III PKSs iteratively utilise a single active site for all stages of the biosynthesis, differing fundamentally from the other two classes as the polyketide chain is built on coenzyme A (CoA), and is thus not always covalently attached to the PKS²⁴.

Increasing numbers of PKSs are being discovered, however, that don't fit the type I, II and III paradigm but fit somewhere in-between these PKS classes²⁵. To help account for this, type I PKSs can be further classified as either iterative, where a single module is used repeatedly to build the polyketide, or modular²⁶. Modular type I PKS can be further classified still as either *cis*-AT PKSs or *trans*-AT PKSs, and will be discussed in greater detail in the following sections.

1.2.1 *Cis*-AT PKSs

The PKS responsible for the biosynthesis of 6-deoxyerythronolide B (6-dEB, **9**), the precursor to the clinically used antibiotic erythromycin (**4**), is an archetypal example of a *cis*-AT PKS and is one of the most-well studied PKSs (Figure 1.4). It consists of three PKS proteins and seven modules; the first module is responsible for the loading of the starter unit propionyl-CoA and the remaining 6 modules catalyse chain elongation, each incorporating a methylmalonyl-CoA extender unit, and modification of the growing polyketide chain¹⁷.

1.2.1.1 The minimal catalytic module

Each module responsible for catalysing chain elongation minimally consists of three domains: an acyl carrier protein (ACP) domain, an AT domain and a ketosynthase (KS) domain. The central component of each module is the ACP domain; a small, non-catalytic protein consisting of a 4-helical bundle²⁷. Despite its simple structure, the ACP domain mediates interactions with all

domains within the module, the KS domain of the downstream module and 4'-phosphopantetheinyl transferase, which converts the inactive *apo*-ACP domain to the active *holo*-form. 4'-phosphopantetheinyl transferase catalyses the addition of the 4'-phosphopantetheine moiety of CoA onto a conserved serine residue of the ACP domain, as shown in Figure 1.5²⁸. This creates a phosphopantetheinyl arm onto which intermediates in the biosynthesis are covalently bound.

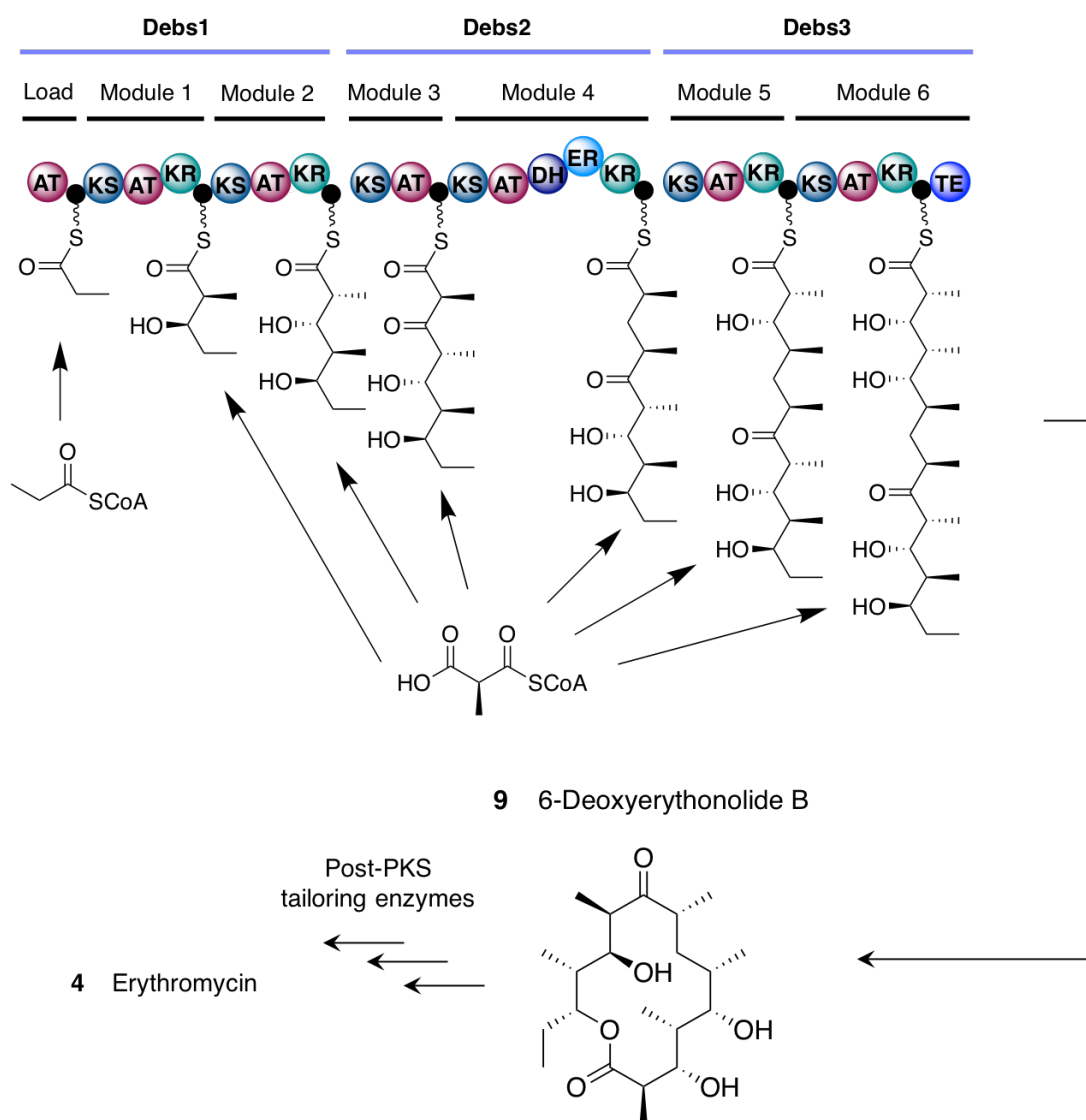


Figure 1.4: The biosynthesis of 6-dEB. The polyketide intermediates after the incorporation of each extender unit and the modifications of the growing polyketide chain are shown. Black circles represent ACP domains.

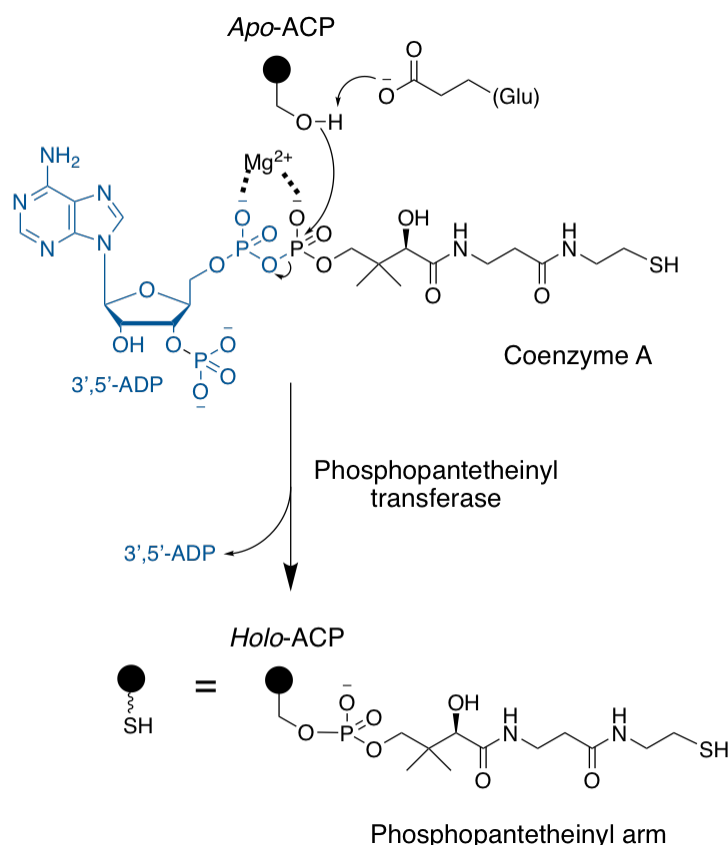


Figure 1.5: Mechanism of the conversion of *apo*-ACP domains to *holo*-ACP domains catalysed by 4'-phosphopantetheinyl transferase, the catalytic glutamic acid residue is indicated.

The role of the AT domain is to select for the appropriate extender (or starter) unit. Nearly all AT domains show high specificity for their cognate unit, with malonyl-CoA and (2*S*)-methylmalonyl-CoA being the most commonly utilised substrates, although many other substrates have been shown to be selected for by AT domains from PKSs (Figure 1.6)^{29,30}. The loading of the extender (or starter) unit onto the ACP domain proceeds via the formation of an acyl-AT domain intermediate, which is formed by attack of the corresponding acyl-CoA by a conserved serine residue in the active site of the AT domain (Figure 1.7)³¹. The acyl-AT domain intermediate is then subject to attack from the thiol moiety of the phosphopantetheinyl arm of the downstream ACP domain.

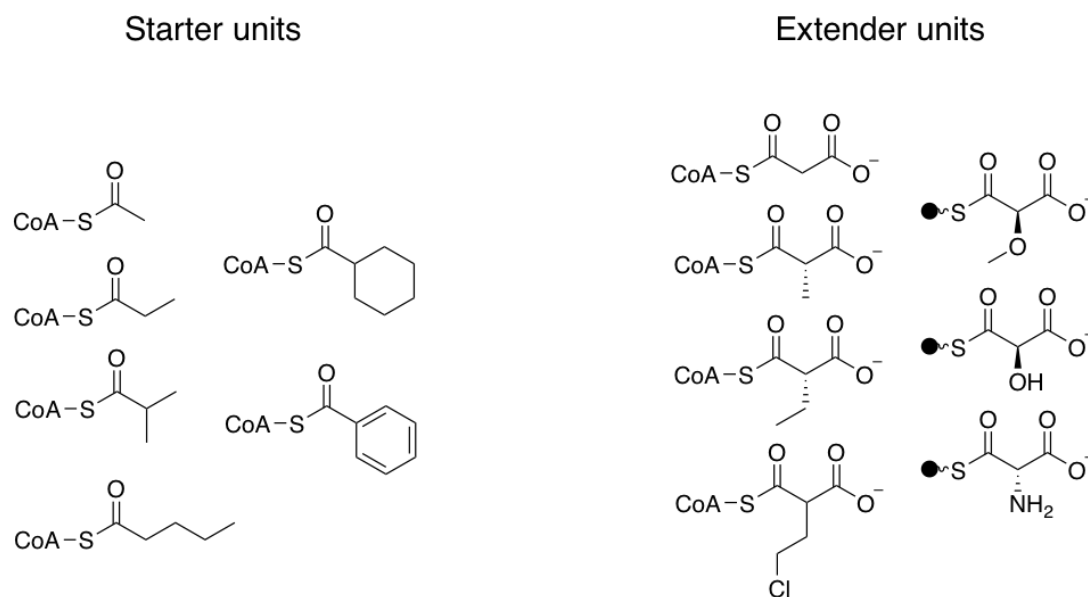


Figure 1.6: Examples of starter and extender units utilised in polyketide biosynthesis by *cis*-AT PKSs. ACP domains are represented by black circles, and where this is the case, it is proposed that the AT domain offloads the substrate from the ACP domain and onto an ACP domain within the PKS.

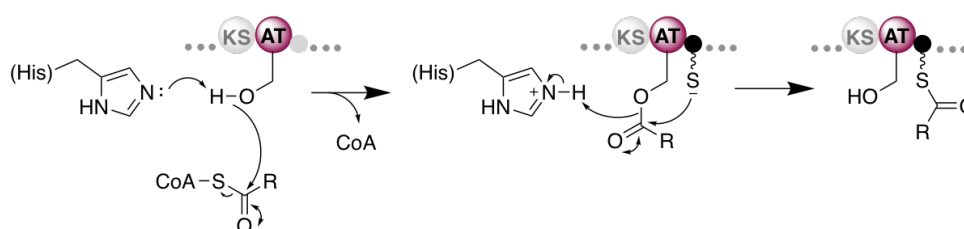


Figure 1.7: Mechanism of the loading of the extender unit (methylmalonyl-CoA for the 6-dEB PKS) or starter unit (propionyl-CoA for the 6-dEB PKS) onto the PKS catalysed by the AT domain, the catalytic histidine residue is indicated. R = starter or extender unit.

The KS domain has two roles, it firstly interacts with the upstream ACP domain and catalyses the intermodular translocation of the polyketide intermediate onto its active site cysteine residue. Once the AT domain has loaded the extender unit onto the downstream ACP domain, the KS domain then catalyses the decarboxylative Claisen-like condensation of the extender unit with the growing polyketide chain, resulting in the formation of a new C-C bond (Figure 1.8). It has been shown, by site-directed mutagenesis of the active site

of the KS domain from module 1 of the 6-dEB PKS, that two conserved histidine residues play an essential role in catalysing these two reactions³². The chain elongation cycle repeats for each module until the end of the PKS is reached.

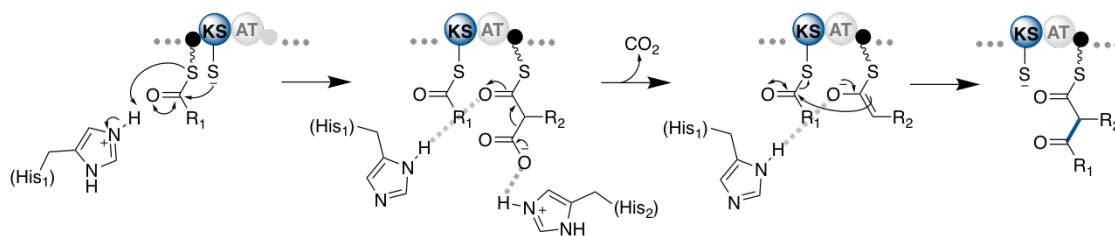
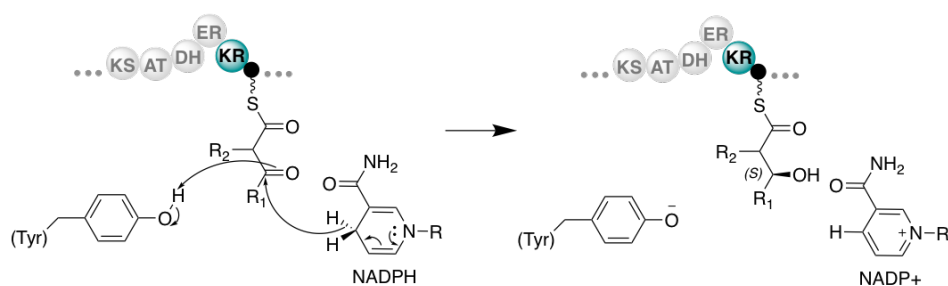


Figure 1.8: Mechanism of intermodular polyketide translocation and chain elongation catalysed by the KS domain. The catalytic histidine residues are indicated (with hydrogen bonds shown in grey) and the new C-C bond formed is shown in bold. R_1 = polyketide chain and R_2 = α -substituent derived from the extender unit.

1.2.1.2 Accessory domains

In addition to the key catalytic domains, each module may also possess a number of accessory domains that are responsible for introducing structural diversity to the polyketide. Ketoreductase (KR) domains act first, and catalyse the β -nicotinamide adenine dinucleotide 2'-phosphate (NADPH)-dependant reduction of the β -keto group, generated by the chain elongation reaction, to a hydroxyl group (Figure 1.9). This reaction proceeds in a stereospecific manner, with KR domains referred to as 'A-type' if they produce a (*S*)-configured hydroxyl group and 'B-type' if they produce a (*R*)-configured hydroxyl group, this can be predicted from the amino acid sequence of the domain^{33,34}. Interestingly, structural studies of A- and B-type KR domains revealed that the relative positions of the catalytic domains and NADPH are maintained in both of the structures, so it is proposed that stereoselectivity is imposed by the polyketide substrate entering the active site from different sides of the catalytic domain³⁵. It has also been demonstrated that some KR domains are capable of epimerising α -substituents, where present³⁶.

A-type KR domain



B-type KR domain

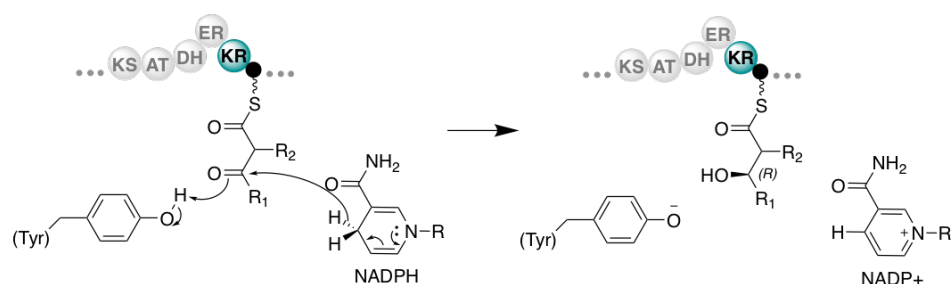


Figure 1.9: Mechanism of the reduction of the β -keto group catalysed by the KR domain. The catalytic tyrosine residue is indicated. R_1 = polyketide chain and R_2 = α -substituent derived from the extender unit. The reaction is stereospecific, with stereochemistry determined by the orientation of the substrate in the active site.

A dehydratase (DH) domain may then follow the KR domain, which generates an α / β -double bond through the *syn*-elimination of water (Figure 1.10)³⁷. The elimination reaction is sensitive to the stereochemical configuration of the β -hydroxyl group, thus DH domains most commonly precede B-type KR domains, generating the thermodynamically more stable *trans*-double bond³⁸.

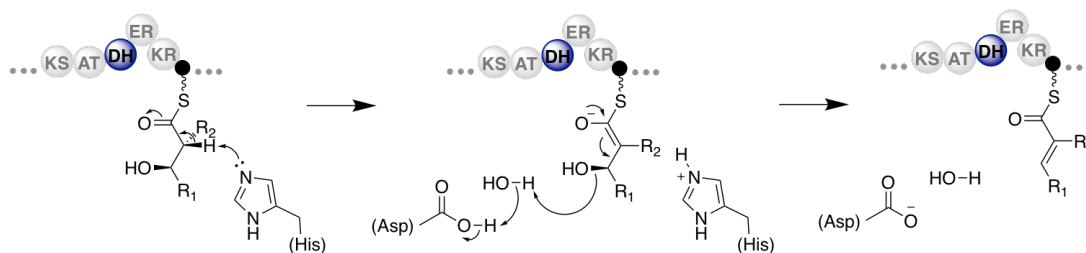


Figure 1.10: Mechanism of the reduction of the β -hydroxyl group catalysed by the DH domain. The catalytic histidine and aspartic acid residues are indicated. R_1 = polyketide chain and R_2 = α -substituent derived from the extender unit.

The double bond generated by the DH domain may then be further reduced by an enoylreductase (ER) domain to give a saturated moiety (Figure 1.11). A crystal structure of the KR-ER didomain from the spinosyn PKS revealed that the NADPH cofactor is positioned to first add a proton to the β -carbon. A second proton is then abstracted from either a tyrosine residue (giving an (*S*)-configured stereocentre at the α -position) or a lysine residue (which transfers a proton if the tyrosine moiety is absent, to give a (*R*)-stereocentre at the α -position), setting the stereochemistry at the α -position, where applicable³⁹.

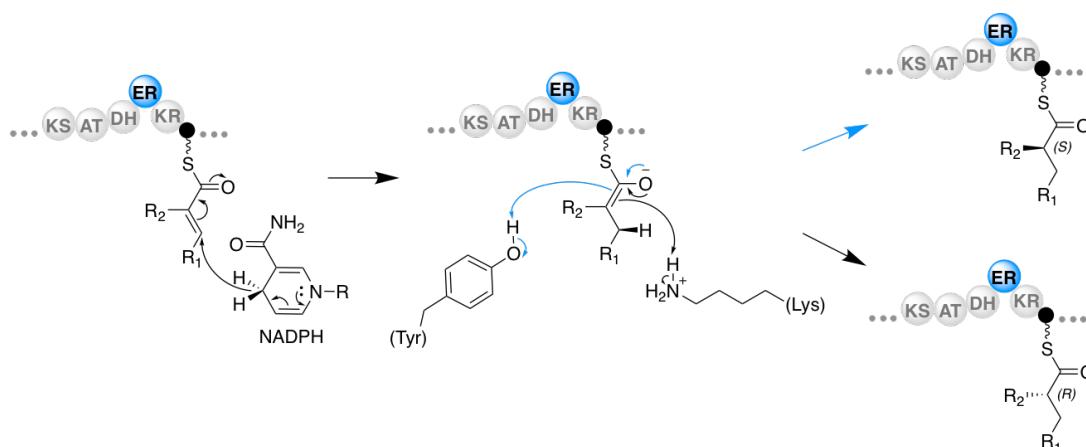


Figure 1.11: Mechanism of the reduction of the α / β -double bond catalysed by the ER domain. The catalytic tyrosine and lysine residues are indicated. R_1 = polyketide chain and R_2 = α -substituent derived from the extender unit.

1.2.1.3 The principal of co-linearity

The 6-dEB PKS demonstrates perfectly the principal of co-linearity, whereby the PKS domain architecture corresponds directly to the polyketide structure. The modules operate in the order they are positioned and each module is used only once. Co-linearity, which is observed for the majority of *cis*-AT PKSs, means that in general the structure of natural products from these systems can be predicted, with a good degree of confidence, from the sequence of the domains. This principal has been exploited for the bioengineering of numerous *cis*-AT PKS assembly lines³⁸. Targeted genetic manipulations, such as inactivation and substitution of the catalytic domains, have been used to create polyketide derivatives with structural modifications correlating directly to the introduced mutations, which in turn have provided extensive experimental evidence for this principal.

Naturally there are exceptions to the rule, and with the number of known PKS assembly lines rapidly expanding, increasing numbers of non-canonical *cis*-AT PKSs are being reported²⁹. Notable deviations from co-linearity in these systems include 'module stuttering' and 'module skipping'. 'Module stuttering' that was essential for biosynthesis was first reported for the aureothin PKS, where module 1 of the PKS was shown to act in an iterative fashion, and catalyse two rounds of chain elongation⁴⁰. This phenomenon has since been identified in numerous other systems. An interesting example of 'module skipping' can be found in the biosynthesis of quartromicin, which is composed of two polyketide chain intermediates, one shorter than the other. Both of these biosynthetic intermediates are built on the same PKS assembly line, which performs programmed module skipping half of the time to generate the shorter polyketide intermediate⁴¹.

1.2.1.4 Post-PKS modifications

Once the end of a polyketide assembly line has been reached the resultant polyketide is commonly released from the PKS *via* hydrolysis or macrocyclization to give a linear or cyclic product, respectively. Chain release is usually catalysed by a thioesterase (TE) domain, which first loads the polyketide onto an active-site serine residue before offloading occurs. Offloading by macrocyclization, like for the 6-dEB PKS, proceeds *via* a high degree of regio- and stereo-selectivity. Combinatorial biosynthesis of PKSs, however, has demonstrated that these domains can be promiscuous towards unnatural substrates, simply hydrolysing the polyketide chain if macrocyclization is no longer feasible⁴².

To introduce further structural diversity, polyketides often undergo a number of post-PKS modifications that are catalysed by standalone enzymes encoded for in the biosynthetic gene cluster (BGC). To obtain the active antibiotic erythromycin (**4**), biologically inactive 6-dEB (**9**) is subjected to glycosylation, with L-mycarose and D-desosamine, hydroxylation at two positions on the macrolide core and *O*-methylation of the L-mycarose moiety⁴³. An extensive array of post-PKS modifications have been observed in the study of these systems, that are catalysed by a diverse range of enzymes, which include; oxidoreductases (that introduce hydroxy, epoxide, aldehyde and keto groups), reductases, glycosyltransferases, methyltransferases (in particular *O*-methyltransferase), acyltransferases, aminotransferases, halogenases and cyclases. Unlike for the tailoring domains that are a part of the PKS, it is often not possible to predict accurately the action of the post-PKS enzymes. Elucidating the function and order in which these enzymes operate is often achieved through the creation of gene knockouts and characterisation of the resulting metabolite, thus the investigation of these enzymes has led to the creation of hundreds of novel polyketide derivatives^{44,45}.

1.2.2 *Trans*-AT PKSs

Recently, a second type of modular PKS was identified, that was found to have evolved independently from *cis*-AT PKSs⁴⁶. *Trans*-AT PKSs differ fundamentally from their *cis*-AT PKS counterpart by way of the activity of the AT domain, hence the chosen nomenclature. *Trans*-AT PKSs lack integrated AT domains and instead AT domain activity is provided *in trans* by a standalone AT domain, which acts iteratively to provide the extender unit to each of the modules (Figure 1.12).

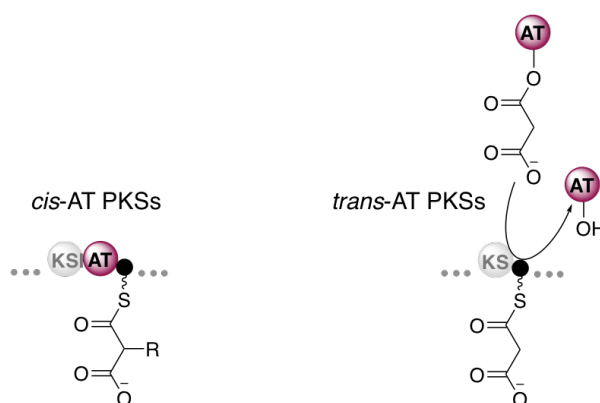


Figure 1.12: The module structure of *trans*-AT PKSs compared to *cis*-AT PKSs.

Trans-AT domains are usually malonyl-CoA specific. Only one *trans*-AT domain has been experimentally characterised that was shown to accept a non-malonyl substrate; an AT domain from the kirromycin PKS was found to utilise ethylmalonyl-CoA⁴⁷. It is proposed, however, that the oxazolomycin and etnangien pathways possess *trans*-AT domains that accept methoxymalonyl-CoA and succinyl-CoA, respectively^{48,49}. This is in huge contrast with the AT domains found in *cis*-AT PKSs, which have been shown to utilise a wide variety of starter and extender units (Figure 1.6). Many *trans*-AT PKS BGCs also encode for acyl hydrolase (AH) domains, that can commonly be found fused to the AT domain⁵⁰. The AH domain from the pederin biosynthetic pathway was shown to hydrolyse a number of different ACP bound acyl units, however it did not cleave malonyl-ACP⁵¹. It is therefore proposed that these

domains play a proofreading role, removing stalled intermediates, during biosynthesis.

Bacillaene (**10**), an unstable polyene diamine antibiotic, was isolated from *Bacillus subtilis* in 1995 and it was found to be biosynthesised by a *trans*-AT PKS, which was the first *trans*-AT PKS to be characterised (Figure 1.13)⁵². Due to the high density of non-canonical modules that often contain novel enzymatic domains in *trans*-AT PKSs, the principle of co-linearity can rarely be applied, and from the domain organisation alone it would not be possible to predict the structure of bacillaene (**10**). This PKS, however, had an interesting feature; stalled polyketide intermediates were hydrolytically released from the PKS. This gave an insight into the function of the many unusual characteristics of this PKS, that are considered to be hallmarks of *trans*-AT PKSs⁵³.

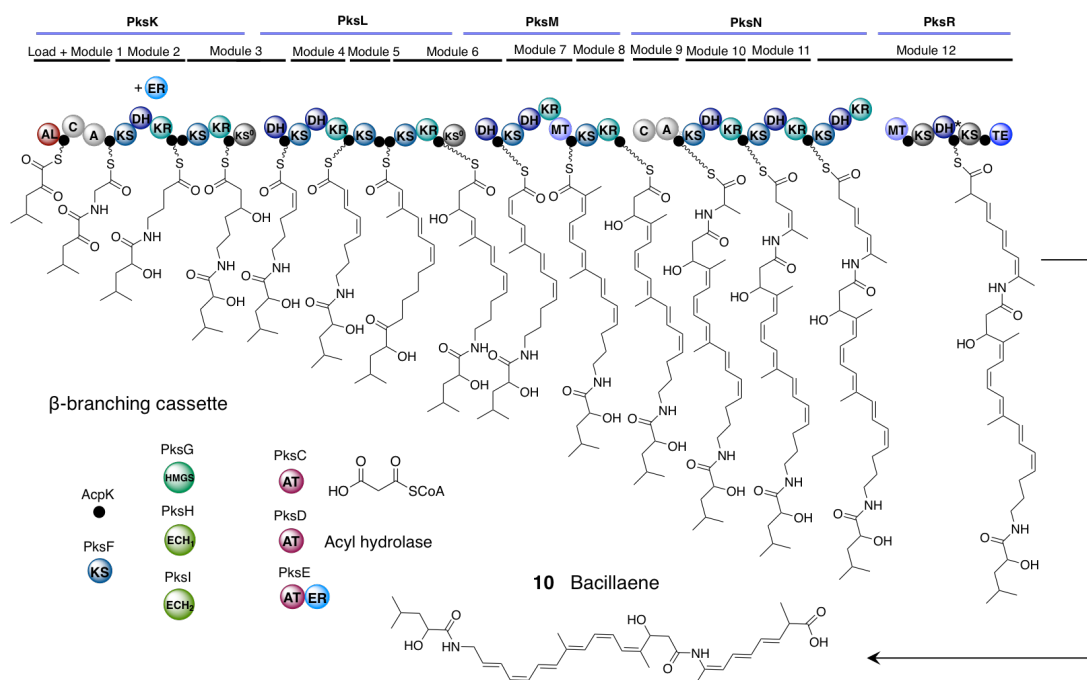


Figure 1.13: Pathway for bacillaene biosynthesis, showing the ACP-bound polyketide intermediates following α - and β -carbon processing. Black circles represent ACP domains. The β -branching cassette shown is responsible for introducing a β -methyl substituent to the growing polyketide chain in module 5. C = condensation domain and A = adenylation domain.

1.2.2.1 Chain initiation mechanisms

In *trans*-AT PKSs a diverse range of starter units are utilised, thus numerous different chain initiation mechanisms have been found. In the bacillaene biosynthetic pathway an acyl ligase (AL) domain is used to load the unusual α -hydroxy-isocaproic acid starter unit onto the first ACP domain of the PKS⁵⁴. AL domains can be found in a few *trans*-AT PKSs and they catalyse the adenosine triphosphate (ATP)-dependant thioesterification of an acid and an ACP domain⁵⁵. These proteins show similarity to aromatic-acid- and long-chain fatty-acid-CoA ligases⁵⁶.

The most commonly seen chain initiation mechanism in *trans*-AT PKSs utilises a GCN5-related *N*-acetyltransferase (GNAT) domain, which catalyses both the decarboxylation of malonyl-CoA to acetyl-CoA and the loading of the acetyl unit onto the first ACP domain of the PKS⁵⁷. These domains can occasionally be observed in *cis*-AT PKSs, where they were first characterised⁵⁷. Starter unit loading in *cis*-AT PKSs is almost exclusively catalysed by an AT domain. The AT domain either directly selects for a starter unit, or loads a typical extender unit, which is then decarboxylated by a KS^Q domain. KS^Q domains possess a mutation of the active site cysteine residue, to either glutamine or serine, meaning they are not capable of catalysing translocation, so act purely as decarboxylase⁵⁸. A comparison of the most commonly used chain initiation mechanisms is shown in Figure 1.14.

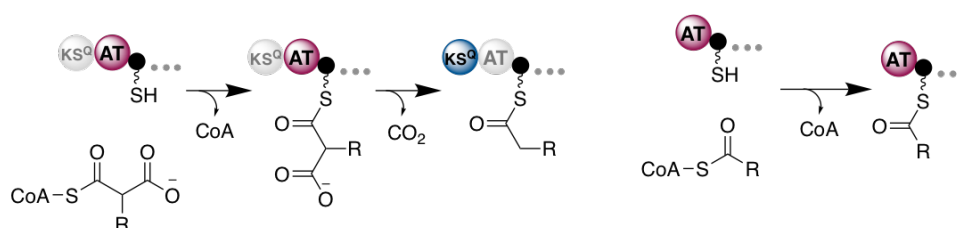
Trans-AT PKSs*Cis*-AT PKSs

Figure 1.14: The most commonly utilised chain initiation mechanisms by *trans*- and *cis*-AT PKSs.

1.2.2.2 Non-elongating KS domains

Four non-elongating KS domains (KS⁰), which are ubiquitous in *trans*-AT PKSs, can be found in the bacillaene PKS. KS⁰ domains simply translocate the growing polyketide from one ACP domain to the next. These KS domains possess a mutation of the conserved histidine residue, from the HGTGT motif, which is required for decarboxylative condensation⁴⁶.

1.2.2.3 Unusual domain architectures

A huge number of module variants have been observed in *trans*-AT PKSs⁵⁰. In the bacillaene biosynthetic pathway this is exemplified by the four different domain architectures used to install double bonds into the growing polyketide chain. The KS-DH-KR-ACP architecture is found in modules 4, 10 and 11 and incorporate double bonds into the growing polyketide chain, however, interestingly in modules 10 and 11 the α / β-double bond generated is migrated, giving a β / γ-double bond. It is proposed that the DH domains in

these modules are likely responsible for the observed double bond topology⁵⁰. A KS-DH-KR-MT-ACP module can also be found in the PKS, which generates a *trans*-double bond with an α -methyl branch. Two modules, modules 3 and 6, are split across two PKS proteins; split modules are regularly observed in *trans*-AT PKSs. Both of these modules have the domain architecture KS-KR-ACP-KS⁰ DH-ACP, which appear to divide reduction and dehydration into discrete steps. Interestingly, it seems that both of these modules install unusual *cis*-double bonds, although dehydration modules with this domain architecture have yet to be biochemically characterised^{50,53}. The final module in the bacillaene PKS has the domain architecture KS-DH-KR-MT-ACP-KS⁰-DH*-KS⁰-TE which also incorporates a β / γ -double bond, but to achieve this the unusual enoyl-isomerase (DH*) domain is utilised⁵⁹.

1.2.2.4 Recruitment of unusual domains

Many of the different domain architectures observed in *trans*-AT PKSs can be attributed to the inclusion of unusual domains, such as the DH* domain. The DH* domain is structurally related to DH domains, however the catalytic histidine is positioned differently, allowing the histidine residue to shuttle a proton from the γ -position to the α -position, *via* the mechanism shown in Figure 1.15⁶⁰. The bacillaene PKS also recruits non-ribosomal peptide synthase (NRPS) domains to integrate the amino acids, glycine and alanine, into the growing polyketide chain. Like PKSs, NRPSs are large modular enzymes, however, they catalyse the condensation and modification of amino acid building blocks⁶¹. The condensation (C) and adenylation (A) domains found in the bacillaene PKS act homologously to KS and AT domains, respectively.

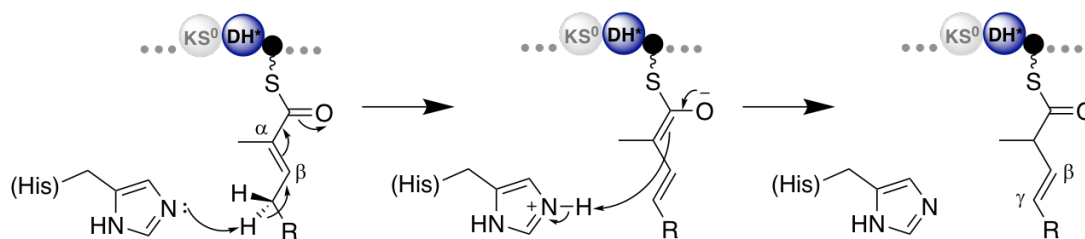


Figure 1.15: Mechanism of double bond migration catalysed by the DH* domain. The catalytic histidine residue is indicated. R = polyketide chain.

Trans-AT PKSs have also been found to utilise many other embedded processing domains during polyketide assembly, which are rarely observed in *cis*-AT PKSs. These include; pyran synthase domains that catalyse the formation of five- or six-membered oxygen containing rings, branching domains that are involved in the formation of glutarimide, lactam and δ -lactone moieties, oxygenase domains and *O*-methyltransferase domains⁵⁰.

1.2.2.5 Introduction of methyl branches during chain assembly

In *trans*-AT PKSs, methyl substituents are commonly incorporated into the polyketide chain during PKS assembly by either a dedicated *C*-methyltransferase (CMT) domain or by a β -branching cassette that acts *in trans*⁵⁰. Both of these mechanisms for methyl branch incorporation can be found in *cis*-AT PKSs, from where they have been characterised, however, they are found much less frequently owing to the use of methylmalonyl-CoA as an extender unit in these PKSs²⁹.

Embedded CMT domains, which can be found in modules 7 and 12 of the bacillaene PKS, catalyse the transfer of a methyl group from *S*-adenosyl methionine (SAM) to the α -position of the β -keto polyketide intermediate that is generated by chain elongation (figure 1.16)⁶². Biochemical studies on the CMT domain from the curacin A biosynthetic pathway (which was the first example of an embedded MT domain to be characterised) revealed that the CMT domain was not able to act upon a β -hydroxyl substrate, indicating that

α -methylation occurs prior to reductive modification of the polyketide intermediate⁶². It has also, recently, been shown that this reaction proceeds stereo-specifically, producing a (2*R*)-2-methyl-3-keto polyketide intermediate⁶³.

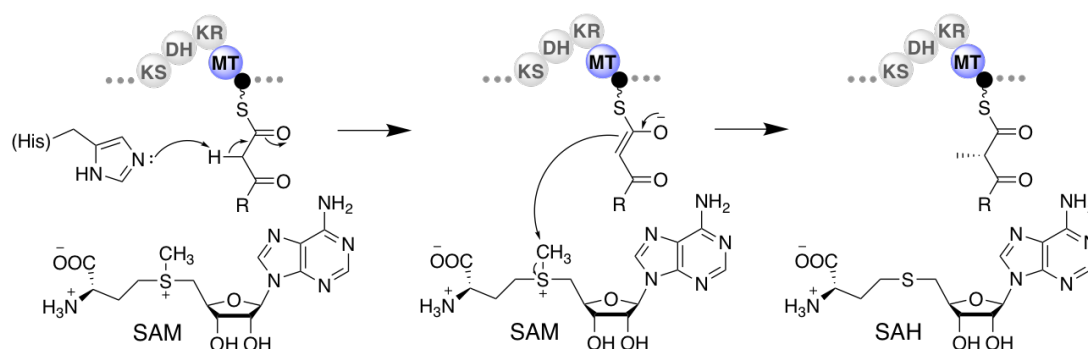


Figure 1.16: Mechanism of *C*-methylation catalysed by embedded *CMT* domains. The catalytic histidine residue is indicated. R = polyketide chain.

Incorporation of a β -methyl branch in the growing polyketide chain requires a set of enzymes that act *in-trans*, which is usually encoded for in the BGC. The set of enzymes are commonly referred to as the β -branching cassette and usually consist of a standalone ACP and KS domain, a 3-hydroxy-3-methyl glutaryl-CoA synthase (HMGS) like enzyme and enoyl-CoA hydratase (ECH) enzymes. To convert the β -keto moiety into a carbon branch, malonyl-CoA is first loaded onto the standalone ACP domain where it is decarboxylated, by the KS domain, to give an acetyl unit. The HMGS enzyme then catalyses the aldol addition of the acetyl unit to the growing polyketide chain, before the aldol adduct is dehydrated and decarboxylated by the ECH enzymes to give a β -methyl branch (Figure 1.17)⁵⁰. The location on the PKS where β -branching occurs can be predicted by the presence of a module that consists of a KS domain and two or more ACP domains, which is module 5 in the bacillaene PKS. It has been established that multiple ACP domains are utilised in PKSs to help overcome rate-limiting steps in the biosynthesis⁶⁴.

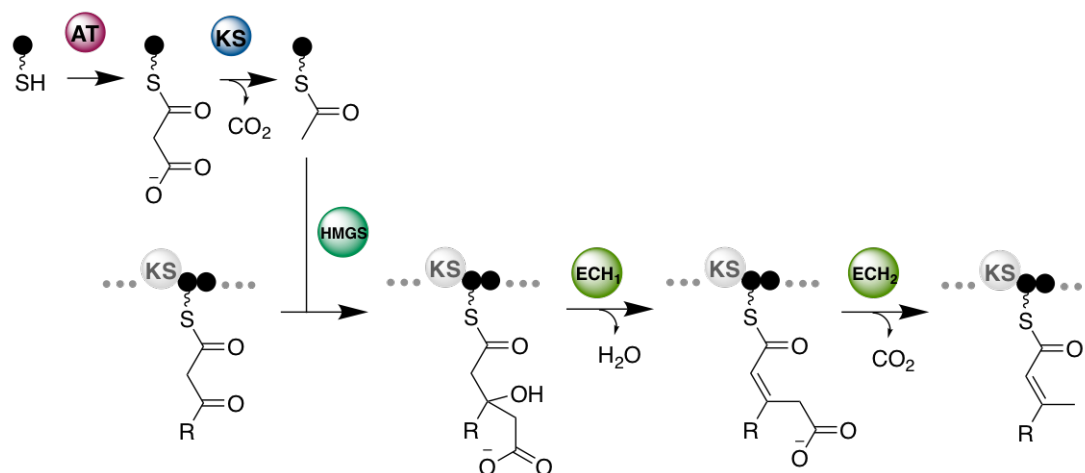


Figure 1.17: General mechanism for the formation of β -branches in PKSs. The role of each enzyme in the β -branching cassette is indicated. R = polyketide chain.

1.3 Gladiolin

Gladiolin (**11**) is a novel macrolide antibiotic that was recently isolated from *Burkholderia gladioli* BCC0238 and is of particular interest as it demonstrates potent antimicrobial activity against *Mycobacterium tuberculosis*, a human pathogen that has received substantial attention in recent years due to the rapidly increasing prevalence of extensively drug resistant strains. It was found to have a minimum inhibitory concentration (MIC) of 0.4 $\mu\text{g/mL}$ against *M. tuberculosis* H37Rv and was also discovered to be active against four *M. tuberculosis* clinical isolates that were resistant to isoniazid and one that was resistant to both isoniazid and rifampicin (**5**)⁶⁵. Like rifampicin (**5**), gladiolin (**11**) was found to inhibit bacterial RNA polymerase, however the precise mode of action has yet to be established⁶⁵.

Gladiolin (**11**) is structurally similar to etnangien (**12**), which is an antibiotic that also demonstrates potent antimicrobial activity by efficiently inhibiting RNA polymerases (Figure 1.18). It was isolated from *Sorangium cellulosum* SO ce750 in 2007⁶⁶. Etnangien (**12**) is a highly unstable antibiotic due to its

hexene moiety which is prone to light induced isomerisation and oxidative degradation⁴⁸. Attempts were made to identify the pharmacore of etnangien (**12**); a series of analogues were synthesised that included the macrolide core with truncated side chains and a series of side chains without the macrolide moiety, however, none of these analogues demonstrated antimicrobial activity^{67,68}. It was, thus, concluded that both the macrolide core and side chain were essential for its pharmacological activity so development of etnangien (**12**) for clinical use was not pursued. The discovery of gladiolin (**11**), however, strongly suggests that modifications can be made to the side chain of these antibiotics without detriment to bioactivity, making gladiolin an attractive compound for further biochemical investigation.

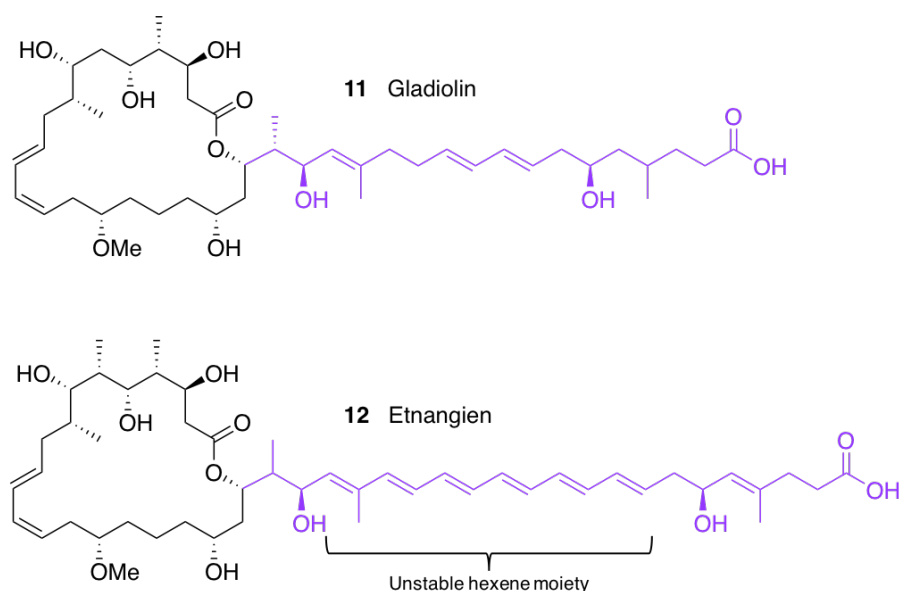


Figure 1.18: The structure of gladiolin and etnangien. The differing side chains are highlighted in purple.

The gladiolin BGC was identified using a combination of single-molecule real-time based genome sequencing and targeted mutagenesis (Figure 1.19). The proposed function of each of the non-PKS proteins encoded for by the BGC are listed in Table 1.1⁶⁵. It was established that gladiolin is biosynthesised by a *trans*-AT PKS, consisting of six PKS proteins, which exhibits many of the functional peculiarities associated with this class of PKS; the proposed

biosynthesis is shown in Figure 1.20⁶⁵. Firstly, it is proposed that the gladiolin PKS contains three non-elongating KS domains, as there are twenty KS domains, while only seventeen chain elongations are required to construct gladiolin. Interestingly, none of the KS domains found in the gladiolin PKS are true KS⁰ domains, as they all possess the histidine residue that is required for catalysing chain elongation, which is normally mutated. Based on the structure of gladiolin (**12**) and the PKS domain architecture, it was tentatively hypothesised that KS domains 12, 14 and 20 are non-elongating⁶⁵. When comparing the structure of gladiolin to the domain architecture it is also apparent that modules 5 and 10 of the PKS lack the expected ER domains, however it has recently been shown that a standalone ER domain, GbnE, is responsible for catalysing enoyl-reduction in both of these modules in the gladiolin PKS⁶⁹. Furthermore, module 11 of the PKS contains an unusual embedded *O*-methyltransferase (OMT) domain. Embedded OMT domains are an uncommon feature of *trans*-AT PKSs, which have yet to be characterised⁵⁰.

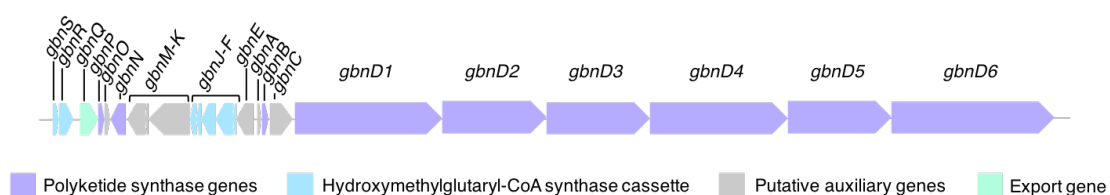


Figure 1.19: The organisation of the gladiolin biosynthetic gene cluster in *B. gladioli* BCC0238.

Table 1.1: The putative functions of the non-PKS-encoding genes in the gladiolin BGC.

Gene/Protein	Length bp/aa	Similar proteins (% Identity)
<i>gbnA</i> /GbnA	264/88	Hypothetical protein: <i>Sorangium cellulosum</i> (60 %) Acyl carrier protein: <i>Streptomyces rimosus</i> (49 %)
<i>gbnB</i> /GbnB	894/298	Malonyl-CoA-ACP transacylase: <i>Sorangium cellulosum</i> (54 %) Hypothetical protein: <i>Desulfovibrio inopinatus</i> (46 %)
<i>gbnC</i> /GbnC	1968/656	Asparagine synthetase: <i>Sorangium cellulosum</i> (65 %) Asparagine synthetase: <i>Labrenzia alba</i> (51 %)
<i>gbnE</i> /GbnE	1389/463	Permease: <i>Sorangium cellulosum</i> (68 %) ER domain BatK: <i>Pseudomonas fluorescens</i> (60 %)
<i>gbnF</i> /GbnF	246/82	Acyl carrier protein: <i>Sorangium cellulosum</i> (68 %) Hypothetical protein: <i>Paenibacillus pinihi</i> (59 %)
<i>gbnG</i> /GbnG	1212/404	Polyketide beta-ketoacyl-ACP synthase: <i>Sorangium cellulosum</i> (65 %) Polyketide beta-ketoacyl-ACP synthase: <i>Pelonsinus fermentans</i> (57 %)
<i>gbnH</i> /GbnH	1260/420	3-hydroxy-3-methylglutaryl-ACP synthase: <i>Sorangium cellulosum</i> (99 %) 3-hydroxy-3-methylglutaryl-ACP synthase: <i>Sorangium cellulosum</i> (76 %)
<i>gbnI</i> /GbnI	789/263	Enoyl-CoA hydratase/isomerase: <i>Sorangium cellulosum</i> So ce56 (63 %) Enoyl-CoA hydratase: <i>Paenibacillus tyrfis</i> (57 %)
<i>gbnJ</i> /GbnJ	747/249	Enoyl-CoA hydratase: <i>Sorangium cellulosum</i> (99 %)

Enoyl-CoA hydratase BatE: *Pseudomonas fluorescens* (60 %)

<i>gbnK</i> /GbnK	3234/1078	Hypothetical protein SorM: <i>Sorangium cellulosum</i> (99 %) Hypothetical protein: <i>Sorangium cellulosum</i> (99 %)
<i>gbnL</i> /GbnL	360/120	Alpha/beta hydrolase: <i>Burkholderia glumae</i> (74 %) Alpha/beta hydrolase: <i>Sorangium cellulosum</i> (58 %)
<i>GbnM</i> /GbnM	1401/467	Amidase: <i>Sorangium cellulosum</i> (56 %) Amidase SorP: <i>Sorangium cellulosum</i> (46 %)
<i>gbnN</i> /GbnN	1185/395	Malonyl-CoA-ACP transacylase: <i>Methylococcaceae bacterium</i> (53 %) Malonyl-CoA-ACP transacylase: <i>Scytonema sp. PCC 10023</i> (48 %)
<i>gbnO</i> /GbnO	849/283	4'-phosphopantetheinyl transferase: <i>Collimonas arenae</i> (47 %) Hypothetical protein: <i>Janthinobacterium sp. CG3</i> (44 %)
<i>gbnP</i> /GbnP	975/325	Malonyl-CoA-ACP transacylase: <i>Burkholderia thailandensis</i> (53 %) Malonyl-CoA-ACP transacylase: <i>Xanthomonas cannabidis</i> (50 %)
<i>gbnQ</i> /GbnQ	1353/451	Multidrug transporter MatE: <i>Chromobacterium subtsugae</i> (72 %) Multidrug transporter MatE: <i>Sorangium cellulosum</i> (68 %)
<i>gbnR</i> /GbnR	792/264	Enoyl-CoA hydratase: <i>Burkholderia glumae</i> (89 %) Enoyl-CoA hydratase: <i>Burkholderia plantarii</i> (86 %)
<i>gbnS</i> /GbnS	1206/402	Enoyl-CoA hydratase: <i>Burkholderia glumae</i> (83 %) Enoyl-CoA hydratase: <i>Burkholderia pseudomallei</i> (73 %)

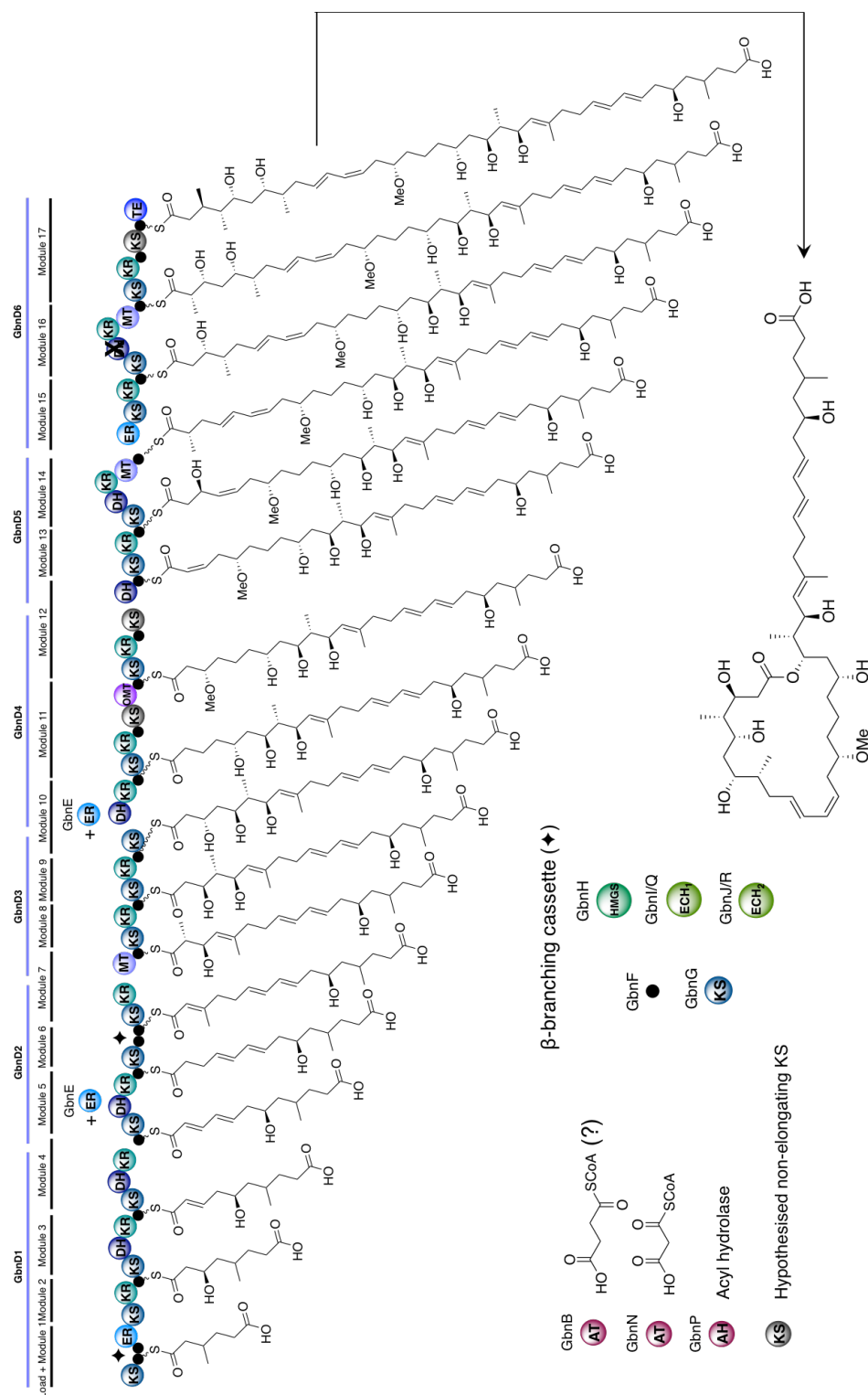


Figure 1.20: Proposed pathway for gladiolin biosynthesis, showing the ACP-bound predicted polyketide intermediates following α - and β -carbon processing. Black circles represent ACP domains. The domain with a cross through it indicates that the domain lacks the key catalytic residues required for activity. The β -branching cassette shown, is responsible for introducing β -methyl substituents to the growing polyketide chain in modules 1 and 6.

Phylogenetic analysis of the AT domains encoded for by the gladiolin BGC indicated that GbnN was an AT domain with specificity for the extender unit malonyl-CoA, whereas the AT domain GbnB claded separately from the malonyl-specific AT domains. It was therefore proposed that this unusual AT domain may show specificity for the hypothesised starter unit, succinyl-CoA⁶⁵. Genes for a HMGS cassette were also identified in the gladiolin BGC, with two sets of ECH enzymes encoded for, which are responsible for the incorporation of β -methyl branches in modules 1 and 6, via the mechanism described in section 1.2.2.5. Additionally a number of putative auxiliary genes were identified in the gladiolin BGC. The role of two genes in particular, *gbc* which encodes for a putative asparagine synthetase and *gbnM* which encodes for a putative amidase, is esoteric owing to the fact that gladiolin contains no nitrogen atoms.

The aim of this study is to investigate some of the unusual features of gladiolin biosynthesis, namely the *O*-methylation module with the dedicated OMT domain, found in module 11 of the PKS, and the role of the cryptic auxiliary genes, *gbc* and *gbnM*.

Chapter 2

O-methylation during chain assembly by *trans*-AT modular polyketide synthases

This work is in collaboration with Christian Hobson, University of Warwick, who synthesised the pantetheine substrates used throughout this chapter.

2.1 Introduction

Trans-AT PKSs have been noted to contain O-methylation modules that have a dedicated OMT domain, however these modules remain one of the peculiarities of *trans*-AT systems that have yet to be characterised⁵⁰. The example found in the gladiolin biosynthetic pathway is shown in Figure 2.1. It is proposed that these modules consist of two parts, with the first part catalysing chain elongation and reduction of the diketo-intermediate to generate a β -hydroxyl moiety, so that methylation can then occur on the second non-elongating segment.

O-methylation is commonly found as a post-PKS tailoring step, with the location of methylation dictated by the structure of the resultant polyketide. These enzymes often show narrow substrate specificity and there are few reported examples where the substrate promiscuity has been successfully expanded by rational engineering^{70,71}. Embedded OMT domains could provide

an alternative approach to engineering methylation, where the position of methylation would be dependent upon the location of the domain within the PKS.

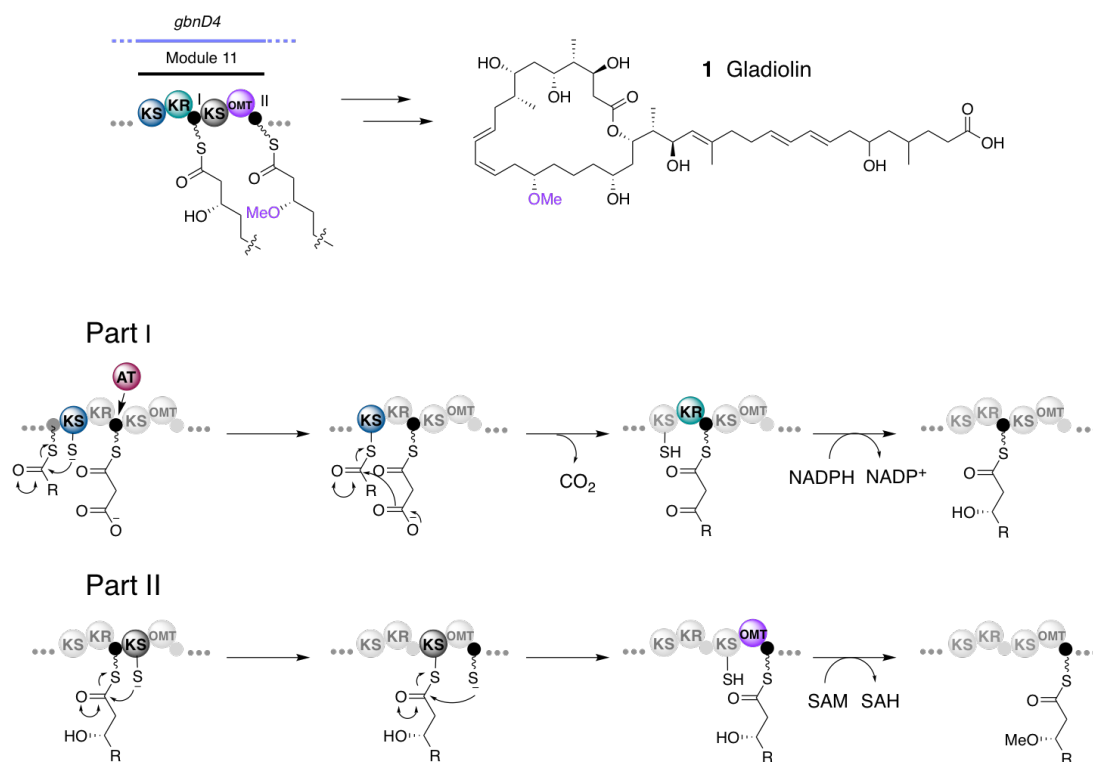


Figure 2.1: A closer look at the *O*-methylation module from the gladiolin biosynthetic pathway. The proposed sequence of enzymatic reactions is shown, with methylation hypothesised to occur on the second non-elongating part of the module. The domains involved in each step are highlighted, with the hypothesised non-elongating KS domain shown in grey. R = polyketide chain.

The aim of this study is to gain full insight into *O*-methylation modules for the prospective of bioengineering, which requires knowledge of the substrate tolerance and structure of the OMT domain as well as an understanding of the role of the hypothesised non-elongating KS domain.

2.2 *In-vitro* reconstitution of embedded OMT domain activity

2.2.1 Purification and characterisation of GldM11 OMT-ACP

In order to study the activity of the OMT domain from the gladiolin biosynthetic pathway a synthetic pET24a construct of the OMT-ACP didomain was designed based upon the domain boundaries determined by antiSMASH 2.0, a web-based tool that is able to analyse PKS domain architecture⁷². This was used to heterologously overproduce GldM11 OMT-ACP as an octa-histidine tagged protein, that was purified by nickel affinity chromatography. An octa-histidine tag was chosen as it has been shown to result in increased protein purity, when compared with the more commonly used hexa-histidine tag, as more stringent wash conditions can be used due to an increased binding strength to the nickel column⁷³. SDS-PAGE analysis showed the protein was pure and of the expected molecular weight (Figure 2.2 (A)), which was confirmed by intact protein mass spectrometry (MS) (Appendix 1). Size exclusion chromatography showed that the protein was monomeric in solution (Figure 2.2 (B)).

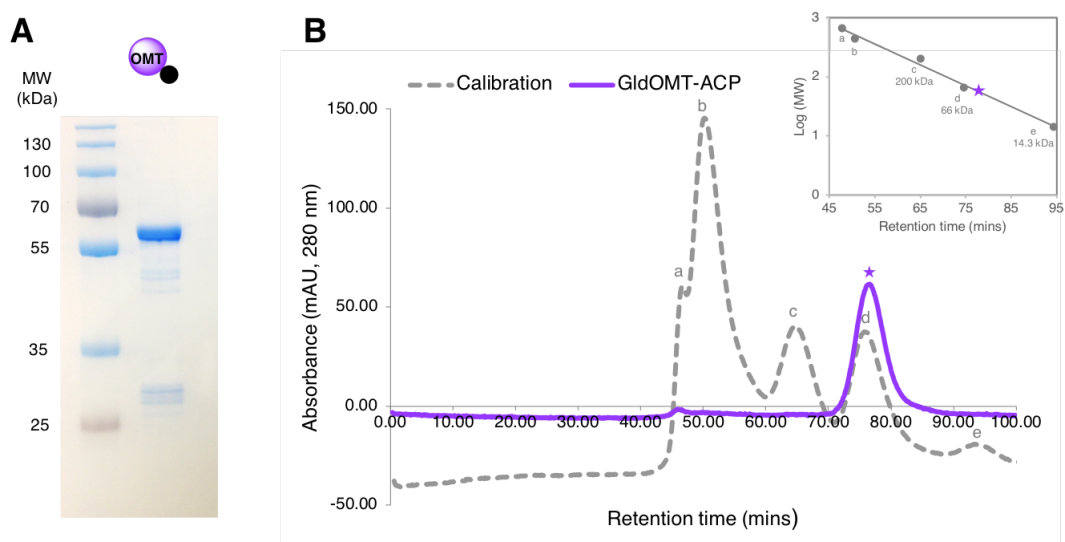


Figure 2.2: (A) 10 % SDS-PAGE gel showing purified GldM11 OMT-ACP. (B) A graph showing the results of size exclusion chromatography for GldM11 OMT-ACP.

2.2.2 Methylation assay design

A one-pot methylation assay was designed and optimised to demonstrate that the excised didomain was catalytically active *in vitro* (Figure 2.3 (A)). Simplified pantetheine analogues of the predicted substrate for the OMT domain were chemo-enzymatically converted to the associated CoA derivatives (Figure 2.3 (B))⁷⁴. This method utilised enzymes from the *E. coli* CoA biosynthetic pathway; first pantothenate kinase (PankK) was used to phosphorylate the substrate mimic so that phosphopantetheine adenylyltransferase (PPAT) could then catalyse the adenylylation to generate dephospho-CoA analogues. A final phosphorylation was carried out by dephosphocoenzyme A kinase (DPCK) producing CoA substrates, that the 4'-phosphopantetheinyl transferase (Sfp) could load onto the conserved serine residue of the ACP domain (as shown in Figure 1.4). 4'-phosphopantetheinyl transferases were found to demonstrate promiscuity towards their CoA substrate and the ACP domain being modified, making them a useful tool for the biochemical studies of PKSs⁷⁵.

The *holo*-protein was incubated with SAM, the biologically abundant cofactor that is essential for methylation, for 1 hour (h) 30 minutes (min). Longer incubation times resulted in minimal increase in observed methylation and also significantly reduced the quality of the MS data due to degradation of the protein. High-resolution intact protein MS, a tool which has been extensively used in the study of PKSs, was chosen as a means of monitoring the outcome of the methylation assays, as this technique can accurately detect biosynthetic transformations that involve small mass changes, from minimal amounts of protein⁷⁶.

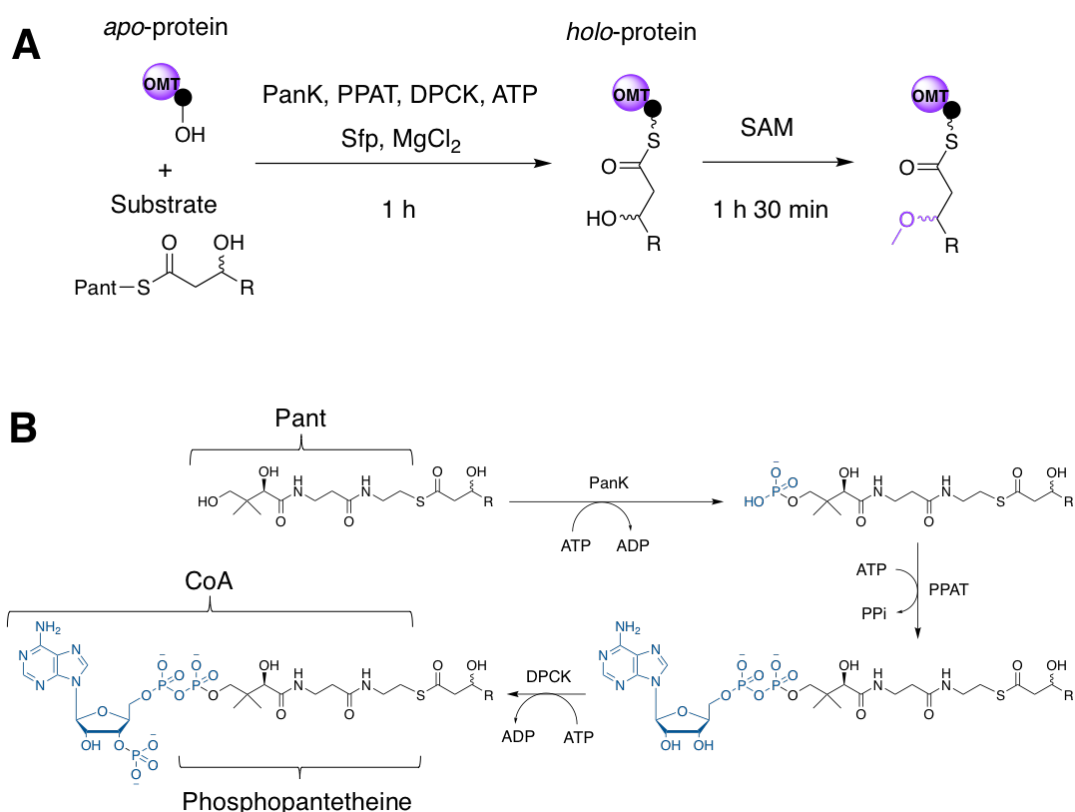


Figure 2.3: (A) Overview of the methylation assays conducted. (B) The enzymatic reactions that were used to convert the pantetheine substrates to CoA analogues.

2.2.3 Confirming the *in-vitro* activity of GldM11 OMT-ACP

Initially, the very simple substrate mimic 3-hydroxybutyryl-ACP was used to assess whether the excised didomain was catalytically active *in vitro*, however it was found that this was not a substrate for the OMT domain (Figure 2.4 (A)). It was then decided to see if the longer-chain substrate mimic 3-hydroxyoctanoyl-ACP would be accepted. Selective methylation of the *S* stereo-configured β -hydroxyl group was observed, which was expected given that sequence analysis of the KR domain from module 11 indicated that this would be the stereochemistry of the hydroxyl group generated *in vivo*⁶⁵ (Figure 2.4 (A)). Methylation was determined by the presence of a peak in the deconvoluted MS spectrum corresponding to a mass increase of 14 Da that was not seen in the absence of SAM (Figure 2.4 (B)). A peak could also be

observed in both the reaction and control experiments corresponding to the loss of water (-18 Da); loss of water is commonly seen during the MS of secondary alcohols. The MS of *holo*-ACP proteins leads to the ejection of the phosphopantetheine arm from the protein preserving the thioester bond to the substrate⁷⁷. The identification of these fragment ions further confirmed that methylation had occurred (Figure 2.4 (C)).

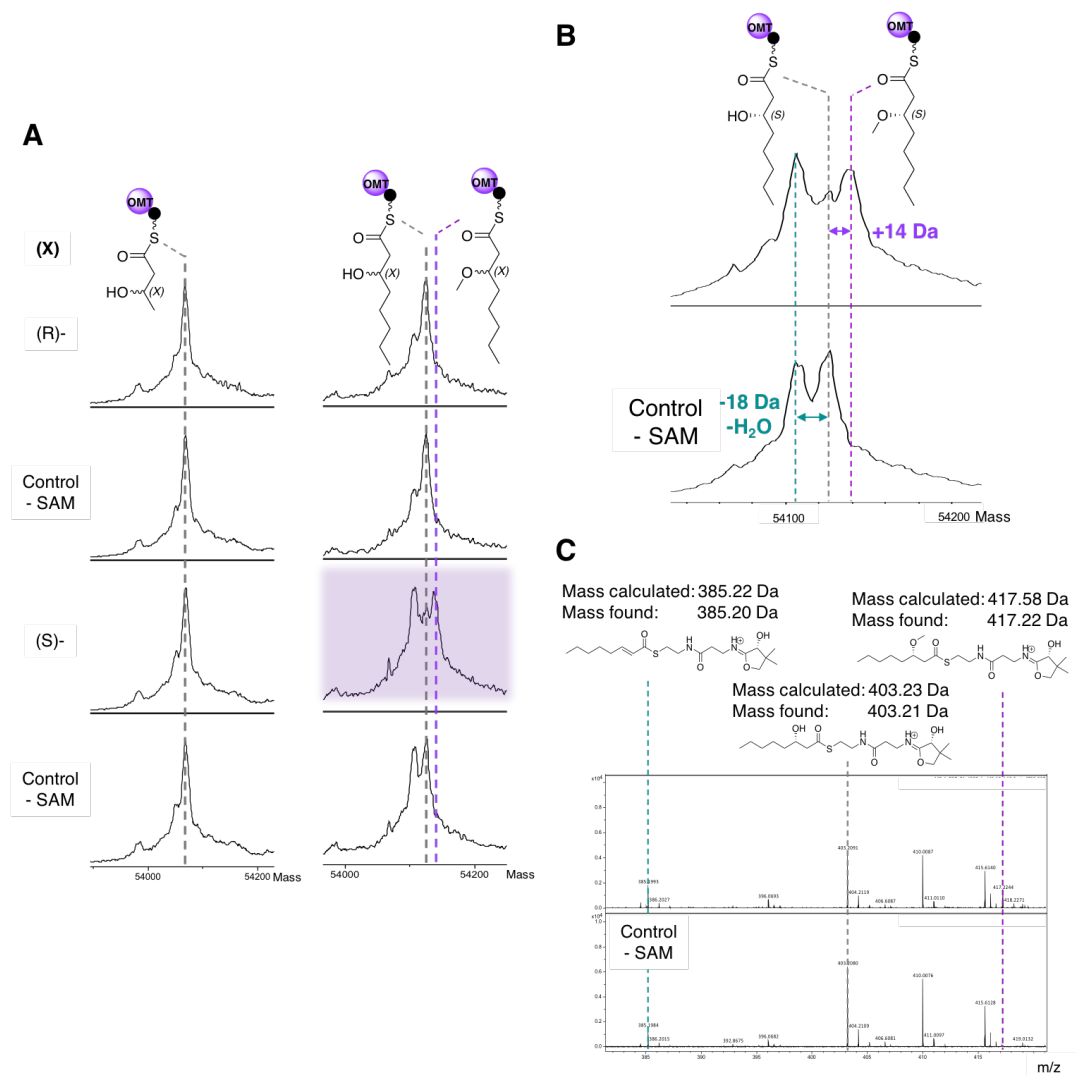


Figure 2.4: (A) Stacked deconvoluted ESI-MS spectra of GldM11 OMT-ACP with ACP bound simplified substrate mimics after incubation with SAM. The spectrum where methylation (+14 Da) was observed is highlighted in purple. (B) A closer look at the deconvoluted spectrum for the (S)-3-hydroxyoctanoyl substrate mimic. (C) Spectrum showing the key phosphopantetheine ejection ions for the (S)-3-hydroxyoctanoyl substrate mimic.

In order to investigate whether the OMT domain showed regioselectivity towards hydroxyl groups at the β -position, the methylation assay was conducted using 3,5-dihydroxyhexanoyl substrates. Only mono-methylation was observed for the (3*S*,5*S*)-3,5-dihydroxyhexanoyl-ACP substrate (Figure 2.5 (A)). The fact that only mono-methylation was observed suggested that the OMT domain was not capable of carrying out methylation of a δ -hydroxyl group, and did indeed shows regioselectivity for β -hydroxyl groups. Unfortunately, data could not be obtained for the (3*S*,5*R*)-3,5-dihydroxyhexanoyl- and (3*R*,5*S*)-3,5-dihydroxyhexanoyl-ACP substrates due to lactonization occurring, which resulted in the loss of the substrate bound to the ACP domain. Lactonization is more rapid for the anti-diols as the reaction proceeds *via* a more stable transition state (Figure 2.5 (B)).

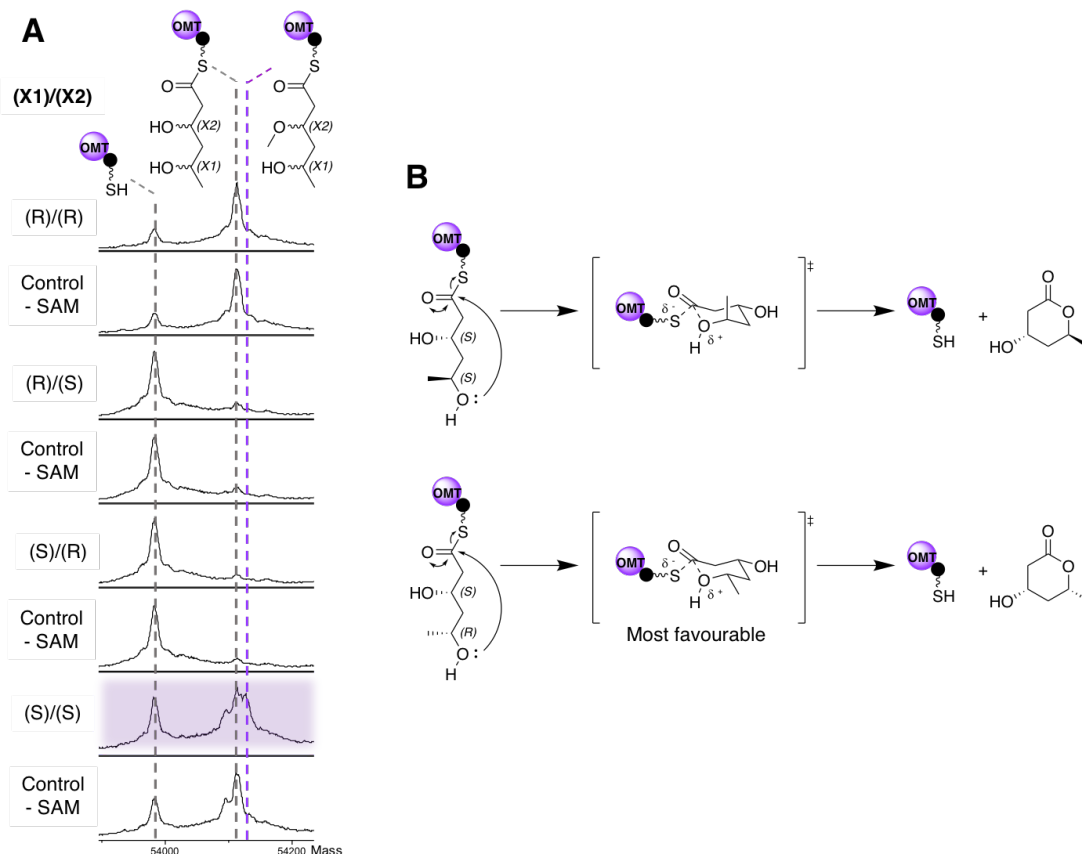


Figure 2.5: (A) Stacked deconvoluted ESI-MS spectra of GldM11 OMT-ACP with the 3,5-dihydroxyhexanoyl ACP bound substrate mimics after incubation with SAM. Highlighted in purple is the only spectrum where methylation (+14 Da) was observed. (B) Mechanism of lactonization of the 3,5-dihydroxyhexanoyl substrates, the transition states are shown. Lactonization is faster for the anti-diols as the reaction proceeds *via* a more favourable transition state that has both substituents in the equatorial position.

2.3 Identification of key catalytic residues

2.3.1 Sequence alignment

ClusterTools, a bioinformatics tool that enables facile searching of genomes for functional elements in close proximity to each other⁷⁸, was used to find additional examples of embedded OMT domains from *trans*-AT PKSs. A database constructed of BGCs from the MiBIG repository⁷⁹ was searched

using a custom HMM rule looking for genes that contained an OMT domain, a PKS KS domain and not a PKS AT domain. It was found that when using the default parameters, OMT domains were not able to be distinguished clearly from CMT domains, so the HMM search E-value was made substantially smaller, so that the returned matches would be more significant. Having been able to identify embedded OMT domains in known BGCs using clusterTools, a search was then ran, using the same settings, on a database containing all reference and representative bacterial genomes from NCBI. This did reveal some cryptic clusters that appear to contain embedded OMT domains, however, in all cases the sequence data was too poor to be able to make any predictions about the products of these clusters. In total 9 additional, characterised *trans*-AT PKSs were found to contain embedded OMT domains, with some pathways utilising more than one embedded OMT domain; the secondary metabolites for the identified pathways are shown in Figure 2.6^{48,80–85}.

In order to identify any conserved residues that may play a key role in catalysis, a sequence alignment was created (Appendix 2). The sequence alignment, along with comparison to the patellazole and luminaolide OMT domains, that sit respectively at the beginning and end of a PKS gene, improved the approximation of the domain boundaries. The sequence alignment showed the expected highly conserved C-terminal GxGxG motif that is considered to be the hallmark SAM-binding region⁸⁶, and also revealed a second region of high sequence conservation, that is a strong candidate to be the region of the active site (Figure 2.7).

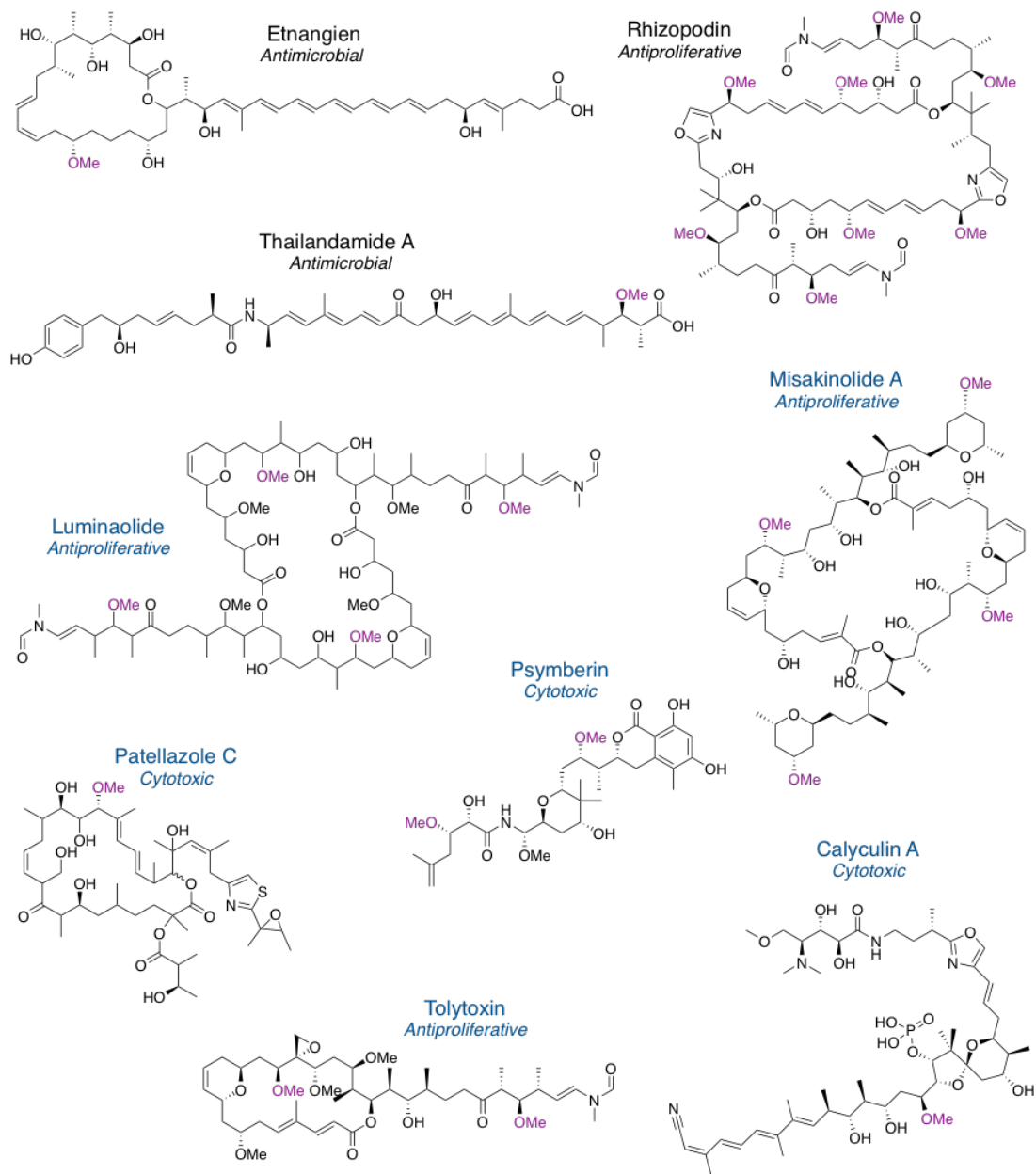


Figure 2.6: The structure of secondary metabolites produced by *trans*-AT PKSs that utilise embedded OMT domains during chain assembly, the O-methyl groups installed by these domains are shown in purple. The titles of natural products derived from marine microorganisms are shown in blue.

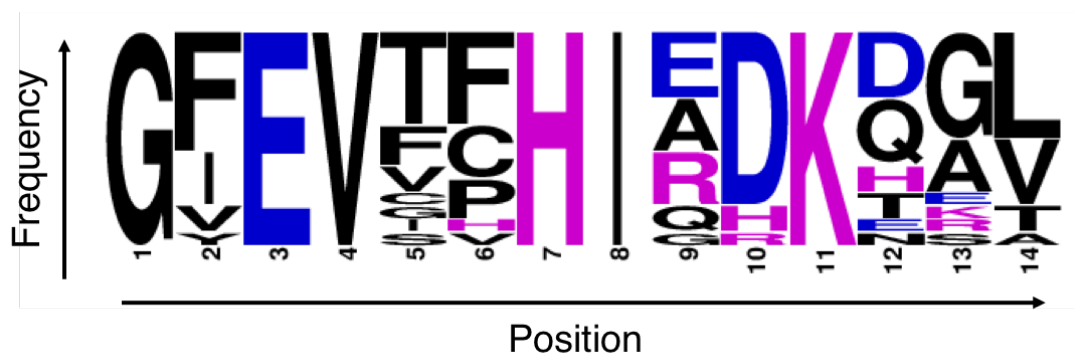


Figure 2.7: Sequence logo, constructed using WebLogo, of a sequence alignment of all reported embedded *OMT* domains from *trans*-AT PKSs, showing the conserved histidine residue proposed to be the key catalytic base. The height of each amino acid indicates the sequence conservation at that position. Basic residues are shown in purple and acidic residues in blue.

The conserved active site motif contained a completely conserved histidine residue. SAM-dependant *OMT*s commonly use histidine, and other basic residues, to deprotonate and thus activate the hydroxyl group for nucleophilic attack on the methyl group of SAM⁸⁷, hence this is the likely role of His153 in the gladiolin *OMT* domain (Figure 2.8).

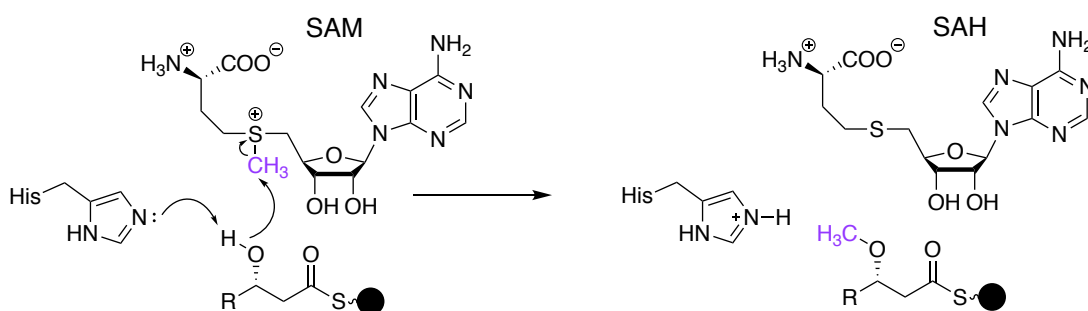


Figure 2.8: The proposed mechanism for SAM dependent methylation in *trans*-AT embedded *OMT*s, • represents the ACP domain. It is not known whether proton transfer happens prior to, or in concert with the transfer of the methyl moiety.

2.3.2 Investigating the activity of GldM11OMT-ACP(H153A)

To probe the role of His153, the residue was mutated to alanine using site-directed mutagenesis. The successful creation of the mutant was confirmed by sequencing and the detected mass change of the resultant protein, GldM11OMT-ACP(H153A) (Appendix 1). The activity of GldM11OMT-ACP(H153A) was evaluated using the methylation assay discussed in section 2.2.2 and no methylation was observed (Figure 2.9). This confirms that this histidine residue is important for catalysis, and also provides an important control reaction to ensure that the observed methylation by GldM11OMT-ACP was carried out by the OMT domain.

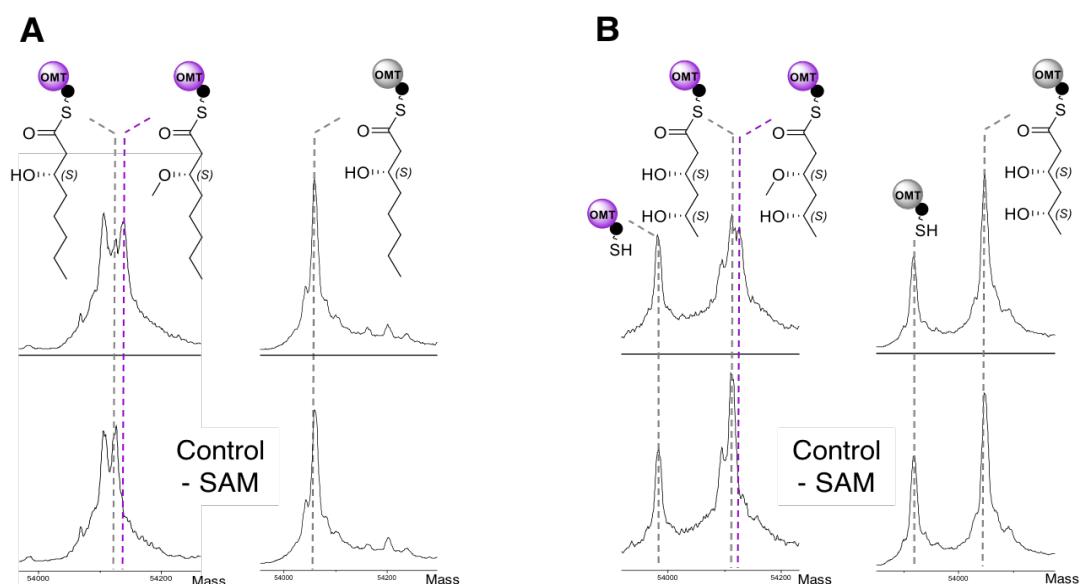


Figure 2.9: Stacked deconvoluted ESI-MS spectra of GldM11OMT-(*S*)-3-hydroxyoctanoyl-ACP (**A**) and GldM11OMT-(3*S*,5*S*)-3,5-dihydroxyhexanoyl-ACP (**B**) after a 1 h 30 min incubation with SAM. GldM11OMT-ACP WT is represented in purple and GldM11OMT-ACP(H153A) is shown in grey. A peak corresponding to methylation (+14 Da) is absent in the GldM11OMT-ACP(H153A) spectra.

2.4 Structural studies

2.4.1 Crystallisation trials

In order to gain further understanding of the active site of embedded OMT domains it was hoped that a crystal structure could be obtained. Structural knowledge of domains is an invaluable tool for prospective PKS engineering, with a thorough understanding of protein interfaces and well defined domain junctions imperative for success⁸⁸. As the domain boundaries for the gladiolin OMT domain were only approximate (based on AntiSMASH predictions), 6 constructs with various *N*- and *C*-termini were created to try and ascertain more precise domain boundaries, which are required for protein crystallography. The *N*- and *C*-termini were chosen based on protein secondary structure predictions made by Phyre²⁸⁹. The constructs were created by cloning the appropriate regions of DNA, that were amplified by PCR from the pET24a-GldM11 OMTACP vector, into the pET28a expression vector using traditional cloning methods. The resultant constructs were tested for the production of soluble protein and only one construct yielded soluble protein, GldM11OMT. It is likely that the other proteins were insoluble due to the chosen domain boundaries disrupting secondary structure, which would prevent the correct folding of the protein.

GldM11OMT was heterologously overproduced on a large scale and purified by both nickel-NTA chromatography and size exclusion chromatography (SEC), as a high level of protein purity was required for crystallography. SDS-PAGE analysis showed that the protein was pure and of the correct molecular weight (Figure 2.10), which was confirmed by intact protein MS (Appendix 1). Proteins overproduced using pET28a constructs undergo *N*-terminal methionine excision and gluconoylation of the histidine-tag, which could be detected in the deconvoluted MS spectra^{90,91}. The activity of this excised OMT domain was demonstrated and the results can be found in section 2.7.

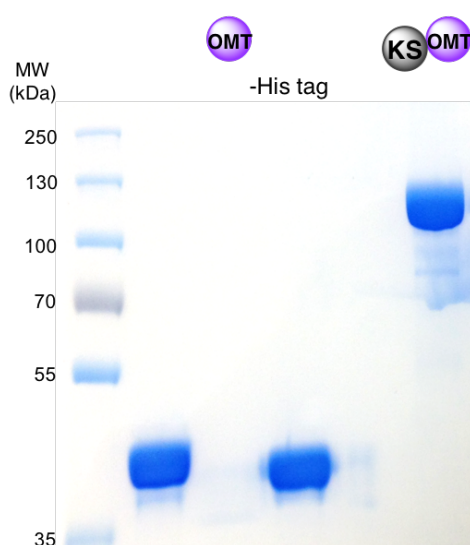


Figure 2.10: 8 % SDS-PAGE gel showing the purified proteins that were used for structural studies. The domain architecture of the proteins is shown above the corresponding lane.

In an attempt to identify conditions that would yield crystals, 7 commercial crystallisations screens were utilised, meaning a variety of pH values, precipitant types and concentrations, as well as different temperatures were tested, however no crystals were obtained. The reason why proteins do not crystallise can be elusive, but it is often found that conformational stability of the protein is paramount⁹². For this reason the crystallisation trials were repeated using GldM11 OMT with the polyhistidine-tag removed as well as the KS-OMT didomain, GldM11KS-OMT. KS domains from PKSs are dimeric and have been demonstrated to crystallise well^{93–97} so it was hoped the presence of this domain would bring additional stability to OMT domain. Unfortunately no conditions were identified that produced protein crystals.

2.4.2 Homology modelling

As a crystal structure could not be obtained a homology model was generated using the intensive modelling mode of Phyre²⁸⁹. The quality of the model was assessed by generating a Ramachandran plot using RAMPAGE software

which revealed that of the 350 residues, 315 were in the favoured region (89.4 %), 22 were in an allowed region (6.3 %) and 15 were in the disallowed region (4.3 %) indicating that the predicted model was not good (Appendix 3). This was not unsurprising given that of the six templates used to build this model the maximum sequence identity was 19 %. A homology model is considered to be good, if a template sequence is used that has at least 30 % sequence identity to the queried sequence⁹⁸, however, literature examples of SAM-dependant MTs share very little sequence identity, but incorporate a highly conserved structural fold⁹⁹. The homology model suggests that like the majority of SAM-dependant MTs, embedded OMT domains from *trans*-AT PKSs belong to the class I MT superfamily. Class I MTs are characterised by alternating α -helices and β -sheets that come together to form a Rossmann-like fold that comprises of a seven-stranded β -sheet flanked by 3 α -helices on each side (Figure 2.11 (A))^{87,99}.

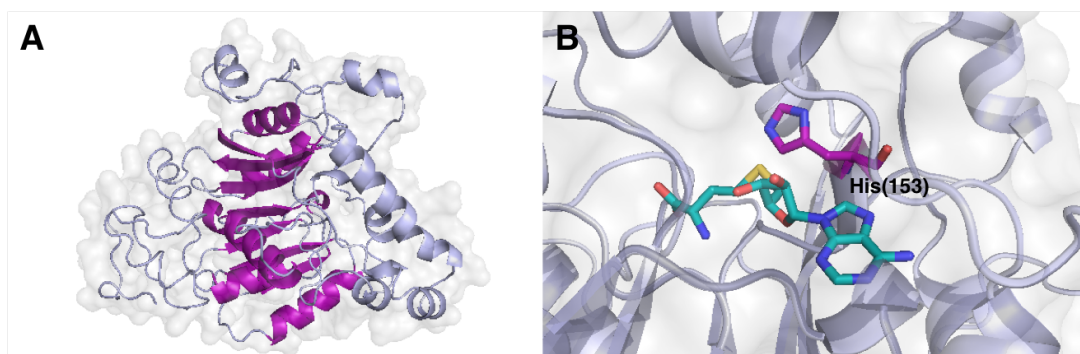


Figure 2.11: Homology model of GldOMT generated using the intensive mode of Phyre2. (A) Overall model with the defining class I MT fold shown in purple. (B) Close up showing the proximity of the proposed key catalytic histidine (purple) to SAM (teal). Images were created using PyMOL¹⁰⁰.

Little information can be gained from this model about substrate binding due to the lack of a good template; unsurprisingly this region varies greatly due to the wide range of substrates class I MTs act upon. Interestingly the model does place the identified key catalytic histidine in close proximity to the sulphur of SAM (Figure 2.11 (B)) which would be required for initiating methyl transfer.

2.5 Phylogenetic analysis

A phylogenetic tree of all embedded OMT domains from known BGCs that were predicted by ClusterTools, and a number of post-PKS tailoring OMTs for which the biosynthetic role has been proven, was generated using ClustalX¹⁰¹ (Figure 2.12).

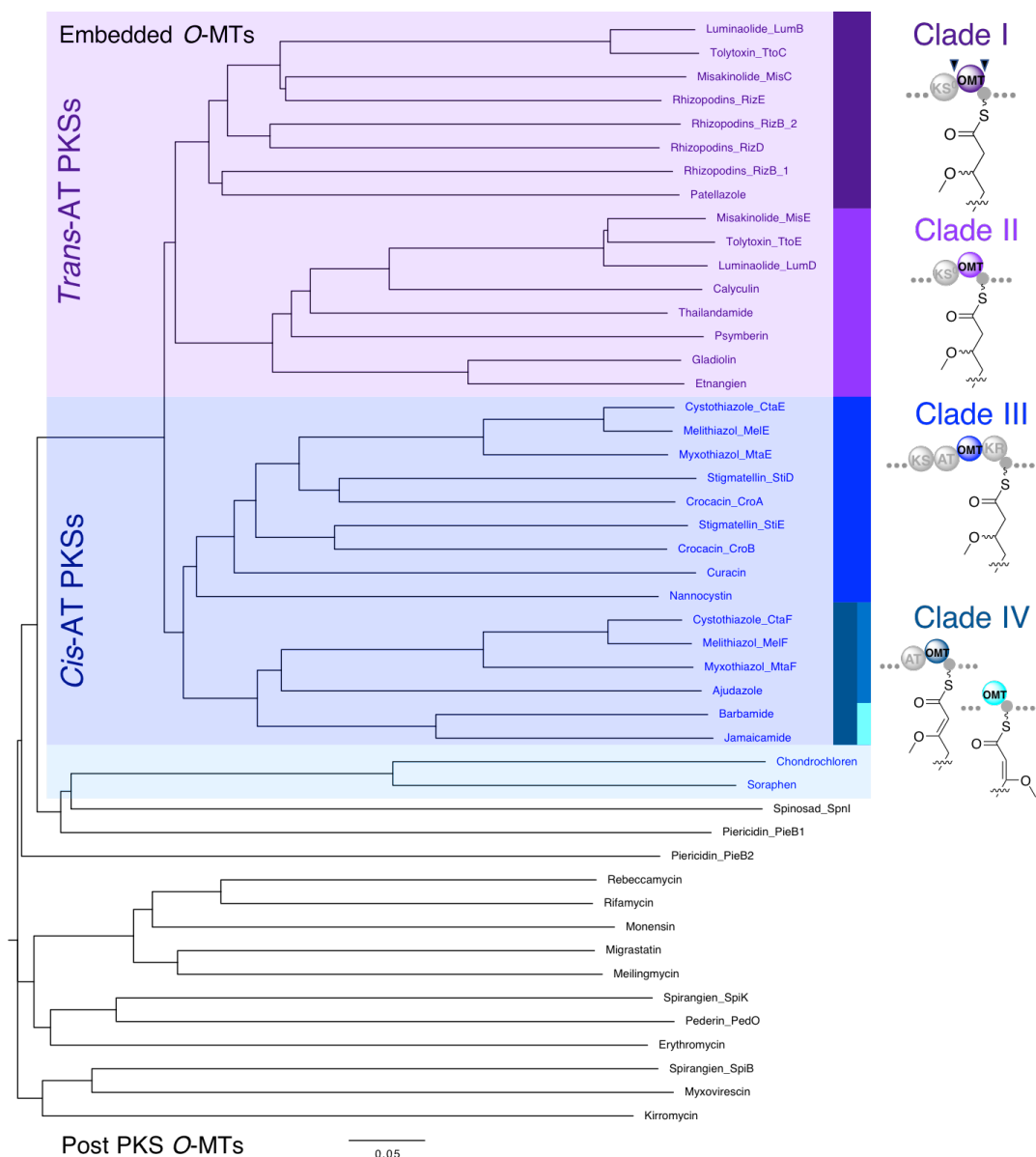


Figure 2.12: Phylogenetic tree of OMT domains from modular type I PKSs. Embedded OMT domains are highlighted, with those found in *trans*-AT PKSs shown in purple and those from *cis*-AT PKSs shown in blue. Where clades are identified the conserved domain architecture is shown.

Embedded *OMT* domains claded separately to post-PKS tailoring enzymes with the phylogenetic analysis suggesting that once an *OMT* domain was recruited into a PKS all other embedded *OMT* domain containing PKSs evolved from this. There are two interesting exceptions that provide potential insight into the evolution of these domains, chondrochloren and soraphen, however the role of these *OMT* domains has yet to be experimentally demonstrated. In the biosynthesis of chondrochloren the *OMT* domain is found at the end of one of the PKS genes and is hypothesised to act across two modules; comparison to similar clusters suggest that this arose from the fusion of an ancestral discrete *OMT* domain¹⁰². The BGC for soraphen contains a unique protein that is comprised of an AT, ACP and *OMT* domain that is proposed to be responsible for the synthesis of the unusual extender unit methoxymalonyl-CoA¹⁰³.

Cis-AT PKSs containing embedded *OMT* domains have been identified exclusively in gram-negative bacteria^{104–113}, from which biosynthetic pathways incorporating novel or atypical features are often found¹¹⁴. These *OMT* domains form two clades based upon whether a KR domain is present within the methylating module. When present it is predicted that the β -hydroxyl substituent generated will be the substrate for the *OMT* domain (Clade III). Where there is no KR domain it is hypothesised that the *OMT* domain gives rise to methylation of the enol form of the β -keto intermediate (Clade IV). Clade IV further splits based upon whether the thermodynamically more stable *trans*-isomer, or the *cis*-isomer, is formed.

The *trans*-AT embedded *OMT* domains also split into two distinct clades, however the reason for this is not apparent. They do not appear to clade based upon surrounding domain architecture or any features of the hypothesised substrates such as stereochemistry of the methylated β -hydroxyl group, substrate length or the presence of α -, γ - or δ - substituents (Figure 2.13).

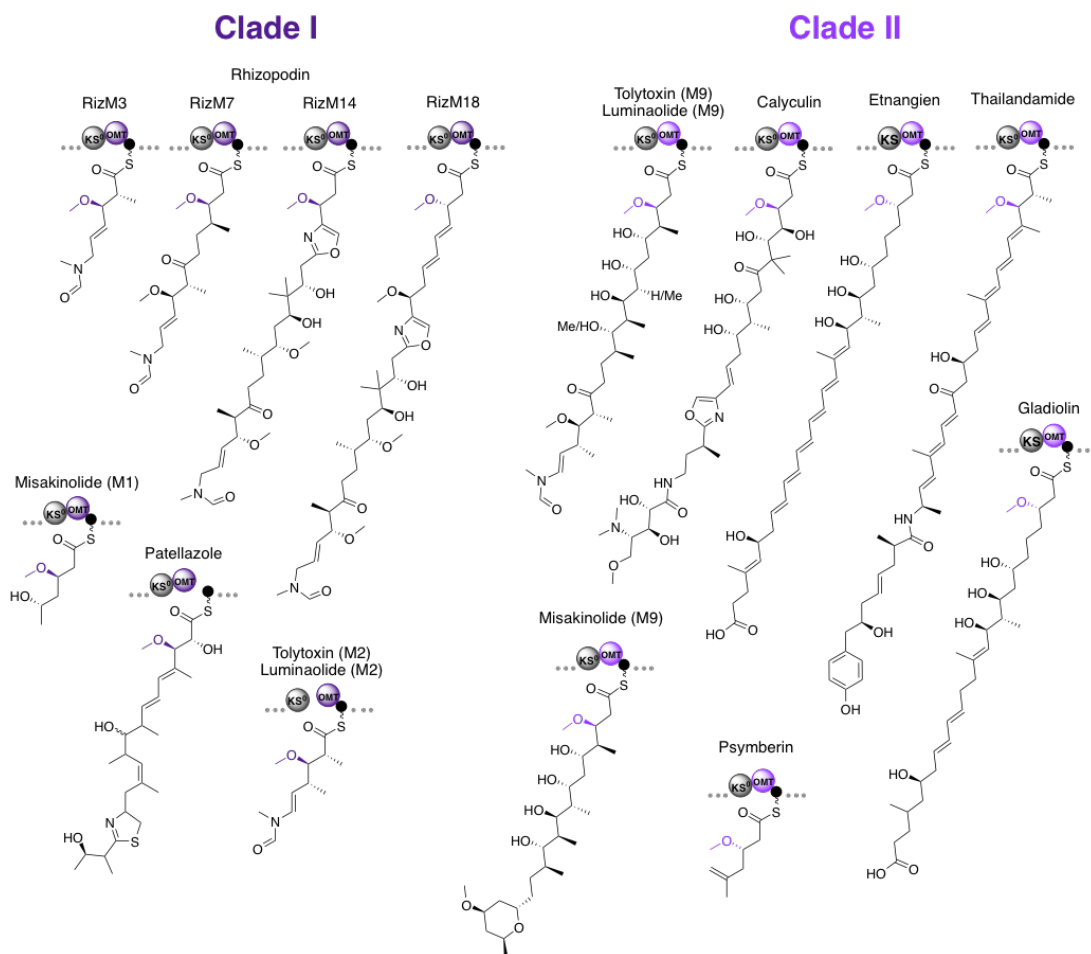


Figure 2.13: The hypothesised substrates for the embedded *OMT* domains from *trans*-AT PKSs. For PKSs that contain more than one embedded *OMT* domain the module containing that *OMT* domain is indicated.

2.6 Comparing the substrate tolerance of embedded *OMT* domains from two different clades

As the phylogenetic analysis revealed that embedded *OMT* domains from *trans*-AT PKSs form two undefined clades, the substrate tolerance of three additional embedded *OMT* domains, two from the misakinolide PKS⁸⁰ and one from the rhizopodin PKS⁸⁵, was explored in an attempt to shed light upon this. The MisM9*OMT* domain clades with the GldM11*OMT* domain (clade II), whilst the MisM1*OMT* domain clades with the RizM18*OMT* domain (clade I). The

DNA sequences of these OMT-ACP didomains were obtained from the MIBiG repository⁷⁹, and were purchased as synthetic DNA fragments (GenScript) that were cloned into the pET28a expression vector. A pET24a-MisM1OMTACP synthetic construct (purchased from Epoch Life Sciences) was used for the overproduction of MisM1OMTACP. The corresponding proteins were overproduced as histidine-tagged proteins and purified by Ni-NTA chromatography. SDS-PAGE analysis showed that all of the proteins were of high purity and of the predicted size (Figure 2.14), which was confirmed by intact protein MS (Appendix I).

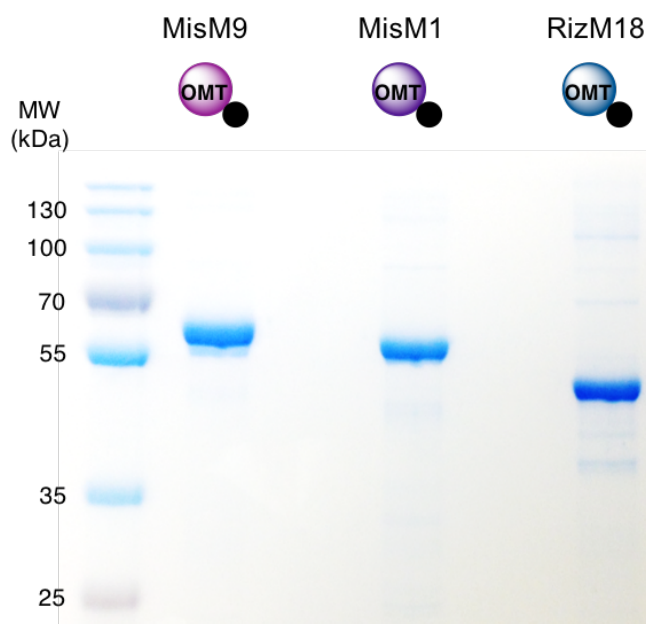


Figure 2.14: 10 % SDS-PAGE gel showing the purified proteins, MisM9OMT-ACP (represented in pink), MisM1OMT-ACP (represented in dark purple) and RizM18OMT-ACP (represented in blue).

A summary of the results from the methylation assays carried out, as discussed in section 2.2.2, is shown in Figure 2.15 and the supporting methylation data can be found in Appendix 4. The data suggests that substrate tolerance is something that is unique to each individual OMT domain and thus unlikely to be a contributing factor towards the formation of the two clades. The MisM1OMT domain demonstrated stereoselectivity towards the methylating hydroxyl group up until the substrate tested became longer than

the natural substrate (shown in Figure 2.13), which interestingly was still accepted as a substrate by the OMT domain. The other two characterised OMT domains did not demonstrate any clear preference for one stereo-configured hydroxyl group over the other, although MisM9 would only accept its proposed natural stereoisomer when a 4-carbon chain substrate was used. Given that the KR domain found within each methylation module would generate this hydroxyl group in a stereospecific manner there would be no evolutionary pressure on the OMT domain to demonstrate selectivity for efficient biosynthesis. The RizM18OMT domain showed no activity for the substrates tested that had a shorter chain length than 8 carbons, however it cannot be determined from the experiments carried out whether the dihydroxyhexanoyl substrates were not methylated due to the chain length, or because the OMT domain could not accept substrates with a δ -hydroxyl group.

In general the studied domains have demonstrated useful substrate promiscuity. The MisM9OMT stands out as the best candidate for future bioengineering as it accepts a range of short substrate mimics used in this study despite the fact that *in vivo* the substrate for this domain is substantially longer (see Figure 2.13) suggesting that it would be capable of methylating a wide range of chain lengths. This OMT domain also showed a lack of stereoselectivity towards the hydroxyl group being methylated meaning theoretically it could be placed after any KR domain in a PKS pathway, and demonstrated a tolerance for substrates with δ -hydroxyl groups, of both stereochemistries, which is a feature often found in growing polyketide chains. For completion it would be interesting, in the future, to determine if this domain can also tolerate substrates with both α - and γ -substituents.

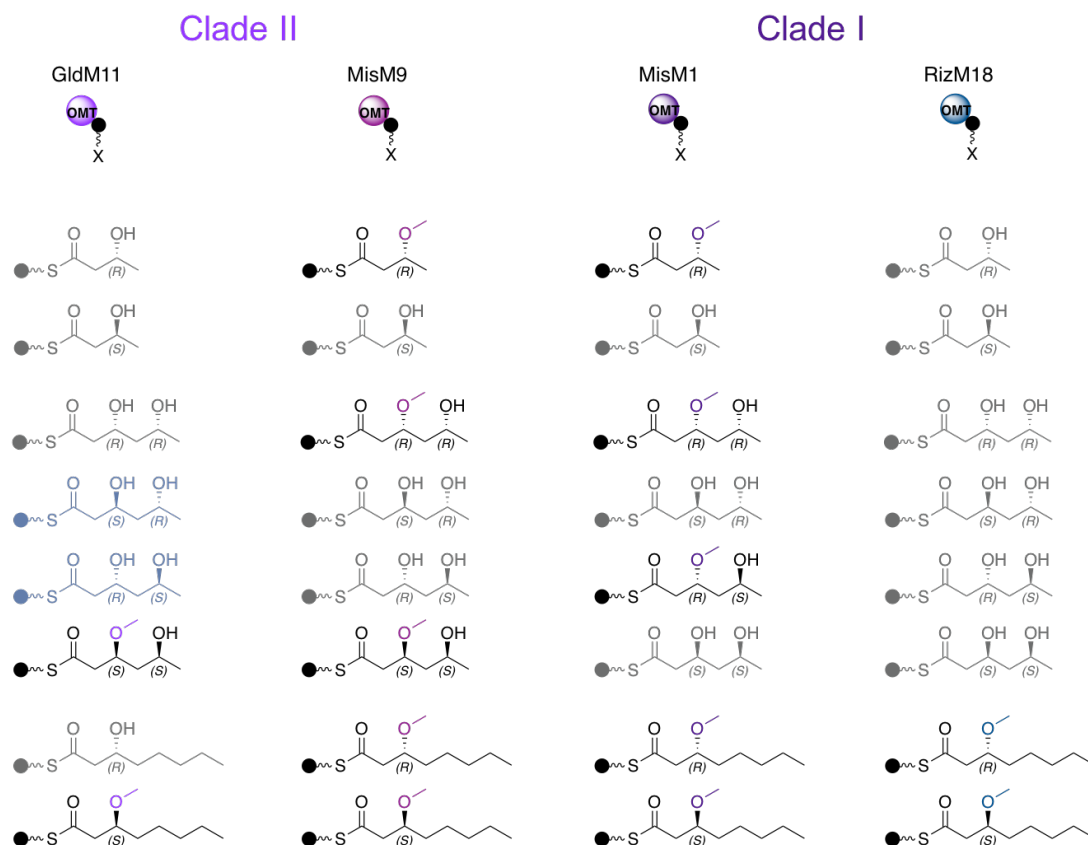


Figure 2.15: A summary of the substrate tolerance determined for several embedded *OMT* domains from *trans*-AT PKSs. Substrates accepted by the *OMT* domain are shown in bold and the predicted site of methylation is indicated. For the substrates shown in blue it is unclear if they can be methylated by the *OMT* domain as the substrates were lost from the protein due to lactonization.

2.7 ACP domain specificity and the role of the KS⁰ domain

2.7.1 The role of the KS⁰ domain

In *trans*-AT PKSs embedded *OMT* domains are always found with hypothesised KS⁰ domains, that are predicted to translocate the growing polyketide chain from one ACP domain to another, so that *O*-methylation can occur (see Figure 2.1). The need for this suggests that there is a unique

interaction required between an OMT domain and an ACP domain for catalytic activity.

In the gladiolin PKS this KS domain is not a true KS⁰ by definition, as it lacks the associated mutation of the histidine residue needed to catalyse chain elongation⁴⁶, however the surrounding residues in the usually conserved HGTGT motif are mutated to HSVGS. These surrounding mutations along with the fact that the gladiolin PKS contains no true KS⁰ domains, but has 3 surplus KS domains when considering the number of chain elongations required to assemble gladiolin, has led to the prediction that this KS domain acts only as a transacylase⁶⁵.

In order to demonstrate the transacylation activity of this KS domain *in vitro*, both of the ACP domains from the gladiolin O-methylation module, GldM11ACP(I) and GldM11ACP(II), were cloned into pET28a, heterologously overproduced in *E. coli* and purified to homogeneity by nickel affinity chromatography. SDS-PAGE analysis showed that both of these proteins were pure and of the correct molecular weight (Figure 2.16), which was confirmed by intact protein MS (Appendix 1).

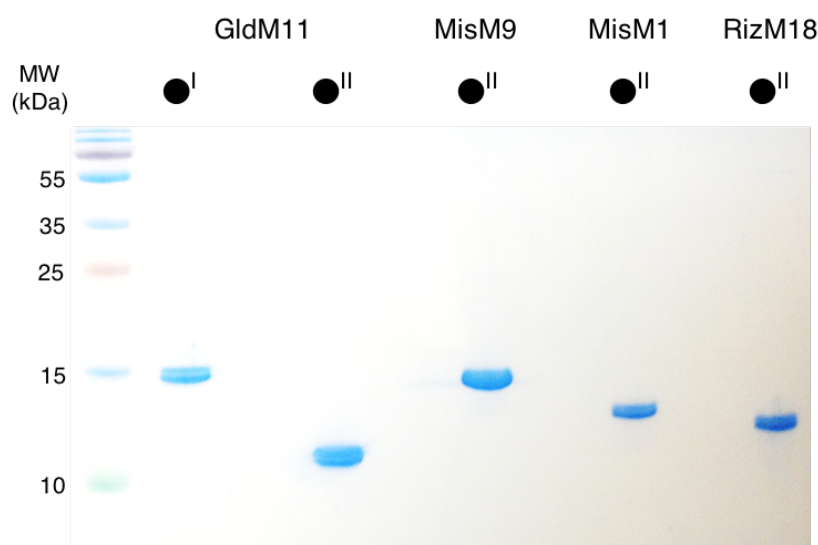


Figure 2.16: 15 % SDS-PAGE gel showing the purified ACP domains used in this study.

To demonstrate transacylation activity of the KS domain, GldM11ACP(I) was first loaded with (*S*)-3-hydroxyoctanoyl-pantetheine as described in section 2.2.2. GldM11ACP(II) was also converted to the *holo*-form by incubating the protein with Sfp and CoA. The resultant *holo*-ACP domains were then incubated with the GldM11KSOMT didomain (refer to section 2.4.1) for 3 h. Intact protein MS analysis of the assay revealed that in the presence of the GldM11KSOMT didomain the (*S*)-3-hydroxyoctanoyl moiety was offloaded from GldM11ACP(I), and partial transfer to GldM11ACP(II) could be observed (Figure 2.17). In the absence of the GldM11KSOMT didomain negligible transfer was seen. This strongly suggests that the KS domain is capable of catalysing translocation, however, to confirm that the KS domain is responsible for the observed transfer, this assay will need to be conducted with a GldM11KSOMT variant, where the KS domain active site cysteine residue is mutated. Work to create this mutant is currently ongoing.

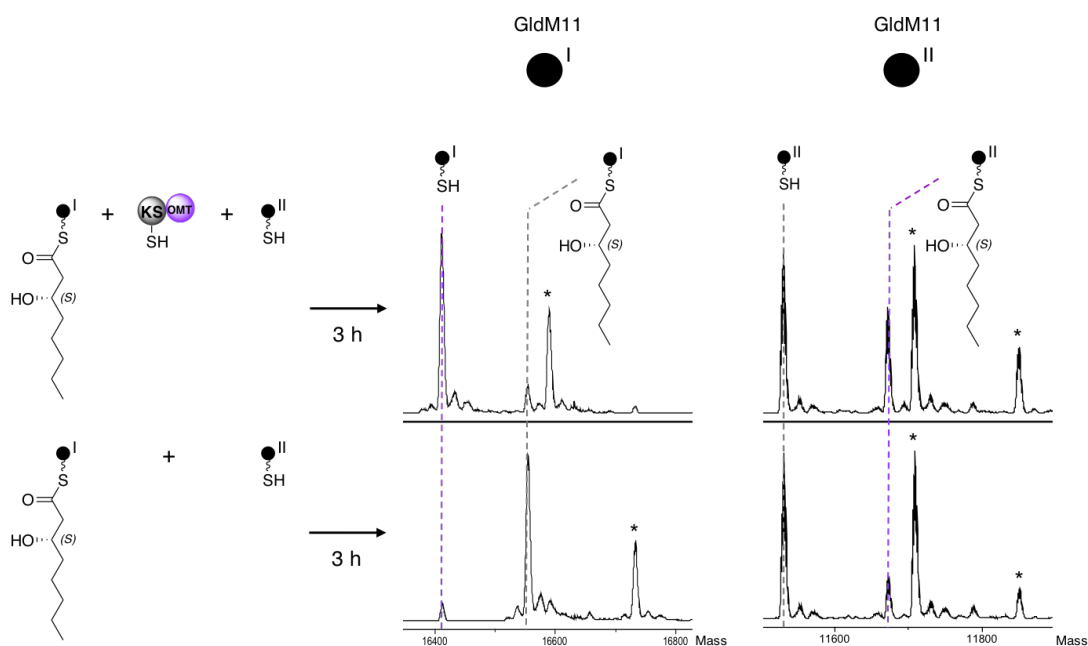


Figure 2.17: Stacked deconvoluted ESI-MS spectra of GldM11-(*S*)-3-hydroxyoctanoyl-ACP(I) and *holo*-GldM11ACP(II) after incubation with and without the GldM11KSOMT didomain. The data shows that the GldM11KSOMT didomain is able to catalyse the transfer of the (*S*)-3-hydroxyoctanoyl moiety from GldM11ACP(I) to GldM11ACP(II).

2.7.2 ACP domain specificity in the gladiolin *O*-methylation module

In order to determine if a specific interaction could be observed between the GldM11OMT domain and the second ACP domain from the gladiolin methylation module, an assay was conducted where both of the ACP domains from the methylation module, loaded with the (*S*)-3-hydroxyoctanoyl pantetheine substrate (see section 2.2.2), were incubated in turn with GldM11OMT in the presence of SAM and the *S*-adenosylhomocysteine (SAH) nucleosidase, Pfs, that prevents potent inhibition of the OMT domain by SAH¹¹⁵. Pfs was included as this assay relies on the OMT domain being able to conduct several catalytic cycles. In each case an excess of the OMT domain was used to account for the fact that the intermolecular interactions of the assay will be less efficient than the intramolecular interactions of the native system. The results are presented in Figure 2.18.

Intact protein MS analysis of the assay showed that *O*-methylation was only able to occur on the second ACP domain of the methylating module, establishing the fact that a specific interaction between an ACP domain and OMT domain is required for methylation. Methylation was not observed for the first ACP domain even when the ratio of ACP domain to OMT domain was increased to 1:4, respectively, which eliminates any potential bias that could occur due to any error in the calculated protein concentrations. This observed ACP domain specificity provides evidence towards this being the reason for incorporation of a non-elongating KS domain as part of the *O*-methylation module.

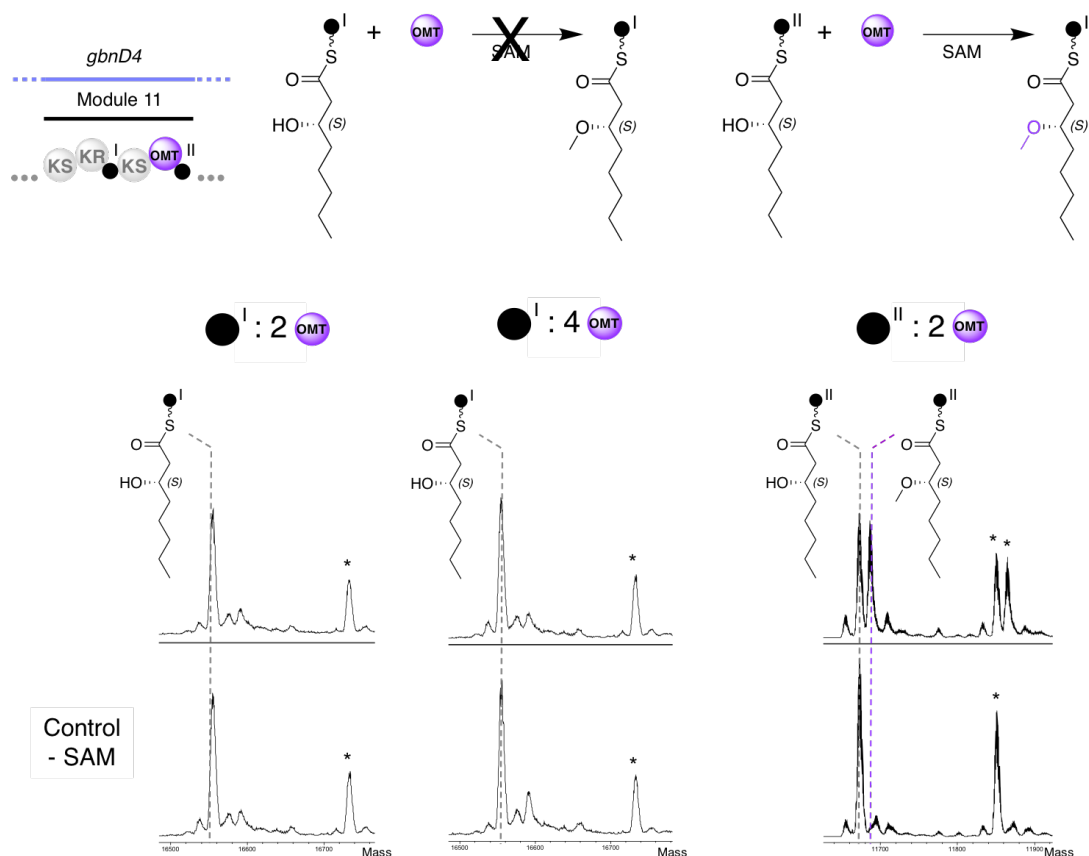
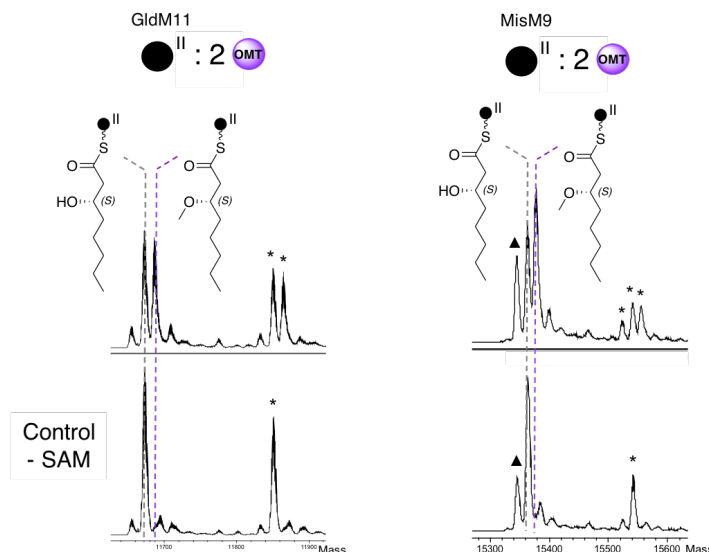


Figure 2.18: Stacked deconvoluted ESI-MS spectra of (*S*)-3-hydroxyoctanoyl-ACP domains from the gladiolin methylation module, after incubation with GldOMT. Methylation (+ 14 Da) is only observed for ACP(II). Peaks marked with an * show the his-tag gluconoylated protein.

2.7.3 ACP domain crosstalk

To assess the ability of the GldOMT domain to interact with the ACP(II) domains from the other *O*-methylation modules looked at in this study, the ACP(II) domains were first cloned into pET28a, heterologously overproduced and purified by nickel affinity chromatography. SDS-PAGE analysis showed that the ACP domains were pure and of the correct molecular weight (Figure 2.16), which was confirmed by intact protein MS (Appendix 1). The crosstalk assays were conducted as described in section 2.7.2 and the results of the analysis of the assays by intact protein MS are presented in Figure 2.19.

Clade II



Clade I

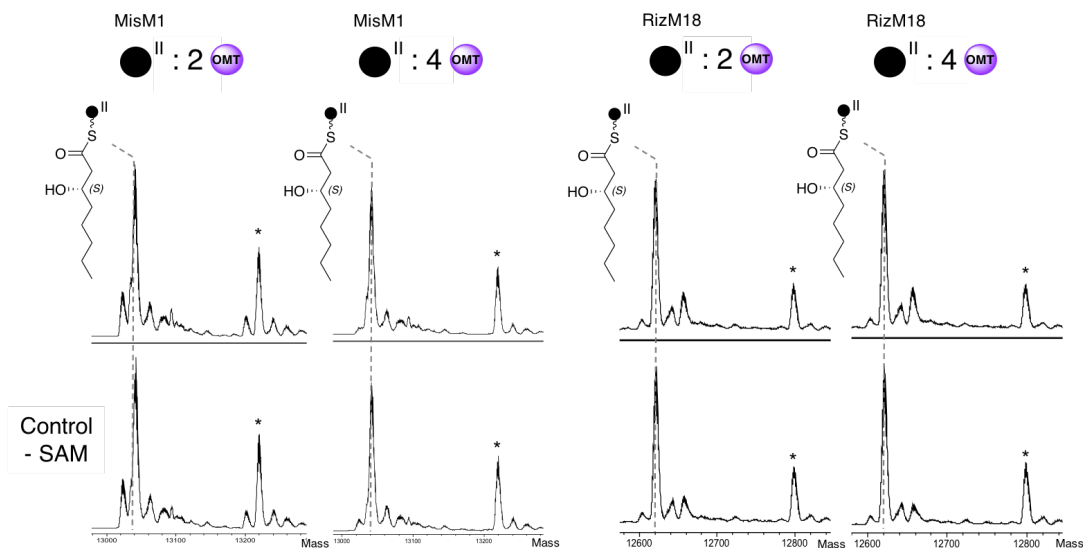


Figure 2.19: Stacked deconvoluted ESI-MS spectra of (*S*)-3-hydroxyoctanoyl-ACP(II) domains, after incubation with GldOMT. Methylation (+ 14 Da) is only observed for GldM11ACP(II) and MisM9ACP(II). Peaks marked with an * are from the gluconoylated protein and those marked with a ▲ are due to the loss of water.

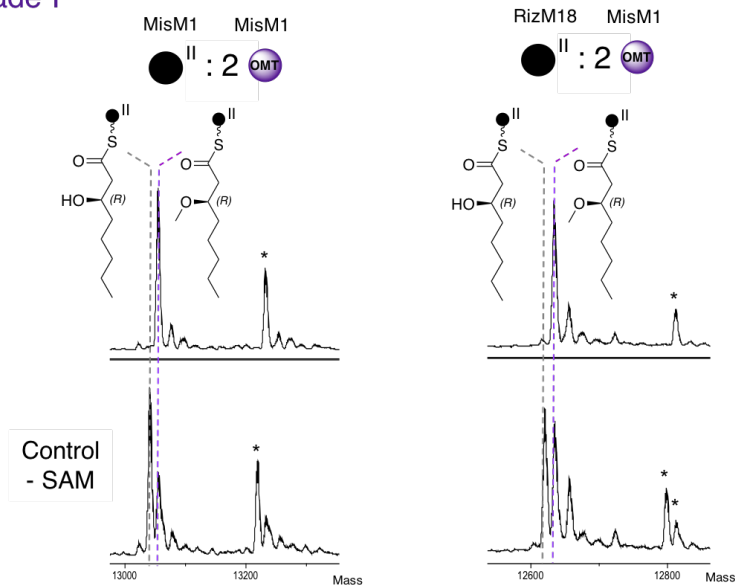
The crosstalk assays revealed that the GldOMT domain was only able to methylate the (*S*)-3-hydroxyoctanoyl substrate when bound to its cognate ACP(II) domain or the MisM9ACP(II), which is an ACP domain associated with an embedded OMT domain from the same clade, clade II. No methylation was observed for the two ACP domains associated with OMT domains from clade

I (even with a 1:4 ratio of ACP domain to OMT domain). This suggested that two different methods of communication could exist between embedded OMT domains and ACP domains.

To investigate this further the MisM1OMT domain was cloned into pET28a, heterologously overproduced and purified to homogeneity by nickel affinity and SEC chromatography (SDS-PAGE analysis and intact protein MS data can be found in Appendix I). The crosstalk assays were then repeated for this OMT domain from clade I, and the results are shown in Figure 2.20. The assays were conducted using (*R*)-3-hydroxyoctanoyl-pantetheine as the substrate, as the MisM1OMT domain utilises an *R*-configured alcohol as its natural substrate. The MisM1OMT domain was able to methylate this substrate when bound to all of the ACP domains. Furthermore, for the ACP domains from clade I OMT domains, methylation was also observed in the control reaction where no SAM was added. This indicates that the MisM1OMT domain retained some SAM in the active site during protein purification. The fact that methylation was observed in the control reaction for these ACP domains and not in the control reaction for ACP domains from clade II OMT domains suggests that a weaker interaction exists between the MisM1OMT domain and ACP domains associated with OMT domains from the other clade.

Overall the crosstalk assays suggest that two different methods of communication exist between embedded OMT domains and ACP domains and that this contributes to the formation of the two clades. As the two different clades were revealed through analysis of the sequence of the OMT domains alone, it is likely that some structural variations exist between the OMT domains from the two clades that facilitate the different interactions, however a crystal structure of an OMT domain from each of the clades will be needed to confirm this.

Clade I



Clade II

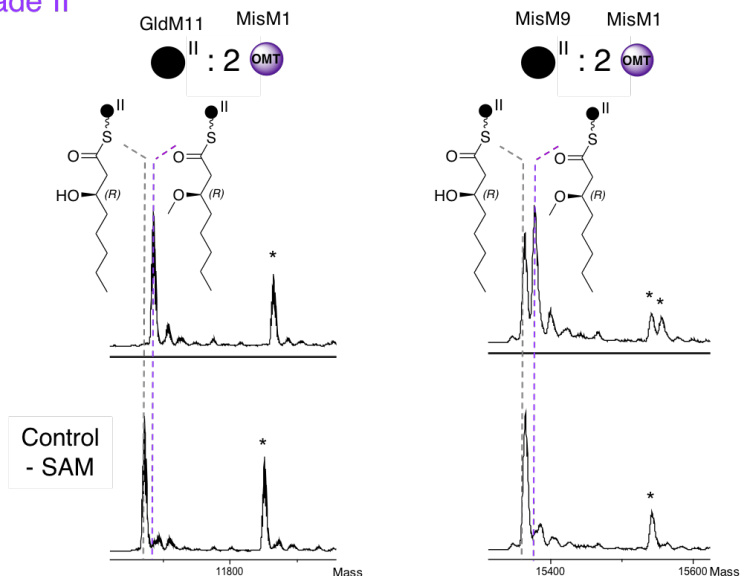


Figure 2.20: Stacked deconvoluted ESI-MS spectra of (*R*)-3-hydroxyoctanoyl-ACP(II) domains, after incubation with MisM1 OMT. Methylation (+ 14 Da) is observed for all ACP domains. Peaks marked with an * are from the gluconoylated protein.

2.8 Summary

This study presents the first characterisation of embedded OMT domains from *trans*-AT PKSs. A mass spectrometry based assay was designed and

implemented to assess the activity of these domains *in vitro*. The substrate tolerance of four embedded OMT domains from three different *trans*-AT PKSs was explored. This revealed the potentially useful substrate promiscuity of these domains. It was established that the embedded OMT domains showed regioselectivity for methylation of β -hydroxyl groups and it was found that three of the studied OMT domains were able to methylate hydroxyl groups of both stereochemistries. A completely conserved histidine residue was identified in OMT domains from *trans*-AT PKSs, and mutation of this residue to an alanine in the gladiolin OMT domain lead to the loss of activity. This suggests that this histidine is likely to act as the key catalytic base in these enzymes.

OMT domains are always found after non-elongating (or hypothesised non-elongating) KS domains in *trans*-AT PKSs. It was proposed that this was so that the KS domain could pass the growing polyketide chain to the downstream ACP domain, so that *O*-methylation could occur. Exploration of this, in the gladiolin *O*-methylation module, revealed that *O*-methylation was only able to occur on the downstream ACP domain, thus providing evidence for ACP domain specificity being the reason for the incorporation of non-elongating KS domains in *O*-methylation modules. Phylogenetic analysis of all known examples of embedded OMT domains from *trans*-AT PKSs revealed that these domains split into two distinct clades. OMT-ACP domain crosstalk assays revealed that the gladiolin OMT domain (clade II) was only able to methylate substrates when bound to ACP domains associated with other embedded OMT domains from the same clade. For the MisM1 OMT domain (clade I) it was found that the OMT domain was able to methylate substrates bound to all ACP domains tested, however, a greater extent of methylation was observed for ACP domain associated with OMT domains from the same clade, suggesting that there is a stronger interaction between these domains. These results indicate that two different methods of communication may exist between embedded OMT domains and ACP domains, and that this is a contributing factor to the formation of the two clades.

Chapter 3

Investigating cryptic starter unit amidation

3.1 Introduction

In the gladiolin BGC there are several putative auxiliary genes of unknown function⁶⁵. Two are of particular interest; *gbnC* that encodes for an asparagine synthetase homologue and *gbnM* which encodes for an amidase (Figure 3.1 (A)). Asparagine synthetase enzymes catalyses the transfer of the amide nitrogen of glutamine to the side chain of aspartic acid in an ATP dependant reaction¹¹⁶, and amidase enzymes usually catalyse the hydrolysis of an amide, generating a carboxylic acid and ammonia¹¹⁷ (Figure 3.1 (B)). As the structure of gladiolin contains no nitrogen atoms this raises the question of what role these genes play in the biosynthesis.

Based on the general activity of these enzymes it is proposed that *gbnC* acts upon the carboxylic acid moiety of gladiolin which is derived from the hypothesised starter unit, succinyl-CoA, and thus *gbnM* would then be responsible for converting the resultant amide derivative to the mature antibiotic.

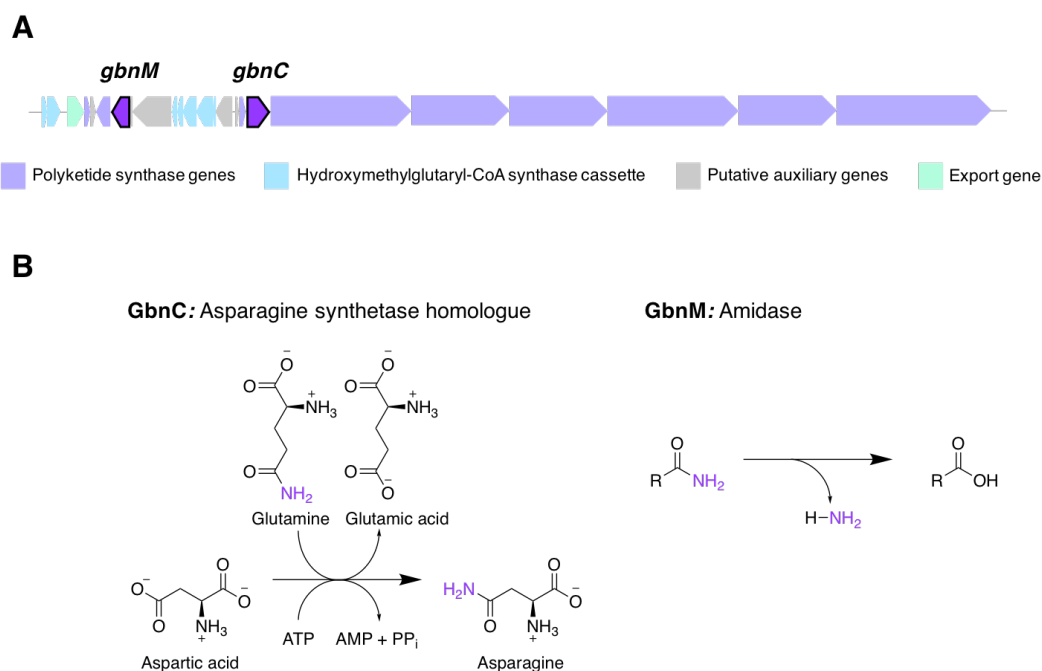


Figure 3.1: (A) The gladiolin biosynthetic gene cluster, the two genes of interest *gbnC* and *gbnM* are highlighted in purple. (B) The general activity of the proposed enzymes.

The use of protecting groups in the biosynthesis of antibiotics is well preceded. A common self-resistance mechanism employed by antibiotic producers to avoid suicide is to either build the antibiotic as a pro-drug, or to enzymatically modify the antibiotic to generate an inactive form, which would then be converted to the active compound immediately prior to or after export from the cell¹¹⁸. An example of this can be found in the biosynthesis of oleandomycin, a polyketide macrolide antibiotic from *Streptomyces antibioticus*. The BGC was found to encode for a glycosyltransferase that was shown to glycosylate oleandomycin, resulting in an inactive derivative. The cluster also encoded for a glycosidase that was able of convert glycosylated oleandomycin to the mature antibiotic¹¹⁹.

There have also been several reported examples where protecting groups have been utilised in the biosynthesis of polyketides where cyclisation of intermediates would lead to the facile formation of either 5- or 6-membered rings^{120,121} (Figure 3.2 (A)). The succinyl-intermediate proposed in the initiation

of gladiolin biosynthesis would be prone to this (Figure 3.2 (B)), thus efficiency of biosynthesis may be improved through the transformation of the carboxylate moiety, to a less nucleophilic amide (the pKa of succinic acid is 4.2 so at biological pH the carboxylic acid moiety would exist deprotonated).

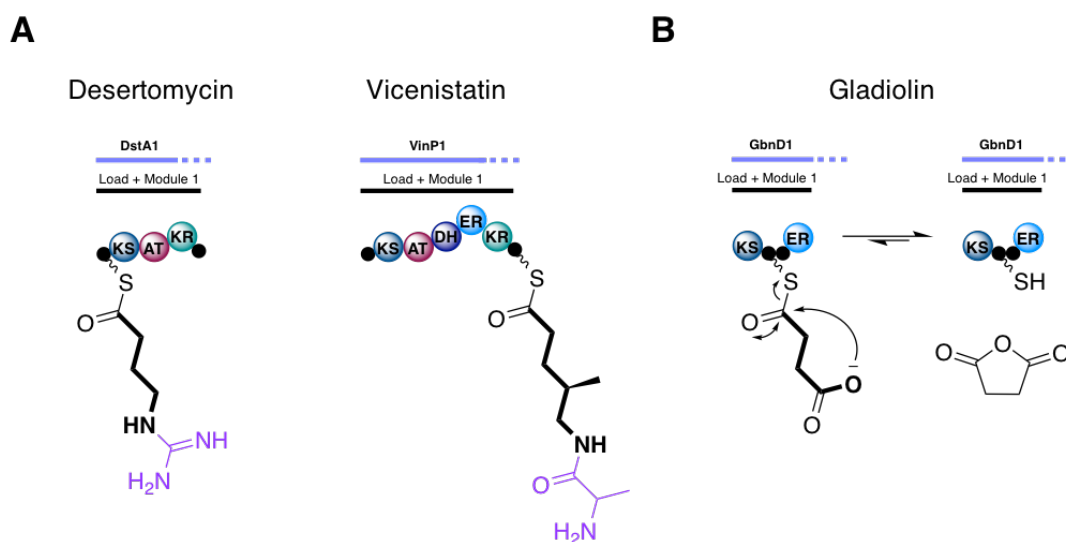


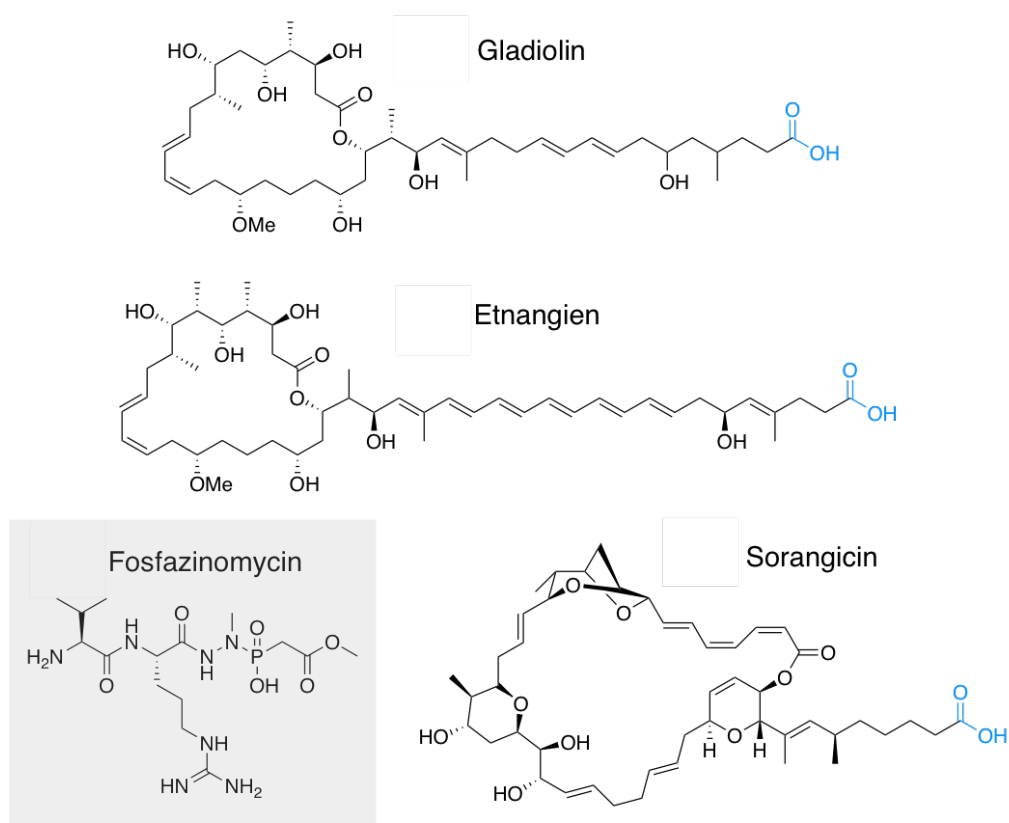
Figure 3.2: (A) Examples of protecting groups used in the biosynthesis of Desertomycin and Vicenistatin. In each case the intermediate shown in bold, unless protected (protecting group shown in purple), would be vulnerable to cyclisation forming a 5- or 6-membered ring. (B) Schematic showing how the intermediate proposed in the initiation of gladiolin biosynthesis would be prone to cyclisation.

The aim of this study is to ascertain the role of the amidase, *gbnM*, and to investigate at what point in the gladiolin biosynthesis the asparagine synthetase, *gbnC*, acts, in order to obtain an insight into the role these genes play in the biosynthesis.

3.2 Identification of BGCs containing homologs of GbnC and GbnM

Using ClusterTools⁷⁸ (discussed in section 2.3.1) the MIBiG repository⁷⁹ was searched for known BGCs containing homologs of both GbnM and GbnC. Three additional BGCs were found and the secondary metabolites for the identified pathways are shown in Figure 3.3 (A). Two share significant similarities with gladiolin; both etnangien and sorangicin are biosynthesised by *trans*-AT PKSs that are predicted to utilise dicarboxylic acid starter units, and they both demonstrate antibiotic activity through inhibition of RNA polymerase^{48,122}. The role of the asparagine synthetase and amidase has not been determined for either of these biosynthetic pathways, however it is presumed analogous to their role in the biosynthesis of gladiolin. The identified exception was the phosphonate natural product fosfazinomycin, although the biosynthesis is not fully understood it is believed that these genes, which demonstrate substantially less sequence identity to the other homologs (Figure 3.3 (B)), are involved in the formation of the hydrazine moiety¹²³.

Genome mining using identified antibiotic resistance genes is a technique that has been successfully used to find novel antibiotics that act upon a chosen target of interest¹²⁴. In an attempt to identify cryptic clusters encoding for potential RNA polymerase inhibitors that share a similar scaffold to gladiolin, a database constructed of all representative and reference genomes from NCBI was searched using ClusterTools. The search looked for biosynthetic pathways that contained both a homolog of GbnM and GbnC as well as a PKS KS domain and *trans*-AT docking domain, however no novel BGCs could be found.

A**B**

Percentage identity (%)				
	Gladiolin	Etnangien	Sorangicin	Fosfazinomycin
GbnC	100	56	45	35
		100	55	34
			100	28
				100
EtnC	65	100		
SorQ	50	52	100	
FzmA	22	22	23	100
	Gladiolin	Etnangien	Sorangicin	Fosfazinomycin
Percentage identity (%)				

Figure 3.3: (A) The structure of secondary metabolites produced by BGCs containing homologs of both GbnC and GbnM; the carboxylic acid moiety that these genes are proposed to act upon is highlighted in blue. (B) Percentage identity matrices showing the similarity between each gene and their identified homologs.

3.3 Investigating the role of *gbnM*

3.3.1 Creation of knock-out mutants in *B. gladioli* BCC1622

In order to study the role of *gbnM*, and the other genes of interest, *in vivo* a mutagenesis system based on the homing endonuclease I-SceI, which had been optimised for use in *Burkholderia* species, was used¹²⁵ (Figure 3.4).

First the sequences flanking the gene targeted for deletion were cloned into the pGPI suicide plasmid. This plasmid was then transferred into the gladiolin producer, *B. gladioli* BCC1622, by triparental mating. This strain was chosen as the target for gene deletions as the original gladiolin producer, *B. gladioli* BCC0238, was not amenable to genetic manipulation due to a high level of antibiotic resistance. Trimethoprim selection was used to drive homologous recombination, as the pGPI plasmid is not able to self-replicate, so targeted insertion of the plasmid into the chromosome is necessary for survival. The pGPI vector contains an I-SceI recognition site, so upon introduction of the pDAI-SceI plasmid that constitutively expressed the I-SceI nuclease, again by triparental mating, a double stranded cut in the DNA was formed. This promoted intramolecular recombination that either introduced the desired mutation or re-formed the wild type (WT).

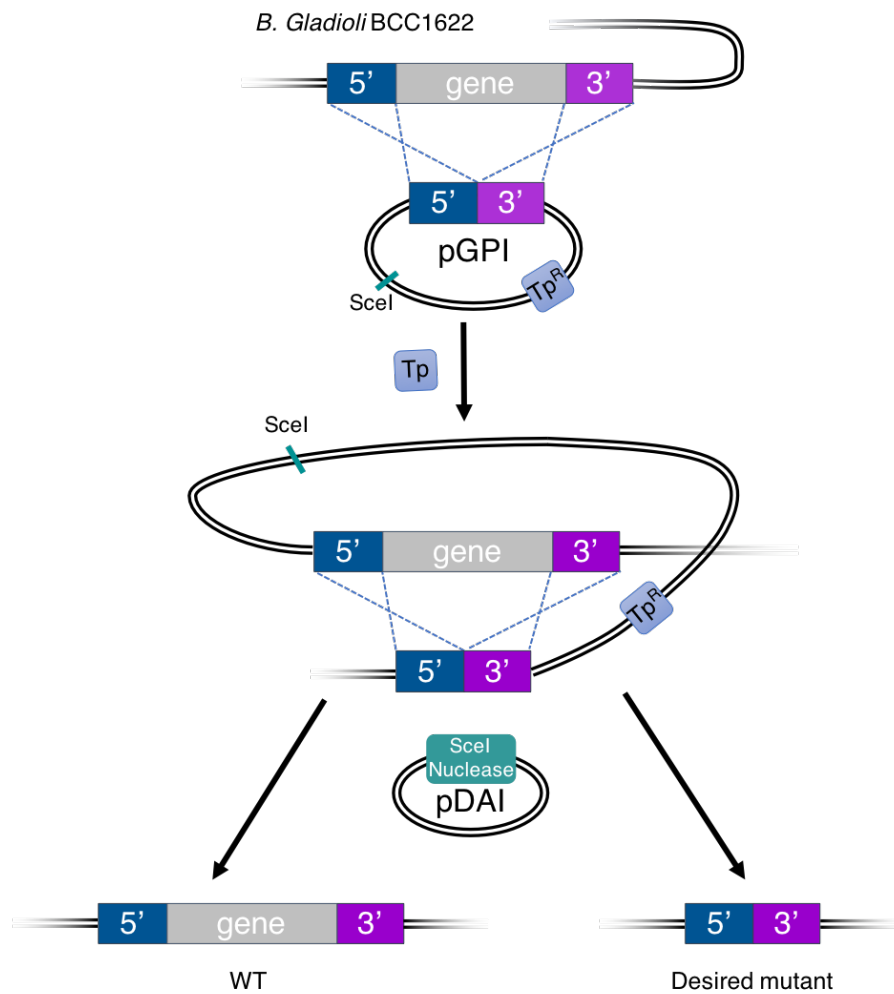


Figure 3.4: Overview of the homing endonuclease I-SceI mutagenesis system used to generate deletion mutants in *B. gladioli* BCC1622.

3.3.2 *In vivo* studies

To ascertain the role of the amidase, *gbnM*, a mutant with the gene cleanly deleted was created using the method discussed in the previous section. PCR analysis was used to identify colonies carrying the appropriate gene deletion (Figure 3.5).

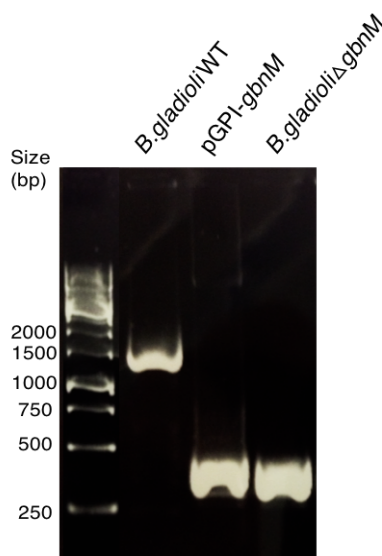


Figure 3.5: Agarose gel electrophoresis showing genetic evidence for the creation of a *gbnM* deletion mutant in *B. gladioli* BCC1622.

The secondary metabolite profile of a successful mutant was analysed by high-resolution LC-MS (Figure 3.6). *B. gladioli* BCC1622 Δ *gbnM* produced a compound with a reduced retention time and mass decrease of 1 Da compared to the WT which was consistent with what would be expected for a more polar amide derivative of gladiolin. To confirm that the introduced mutation was responsible for this observation, the mutant was complemented by *in trans* expression of *gbnM*, which resulted in partial restoration of gladiolin production. The complementation was achieved using the arabinose-inducible pMLBAD vector, which *gbnM* was cloned into. This vector was chosen as it had been shown to effectively express genes in *Burkholderia cepacia*¹²⁶. This result establishes that amidation of gladiolin or a gladiolin precursor plays a role in the biosynthesis.

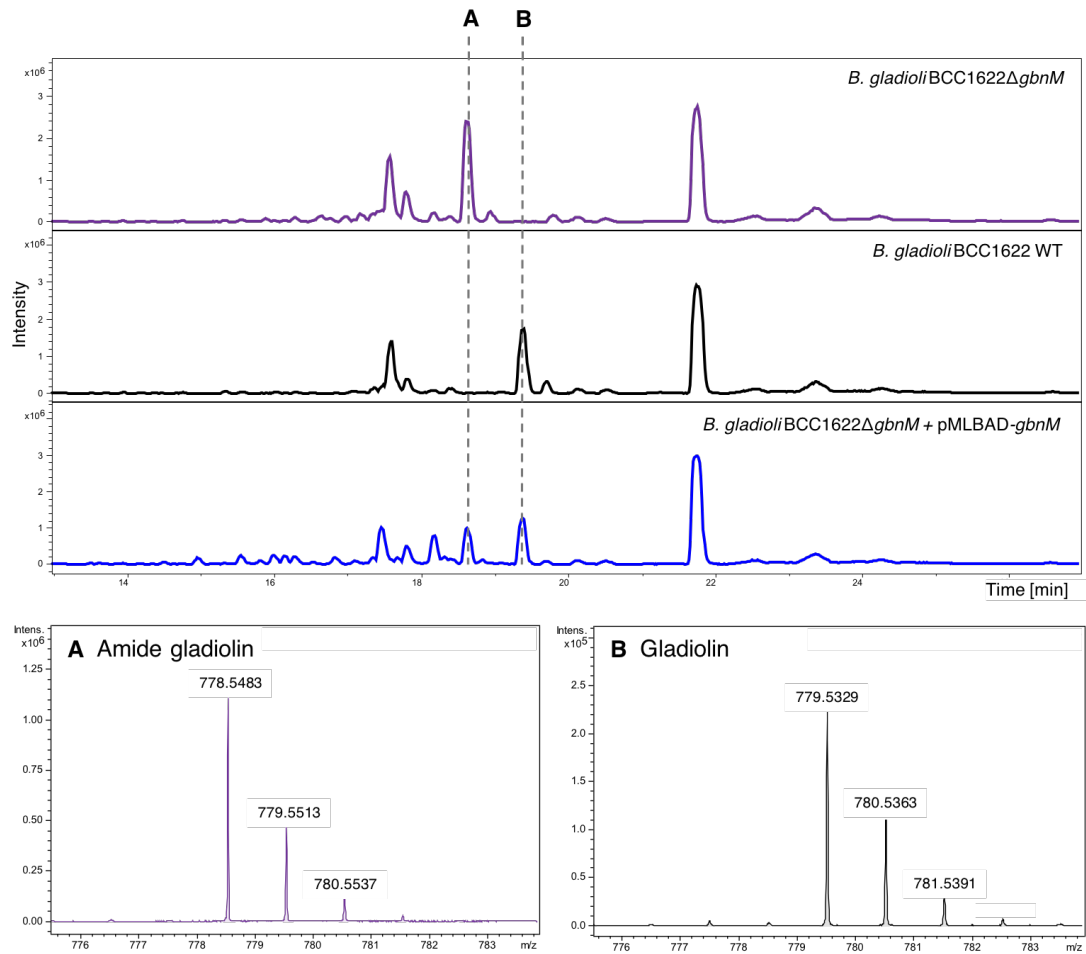


Figure 3.6: Base peak chromatograms from LC-MS analyses of extracts from agar-grown cultures of *B. gladioli* BCC1622 WT (black), *B. gladioli* BCC1622 Δ *gbnM* (purple) and *B. gladioli* BCC1622 Δ *gbnM* following *in trans* expression of *gbnM* (blue). The *gbnM* deletion mutant produced an amide derivative of gladiolin for which the high-resolution MS data is shown, alongside the MS data of gladiolin.

3.3.3 Isolation and structure elucidation of amide gladiolin

This work is in collaboration with Dr Yousef Dashti, University of Warwick, who assisted in performing and assigning the NMR of amide gladiolin and who also assigned the NMR of iso-amide gladiolin.

To isolate amide gladiolin for characterisation, *B. gladioli* BCC1622 Δ *gbnM* was grown on BSM agar (1 L) for 2.5 days. The crude ethyl acetate extract was purified by reverse-phase high-performance liquid chromatography (HPLC) yielding amide gladiolin (14 mg) as a white solid (Appendix 5). Gladiolin was also purified from a culture of *B. gladioli* BCC1622 WT, which was used for comparison during this study.

High resolution MS analysis of amide gladiolin showed a peak corresponding to $[M+H]^+$ at $m/z = 778.5483$ that was consistent with the predicted molecular formula $C_{44}H_{75}NO_{10}$ (calculated mass for $C_{44}H_{76}NO_{10}^+$: 778.5464, Appendix 6). The NMR assignments for amide gladiolin were very similar to that reported for gladiolin⁶⁵, with distinct differences seen only for chemical shifts related to the amide moiety and the adjacent methylene group (Table 3.1, the NMR spectra can be found in Appendix 7). The proton spectrum revealed two distinctive exchangeable amide protons at δ_H 6.65 and 7.22, and in the HMBC spectrum both of these protons showed $^2J_{CH}$ coupling to the carbon of the side chain carbonyl (δ_C 174.7). In addition, the proton at δ_H 6.65 showed $^3J_{CH}$ coupling to the carbon of the expected adjoining methylene group (δ_C 32.93), thus confirming that amidation occurs on the carboxylic acid moiety of gladiolin (Figure 3.7). Similarity of the rest of the chemical shift assignments to gladiolin established that deletion of *gbnM* caused no other structural changes to the molecule.

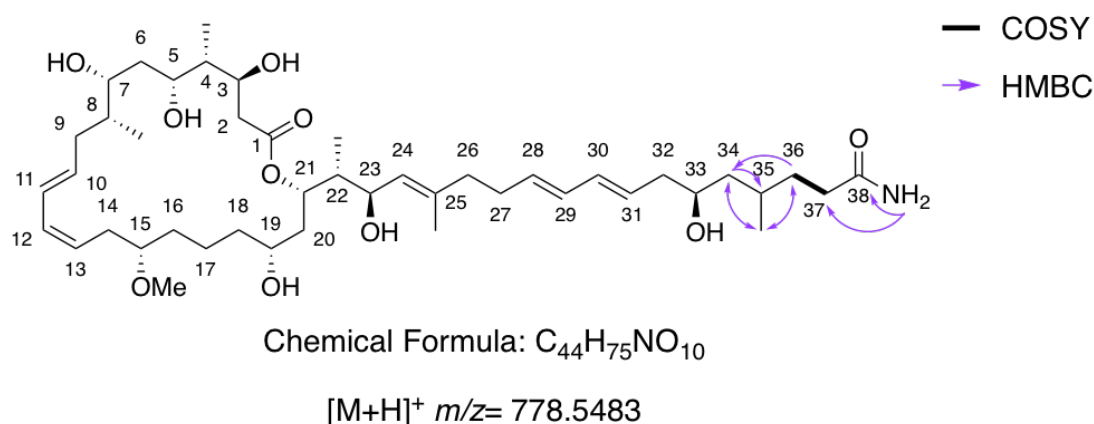


Figure 3.7: Chemical structure of amide gladiolin showing the key HMBC and COSY correlation confirming the position of the amide moiety.

Table 3.1: NMR assignments for amide gladiolin (DMSO- d_6 , 600 Hz) compared to those published for gladiolin⁶⁵. * indicates that the signal was found under the DMSO- d_6 peak, and ** that the signal was found under the water peak. Highlighted in blue are the chemical shift assignments that differ most significantly. Refer to figure 3.7 for the positions. Multiplicity is denoted as follows: s = singlet, d = doublet, dd = doublet of doublets, dt = doublet of triplets, br = broad, m = multiplet and ol = overlapped.

Amide gladiolin			Gladiolin	
Position	δ_H (J, Hz)	δ_C	δ_H (J, Hz)	δ_C
1	-	171.3	-	170.9
2	2.22, 2.40, 2 x 1H, 2 x m	39.6*	2.22, 2.40, 2 x 1H, 2 x m	39.6
3	3.94, 1H, m	69.3	3.95, 1H, m	68.9
3-OH	4.69, 1H, d, (5.6)	-	4.05, ol	-
4	1.46, 1H, m	43.7	1.46, 1H, m	43.8
4-Me	0.79, 3H, d, (6.4)	9.7	0.79, 3H, d, (6.5)	9.6
5	3.72, 1H, m	69.1	3.71, 1H, m	68.8
5-OH	4.35, 1H, ol	-	4.02, ol	-
6	1.46, 2H, m, ol	39.2*	1.46, 2H, m	39.6
7	3.54, 1H, m	70.7	3.54, 1H, m	70.6
7-OH	4.36, 1H, ol	-	4.02, 1H, br s	-

8	1.62, 1H, m	36.5	1.62, 1H, m	37.0
8-Me	0.79, 3H, d, (6.7)	13.6	0.78, 3H, d, (6.0)	13.3
9	1.94, 2.18, 2 x 1H, 2 x m	35.9	1.92, 2.18, 2 x 1H, 2 x m	35.8
10	5.71, 1H, dt, (6.6, 14.5)	133.0	5.71, 1H, dt, (6.5, 15.0)	133.0
11	6.29, 1H, dd, (10.96, 15.00)	126.5	6.29, 1H, dd, (11.0, 15.0)	126.1
12	6.01, 1H, ol	130.0	6.02, 1H, (11.0, 11.0)	129.7
13	5.31, 1H, ol	125.4	5.31, 1H, dt, (7.0, 11.0)	125.0
14	2.26, 2.36, 2 x 1H, 2 x m	30.6	2.26, 2.38, 2 x 1H, 2 x m	30.0
15	3.18, 1H, m	80.1	3.18, 1H, dd, (5.5, 5.5)	79.6
15-OMe	3.23, 3H, s	55.8	3.23, 3H, s	55.5
16	1.31, 1.46, 2 x 1H, 2 x m	33.2	1.32, 1.48, 2 x 1H, 2 x m	32.2
17	1.21, 1.36, 2 x 1H, 2 x m	21.2	1.23, 1.36, 2 x 1H, 2 x m	21.1
18	1.24, 1.28, 2 x 1H, m, ol	38.3	1.32, 1.30, 2 x 1H, 2 x m	36.2
19	3.38**	66.4	3.38, 1H, m	66.1
19-OH	4.17, 1H, d, (5.9)	-	3.54, 1H, ol	-
20	1.50, 2H, m	36.4	1.49, 1.61, 2 x 1H, 2 x m	36.3
21	5.30, 1H, m, ol	72.0	5.30, 1H, m	71.8
22	1.80, 1H, m	42.2	1.80, 1H, m	42.1
22-Me	0.67, 3H, d, (7.03)	10.2	0.69, 3H, d, (6.5)	10.4
23	3.99, 1H, m	68.1	3.98, 1H, m	67.8
23-OH	4.66, 1H, d, (4.03)	-	4.46, 1H, br s	-
24	5.06, 1H, d, (8.90)	127.9	5.06, 1H, d, (8.5)	127.6
25	-	135.5	-	134.7
25-Me	1.59, 3H, s	16.5	1.59, 3H, s	16.2
26	2.02, 2H, ol	39.1*	2.02, 2H, m	38.9
27	2.14, 2H, m	30.5	2.12, 2H, m	30.1

28	5.54, 1H, ol	131.6	5.54, 1H, dt, (7.0, 15.0)	130.2
29	5.98, 1H, m, ol	130.6	5.98, 1H, dd, (11.0, 15.0)	131.1
30	5.95, 1H, ol	131.7	5.96, 1H, dd, (11.0, 15.0)	131.9
31	5.57, 1H, ol	129.4	5.56, 1H, dt, (7.0, 15.0)	129.2
32	2.08, 2H, m	41.6	2.08, 2H, m	41.3
33	3.99, 1H, m	67.6	3.49, 1H, m	67.1
33-OH	4.37, 1H, ol	-	3.80, 1H, br s	-
34	1.04, 1.28, 2 x 1H, 2 x m	44.0	1.04, 1.28, 2 x 1H, 2 x m	43.6
35	1.61, 1H, m, ol	28.4	1.61, 1H, m	28.1
35-Me	0.80, 3H, d, (6.3)	18.9	0.80, 3H, d, (6.5)	17.9
36	2.02, 2H, m, ol	32.5	1.32, 1.47, 2 x 1H, 2 x m	32.3
37	1.29, 1.46, 2 x 1H, 2 x m	32.9	2.14, 2.19, 2 x 1H, 2 x m	29.9
38	-	174.7	-	174.9
NH₂	6.65, 7.22, 2 x 1H, 2 x br s	-	-	-

3.3.4 Stability of amide gladiolin in protic solvents

Gladiolin was found to undergo slow conversion to a 24-membered lactone isomer in protic solvents such as methanol. This isomerisation occurred *via* the nucleophilic addition of the C-23 hydroxyl group to the ester bond followed by elimination of the C-21 hydroxyl group⁶⁵ (Figure 3.8 (A)). This conversion was also seen for amide gladiolin, generating *iso*-amide gladiolin (Figure 3.8 (A), characterisation of which can be found in Appendix 8). The rate of this conversion appeared to be faster for amide gladiolin (Figure 3.8 (B)); the reason for this is not apparent but it is possible that the presence of the amide moiety changes the conformation of the molecule in solution, bringing the C-23 hydroxyl group closer to the ester bond. This indicates that amide

hydrolysis, catalysed by GbnM, is responsible for generating a comparatively more stable antibiotic.

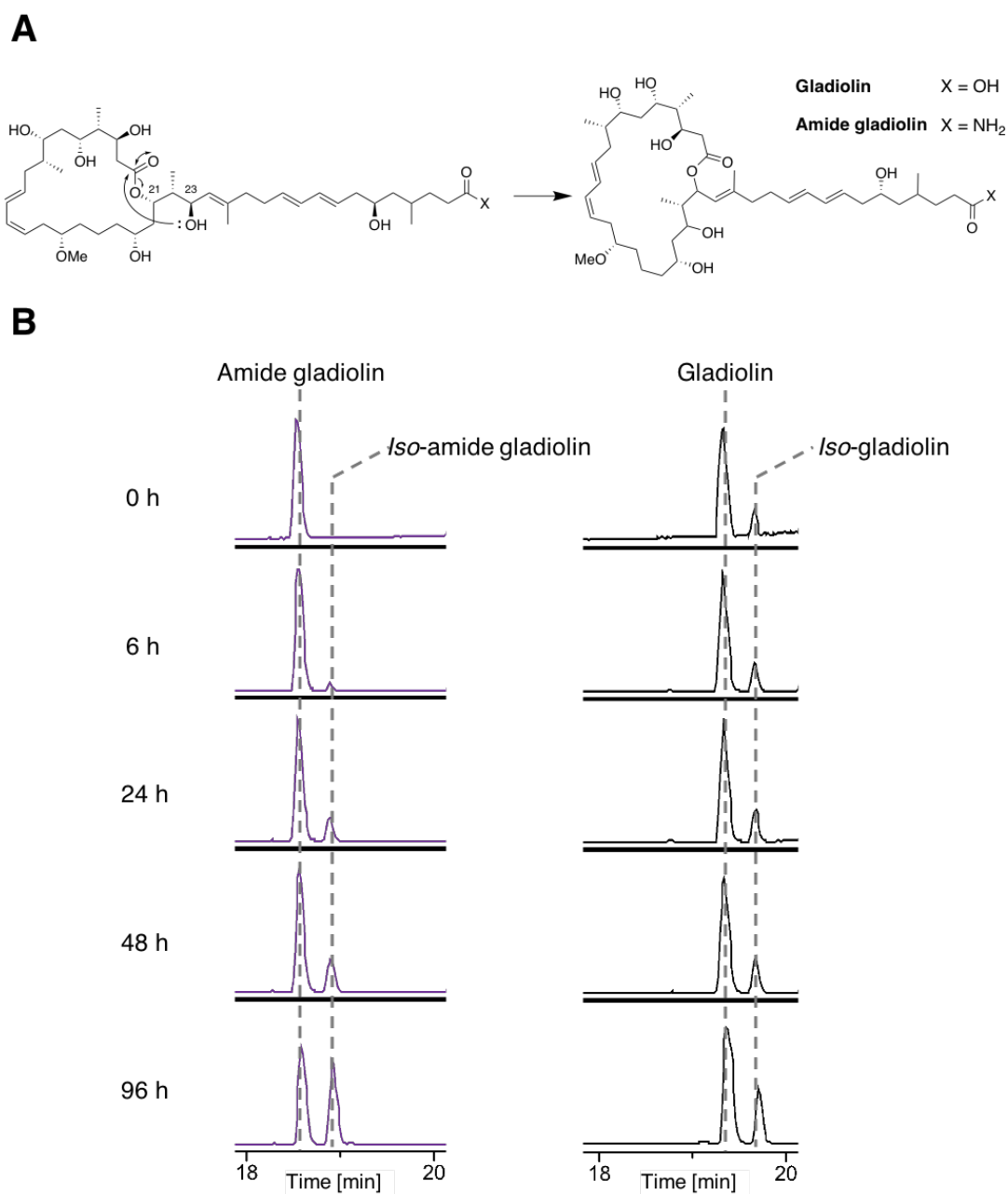


Figure 3.8: (A) The conversion of amide gladiolin to *iso*-amide gladiolin. (B) Base peak chromatograms from the LC-MS time-course analysis showing the stability of amide gladiolin and gladiolin in methanol.

3.3.5 *In vitro* studies

To further verify the activity of GbnM, complementary *in vitro* studies were conducted. *GbnM* was cloned into the vector pET28a for protein expression, however the amidase was poorly overproduced in *E. coli* and was almost exclusively insoluble. Secondary structure prediction by Phyre²⁸⁹ suggested that the amidase may contain a *trans*-membrane helical domain which could explain the difficulty in overproducing soluble protein (the same prediction was also made for the homologues from etnangien and sorangicin biosynthesis); membrane association would not be unexpected should this protein play a role in self-protection. In an attempt to overcome this buffers containing the detergent Triton X-100 or CHAPS were trialled during protein extraction, as detergents have often proved useful in the solubilisation of membrane associated proteins¹²⁷, however this was not successful. For this reason the assay was conducted using cell lysate. The presence of GbnM in the cell lysate was confirmed by SDS-PAGE analysis (Figure 3.9 (B)).

In vitro activity of GbnM was demonstrated by incubating purified amide gladiolin (see section 3.3.3 for details on the purification) with cell lysate containing overproduced GbnM (Figure 3.9 (A)). As expected, LC-MS analysis revealed the appearance of a product peak with an identical mass and retention time to the gladiolin standard (Figure 3.9 (C)). In order to ensure that the observed activity was not the result of any other proteins present in the cell lysate a control experiment was conducted using cell lysate from *E. coli* BL21 containing the empty pET28a vector. Conversion to gladiolin was not detected, thus verifying that GbnM was responsible for the observed conversion.

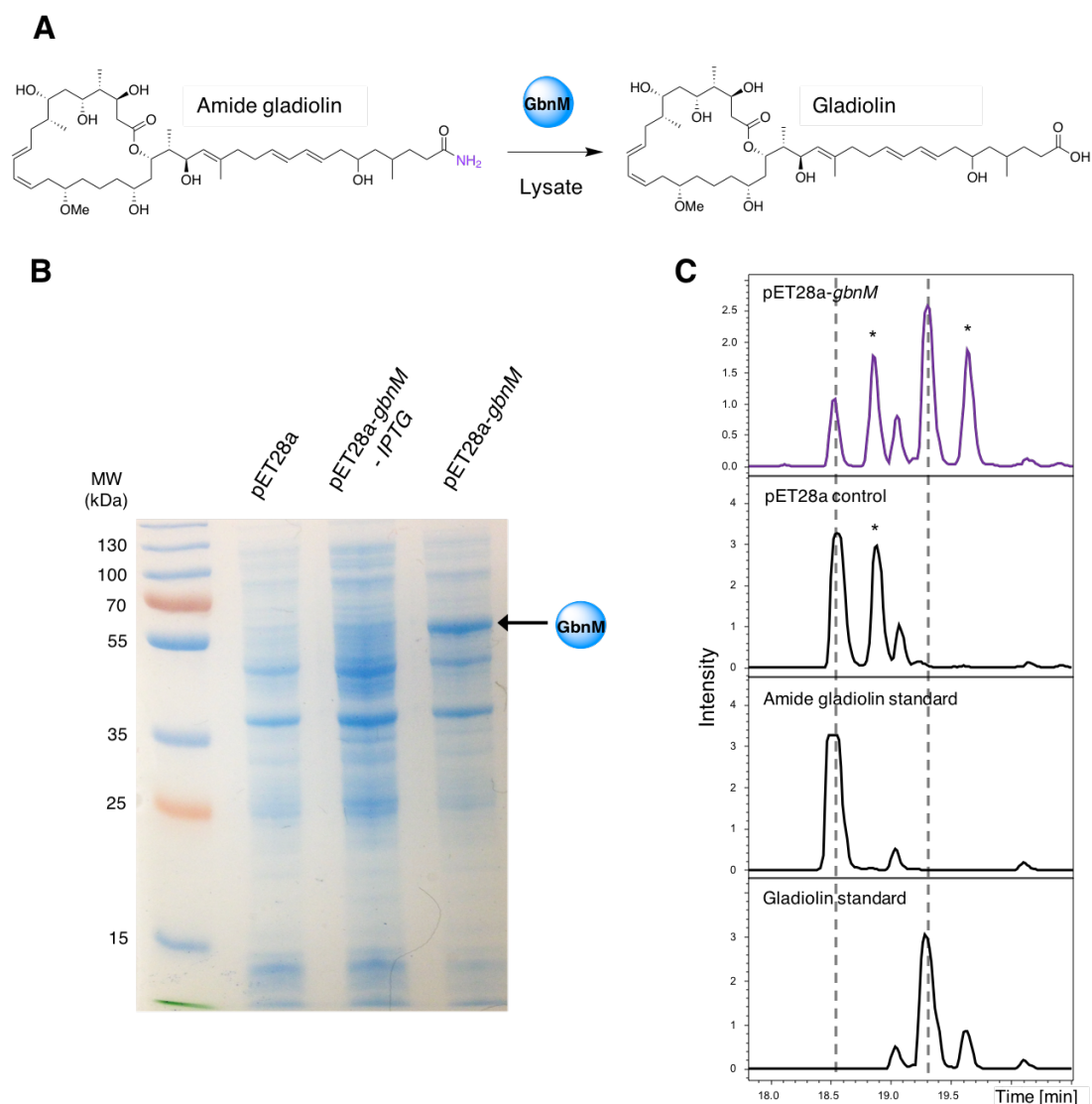


Figure 3.9: (A) Overview of the assay used to demonstrate activity of GbnM *in vitro*. (B) 10 % SDS-PAGE analysis showing the presence of GbnM in the cell lysate. (C) Base peak chromatograms from LC-MS analysis of the assay showing the conversion of amide gladiolin to gladiolin by GbnM, the peaks marked with * result from isomerisation.

3.4 Antimicrobial activity of amide gladiolin

Purified amide gladiolin was tested for its antimicrobial activity against a range of clinically relevant pathogens, which included representative members of the ESKAPE panel. The ESKAPE panel consists of bacteria that demonstrate concerning levels of antibiotic resistance¹²⁸. The minimum inhibitory

concentrations (MICs) were determined using the broth microdilution method (Table 3.2).

Table 3.2: MIC values determined for amide gladiolin against *C. albicans* and a range of gram-negative and gram-positive bacteria. The MIC values determined for gladiolin in this study are also shown as a comparison.

Test organism	Amide gladiolin MIC ($\mu\text{g/mL}$)	Gladiolin MIC ($\mu\text{g/mL}$)
Gram-negative bacteria		
<i>Klebsiella pneumonia</i> DSM26371	>64	>64
<i>Acinetobacter baumannii</i> DSM25645	8	32
<i>Pseudomonas aeruginosa</i> DSM29239	64	>64
<i>Enterobacter cloacae</i> DSM16690	64	>64
<i>Escherichia coli</i> SY327	16	>64
<i>Burkholderia gladioli</i> BCC1622	>64	>64
Gram-positive bacteria		
<i>Enterococcus faecium</i> DSM25390	2	2
<i>Staphylococcus aureus</i> DSM21979	4	8
Fungi		
<i>Candida albicans</i> SC 5314	2	4

The MIC assays revealed that amide gladiolin was moderately active against gram-positive bacteria and *C. albicans*. Interestingly, amide gladiolin, unlike gladiolin, demonstrated moderate antimicrobial activity against gram-negative bacteria, including *A. baumannii* and *E. coli*. This suggests that amidation of gladiolin improved the permeative ability of the antibiotic to the often impenetrable outer membrane of these bacteria. Amide gladiolin was also tested against *B. gladioli* BCC1622, the producer of the gladiolins, in an attempt to see if it displayed resistance, however no activity was observed for either amide gladiolin or gladiolin. Presumably, the outer membrane in *B. gladioli* acts as a barrier to the influx of the antibiotics.

As gladiolin demonstrated potent antimicrobial activity against *M. tuberculosis*⁶⁵, and the structurally related etnangien was found to be active against a range of other *Mycobacterium* species including *M. smegmatis*⁶⁶, it was thus decided to evaluate the *in-vitro* activity of amide gladiolin against *M. smegmatis*. *M. smegmatis* is often used as an *in vitro* model for determining if a compound is likely to demonstrate activity towards *M. tuberculosis* owing to its much shorter generation time and the fact that it is considered to be non-pathogenic¹²⁹. The resazurin microtiter assay was used, which is a simple and inexpensive technique for determining the MICs of *Mycobacteria* using the cell permeable redox indicator resazurin. Viable cells with an active metabolism will reduce resazurin, which is blue, to resorufin, which is pink (Figure 3.10)¹³⁰. The MIC was defined as the lowest antibiotic concentration that prevented the colour change from blue to pink. It was found that amide gladiolin was slightly more active than gladiolin against *M. smegmatis*, 16 µg/mL compared to 32 µg/mL respectively (Table 3.3). It was thus tested for activity against *M. tuberculosis*, which was conducted by the Fullam Lab (School of Life Sciences, University of Warwick) as part of a TB screening initiative funded by INTERGRATE EPSRC (grant number EP/M027503/1). The MIC for *M. tuberculosis* H37Rv was determined to be 2 µg/mL.

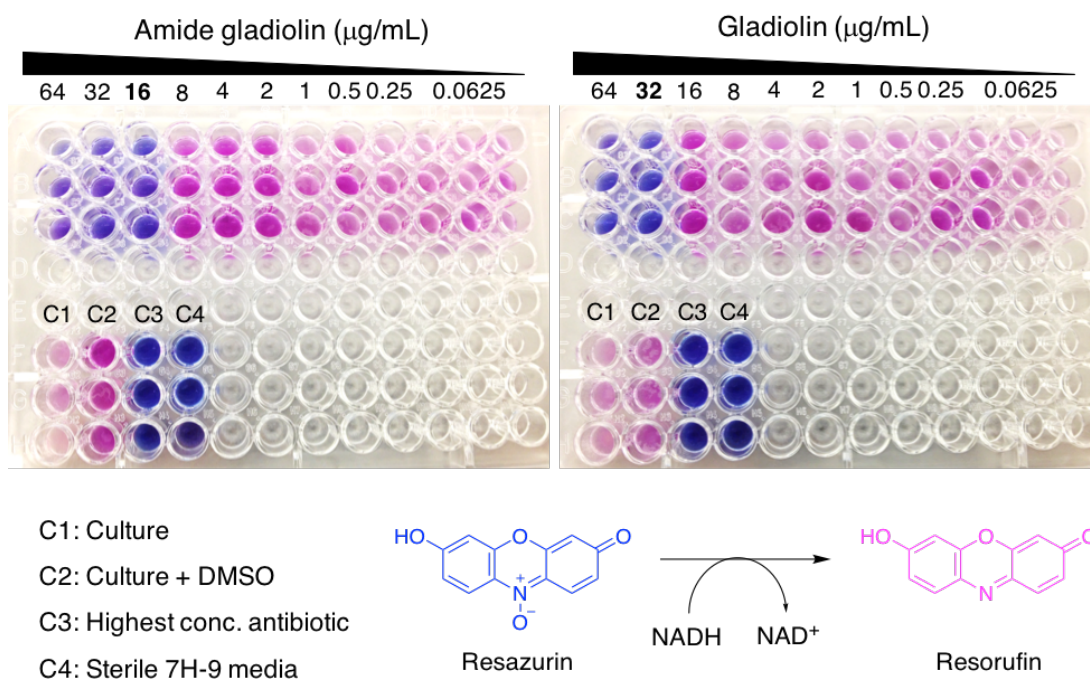


Figure 3.10: The rezazurin microtiter assay used to determine the MIC of amide gladiolin and gladiolin for *M. smegmatis*.

Table 3.3: MIC values determined for amide gladiolin against *M. smegmatis* and *M. tuberculosis*. The value shown in blue is the value reported by Song *et al.* and was not determined as part of this study⁶⁵.

Test organism	Amide gladiolin MIC (μg/mL)	Gladiolin MIC (μg/mL)
<i>Mycobacterium smegmatis</i> DSM43756	16	32
<i>Mycobacterium tuberculosis</i> H37Rv	2	0.4 ⁶⁵

The antimicrobial activity of amide gladiolin suggests that this antibiotic is still capable of efficiently inhibiting RNA polymerase, although this will need to be experimentally investigated to confirm the target for the compound was not affected by the amidation. It was reported that for sorangicin, this was not the case as a chemically synthesised amide derivative maintained the ability to inhibit *E. coli* RNA polymerase¹³¹. This suggests that amidation alone is not able to confer resistance. Interestingly, a crystal structure of *Thermus*

aquaticus RNA polymerase in complex with sorangicin revealed that the side chain was not involved in RNA polymerase interactions and that the carbonyl of the carboxylic acid moiety only formed a hydrogen bond with the macrolide core, possibly to maintain a certain molecular conformation, which an amide would also be capable of (Figure 3.11)¹. This calls into question how the RNA polymerase of a producing organism could possess an adaption that provided resistance to the amide derivative and not the mature antibiotic, and thus whether the role of amidation of these antibiotics during biosynthesis is in fact self-resistance.

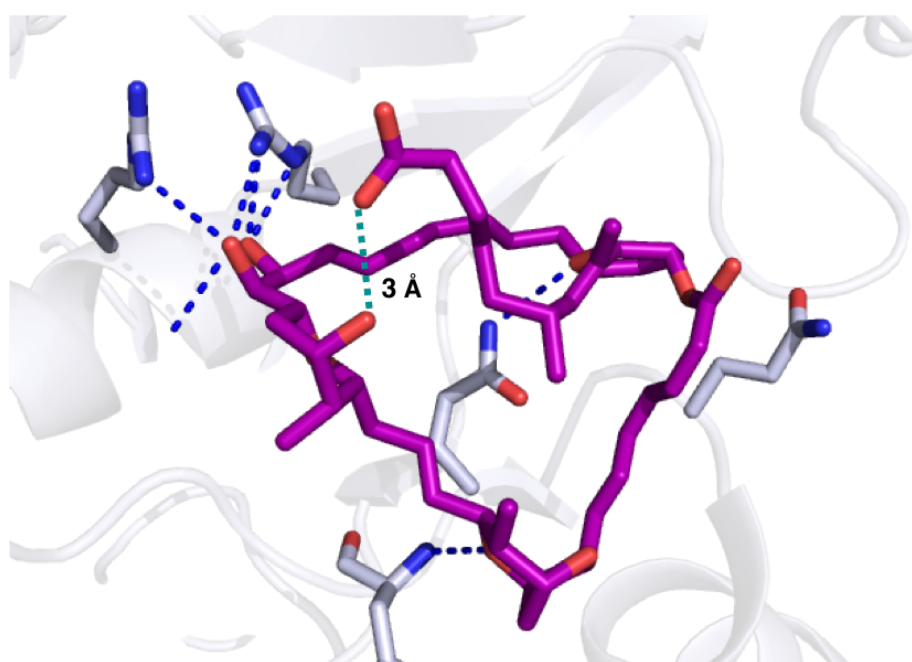


Figure 3.11: Image of sorangicin (purple) in complex with *Thermus aquaticus* RNA polymerase (PDB: 1YNJ¹). Hydrogen bonds are shown with dashed lines with those to the protein shown in blue, and the intramolecular hydrogen bond between the carboxylic acid moiety and the macrolide core shown in teal. This image was generated using PyMOL¹⁰⁰.

3.5 Investigating the action of *gbc*

3.5.1 *In vivo* studies

To investigate the role of the asparagine synthetase, *gbc*, in the biosynthesis of gladiolin the gene was inactivated through the creation of an in-frame deletion using the method discussed in section 3.3.1, and a successful mutant was identified by PCR analysis (Figure 3.12). High-resolution LC-MS analysis of the extract of an agar culture of *B. gladioli* BCC1622 Δ *gbc* revealed that production of gladiolin was completely abolished (Figure 3.13). The deletion mutant was complemented by *in trans* expression of *gbc* which restored gladiolin production to a limited extent. A better level of complementation could not be achieved even with increasing the concentration of arabinose, which induces expression, in the media (Figure 3.13 (B)). This could be due to the fact that arabinose offers the bacteria an alternative carbon source, leading to a shift in the secondary metabolite profile, which could disfavoured gladiolin production. This result established that *gbc*, plays an essential role in the biosynthesis of gladiolin.

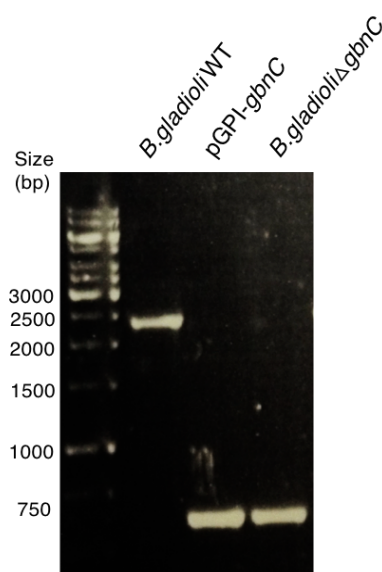


Figure 3.12: Agarose gel electrophoresis showing genetic evidence for the creation of a *gbc* deletion mutant in *B. gladioli* BCC1622.

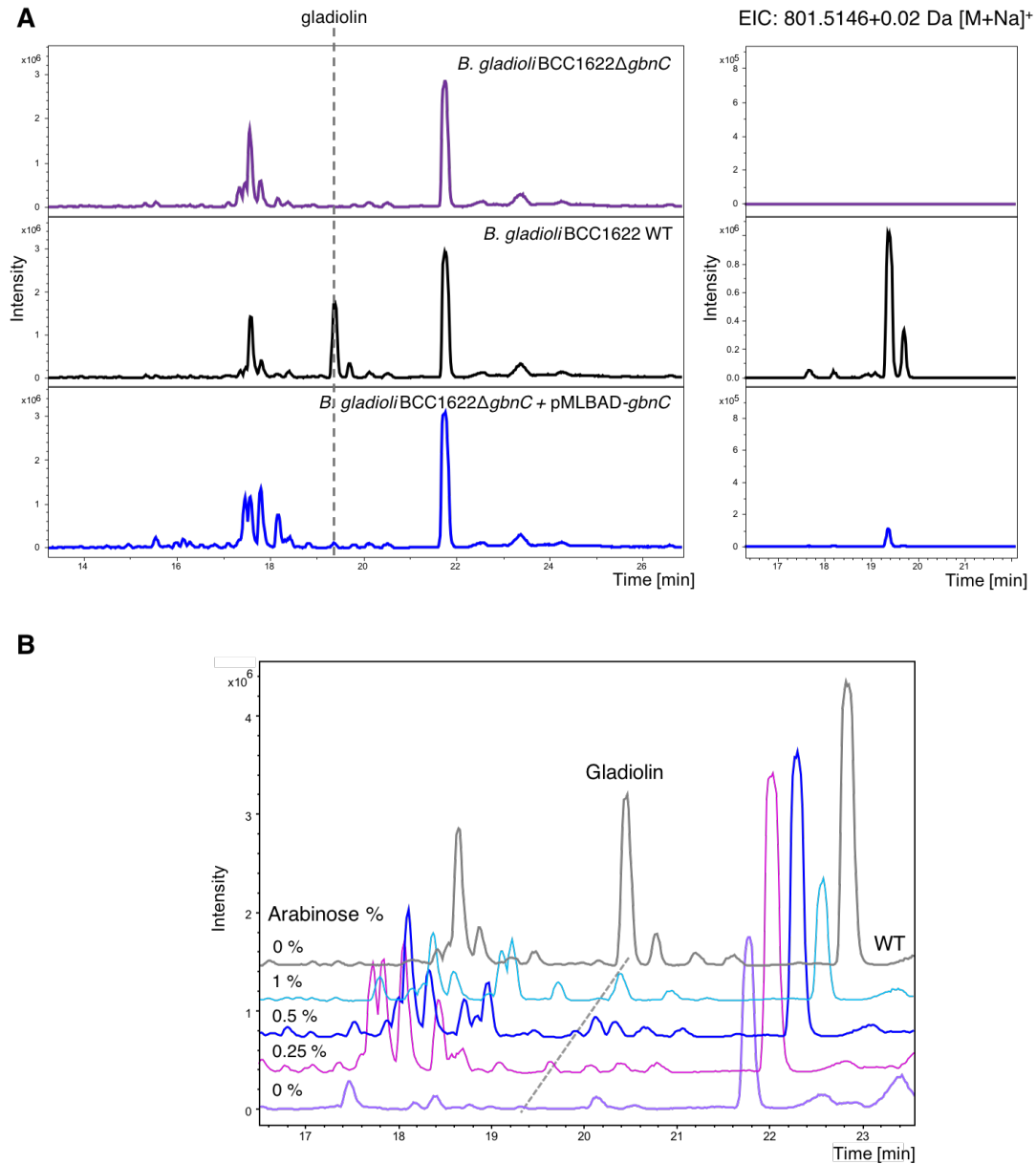


Figure 3.13: (A) Base peak chromatograms from LC-MS analyses of extracts from agar-grown cultures of *B. gladioli* BCC1622 WT (black), *B. gladioli* BCC1622Δ*gbcN* (purple) and *B. gladioli* BCC1622Δ*gbcN* following *in trans* expression of GbcN (blue). The extracted ion chromatogram for the [M+Na]⁺ ion is also shown. (B) Base peak chromatograms from LC-MS analyses of extracts from cultures of *B. gladioli* BCC1622Δ*gbcN* where the media was supplemented with increasing concentrations of arabinose (blue). For comparison the WT is shown in grey.

3.5.2 Complementation with the homolog SorQ

Attempts to overproduce the asparagine synthetase, GbnC, as a hexahistidine-tagged recombinant protein were unsuccessful; the resultant protein was completely insoluble, and the solubility of the protein could not be improved, even with the inclusion of the small ubiquitin-like modifier (SUMO) or maltose-binding protein (MBP) solubility tag, which have been reported to enhance the overproduction, stability and solubility of numerous proteins¹³². The same insolubility was also found for SorQ, a homologue of GbnC, from the sorangicin biosynthetic pathway, that shares 50% amino acid sequence identity (Figure 3.3 (B)).

As *in vitro* studies with GbnC were not possible, complementation of *B. gladioli* BCC1622 Δ *gbnC* by *in trans* expression of *sorQ* was attempted. The structure of sorangicin and gladiolin are largely different but intermediates in polyketide chain initiation are likely to be similar. The gladiolin starter unit is proposed to be succinyl-CoA and for sorangicin it is hypothesised to be malonyl-CoA. If complementation could be achieved this would suggest that the asparagine synthetase is likely to act pre-polyketide assembly. Unfortunately, complementation was not observed (Figure 3.14). The reason for this could be that *B. gladioli* was unable to produce SorQ as a functional and soluble protein given that its origin is from a very different organism, or that succinyl-CoA, ACP bound succinyl derivatives or gladiolin were not substrates for this protein.

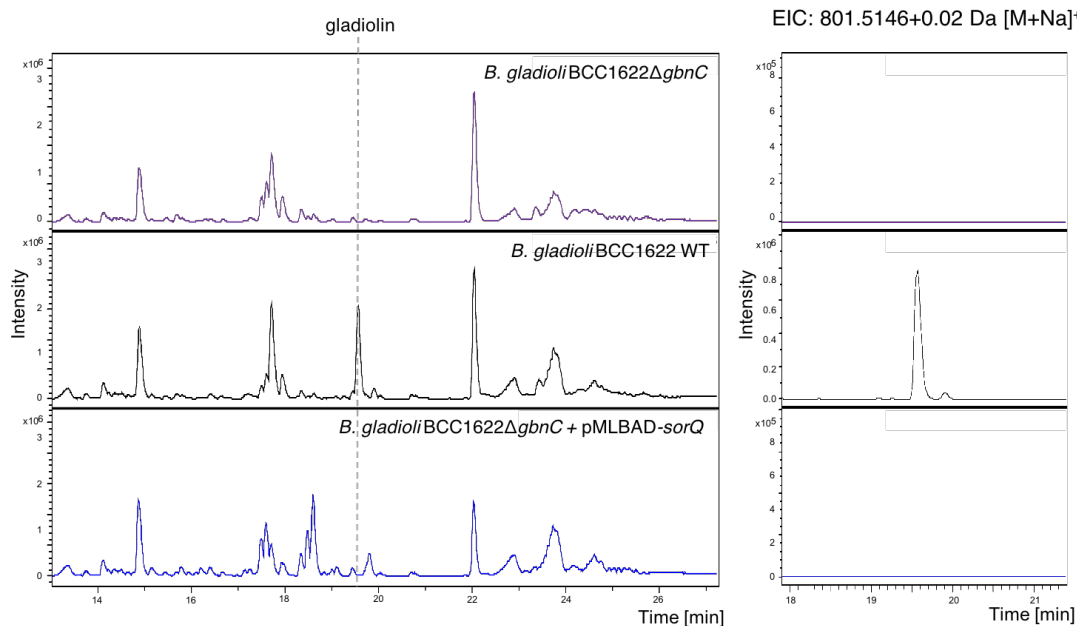


Figure 3.14: Base peak chromatograms from LC-MS analyses of extracts from agar-grown cultures of *B. gladioli* BCC1622 WT (black), *B. gladioli* BCC1622Δ*gbcC* (purple) and *B. gladioli* BCC1622Δ*gbcC* following *in trans* expression of *sorQ* (blue). The extracted ion chromatogram for the $[M+Na]^+$ ion is also shown.

3.6 Exploration of polyketide chain initiation in the biosynthesis of gladiolin

3.6.1 Proposed mechanisms of chain initiation

As the action of the asparagine synthetase, GbnC, could not be investigated directly the cryptic initiation of gladiolin biosynthesis was studied, in hope of revealing whether amidation is necessary for initiation of polyketide biosynthesis. The gladiolin PKS lacks a conventional clearly demarcated loading module, however, there is a standalone ACP domain, GbnA, that has yet to be accounted for in the biosynthesis of gladiolin, which could act as the loading module. The gladiolin BCG cluster was also found to encode for an AT-like enzyme, GbnB, which clades independently to malonyl-CoA specific

AT domains (Appendix 9). As KirCII, from the kirromycin pathway, which demonstrates selectivity for ethyl-malonyl-CoA *in vitro*⁴⁷ and OzmC, from the oxazolomycin pathway, that is proposed to utilise methoxymalonyl-CoA⁴⁹, also clade separately, it is proposed that GbnB will show selectivity towards the starter unit, succinyl-CoA. Homologs of both GbnA and GbnB can be found in the etnangien BGC. It is therefore hypothesised that GbnB loads succinyl-CoA onto the ACP domain, GbnA, and the succinyl moiety is subsequently transferred from GbnA to the KS domain of GbnD1, from which polyketide chain assembly can proceed (Figure 3.16 (A)). A similar mechanism to this is proposed for the chain initiation in the biosynthesis of the glutarimide-containing polyketides (Figure 3.15)⁵⁰. Amidation (also of a terminal carboxylate moiety) must play a role at some stage during early biosynthesis, as the domains that have been shown to install the glutarimide moiety can always be found in the second non-elongating part of module 1 of the PKS^{133,134}. Amidation, however, has yet to be biochemically characterised in these PKSs.

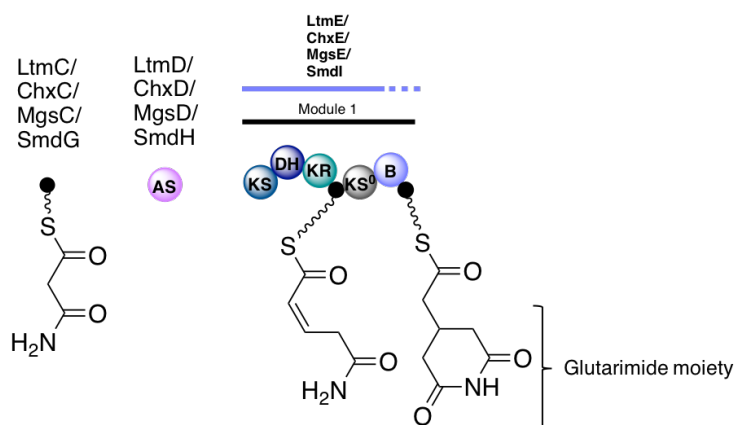


Figure 3.15: Proposed chain initiation mechanism in the biosynthesis of the glutarimide-containing polyketides. AS = asparagine synthetase homolog. This proposed initiation mechanism can be found in the lactimidomycin (Ltm) PKS¹³⁵, cycloheximide (Chx) PKS¹³⁶, *iso*-migrastatin (Mgs) PKS¹³⁴ and the 9-methyl-streptimidone (Smd) PKS¹³⁷.

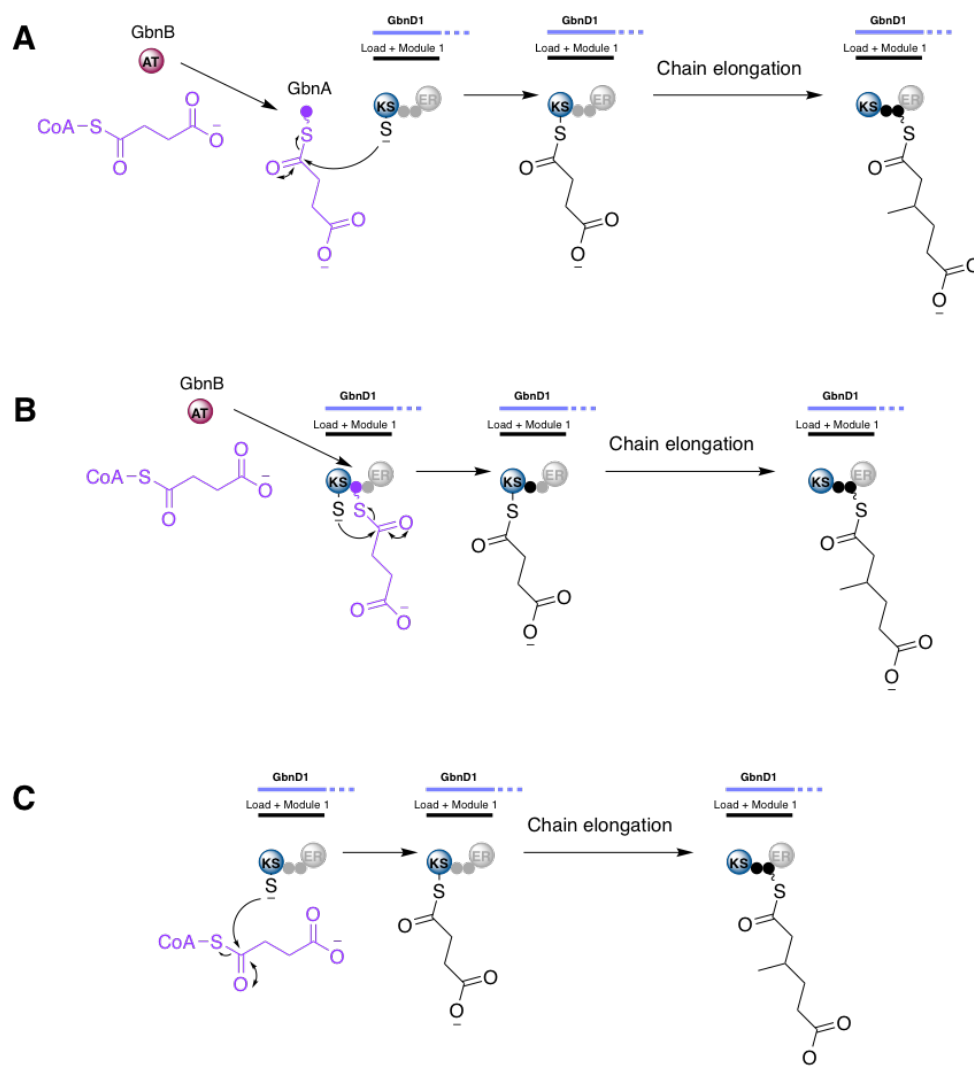


Figure 3.16: Proposed alternative mechanisms of polyketide chain initiation in the biosynthesis of gladiolin. Potential substrates for the asparagine synthetase (GbnC) are shown in purple. In mechanism (A) the AT domain (GbnB) loads succinyl-CoA onto the standalone ACP domain (GbnA), prior to the succinyl moiety being transferred to the KS domain of GbnD1. In mechanism (B) GbnB loads succinyl-CoA directly onto the first ACP domain of GbnD1 and the starter unit is then back transferred onto the KS domain so that chain elongation can occur. In mechanism (C) the first KS domain of GbnD1 directly accepts succinyl-CoA as a substrate.

It is also, however possible that succinyl-CoA is directly loaded onto the first ACP domain of GbnD1 by GbnB, and subsequently back transferred onto the active site of the KS domain so that chain elongation can occur (Figure 3.16 (B)), or that neither of these enzymes are required and the KS domain of

GbnD1 can directly accept succinyl-CoA (Figure 3.16 (C)). Self-loading of the starter unit onto the first KS domain has been demonstrated for the erythromycin pathway, where deletion of the loading module did not terminate biosynthesis, although a notable drop in the biosynthetic efficiency was observed¹³⁸.

Should amidation of the carboxylic acid moiety act as a biosynthetic protecting group it would be anticipated that the asparagine synthetase, GbnC, would act either upon Succinyl-CoA, creating a pool of substrate for the PKS, or on the ACP bound succinyl moiety which would be vulnerable to cyclisation (Figure 3.2 (B)).

3.6.2 Establishing the importance of *gbnA* and *gbnB* in the biosynthesis of gladiolin

To confirm the involvement of the ACP domain (GbnA) and the AT domain (GbnB) in the biosynthesis of gladiolin, gene deletion mutants were created using the method discussed in section 3.3.1. Mutants with the correct genotype were identified by PCR analysis (Figure 3.17 (A)). Both of these mutants lost the ability to produce gladiolin, indicating that both GbnA and GbnB play an essential role in the biosynthesis (Figure 3.17 (B)). To fully confirm this complementation by *in trans* expression of the deleted genes is ongoing.

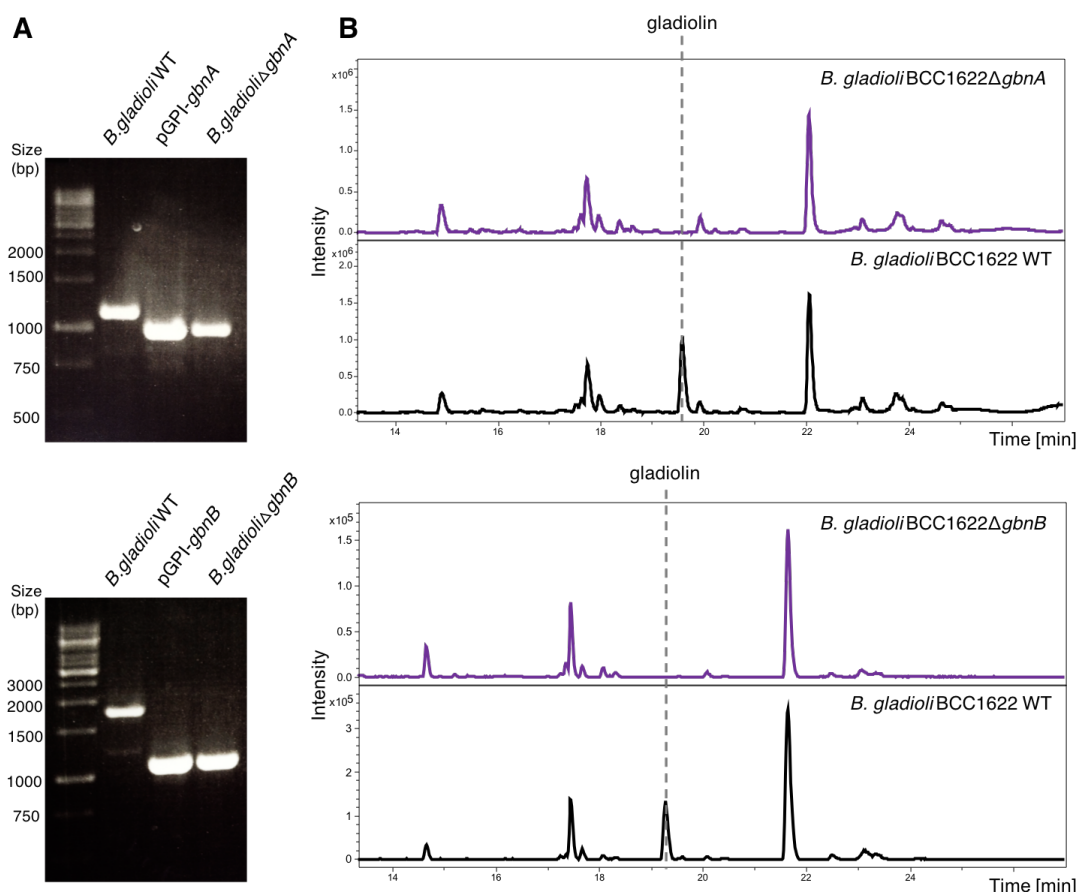


Figure 3.17: (A) Agarose gel electrophoresis showing genetic evidence for the creation of a *gbnA* and *gbnB* deletion mutant in *B. gladioli* BCC1622. (B) Base peak chromatograms from LC-MS analyses of extracts from agar-grown cultures of *B. gladioli* BCC1622 WT (black), and the generated mutants (purple). In both deletion mutants the production of gladiolin was abolished.

3.6.3 Overproduction and purification of proteins involved in chain initiation

In order to test the hypothesised mechanism for chain initiation *in vitro*, GbnA, GbnB, and the first ACP domain from the gladiolin biosynthetic pathway, GldM1ACP(I), were heterologously overproduced as hexa-histidine tagged proteins, which were purified by nickel-NTA chromatography. This was achieved by cloning the gene, or section of DNA, of interest into the expression vector pET28a as discussed in section 2.4.1. The malonyl-specific AT domain, EtnK, was also overproduced and purified, from a synthetic pET24a construct.

This was used for comparison with GbnB as the malonyl-specific AT domain for the gladiolin pathway, GbnN, did not give soluble protein when overproduced in *E. coli*. A second malonyl-specific AT domain was also used for comparison, PksC from the bacillaene biosynthetic pathway (the purified protein was kindly provided by Dr Matthew Jenner). This was used as the bacillaene pathway does not utilise succinyl-CoA at any stage during biosynthesis⁵³. SDS-PAGE analysis showed that the proteins were of a reasonable level of purity and of the expected molecular weight (Figure 3.18), which was confirmed by intact protein MS (Appendix 1).

An attempt was made to obtain the first KS domain from the gladiolin PKS, GldM1KS, as a purified protein, however overproduction yielded only insoluble protein. It is possible that the protein was insoluble due to disruption of the secondary structure. The chosen domain boundaries were based on predictions made by antiSMASH 2.0, and it is hoped that through trialling different domain boundaries a soluble protein could be obtained.

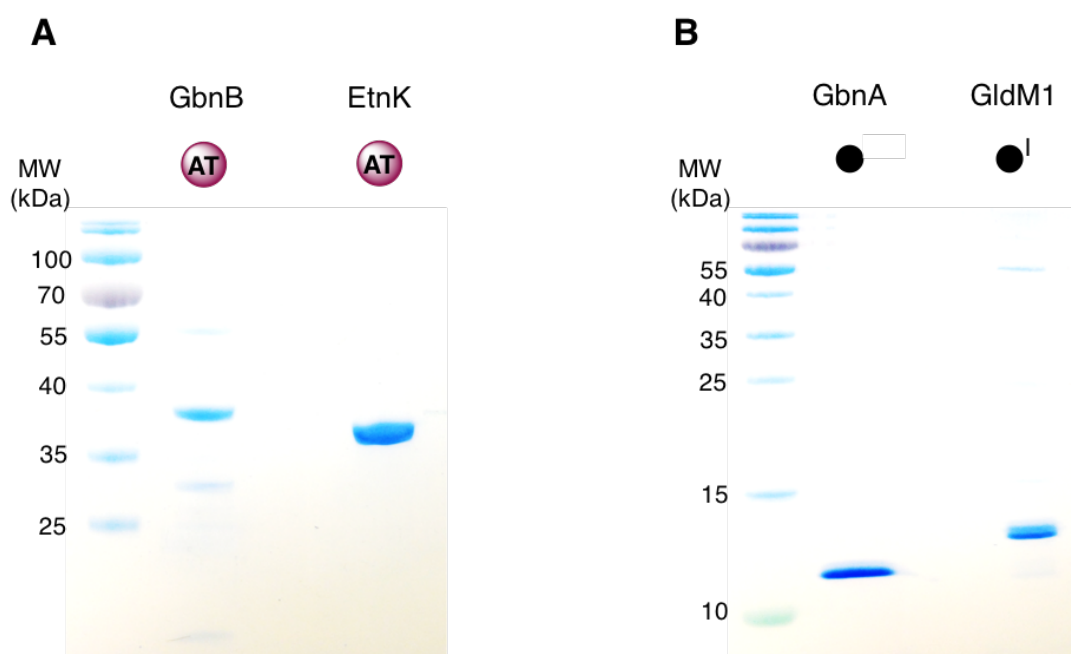


Figure 3.18: (A) 12 % SDS-PAGE analysis of the purified AT domains GbnB and EtnK. (B) 15 % SDS-PAGE analysis of the purified ACP domains GbnA and GldM1ACP(I).

3.6.4 Investigating the substrate tolerance of GbnB

The activity of the AT domain, GbnB, was tested by incubating it with succinyl-CoA. For comparison, the same assay was performed with the proposed malonyl-specific AT domains, EtnK and PksC, from the etnangien⁴⁸ and bacillaene biosynthetic pathway⁵³, respectively. Intact protein MS was used to analyse whether the succinyl-moiety could be loaded onto the active site serine residue of the AT domains. Interestingly, it was found that all three of the AT domains were able to accept succinyl-CoA, determined by an observed mass increase of 100 Da, although complete loading was observed only for GbnB (Figure 3.19). Malonyl-CoA was also tested as a potential substrate for GbnB, and complete loading of all three of the AT domains was observed.

The ability of GbnB to accept succinamic-CoA was next investigated. As succinamic-CoA was not commercially available and would be challenging to synthesise, a biosynthetic approach was chosen that utilised the class III CoA transferase, VbxP, from the vibroxin biosynthetic pathway, which was found to be capable of generating a wide scope of CoA analogues from the corresponding acids and acetyl-CoA (purified VbxP was kindly provided by Christian Hobson)¹³⁹. Overnight incubation of an (5X) excess of succinamic acid with VbxP and acetyl-CoA lead to the complete conversion of acetyl-CoA to succinamic-CoA (Figure 3.20). VbxP was removed from the reaction mixture using a 5 kDa MWCO spin filter and the resultant flow through was incubated with GbnB. It was discovered that succinamic-CoA was not a substrate for this AT domain, and as expect it was also not accepted by the malonyl-specific AT domains (Figure 3.19).

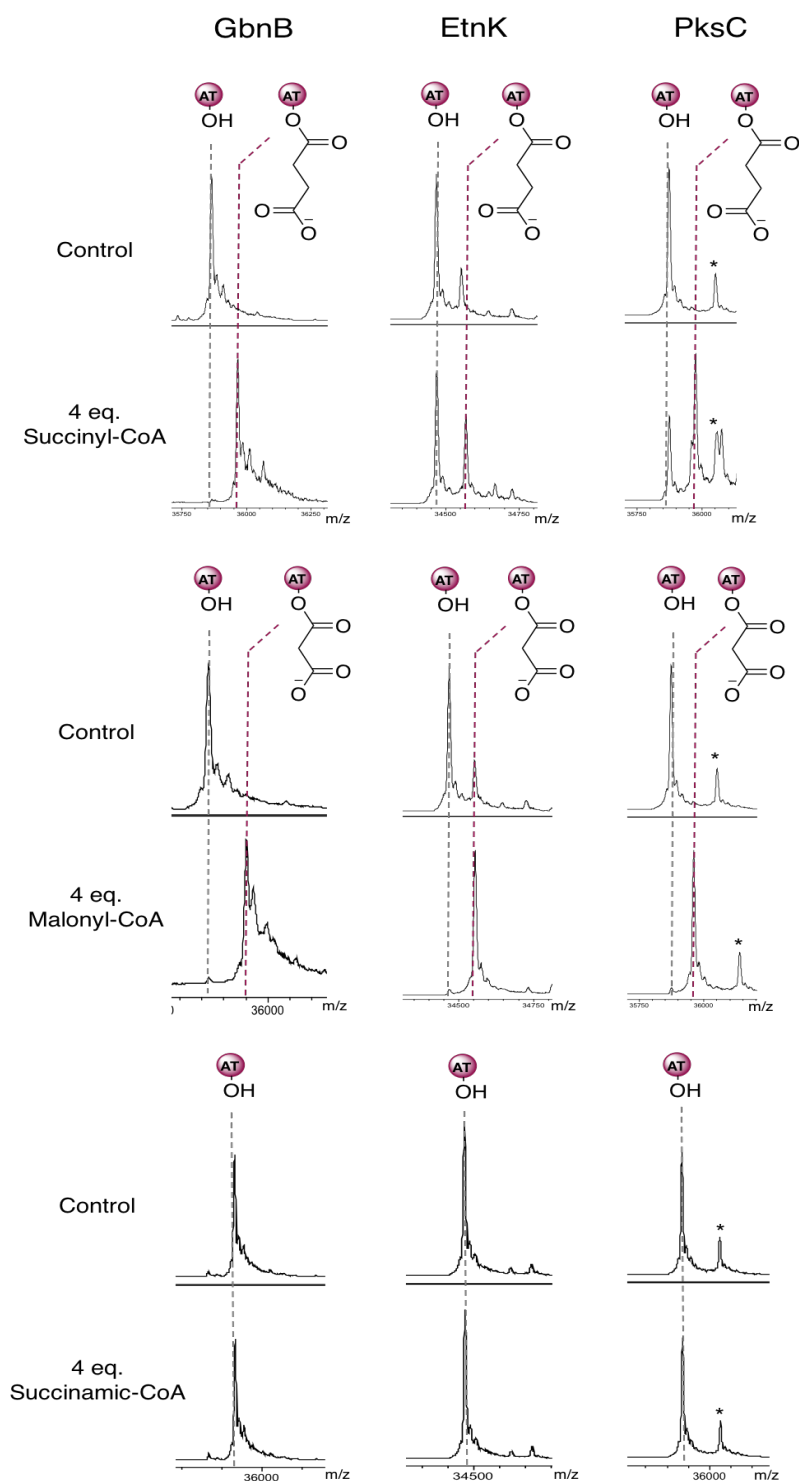


Figure 3.19: Stacked deconvoluted ESI-MS spectra of the AT domains GbnB, EtnK and PksC showing whether they are capable of accepting succinyl-CoA (first), malonyl-CoA (middle) or succinamic-CoA (last) as substrates. Succinamic-CoA was not a substrate for any of the AT domains. For the control reactions the CoA derivatives were omitted (for succinamic-CoA the VbxP control reaction was used for the control to demonstrate that free succinamic acid could not self-load onto the AT domains).

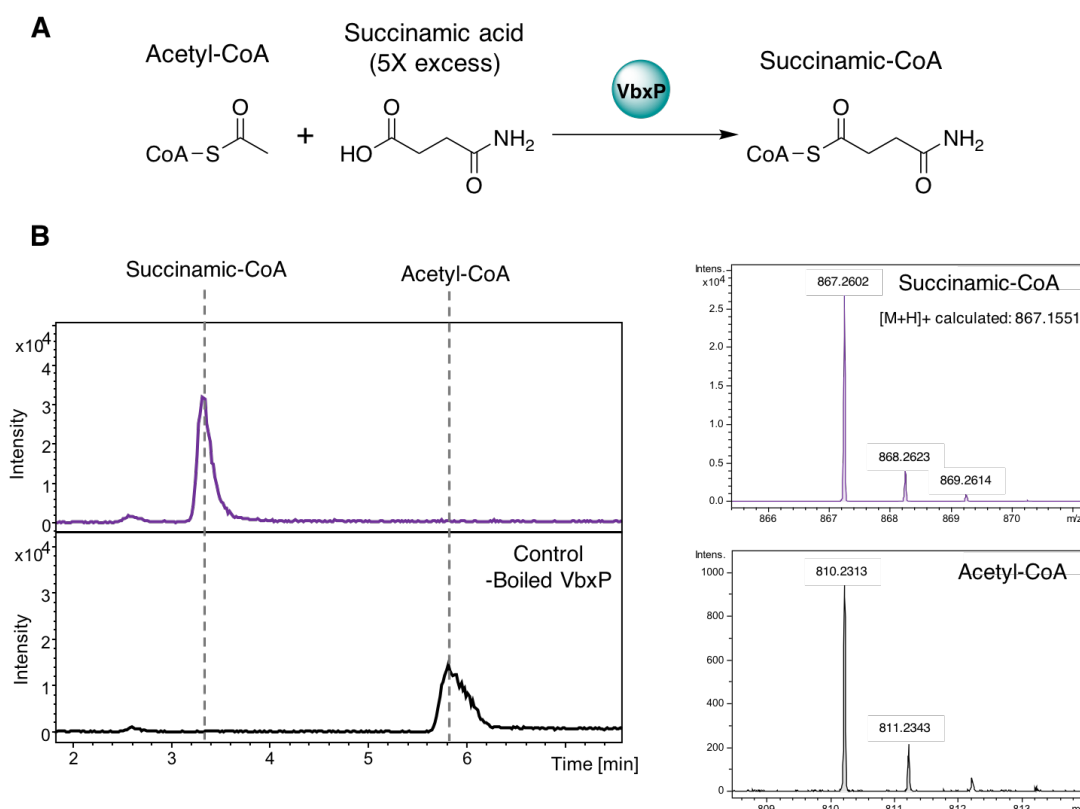


Figure 3.20: (A) Overview of the VbxP catalysed reaction used to generate succinamic-CoA. (B) Base peak chromatograms from the LC-MS analysis of the CoA transferase, VbxP, catalysed reaction of acetyl-CoA with succinamic acid (purple) and the control reaction using boiled VbxP (black). VbxP was able to completely convert acetyl-CoA to succinamic-CoA for which the high-resolution MS data is shown, alongside the MS data of acetyl-CoA for comparison.

As succinyl-CoA was found to be a substrate for GbnB, the next step will be to determine if GbnB is capable of transferring the succinyl unit onto the phosphopantetheinyl moiety of either *holo*-GbnA or *holo*-GldM1ACP(I). This work is currently ongoing.

3.7 Summary

In this study, the gene, *gbnM*, that encodes for a putative amidase, which was unaccounted for in the gladiolin biosynthetic pathway, was cleanly deleted generating a mutant strain *B. gladioli* BCC1622 Δ *gbnM* that produced a novel, amidated analogue of gladiolin. This revealed that amidation of gladiolin or a gladiolin precursor plays a role in the biosynthesis. Characterisation of amide gladiolin revealed that amidation, which was shown to occur on the carboxylic acid moiety of gladiolin, did not result in the loss of antimicrobial activity. In fact amide gladiolin was found to display moderate antimicrobial activity against the gram-negative bacteria, *A. baumannii* and *E. coli*, which was not seen for gladiolin. The fact that amide gladiolin maintained a similar level of antimicrobial activity to gladiolin suggests that amidation may not play a primary role in self-resistance, as was initially hypothesised.

Studies on the complementary gene, *gbnC*, which encodes for a putative asparagine synthetase, that is proposed to amidate the carboxylic acid moiety, proved to be not straightforward. Inactivation of *gbnC* *in vivo* abolished production of gladiolin, which established that GbnC plays an essential role in the biosynthesis. Unfortunately, *in vitro* studies were not possible due to GbnC being completely insoluble when recombinantly overproduced in *E. coli*. To ascertain whether GbnC acts prior to polyketide assembly, investigation into the cryptic chain initiation in gladiolin biosynthesis was begun. It was established that the unusual AT domain was able to accept the proposed starter unit, succinyl-CoA, however, it will need to be determined if this AT domain is capable of loading the succinyl-unit onto either of the ACP domains that could be involved in chain initiation.

Chapter 4

Conclusions and perspectives

4.1 O-methylation during chain assembly by *trans*-AT modular polyketide synthases

This study presents the first biochemical characterisation of embedded OMT domains from *trans*-AT PKSs. Embedded OMT domains could provide a novel approach to engineering methylation, where the position of methylation would depend only upon the location of the domain within the PKS. For successful bioengineering a comprehensive understanding of the biochemical activity and structure of the domain, as well as knowledge of the domain interactions is required.

4.1.1 Exploration of the substrate tolerance of embedded OMT domains

In order to assess the activity of embedded OMT domains *in vitro* a mass spectrometry based assay was designed and implemented. The substrate tolerance of four embedded OMT domains from three different polyketide assembly lines was explored, revealing potentially useful substrate promiscuity. It was found that each OMT domain showed its own unique substrate tolerance profile, however all of the OMT domains were found to accept at least some of the short substrate mimics used in this study, despite the proposed *in vitro* substrates being substantially longer. Three of the OMT

domains were able to methylate β -hydroxyl groups of both stereochemistries. This is likely due to the fact that there would be no evolutionary pressure on the *OMT* domains to demonstrate stereoselectivity, as the *KR* domains found within each methylation module would generate the hydroxyl group to be methylated in a stereospecific manner. It is probable that substituents surrounding the methylating hydroxyl groups will have the biggest influence over whether the substrate will be accepted by the *OMT* domain. It was demonstrated that three of the *OMT* domains were able to accept substrates that also had a δ -hydroxyl group. In growing polyketide chains substituents, most commonly methyl groups, can often be found in the α - and γ - positions. Future work could, therefore, involve synthesising short pantetheine substrate mimics that possess a β -hydroxyl group as well as an α - or γ - methyl group, and assessing their ability to be methylated by the *OMT* domains (Figure 4.1). Comprehensive knowledge of the substrate tolerance of embedded *OMT* domains will allow for an understanding of where these domains can be placed successfully in a polyketide assembly line.

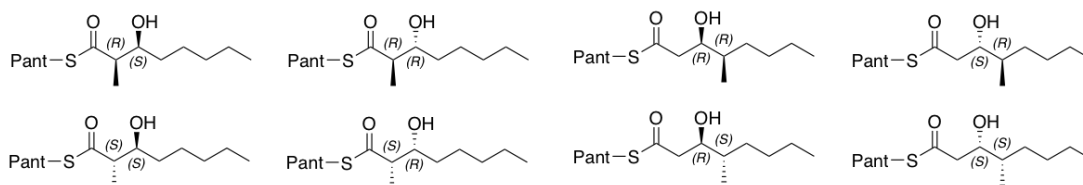


Figure 4.1: Proposed substrates for further exploring the substrate tolerance of embedded *OMT* domains from *trans*-AT PKSs.

4.1.2 Characterising the interactions between embedded *OMT* domains and ACP domains

Within *O*-methylation modules in *trans*-AT PKSs there is always a non-elongating (or hypothesised non-elongating) *KS* domain found before the *OMT* domain, which is proposed to just translocate the polyketide chain to the downstream ACP domain so that *O*-methylation can occur. For the gladiolin *O*-methylation module the transacylation activity of the hypothesised non-

elongating KS domain was proven. It was also demonstrated that methylation was only able to occur on the second ACP domain, thus establishing that a specific interaction is required between the *OMT* domain and ACP domain for catalytic activity. The fact that *O*-methylation could only be observed for the downstream ACP domain provides evidence towards ACP domain specificity being the reason for the incorporation of a non-elongating KS domains as part of the methylation module.

Phylogenetic analysis revealed that embedded *OMT* domains from *trans*-AT PKSs formed two clades. *OMT*-ACP domain crosstalk assays revealed that the gladiolin *OMT* domain, from clade II, was only able to catalyse *O*-methylation of substrates bound to ACP domains that were associated with other clade II *OMT* domains. Conversely, methylation was observed for the MisM1*OMT* domain, from clade I, when incubated with ACP domains associated with *OMT* domains from both clades, however, a greater extent of methylation was observed for ACP domains affiliated with other clade I *OMT* domains. This suggests that there are more significant *OMT*-ACP domain interaction between the domains associated with the same clade, and indicates that two different *OMT*-ACP domain communication mechanisms may have evolved.

Future work could involve characterisation of these interactions. There are a number of techniques available for quantifying protein-protein interaction; one of the most commonly used techniques to determine binding affinities and association/dissociation kinetics is surface plasmon resonance. This involves immobilisation of one of the interacting proteins on a dedicated sensor surface, prior to the partner protein being injected across the surface of the immobilised protein. The binding is monitored by changes in the refractive index of the medium close to the surface, upon addition of the second protein¹⁴⁰. It would also be interesting to map the binding sites for the *OMT*-ACP domain interactions to establish how they differ from each other, and from the interactions between ACP domains and other tailoring domains. This could be

achieved by carbene footprinting, which utilises diazirine reagents to carbene-label the surface of proteins upon photoirradiation; the labelled residues are identified by MS analysis of the digested protein¹⁴¹. Binding sites are determined by comparing the labelling of a protein in the presence and absence of the interacting partner protein. To gain the most from this, however, a structure of an embedded OMT domain would be required.

4.1.3 Structural studies of embedded OMT domains

This work is in collaboration with Dr Hussain Buhkya, Monash University, who conducted and analysed the negative stain TEM imaging of GldM11KS-OMT-ACP.

Unfortunately, a crystal structure of the embedded OMT domain from the gladiolin biosynthetic pathway could not be obtained, as no conditions were identified that produced protein crystals. Inherent biochemical features of the protein can influence whether or not a protein is able to crystallise¹⁴², thus future work could involve setting up crystallisation trials for the other embedded OMT domains looked at in the study.

Cryo-EM is an alternative structural biology technique that is able to determine high resolution (3-5 Å) structures of protein complexes that are 200 kDa or larger¹⁴³, which has recently been applied to the study of the structure of a PKS module¹⁴⁴. As cryo-EM is also able to provide information about the spatial organisation of domains and identify interaction surfaces between the proteins, it was decided that this would be a useful technique to apply to the KS⁽⁰⁾-OMT-ACP subset of domains, that were determined to be essential for O-methylation on *trans*-AT PKS assembly lines. GldM11KS-OMT-ACP was cloned into pET28a, and it was demonstrated that the tri-domain could be overproduced and purified to a high level of homogeneity, which was characterised by SDS-PAGE analysis and intact protein MS (Figure 4.2 (A), Appendix 1). The construct was subsequently provided to Dr Hussain Buhkya

(Challis group, Monash University), who is attempting to obtain high resolution cryo-EM images of this protein. Initial negative stain TEM data revealed that the protein (which was found to be around 14-17 nm in size, with the KS domain being approximately 6-7 nm and the OMT domain 8-9 nm) is suitable for cryo-EM studies, with a number of different orientations of the molecule being able to be observed (Figure 4.2 (B)).

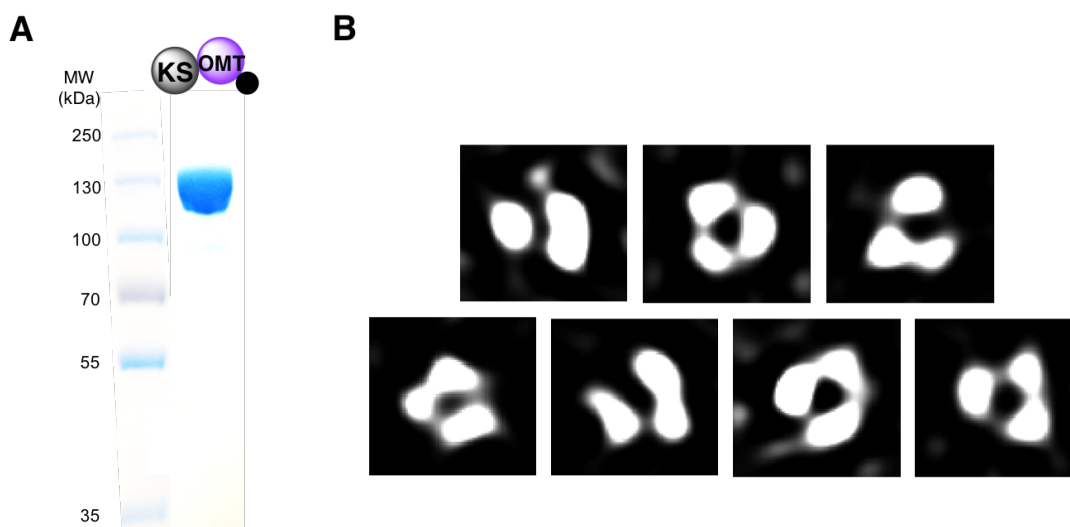


Figure 4.2: (A) 8 % SDS-PAGE gel showing purified GldM11KS-OMT-ACP. (B) 2D class averages of the negative stain TEM images of the tri-domain protein obtained by classification of different conformations using EMAN 2.1¹⁴⁵.

4.1.4 Bioengineering *O*-methylation

The understanding of *O*-methylation on *trans*-AT PKS assembly lines gained from this study will hopefully provide a basis for future efforts towards rational engineering of *O*-methylation in these systems. The biosynthesis of gladiolin presents an application for this. Gladiolin (and amide gladiolin) undergoes conversion to a 24-membered lactone isomer in protic solvents due to nucleophilic addition of the C-23 hydroxyl group to the ester, followed by elimination of the C-21 hydroxyl group (Figure 3.9 (A)). Future work could involve attempting selective methylation of the C-23 hydroxyl group, to

increase the stability of gladiolin, via the introduction of the subset of domains essential for *O*-methylation, to module 9 of the PKS (Figure 4.3).

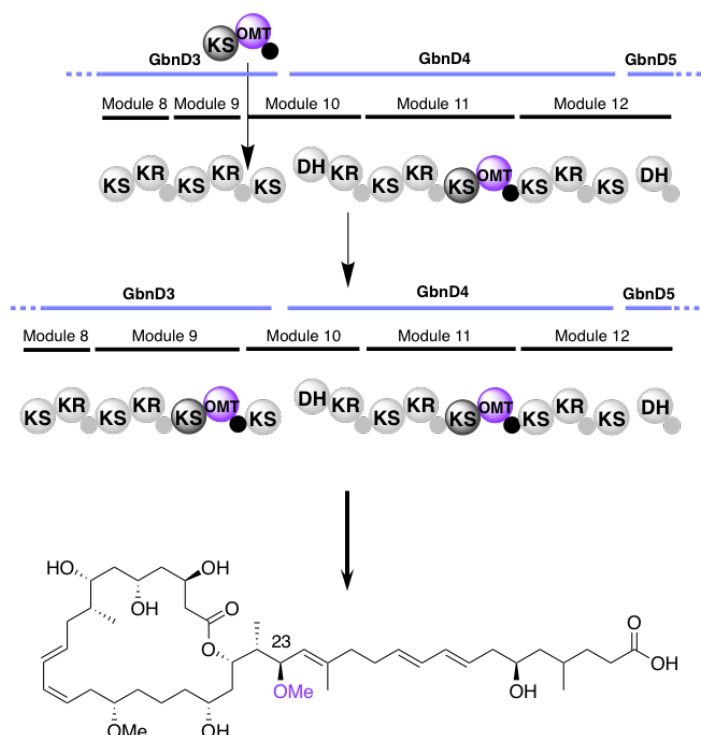


Figure 4.3: Proposed bioengineering of the gladiolin PKS to introduce methylation of the hydroxyl-group at position 23.

4.2 Investigating cryptic starter unit amidation

4.2.1 Amide gladiolin

In vivo inactivation of the putative amidase encoding gene, *gbnM*, from the gladiolin BGC lead to the discovery of a novel amide derivative of the antibiotic, and also provided evidence that GbnM is responsible for converting it to the mature antibiotic. The ability of GbnM to hydrolyse the amide was confirmed by *in vitro* studies. The knowledge that amidation of gladiolin or a gladiolin precursor plays a role in the biosynthesis will prove useful for future biosynthetic studies, on this and similar PKS systems, as amidated biosynthetic intermediates may be required or observed.

The antimicrobial activity of amide gladiolin was determined for representative members of the ESKAPE panel as well as *C. albicans* and *M. tuberculosis*. It was found that the amidation of gladiolin did not result in the loss of antimicrobial activity and interestingly, lead to moderate activity against the gram-negative bacteria, *A. baumannii* and *E. coli*, that was not seen for gladiolin. Multidrug-resistant strains of *A. baumannii*, *E. coli*, and *M. tuberculosis* (for which the MIC of amide gladiolin was determined to be 2 µg/mL) are of increasing clinical prevalence, so there is a pressing need for new antibiotics to treat such infections. Amide gladiolin could provide the basis for the development of an antibiotic to address this problem.

4.2.2 The role of GbnC in the biosynthesis of gladiolin

In vivo deletion of the putative asparagine synthetase encoding gene, *gbnC*, which is proposed to be responsible for the amidation of the carboxylic acid moiety of gladiolin, abolished the production of gladiolin and related metabolites. This established that *gbnC* plays an essential role in the biosynthesis. Unfortunately, *in vitro* studies to assess the activity of GbnC were not possible, as GbnC was completely insoluble when heterologously overproduced in *E. coli*. Bioinformatics revealed that the BGCs responsible for the biosynthesis of etnangien and sorangicin, both of which share numerous similarities to gladiolin, contain homologs of both GbnM and GbnC. It was established that SorQ, the homolog of GbnC from the sorangicin biosynthetic pathway was also completely insoluble when heterologously overproduced in *E. coli*. Future work could involve determining if soluble protein can be obtained from the overproduction of the homolog, EtnC, from the etnangien biosynthetic pathway.

4.2.3 Exploration of polyketide chain initiation in gladiolin biosynthesis

To determine if amidation is necessary for initiation of polyketide biosynthesis, the cryptic chain initiation mechanism for gladiolin biosynthesis was investigated. The gladiolin PKS lacks a clearly demarcated loading module, however, there is a standalone ACP domain, GbnA, which was determined by *in vivo* inactivation to be essential for biosynthesis, that could act as the loading module. Polyketide chain initiation by a standalone ACP domain has been suggested to occur in a number of *trans*-AT PKSs⁵⁰.

The gladiolin BGC also encoded for a putative AT domain, GbnB, which was found to clade independently from malonyl-specific AT domains, thus it was proposed that this AT domain could be responsible for the loading of a succinyl-CoA starter unit. To investigate this, the substrate tolerance of GbnB was explored, and it was found that both succinyl- and malonyl-CoA were accepted substrates, but succinamic-CoA was not. If it can be confirmed that GbnB is responsible for the loading of the starter unit, then this will establish that GbnC does not act directly upon succinyl-CoA. Future work will involve determining if GbnB is able to transfer the succinyl-unit to GbnA, or the first ACP domain of GbnD1, and whether any selectivity is displayed.

KS domains from *trans*-AT PKSs are known to play a gatekeeping role, by demonstrating specificity for the downstream polyketide intermediate¹⁴⁶. This means that determining the substrate specificity of the *N*-terminal KS domain of GbnD1 will be key to proving if amidation is important for polyketide chain initiation. Unfortunately, initial attempts to overproduce this KS domain yielded only insoluble protein. Future work will involve testing a number of different domain boundaries for the KS domain to see if soluble protein can be obtained.

4.2.4 Revised gladiolin biosynthesis

This research has revealed that the previously overlooked genes in the gladiolin BGC encoding for an asparagine synthetase (GbnC), an amidase (GbnM) and a standalone ACP domain (GbnA) play an essential role in the biosynthesis of gladiolin. It was initially hypothesised that the asparagine synthetase and amidase could play a role in antimicrobial self-resistance, however the observed antimicrobial activity of amide gladiolin suggests that amidation alone is not able to confer resistance. To ensure that amidation of gladiolin did not change the target of the antibiotic, a collaboration has been established with the Ebright Lab (Waksman Institute of Microbiology, Rutgers), who are looking to confirm that amide gladiolin acts by inhibition of RNA polymerase, and to also biochemically analyse the binding site and mechanism of the gladiolins. It is therefore probable that amidation of gladiolin acts as a biosynthetic protecting group. It is thus proposed that the biosynthesis of gladiolin proceeds in a similar fashion to the biosynthesis of the glutarimide-containing polyketides (Figure 3.15). The starter unit succinyl-CoA is first loaded onto the ACP domain GbnA, most likely by the AT domain GbnB, before being amidated by the asparagine synthetase, GbnC. The succinamic moiety is then passed onto the KS domain of GbnD1 and gladiolin is built as the amidated derivative. It was demonstrated, both *in vivo* and *in vitro*, that the amidase, GbnM, is able to convert amide gladiolin to the mature antibiotic. The revised biosynthesis of gladiolin is shown in Figure 4.4.

As the BGCs of both etnangien and sorangicin contain homologs of the asparagine synthetase and amidase, and the biosynthesis of these polyketides also utilise a dicarboxylic acid starter unit it is very likely that these biosyntheses proceed in the same way as proposed for the biosynthesis of gladiolin. The sorangicin BGC is missing a homolog of the standalone ACP domain (GbnA), however it is suggested that the sorangicin BGC has a split cluster organisation as the β -branching cassette required for the biosynthesis of sorangicin is absent from the identified BGC¹⁴⁷.

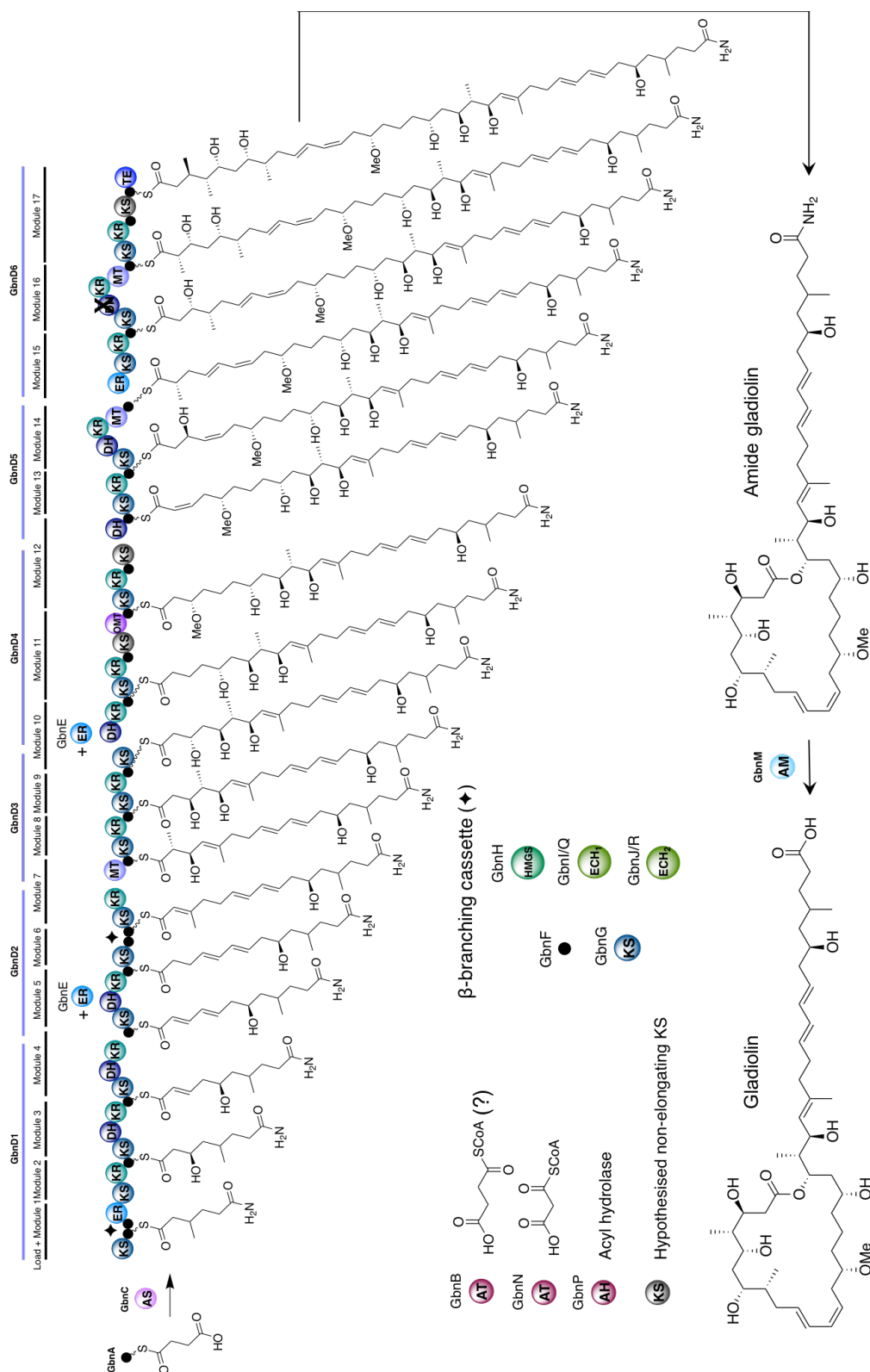


Figure 4.4: Revised pathway for gladiolin biosynthesis, showing the ACP-bound predicted polyketide intermediates following α - and β -carbon processing. Black circles represent ACP domains. The domain with a cross through it indicates that the domain lacks the key catalytic residues required for activity. AS= asparagine synthetase and AM= amidase.

Chapter 5

Experimental

5.1 Microbial strains and culture conditions

5.1.1 Microbial strains

Table 5.1: Bacterial and fungal strains used within this project. Strains shown in dark blue are risk group 2 microorganisms.

Microbial strain	Growth temperature	Use	Source
<i>Escherichia coli</i>			
TOP10 (chemically competent)	37 °C	Cloning and plasmid propagation	Lab stock
NEB® 5-alpha (chemically competent)	37 °C	Cloning and plasmid propagation	NEB
SY327 (electrocompetent)	37 °C	Cloning, plasmid propagation and donor strain for conjugation between <i>E. coli</i> and <i>B. gladioli</i>	Lab stock
HB101 (pRK 2013)	37 °C	Helper strain for conjugation between <i>E. coli</i> and <i>B. gladioli</i>	Lab stock
BL21(DE3) (chemically competent)	37 °C	Expression of recombinant proteins	Lab stock
C43(DE3) (chemically competent)	37 °C	Expression of recombinant proteins	Lab stock

Other			
<i>Burkholderia gladioli</i> BCC1622	30 °C	Genetic manipulation and production of gladiolin	Cardiff Burkholderia collection
<i>Burkholderia gladioli</i> BCC1622 Δ <i>gbnM</i>	30 °C	Production of amide gladiolin	Genetically engineered in this study
<i>Burkholderia gladioli</i> BCC1622 Δ <i>gbnC</i>	30 °C	Biosynthetic studies of gladiolin	Genetically engineered in this study
<i>Burkholderia gladioli</i> BCC1622 Δ <i>gbnA</i>	30 °C	Biosynthetic studies of gladiolin	Genetically engineered in this study
<i>Burkholderia gladioli</i> BCC1622 Δ <i>gbnB</i>	30 °C	Biosynthetic studies of gladiolin	Genetically engineered in this study
<i>Klebsiella pneumonia</i> DSM26371	30 °C	Antimicrobial susceptibility testing	DSMZ culture collection
<i>Acinetobacter baumannii</i> DSM25645	30 °C	Antimicrobial susceptibility testing	DSMZ culture collection
<i>Pseudomonas aeruginosa</i> DSM29239	30 °C	Antimicrobial susceptibility testing	DSMZ culture collection
<i>Enterobacter cloacae</i> DSM16690	30 °C	Antimicrobial susceptibility testing	DSMZ culture collection
<i>Enterococcus faecium</i> DSM25390	30 °C	Antimicrobial susceptibility testing	DSMZ culture collection
<i>Staphylococcus aureus</i> DSM21979	30 °C	Antimicrobial susceptibility testing	DSMZ culture collection
<i>Candida albicans</i> SC 5314	30 °C	Antimicrobial susceptibility testing	ATCC culture collection
<i>Mycobacterium smegmatis</i> DSM43756	37 °C	Antimicrobial susceptibility testing	DSMZ culture collection

5.1.2 Plasmids

Table 5.2: Plasmids used within this project (plasmid maps can be found in Appendix 10).

Plasmid	Use	Resistance	Source
pET24a	Overproduction of proteins with an <i>N</i> -terminal His ₈ -tag	Kanamycin	Epoch Life Sciences
pET28a	Overproduction of proteins with an <i>N</i> -terminal His ₆ -tag	Kanamycin	Novagen
H-MBP-3C	Overproduction of proteins with an <i>N</i> -terminal His ₆ -tag and MBP fusion	Ampicillin	Alexandrov <i>et al.</i> ¹⁴⁸
pET151	Overproduction of proteins with an <i>N</i> -terminal His ₆ -tag	Ampicillin	Thermo Scientific
pET SUMO	Overproduction of proteins with an <i>N</i> -terminal His ₆ -tag and SUMO fusion	Kanamycin	Thermo Scientific
pGPI-Scel	Suicide plasmid for introducing a targeted I-SceI restriction site in the genome of <i>Burkholderia</i>	Trimethoprim	Flannagan <i>et al.</i> ¹²⁵
pRK 2013	Plasmid vehicle for DNA transposition	Kanamycin	Figurski and Helinski ¹⁴⁹
pDAI-Scel	Yeast homing endonuclease I-SceI expression vector	Tetracycline	Flannagan <i>et al.</i> ¹²⁵
pMLBAD	Broad host range vector for arabinose-inducible <i>in trans</i> expression of proteins	Trimethoprim	Lefebvre and Valvano ¹²⁶

5.1.3 Media

All media was made using deionised water (dH₂O) and sterilised using an autoclave (121 °C, 15 min).

LB media 25 g/L lysogeny broth (+15 g/L agar for LB agar plates).

M9 minimal media	250 mL sterile agar (6 g agar in 250 mL dH ₂ O, autoclave sterilised), 40 mL 10X M9 salts (30 g anhydrous Na ₂ HPO ₄ , 15 g anhydrous KH ₂ PO ₄ , 5 g NaCl and 10 g NH ₄ Cl dissolved in 500 mL dH ₂ O, autoclave sterilised), 0.4 mL 1 M MgSO ₄ (in dH ₂ O, filter sterilised), 0.4 mL 0.1 M CaCl ₂ (in dH ₂ O, filter sterilised), 50 % sucrose (in dH ₂ O, filter sterilised) and sterile dH ₂ O (to 400 mL) were all warmed to 50 °C and combined, before the plates were poured ¹⁵⁰ .
BSM-G media	50 mL/L 20X phosphate salts (85 g/L KH ₂ PO ₄ ·3H ₂ O and 20 g/L Na ₂ HPO ₄ ·H ₂ O), 50 mL/L 20X ammonium chloride (40 g/L), 10 mL/L 100X metal salts (20 g/L MgSO ₄ ·7H ₂ O, 1.2 g/L FeSO ₄ ·7H ₂ O, 0.3 g/L MnSO ₄ ·H ₂ O, 0.3 g/L ZnSO ₄ ·7H ₂ O and 0.1 g/L CoSO ₄ ·7H ₂ O, light sensitive), 0.5 g/L CAS amino acids, 0.5 g/L yeast extract, 0.1 g/L nitrilotriacetic acid, 4 g/L glycerol and 15 g/L agar (supplemented with 12.5 mL/L 20 % arabinose after autoclave sterilisation where required) ¹⁵¹ .
Mueller Hinton media	22 g/L Mueller Hinton II broth.
Mycobacteria media	2 g yeast extract, 2 g proteose peptone No.3, 2 g peptone from casein (tryptic digest), 2.5 g Na ₂ HPO ₄ ·12H ₂ O, 1 g KH ₂ PO ₄ , 1.5 g sodium citrate, 0.6 g MgSO ₄ ·7H ₂ O, 50 mL glycerol, 0.5 g Tween 80 (+15 g/L agar for mycobacterium agar plates) in 1 L.

7-H9 media 5 g/L Middlebrook 7-H9 broth base was supplemented with 0.5 % glycerol. This was enriched with OADC growth supplement after autoclave sterilisation.

5.1.4 Culture conditions

Unless stated otherwise cultures were grown overnight in LB at the appropriate growth temperature (Table 5.1). Liquid cultures were shaken at 180 rpm. Antibiotics, where required, were used at the working concentration and all antibiotic stock solutions were sterilised by filtering through a 0.2 µm MiniStart® syringe filter (sartorius), and stored at -20 °C (table 5.3). All work with bacteria was carried out under aseptic conditions and any work with risk group 2 microorganisms (Table 5.1) was carried out in a category 2 containment laboratory equipped with a Walkers Class 2 MSC safety cabinet.

Table 5.3: Antibiotics used within this project.

Antibiotic	Stock conc.	Solvent	Working conc.
Kanamycin (Kan)	50 mg/mL	H ₂ O	50 µg/mL
Ampicillin (Amp)	100 mg/mL	H ₂ O	100 µg/mL
Trimethoprim (Tp)	50 mg/mL	DMSO	50 µg/mL
Tetracycline (Tc)	20 mg/mL	70 % EtOH	20 µg/mL
Polymyxin (PMX)	300000 units/mL	H ₂ O	600 units/mL

5.2 Buffers

Buffers were made using dH₂O. The pH was adjusted with either HCl or NaOH and filtered using an EDM Millipore™ Steritop™ vacuum bottle-top filter where required. The loading, storage, gel filtration, HEPES buffer A, HEPES buffer B and crystallisation buffers were stored at 4 °C. The SDS-PAGE loading buffer was stored at -20 °C.

TBE buffer	89 mM Tris-base, 89 mM boric acid, 2 mM EDTA, pH 8.3.
Loading buffer	20 mM Tris-HCl, 100 mM NaCl, 20 mM imidazole, pH 8.
Storage buffer	20 mM Tris-HCl, 100 mM NaCl, pH 8.
300 mM imidazole buffer	20 mM Tris-HCl, 100 mM NaCl, 300 mM imidazole, pH 8.
	This buffer was diluted with storage buffer to give 50 mM, 100 mM and 200 mM imidazole buffers.
SDS-PAGE loading buffer (5X)	250 mM Tris-HCl (pH6.8), 10 % (w/v) SDS, 30 % (v/v) glycerol, 10 mM β -mercaptoethanol, 0.02 % (w/v) bromophenol blue.
SDS-PAGE running buffer	25 mM Tris-base, 192 mM glycine, 0.1 % (w/v) SDS, pH 8.3.
Gel filtration buffer	20 mM Tris-HCl, 150 mM NaCl, pH 8.
HEPES buffer A	30 mM HEPES, 500 mM NaCl, 10 % glycerol, pH 7.5.
HEPES buffer B	30 mM HEPES, 500 mM NaCl, 150 mM imidazole, 10 % glycerol, pH 7.5.
Crystallisation buffer	10 mM HEPES, 150 mM NaCl, pH 7.5.
300 mM sodium phosphate buffer	13.44 mL 1M Na_2HPO_4 and 1.56 mL 1 M NaH_2PO_4 were diluted to 50 mL (pH 7.8).

5.3 DNA manipulation and cloning procedures

5.3.1 DNA isolation and purification

B. gladioli BCC1622 genomic DNA was extracted from a 5 mL overnight culture using the GeneJET Genomic DNA Purification Kit (Thermo Scientific), following the manufacturer's recommended protocol for gram-negative bacteria.

For extraction of plasmid DNA, overnight cultures (7 mL), with the appropriate antibiotic, were inoculated using single colonies of the propagating *E.coli* strain. The DNA was extracted using a GeneJET Plasmid Miniprep Kit (Thermo Scientific), according to the manufacturer's instructions.

Where required, DNA fragments were excised from agarose gels and purified using a GeneJET Gel Extraction Kit (Thermo Scientific) following the manufacturer's recommended protocol. All purified DNA was stored at -20 °C.

5.3.2 PCR

PCRs were carried out using an Eppendorf Mastercycler® nexus GX2 and the reactions were set up on ice in 0.2 mL PCR tubes as shown in Table 5.4.

Table 5.4: Volumes of reagents used for 25 μ L PCR reactions.

General PCR		Colony PCR	
Reagent	Volume (μ L)	Reagent	Volume (μ L)
Template DNA*	0.5-1	Overnight culture/ pGPI vector	1
Phusion High-Fidelity 2X Master Mix (GC Buffer, NEB)	12.5	OneTaq Hot Start Quick-Load 2X Master Mix (GC buffer, NEB)	12.5
10 μ M Forward primer	1.25	10 μ M Forward primer	1.25
10 μ M Reverse primer	1.25	10 μ M Reverse primer	1.25
DMSO (5 %)	1.25	DMSO (5 %)	1.25
Sterile H ₂ O	8.25-7.75	Sterile H ₂ O	7.75

* Template DNA was either *B. gladioli* BCC1622 genomic DNA (30 ng/ μ L), plasmid DNA (10 ng/ μ L, diluted with sterile H₂O) or a synthetic DNA fragment (GenScript, 10 ng/ μ L, diluted with sterile H₂O).

The thermocycling conditions used were as follows:

<u>General PCR</u>	<u>Colony PCR</u>
98 °C – 3 minutes	94 °C – 3 minutes
98 °C – 10 seconds	94 °C – 30 seconds
X35 T_A — 30 seconds	X35 T_A — 60 seconds
72 °C – 45-60 seconds/kb	68 °C – 60 seconds/kb
72 °C – 10 minutes	68 °C – 10 minutes
4 °C – Hold	4 °C – Hold

The appropriate annealing temperature (T_A) for each set of primers (all of which were purchased from Sigma-Aldrich) was determined from gradient PCR trials. The product(s) of the reactions were analysed as discussed in section 5.3.3.

5.3.3 DNA size and concentration analysis

PCR products and DNA fragments were analysed by agarose gel electrophoresis. 1 % agarose gels were made up in TBE buffer supplemented with 0.01 $\mu\text{L/mL}$ GelRed™ nucleic acid gel stain (Biotium). Prior to electrophoresis, samples were prepared by mixing with 6X DNA gel loading dye (with the exception of colony PCRs where the loading dye was included in the polymerase master mix) and the gels were run in TBE buffer at 90 V for 60-90 min, depending on the required separation, using a Bio-Rad Sub-Cell® GT Cell and visualised using a UVP BioDoc-It® Imaging System. GeneRuler 1 kb DNA ladder (Thermo Scientific) was used as a reference to predict the size of the DNA. When required the DNA was purified from the gel as discussed in section 5.3.1.

DNA concentrations were determined using a Thermo Scientific™ NanoDrop™ Lite spectrophotometer.

5.3.4 Recombinant plasmid creation

5.3.4.1 Traditional cloning

5.3.4.1.1 Restriction digestion

Restriction recognition sites were included in the designed oligonucleotide primers, which are shown in Table 5.5, that were used in the PCRs (section 5.3.2). The resultant DNA fragments (and required vector) were incubated with the appropriate restriction enzymes (Thermo Scientific), utilising the conditions recommended by the Thermo Scientific DoubleDigest Calculator, or incubated with FastDigest buffer for FastDigest enzymes. The reaction mixtures were incubated at 37 °C for 30 min (for FastDigest enzymes) to 1 h 30 min. The linearized vector was purified as described in section 5.3.3, and digested PCR products were purified directly using the GeneJET Gel Extraction Kit by adding

a 1:1 volume of binding buffer to the digestion reaction, and then proceeding with the recommended protocol.

Table 5.5: Oligonucleotide primers used for the creation of recombinant plasmids, the restriction sites used are underlined.

Target Plasmid	Primers
	For_GldM11OMT+NdeI
pET28a-GldM11 OMT	(5'-GAAAGG <u>CATATG</u> GGCGGTCCGGGTGACAAG-3')
	Rev_GldM11OMT+HindIII
	(5'-AAAGGG <u>AAGCTT</u> CACGGCGCCTGCCGCGA-3')
	For_GldM11KSOMT+NdeI
pET28a-GldM11KSOMT	(5'-AAAGGG <u>CATATG</u> GATCAGGGCGGGGTAGCG-3')
	Rev_GldM11OMT+HindIII
	For_GldM11KSOMTACP+NdeI
pET28a-GldM11KSOMTACP	(5'-TAAAC <u>CATATG</u> AACCGGAGCGAATCGACAC-3')
	Rev_GldM11KSOMTACP+HindIII
	(5'-AACTA <u>AAGCTT</u> CAATAGCTGATCCAGTTCTG-3')
	For_MisM9OMTACP+NheI
pET28a-MisM9OMTACP	(5'-AAAGGGG <u>CTAGCG</u> GCGAGGAA-3')
(**)	Rev_MisM9OMTACP+HindIII
	(5'-GAAGAGA <u>AAGCTT</u> CACCTGGA-3')
	For_RizM18OMTACP+NdeI
pET28a-RizM18OMTACP	(5'-AGAGAG <u>CATATG</u> CGTGTCTGAG-3')
(**)	Rev_RizM18OMTACP+HindIII
	(5'-TTTCCCA <u>AAGCTT</u> CACTGCGA-3')
	For_GldM11ACP(I)+NdeI
pET28a-GldM11ACP(I)	(5'-AGGG <u>CATATG</u> GACGAGGAAGACGAAACCATG-3')
	Rev_GldM11ACP(I)+SalI
	(5'-AGGGG <u>TCTG</u> ACTCACCGAGCGAGATCGGCCCG-3')
	For_GldM11ACP(II)+NdeI
pET28a-GldM11ACP(II)	(5'-AGGAAG <u>CATATG</u> CCGTCTGGCCGGAGCGCGC-3')
	Rev_GldM11ACP(II)+HindIII
	(5'-AAAGGG <u>AAGCTT</u> CAGCGGGCGCCACGCGTTTC-3')
	For_MisM9ACP(II)+NdeI
pET28a-MisM9ACP(II)	(5'-AAAGGG <u>CATATG</u> CAGCCGATGGACGGTCAG-3')
	Rev_MisM9ACP(II)+HindIII

	(5'-AGAGAGAAGCTTTCACCTGGATTGAGCCTC-3')
	For_MisM1ACP(II)+NdeI
pET28a-MisM1ACP(II)	(5'-AAAGGGGCATATGACATCCGTGGCAGATATC-3')
	Rev_MisM1ACP(II)+HindIII
	(5'-AGAGAGAAGCTTTCATGAGACAACGGGTACAGA-3')
	For_RizM18ACP(II)+NdeI
pET28a-RizM18ACP(II)	(5'-AAAGGGGCATATGGCTTCGTGCGCGGATATC-3')
	Rev_RizM18ACP(II)+HindIII
	(5'-AGAGAGAAGCTTTCAGTGCAGACGAC-3')
	For_MisM1OMT+NdeI
pET28a-MisM1OMT	(5'-GAAAGGGCATATGCAAAGTATGCCTGATCTC-3')
	Rev_MisM1OMT+HindIII
	(5'-AGGAAGAAGCTTTCAGACGATTTGCTCTATATG-3')
	For_gbnM+NdeI
pET28a- <i>gbnM</i> (*)	(5'-AAAAAACATATGACTCCCGAACTCGACCTC-3')
	Rev_gbnM+HindIII
	(5'-AAAAAAAAGCTTGGCCGGCGTGTCAGTTGCC-3')
	For_gbnC+NdeI
pET28a- <i>gbnC</i>	(5'-AAAAAACATATGGTGACCGTGTGCGGAATC-3')
	Rev_gbnC+HindIII
	(5'-AAAAATAAGCTTTCATACCGTCACGCGTCC-3')
H-MBP-3C- <i>gbnC</i>	For_H-MBP-3C- <i>gbnC</i> +XbaI
	(5'-AAAGGGTCTAGAAATGACTCCCGAACTCGACCTC-3')
	Rev_H-MBP-3C- <i>gbnC</i> +HindIII
	(5'-AGAGAGAAGCTTGGCCGGCGTGTCAGTTGCC-3')
	For_gbnA+NdeI
pET28a- <i>gbnA</i>	(5'-AAAGGGGCATATGCAAGACAAAATTCAGCAATTC-3')
	Rev_gbnA+HindIII
	(5'-AAAGGGAAGCTTGGAACAGGAGCGCGGTCA-3')
	For_GldM1ACP(I)+NdeI
pET28a-GldM1ACP(I)	(5'-AAAGGGGCATATGTTGCCGAAGGCCGCGGTC-3')
	Rev_GldM1ACP(I)+HindIII
	(5'-TATGGGAAGCTTTCACGGTTCCGGAGCAGCCAG-3')
	For_GldM1KS+NdeI
pET28a-GldM1KS	(5'-AAAGGGGCATATGAGCAAGGACCTCAAGGACAGC-3')
	Rev_GldM1KS+HindIII
	(5'-AGAGAGAAGCTTTCACGACAGCAGCGCGATTCC-3')

* 8 % DMSO was required for the PCR reaction.

** GenPart synthetic DNA fragments (GenScript) were used as template DNA.

5.3.4.1.2 Ligation

Each ligation reaction consisted of 50 ng of linearized vector DNA and a 4:1 molar ratio of the digested insert to vector (in some cases this was raised to a ratio of 8:1 where a 4:1 ratio was unsuccessful). 1 μ L of T4 DNA Ligase and 4 μ L of 5X Rapid DNA Ligation Buffer (both purchased from Roche) were also added, with dH₂O being used to give a final volume of 20 μ L. Ligation reactions were incubated overnight at RT.

5.3.4.2 TOPO cloning

TOPO cloning utilizes blunt-end PCR products with the addition of a CACC overhang at the 5' end of the insert to impart directionality, so restriction digestion was not needed (the designed oligonucleotide primers are shown in Table 5.6). The Champion™ pET151 Directional TOPO® Expression Kit (Thermo Scientific) was used and in general the recommended protocol was followed, however a higher insert: vector ratio (6:1) was used, and the ligation reactions were incubated overnight at RT.

Table 5.6: Oligonucleotide primers used for the creation of pET151 constructs, the added CACC overhang is shown in purple.

Plasmid	Primers
pET151- <i>glnB</i>	For_ <i>glnB</i> (5'-CACCATGACCGCGCTCCTGTTC-3')
	Rev_ <i>glnB</i> (5'-TCAATCGTAAGTGGTGCTGAAT-3')
pET151- <i>pfs</i>	For_ <i>pfs</i> (5'-CACCATGAAAATCGGCATCATTGGTG-3')
	Rev_ <i>pfs</i> (5'-TCAGCCATGTGCAAGTTTCTGCAC-3')

5.3.4.3 TA Cloning

The creation of the pETSUMO-*gbc* construct, for the overproduction of the *N*-terminal histidine and SUMO tagged protein, was achieved using the Champion™ pET SUMO Protein Expression System (Thermo Scientific) using the manufacturer's recommended procedure. The kit utilises TA cloning, which relies on the single A residues added to the 3' ends of the PCR product by *Taq* polymerase for ligation into the vector, which has single 3' T residue overhangs, so the PCR was performed using OneTaq Hot Start Quick-Load 2X Master Mix (GC buffer, NEB) as discussed in section 5.3.2, and the oligonucleotide primers used are shown in Table 5.7.

Table 5.7: Oligonucleotide primers used for the creation of the pETSUMO-*gbc* construct.

Plasmid	Primers
pETSUMO- <i>gbc</i>	For_pETSUMO-gbc
	(5'-AGCGTGACCGTGTGCGGAATC-3')
	Rev_pETSUMO-gbc
	(5'-TCATACCGTCACGCGTCC-3')

5.3.4.4 Transformation of *E.coli*

5.3.4.4.1 Chemical transformation

To chemically competent *E. coli* (50 µL) was added either the ligation reaction (10 µL, or all 6 µL for the TOPO ligation) or plasmid DNA (1 µL). The cells were then incubated on ice for 20 min before being heat-shocked at 42 °C for 30 seconds and once again returned to the ice for 20 min. 250 µL of LB was then added and the cells shaken, horizontally, for 1 h, before 50 and 100 µL were plated on LB agar plates containing the appropriate antibiotic, which were incubated at 37 °C overnight. For details on the purification of the resultant plasmids see section 5.3.1.

5.3.4.4.2 Electroporation

Prior to electroporation the ligation reactions were purified. The T4 DNA ligase was deactivated by heating at 65 °C for 20 min, before 1/10 of the volume of 3 M NaAc followed by 2.5X the volume of ice-cold 95 % EtOH was added. This was incubated at -20 °C for 5 min. The sample was then centrifuged (20 min, 13000 rpm, 4 °C) and the supernatant removed before 250 µL of ice-cold 70 % EtOH was added. After centrifugation (5 min, 13000 rpm, 4 °C) the supernatant was again removed and the precipitated DNA allowed to dry.

The purified ligation reactions were dissolved in 10 µL dH₂O, all of which was used in the transformation. For each transformation, the DNA was added to electrocompetent *E.coli* SY327 (60 µL), in an electroporation cuvette that had been cooled on ice. Electroporation was carried out at 1.8 kV, 100 Ω and 25 µF using a BioRAD Gene Pulsar® II in combination with a BioRad Pulse Controller Plus, and immediately after 500 µL of LB was added and the cells shaken, horizontally, for 45 min. Different volumes of the transformation reactions (50 and 100 µL) were plated across two LB agar plates (+ Tp). For details on the purification of the resultant plasmids see section 5.3.1

5.3.4.5 Sequencing

The presence of the insert(s) was confirmed for each plasmid using restriction digestion (section 5.3.4.1.1), which was analysed by agarose gel electrophoresis (section 5.3.3), and successful recombinant plasmids were sequenced by GATC biotech. The sequencing samples were prepared according to the company's requirements.

5.3.5 Mutagenesis

Site-directed mutagenesis for the GldM11OMTACP(H153A) mutant was performed using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs) following the manufacturers' recommended protocol. The pET24a-

GldM11OMTACP plasmid was used as the template for the PCR and the oligonucleotide primers used are shown in Table 5.8.

Table 5.8: Oligonucleotide primers used for the creation of the GldM11OMTACP(H153A) mutant, the mutated residues are shown in purple.

Primers	Mutation
For_GldM11OMTACP(H153A) (5'-CGGCGATTGCGCAGCTCACCTCGAA-3')	CAC → TGC
Rev_GldM11OMTACP(H153A) (5'-ACAAGACGGCCCTGTTCTCGAACATCC-3')	(His → Ala)

5.3.6 Gene deletion and complementation

5.3.6.1 Gene deletion

In-frame deletions in *B. gladioli* BCC1622 were introduced *via* double homologous recombination using the suicide plasmid pGPI and the I-SceI nuclease expression plasmid pDAI-SceI¹²⁵. To create the pGPI plasmids, 800-1000 bp sequences flanking each of the genes were amplified by PCR (section 5.3.2), using the oligonucleotide primers listed in Table 5.9, and the resultant PCR products were cloned into the pGPI vector (section 5.3.4.1). The ligation reactions were transformed into *E. coli* SY327 by electroporation (section 5.3.4.4.2) and the plasmids were purified and analysed as discussed in section 5.3.1 and 5.3.4.5, respectively.

Table 5.9: Oligonucleotide primers used for the creation of pGPI vectors, the restriction sites used are underlined.

Plasmid	Primers
pGPI- <i>gbnA</i>	For_ gbnA_5'+XbaI (5'-AAAGGGTCTAGATGCCAAGCGATGATGCGGTAA-3')
	Rev_ gbnA_5'+HindIII (5'-AAAGGGAAGCTTGCGGTCCGACAGGAATTGCTG-3')
	For_ gbnA_3'+HindIII (5'-AAGAGGAAGCTTCTGTCCGGAATCGTGTCTGTTTC-3')
	Rev_ gbnA_3'+KpnI (5'-AGGAAGGGTACCTCAACTGCTTGCCGAAGCCGA-3')
	For_ gbnB_5'+XbaI (5'-AAAGGGTCTAGATTTTGATCGATAAGCTCGCGT-3')
	Rev_ gbnB_5'+HindIII (5'-AAAGGGAAGCTTGAACCGATAACGGGATGCAAA-3')
pGPI- <i>gbnB</i>	For_ gbnB_3'+HindIII (5'-AAGAGGAAGCTTCGCTTGGTTCGAGGTCGGCTTC-3')
	Rev_ gbnB_3'+KpnI (5'-AAAGGGGGTACCCAGCCGCACGCTATCGGACAG-3')
pGPI- <i>gbnC</i>	For_ gbnC_5'+XbaI (*) (5'-AAAGGGTCTAGAGATATTCTCGACGCCGTC-3')
	Rev_ gbnC_5'+NdeI (*) (5'-GGAGAACATATGCAGCATCCCCTTCAGGCA-3')
	For_ gbnC_3'+NdeI (5'-AAAGGGCATATGGAGAACCAGACCTTGTTG-3')
	Rev_ gbnC_3'+EcoRI (5'-AAAGGGGAATTCGGATCACGCCATACACCT-3')
	For_ gbnM_5'+XbaI (5'-ATATCTAGAGGCGGCAGGCTTCGTCTCGCG-3')
	Rev_ gbnM_5'+HindIII (5'-ATAAAGCTTCGATTTCGAGGTCGAGTTCGG-3')
pGPI- <i>gbnM</i> (*)	For_ gbnM_3'+HindIII (5'-ATAAAGCTTATCGGCACGGTGCGAACG-3')
	Rev_ gbnM_3'+KpnI (5'-ATAGGTACCGCGACCAGGCGTGTCTCAGG-3')

* 8 % DMSO was required for the PCR reactions.

For each in-frame deletion, the appropriate pGPI plasmid was transformed into *E. coli* SY327, as described in section 5.3.4.4.2, and an overnight culture was grown (+Tp), along with overnight cultures of *B. gladioli* BCC1622 and *E. coli* HB101 +PRK2013 (+Kan), for triparental mating. The cultures were centrifuged (8 min, 4000 rpm) before the cell pellets were resuspended in 5 mL LB with 10 mM MgCl₂, this was then repeated to remove any traces of antibiotics. 100 µL of each re-suspension was then combined and mixed, before 100 µL was spread onto a nitrocellulose membrane (Millipore) placed on a LB agar plate with 10 mM MgCl₂ which was incubated at 30 °C overnight. The cell mixture was removed from the filter using 1 mL sterile 0.9 % NaCl and this was spread across several LB plates (+PMX₆₀₀ and Tp₁₅₀) which were left to grow for 2.5 days (30 °C). A control where each component was plated individually was also created to ensure they could not grow on the LB plates (+PMX₆₀₀ and Tp₁₅₀). Colony PCR was used to screen for successful single-crossover mutants and the oligonucleotide primers used can be found in Table 5.10 (section 5.3.2).

Table 5.10: Oligonucleotide primers used for colony PCR to screen for successful crossover mutants.

Mutant	Primers
<i>gbnA</i> (*)	For_ <i>gbnA</i> _5'
	(5'-TCCAGATCGGATATTTTCGGAA-3')
	Rev_ <i>gbnA</i> _3'
	(5'-TGATATTGCTGATGAACACCG-3')
<i>gbnB</i>	For_ <i>gbnB</i> _5'
	(5'-CGCGCCATGCAGTCATGATGG-3')
	Rev_ <i>gbnB</i> _3'
	(5'-GAACAGCGGACGCTTGCCGAA-3')
<i>gbnC</i>	For_ <i>gbnC</i> _5'
	(5'-TCGGCTTCGGCAAGCAGTTGA-3')
	Rev_ <i>gbnC</i> _3'
	(5'-AGCAATCGGCATGGTCGATGA-3')

<i>gbnM</i>	For_gbnM_5' (5'-GGCGTGGCCGGTGAAATCCT-3') Rev_gbnM_3' (5'-CGGCGGTCAGTTCCGGATA-3')
-------------	---

* No DMSO was required for the PCR reaction.

For the second homologous recombination this process was repeated but cultures of a successful mutant from the first step, *E. coli* SY327 +PDAI (+Tc) and *E. coli* HB101 +PRK2013 (+Kan) were used, with the mixture after triparental mating being plated on LB agar plates with the antibiotics, PMX₆₀₀ and Tc₂₀₀. Colony PCR, using the oligonucleotide primers listed in Table 5.10, was used to identify colonies carrying the appropriate gene deletion (section 5.3.2). The successful deletion mutants were cured of the pDAI plasmid by plating the cells on M9 minimal media¹⁵⁰. The effects of the gene deletions were analysed as described in section 5.6.3.3.

5.3.6.2 Gene complementation

For genetic complementation the genes were amplified by PCR (section 5.3.2) using the oligonucleotide primers listed in Table 5.11, and cloned into the pMLBAD vector as discussed in section 5.3.4.1. The resultant plasmids were introduced into *E. coli* SY327 using electroporation (section 5.3.4.4.2) and transferred into the *B. gladioli* BCC1622 mutants using triparental mating as discussed in section 5.3.6.1. Mutants carrying the complementation plasmid were selected for using LB agar plates supplemented with PMX₆₀₀ and Tp₁₅₀. As a control for each complementation reaction, the empty pMLBAD vector was also introduced into each mutant.

Table 5.11: Oligonucleotide primers used for the creation of pMLBAD vectors, the restriction sites used are underlined.

Plasmid	Primers
pMLBAD- <i>gbnM</i> (*)	For_pMLBAD- <i>gbnM</i> +NcoI
	(5'-AAAAA <u>CCATGG</u> ATGACTCCCGAACTCGACCTC-3')
	Rev_ <i>gbnM</i> +HindIII
	(5'-AAAAAA <u>AAGCTT</u> GCCGGCGTGTCAGTTGCC-3')
pMLBAD- <i>gbnC</i>	For_pMLBAD- <i>gbnC</i> +NcoI
	(5'-AAAGGG <u>CCATGG</u> GTGACCGTGTGCGGAATC-3')
	Rev_ <i>gbnC</i> +HindIII
	(5'-AAAAA <u>TAA</u> GCTTTCATACCGTCACGCGTCC-3')
pMLBAD- <i>sorQ</i>	For_pMLBAD- <i>sorQ</i> +NcoI
	(5'-AAAGGG <u>CCATGG</u> ATGTGCGGCATAGCGGGG-3')
	Rev_ pMLBAD- <i>sorQ</i> +HindIII
	(5'-AAAGGGA <u>AAGCTT</u> TCAGCGCTCTTCGAGCCC-3')

The ability of the complemented mutants to produce gladiolin was determined by high-resolution LC-MS, of the MeCN extracts from BSM-G, supplemented with 0.25 % arabinose, agar-grown cultures. The cultures were prepared as described in section 5.3.6.3.

5.3.6.3 Analysis of the metabolite profile of knockout mutants

To examine the effects of the gene deletions on gladiolin production and the results of complementation, overnight cultures of the *B. gladioli* BBCC1622 mutants and WT were grown, and the cells were pelleted by centrifugation (8 min, 4000 rpm) before being resuspended in 4 mL of sterile 0.9 % NaCl. For each strain, a sterile cotton bud was used to draw streaks across a BSM-G agar plate. The plates were incubated at 30 °C for 2.5 days. Prior to analysis the bacterial cells were removed from the plate and the solid medium chopped up into small pieces and placed in a glass vial. 4 mL of MeCN was added and this was shaken and left for 1 h, before the supernatant was filtered using 0.45 µm nylon micro-centrifugal filters (Thermo Scientific) and analysed by high-resolution LC-MS as described in section 5.3.6.4.

5.3.6.4 High-resolution LC-MS

High-resolution LC-MS was performed using a Bruker MaXis™ Impact Ultra-High Resolution ESI-Q-TOF instrument (ESI in positive ion mode; full scan 50-2500 m/z; end plate offset, -500 V; capillary, -4500 V; nebulizer gas (N₂), 1.4 bar; dry gas (N₂), 8 L/min and dry temperature, 200 °C) connected to a Dionex UltiMate 3000 UHPLC fitted with a ZORBAX Eclipse Plus C18 column (1.8 µm, 2.1 x 100 mm, Agilent). The UHPLC elution profile is detailed in Table 5.12 and the absorbance was monitored at 210 and 254 nm. Calibration was performed with 10 mM sodium formate through a loop injection of 20 µL at the start of each run.

Table 5.12: Elution conditions for high-resolution LC-MS analysis.

Time (mins)	H₂O / 0.1 % FA (%)	MeCN / 0.1 % FA (%)	Flow rate (mL/min)
0.0	95	5	0.2
5.3	95	5	0.2
17.3	0	100	0.2
22.3	0	100	0.2
25.3	95	5	0.2
33.3	95	5	0.2

5.4 Isolation and characterisation of amide gladiolin

5.4.1 Production and purification of gladiolin and amide gladiolin

For the production of amide gladiolin and gladiolin, 5 mL overnight cultures of *B. gladioli* BCC1622Δ*gbnM* and *B. gladioli* BCC1622 WT, respectively, were pelleted by centrifugation (8 min, 4000 rpm) and the pellets resuspended in 4 mL of sterile 0.9 % NaCl. For each strain, a sterile cotton bud was used to draw streaks across 30 BSM-G agar plates (1 L of medium) which were incubated at 30 °C for 2.5 days. Prior to extraction the bacterial cells were

removed from the plates and the solid medium was chopped into small pieces. To extract the metabolites of interest the small pieces of solid medium were covered with ethyl acetate. This was left for 2 h at RT before being filtered and dried *in vacuo*. The solid medium was again covered with ethyl acetate and the extraction process repeated, yielding a dark brown crude extract.

High-performance liquid chromatography (HPLC) was used to purify the antibiotics, and was carried out using an Agilent Technologies 1260 Infinity LC system fitted with an ELSD and fraction collector. The crude extracts were first dissolved in a minimum volume of MeOH and loaded onto dental cotton tubes that were dried *in vacuo*. The dried dental cotton was put into a cartridge before being loaded onto a BETASIL C18 column (150 x 21.2 mm, Thermo Scientific). The HPLC elution profile is detailed in Table 5.13. The fractions containing the pure antibiotics were identified using high-resolution LC-MS (section 5.3.6.4), and the desired fractions for amide gladiolin were dried using a Genevac EZ-2 (SP Scientific), and for gladiolin using a ScanVac CoolSafe freeze dryer (LaboGene). This yielded amide gladiolin (14 mg) and gladiolin (18 mg) as white solids.

Table 5.13: Elution conditions for HPLC purification of amide gladiolin and gladiolin.

Time (mins)	H ₂ O (%)	MeCN (%)	Flow rate (mL/min)
0	65	35	9
5	65	35	9
50	35	65	9
51	0	100	9
55	0	100	9
56	65	35	9
60	65	35	9

5.4.2 NMR spectroscopy

High-field 1- and 2-D NMR spectra for structure elucidation were recorded using a Bruker Avance III 600 MHz instrument. The instrument operated at 600 MHz for ^1H NMR and 150 MHz for ^{13}C NMR with all spectra being recorded at 298 K. Purified amide gladiolin (14 mg) was dissolved in 230 μL of $\text{DMSO}-d_6$ (Sigma Aldrich) and the ^1H and ^{13}C NMR chemical shifts were referenced to the solvent peaks at δ_{H} 2.50 and δ_{C} 39.5, respectively. For iso-amide gladiolin the NMR analysis was performed using 5 mg dissolved in 230 μL of $\text{DMSO}-d_6$. Bruker Topspin 2.1 software was used to process the spectra. The assignments for amide gladiolin can be found in Table 3.1 and the assignments for *iso*-amide gladiolin can be found in table S1 (appendix 8).

5.4.3 Stability analysis

Amide gladiolin and gladiolin were dissolved in MeOH (1 mg/mL) and left at 15 °C for 0, 6, 24, 48 and 96 h, prior to high-resolution LC-MS analysis (section 5.3.6.4).

5.5 Protein overproduction, purification & characterisation

5.5.1 Protein overproduction

For protein overproduction, 1 μL of plasmid DNA was used to chemically transform the chosen *E.coli* strain (section 5.3.4.4.1). One colony was picked and a 10 mL overnight culture grown, which was used to inoculate 1 L of LB (in a 5 L conical flask) that was supplemented with the appropriate antibiotic. This was grown until the $\text{OD}_{600} = 0.6\text{--}0.8$. The culture was then cooled before the inducer, IPTG (1 M stock dissolved in sterile dH_2O , stored at -20 °C), was added. The culture was left overnight at 15 °C (180 rpm). A comprehensive

list of the *N*-terminal his-tagged proteins produced as part of this study, and details of the precise expression conditions can be found in Table 5.14.

Table 5.14: Overproduction conditions for recombinant proteins. Rows in blue indicate that the overproduced protein was insoluble.

Protein	Plasmid (Source)	Resistance	<i>E.coli</i> strain	IPTG Conc.
GldM11OMTACP	pET24a- GldM11OMTACP (Synthetic construct*)	Kan	BL21	1 mM
GldM11OMTACP (H153A)	pET24a- GldM11OMTACP (H153A) (This study)	Kan	BL21	1 mM
GldM11OMT	pET28a- GldM11OMT (This study)	Kan	BL21	0.5 mM
GldM11KSOMT	pET28a- GldM11KSOMT (This study)	Kan	BL21	0.5 mM
GldM11KSOMTACP	pET28a- GldM11KSOMTACP (This study)	Kan	BL21	0.5 mM
MisM9OMTACP	pET28a- MisM9OMTACP (This study)	Kan	BL21	0.5 mM
MisM1OMTACP	pET24a- MisM1OMTACP (Synthetic construct*)	Kan	BL21	0.5 mM
RizM18OMTACP	pET28a- RizM18OMTACP (This study)	Kan	BL21	0.5 mM
GldM11ACP(I)	pET28a- GldM11ACP(I) (This study)	Kan	C43	0.5 mM
GldM11ACP(II)	pET28a- GldM11ACP(II)	Kan	BL21	1 mM

	(This study)			
MisM9ACP(II)	pET28a-MisM9ACP(II) (This study)	Kan	BL21	0.5 mM
MisM1ACP(II)	pET28a-MisM1ACP(II) (This study)	Kan	BL21	0.5 mM
RizM18ACP(II)	pET28a-RizM18ACP(II) (This study)	Kan	BL21	0.5 mM
MisM1 OMT	pET28a-MisM1 OMT (This study)	Kan	BL21	0.5 mM
Pfs	pET151- <i>pfs</i> (This study)	Amp	BL21	1 mM
PanK	pET29b- <i>panK</i> (Tosin <i>et al.</i> ¹⁵²)	Kan	BL21	0.5 mM
PPAT	pET29b-PPAT (Tosin <i>et al.</i> ¹⁵²)	Kan	BL21	0.5 mM
DPCK	pET20b(+)-DPCK (Tosin <i>et al.</i> ¹⁵²)	Amp	BL21	1 mM
Sfp	pET28a- <i>sfp</i> (Kindly provided by Dr M. Tosin)	Kan	BL21	1 mM
GbnM	pET28a- <i>gbnM</i> (This study)	Kan	BL21	0.25 mM
GbnC	pET28a- <i>gbnC</i> (This study)	Kan	BL21/ C43	0.1-0.5 mM
	H-MBP-3C- <i>gbnC</i> (This study)	Amp		0.5 mM
	pETSUMO- <i>gbnC</i> (This study)	Kan		0.5 mM
SorQ	pET24a- <i>sorQ</i> (Synthetic construct*)	Kan	BL21/ C43	0.5 mM
GbnB	pET151- <i>gbnB</i> (This study)	Amp	C43	0.5 mM
EtnK	pET24a- <i>etnK</i>	Kan	BL21	1 mM

	(Synthetic construct*)			
GbnA	pET28a- <i>gbnA</i> (This study)	Kan	BL21	0.5 mM
GldM1ACP(I)	pET28a-GldM1ACP(I) (This study)	Kan	BL21	1 mM
GldM1KS	pET28a-GldM11KS (This study)	Kan	BL21/ C43	0.5 mM

* Synthetic constructs were purchased from Epoch Life Sciences and were designed with an octa-histidine tag.

5.5.2 Protein purification

The cells were harvested from the culture using centrifugation (30 min, 5000 rpm, 4 °C) and the cell pellets were resuspended in loading buffer (10 mL/L). The cells were lysed using a Constant Systems Ltd TS-Series Cabinet cell disruptor, set to a pressure of 20 psi, and the cell debris was removed by centrifugation (40 min, 17000 rpm, 4 °C). The supernatant was filtered using a 0.45 µm syringe filter (GE Healthcare), before being applied to a 1 mL HiTrap™ HP affinity column (GE Healthcare), which was washed with 15 mL loading buffer. The protein was then eluted from the column in a stepwise approach by first using 50 mM imidazole buffer (5 mL), followed by 100 mM imidazole buffer (3 mL), 200 mM imidazole buffer (3 mL) and finally 300 mM imidazole buffer (3 mL).

The fractions containing the protein (as determined by SDS-PAGE analysis, section 5.5.3) were pooled and concentrated to 1-2 mL, using an Amicon® Ultra Centrifugal Filter (Millipore) with the appropriate molecular weight cut-off (4000 rpm, 4 °C), before being buffer exchanged by the addition of 15 mL of storage buffer and concentrated to approximately 250-500 µL. Glycerol was added to give a 10 % final concentration and the proteins were divided into aliquots (30-50 µL), flash frozen in liquid nitrogen and stored at -80 °C.

5.5.3 SDS-PAGE analysis

SDS-PAGE gels were prepared as shown in Table 5.15.

Table 5.15: Volumes of reagents used in the preparation of SDS-PAGE gels.

Reagent	Stacking gel (1 mL)	Resolving gel (5 mL)			
	Volume (mL)	8 % Volume (mL)	10 % Volume (mL)	12 % Volume (mL)	15 % Volume (mL)
H ₂ O	0.68	2.3	1.9	1.6	1.1
30 % Acrylamide mix	0.17	1.3	1.7	2.0	2.5
1.0 M Tris (pH 6.8)	0.13	-	-	-	-
1.5 M Tris (pH 6.8)	-	1.3	1.3	1.3	1.3
10 % SDS	0.01	0.05	0.05	0.05	0.05
10 % APS	0.01	0.05	0.05	0.05	0.05
TEMED	0.001	0.003	0.002	0.002	0.002

Protein samples were prepared by mixing 20 μ L of each fraction (pure protein samples were first diluted to 100 ng/ μ L with loading buffer) with 5 μ L 5X SDS-PAGE loading buffer. 10 μ L of each sample was loaded onto the gel, along with PageRuler™ Plus Prestained Protein Ladder (4 μ L, Thermo Scientific) which was used as a molecular weight standard. Gels were run using a Mini-PROTEAN® Tetra cell (Bio-Rad, 180 V, 50 min), stained with EZBlue™ gel staining reagent (Sigma Aldrich) for 30 min and destained overnight in dH₂O.

5.5.4 Determining protein concentration

In general protein concentrations were determined by measuring the absorbance at 280 nm, using a Thermo Scientific™ NanoDrop™ Lite spectrophotometer, and calculating the concentration using the equation:

$$C = \frac{A}{\epsilon l}$$

A is the absorbance, l is the pathlength (1 cm) and ϵ is the extinction coefficient as determined by the ExPASy ProtParam tool.

For GldM11ACP(II) and MisM9ACP(II) the proteins did not have extinction coefficients, so the protein concentrations were determined using the Bradford assay. The Quick Start™ Bradford Protein Assay Kit (BIO-RAD) was used according to the manufacturer's recommended protocol using BSA as the protein standard. A Thermo BioMate™ 3 UV-Vis spectrophotometer was used to measure the absorbances at 595 nm.

5.5.5 Determining protein oligomerisation state

To determine protein oligomerisation states SEC, which separates proteins in solution based on their size, was used. SEC was performed using an AKTA pure chromatography system fitted with a GE Healthcare HiLoad® 16/600 Superdex® 200 pg column. Gel filtration buffer was used as the mobile phase (1.2 column volumes per run) with a flow rate of 1 mL/min. A protein standard consisting of 6 mg of thyroglobin (669 kDa), 6 µL of apoferritin (443 kDa), 3 mg of β -amylase (200 kDa), 1 mg of BSA (66 kDa) and 1 mg of lysozyme (14.3 kDa) was made up in 200 µL of gel filtration buffer and was run first through the column to give a comparison between protein mass and retention time under the conditions of the experiment. Approximately 3 mg of the protein of interest (diluted to 200 µL using gel filtration buffer) was then run on the column.

5.5.6 Intact protein MS

All proteins were diluted with HPLC grade H₂O to a concentration of 10-20 μ M for analysis on the Bruker MaXis™ II Ultra-High Resolution ESI-Q-TOF instrument (ESI in positive ion mode; full scan 50-2500 m/z; end plate offset, -500 V; capillary, -4500 V; nebulizer gas (N₂), 1.4 bar; dry gas (N₂), 8 L/min and dry temperature, 200 °C) connected to a Dionex 3000 RS UHPLC fitted with an ACE C4-300 RP column (100 x 2.1 mm, 5 μ m, 30 °C). The UHPLC elution profile is detailed in Table 5.16.

Table 5.16: Elution conditions for intact protein MS.

Time (mins)	H ₂ O / 0.1 % FA (%)	MeCN / 0.1 % FA (%)	Flow rate (mL/min)
0.0	95	5	0.2
5.0	95	5	0.2
35.0	0	100	0.2
40.0	0	100	0.2
45.0	95	5	0.2

5.6 Protein crystallisation trials

5.6.1 Protein overproduction and purification

For protein crystallisation trials a large volume of high-purity protein was required. Protein overproduction was carried out on a large scale (3 L, 1 L per 5 L flask), but was otherwise performed as discussed in section 5.5.1. The cells were harvested by centrifugation (30 min, 5000 rpm, 4 °C) and the cell pellets were resuspended in HEPES buffer A (10 mL/L of culture), before the resuspended cells were added to a SONOPULS rosett cell RZ 3 (BANDELIN) which was placed on ice. The cells were lysed by sonication (amplitude: 50, 3 min, 3 pulsar, X3) and the lysate was clarified by centrifugation (40 min, 17000

rpm, 4 °C). The cell free extract was then passed through a 0.45 µm syringe filter.

Protein purification was performed using an AKTA pure chromatography system, fitted with a F9-C fraction collector, using a flow rate of 1 mL/min. First, the protein was loaded onto a 1 mL HiTrap™ HP affinity column, which was washed with 25 mL 10 % HEPES buffer B (90 % HEPES buffer A). The protein was then eluted from the column by increasing the concentration of HEPES buffer B to 100 % over 15 mL (collecting 2 mL fractions). The fractions containing the protein (as determined by SDS-PAGE analysis, section 5.5.3) were pooled and concentrated to 2 mL, using an Amicon® Ultra Centrifugal Filter (Millipore) with the appropriate molecular weight cut-off (4000 rpm, 4 °C). The protein was then further purified and buffer exchanged using SEC (see section 5.5.5, crystallisation buffer). The protein was loaded onto the SEC column using a 2 mL loop and 2 mL fractions were collected when a peak was detected. The fractions containing the protein were concentrated to 10 mg/mL. The protein was stored at 4 °C between experiments, and some protein was flash frozen in liquid nitrogen and stored at -80 °C for use in assays.

5.6.2 Crystallisation trials

To the purified protein, SAM (150 mM, dissolved in 1 M HEPES, pH 7.5) was added to give a final concentration of 2 mM. The protein was then used to set up 100 nL sitting drop crystallisation trials using a mosquito® LCP robot (ttp:labtech). Seven 96-well commercially-available crystallisation screens (JCSG cores I-IV, PACT premier, Morpheus® and SG1™, purchased from Molecular Dimensions) were trialled, testing both a 1:1 protein: buffer ratio and a 2:1 protein: buffer ratio. The crystallisation trials were incubated at both RT and 18 °C.

5.6.3 Cleavage of the hexahistidine tag

The His-tag was cleaved from GldM11 OMT after purification by Ni-NTA (which was carried out as described in section 5.6.1). The protein (3 mL) was incubated with thrombin protease (1 U/ μ L, Sigma Aldrich) so that there was 1 U/ 100 μ g of protein to be cleaved, and left for 16 h at 4 °C. Successful cleavage was confirmed by SDS-PAGE analysis (section 5.5.3) and the cleavage reaction was passed over a column (0.5 mL/min flow rate) packed with half benzamidine sepharose (which binds thrombin protease) and half Ni-NTA (which binds any protein where the His-tag had not been cleaved), washing with 10 mL HEPES buffer A. The flow through was concentrated using an Amicon® Ultra Centrifugal Filter (Millipore) with the appropriate molecular weight cut-off (4000 rpm, 4 °C), before the protein was purified to homogeneity by SEC chromatography as described in section 5.6.1, and used in crystallisation trials as discussed in section 5.6.2.

5.7 Assays

5.7.1 Biochemical assays

5.7.1.1 Methylation assays

The composition of the methylation assays is shown in Table 5.17. In order to prevent a significant lowering of the pH of the assay, SAM iodide (Sigma-Aldrich) was dissolved in 300 mM phosphate buffer (pH 7.8) to give a 20 mM stock solution (stored at -80 °C) and the 80 mM aqueous ATP stock solution was adjusted to pH 7 using NaOH. All components of the assay except for SAM were added to the reaction, with the pantetheine substrate (dissolved in DMSO) being added last. The assay was mixed by pipetting up and down and incubated for 1 h at RT to allow the substrate to be loaded onto the ACP domain. SAM was then added, and the assay mixed before being incubated

for a further 1 h 30 min (RT). For each pantetheine substrate, and for each OMT-ACP didomain, the assay was repeated in triplicate with controls being performed by adding the appropriate volume of storage buffer in place of SAM.

Table 5.17: The composition of the methylation assays. The * indicates that the assay component acts catalytically.

Component	Concentration	Volume
OMT-ACP Protein	100 μ M	10-14 μ L
Pantetheine substrate	800 μ M	2 μ L
PanK	8 μ M *	1 μ L
PPAT	8 μ M *	1 μ L
DPCK	8 μ M *	5 μ L
ATP	4 mM	2.5 μ L
Sfp	10 μ M *	2 μ L
MgCl ₂	10 mM	5 μ L
SAM	600 μ M	1.5 μ L
Storage buffer	-	to 50 μ L

The assays were then diluted with HPLC grade water, so that the final concentration of the OMT-ACP protein was 10-18 μ M, before being analysed using intact protein MS (see section 5.5.6).

5.7.1.2 Transacylation assay

First, the *apo*-ACP domains, GldM11ACP(I) and GldM11ACP(II), were converted to (*S*)-3-hydroxyoctanoyl- *holo*-GldM11ACP(I) and *holo*-GldM11ACP(II), respectively. The reactions were set up as shown in Table 5.18, and incubated at RT for 1 h 30 min. Following incubation, excess CoA or pantetheine substrate were removed by buffer exchanging into storage buffer using Amicon Ultra centrifugal filters with a 3 kDa MWCO membrane (Millipore), and concentrated to 50 μ L.

Table 5.18: The composition of the ACP domain loading reactions. The * indicates that the assay component acts catalytically.

Loading of GldM11ACP(I)		Loading of GldM11ACP(II)	
Component	Volume	Component	Volume
ACP protein (200 μ M)	71 μ L	ACP protein (200 μ M)	25 μ L
3-hydroxyoctyl pantetheine (600 μ M)	2 μ L	CoA (500 μ M)	1.26 μ L
PanK (10 μ M *)	1.5 μ L	Sfp (10 μ M *)	4 μ L
PPAT (10 μ M *)	1.5 μ L	MgCl ₂ (10 mM)	10 μ L
DPCK (10 μ M *)	5 μ L		
ATP (4 mM)	5 μ L		
Sfp (10 μ M *)	4 μ L		
MgCl ₂ (10 mM)	10 μ L		
Storage buffer	to 100 μ L	Storage buffer	to 100 μ L

The transacylation assay was then carried out by combining (*S*)-3-hydroxyoctanoyl- *holo*-GldM11ACP(I) (100 μ M) and *holo*-GldM11ACP(II) (100 μ M) with GldM11KSOMT (25 μ M) in storage buffer (100 μ L final volume). A control reaction was performed by adding the appropriate volume of storage buffer in place of GldM11KSOMT. The reactions were incubated at RT for 3 h before the reaction mixtures were diluted 10-fold in HPLC grade water for analysis by intact protein MS (see section 5.5.6). The assay was repeated in triplicate.

5.7.1.3 ACP-OMT interaction assays

For the ACP-OMT interaction assays the *apo*-ACP domains were loaded with either (*S*)-3-hydroxyoctanoyl- or (*R*)-3-hydroxyoctanoyl-pantetheine to assess interactions with GldM11OMT and MisM1OMT, respectively. The ACP domains were loaded with the pantetheine substrate as shown in Table 5.18 for GldM11ACP(I). The *holo*-form of the ACP domains (40 μ M) were then incubated with the OMT domain (80 μ M), the *S*-adenosylhomocysteine

nucleosidase, Pfs (1 μ L), to prevent potent inhibition of the OMT domain by SAH, and SAM (600 μ M) in storage buffer (total volume 50 μ L), for 2 h at RT. The assays were diluted with HPLC grade water to give a final ACP domain concentration of 10 μ M, before being analysed by intact protein MS (see section 5.5.6). Control reactions were performed by adding the appropriate volume of storage buffer in place of SAM. Where no ACP-OMT domain interactions could be observed, the assays were repeated with 160 μ M OMT domain, to ensure that the assays contained no potential bias that could occur due to any error in the calculated protein concentrations.

5.7.1.4 GbnM activity assay

This assay was conducted using cell lysate. To prepare the cell lysate pET28a-*gbnM* and the empty expression vector, pET28a, were separately transformed into *E. coli* BL21 (section 5.3.4.4.1). Protein overproduction was performed on a small scale (10 mL LB in a sterile 50 mL Falcon™ tube), but was otherwise performed as discussed in section 5.5.1. IPTG was also added to the pET28a control for consistency. The cells were collected from the cultures by centrifugation (15 min, 4000 rpm) and the cell pellets were resuspended in 400 μ L storage buffer prior to the cells being lysed by sonication (amplitude: 30, 2 min, 2 pulsar, on ice). The resultant lysates were used directly in the assay and 10 % SDS-PAGE analysis was performed to assess the composition of the lysates (section 5.5.3).

For the assay, each of the lysates (5 μ L, 10 % of the final volume) was combined with amide gladiolin, which was dissolved in DMSO (2 μ L, 800 μ M) and made up to a total volume of 50 μ L with storage buffer. Standards of amide gladiolin and gladiolin were created by adding the antibiotic to storage buffer in the absence of lysate. The reactions were incubated at RT for 3 h, before 100 μ L of MeCN was added to precipitate the proteins, which were removed by centrifugation (5 min, 13000 rpm). The supernatants were filtered using

0.45 μ m nylon micro-centrifugal filters (Thermo Scientific™), before being analysed by high-resolution LC-MS (section 5.3.6.4).

5.7.1.5 AT domain loading assays

Stocks of acetyl-, succinyl- and malonyl-CoA (20 mM in HPLC grade water) were flash frozen and stored at -80 °C. 20 mM succinamic acid was also prepared in HPLC grade water and was pH adjusted, using concentrated NaOH (pH 7). Succinamic-CoA was made by combining the CoA transferase, VbxP (150 μ M, supplied by Christian Hobson), acetyl-CoA (1.5 mM), succinamic acid (6 mM) and storage buffer in a final volume of 100 μ L. A boiled enzyme control (10 min, 95 °C) was also performed. After overnight incubation at RT, 30 μ L of the assay was added to 60 μ L MeOH to precipitate the enzyme, and the precipitant was removed by centrifugation (5 min, 13000 rpm). Complete conversion of acetyl-CoA to succinamic-CoA was confirmed by high-resolution LC-MS (section 5.3.6.3). VbxP was removed from the remainder of the assay using an Amicon Ultra-0.5 mL centrifugal filter with a 3 kDa MWCO (Millipore) and the flow through was used directly in future assays.

To investigate the substrate tolerance of GbnB, compared to the malonyl-specific AT domains EtnK and PksC (supplied by Dr Mathew Jenner), the AT domains (60 μ M) were incubated with either succinyl-CoA, succinamic-CoA or malonyl-CoA (240 μ M) in storage buffer (50 μ L final volume). After 15 min, the assays were diluted with HPLC grade water, so that the final concentration of the AT domains was 15 μ M, before being analysed by intact protein MS (section 5.5.6). Control reactions were performed by eliminating the CoA derivatives, which for succinamic-CoA involved using the supernatant of the boiled VbxP reaction to ensure that succinamic acid was not capable of self-loading onto the AT domains.

5.7.2 Antimicrobial activity assays

5.7.2.1 Broth microdilution method

In general MICs were determined using the broth microdilution method (National Committee for Clinical Laboratory Standards, document M7-A7¹⁵³). Overnight cultures of the test strains (refer to Table 5.1) were grown in 5 mL Mueller Hinton media at 30 °C. The antibiotics were dissolved in DMSO to a concentration of 5 mg/mL. In a 96-well microtiter plate 50 µL of serial 2-fold dilutions of the antibiotic (128 µg/mL - 0.0025 µg/mL) in Mueller Hinton media were mixed with 50 µL of microbial suspension that was made up, from the overnight cultures, to McFarland standard No. 0.5 and then diluted 100-fold, also in Mueller Hinton media. The assays were incubated for 16 h at 30 °C, and the MIC was defined as the lowest concentration that visibly inhibited bacterial growth. Each assay was performed in triplicate.

5.7.2.2 Resazurin microtiter assay

MICs against *M. smegmatis* were determined using the rezasurin microtiter assay¹³⁰. A culture of *M. smegmatis* was grown in mycobacteria media for 2 days at 37 °C while being shaken at 220 rpm. In a 96-well microtiter plate 50 µL of a bacterial suspension made up to McFarland standard No. 1, in 7H-9 media and diluted 20-fold was mixed with 50 µL of serial 2-fold dilutions of the antibiotic (128 µg/mL - 0.0025 µg/mL) also in 7H-9 media (initially the antibiotic was dissolved in DMSO to a concentration of 5 mg/mL). The assays were incubated at 37 °C for 2.5 days. 15 µL of resazurin solution (Acros Organics, made to a 0.01 % weight/volume solution in sterile water) was then added to each well and incubated overnight, the MIC was defined as the lowest concentration to prevent a complete colour change from blue to pink. Each assay was performed in triplicate and controls were carried out to ensure that no components of the assay other than *M. smegmatis* could induce a colour change.

5.8 Computational methods

5.8.1 ClusterTools searches

ClusterTools⁷⁸ was used to search for embedded OMT domains from PKSs. First the MiBIG database⁷⁹ (constructed on 03-09-2016 by Dr Emzo de los Santos) was searched for embedded OMT domains from *trans*-AT PKSs using the custom HMM rule 'OMT and PKS_KS and not PKS_AT' using the PKS annotation HMMs from antiSMASH⁷². The default parameters were not able to clearly distinguish OMT domains from CMT domains, so the HMM search E-value was set to 1.0e-30. Using the adjusted parameters this database was then searched for embedded OMT domains from *cis*-AT PKS by specifying that an OMT, PKS_KS and PKS_AT were all required. To look for additional and cryptic clusters containing embedded OMT domains these two searches were then performed on a database made from all reference and representative genomes from NCBI (created on 05-12-2017 by Dr Emzo de los Santos).

ClusterTools was also used to search these two databases for BGCs that contained homologues of the putative asparagine synthetase, *gbc*, and the putative amidase *gbcM*, from the gladiolin biosynthetic pathway. To do this the protein sequences of the genes were manually imputed to be used in a query that specified that both of these genes were required. Using the default parameters ClusterTools identified three biosynthetic clusters, the etnangien⁶⁶, sorangicin¹⁴⁷ and fosfazinomycin¹²³ clusters, from the MiBIG database. When searching the NCBI database it was also specified that a PKS_KS and *trans*-AT docking domain were required; no additional BGCs were identified.

5.8.2 Sequence alignments

All multiple sequence alignments were generated using ClustalX 2.1 by performing a complete alignment with default parameters¹⁰¹. The protein sequences used were retrieved from either the MIBiG repository⁷⁹ or NCBI databases.

5.8.3 Phylogenetic analysis

The phylogenetic trees presented in this study were also created using ClustalX 2.1¹⁰¹, from the generated sequence alignments using the neighbour joining clustering algorithm. The resultant phylogenetic trees were visualised using FigTree v1.4.3, from which the presented images were generated.

5.8.4 Homology modelling of GldM11 OMT

The homology model of GldOMT was constructed using the intensive modelling mode of Phyre²⁸⁹. The published structures of the cyclopropane fatty acid synthase (6BQC)¹⁵⁴, pavine (5KN4)¹⁵⁵ and coclaurine (6GKV)¹⁵⁶ MMTs and the mycolic acid cyclopropane synthases CmaA1 (1KPG), CmaA2 (1KPI) and PcaA (1L1E)¹⁵⁷ were used as templates. To assess the quality of the model a Ramachandran plot was produced using the RAMPAGE server. SAH was added to the active site by aligning the model to the published structure of the coclaurine MMT, that was in complex with SAH using PyMOL¹⁰⁰.

References

- 1 E. A. Campbell, O. Pavlova, N. Zenkin, F. Leon, H. Irschik, R. Jansen, K. Severinov and S. A. Darst, *EMBO J.*, 2005, **24**, 674-682.
- 2 R. Gaynes, *Emerg. Infect. Dis.*, 2017, **23**, 849-853.
- 3 A. Fleming, *Br. J. Exp. Pathol.*, 1929, **10**, 226-236.
- 4 R. Aminov, *Biochem. Pharmacol.*, 2017, **133**, 4-19.
- 5 K. Lewis, *Nat. Rev. Drug Discov.*, 2013, **12**, 371-387.
- 6 A. Schatz, E. Bugie and S. A. Waksman, *Clin. Orthop. Relat. Res.*, 1944, **55**, 66-69.
- 7 A. Tomasz, *Annu. Rev. Microbiol.*, 1979, **33**, 113-137.
- 8 D. Kahne, C. Leimkuhler, W. Lu and C. Walsh, *Chem. Rev.*, 2005, **105**, 425-448.
- 9 E. A. Campbell, N. Korzheva, A. Mustaev, K. Murakami, S. Nair, A. Goldfarb and S. A. Darst, *Cell*, 2001, **104**, 901-912.
- 10 I. Chopra and M. Roberts, *Microbiol. Mol. Biol. Rev.*, 2001, **65**, 232-260.
- 11 J. M. Munita, C. A. Arias, A. R. Unit and A. De Santiago, *Microbiol. Spectr.*, 2016, **4**, VMBF-0016-2015.
- 12 X. Jiang, M. M. H. Ellabaan, P. Charusanti, C. Munck, K. Blin, Y. Tong, T. Weber, M. O. A. Sommer and S. Y. Lee, *Nat. Commun.*, 2017, **8**, 15784.
- 13 J. M. A. Blair, M. A. Webber, A. J. Baylay, D. O. Ogbolu and L. J. V. Piddock, *Nat. Rev. Microbiol.*, 2011, **13**, 42-51.
- 14 G. D. Wright, *Chem. Commun.*, 2011, **47**, 4055-4061.
- 15 A. Milshteyn, J. S. Schneider and S. F. Brady, *Chem. Biol.*, 2014, **21**, 1211-1223.
- 16 J. Handelsman, *Microbiol. Mol. Biol. Rev.*, 2005, **68**, 669-685.
- 17 D. J. Bevitt, J. Cortes, S. F. Haydock and P. F. Leadlay, *Eur. J. Biochem.*, 1992, **204**, 39-49.
- 18 T. W. Yu, Y. Shen, Y. Doi-Katayama, L. Tang, C. Park, B. S. Moore, C. Richard Hutchinson and H. G. Floss, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9051-9056.
- 19 L. B. Pickens and Y. Tang, *Metab. Eng.*, 2009, **11**, 69-75.

- 20 H. Jenke-Kodama, A. Sandmann, R. Müller and E. Dittmann, *Mol. Biol. Evol.*, 2005, **22**, 2027-2039.
- 21 E. Schweizer and J. Hofmann, *Microbiol. Mol. Biol. Rev.*, 2004, **68**, 501-517.
- 22 T. Robbins, Y.-C. Liu, D. E. Cane and C. Khosla, *Curr. Opin. Struct. Biol.*, 2016, **41**, 10-18.
- 23 C. Hertweck, A. Luzhetskyy, Y. Rebets and A. Bechthold, *Nat. Prod. Rep.*, 2007, **24**, 162-190.
- 24 J. A. Chemler, T. J. Buchholz, T. W. Geders, D. L. Akey, M. Rath, G. E. Chlipala, J. L. Smith and D. H. Sherman, *J. Am. Chem. Soc.*, 2012, **134**, 7359–7366.
- 25 B. Shen, *Curr. Opin. Chem. Biol.*, 2003, **7**, 285-295.
- 26 H. Chen and L. Du, *Appl. Microbiol. Biotechnol.*, 2016, **100**, 541-557
- 27 V. Y. Alekseyev, C. W. Liu, D. E. Cane, J. D. Puglisi and C. Khosla, *Protein Sci*, 2007, **16**, 2093-2107.
- 28 J. Beld, E. C. Sonnenschein, C. R. Vickery, J. P. Noel and M. D. Burkart, *Nat. Prod. Rep.*, 2014, **31**, 61-108.
- 29 A. T. Keatinge-Clay, *Chem. Rev.*, 2017, **117**, 5334-5366.
- 30 Y. A. Chan, A. M. Podevels, B. M. Kevany and M. G. Thomas, *Nat. Prod. Rep.*, 2009, **26**, 90-114.
- 31 B. J. Dunn, D. E. Cane and C. Khosla, *Biochemistry*, 2013, **52**, 1839-1841.
- 32 T. Robbins, J. Kapilivsky, D. E. Cane and C. Khosla, *Biochemistry*, 2016, **55**, 4476-4484.
- 33 A. T. Keatinge-Clay, *Chem. Biol.*, 2007, **14**, 898-908.
- 34 P. Caffrey, *ChemBioChem*, 2003, **4**, 654-657.
- 35 R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. Richard Hutchinson and R. McDaniel, *Biochemistry*, 2003, **42**, 72-79.
- 36 J. Zheng, S. K. Piasecki and A. T. Keatinge-Clay, *ACS Chem. Biol.*, 2013, **8**, 1964-1971.
- 37 D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick and J. L. Smith, *Structure*, 2010, **18**, 94-105.
- 38 J. F. Barajas, J. M. Blake-Hedges, C. B. Bailey, S. Curran and J. D. Keasling, *Synth. Syst. Biotechnol.*, 2017, **2**, 147-166.
- 39 J. Zheng, D. C. Gay, B. Demeler, M. A. White and A. T. Keatinge-Clay, *Nat. Chem. Biol.*, 2012, **8**, 615-621.
- 40 J. He and C. Hertweck, *Chem. Biol.*, 2003, **10**, 1225-1232.
- 41 H. Y. He, H. X. Pan, L. F. Wu, B. B. Zhang, H. B. Chai, W. Liu and G. L.

- Tang, *Chem. Biol.*, 2012, **19**, 1313-1323.
- 42 M. E. Horsman, T. P. A. Hari and C. N. Boddy, *Nat. Prod. Rep.*, 2016, **33**, 183-202.
- 43 J. Staunton and B. Wilkinson, *Chem. Rev.*, 1997, **97**, 2611-2630.
- 44 U. Rix, C. Fischer, L. L. Remsing and J. Rohr, *Nat. Prod. Rep.*, 2002, **19**, 542-580.
- 45 C. Olano, C. Méndez and J. A. Salas, *Nat. Prod. Rep.*, 2010, **27**, 571-616.
- 46 T. A. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck and J. Piel, *Nat. Biotechnol.*, 2008, **26**, 225-233.
- 47 E. M. Musiol, T. Härtner, A. Kulik, J. Moldenhauer, J. Piel, W. Wohlleben and T. Weber, *Chem. Biol.*, 2011, **18**, 438-444.
- 48 D. Menche, F. Arikan, O. Perlova, N. Horstmann, W. Ahlbrecht, S. C. Wenzel, R. Jansen, H. Irschik and R. Müller, *J. Am. Chem. Soc.*, 2008, **130**, 14234-14243.
- 49 C. Zhao, J. M. Coughlin, J. Ju, D. Zhu, E. Wendt-Pienkowski, X. Zhou, Z. Wang, B. Shen and Z. Deng, *J. Biol. Chem.*, 2010, **285**, 20097-20108.
- 50 E. J. N. Helfrich and J. Piel, *Nat. Prod. Rep.*, 2016, **33**, 231-316.
- 51 K. Jensen, H. Niederkrüger, K. Zimmermann, A. L. Vagstad, J. Moldenhauer, N. Brendel, S. Frank, P. Pöplau, C. Kohlhaas, C. A. Townsend, M. Oldiges, C. Hertweck and J. Piel, *Chem. Biol.*, 2012, **19**, 329-339.
- 52 P. S. Patel, S. Huang, S. Fisher, D. Pirnik, C. Aklonis, L. Dean, E. Meyers, P. Fernandes and F. Mayerl, *J. Antibiot. (Tokyo)*, 1995, **48**, 997-1003.
- 53 J. Moldenhauer, X. H. Chen, R. Borriss and J. Piel, *Angew. Chemie - Int. Ed.*, 2007, **46**, 8195-8197.
- 54 R. A. Butcher, F. C. Schroeder, M. A. Fischbach, P. D. Straight, R. Kolter, C. T. Walsh and J. Clardy, *Proc. Natl. Acad. Sci.*, 2007, **104**, 1506-1509.
- 55 J. M. Fritzler and G. Zhu, *Int. J. Parasitol.*, 2007, **37**, 307-316.
- 56 G. Huang, L. Zhang and R. G. Birch, *Microbiology*, 2001, **147**, 631-642.
- 57 L. Gu, T. W. Geders, B. Wang, W. H. Gerwick, K. Håkansson, J. L. Smith and D. H. Sherman, *Science*, 2007, **318**, 970-974.
- 58 A. Witkowski, A. K. Joshi, Y. Lindqvist and S. Smith, *Biochemistry*, 1999, **38**, 11643-11650.
- 59 J. Moldenhauer, D. C. G. Götz, C. R. Albert, S. K. Bischof, K. Schneider, R. D. Süssmuth, M. Engeser, H. Gross, G. Bringmann and J. Piel, *Angew. Chemie - Int. Ed.*, 2010, **49**, 1465-1467.

- 60 D. C. Gay, P. J. Spear and A. T. Keatinge-Clay, *ACS Chem. Biol.*, 2014, **9**, 2374-2381.
- 61 R. D. Süßmuth and A. Mainz, *Angew. Chemie - Int. Ed.*, 2017, **56**, 3770-3821.
- 62 M. A. Skiba, A. P. Sikkema, W. D. Fiers, W. H. Gerwick, D. H. Sherman, C. C. Aldrich and J. L. Smith, *ACS Chem. Biol.*, 2016, **11**, 3319-3327.
- 63 X. Xie, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2017, **139**, 6102-6105.
- 64 A. S. Rahman, J. Hothersall, J. Crosby, T. J. Simpson and C. M. Thomas, *J. Biol. Chem.*, 2005, **280**, 6399-6408.
- 65 L. Song, M. Jenner, J. Masschelein, C. Jones, M. J. Bull, S. R. Harris, R. C. Hartkoorn, A. Vocat, I. Romero-Canelon, P. Coupland, G. Webster, M. Dunn, R. Weiser, C. Paisey, S. T. Cole, J. Parkhill, E. Mahenthiralingam and G. L. Challis, *J. Am. Chem. Soc.*, 2017, **139**, 7974-7981.
- 66 H. Irschik, D. Schummer, G. Höfle, H. Reichenbach, H. Steinmetz and R. Jansen, *J. Nat. Prod.*, 2007, **70**, 1060-1063.
- 67 D. Menche, P. Li and H. Irschik, *Bioorganic Med. Chem. Lett.*, 2010, **20**, 939-941.
- 68 M. Altendorfer, A. Raja, F. Sasse, H. Irschik and D. Menche, *Org. Biomol. Chem.*, 2013, **11**, 2116-2139.
- 69 X. Jian, unpublished work.
- 70 H. A. Cooke, E. L. Guenther, Y. Luo, B. Shen and S. D. Bruner, *Biochemistry*, 2009, **48**, 9590-9598.
- 71 S. M. Bernard, D. L. Akey, A. Tripathi, S. R. Park, J. R. Konwerski, Y. Anzai, S. Li, F. Kato, D. H. Sherman and J. L. Smith, *ACS Chem. Biol.*, 2015, **10**, 1340-1351.
- 72 K. Blin, M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano and T. Weber, *Nucleic Acids Res.*, 2013, **41**, W204-212.
- 73 J. A. Bornhorst and J. J. Falke, *Methods Enzymol.*, 2000, **326**, 245-254.
- 74 I. Nazi, K. P. Koteva and G. D. Wright, *Anal. Biochem.*, 2004, **324**, 100-105.
- 75 L. E. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber and C. T. Walsh, *Biochemistry*, 1998, **37**, 1585-1595.
- 76 S. B. Bumpus and N. L. Kelleher, *Curr. Opin. Chem. Biol.*, 2008, **12**, 475-482.
- 77 D. Meluzzi, W. H. Zheng, M. Hensler, V. Nizet and P. C. Dorrestein, *Bioorganic Med. Chem. Lett.*, 2008, **18**, 3107-3111.
- 78 E. L. de los Santos and G. L. Challis, *bioRxiv*, 2017, DOI:

- 10.1101/119214.
- 79 M. H. Medema, R. Kottmann, P. Yilmaz, M. Cummings, J. B. Biggins, K. Blin, I. De Bruijn, Y. H. Chooi, J. Claesen, R. C. Coates, P. Cruz-Morales, S. Duddela, S. Düsterhus, D. J. Edwards, D. P. Fewer, N. Garg, C. Geiger, J. P. Gomez-Escribano, A. Greule, M. Hadjithomas, A. S. Haines, E. J. N. Helfrich, M. L. Hillwig, K. Ishida, A. C. Jones, C. S. Jones, K. Jungmann, C. Kegler, H. U. Kim, P. Kötter, D. Krug, J. Masschelein, A. V. Melnik, S. M. Mantovani, E. A. Monroe, M. Moore, N. Moss, H. W. Nützmann, G. Pan, A. Pati, D. Petras, F. J. Reen, F. Rosconi, Z. Rui, Z. Tian, N. J. Tobias, Y. Tsunematsu, P. Wiemann, E. Wyckoff, X. Yan, G. Yim, F. Yu, Y. Xie, B. Aigle, A. K. Apel, C. J. Balibar, E. P. Balskus, F. Barona-Gómez, A. Bechthold, H. B. Bode, R. Borriss, S. F. Brady, A. A. Brakhage, P. Caffrey, Y. Q. Cheng, J. Clardy, R. J. Cox, R. De Mot, S. Donadio, M. S. Donia, W. A. Van Der Donk, P. C. Dorrestein, S. Doyle, A. J. M. Driessen, M. Ehling-Schulz, K. D. Entian, M. A. Fischbach, L. Gerwick, W. H. Gerwick, H. Gross, B. Gust, C. Hertweck, M. Höfte, S. E. Jensen, J. Ju, L. Katz, L. Kaysser, J. L. Klassen, N. P. Keller, J. Kormanec, O. P. Kuipers, T. Kuzuyama, N. C. Kyrpides, H. J. Kwon, S. Lautru, R. Lavigne, C. Y. Lee, B. Linqun, X. Liu, W. Liu, A. Luzhetskyy, T. Mahmud, Y. Mast, C. Méndez, M. Metsä-Ketelä, J. Micklefield, D. A. Mitchell, B. S. Moore, L. M. Moreira, R. Müller, B. A. Neilan, M. Nett, J. Nielsen, F. O’Gara, H. Oikawa, A. Osbourn, M. S. Osburne, B. Ostash, S. M. Payne, J. L. Pernodet, M. Petricek, J. Piel, O. Ploux, J. M. Raaijmakers, J. A. Salas, E. K. Schmitt, B. Scott, R. F. Seipke, B. Shen, D. H. Sherman, K. Sivonen, M. J. Smanski, M. Sosio, E. Stegmann, R. D. Süssmuth, K. Tahlan, C. M. Thomas, Y. Tang, A. W. Truman, M. Viaud, J. D. Walton, C. T. Walsh, T. Weber, G. P. Van Wezel, B. Wilkinson, J. M. Willey, W. Wohlleben, G. D. Wright, N. Ziemert, C. Zhang, S. B. Zotchev, R. Breitling, E. Takano and F. O. Glöckner, *Nat. Chem. Biol.*, 2015, **11**, 625-631.
- 80 R. Ueoka, A. R. Uria, S. Reiter, T. Mori, P. Karbaum, E. E. Peters, E. J. N. Helfrich, B. I. Morinaka, M. Gugger, H. Takeyama, S. Matsunaga and J. Piel, *Nat. Chem. Biol.*, 2015, **11**, 705–712.
- 81 J. C. Kwan, M. S. Donia, A. W. Han, E. Hirose, M. G. Haygood and E. W. Schmidt, *Proc. Natl. Acad. Sci.*, 2012, **109**, 20655–20660.
- 82 T. Wakimoto, Y. Egami, Y. Nakashima, Y. Wakimoto, T. Mori, T. Awakawa, T. Ito, H. Kenmoku, Y. Asakawa, J. Piel and I. Abe, *Nat. Chem. Biol.*, 2014, **10**, 648–655.
- 83 K. M. Fisch, C. Gurgui, N. Heycke, S. A. Van Der Sar, S. A. Anderson,

- V. L. Webb, S. Taudien, M. Platzer, B. K. Rubio, S. J. Robinson, P. Crews and J. Piel, *Nat. Chem. Biol.*, 2009, **5**, 494–501.
- 84 K. Ishida, T. Lincke, S. Behnken and C. Hertweck, *J. Am. Chem. Soc.*, 2010, **132**, 13966–13968.
- 85 D. Pistorius and R. Müller, *ChemBioChem*, 2012, **13**, 416–426.
- 86 P. Z. Kozbial and A. R. Mushegian, *BMC Struct. Biol.*, 2005, DOI: 10.1186/1472-6807-5-19.
- 87 D. K. Liscombe, G. V. Louie and J. P. Noel, *Nat. Prod. Rep.*, 2012, **29**, 1238–1250.
- 88 M. Klaus and M. Grininger, *Nat. Prod. Rep.*, 2018, DOI: 10.1039/c8NP0030a.
- 89 L. A. Kelly, S. Mezulis, C. Yates, M. Wass and M. Sternberg, *Nat. Protoc.*, 2015, **10**, 845–858.
- 90 K. F. Geoghegan, H. B. F. Dixon, P. J. Rosner, L. R. Hoth, A. J. Lanzetti, K. A. Borzilleri, E. S. Marr, L. H. Pezzullo, L. B. Martin, P. K. Lemotte, A. S. McColl, A. V. Kamath and J. G. Stroh, *Anal. Biochem.*, 1999, **267**, 169–184.
- 91 S. Bonissone, N. Gupta, M. Romine, R. A. Bradshaw and P. A. Pevzner, *Mol. Cell. Proteomics*, 2013, **12**, 14–28.
- 92 M. C. Deller, L. Kong, B. Rupp, *Acta Crystallogr. Sect. F Struct. Biol. Commun.*, 2016, **72**, 72–95.
- 93 Y. Tang, C.-Y. Kim, I. I. Mathews, D. E. Cane and C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **72**, 72–95.
- 94 D. C. Gay, G. Gay, A. J. Axelrod, M. Jenner, C. Kohlhaas, A. Kampa, N. J. Oldham, J. Piel and A. T. Keatinge-Clay, *Structure*, 2014, **22**, 444–451.
- 95 J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W. H. Gerwick, D. H. Sherman and J. L. Smith, *Chem. Biol.*, 2013, **20**, 1340–1351.
- 96 T. Bretschneider, J. B. Heim, D. Heine, R. Winkler, B. Busch, B. Kusebauch, T. Stehle, G. Zocher and C. Hertweck, *Nature*, 2013, **502**, 124–128.
- 97 J. R. Lohman, M. Ma, J. Osipiuk, B. Nocek, Y. Kim, C. Chang, M. Cuff, J. Mack, L. Bigelow, H. Li, M. Endres, G. Babnigg, A. Joachimiak, G. N. Phillips and B. Shen, *Proc. Natl. Acad. Sci.*, 2015, **112**, 12693–12698.
- 98 E. Krieger, S. B. Nabuurs and G. Vriend, *Struct. Bioinforma.*, 2003, **44**, 509–523.
- 99 J. L. Martin and F. M. McMillan, *Curr. Opin. Struct. Biol.*, 2002, **12**, 783–793.
- 100 W. DeLano, *CCP4 Newsl. Protein Crystallogr.*, 2002.

- 101 M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. Mcgettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins, *Bioinformatics*, 2007, **23**, 2947-2948.
- 102 S. Rachid, M. Scharfe, H. Blöcker, K. J. Weissman and R. Müller, *Chem. Biol.*, 2009, **16**, 70–81.
- 103 J. Ligon, S. Hill, J. Beck, R. Zirkle, I. Molnár, J. Zawodny, S. Money and T. Schupp, *Gene*, 2002, **285**, 257-267.
- 104 Z. Feng, J. Qi, T. Tsuge, Y. Oba, T. Kobayashi, Y. Suzuki, Y. Sakagami and M. Ojika, *Biosci. Biotechnol. Biochem.*, 2005, **69**, 1372–80.
- 105 O. Perlova, J. Fu, S. Kuhlmann, D. Krug, A. F. Stewart, Y. Zhang and R. Müller, *Appl. Environ. Microbiol.*, 2006, **72**, 7485–7494.
- 106 N. Gaitatzis, B. Silakowski, B. Kunze, G. Nordsiek, H. Blöcker, G. Höfle and R. Müller, *J. Biol. Chem.*, 2002, **277**, 13082–13090.
- 107 S. Müller, S. Rachid, T. Hoffmann, F. Surup, C. Volz, N. Zaburannyi and R. Müller, *Chem. Biol.*, 2014, **21**, 855-865.
- 108 Z. Chang, N. Sitachitta, J. V. Rossi, M. A. Roberts, P. M. Flatt, J. Jia, D. H. Sherman and W. H. Gerwick, *J. Nat. Prod.*, 2004, **67**, 1356–1367.
- 109 P. Krastel, S. Roggo, M. Schirle, N. T. Ross, F. Perruccio, P. Aspesi, T. Aust, K. Buntin, D. Estoppey, B. Liechty, F. Mapa, K. Memmert, H. Miller, X. Pan, R. Riedl, C. Thibaut, J. Thomas, T. Wagner, E. Weber, X. Xie, E. K. Schmitt and D. Hoepfner, *Angew. Chemie - Int. Ed.*, 2015, **54**, 10149–10154.
- 110 K. Buntin, K. J. Weissman and R. Müller, *ChemBioChem*, 2010, **11**, 1137–1146.
- 111 Z. Chang, P. Flatt, W. H. Gerwick, V. A. Nguyen, C. L. Willis and D. H. Sherman, *Gene*, 2002, **296**, 235–247.
- 112 D. J. Edwards, B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Goeger, M. A. Roberts and W. H. Gerwick, *Chem. Biol.*, 2004, **11**, 817-833.
- 113 S. Weinig, H. J. Hecht, T. Mahmud and R. Müller, *Chem. Biol.*, 2003, **10**, 939-952.
- 114 J. Masschelein, M. Jenner and G. L. Challis, *Nat. Prod. Rep.*, 2017, **34**, 712–783.
- 115 C. L. Hendricks, J. R. Ross, E. Pichersky, J. P. Noel and Z. S. Zhou, *Anal. Biochem.*, 2004, **326**, 100-105.
- 116 T. M. Larsen, S. K. Boehlein, S. M. Schuster, N. G. J. Richards, J. B. Thoden, H. M. Holden and I. Rayment, *Biochemistry*, 1999, **38**, 16146-16157.
- 117 M. Sharma, N. N. Sharma and T. C. Bhalla, *Rev. Environ. Sci.*

- Biotechnol.*, 2009, **8**, 343.
- 118 E. Cundliffe and A. L. Demain, *J. Ind. Microbiol. Biotechnol.*, 2010, **37**, 643-672.
 - 119 L. M. Quirós, I. Aguirrezabalaga, C. Olano, C. Méndez and J. A. Salas, *Mol. Microbiol.*, 1998, **28**, 1177-1188.
 - 120 H. Hong, M. Samborsky, F. Lindner and P. F. Leadlay, *Angew. Chemie - Int. Ed.*, 2016, **55**, 1118-1123.
 - 121 Y. Shinohara, F. Kudo and T. Eguchi, *J. Am. Chem. Soc.*, 2011, **133**, 18134-18137.
 - 122 H. Irschik, M. Kopp, K. J. Weissman, K. Buntin, J. Piel and R. Müller, *ChemBioChem*, 2010, **11**, 1840-1849.
 - 123 Z. Huang, K. K. A. Wang and W. A. Van Der Donk, *Chem. Sci.*, 2016, **7**, 5219-5223.
 - 124 M. N. Thaker, W. Wang, P. Spanogiannopoulos, N. Waglechner, A. M. King, R. Medina and G. D. Wright, *Nat. Biotechnol.*, 2013, **31**, 922-927.
 - 125 R. S. Flannagan, T. Linn and M. A. Valvano, *Environ. Microbiol.*, 2008, **10**, 1652-1660.
 - 126 M. D. Lefebvre and M. A. Valvano, *Appl. Environ. Microbiol.*, 2002, **68**, 5956-5964.
 - 127 M. Orwick-Rydmark, T. Arnold and D. Linke, *Curr. Protoc. Protein Sci.*, 2016, DOI: 10.1002/0471140864.ps0408s84.
 - 128 S. Santajit and N. Indrawattana, *Biomed Res. Int.*, 2016, DOI: 10.1155/2016/2475067.
 - 129 M. Altaf, C. H. Miller, D. S. Bellows and R. O'Toole, *Tuberculosis*, 2010, **90**, 333-337.
 - 130 J.-C. Palomino, A. Martin, M. Camacho, H. Guerra, J. Swings and F. Portaels, *Antimicrob. Agents Chemother.*, 2002, **46**, 2720-2722.
 - 131 R. Jansen, D. Schummer, H. Irschik and G. Höfle, *Liebigs Ann. der Chemie*, 1990, **10**, 975-988.
 - 132 D. Esposito and D. K. Chatterjee, *Curr. Opin. Biotechnol.*, 2006, **17**, 353-358.
 - 133 B. Kusebauch, B. Busch, K. Scherlach, M. Roth and C. Hertweck, *Angew. Chemie - Int. Ed.*, 2009, **48**, 5001-5004.
 - 134 J. Ju, S. R. Rajski, S. K. Lim, J. W. Seo, N. R. Peters, F. M. Hoffmann and B. Shen, *J. Am. Chem. Soc.*, 2009, **131**, 1370-1371.
 - 135 J. W. Seo, M. Ma, T. Kwong, J. Ju, S. K. Lim, H. Jiang, J. R. Lohman, C. Yang, J. Cleveland, E. Zazopoulos, C. M. Farnet and B. Shen, *Biochemistry*, 2014, **53**, 7854-7865.
 - 136 M. Yin, Y. Yan, J. R. Lohman, S. X. Huang, M. Ma, G. R. Zhao, L. H. Xu,

- W. Xiang and B. Shen, *Org. Lett.*, 2014, **16**, 3072-3075.
- 137 B. Wang, Y. Song, M. Luo, Q. Chen, J. Ma, H. Huang and J. Ju, *Org. Lett.*, 2013, **15**, 1278-1281.
- 138 A. Pereda, R. G. Summers, D. L. Stassi, X. Ruan and L. Katz, *Microbiology*, 1998, **144**, 543-553.
- 139 C. Hobson, unpublished work.
- 140 B. Douzi, in *Bacterial Protein Secretion Systems: Methods and Protocols*, eds. L. Journet and E. Cascales, Springer New York, New York, NY, 2017, pp. 257-275.
- 141 L. Manzi, A. S. Barrow, D. Scott, R. Layfield, T. G. Wright, J. E. Moses and N. J. Oldham, *Nat. Commun.*, 2016, DOI: 10.1038/ncomms13288.
- 142 A. McPherson and J. A. Gavira, *Acta Crystallogr. Sect. FStructural Biol. Commun.*, 2014, **70**, 2-20.
- 143 E. V. Orlova and H. R. Saibil, *Chem. Rev.*, 2011, **111**, 7710-7748.
- 144 S. Dutta, J. R. Whicher, D. A. Hansen, W. A. Hale, J. A. Chemler, G. R. Congdon, A. R. H. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, 2014, **510**, 512-517.
- 145 S. J. Ludtke, in *Methods in Enzymology*, 2016, **579**, 159-189.
- 146 M. Jenner, S. Frank, A. Kampa, C. Kohlhaas, P. Pöplau, G. S. Briggs, J. Piel and N. J. Oldham, *Angew. Chemie - Int. Ed.*, 2013, **52**, 1143-1147.
- 147 H. Irschik, M. Kopp, K. J. Weissman, K. Buntin, J. Piel and R. Müller, *ChemBioChem*, 2010, **11**, 1840-1849.
- 148 A. Alexandrov, K. Dutta and S. M. Pascal, *Biotechniques*, 2001, **30**, 1194-1198.
- 149 D. H. Figurski and D. R. Helinski, *Proc. Natl. Acad. Sci.*, 1979, **76**, 1648-1652.
- 150 K. Agnoli, S. Schwager, S. Uehlinger, A. Vergunst, D. F. Viteri, D. T. Nguyen, P. A. Sokol, A. Carlier and L. Eberl, *Mol. Microbiol.*, 2012, **83**, 362-378.
- 151 L. A. O'Sullivan, A. J. Weightman, T. H. Jones, A. M. Marchbank, J. M. Tiedje and E. Mahenthiralingam, *Environ. Microbiol.*, 2007, **9**, 1017-1034.
- 152 M. Tosin, D. Spiteller and J. B. Spencer, *ChemBioChem*, 2009, **10**, 1714-1723.
- 153 Clinical and Laboratory Standards Institute, *Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically; Approved Standard*, Clinical and Laboratory Standards Institute, Pennsylvania, 7th edn., 2006.

- 154 S. B. Hari, R. A. Grant and R. T. Sauer, *Structure*, 2018, **26**, 1251–1258.
- 155 M. A. Torres, E. Hoffarth, L. Eugenio, J. Savtchouk, X. Chen, J. S. Morris, P. J. Facchini and K. K. S. Ng, *J. Biol. Chem.*, 2016, **291**, 23403-23415.
- 156 M. R. Bennett, M. L. Thompson, S. A. Shepherd, M. S. Dunstan, A. J. Herbert, D. R. M. Smith, V. A. Cronin, B. R. K. Menon, C. Levy and J. Micklefield, *Angew. Chemie - Int. Ed.*, 2018, **57**, 10600-10604.
- 157 C. C. Huang, C. V. Smith, M. S. Glickman, W. R. Jacobs and J. C. Sacchettini, *J. Biol. Chem.*, 2002, **277**, 11559-11569.

Appendices

Appendix 1: Protein sequences and characterization by MS

For proteins overproduced using the pET28a vector post-translational cleavage of the *N*-terminal methionine, by methionine aminopeptidases, occurs. The His-tag also becomes gluconoylated; the peaks in the MS corresponding to this are marked with a *. Mutated residues are shown in red and the cleaved His-tag in teal.

GldM11 OMT-ACP

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MKHHHHHHHH	GGLVPRGSHG	SEASPAWTGG	TGGTGGPGDK	AAEYYTFDAQ	RRSDTYHEEY
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
LTFCPFEREQ	PGFSMTRVLS	APEAYPDEYR	QVLDKQKEMR	AVTFRHERFE	RVGRVLDIGC
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
GCGSDLIELA	LRHPALHADG	FTITPAQAEI	GNRRIANLGI	AERVRI RHAD	SSRDAFP HRY
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
DLVIGFEVSC	HIADKTALFS	NIRSLAPTG	CILLMDFIAN	LRGSIVDDKV	DIAISPQAEW
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>
ANVLAEQGLV	IDDLVDLSEG	IANYVFDPDV	ERNIAHFPEA	ARRSSRN FAN	MAVSLRQ NWI
<u>310</u>	<u>320</u>	<u>330</u>	<u>340</u>	<u>350</u>	<u>360</u>
SYCLFRIVVD	RHGRSTEALR	AANA AVIARR	TPYREALADL	FARRDGSGDI	GLTAAVMAAA
<u>370</u>	<u>380</u>	<u>390</u>	<u>400</u>	<u>410</u>	<u>420</u>
SRQAPAREAA	KEAVQQAAPA	PQPAAAPSAG	ARAATVREAL	ERVMIDTLGV	SPAELRAARS
<u>430</u>	<u>440</u>	<u>450</u>	<u>460</u>	<u>470</u>	<u>480</u>
FAELGVDSL L	AVRLLEAINV	AFDLHEPTSV	MFEFHDTQAL	AAHLETRGAR	VAALPAEAAV
<u>490</u>					
ALAGVPAARA	DSR				

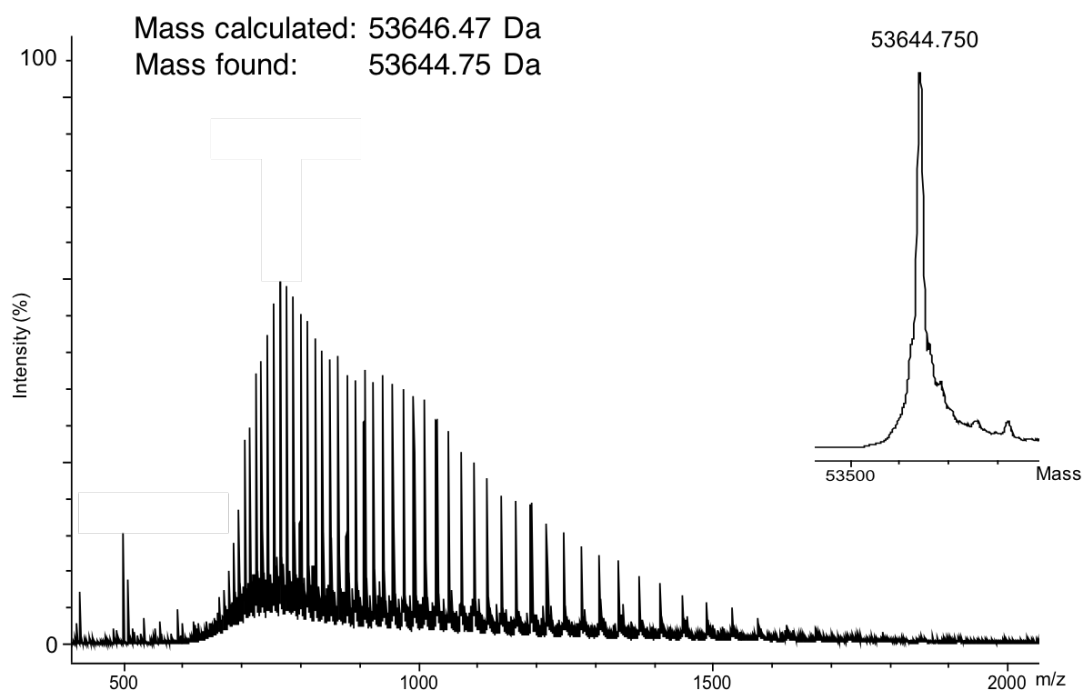


Figure S1: Raw and deconvoluted ESI-MS spectra for GldM11 OMT-ACP.

GldM11 OMT-ACP(H153A)

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MKHHHHHHHH	GGLVPRGSHG	SEASPAWTGG	TGGTGGPGDK	AAEYYTFDAQ	RRSDTYHEEY
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
LTFCPFEREQ	PGFSMTRVLS	APEAYPDEYR	QVLDKQKEMR	AVTFRHERFE	RVGRVLDIGC
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
GCGSDLIELA	LRHPALHADG	FTITPAQAEL	GNRRIANLGI	AERVRIRHAD	SSRDAFPHRY
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
DLVIGFEVSC	AIADKTALFS	NIRRLAPTG	CILLMDFIAN	LRGSIVDDKV	DIAISPQAEW
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>
ANVLAEQGLV	IDDLVDLSEG	IANYVFDPDV	ERNIAHFPEA	ARRSSRNFAN	MAVSLRQNW
<u>310</u>	<u>320</u>	<u>330</u>	<u>340</u>	<u>350</u>	<u>360</u>
SYCLFRIVVD	RHGRSTEALR	AANA AVIARR	TPYREALADL	FARRDGS GDI	GLTAAVMAAA
<u>370</u>	<u>380</u>	<u>390</u>	<u>400</u>	<u>410</u>	<u>420</u>
SRQAPAREAA	KEAVQQAAPA	PQPAAAPSAG	ARAATVREAL	ERVMIDTLGV	SPAELRAARS
<u>430</u>	<u>440</u>	<u>450</u>	<u>460</u>	<u>470</u>	<u>480</u>
FAELGVDSLL	AVRLLEAINV	AFDLHEPTSV	MFEFHDTQAL	AAHLETRGAR	VAALPAEAAV
<u>490</u>					
ALAGVPAARA	DSR				

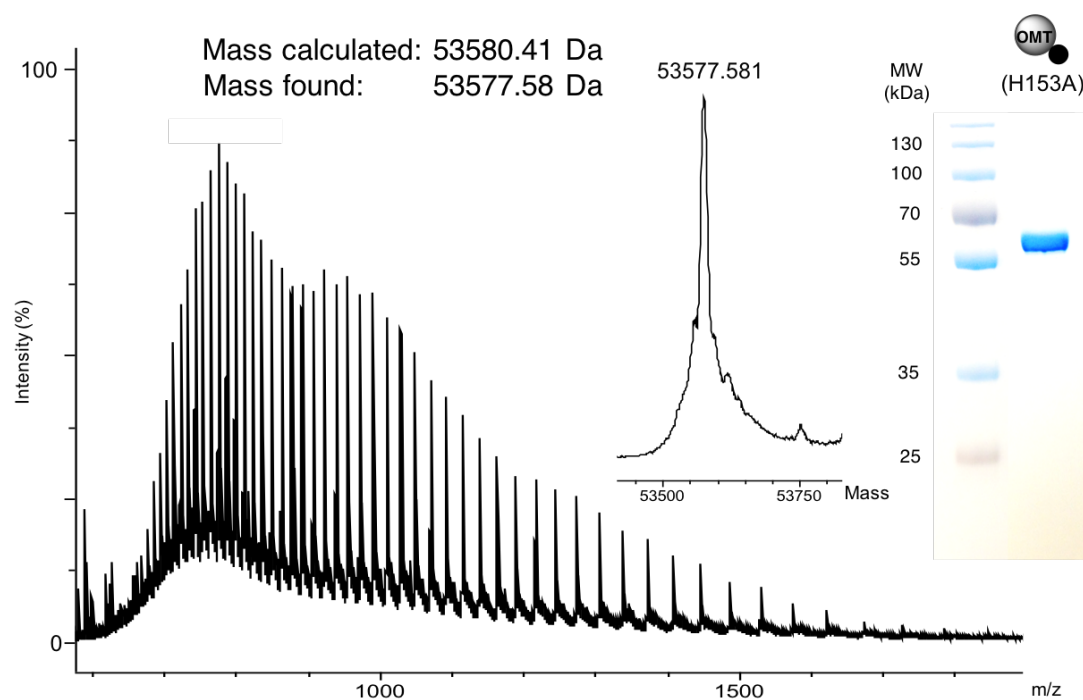


Figure S2: Raw and deconvoluted ESI-MS spectra for GldM11OMT-ACP(H153A). 10 % SDS-PAGE analysis of the protein is also shown.

GldM11OMT

10	20	30	40	50	60
MGSSHHHHHH	SSGLVPRGSH	MGGPGDKAAE	YYTFDAQRRS	DTYHEEYLTf	CPFEREQPGF
70	80	90	100	110	120
SMTRVLSAPE	AYPDEYRQVL	DKQKEMRAVT	FRHERFERVG	RVLDIGCGCG	SDLIELALRH
130	140	150	160	170	180
PALHADGFTI	TPAQAELGNR	RIANLGIAER	VRIRHADSSR	DAFPHRYDLV	IGFEVSCHIA
190	200	210	220	230	240
DKTALFSNIR	RSLAPTGCIL	LMDFIANLRG	SIVDDKVDIA	ISPQAEWANV	LAEQGLVIDD
250	260	270	280	290	300
LVDLSEGAN	YVFDPDVERN	IAHFPEAARR	SSRNFAANMAV	SLRQNWISYC	LFRIVVDRHG
310	320	330	340	350	
RSTEALRAAN	AAVIARRTPY	REALADLFAR	RDGSGDIGLT	AAVMAAASRQ	AP

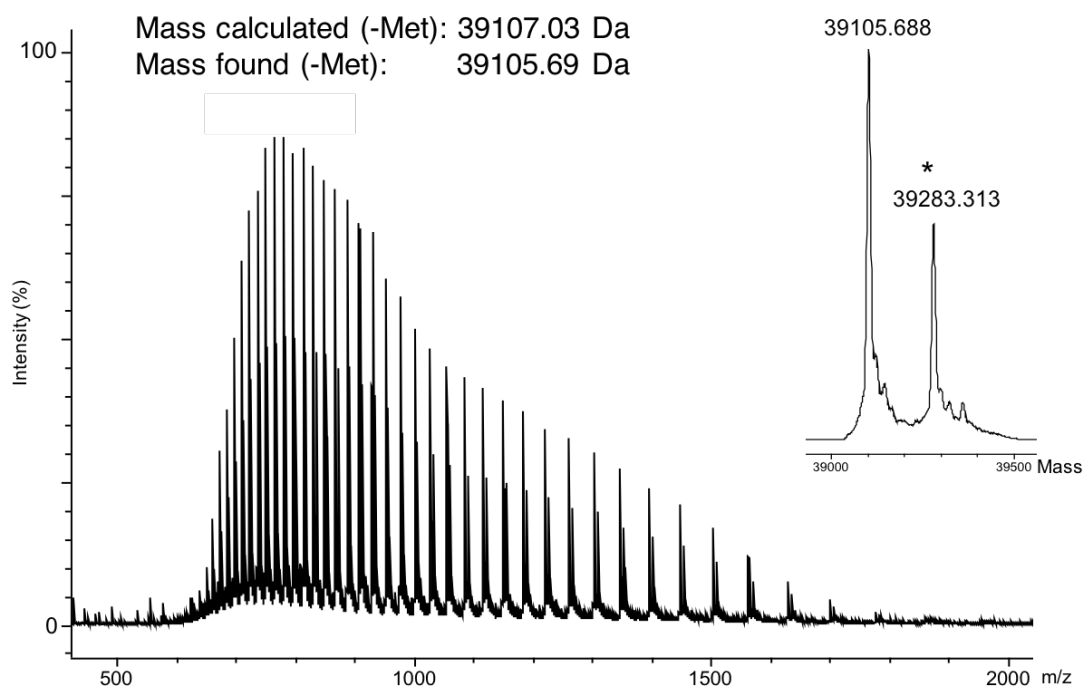


Figure S3: Raw and deconvoluted ESI-MS spectra for GldM11 OMT.

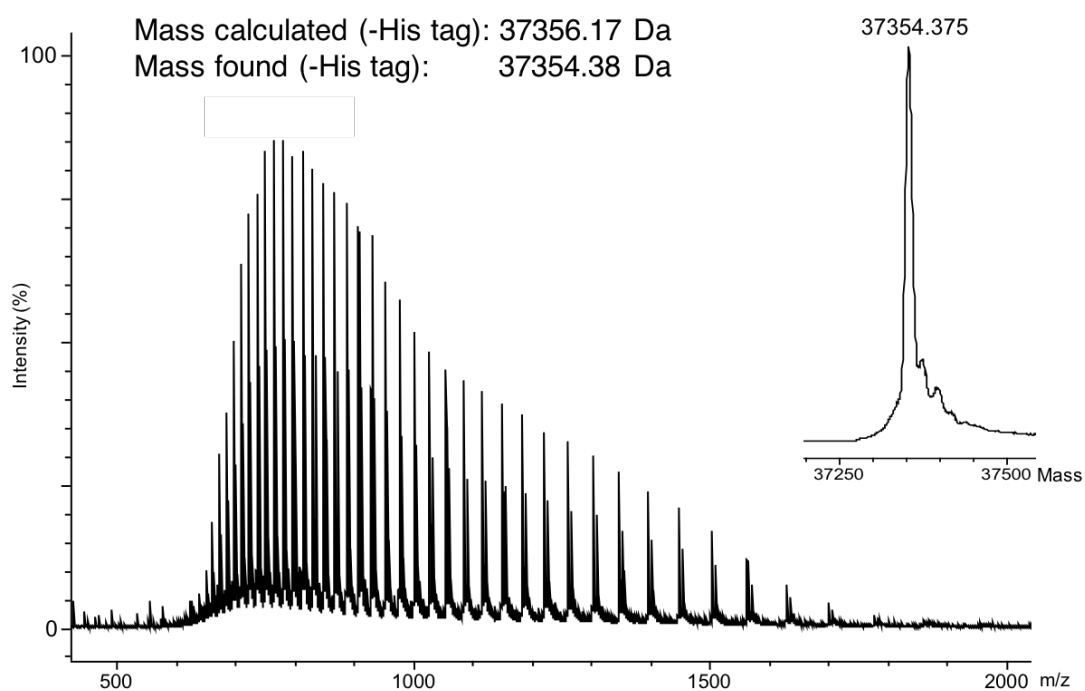


Figure S4: Raw and deconvoluted ESI-MS spectra for GldM11 OMT after cleavage of the His-tag.

GldM11KS-OMT

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>	
MGSSHHHHHH	SSGLVPRGSH	MDQGGVAII	G	VSVRTAGAND	AGELWELLRS	GRRRAIGEVP
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>	
SRWDWRPYFS	GPGEASNRIA	THRGAFIEGL	DGFDPLFFEI	SPREAQWMDP	RQRLILEEAW	
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>	
RAFEDAGYAG	ERLRGSRGCV	FVGVEEGVPG	EAADGLATSH	HNGILAARIS	YVLDLKGPNL	
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>	
AINTACSSGL	VAVHTACQSV	QRGECELALA	GGVNVLSNPL	TYVALTQGGM	LSSSGECHAF	
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>	
DARADGMVPG	EAVAAVVLKD	LARAQADGDP	IHGVIASGV	NYDGRNTGIT	APSARSQAL	
<u>310</u>	<u>320</u>	<u>330</u>	<u>340</u>	<u>350</u>	<u>360</u>	
ISEVLERGGI	EAERIEAVLA	HSVGSPLGDP	IEARALCEAL	GEGLGEGLINE	GLNEGTQTRV	
<u>370</u>	<u>380</u>	<u>390</u>	<u>400</u>	<u>410</u>	<u>420</u>	
LGSVKPQIGH	TFAASGVVNV	IAMCASLRHG	LRLGIANHEV	ANPDLRIGDG	ALSLGAGAQP	
<u>430</u>	<u>440</u>	<u>450</u>	<u>460</u>	<u>470</u>	<u>480</u>	
WPKRAGVARC	GLVSATGMSG	TNACVVIEEA	PEETSGAVGK	ETRREWLVLK	SARESGALRE	
<u>490</u>	<u>500</u>	<u>510</u>	<u>520</u>	<u>530</u>	<u>540</u>	
SARRLRAWLG	GEAGAEVELD	ALSLTLDVGR	EAMRYRLAMV	VGAGGSEEGR	SGLMAALDAY	
<u>550</u>	<u>560</u>	<u>570</u>	<u>580</u>	<u>590</u>	<u>600</u>	
LEGAEGAALA	GKGVYAGEVA	DEGMEEAGSA	ALSKASALSG	QARDWCEGLR	TDDAVLHRGG	
<u>610</u>	<u>620</u>	<u>630</u>	<u>640</u>	<u>650</u>	<u>660</u>	
RPRRLAGLPT	YPFARRPLGQ	PPAAEASPAW	TGGTGGTGGP	GDKAAEYYTF	DAQRRSDTYH	
<u>670</u>	<u>680</u>	<u>690</u>	<u>700</u>	<u>710</u>	<u>720</u>	
EEYLTFPCFE	REQPGFSMTR	VLSAPEAYPD	EYRQVLDKQK	EMRAVTFRHE	RFERVGRVLD	
<u>730</u>	<u>740</u>	<u>750</u>	<u>760</u>	<u>770</u>	<u>780</u>	
IGCGCGSDLI	ELALRHPALH	ADGFTITPAQ	AELGNRRIAN	LGIAERVIR	HADSSRDAFP	
<u>790</u>	<u>800</u>	<u>810</u>	<u>820</u>	<u>830</u>	<u>840</u>	
HRYDLVIGFE	VSCHIADKTA	LFSNIRRLA	PTGCILLMDF	IANLRGSIVD	DKVDIAISPQ	
<u>850</u>	<u>860</u>	<u>870</u>	<u>880</u>	<u>890</u>	<u>900</u>	
AEWANVLAEQ	GLVIDDLVDL	SEGIANYVFD	PDVERNIAHF	PEAARRSSRN	FANMAVSLRQ	
<u>910</u>	<u>920</u>	<u>930</u>	<u>940</u>	<u>950</u>	<u>960</u>	
NWISYCLFRI	VVDRHGRSTE	ALRAANAABI	ARRTPYREAL	ADLFARRDGS	GDIGLTAAM	

AAASRQAP

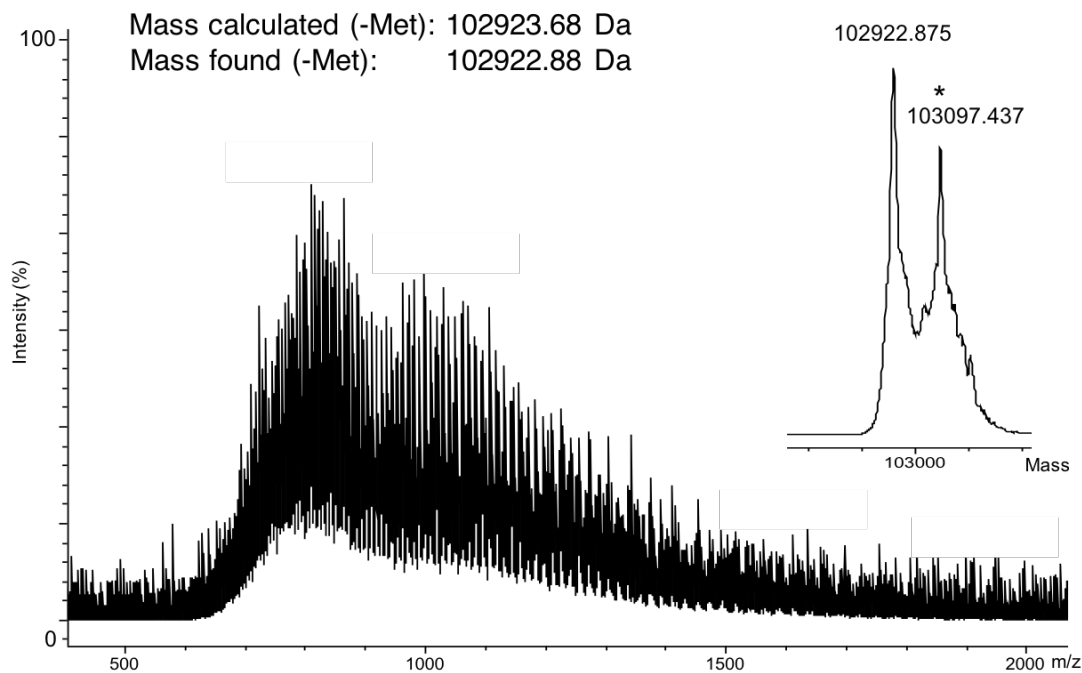


Figure S5: Raw and deconvoluted ESI-MS spectra for GldM11KS-OMT.

GldM11KS-OMT-ACP

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MGSSHHHHHS	SGLVPRGSHM	NRSESTRPDQ	GGVAIIGVSV	RTAGANDAGE	LWELLRSGRR
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
AIGEVPASRW	DWRPYFSGPG	EASNRIATHR	GAFIEGLDGF	DPLFFEISPR	EAQWMDPRQR
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
LILEEAWRAF	EDAGYAGERL	RGSRCGVFVG	VEEGVPGEAA	DGLATSHHNG	ILAARISYVL
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
DLKGPNLAIN	TACSSGLVAV	HTACQSVQRG	ECELALAGGV	NVLNSPLTYV	ALTQGGMLSS
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>
SGECHAFDAR	ADGMVPGEAV	AAVVLKDLAR	AQADGDPIHG	VIRASGVNYD	GRTNGITAPS
<u>310</u>	<u>320</u>	<u>330</u>	<u>340</u>	<u>350</u>	<u>360</u>
ARSQRALISE	VLERGGIEAE	RIEAVLAHSV	GSPLGDPIEA	RALCEALGEG	LGEGLNEGLN
<u>370</u>	<u>380</u>	<u>390</u>	<u>400</u>	<u>410</u>	<u>420</u>
EGTQTRVLGS	VKPQIGHTFA	ASGVVNVIAM	CASLRHGLRL	GIANHEVANP	DLRIGDGALS
<u>430</u>	<u>440</u>	<u>450</u>	<u>460</u>	<u>470</u>	<u>480</u>
LGAGAQPWPK	RAGVARCGLV	SATGMSGTNA	CVVIEEAPPEE	TSGAVGKETR	REWLVKLSAR
<u>490</u>	<u>500</u>	<u>510</u>	<u>520</u>	<u>530</u>	<u>540</u>
ESGALRESAR	RLRAWLGGEA	GAEVELDALIS	LTLDVGREAM	RYRLAMVVGA	GGSEEGRSGL

550	560	570	580	590	600
MAALDAYLEG	AEGAALAGKG	VYAGEVADEG	MEEAGSAALS	KASALSGQAR	DWCEGLRTDD
610	620	630	640	650	660
AVLHRGGRPR	RLAGLPTYPF	ARRPLGQPPA	AEASPAWTGG	TGGTGGPGDK	AAEYYTFDAQ
670	680	690	700	710	720
RRSDTYHEEY	LTFCPFEREQ	PGFSMTRVLS	APEAYPDEYR	QVLDKQKEMR	AVTFRHERFE
730	740	750	760	770	780
RVGRVLDIGC	GCGSDLIELA	LRHPALHADG	FTITPAQAEL	GNRRIANLGI	AERVRIRHAD
790	800	810	820	830	840
SSRDAFPHRY	DLVIGFEVSC	HIADKTALFS	NIRRLAPTGT	CILLMDFIAN	LRGSIVDDKV
850	860	870	880	890	900
DIAISPQAEW	ANVLAEQGLV	IDDLVDLSEG	IANYVFDPDV	ERNIAHFPEA	ARRSSRNFAH
910	920	930	940	950	960
MAVSLRQNWV	SYCLFRIVVD	RHGRSTEALR	AANA AVIARR	TPYREALADL	FARRDGSGLI
970	980	990	1000	1010	1020
GLTAAVMAAA	SRQAPAREAA	KEAVQQAAPA	PQPAAAPSAG	ARAATVREAL	ERVMIDTLGV
1030	1040	1050	1060	1070	
SPAELRAARS	FAELGVDSL	AVRLLEAINV	AFDLHEPTSV	MFEFHDTQAL	AAHLETRG

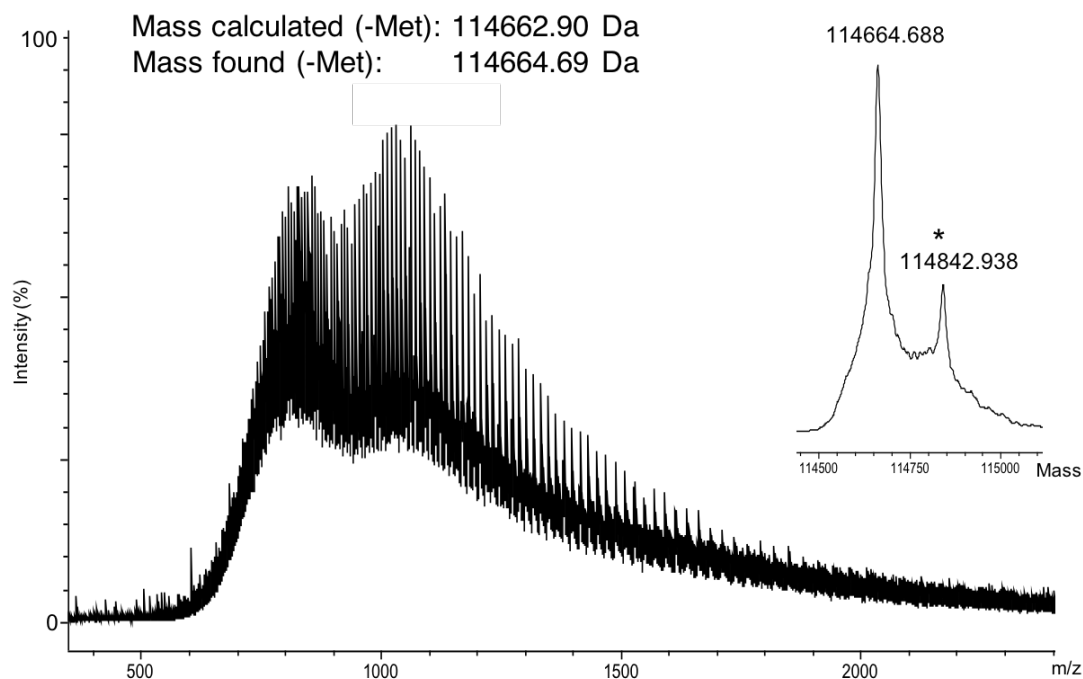


Figure S6: Raw and deconvoluted ESI-MS spectra for GldM11KS-OMT-ACP.

MisM9OMT-ACP

10	20	30	40	50	60
MGSSHHHHHH	SSGLVPRGSH	MASGEEKSDV	EPRIDHTLDE	NKAVEFYTRM	TQDCDEAFQE
70	80	90	100	110	120
EYLTFGAFPR	RIPGFSMTRV	AMNAERYPEE	LEMIKESQRE	LRQVLFCRED	FDRVHTFLDI
130	140	150	160	170	180
GCGHGTVDVIQ	VAALYPHIQT	HGFTITESQA	ILGNRRIEHM	NLGSRRARIYH	KDSSKEPPFD
190	200	210	220	230	240
NCDMILGIEV	TFHIRDKDAL	LQKMADALSD	DGKILLMDYI	ANLRGSIVDP	NVEISIPTQS
250	260	270	280	290	300
QWIDLLSQHD	LIIDEIIDVS	TEVSNYLYDP	EHAHTKHLP	KVSQDSYRNY	ANQAIALEKG
310	320	330	340	350	360
WITYCLLKLQ	RDTQRSEQER	REYNAWKIAN	QTPYAQALED	MLRVGRMPYP	LSGGLKMAVQ
370	380	390	400	410	420
PMDGQPPSTL	DESVPDLGQI	QTALAGIFVA	ILGFETRELD	FIDQLPLTEL	GISSVNAVEL
430	440	450	460	470	
LEAINTFDL	HLPTSLVFEC	HRLEDLAGHI	QVELGRKPIA	PAMSNASVAV	QAEAESR

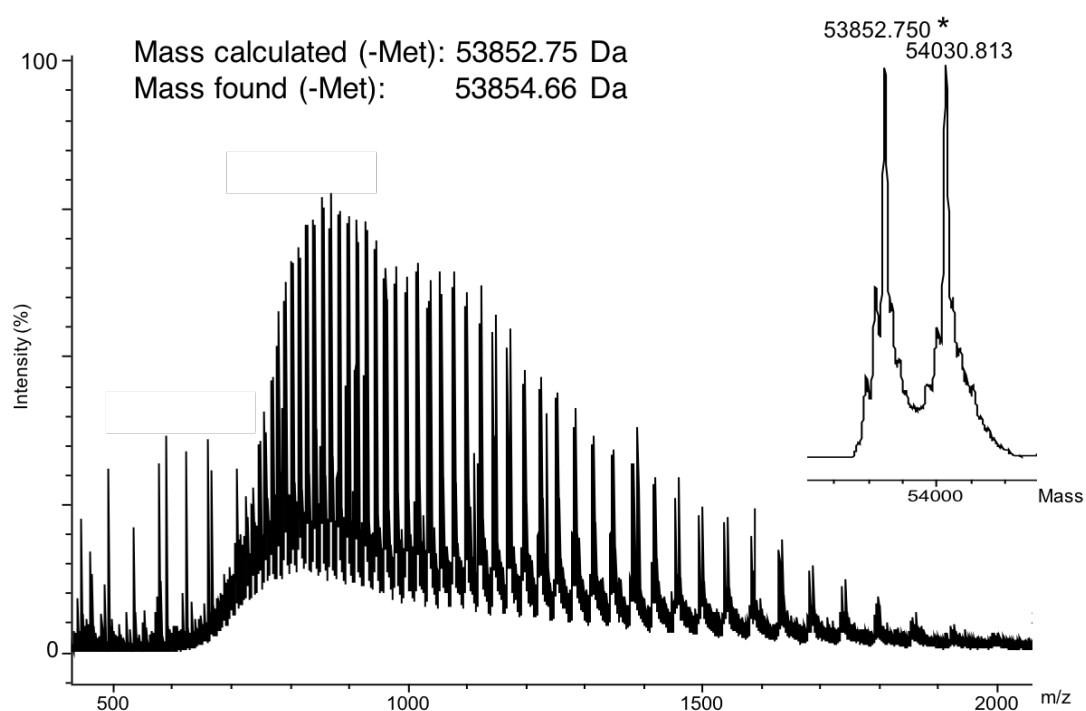


Figure S7: Raw and deconvoluted ESI-MS spectra for MisM9OMT-ACP.

MisM1 OMT-ACP

10 20 30 40 50 60
 MKHHHHHHHH GGLVPRGSHG SQSMPDLISQ KTHETRVNMS DIGQVVADYY DQVTESLKQK
 70 80 90 100 110 120
 IGSQEVYLLF APFQERVEGF SWLKMFFEPE KHREHFELTL QKQHELKSML YRLVEFANVH
 130 140 150 160 170 180
 RIFDIGCGYA TDLIQLAKRH PQVRGWGFSI SLNQVEFGQQ RIQDEGLGER VQLSCNDSTK
 190 200 210 220 230 240
 DPLPGLFDLI IGFEVVVHIA DKQGVFGNMA RHLSNNGTIV LADCVANTVT SINTSYIGQY
 250 260 270 280 290 300
 TSTAEYSDI LAQNGLAIVD CVDVGPEISN FLYDPNYLEN AAYLSSIYPE MVNIEKEHQG
 310 320 330 340 350 360
 WSNFGKALEV NLVRYVLLTI RKATASQTYD HLLQINLERL GSGALSYAEA LQAYPHIEQI
 370 380 390 400 410 420
 VIEQIAETKS PDSNATTTSV ADIEQTIVEL AAQVLEMPD EVNTQARFID FGVDSLRGLI
 430 440 450 460 470 480
 LLDVNRRLG LTLQIPVLFD HSSIHELSTY IASQLPSSPQ AVVSAPAASV PVVSPAVVS

SVRS

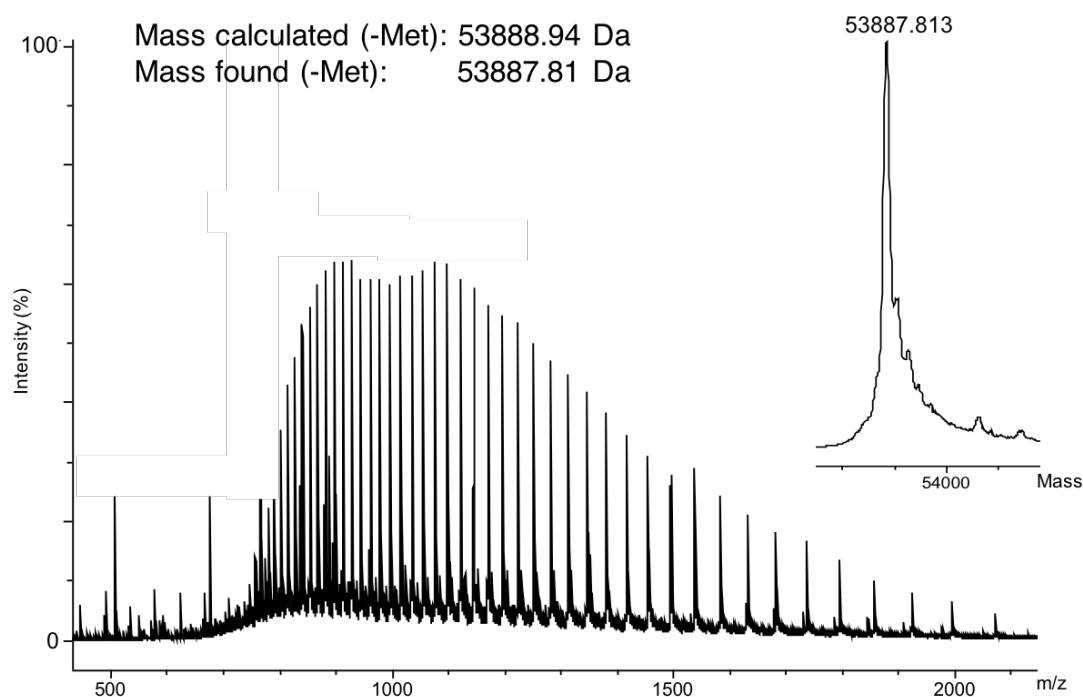


Figure S8: Raw and deconvoluted ESI-MS spectra for MisM1 OMT-ACP.

RizM18OMT-ACP

10 20 30 40 50 60
 MGSSHHHHHH SSGLVPRGSH MRVETGAKVE VKEESPNVVS GYYNRFAAEV DPKEEIYLI^F
 70 80 90 100 110 120
 PFFPEKLPGF SWLMTFFEPH RRPEHARYML ARQKDMKAVL YRHVDFSRVR TVMDIGCGLA^I
 130 140 150 160 170 180
 TDLIQLAQRH PKLSADGFTI AQKQVDVART RIARAGLSER LHIDCRDSSR DPFPRRYDLI^I
 190 200 210 220 230 240
 IGFEVVFHIE DKNGVFSNIA EHLEEGGLVV LADGVANTVT EINLPHLGQF TSTVPQLAEI^I
 250 260 270 280 290 300
 LGRHGLRVVS CVDTSEGVSR FLDDDRLEEN LARLRSEYPE SAHDEGQHRG WANFGKALGA^I
 310 320 330 340 350 360
 RLIRYLLFTL EKAPAGEDRA GLVLLNQSRI EGAQPFPSAL QALEAKRLDP VRTASSADIS^I
 370 380 390 400 410 420
 IERQLVAVAA EVFEMAPTEL DPTASFKDYG IGSLTGLRFI DAVNRRRLHLR LKMEHFFQCP^I
 430 440
 DLRSLSAFIR EHEEVRAAVV SQ

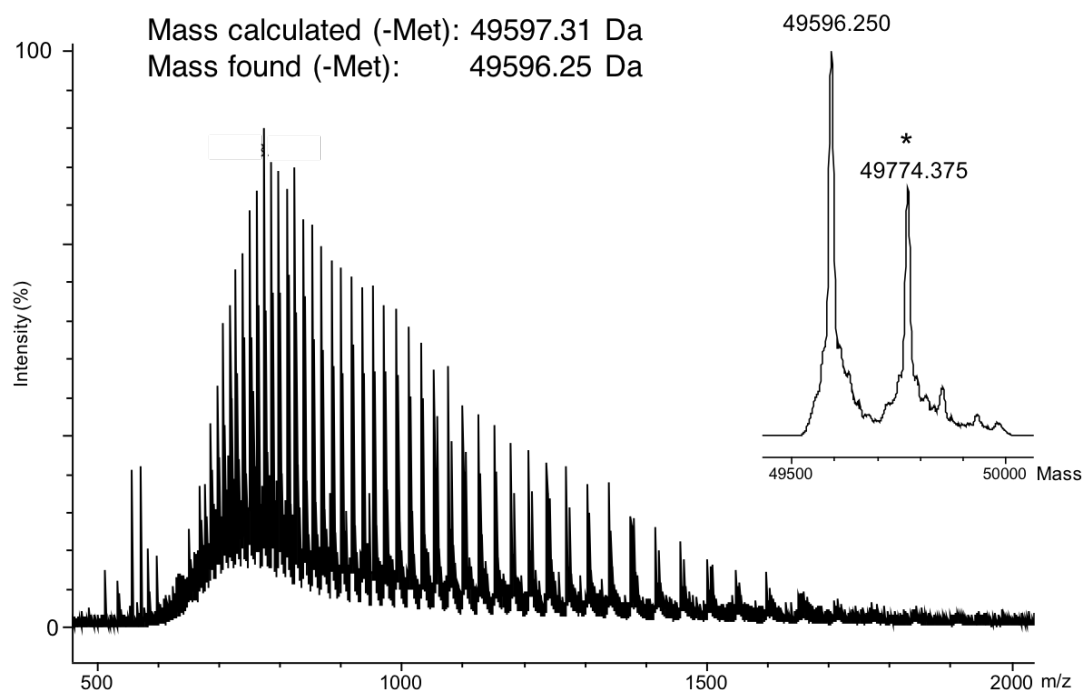


Figure S9: Raw and deconvoluted ESI-MS spectra for RizM18OMT-ACP.

GldM11ACP(I)

10 20 30 40 50 60
 MGSSHHHHHH SSGLVPRGSH MDEEDETMSA PLQVAADHDA TDAFQRWLHA GIRAGVESLL
 70 80 90 100 110 120
 KLSGERLRDT ANFADLGFDS IGLAHFGRWL SRRFGIDVTP DLFFDHSTIV ALVAHLSAAS
 130 140
 REALIERWRA DTRGAAPPT AARADLAR

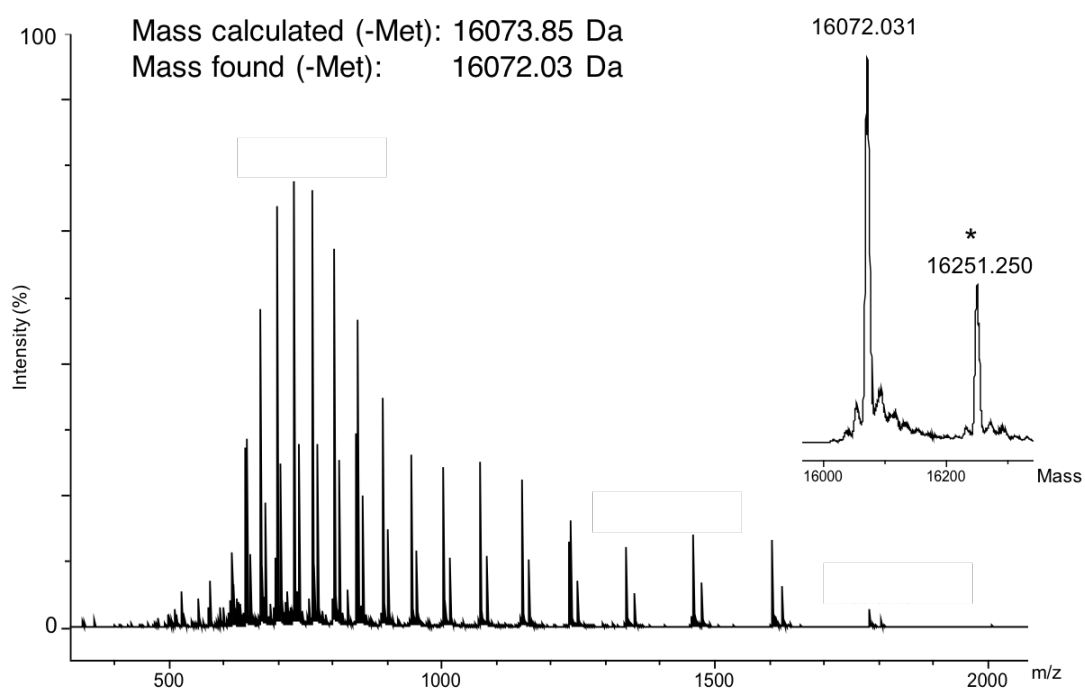


Figure S10: Raw and deconvoluted ESI-MS spectra for GldM11ACP(I).

GldM11ACP(II)

10 20 30 40 50 60
 MGSSHHHHHH SSGLVPRGSH MPSAGARAAT VREALERVMI DTLGVSPAEL RAARSFAELG
 70 80 90 100
 VDSSLAVRLL EAINVAFDLH EPTSVMFEFH DTQALAAHLE TRGAR

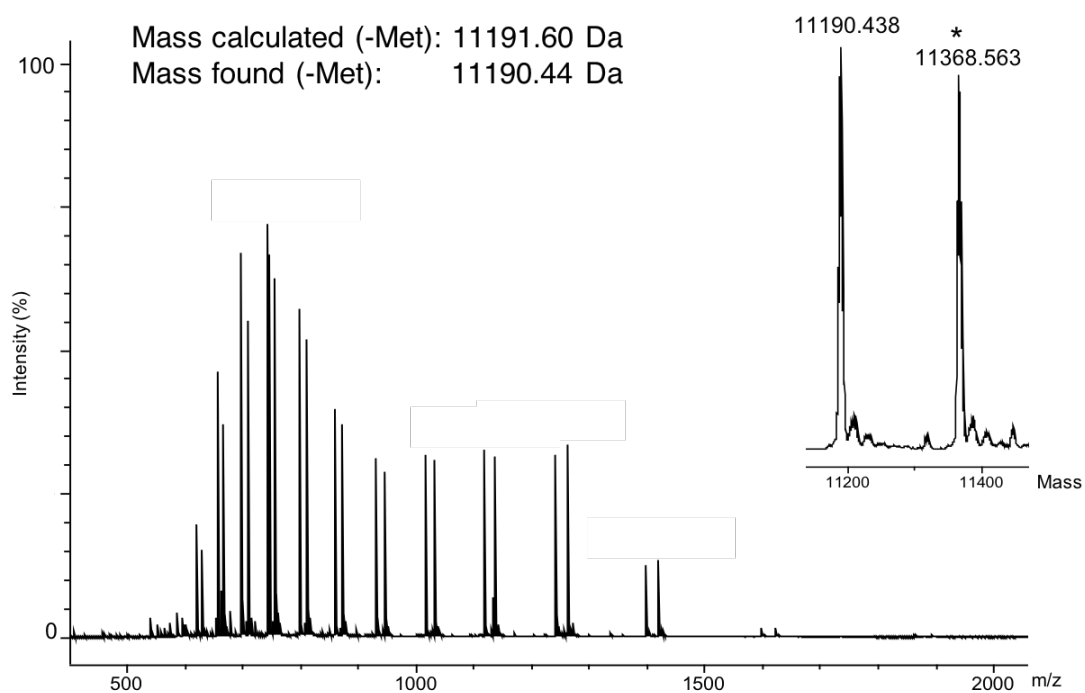


Figure S11: Raw and deconvoluted ESI-MS spectra for GldM11ACP(II).

MisM9ACP(II)

10 20 30 40 50 60
 MGSSHHHHH SSGLVPRGSH MQPMDGQPPS TLDESVPDLG QIQTALAGIF VAILGFETRE
 70 80 90 100 110 120
 LDFIDQLPLT ELGISSVNAV ELLEAINTF DLHLPTSLVF ECHRLEDLAG HIQVELGRKP
 130
 IAPAMSNASV AVQAEAESR

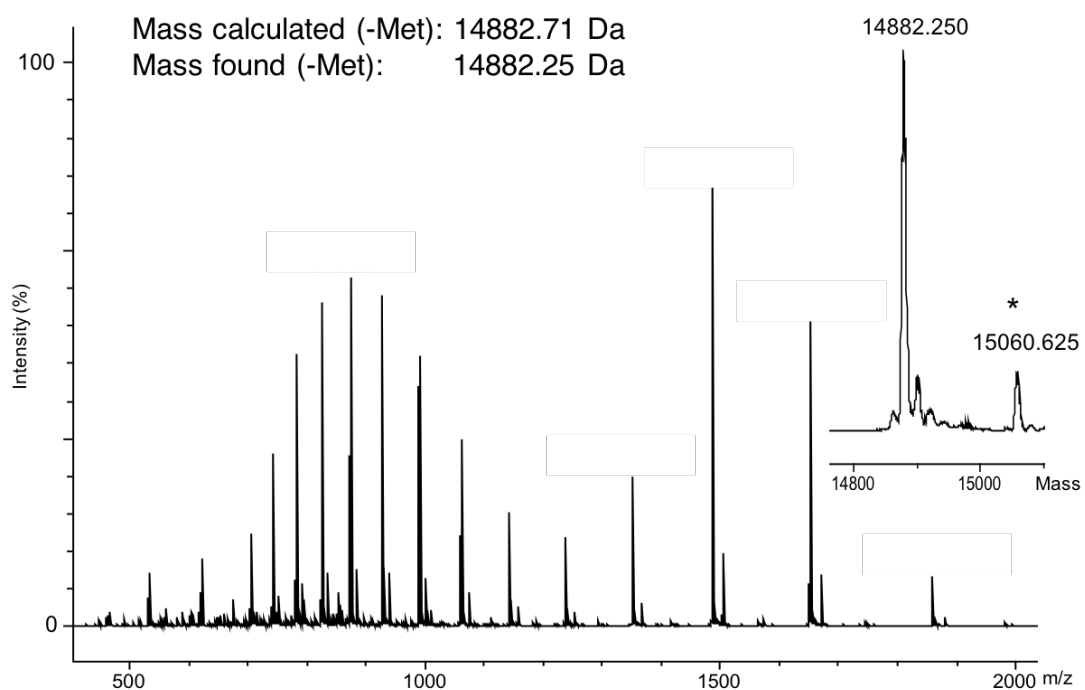


Figure S12: Raw and deconvoluted ESI-MS spectra for MisM9ACP(II).

MisM1ACP(II)

10 20 30 40 50 60
 MGSSHHHHH SGLVPRGSH MTSVADIEQT IVELAAQVLE MMPDEVNTQA RFIDFGVDSL
 70 80 90 100 110
 RGLILLDAVN RRLGLTLQIP VLFDHSSIHE LSIYIASQLP SSPQAVVSAP AASVPVVS

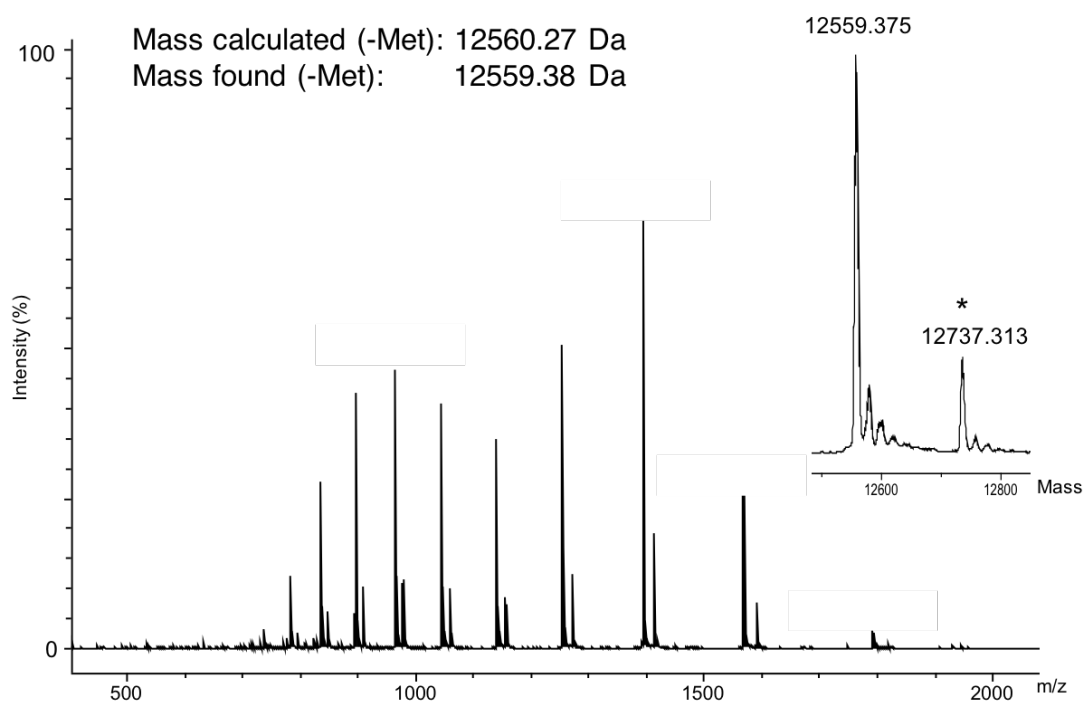


Figure S13: Raw and deconvoluted ESI-MS spectra for MisM1ACP(II).

RizM18ACP(II)

10 20 30 40 50 60
 MGSSHHHHHH SSGLVPRGSH MASSADISIE RQLVAVAAEV FEMAPTELDP TASFKDYGIG
 70 80 90 100 110
 SLTGLRFIDA VNRRLHLRLK MEHFFQCPDL RSLSAFIREH EEVRAAVVSQ

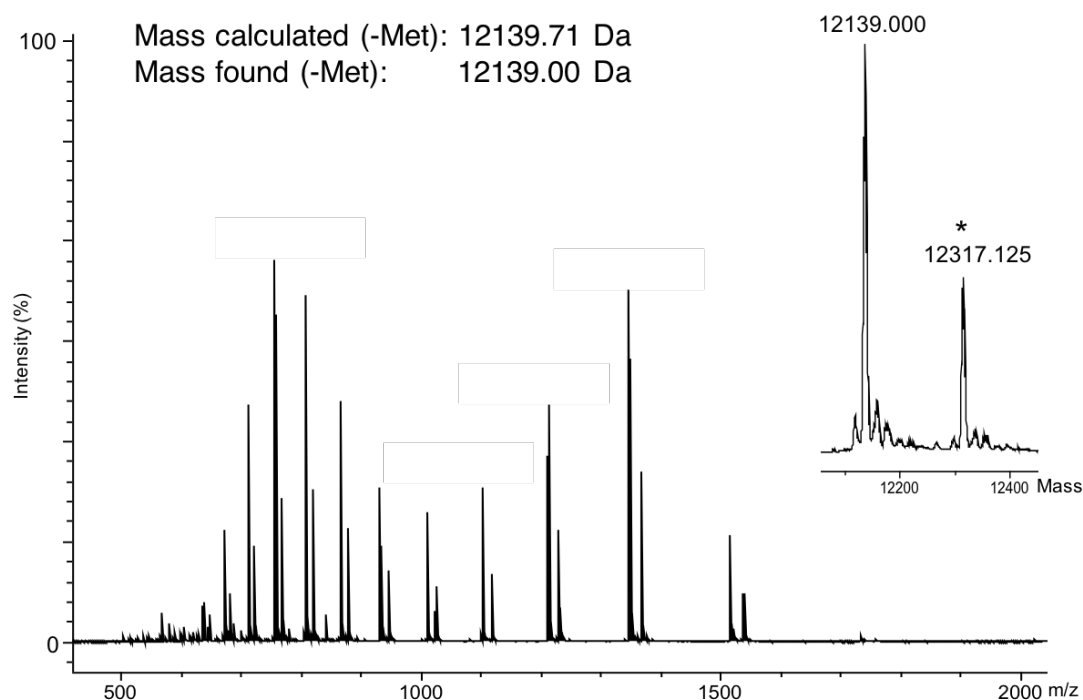


Figure S14: Raw and deconvoluted ESI-MS spectra for RizM18ACP(II).

MisM1 OMT

10 20 30 40 50 60
 MGSSHHHHHH SSGLVPRGSH MQSMPDLISQ KTHETRVNMS DIGQVVADYY DQVTESLKQK
 70 80 90 100 110 120
 IGSQEVYLLF APFQERVEGF SWLKMFFEPE KHREHFELTL QKQHELKSML YRLVEFANVH
 130 140 150 160 170 180
 RIFDIGCGYA TDLIQLAKRH PQVRGWGFSI SLNQVEFGQQ RIQDEGLGER VQLSCNDSTK
 190 200 210 220 230 240
 DPLPGLFDLI IGFEVVVHIA DKQGVFGNMA RHLSNNGTIV LADCVANTVT SINTSYIGQY
 250 260 270 280 290 300
 TSTAEEYSDI LAQNGLAIVD CVDVGPEISN FLYDPNYLEN AAYLSSIYPE MVNIEKEHQG
 310 320 330 340 350 360
 WSNFGKALEV NLVRYVLLTI RKATASQTYD HLLQINLERL GSGALSYAEA LQAYPHIEQI

V

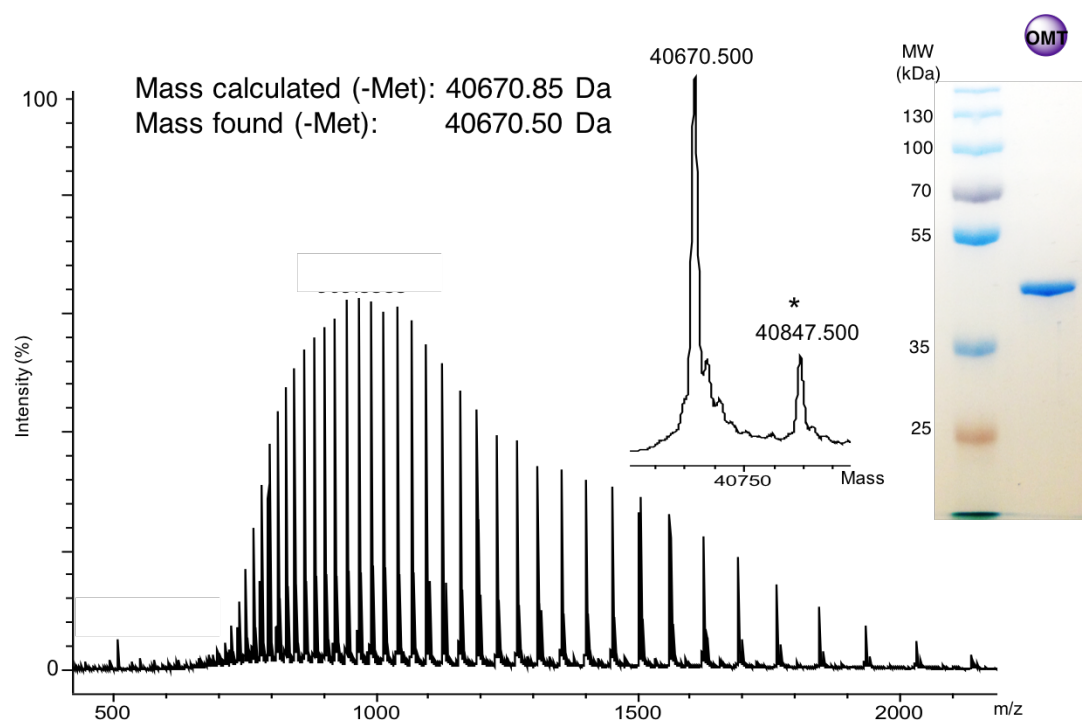


Figure S15: Raw and deconvoluted ESI-MS spectra for MisM1 OMT. 10 % SDS-PAGE analysis of the protein is also shown.

GbnM

10 20 30 40 50 60
 MGSSHHHHHH SSSLVPRGSH MMTPELDLES TDASALLADL RGGRRTAQDA VAASLARIER
 70 80 90 100 110 120
 LNPRVNIAIVT LDAERAMNEA RAADARLAAG QPAGLLHGLP ITVKDAIFTE GMRTTAGHAP
 130 140 150 160 170 180
 LRDQVPERDA TAVARLRAAG AIVIGKTNCA TLCSDLQTNV PVFGRTSNPW DASRTAGGSS
 190 200 210 220 230 240
 GGDAAVALG LVPLGIGTDT GGSIRVPAAY CGVYGFKPSL AAVPGDGVIP PLASGEGHAD
 250 260 270 280 290 300
 SLTTIGPIAR SVRDLRLAYR VLSGDALADA PAPAATRIAW TDPFAAVPVD DSVRAALARA
 310 320 330 340 350 360
 RDLLSATGLQ VERREPPFDP RRVNYHYLLL SLYEFTPREL KPGVHFAPRL AETLRSLFGG
 370 380 390 400 410 420
 GLAGSYAKVK AWQRTQIDAY QAFADGIDAW VIPATPTVAF PHQSGARALN LVSEGKRVQA
 430 440 450 460 470 480
 GYWNAAGGYA FAANLLGCPS VTIPIGQDAA GLPIALQALG RPGGDAALLR AAEAIAERIG
 TVRTPPGN

GbnB

10 20 30 40 50 60
 MHHHHHHGKP IPNPLLGLDS TENLYFQGID PFTMTALLFP GQGAQEASMC EGVAAMPAPA
 70 80 90 100 110 120
 SRYRFVCDAA GADILDAVAR HGERVLHRNL FSSLLTVLVA SLSLDAFRAR EAAAPEAVAG
 130 140 150 160 170 180
 YSVGQWTALY AAGVVDFEQL VGLIVERARA MDACFEETAG GMSAVIGVRE EALERELASL
 190 200 210 220 230 240
 RESGHA VFIS NINCVGQYSI AGTLDALDRA EQRISGLRPK KLVRLPVSGA WHCPLLAEAE
 250 260 270 280 290 300
 QRFAAVLARV RLASPRLPVA DNVTTGGWLPE APDALRATLA AHLTRPVWRW DCVRTLIAAG
 310 320 330
 STRLVEVGFG KQLSRFGIFI DRQVEFSTTY D

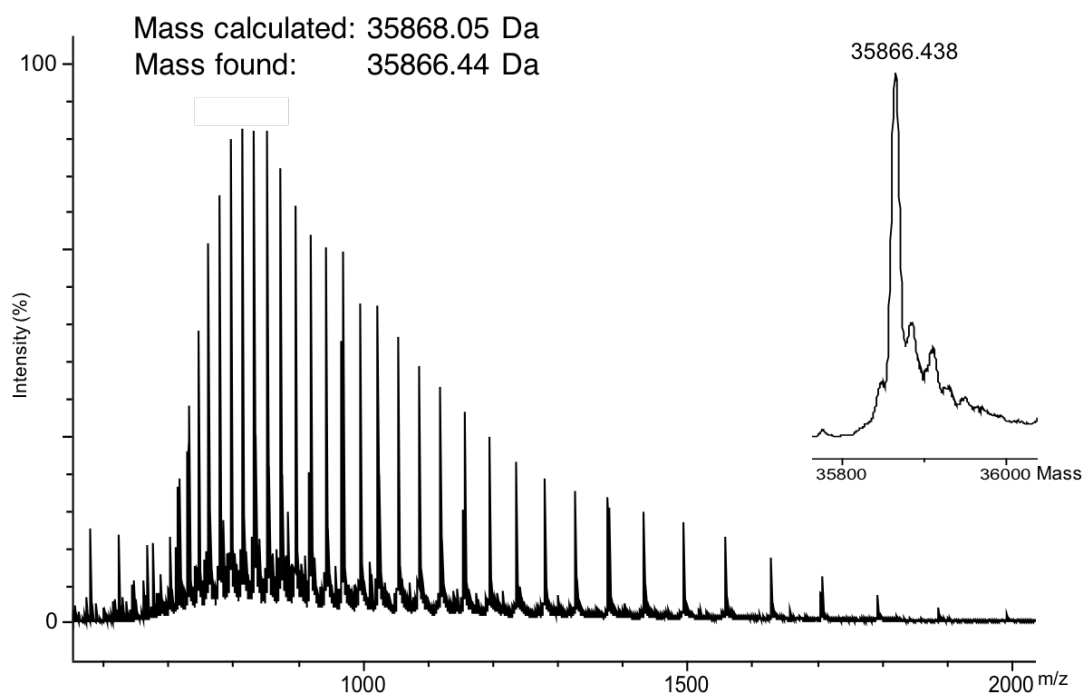


Figure S16: Raw and deconvoluted ESI-MS spectra for GbnB.

EtnK

10 20 30 40 50 60
 MKHHHHHHHH GGLVPRGSHG SHAGPRATVA FMFPGQGSQK RGMGAGLFDS VPEYAAVERE
 70 80 90 100 110 120
 VDALLGYSMR ALCHEDPDGR LKETQYTQPA LYVVNALHYY DAIARGQRPD YVAGHSLGEY
 130 140 150 160 170 180
 NALLAAGAFD LLTGLRLVKK RGELMAASVS GGMAAVIGMD EGRIKQVLTE NGLGTIDVAN
 190 200 210 220 230 240
 FNSPSQIVIS GPVADIARGH TVFEDAGART YVILPVSAAF HSRIVEETGR AFADFIAPMR
 250 260 270 280 290 300
 FEALRIPVIS NVSARPYEAK DPSATIKSL L VQQITRPVQW VQSVSFLMAQ GVKEFREIGP
 310
 GNVLTRLVQQ IERQPSGA

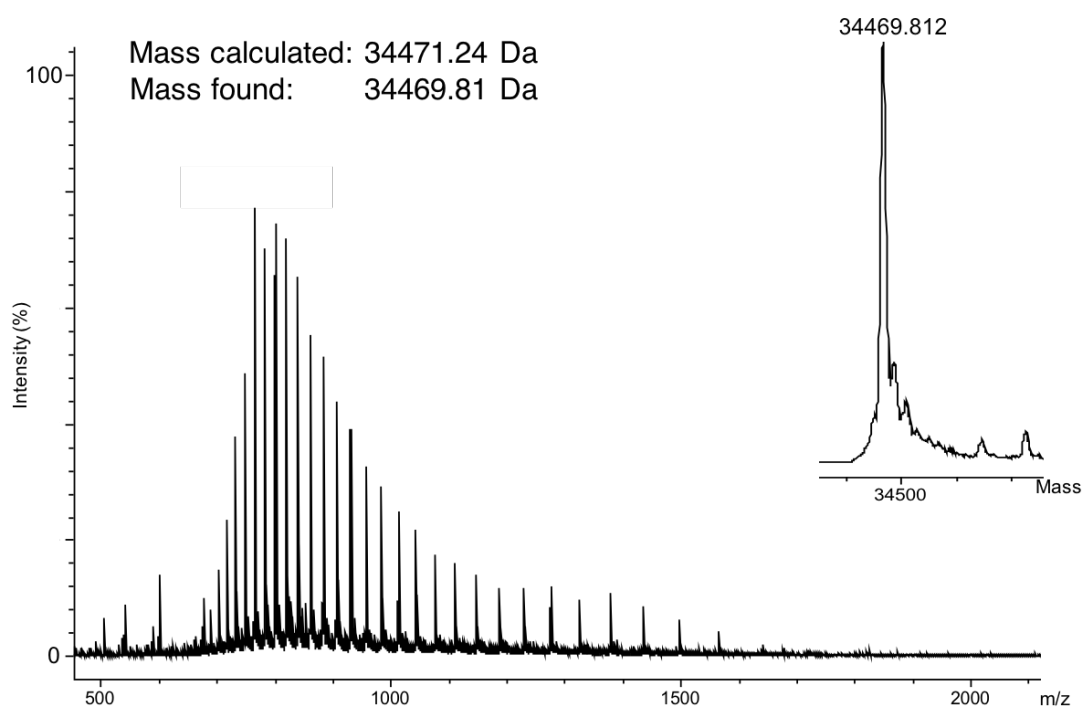


Figure S17: Raw and deconvoluted ESI-MS spectra for EtnK.

GbnA

10 20 30 40 50 60
 MGSSHHHHHH SSGLVPRGSH MQDKIQQFLS DRLLVDFDRD VDADSDLFQL GLMDSETYAE
 70 80 90 100
 LIQFIEKTFH LQFSDEEILS NVVASLSGIV SFVSEALEHR AAGSFVRA

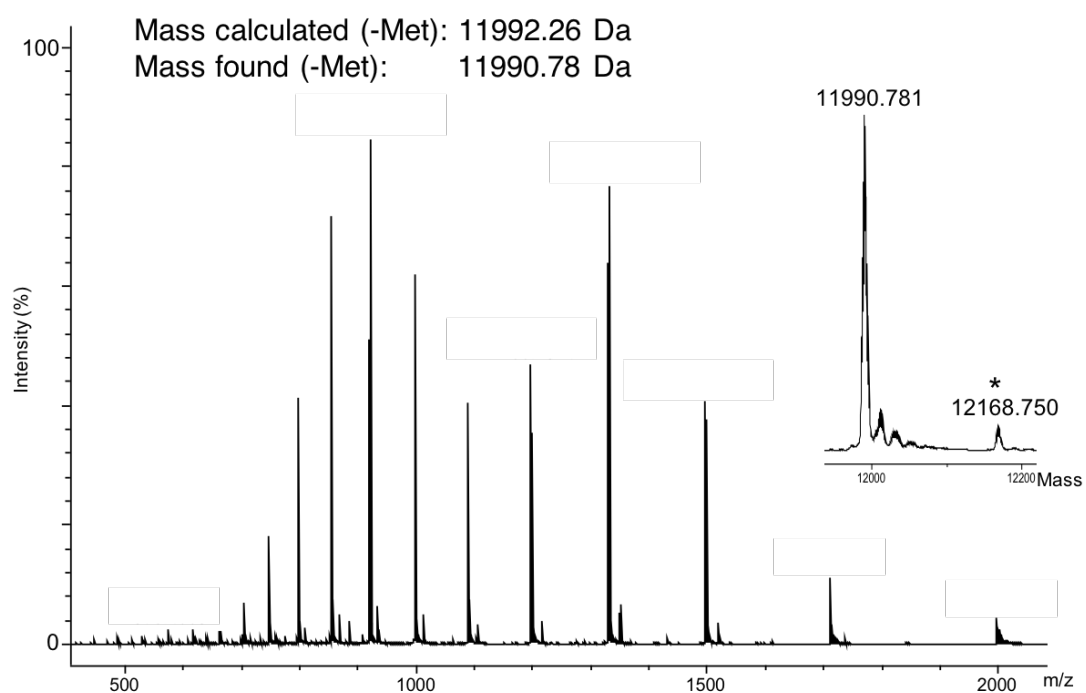


Figure S18: Raw and deconvoluted ESI-MS spectra for GbnA.

GldM1ACP(I)

10 20 30 40 50 60
 MGSSHHHHHH SGLVPRGSH MLPKAAVAPA SRAVVAASAK PVRDLAAELA RGLAQALCME
 70 80 90 100 110 120
 TDDETGVDDAD QPFFELGLDS ITGVEWTRAI NRAYGLQLPA TRLYDHPTIA QLTRYLETLL

AAPEP

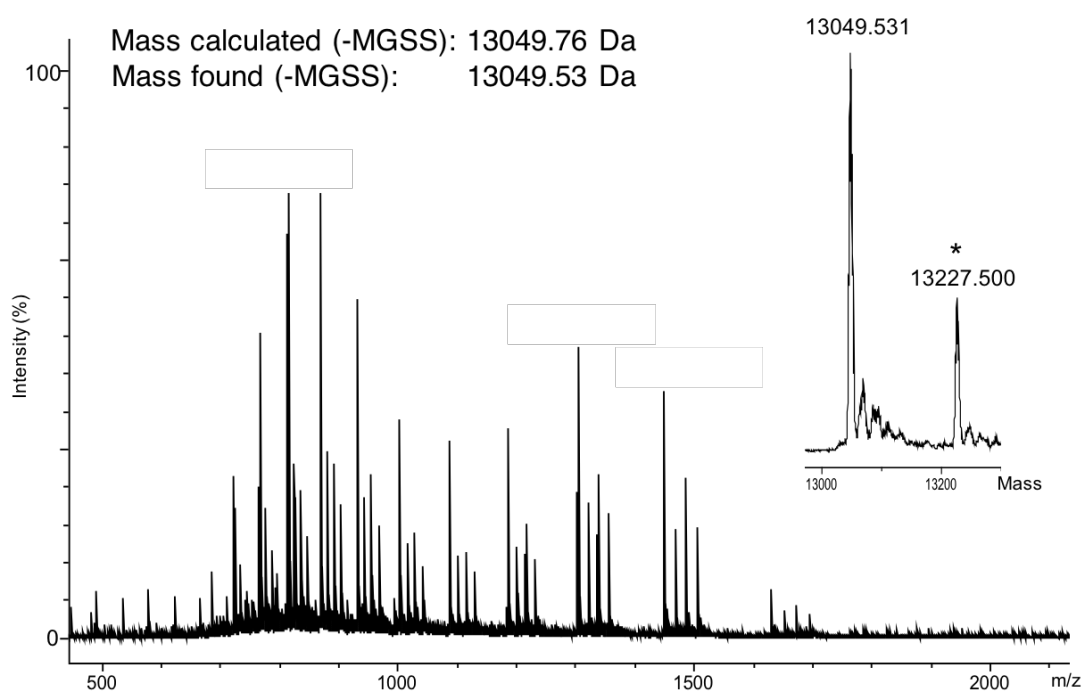


Figure S19: Raw and deconvoluted ESI-MS spectra for GldM1ACP(I).

Appendix 2: Sequence alignment of embedded OMT domains

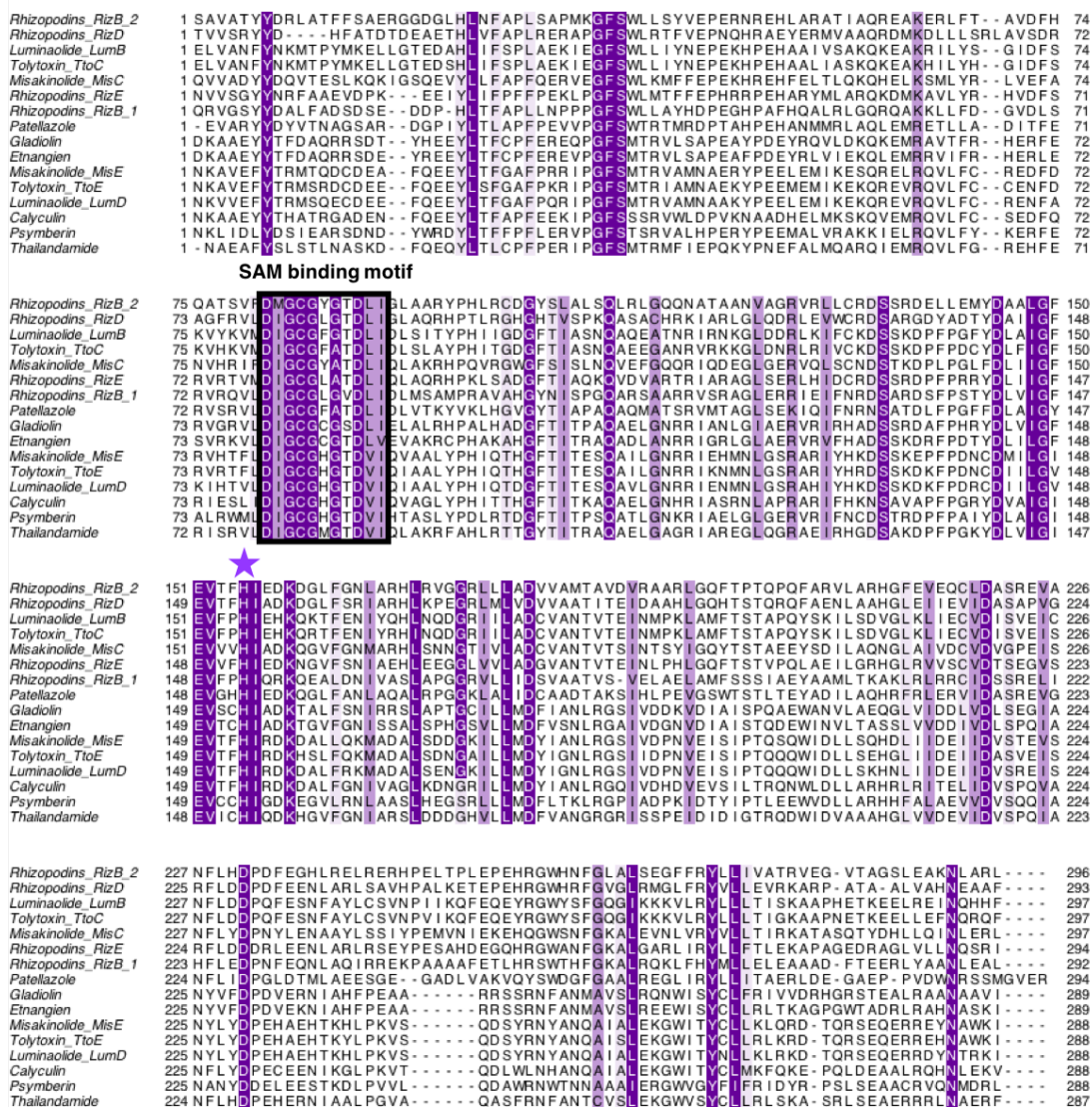


Figure S20: Sequence alignment of all embedded OMT domains from *trans*-AT PKSs. Residues are shaded if there is greater than 60% sequence conservation and the darker the shade of purple the greater the extent of conservation. The star indicates the proposed key catalytic base. This image was generated using JalView.

Appendix 3: Ramachandran plot for the GldOMT homology model

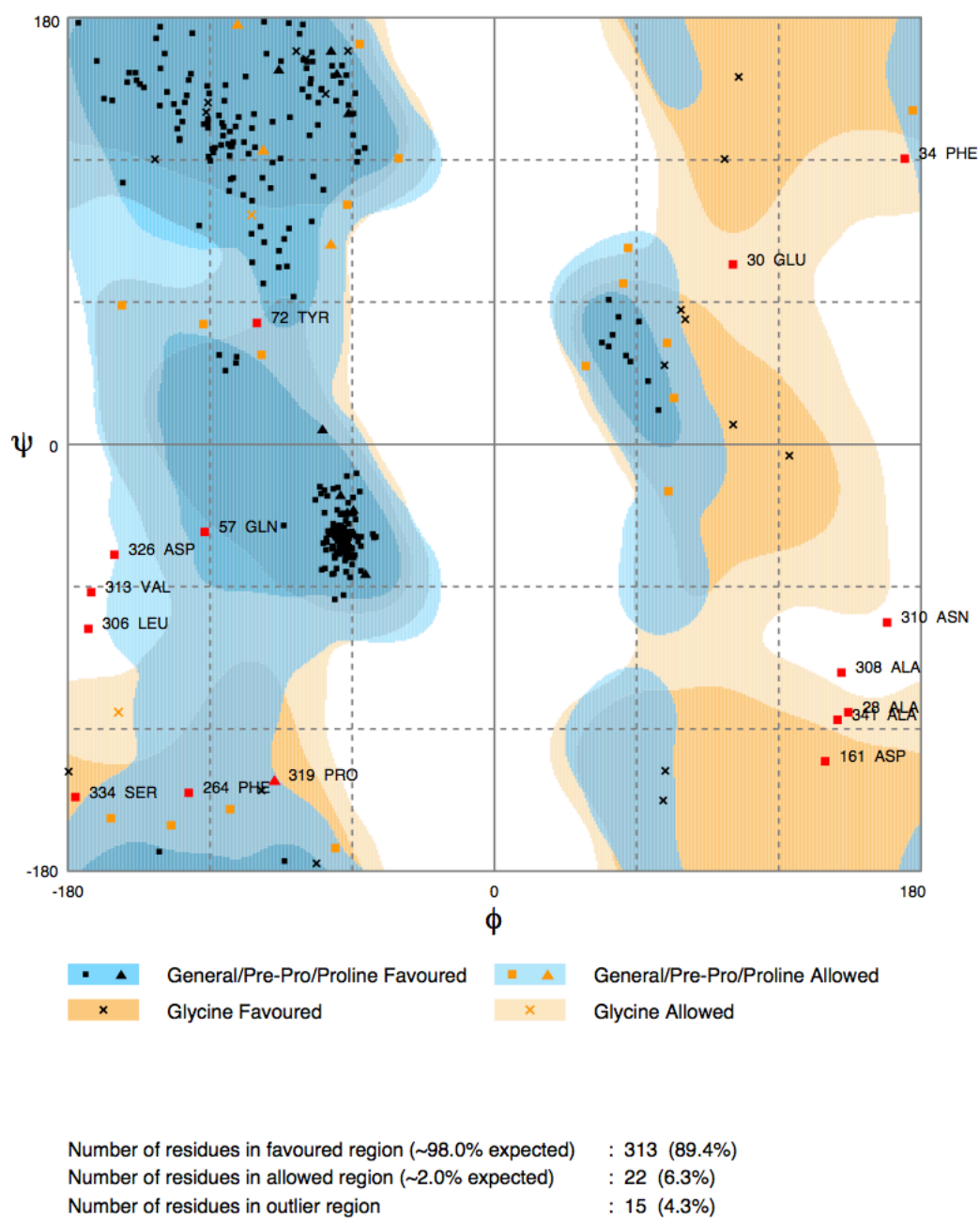


Figure S21: Ramachandran plot created using RAMPAGE software to assess the quality of the homology model generated for GldOMT.

Appendix 4: MS data for assays looking at the substrate tolerance of embedded OMT domains from two different clades

MisM9OMTACP

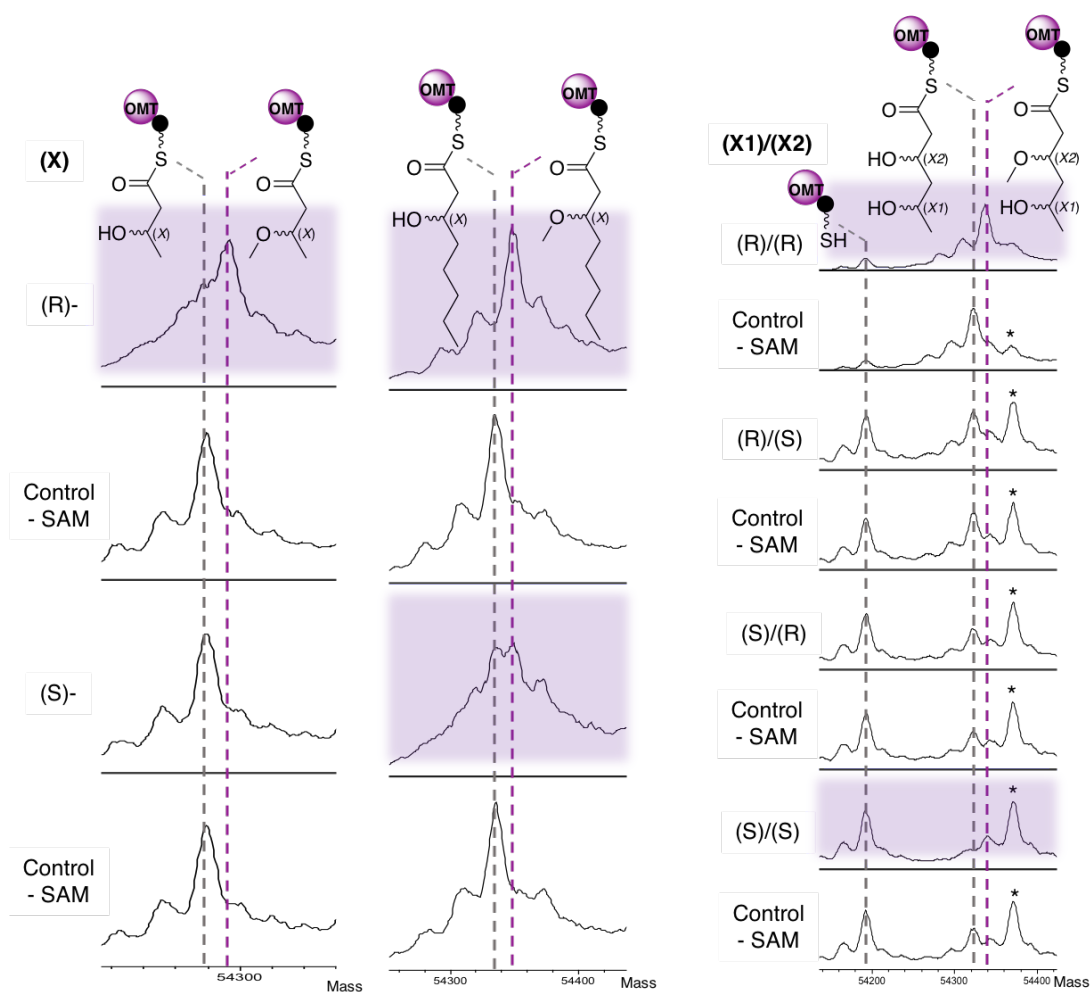


Figure S22: Stacked deconvoluted ESI-MS spectra of MisM9OMT-ACP with ACP bound simplified substrate mimics after incubation with SAM. The spectra where methylation (+14 Da) is observed are highlighted in purple. The peaks marked with an * are from the gluconoylated protein where the substrate has been lost through lactonization.

MisM1 OMTACP

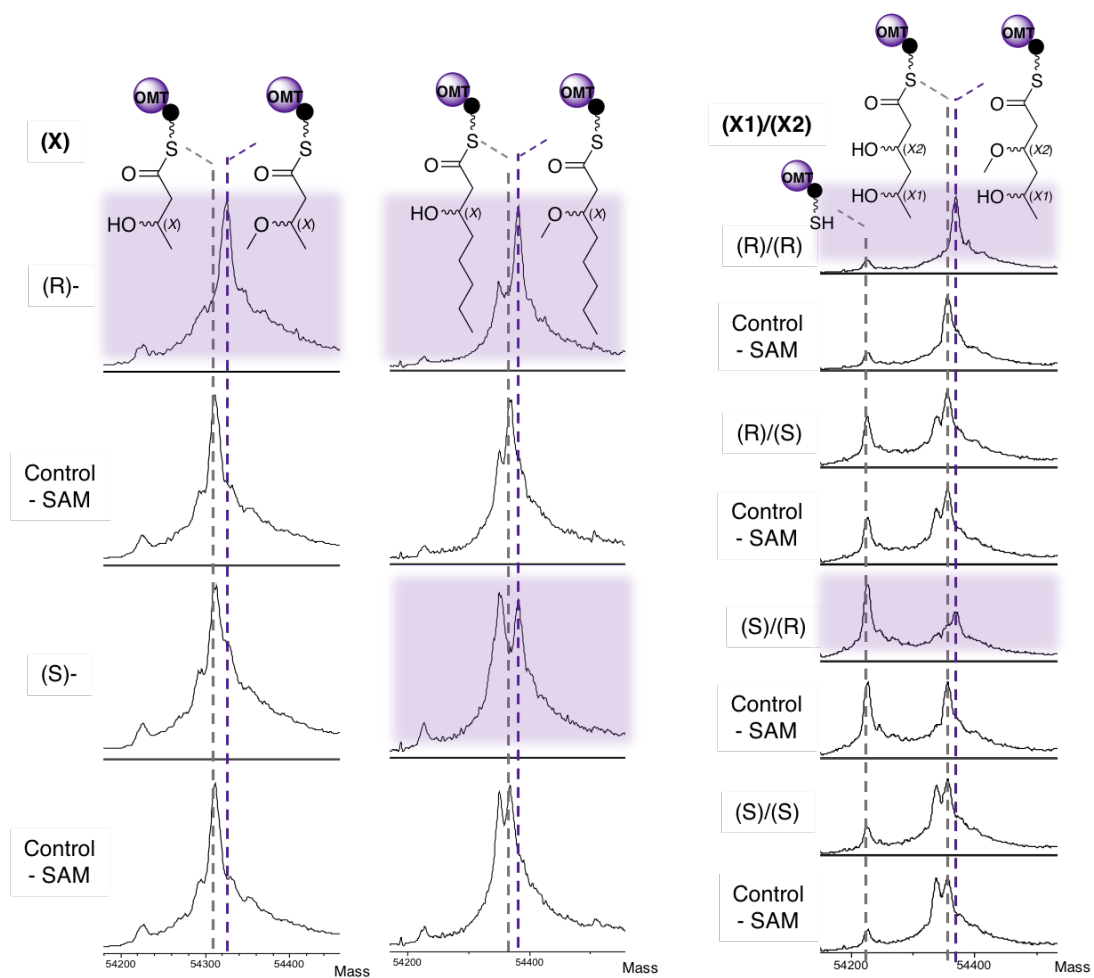


Figure S23: Stacked deconvoluted ESI-MS spectra of MisM1 OMT-ACP with ACP bound simplified substrate mimics after incubation with SAM. The spectra where methylation (+14 Da) is observed are highlighted in purple.

RizM18OMTACP

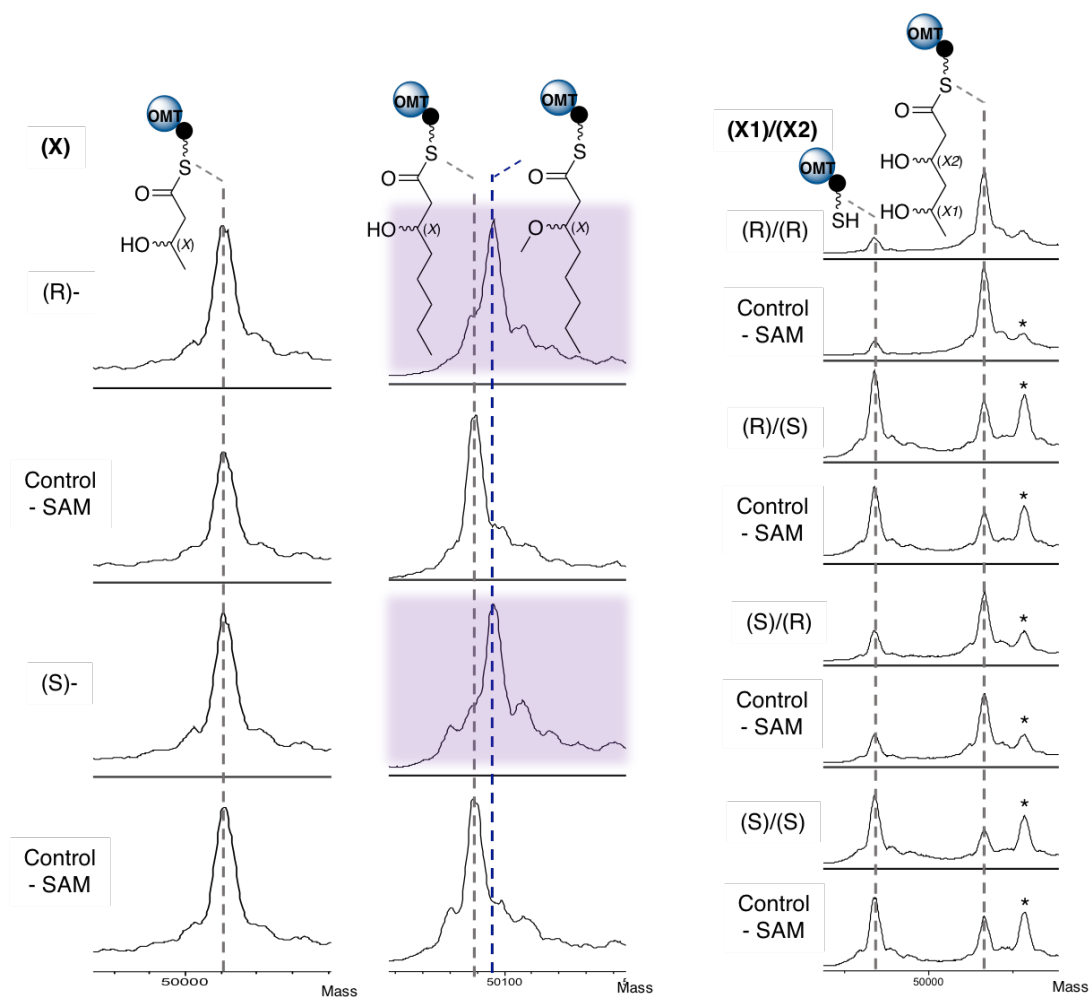


Figure S24: Stacked deconvoluted ESI-MS spectra of RizM18OMT-ACP with ACP bound simplified substrate mimics after incubation with SAM. The spectra where methylation (+14 Da) is observed are highlighted in purple. The peaks marked with an * are from the gluconoylated protein where the substrate has been lost through lactonization.

Appendix 5: Purification of the gladiolins

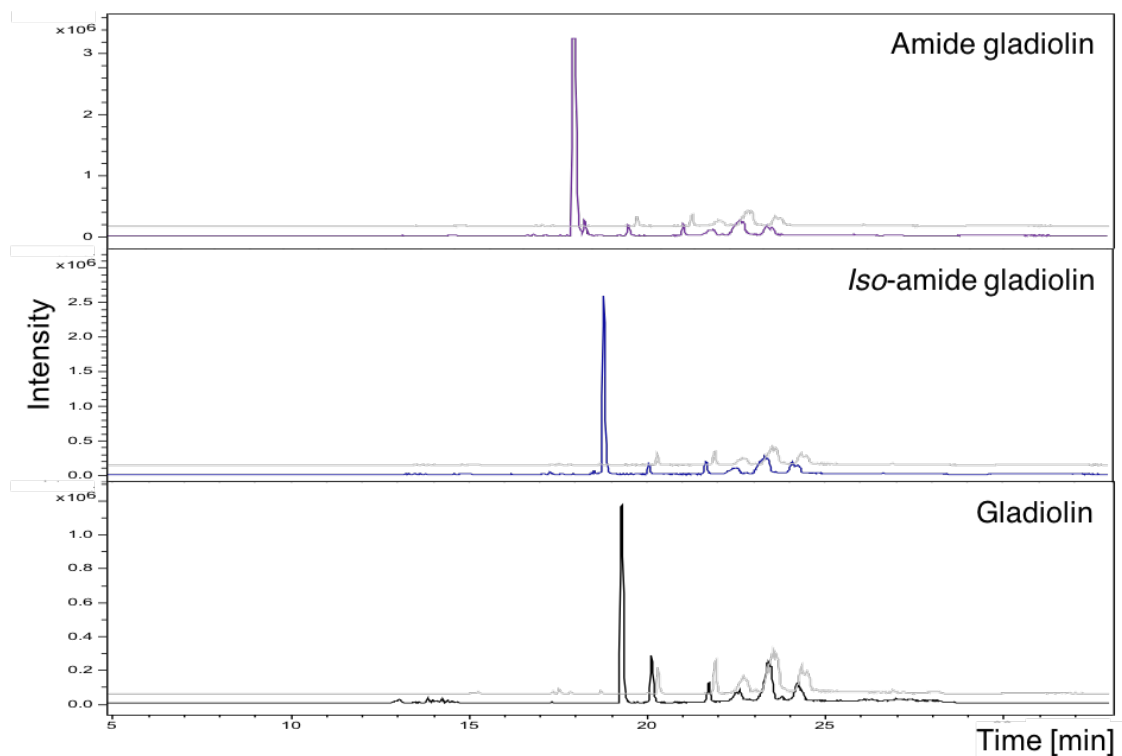


Figure S25: Base peak chromatograms from LC-MS analyses of purified amide gladiolin (purple), *iso*-amide gladiolin (blue) and gladiolin (black). For each chromatogram a blank is shown in grey to indicate peaks that are contamination from the column.

Appendix 6: High-resolution MS of amide gladiolin

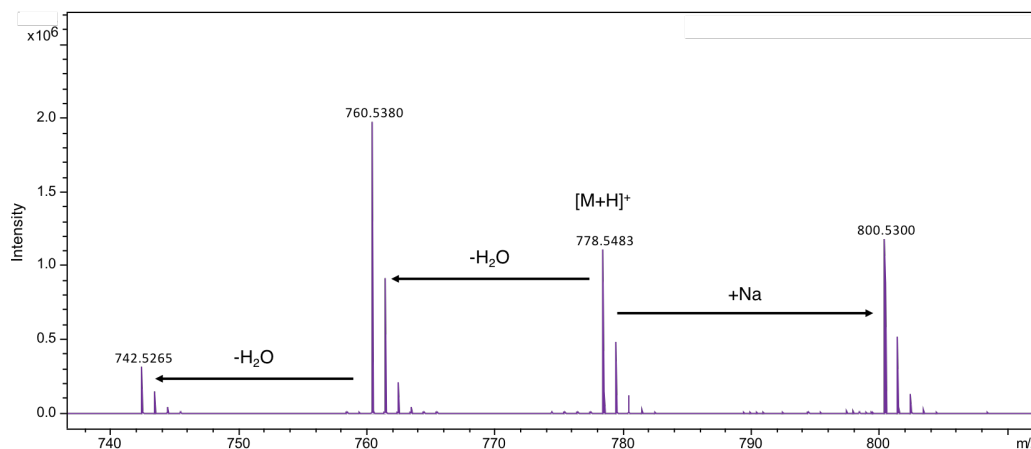


Figure S26: High-resolution MS analysis of amide gladiolin in positive ion mode.

Appendix 7: NMR spectra of amide gladiolin

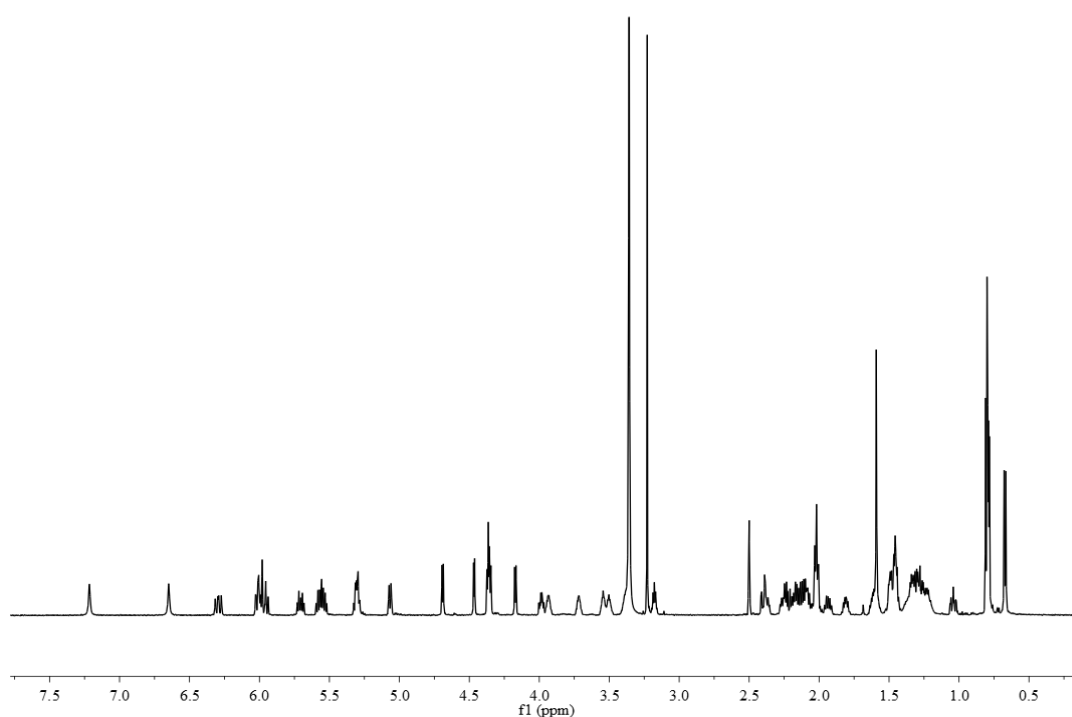


Figure S27: 1H NMR spectrum of amide gladiolin recorded in $DMSO-d_6$.

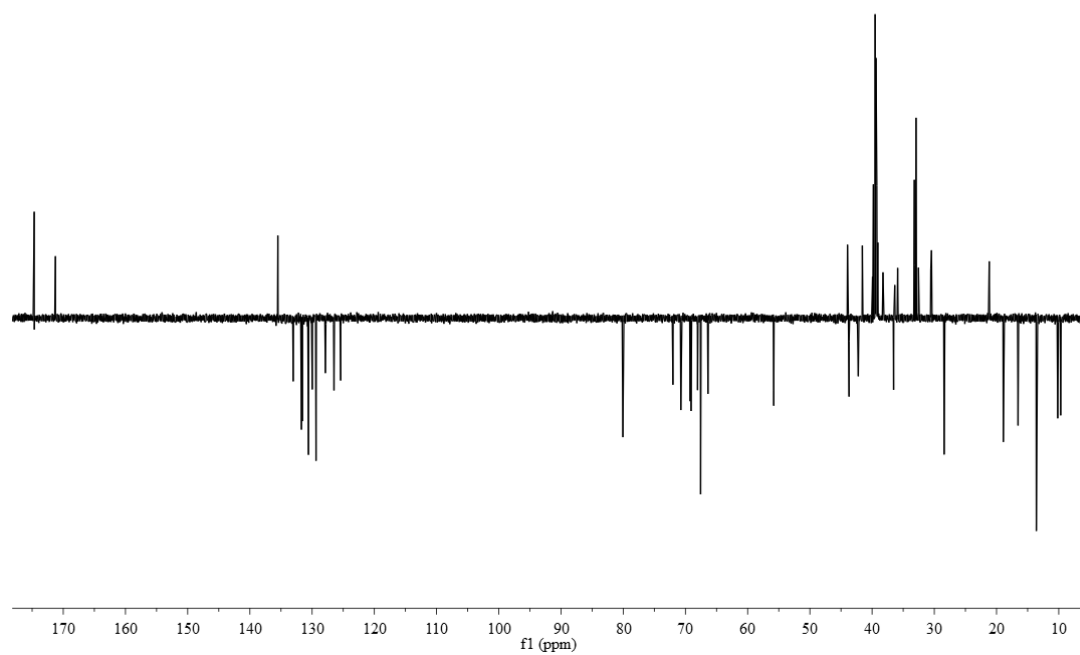


Figure S28: ^{13}C NMR spectrum of amide gladiolin recorded in $\text{DMSO}-d_6$.

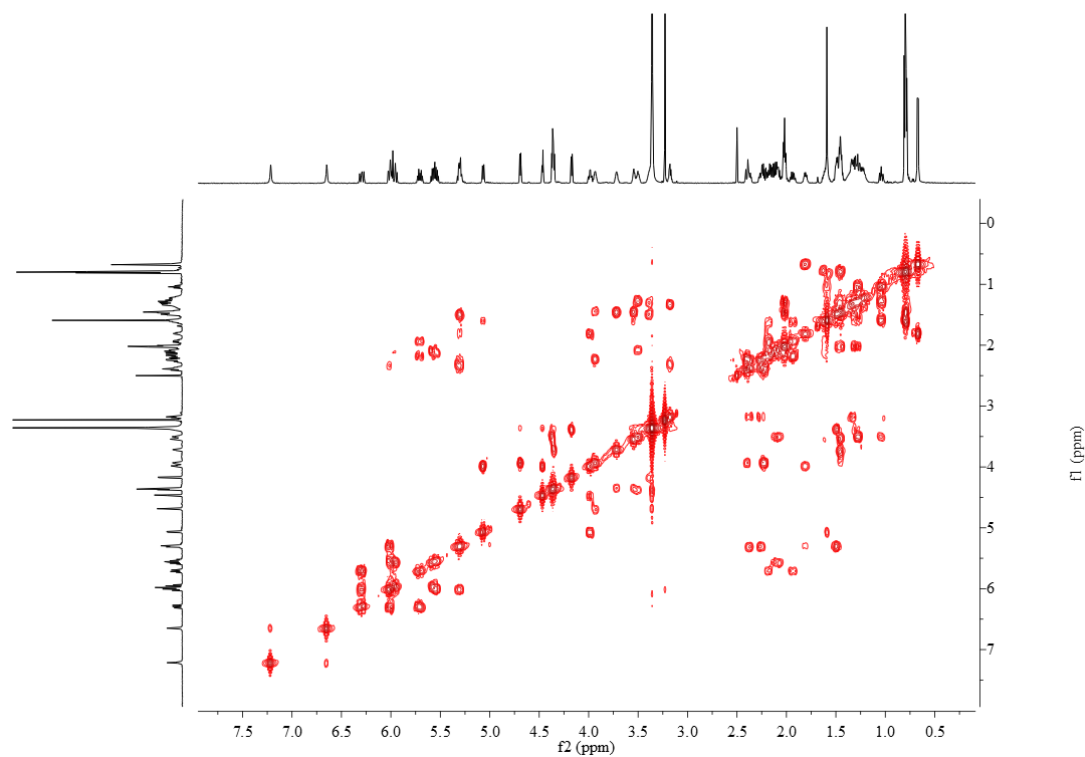


Figure S29: COSY spectrum of amide gladiolin recorded in $\text{DMSO}-d_6$.

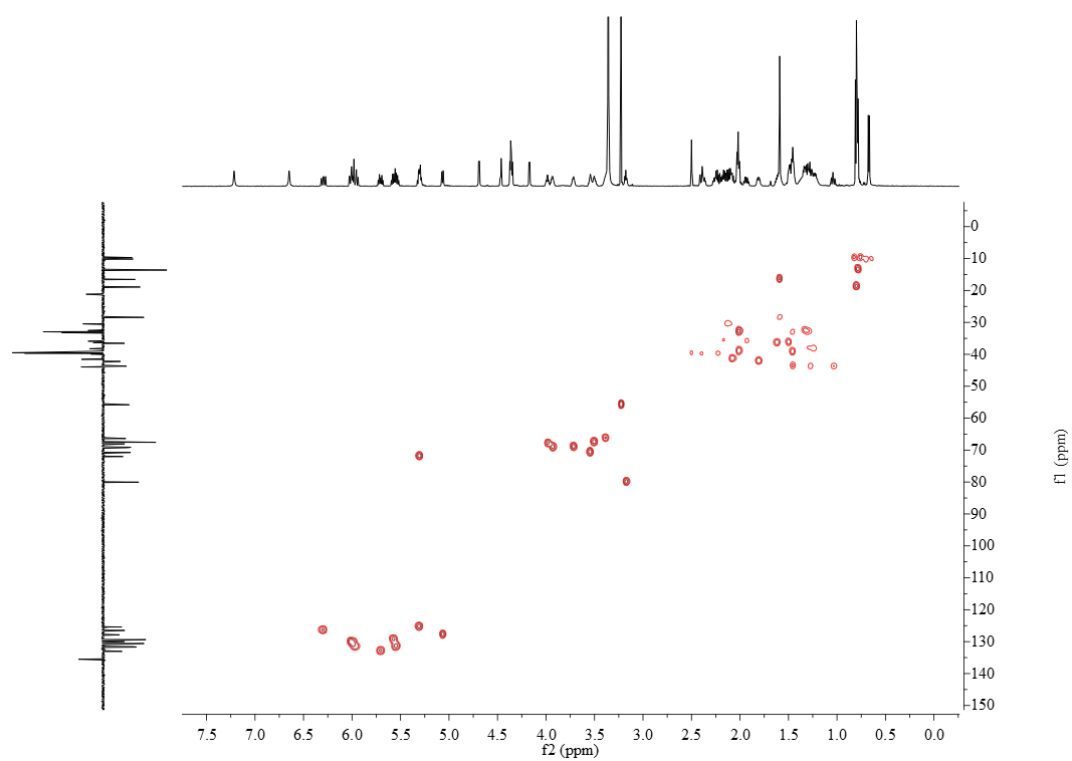


Figure S30: HSQC spectrum of amide gladiolin recorded in $\text{DMSO}-d_6$.

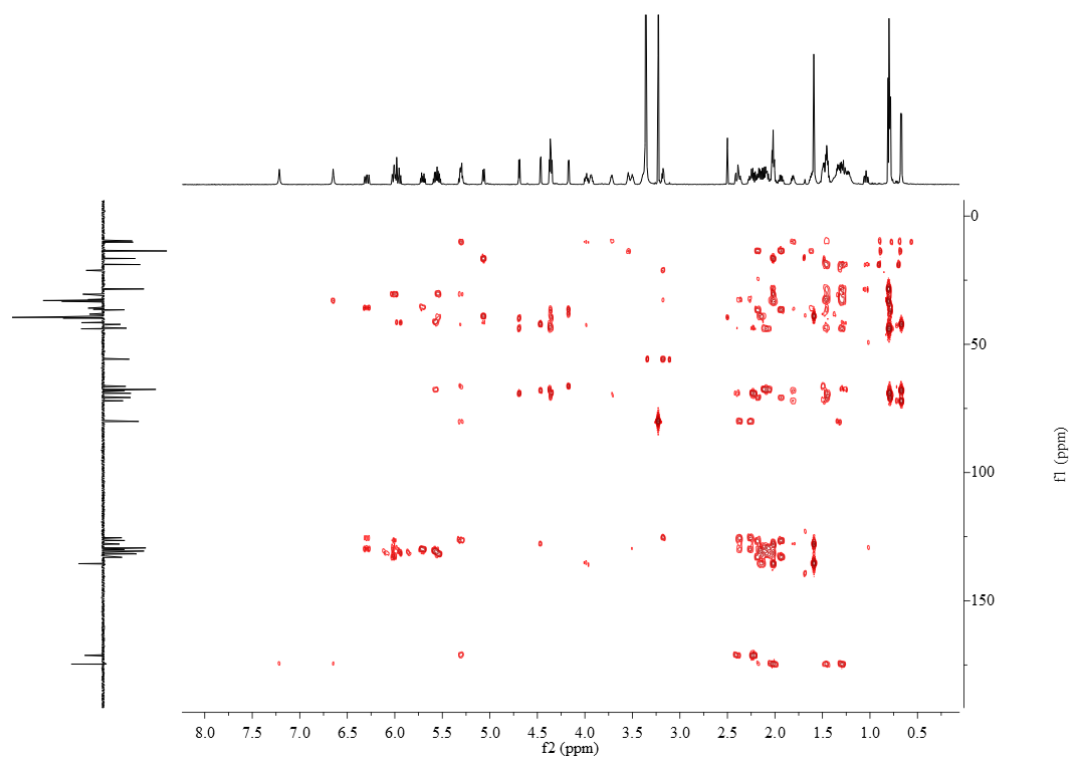


Figure S31: HMBC spectrum of amide gladiolin recorded in $\text{DMSO}-d_6$.

Appendix 8: Characterisation of *iso*-amide gladiolin

NMR spectra

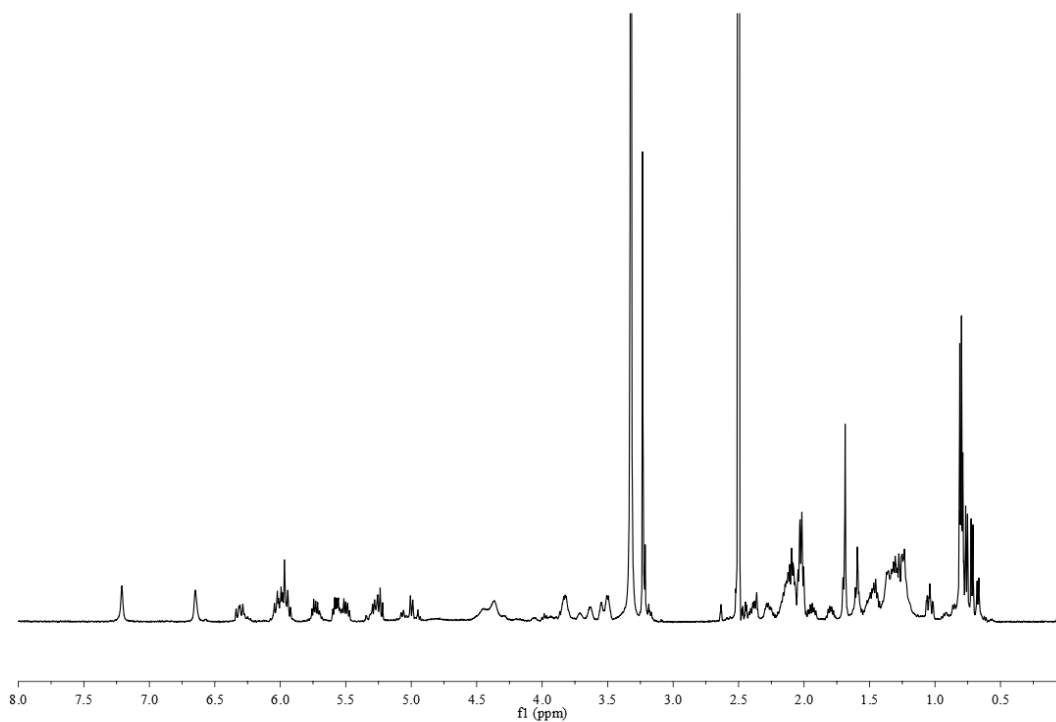


Figure S32: ^1H NMR spectrum of *iso*-amide gladiolin recorded in $\text{DMSO}-d_6$.

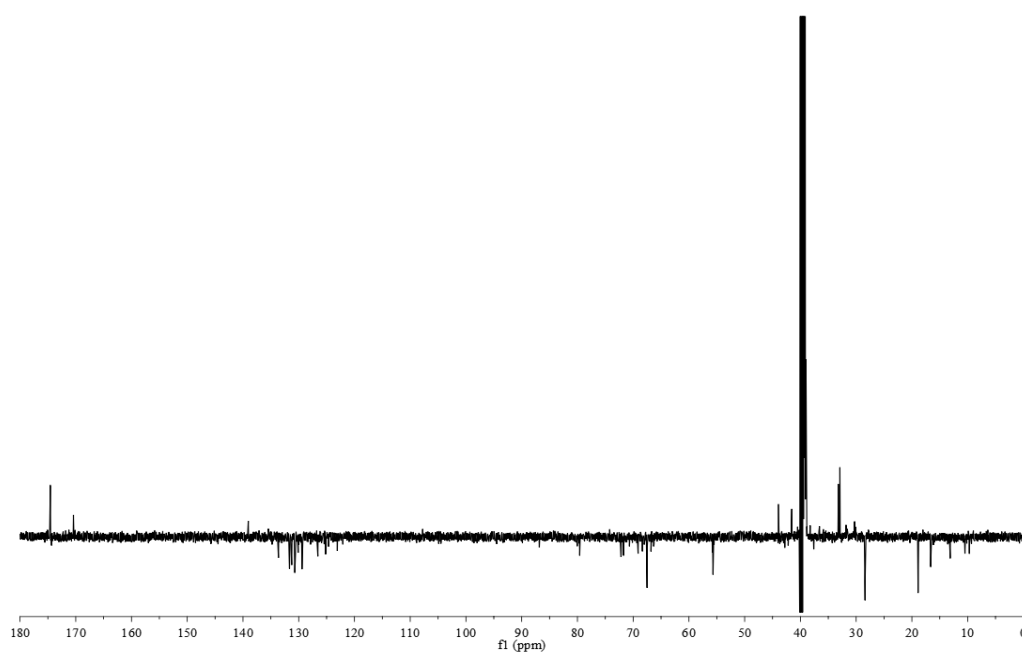


Figure S33: ^{13}C NMR spectrum of *iso*-amide gladiolin recorded in $\text{DMSO}-d_6$.

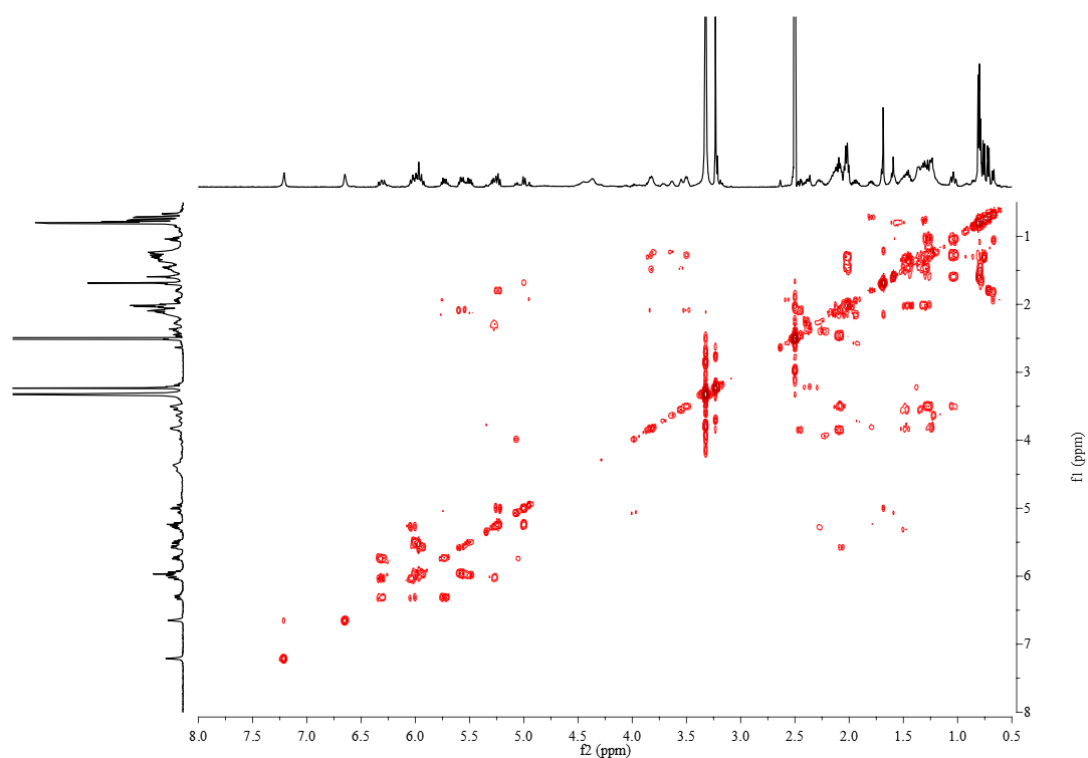


Figure S34: COSY spectrum of *iso*-amide gladiolin recorded in DMSO- d_6 .

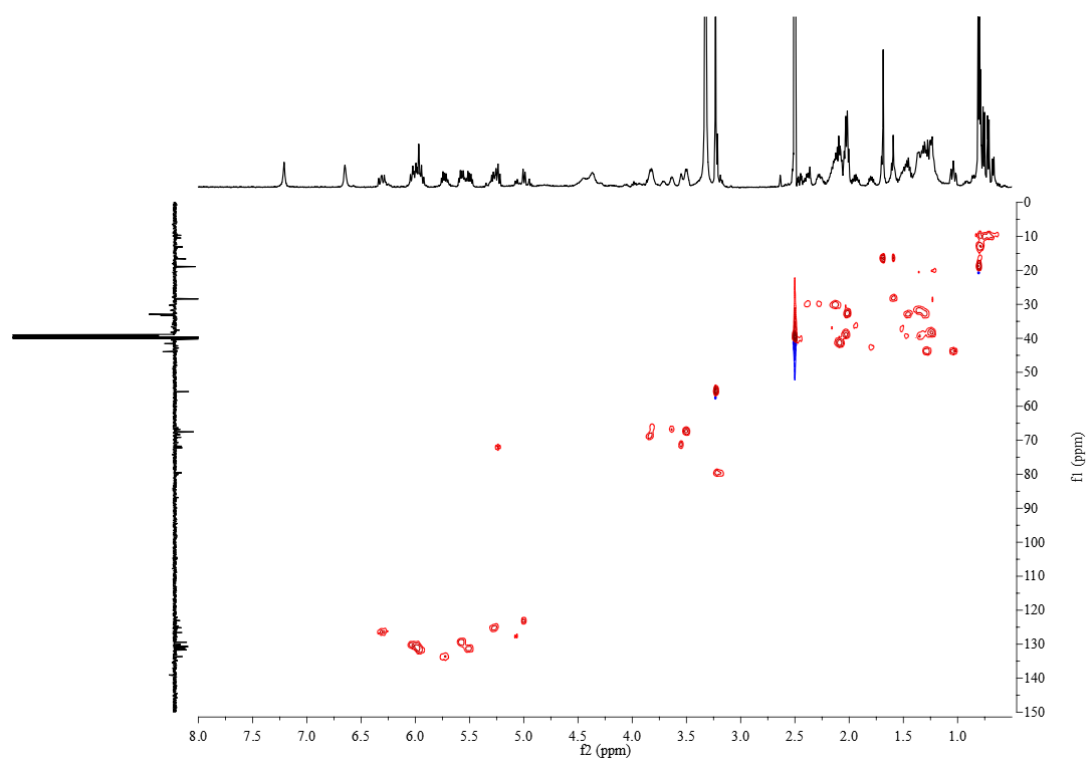


Figure S35: HSQC spectrum of *iso*-amide gladiolin recorded in DMSO- d_6 .

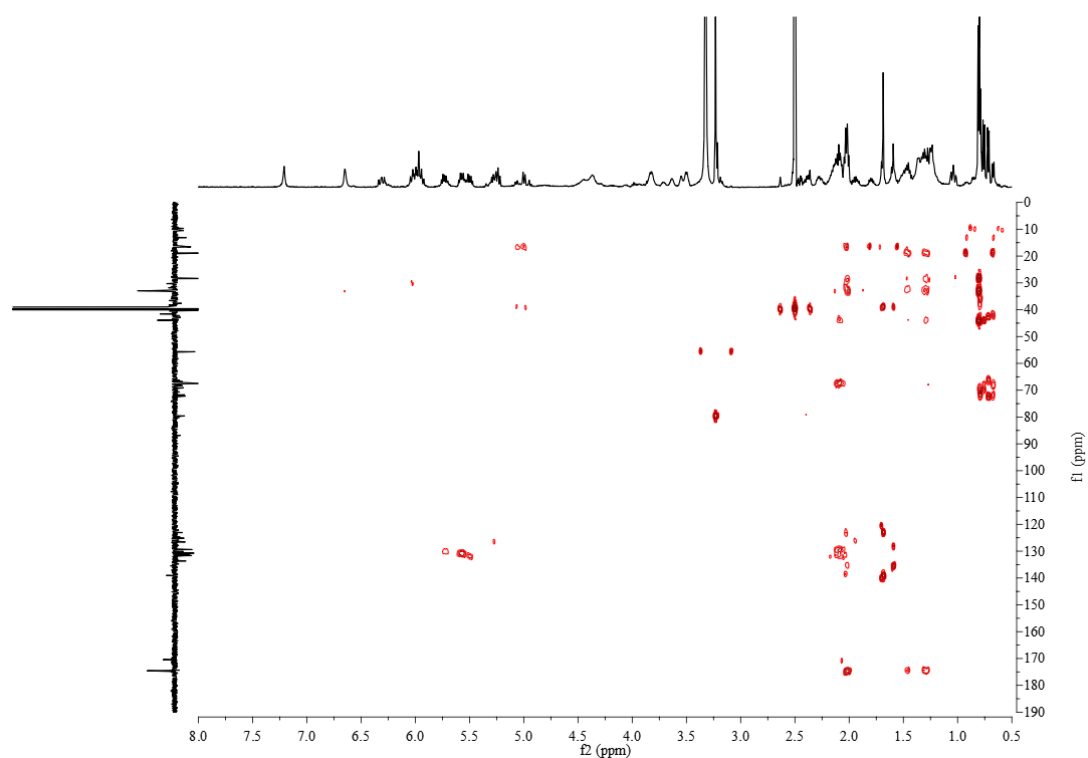
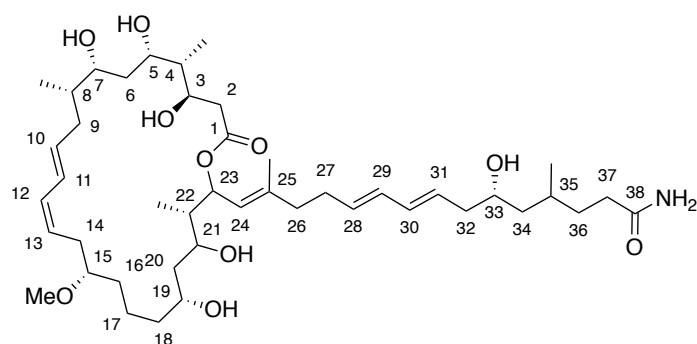


Figure S36: HMBC spectrum of *iso*-amide gladiolin recorded in DMSO- d_6 .

NMR assignment

Table S1: NMR assignments for *iso*-amide gladiolin (DMSO- d_6 , 600 Hz). * indicates that the signal was found under the DMSO- d_6 peak. Multiplicity is denoted as follows: s = singlet, d = doublet, dd = doublet of doublets, dt = doublet of triplets, br = broad, m = multiplet and ol = overlapped.



Position	δ_H (J, Hz)	δ_C
1	-	170.41
2	2.15, 2.45, 2 x 1H, 2 x m	39.6*
3	3.72, 1H, m	68.35
4	1.45, 1H, m	43.5
4-Me	0.76, 3H, d, (6.9)	9.7
5	3.84, 1H, m	69.1
6	1.35, 1.48, 2H, m, ol	39.3*
7	3.55, 1H, m	71.7
8	1.51, 1H, m	37.7
8-Me	0.79, 3H, d, (6.7)	13.1
9	1.93, 2.16, 2 x 1H, 2 x m	36.5
10	5.73, 1H, dt, (6.8, 14.3)	133.6
11	6.31, 1H, dd, (11.0, 14.8)	126.57
12	6.01, 1H, ol	130.1
13	5.27, 1H, ol	125.2
14	2.28, 2.38, 2 x 1H, 2 x m	30.1
15	3.21, 1H, m	79.6
15-OMe	3.23, 3H, s	55.7
16	1.36, 1H, m	31.8
17	1.22, 1.36, 2 x 1H, 2 x m	20.6
18	1.24, 1.32, 2H, m, ol	38.2
19	3.64	66.8
20	1.35, 1.47, 2 x 1H, 2 x m	39.3

21	3.82, 1H, m	66.3
22	1.80, 1H, m	42.2
22-Me	0.71 3H, d, (7.0)	10.6
23	5.24, 1H, m	72.2
24	5.00, 1H, d, (9.2)	123.1
25	-	139.1
25-Me	1.69, 3H, s	16.6
26	2.03, 2H, ol	39.06*
27	2.13, 2H, m	30.3
28	5.51, 1H, m	131.3
29	5.96, 1H, m, ol	130.7
30	5.96, 1H, ol	131.7
31	5.57, 1H, m	129.4
32	2.09, 2H, m	41.6
33	3.50, 1H, m	67.5
34	1.04, 1.28, 2 x 1H, 2 x m	43.9
35	1.60, 1H, m, ol	28.4
35-Me	0.80, 3H, d, (6.5)	18.9
36	2.02, 2H, m, ol	32.5
37	1.30, 1.46, 2 x 1H, 2 x m	32.9
38	-	174.5
NH₂	6.65, 7.21, 2 x 1H, 2 x br s	-

Antimicrobial activity

Table S2: MIC values determined for *iso*-amide gladiolin against *C. albicans*, a range of gram-negative and gram-positive bacteria and *M. tuberculosis*.

Test organism	<i>iso</i> -amide gladiolin MIC (µg/mL)
Gram-negative bacteria	
<i>Klebsiella pneumonia</i> DSM26371	>64
<i>Acinetobacter baumannii</i> DSM25645	16
<i>Pseudomonas aeruginosa</i> DSM29239	>64
<i>Enterobacter cloacae</i> DSM16690	>64
<i>Escherichia coli</i> SY327	32
<i>Burkholderia gladioli</i> BCC1622	>64
Gram-positive bacteria	
<i>Enterococcus faecium</i> DSM25390	2
<i>Staphylococcus aureus</i> DSM21979	8
Mycobacterium	
<i>Mycobacterium tuberculosis</i> H37Rv	4
Fungi	
<i>Candida albicans</i> SC 5314	4

Appendix 9: Phylogenetic analysis of AT-like domains from *trans*-AT PKSs

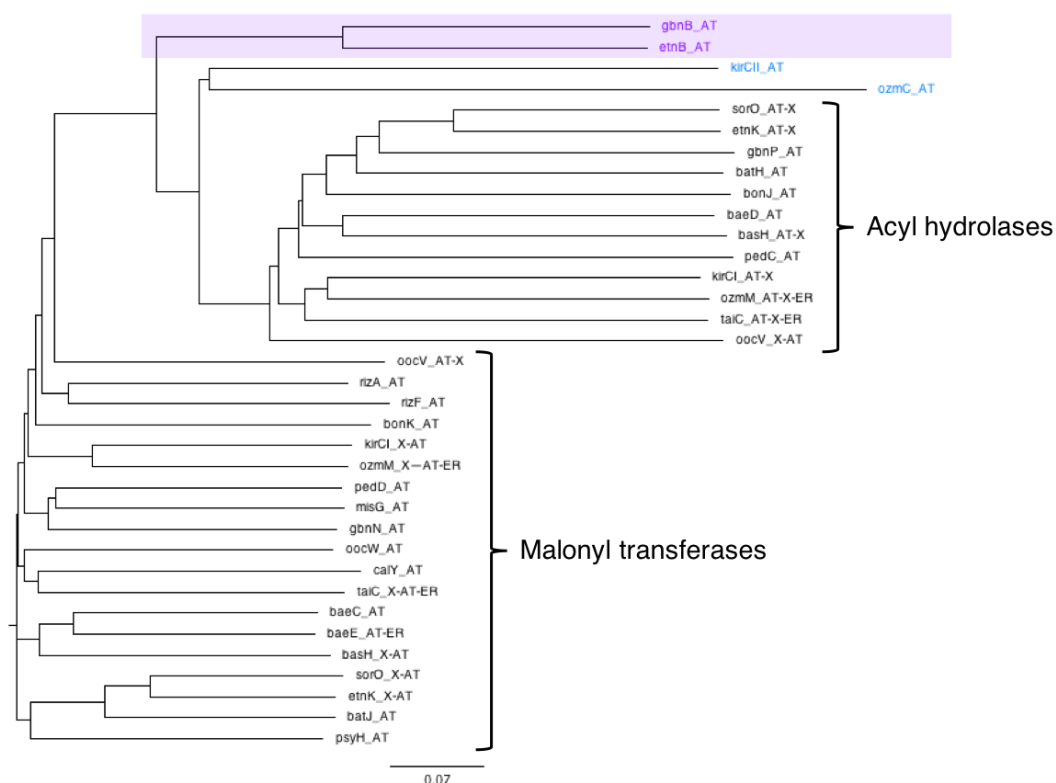
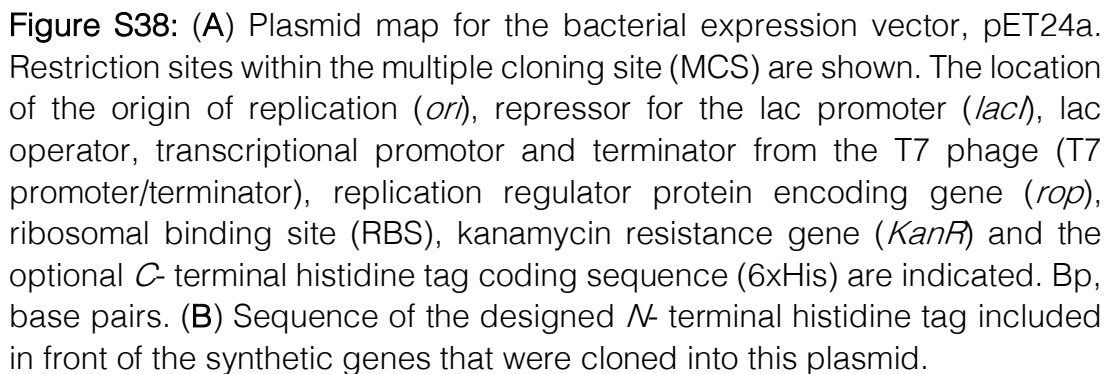


Figure S37: Phylogenetic analysis of AT-like domains from *trans*-AT PKSs. Tip labels consist of the protein name and domain architecture (in the case of proteins with more than one AT domain the appropriate AT domain is shown and the other is denoted with X). Highlighted in purple are the hypothesised succinyl-specific AT domains and tip labels in blue indicate that the AT domain utilizes a non-malonyl substrate. Sequences from the following pathways were used: Gbn, gladiolin (*Burkholderia gladioli*); Etn, etnangien (*Sorangium cellulosum* So ce56); Kir, kirromycin (*Streptomyces collinus*); Ozm, oxazolomycin (*Streptomyces albus*); Sor, sorangicin (*Sorangium cellulosum* So ce12); Bat, batumin (*Pseudomonas fluorescens*); Bon, bongkreikic acid (*Burkholderia gladioli*); Bae, bacillaene (*Bacillus amyloliquefaciens*); Bas, basiliskamides (*Bacillus laterosporus*); Ped, pederin (*Pseudomonas* sp.); Tai, thailandamide (*Burkholderia thailandensis*); Ooc, oocydins (*Serratia marcescens*); Riz, rhizopodin (*Myxococcus stipitatus*); Mis, misakinolide (uncultivated symbiont of the marine sponge *Theonella swinholid*); Cal, calyculin (uncultivated symbiont of the marine sponge *Discodermia calyx*) and Psy, psymberein (uncultivated symbiont of the marine sponge *Psammocinia* aff. *bulbosa*).



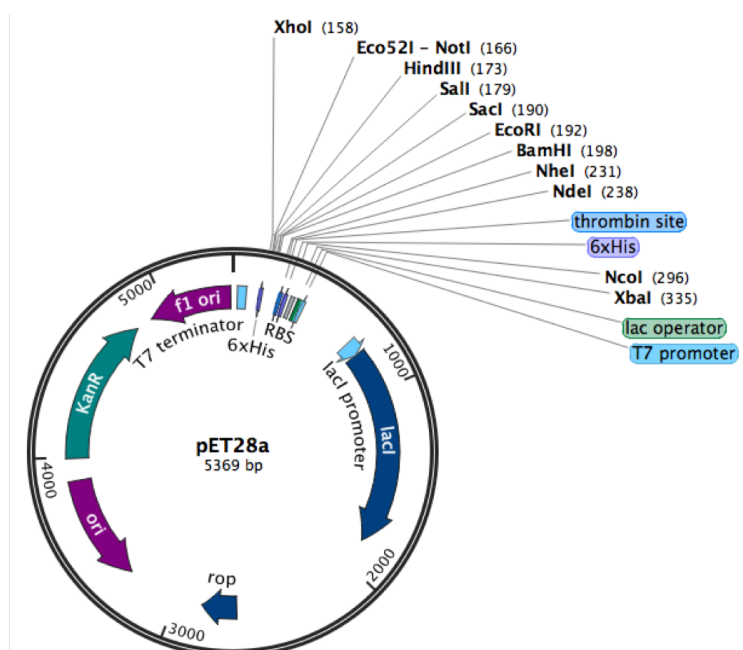


Figure S39: Plasmid map for the bacterial expression vector, pET28a. Restriction sites within the MCS are shown. The location of the origin of replication (*ori*), repressor for the lac promoter (*lacI*), lac operator, transcriptional promoter and terminator from the T7 phage (T7 promoter/terminator), replication regulator protein encoding gene (*rop*), ribosomal binding site (RBS), kanamycin resistance gene (*KanR*) and *N*- (and optional *C*-) terminal histidine tag coding sequence (6xHis) are indicated.

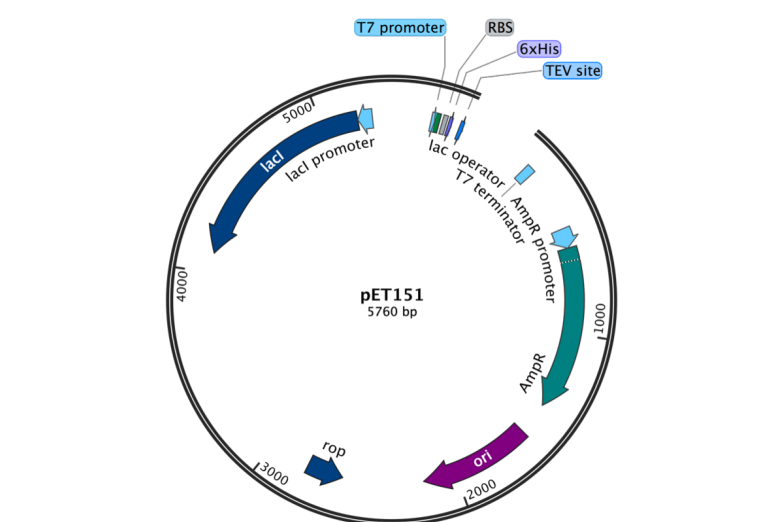


Figure S40: Linearized plasmid map for the bacterial expression vector, pET151. Blunt-end PCR products with a 5' CACC overhang are cloned into the vector at the DNA break shown. The location of the origin of replication (*ori*), repressor for the lac promoter (*lacI*), lac operator, transcriptional promoter and terminator from the T7 phage (T7 promoter/terminator), replication regulator protein encoding gene (*rop*), ribosomal binding site (RBS), ampicillin resistance gene (*AmpR*) and *N*- terminal histidine tag coding sequence (6xHis) are indicated.

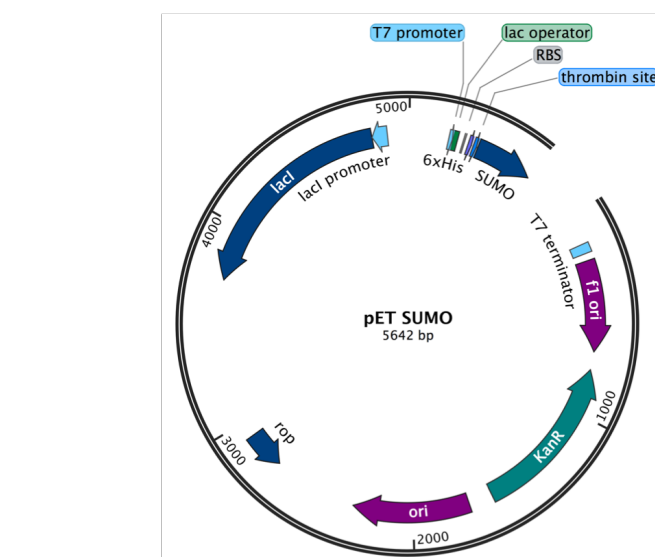


Figure S41: Linearized plasmid map for the bacterial expression vector, pET-SUMO. PCR products with a 3' A overhang (added by *Taq* polymerase) are cloned into the vector at the DNA break shown. The location of the origin of replication (*ori*), repressor for the *lac* promoter (*lacI*), *lac* operator, transcriptional promoter and terminator from the T7 phage (T7 promoter/terminator), replication regulator protein encoding gene (*rop*), ribosomal binding site (RBS), kanamycin resistance gene (*KanR*), SUMO solubility tag (SUMO) and *N*-terminal histidine tag coding sequence (6xHis) are indicated.

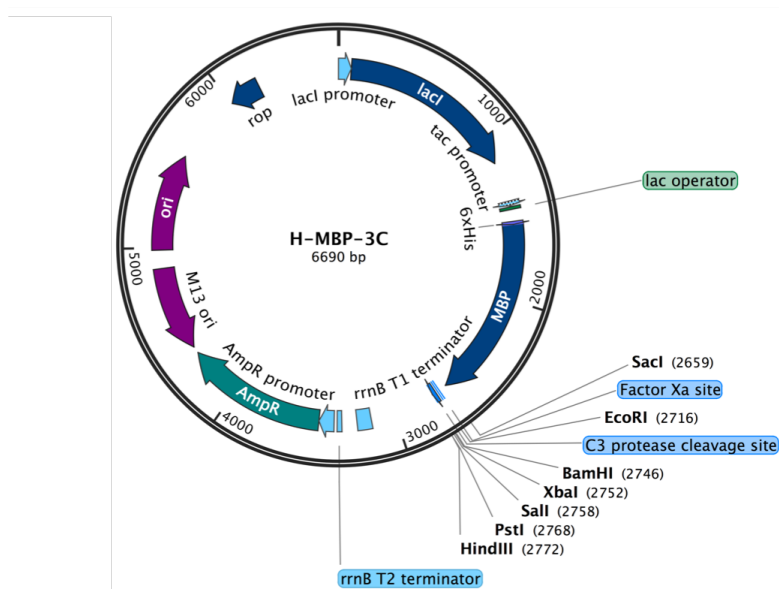


Figure S42: Plasmid map for the bacterial expression vector, H-MBP-3C. Restriction sites within the MCS are shown. The location of the origin of replication (*ori*), repressor for the *lac* promoter (*lacI*), *lac* operator, translational *tac* promoter and *rrnB* terminator, replication regulator protein encoding gene (*rop*), ampicillin resistance gene (*AmpR*), MBP solubility tag (MBP) and *N*-terminal histidine tag coding sequence (6xHis) are indicated.

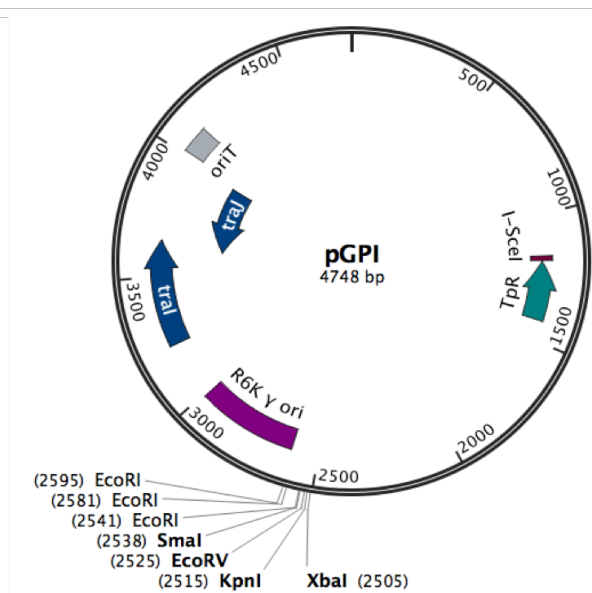


Figure S43: Plasmid map for the mutagenesis vector, pGPI. Restriction sites within the MCS are shown. The location of the defective origin of replication (*R6K γ ori*), origin of transfer (*oriT*), transfer genes (*traJ* and *traI*), trimethoprim resistance gene (*Tpr*) and I-SceI recognition site are indicated.

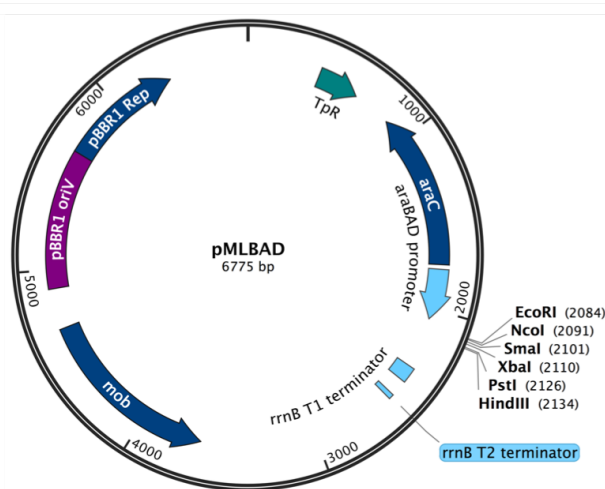


Figure S44: Plasmid map for the bacterial expression vector, pMLBAD. Restriction sites within the MCS are shown. The location of the origin of replication (*pBBR1 oriV*), arabinose inducible promoter (*araBAD promoter*), translational regulatory gene (*araC*) and *rrnB* terminator, replication regulator protein (*pBBR1 Rep*), gene for conjugal transfer of plasmid (*mob*) and trimethoprim resistance gene (*Tpr*) are indicated.