Elucidation of molecular mechanisms of antibiotics biosynthesis in *Burkholderia gladioli*

by

Xinyun Jian

Thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Chemistry

University of Warwick, Department of Chemistry

September 2018

List of contents

List of Figures, Schemes and Tables VI

Acknowledgements XIV

Declaration XV

Abbreviations XVI

Abstract XVIII

Chapter 1 Introduction 1

1.1 *Burkholderia* 1

1.1.1 The diversity and versatility of *Burkholderia* 1

1.1.2 Natural products discovery in *Burkholderia* 3

1.1.3 Discovery of gladiolin from *Burkholderia gladioli* BCC0238 3

1.1.4 Re-discovery of icosalides from *Burkholderia gladioli* BCC0238 5

1.2 Polyketides 6

1.2.1 Polyketide natural products 6

1.2.2 Polyketide biosynthesis 6

1.2.3 Biosynthesis of gladiolin and etnangien 19

1.3 Non-ribosomal peptides 23

1.3.1 Non-ribosomal peptide natural products 23

1.3.2 Non-ribosomal peptide biosynthesis 24

1.3.3 Biosynthesis of icosalides 30

1.4 Study aims 32

Chapter 2 Characterization of *cis*- and *trans*-acting enoyl reductases in the gladiolin polyketide synthase 33

2.1 Characterization of the GbnD1 ER1 domain 35

2.1.1 *In vivo* characterization of the GbnD1 ER1 domain 35

2.1.2 *In vitro* characterization of the GbnD1 ER1 domain 44

2.2 Characterization of GbnE 50

2.2.1 *In vivo* characterization of GbnE 50

2.2.2 *In vitro* characterization of GbnE 57

2.3 Conclusions and future work 65

Chapter 3 Probing the molecular mechanism of chain length control in gladiolin biosynthesis 67

3.1 Iterative nature of GbnD2 Module 5 68

3.1.1 Production of recombinant GbnD2 Module 5 68

3.1.2 Investigating the activity of GbnD2 Module 5 *in vitro* 69

3.1.3 Probing the ACP-selectivity of GbnD2 Module 5 iteration 72

3.2 The effect of GbnE on GbnD2 Module 5 iteration 75

3.2.1 GbnE abolishes iteration of GbnD2 Module 5 75

3.2.2 Protein-protein interactions between GbnE and the GbnD2 ACP5 domain prevent re-acylation of the GbnD2 KS5 domain 77

3.2.3 The 'gate-keeping' role of the GbnD2 KS6 domain 78

3.3 Conclusions and future work 84

Chapter 4 Investigation of a double chain initiation mechanism in icosalides biosynthesis .88

4.1 Characterization of the IcoA C4 domain.....89

4.1.1 Production of recombinant IcoA C4 domain 89

4.1.2 Characterization of the N-acylation activity of the IcoA C4 domain 90

4.2 Characterization of the IcoA C4-A3-PCP3 tri-domain 92

4.2.1 Production of recombinant IcoA C4-A3-PCP3 tri-domain 92

4.2.2 Characterization of the chain initiation activity of the IcoA C4-A3-PCP3 tri-domain.....93

4.3 Characterization of IcoA Module 3.....96

4.4 Trails for *in vitro* reconstitution of icosalides derivatives biosynthesis 99

4.4.1 Strategy 1: two-piece reconstitution.....100

4.4.2 Strategy 2: one-piece reconstitution.....101

4.5 Conclusions and future work.....102

Chapter 5 Conclusions and future perspectives 104

Chapter 6 Materials and methods 108

6.1 Materials 108

6.1.1 Growth Media 108

6.1.2 Buffers 108

6.1.3 Reagents and Kits 109

6.1.4 Bacterial strains.....110

6.1.5 Vectors.....111

6.2 General DNA manipulation 112

6.2.1 Genomic DNA isolation 112

6.2.2 Polymerase chain reaction (PCR) 113

6.2.3 Fragment DNA purification 114

6.2.4 DNA restriction digestion and ligation 114

6.2.5 TOPO® cloning 115

6.2.6 TA® cloning 116

6.2.7 Gibson assembly and GeneArt assembly 117

6.2.8 DNA purification for electroporation 118

6.2.9 Transformation and electroporation 118

6.2.10 Plasmid DNA mini-prep 119

6.3 Genetic manipulation of *Burkholderia* 120

6.3.1 Targeted *in-frame* gene deletion in *Burkholderia* 120

6.3.2 Site-directed mutagenesis in *Burkholderia* 120

6.3.3 Gene complementation in *Burkholderia* 121

6.3.4 General tri-parental mating procedures 123

6.4 *Burkholderia* metabolite extraction and LC-MS analysis 124

6.4.1 Metabolite production and extraction 124

6.4.2 Metabolite LC-MS analysis 124

6.5 Recombinant protein overproduction and purification 125

6.5.1 Cloning of recombinant protein expression constructs: 125

6.5.2 Site-directed mutagenesis 127

6.5.3 Recombinant protein overproduction 128

6.5.4 Recombinant protein purification 129

6.5.5 SDS-PAGE analysis of protein....131

6.6 *In vitro* biochemical assays....132

6.6.1 General PPTase-catalysed phosphopantetheinyl loading ACP assays....132

6.6.2 Characterization of GbnD1 ER1 domain....133

6.6.3 Characterization of GbnE....134

6.6.4 Characterization of GbnD2 Module5....134

6.6.5 Characterization of GbnD2 Module 6 (S941A)....135

6.6.6 Characterization of IcoA C4 domain....135

6.6.7 Characterization of IcoA C4-A3-PCP3 tri-domain....136

6.6.8 Characterization of IcoA Module 3....136

6.6.9 Two-piece *In vitro* reconstitution of icosalides biosynthesis....136

6.7 LC-MS analysis of proteins and *in vitro* intact protein assays....137

6.8 Full length icosalide NRPS encoded gene cloning....137

6.8.1 Capturing plasmid construction....137

6.8.2 Yeast transformation....138

6.8.3 Candidate yeast colonies screening....138

6.8.4 *IcoA* containing plasmid DNA extraction and confirmation....139

6.9 Icosalides full length NRPS heterologous expression in *E. coli*....139

References....141

Appendix....159

List of Figures

Figure 1.1 Beneficial and problematic traits of the *Burkholderia* species.

Figure1.2 Representative examples of natural products with biological activities from *Burkholderia* species.

Figure 1.3 Structural comparison of gladiolin 7 and etnangien 8.

Figure 1.4 Structures and activities of icosalide A1 (9), A2 (10) and B (11).

Figure 1.5 Representative bioactive polyketide natural products.

Figure 1.6 Schematic view of the representative type I modular PKS responsible for erythromycin A biosynthesis.

Figure1.7 Schematic overview of chain initiation, extension and termination by type I modular PKSs.

Figure 1.8 Schematic view of the representative *trans*-AT PKS responsible for leinamycin biosynthesis.

Figure1.9 Examples of module iteration in type I modular PKSs.

Figure1.10 Mechanism of acyl transfer reaction catalyzed by the AT domains from modular PKSs.

Figure 1.11 Mechanism of modular PKS KS domains catalyzed chain elongation reaction.

Figure 1.12 Catalytic and stereo control mechanisms of KR and DH domains from modular PKSs.

Figure 1.13 Catalytic mechanism, stereo control and structural overview of of *cis*-acting ER domains from type I modular PKSs.

Figure 1.14 Phylogenetic, functional and structural overview of *trans*-acting ER domains from type I modular PKSs.

Figure 1.16 Schematic view of PPtase catalyzed *holo* modification of ACP domains (A) and distinct interfaces mediate KS-ACP protein–protein interactions (B).

Figure 1.17 Comparison of the gladiolin and etnangien biosynthetic gene clusters.

Figure1.18 Comparison of the *trans*-AT PKSs responsible for etnangien and gladiolin assembly.

Figure 1.19 Comparison of the proposed biosynthesis in the first module of gladiolin (red box) and etnangien (blue box) PKSs.

Figure 1.20 Representative bioactive non-ribosomal peptides.

Figure 1.21 Schematic view of the representative modular NRPS that responsible for vancomycin biosynthesis.

Figure 1.22 Schematic overview of non-ribosomal peptide chain elongation relations catalyzed by the three core domains.

Figure 1.23 Ribbon diagrams of the NMR solution structures of the TycC3-PCP (from tyrocidine A synthetase) showing three conformers in the A, A/H, and H states.

Figure 1.24 Crystal structure and detailed view of active site of CDA-C1 domain.

Figure 1.25 Classification of the C domains from NRPS based on the phylogenetic and bioinformatic analysis.

Figure 1.26 Identification of the icosalide NRPS biosynthetic gene and phylogenetic affiliation catalytic domains (unpublished data from Challis group).

Figure 1.28 NRPS and proposed biosynthesis of icosalides.

Figure 2.1 Partial biosynthetic pathways of the gladiolin (A) and etnangien (B) PKSs.

Figure 2.2 Comparison of catalytic domain architectures of the gladiolin and etnangien PKSs showing the hypothetic catalytic origin of the additional enoyl reduction event in module 5 of the gladiolin PKS.

Figure 2.3 Work flow of the mutagenesis system applied in *in-frame* deletion in *Burkholderia*.

Figure 2.4 Structure comparison of gladiolin and its analogues.

Figure 2.5 Schematic representation of generation and PCR confirmation of *B. gladioli* BCC1622_Δ*gbnD1*_ER1.

Figure 2.6 LC-MS analysis of metabolite extracts from BCC1622 wild type and BCC1622_Δ*gbnD1*_ER1.

Figure 2.7 Structural comparison of gladiolin and proposed derivatives arising from inactivation of GbnD1 ER1 *in vivo*.

Figure 2.8 Proposed structure of unreduced gladiolin derivatives.

Figure 2.9 Structure of NADPH binding site in SpnER2 domain from spinosyn PKS.

Figure 2.10 Partial result of multiple sequence alignment between GbnD1 ER1 domain and reported ER domains from type I modular PKSs.

Figure 2.11 Schematic representation of generation and PCR confirmation of *B. gladioli* BCC1622_ *gbnD1*_ER1 (G388S/G389P)

Figure 2.12 LC-MS analysis of metabolite extract from BCC1622 wild type and BCC1622_*gbnD1*_ER1*(GG388-399SP).

Figure 2.13 SDS-PAGE and LC-MS analysis of purified N-His_6-GbnD1 ACP1-ER1, N-His_6-GbnD1 ER1, N-His_6-GbnD2 ACP5.

Figure 2.14 LC-MS analysis of *in vitro* GbnD1 ER1 domain *intra*-module enoyl reduction activity.

Figure 2.15 LC-MS analysis of *in vitro* GbnD1 ER1 domain *inter*-modular enoyl reduction activity in module 5.

Figure 2.16 LC-MS analysis of *in vitro* assay demonstrating enoyl reduction activity of the isolated GbnD1 ER1 domain.

Figure 2.17 Schematic representation of generation and PCR confirmation of *B. gladioli* BCC1622_Δ*gbnE*.

Figure 2.18 LC-MS analysis of metabolite extracts from *B. gladioli* BCC1622 wild type and BCC1622_Δ*gbnE.*

Figure 2.19 Structural comparison of gladiolin and proposed derivatives produced by *gbnE* deletion mutant.

Figure 2.20 PCR confirmation of BCC1622_Δ*gbnE*::*gbnE.*

Figure 2.21 LC-MS analysis of gladiolin production in *B. gladioli* BCC1622 wild type, BCC1622_Δ*gbnE*, BCC1622_Δ*gbnE*::*gbnE* and BCC1622_Δ*gbnE*::pMLBAD.

Figure 2.22 Schematic representation of generation and PCR confirmation of *B. gladioli* BCC1622_Δ*gdsB*_ER and BCC1622_Δ*gbnE*Δ*gdsB*_ER.

Figure 2.23 LC-MS analysis of gladiolin production in *B. gladioli* BCC1622 wild type, BCC1622_Δ*gbnE*, BCC1622_Δ*GdsB*_ER and BCC1622_Δ*gbnE*Δ*GdsB*_ER.

Figure 2.24 SDS-PAGE and LC-MS analysis of purified N-His_6-SUMO-GbnE.

Figure 2.25 LC-MS analysis of in vitro enoyl reduction activity of GbnE.

Figure 2.26 SDS-PAGE and LC-MS analysis of N-His_6-GbnD1 ACP3, N-His_6-GbnD4 ACP10 and N-His_6-GbnD5 ACP12.

Figure 2.27 LC-MS analysis of 2,4-hexadienoyl-PPant loading onto the GbnD1 ACP3 (A), GbnD2 ACP5 (B), GbnD4 ACP10 (C) and GbnD5 ACP12 domain (D).

Figure 2.28 LC-MS analysis of GbnE *in vitro* enoyl reduction with GbnD2 ACP5, GbnD4 ACP10, GbnD1 ACP3, and GbnD5 ACP12 domain.

Figure 2.29 Structure of active site in FabK from *Streptococcus pneumoniae*.

Figure 2.30 Partial result of multiple sequences alignment between GbnE and homologous enzymes.

Figure 2.31 LC-MS analysis of purified N-His_6-SUMO-GbnE, N-His_6-SUMO-GbnE(H198V).

Figure 2.32 LC-MS analysis of *in vitro* enoyl reduction activity of GbnE (H198V).

Figure 3.1 Comparison of partial biosynthetic pathways of gladiolin (A) and etnangien (B).

Figure 3.2 SDS-PAGE and LC-MS analysis of purified N-His_6-GbnD2 Module 5.

Figure 3.3 LC-MS analysis of *in vitro* reconstitution of iterative activity of GbnD2 Module 5.

Figure 3.4 SDS-PAGE analysis and LC-MS analysis of N-His_6-tag cleaved GbnD1 ACP3 and GbnD4 ACP10 domain.

Figure 3.5 LC-MS analysis of *in vitro* iteration control assays of GbnD2 Module 5.

Figure 3.6 LC-MS analysis of *in vitro* GbnD2 Module 5 iteration assays with and without addition of GbnE.

Figure 3.7 LC-MS analysis of *in vitro* GbnD2 Module 5 iteration assays with and without addition of GbnE (H198V).

Figure 3.8 SDS-PAGE and LC-MS analysis of purified N-His_6-GbnD2 Module 6(S941A).

Figure 3.9 LC-MS analysis of *in vitro* PPtase catalyzed 2,4-hexadienoyl-PPant and 4-hexenoyl-PPant loading of GbnD2 ACP5 domain.

Figure 3.10 LC-MS analysis of *in vitro* chain translocation between GbnD2 ACP5 domain and GbnD2 KS6 domain.

Figure 3.11 LC-MS analysis of acylation of the KS domain in GbnD2 Module 6(S941A).

Figure 3.12 LC-MS analysis of GbnD2 Module6 (S941A) *in vitro* chain elongation.

Figure 4.1 SDS-PAGE and LC-MS analysis of purified N-His_6-IcoA PCP3 and N-His_6-IcoA C4.

Figure 4.2 LC-MS analysis of the N-acylation activity of the IcoA C4 domain.

Figure 4.3 SDS-PAGE and LC-MS analysis of purified N-His_6-IcoA C4-A3-PCP3 tri-domain and N-His_6-IcoA C4*-A3-PCP3(H143A) tri-domain.

Figure 4.4 LC-MS analysis of the L-serine adenylation activity of A domain in IcoA C4-A3-PCP3 and IcoA C4*-A3-PCP3 (H143A).

Figure 4.5 LC-MS analysis of *in vitro* reconstitution of chain initiating activity of IcoA C4-A3-PCP3.

Figure 4.6 SDS-PAGE and LC-MS analysis of purified N-His_6-IcoA Module 3.

Figure 4.7 LC-MS analysis following *apo* to *holo* modification and L-Serine loading of IcoA module

3 and proposed identity of the unknown derivative of IcoA module 3.

Figure 4.8 Schematic view of *in vitro* reconstituted biosynthesis of novel icosalides derivatives.

Figure 4.9 *In vitro* reconstitution of IcoA by a two-part strategy.

Figure 4.10 Restriction digestion map of the pET28a(+)-*icoA*

Figure 5.1 Future perspectives of different chain length control mechanisms employed by the gladiolin and etnangien assembly lines.

Figure 5.2 Schematic view of rational engineering of novel icosalides derivatives via a combinatorial biosynthesis strategy.

List of schemes

Scheme 2.1 Design of *in vitro* assay for investigating the GbnD1 ER1 domain's *intra*-modular enoyl reduction activity.

Scheme 2.2 Design of *In vitro* assay for investigating putative *inter*-module enoyl reduction activity of GbnD1 ER1 domain in module 5.

Scheme 2.3 Design of *in vitro* assay for monitoring activity of isolated GbnD1 ER1 domain.

Scheme 2.4 Design of *In vitro* assay for examination of enoyl reduction activity of GbnE.

Scheme 3.1 Design of *in vitro* assay (A) and proposed mechanism (B) for examining the iterative nature of GbnD2 Module 5.

Scheme 3.2 Design of *in vitro* assays for monitoring GbnD2 Module 5 iteration with GbnE presence.

Scheme 3.3 Design of *in vitro* assays for monitoring substrate specificity of GbnD2 KS6 domain in the translocation process.

Scheme 3.4 Design of *in vitro* assays for monitoring substrate specificity of GbnD2 KS6 domain in chain elongation process.

Scheme 4.1 Proposed biosynthetic events in module 3 of IcoA

Scheme 4.2 Design of *in vitro* assays for testing the N-acylation activity of the IcoA C4 domain.

Scheme 4.3 Design of *in vitro* assay for examination of the chain initiating activity of IcoA C4-A3-PCP3.

Scheme 4.4. Optimized *in vitro* one-pot assay for reconstitution of the chain initiating activity of IcoA C4-A3-PCP3.

Scheme 4.6 Design of on-pot *in vitro* assay for reconstitution of the activity of IcoA Module 3.

List of tables

Table 6.1 List of antibiotics

Table 6.2 List of bacterial strains

Table 6.3 List vectors

Table 6.4 Components of typical PCR reactions

Table 6.5 Thermocycling conditions for a routine PCR

Table 6.6 Components of DNA restriction digestion reactions

Table 6.8 Components of TOPO® cloning reactions

Table 6.9 Components of typical TA® cloning reactions

Table 6.10 Components of typical Gibson assembly reactions

Table 6.11 Components of typical GeneArt assembly reactions

Table 6.12 Overview of constructs used in in-frame deletion, site-directed mutagenesis and complementation in Burkholderia

Table 6.13 UHPLC elution profile metabolite LC-MS analysis

Table 6.14 Overview of primers and cloning methods for constructs used for recombinant protein expression

Table 6.15 Components of Q5® Site-Directed Mutagenesis reactions

Table 6.16 Overview of the primers and mutations for mutant constructs

Table 6.17 Overview of tag type, size and purification method for recombinant protein

Table 6.18 Recipe of Tris-glycine SDS-polyacrylamide gels

Table 6.19 UHPLC elution profile for protein and intact protein LC-MS analysis

Table 6.20 Overview of primers for construct pCAP1000-icoA

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor, Prof. Gregory Challis for giving me the opportunity to work on such an interesting and varied project, and for providing such a great multidisciplinary platform for our studies. His sound guidance, deep insights, kind support and encouragement are highly appreciated.

I would like to specially thank Dr. Matthew Jenner, for his enormous help to many aspects throughout my PhD. These include teaching me biochemistry and protein mass spectrometry techniques and providing valuable discussions and comments to my experiments, reports, presentations and thesis. I would also like to thank Dr. Joleen Masschelein for teaching me *Burkholderia* genetic manipulation and offering useful suggestions to my project in the first two years of my PhD. I am grateful to Dr. Lona Alkahalaf, for her patiently reading through and correcting my thesis, as well as helpful advice to my project. I am also appreciative to Dr. Douglas Roberts and Christian Hobson, for synthesizing the substrates used in this study and for sometimes explaining chemistry questions to me.

My thanks also go to my advisory panel: Dr. Christophe corre and Dr. Claudia Blindauer, Dr. Lijiang Song, Dr. Simone Kosol, Dr. Yousef Dashti, Dr. Chris Fage, Dr. Daniel Zabala, Dr. Emzo De Los Santos, Dr. Shanshan Zhou, for their advice and help during my PhD. Furthermore, I would like to thank everyone in the Challis group for creating a positive and lovely working environment. Thanks to the CSC-Warwick joint scholarship for funding.

Finally, I would like to thank my parents and family for their Love, support and encouragements. I want to give my special thanks to Chuan, who has always been there for me. I couldn't have done it without you.

Declaration

I confirm this thesis has been prepared in accordance with the university's guidelines on the presentation of a research thesis for the degree of Doctor of Philosophy. The experimental work reported in this thesis is original research carried out by myself, unless otherwise stated. No material has been submitted in any application for any other degree.

Results from other authors are referenced in the usual manner throughout the thesis.

_____ Date:_____

Xinyun Jian

Abbreviations

A-Domain	Adenylation Domain
ACP	Acyl Carrier Protein
ADP	Adenosine Diphosphate
AL	Acyl Ligase
Ala	Alanine
AMP	Adenosine Monophosphate
Arg	Arginine
Asn	Asparagine
AT	Acyltransferase
ATP	Adenosine Triphosphate
Bcc	*Burkholderia cepacia* complex
C-Domain	Condensation Domain
CoA	Coenzyme A
Cys	Cystine
DMSO.	Dimethylsulphoxide
DNA	Deoxyribonucleic Acid
DPCK	Dephosphocoenzyme A Kinase
E-domain	Epimerization Domain
ECH	Enoyl-CoA Dehydratase
EDTA	Ethylenediaminetetraacetic Acid
ER	Enoyl Reductase
ESI	Electrospray Ionisation
FAS	Fatty Acid Synthase
FMN	Flavin Mononucleotide
Gln	Glutamine
Glu	Glutamic Acid Glycine
Gly	Glycine
GNAT	GCN5-related N-acetyltransferase
His	Histidine
HMGS	3-hydroxy-3-methylglutaryl-CoA
IPTG	Isopropylthio-β-galactoside
KR	Ketoreductase
KS	Ketosynthase
LB	Luria Broth

LC	Liquid Chromatography
Leu	Leucine
Lys	Lysine
Met	Methionine
MS/MS	Tandem Mass Spectrometry
MS	Mass Spectrometry
MT	Methyl Transferase/ Mutant
MIC	Minimum Inhibitory Concentration
NADH	Nicotinamide Adenine Dinucleotide
NADPH	Nicotinamide Adenine Nucleotide Phosphate
NMR	Nuclear Magnetic Resonance
NRPS	Non-Ribosomal Peptide Synthase
PanK	Pantethenate Kinase
Phe	Phenylalanine
PKS	Polyketide Synthase
Pant	(Phospho)pantetheine
PPant	Phosphopantetheinyl
PPAT	Phosphopantetheine Adenylyltransferase
PPTase	4'-Phosphopantetheinyl Transferase
Pro	Proline
PUFA	Poly Unsaturated Fatty Acid
Ser	Serine
NAC	N-Acetyl Cysteine
TAR	Transformation-Associated Recombination
TE	Thioesterase
TFA	Trifluoroacetic Acid
Thr	Threonine
TOF	Time of Flight
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine
UHPLC	Ultra-high Performance Liquid Chromatography
WT	Wild Type

Abstract

Burkholderia is a multi-talented genus of Gram-negative bacteria that has been recently shown to be a promising, untapped source of antibiotics with the potential to overcome antimicrobial resistance. An antimicrobial activity screen of a clinical isolate, *B. gladioli* BCC0238, identified a novel polyketide macrolide antibiotic, gladiolin, that is structurally related to a known antibiotic etnangien and exhibits potent activity against *Mycobacterium tuberculosis.* Gladiolin shows significantly increased stability towards light and air compared to etnangien, due to the absence of a highly unstable hexaene moiety present in the side chain of etnangien, which is the key structural difference between the two metabolites. Comparison of the polyketide synthases (PKSs) responsible for the biosynthesis of gladiolin and etnangien, however, reveals a strikingly similar domain architecture. This thesis reports the elucidation of the catalytic origins for the main structural differences between the two metabolites, revealing a *trans*-acting enoyl reduction event involved in shutting down a programmed iteration in gladiolin biosynthesis, which is proposed for the polyene formation in etnangien biosynthesis.

In addition to gladiolin, a set of lipopeptodiolides, known as icosalides, which were originally reported as fungal metabolites, were also discovered from *B. gladioli* BCC0238. The non-ribosomal peptide synthase (NRPS) responsible for the biosynthesis of icosalides exhibits an unprecedented domain architecture revealing two directly adjacent condensation (C) domains embedded in the middle of the NRPS. In this study, the enzymology of these C domains has been investigated *in vitro*, elucidating an unusual double chain initiation mechanism for asymmetric peptidolide biosynthesis. Additionally, efforts towards *in vitro* reconstitution of the entire icosalide NRPS is reported, which may allow access to novel antibiotic analogues.

Chapter 1

Introduction

1.1 *Burkholderia*

1.1.1 The diversity and versatility of *Burkholderia*

Burkholderia is a genus of Gram-negative β-proteobacteria. Species belonging to this genus which occupies remarkably diverse ecological niches, ranging from environmental water, soil, the rhizosphere of plants to animals and humans.[1,2,3] The first described species from this genus was identified by William Burkholder in 1950 as a plant pathogen that caused sour rot of onion and named as *Pseudomonas cepacian* due to the phenotypic similarity to *Pseudomonas* species.[4] The *Burkholderia* species were included in the *Pseudomonas* genus till 1992 when the new genus *Burkholderia* was generated.[5] This genus underwent a rapid expansion subsequently and now composes over 80 identified species, which are classified into two major clades.[6] A particular important group of species is the *Burkholderia cepacia* complex (Bcc) located in clade I. Bcc is a collection of genetically distinct but phenotypically similar *Burkholderia* species, which have attracted considerable attention from researchers due to their challenging taxonomy, particular importance in clinical epidemiology as well as exceptional metabolic versatility among *Burkholderia* species.[7]

Burkholderia species are now known as one of the most versatile groups of Gram-negative bacteria that are capable of forming a variety of complex interactions with hosts.[1] These interactions form the basis of both the problematic and beneficial traits of *Burkholderia* species as shown in Figure 1.1. Apart from some species identified as plants pathogens, for example *B. plantarii* causes seedling blight of rice[8] and *B. andropogonis* causes stripe disease of sorghum[9],

several species are also capable of infecting animals and humans. The primary pathogens include *B.mallei* and *B. pseudomallei* which cause glanders mainly in horses and melioidosis in both animals and humans respectively.[10,11] Species like *B. gladioli*, *B. fungorum* and Bcc are opportunistic pathogens that are considered as serious threats to vulnerable individuals, especially cystic fibrosis (CF) patients.[12,13,14]

In contrast to the pathogenic traits that led to the original identification, some *Burkholderia* species have been demonstrated to develop beneficial interactions with plants leading to a general growth promotion of crops. These species are able to fix atmospheric nitrogen, such as *B. vietnamiensis* and *B. tropica*,[15,16] or produce bioactive metabolite to protect seedling plants from attack by pathogenetic fungi and nematodes or insects.[17,18,19] Additionally, several species have also showed great biotechnological potential as bioremediation agents, due to their ability to degrade polyaromatic hydrocarbon pollutants, such as *B.cepacia*[20,21].

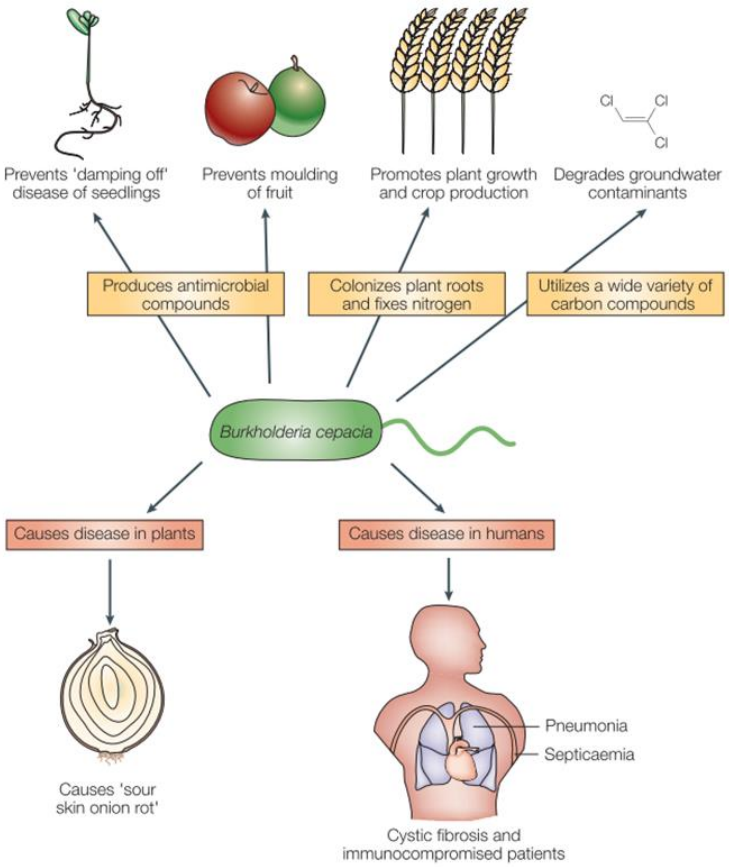


Figure 1.1 Beneficial and problematic traits of the *Burkholderia* species (reproduced from ref.[22]). The pathogenicity to plants, animals and humans, particularly evident for cystic fibrosis (CF) patients, represent the problematic traits of *Burkholderia* species. The ability to biodegrade pollutants and promote plant growth via

nitrogen fixing or producing antimicrobial metabolites form the beneficial basis of this genus.

1.1.2 Natural products discovery in *Burkholderia*

Burkholderia species exhibit a remarkable ability to produce a variety of specialized metabolites with diverse bioactivities (Figure 1.2), the majority of which are antimicrobial metabolites, such as antibiotic cepacins **1**[23] and enacyloxins IIa **5**[24] and antifungal agent pyrrolnitrin **6**[25] and occidiofungin **2**[26] .

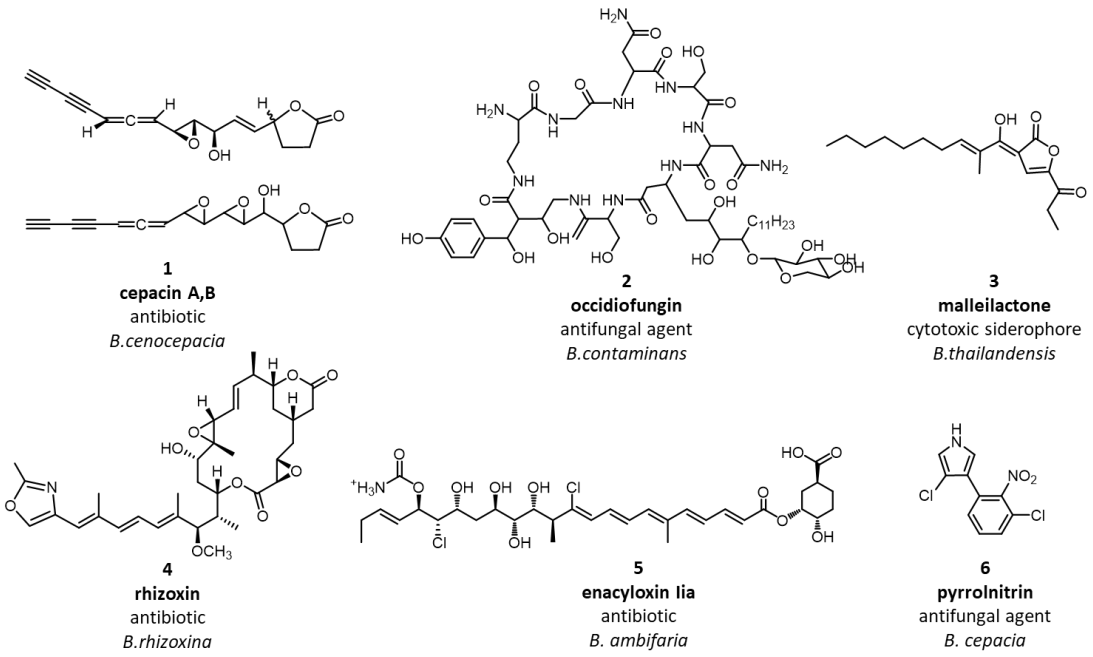
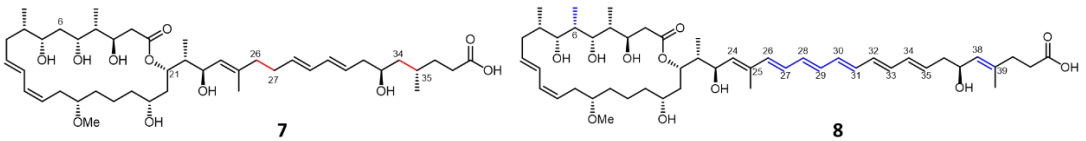


Figure1.2 Representative examples of bioactive natural products produced by *Burkholderia* species.

In recent years, the availability of vast genome sequence data from *Burkholderia* species has uncovered their surprisingly huge genetic capacity with the genome size ranging between 6 and 9 Mb, which underpins their extraordinary specialized metabolic potential.[27] Aided by the advances in genomic-driven natural product discovery technologies, numerous natural product-linked as well as cryptic biosynthetic gene clusters have been identified from *Burkholderia* species, proving them as an untapped and promising source of novel natural products.

1.1.3 Discovery of gladiolin from *Burkholderia gladioli* BCC0238

During previous screens of *Burkholderia* clinical isolates for antimicrobial metabolite production in the Mahenthiralingam group, *B. gladioli* BCC0238, isolated in 1996 from the sputum of a cystic fibrosis patient, was found to exhibit a broad spectrum of antimicrobial activity.[28] HPLC purification coupled with structural elucidation of the active metabolite from *B. gladioli* BCC0238 resulted in the identification of a novel polyketide macrolide, named gladiolin **7** (Figure 1.3) by the Challis group.[28] Gladiolin is structurally similar to the known polyketide antibiotic etnangien **8** (Figure 1.3), produced by *Sorangium cellulosum*.[29,30] Structural comparison of gladiolin and etnangien reveals three differences: 1) an extra methyl group at C6 in etnangien, 2) the C38-C39 double bound in etnangien is saturated at the corresponding C34-C35 position in gladiolin and 3) the C26-C31 triene moiety in etnangien is replaced by saturated a C26-C27 bond in gladiolin. The conjugated hexaene moiety in the etnangien side chain significantly destabilizes the molecule, making it prone to light-induced isomerization and rapid oxidative degradation.[30] Gladiolin lacking this moiety and thus has significantly increased stability towards light and air. However, the lactone of gladiolin is prone to rearrangement/hydrolysis (detailed in section 2.1.1.1).



7
gladiolin
Chemical Formula: $C_{44}H_{74}O_{11}$
Exact Mass: 778.5231

8
etnangien
Chemical Formula: $C_{49}H_{76}O_{11}$
Exact Mass: 840.5388

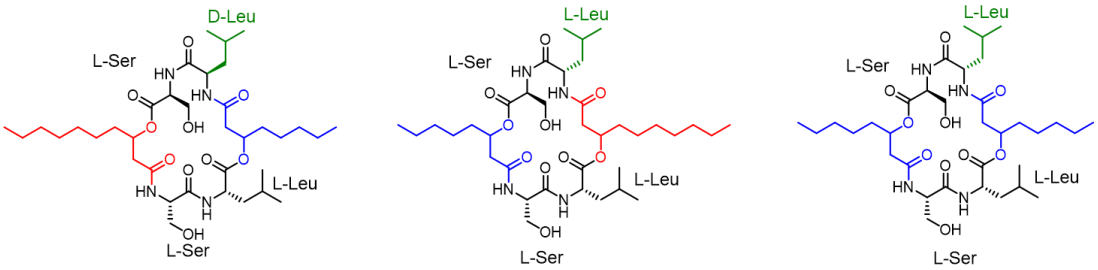
Figure 1.3 Structural comparison of gladiolin 7 and etnangien 8. Structural differences are highlighted in red (gladiolin) and blue (etnangien).

Etnangien was reported to be active against a broad panel of Gram-positive bacteria, including *Mycobacterium smegmatis* (MIC = 1 µg/ml), a non-pathogenic species, which is phylogenetically close to a known tuberculosis causing pathogenic species, *Mycobacterium tuberculosis*. As an analog of etnangien, gladiolin was also tested against mycobacteria, revealing it possesses potent activity against *M. tuberculosis* (MIC = 0.4 µg/ml against the H37Rv strain).[28] Both compounds have been shown to be capable of inhibiting the mycobacterial RNA polymerase, which is a validated drug target of *M. tuberculosis,* while having low mammalian cytotoxicity.[28-31] With the significantly increased stability towards light and air compared with etnangien, potent activity

including against several isoniazid and isoniazid/rifampicin-resistant clinical isolates of *M. tuberculosis*,[28] gladiolin offers more promise for development of novel drug to confront multidrug-resistant tuberculosis.

1.1.4 Re-discovery of icosalides from *Burkholderia gladioli* BCC0238

In the process of screening for other bioactive metabolites from *B. gladioli* BCC0238 in the Challis Group, a set of lipopeptodiolides, known as icosalides A1 (**9**), A2 (**10**) and B (**11**) (Figure 1.4) were identified from the extracts of this strain. Interestingly, icosalides were originally isolated from the extract of a fungi species[32].



9
icosalide A1
Chemical Formula: $C_{36}H_{64}N_4O_{10}$
Exact Mass: 712.4622

10
icosalide A2
Chemical Formula: $C_{36}H_{64}N_4O_{10}$
Exact Mass: 712.4622

11
icosalide B
Chemical Formula: $C_{36}H_{60}N_4O_{10}$
Exact Mass: 684.4309

Antimicrobial Agent

Cytotoxic to Proliferating Cells

Figure 1.4 Structures and activities of icosalide A1 (9), A2 (10) and B (11). The amino acid building blocks are denoted, with the different configurations of the second leucine residue highlighted in green. The incorporated 3-hydroxyacyl units are highlighted in red and blue.

Structurally, all three icosalide analogs contain two serine and two leucine amino acid residues, with two fatty acid-derived 3-hydroxyacyl units incorporated as part of the central twenty-member ring. A distinguishing feature of icosalide A1 is the presence of a single D-Leucine residue, whereas A2 and B universally contain L-configured amino acid residues. Besides, icosalides A1 and A2 integrate eight and ten carbon 3-hydroxyacyl chains into the cyclic core generating an asymmetric peptolide, whereas icosalide B obtains structural symmetry via incorporation of two eight carbon 3-hydroxyacyl units. With these structural differences, the three Icosalide analogs display very different bioactivities: icosalide A1 shows modest antimicrobial activity against

Streptococcus pyogenes, *S. pneumoniae*, and *Enterococcus faecalis* while icosalides A2 and B are cytotoxic to replicating Madin-Darby canine kidney (MDCK) cells.

1.2 Polyketides

1.2.1 Polyketide natural products

Polyketides (PKs) are a large family of natural products that possess impressive structural complexity and diversity, as well as a broad range of biological activities, such as antibiotic erythromycin A **12**, antifungal agent amphotericin B **13**, anti-cancer agent epothilone B **16**, cholesterol-lowering agent lovastatin **14**, immunosuppressant rapamycin **15** and pesticide spinosyn A **17** (Figure 1.5). Their pharmaceutical and economic value have motivated interest in novel structure discovery, elucidation of the enzymology and structure of biosynthetic machineries as well as genetic and chemical engineering of novel derivatives for the last three decades.

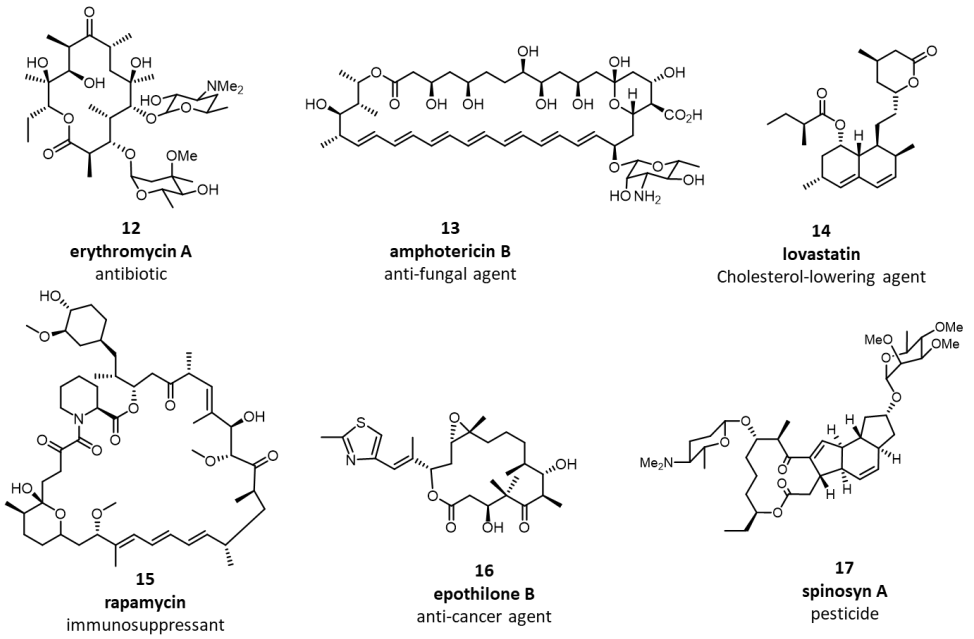


Figure 1.5 Representative bioactive polyketide natural products.

1.2.2 Polyketide biosynthesis

Despite polyketide natural products having remarkable structural diversity, they are all assembled from simple acyl-coenzyme A (CoA) derived building blocks by a group of enzymes called polyketide synthases (PKSs).[33] Three distinct types of PKSs exist in various organisms, including bacteria, fungi and plants, and are responsible for biosynthesis of different classes of polyketides, including aromatic and complex polyketides.[34] This study focuses on the bacterial polyketides complex, the majority of which are assembled by type I modular PKSs.

1.2.1.1 Type I modular polyketide synthases

Type I modular PKSs are large multifunctional enzyme complexes, in which the catalytic domains are organized into multiple sequential modules, whereby a module is responsible for one catalytic cycle of the building block selection and installation.[35] The best studied type I modular PKSs is the 6-deoxyerythromycin B synthase (DEBS) involved in erythromycin biosynthesis, which cotains six chain extension modules flanked by a loading and module and a chain releasing thioesterase (TE) domain into three gigantic multienzymes (DEBS 1, DEBS 2 and DEBS 3) as shown in Figure 1.6.

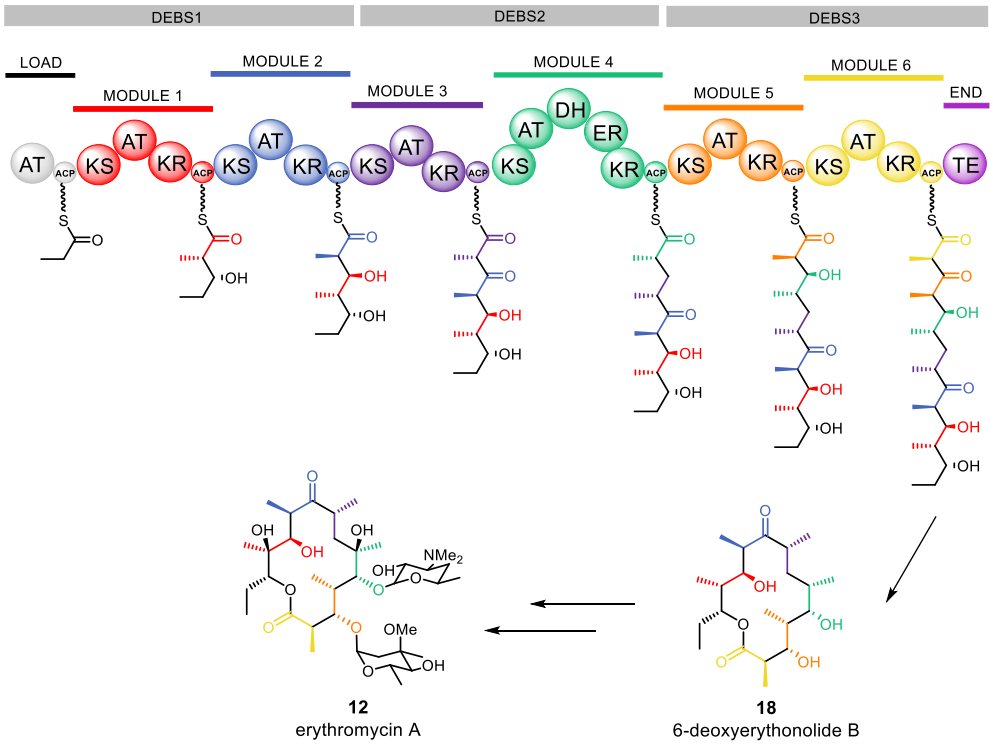


Figure 1.6 Schematic view of the type I modular PKS responsible for erythromycin A biosynthesis. The

correspondence between the structures of the assembled intermediates and final structure id highlighted by the color coding the domain organization of the PKS.

1.2.1.2 Biosynthetic overview of type I modular PKSs

Type I modular PKSs employ the biosynthetic logic resembling that of mammalian fatty acid synthases (FASs). In general, both biosynthetic machineries incorporate the simple acyl-CoA building blocks in an assembly line fashion through three steps: chain initiation, chain extension and chain termination (Figure 1.7).

Chain assembly is normally initiated by an acyltransferase (AT) domain in the loading module selecting and loading a starter unit from the corresponding CoA thioester (commonly acetyl-CoA) onto the phosphopantetheinyl (PPant) moiety of an acyl carrier protein (ACP) domain. This PPant moiety is attached to the ACP domain by a phosphopantetheinyl transferase (PPase) prior to polyketide assembly. An alternative chain initiating mechanism involves an AT domain loading a malonyl or methylmalonyl unit, followed by a decarboxylation catalyzed by a ketosynthase domain (KS) variant KS^{Q} (discussed in section 1.2.2.5 'Ketosynthase domains').

In each chain elongation step, a KS domain catalyses a decarboxylative Claisen condensation between the starter unit or the growing polyketide chain it accepts from the upstream ACP domain and the extender unit (commonly methymalonyl- or manlonyl-CoA) loaded on the downstream ACP domain by the adjacent AT domain. The minimal chain elongation module thus consists of KS, AT and ACP domains which extend the polyketide backbone by one ketide unit. The resulting β-keto-acyl thioester can undergo optional modification by ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains, leading successively to β-hydroxyl, α,β-unsaturated and fully saturated intermediates respectively. The polyketide chain is passed along the assembly line between ACP domains and KS domains and built up through repeated chain extension.

The chain assembly is normally terminated by a TE domain located at the end of the PKS releasing the product by hydrolysis or more commonly macrocyclization. Further post-PKS modifications, such as glycosylation, methylation and hydroxylation are sometimes employed to introduce

additional structural complexity and diversity.

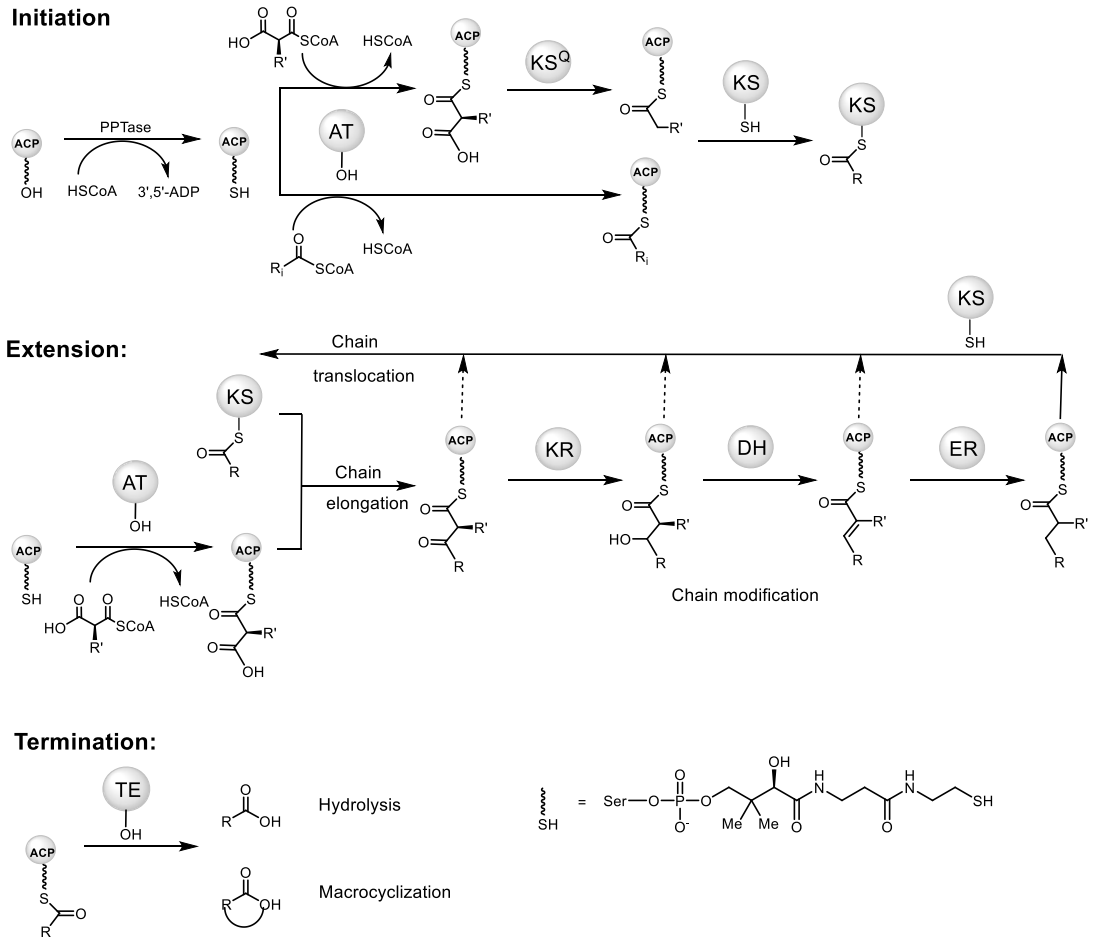


Figure 1.7 Schematic overview of chain initiation, extension and termination by type I modular PKSs (adapted from ref.[36]).

1.2.1.3 Subdivision of type I modular PKSs

Given the above biosynthetic logic employed by type I modular PKSs, the structures of their polyketide products are often found to be directly correlated to their domain organizations (Figure 1.6). This correlation is referred to as 'colinearity' and has been applied to effective prediction of final product structure from the PKS architecture.[37]

In recent years, however, another class of modular system has emerged, in which the colinearity is normally not conformed to. In these particular systems, the most obvious feature is that integrated AT domains are absent from the assembly lines. Instead, the AT activity required by every module is supplied by one or more *trans*-acting AT domains.[38] They are thus termed as

trans-AT PKSs, with DEBS representing modular systems termed as *cis*-AT PKSs.[38] Examples of *trans*-AT PKSs include those responsible for the biosynthesis of pederin,[39] bacillaene,[40] viginamycin[41] and leinamycin **19**[42] (Figure1.8).

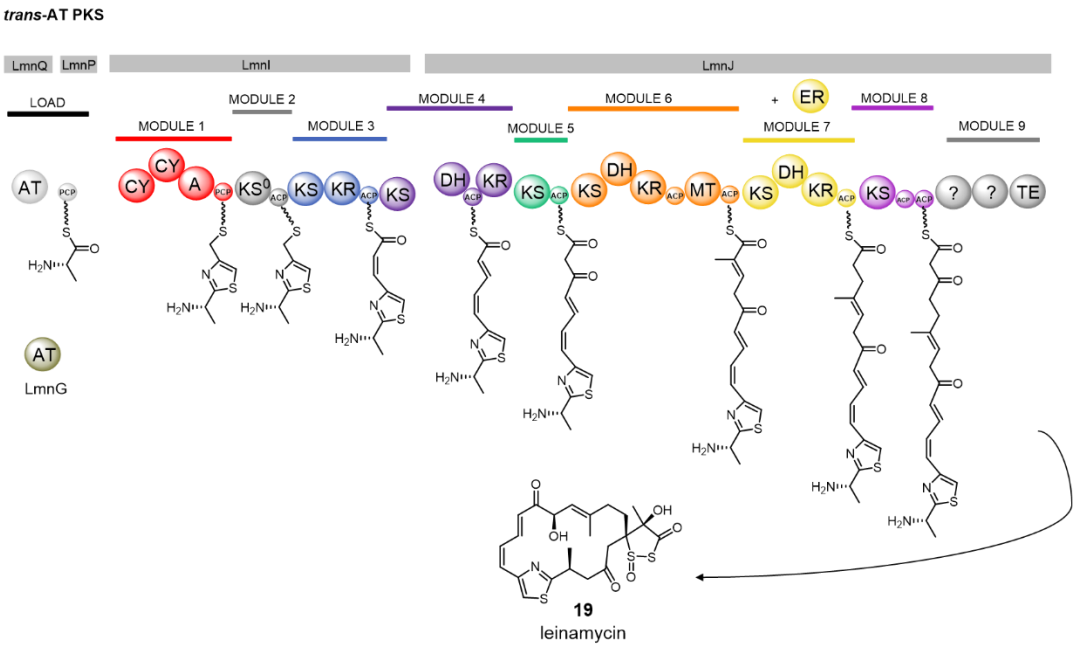


Figure 1.8 Schematic view of the *trans*-AT PKS responsible for leinamycin biosynthesis.

In addition to *trans*-acting AT domains as a distinguishing feature, *trans*-AT PKSs exhibit several other non-canonical architectural and enzymatic characteristics. These include unusual domain architectures, modules split into across subunits, inactive domains or and *trans*-acting enzymes other than AT domains as shown in Figure 1.8. Phylogenetic studies have revealed that the *cis*- and *trans*-AT PKSs have distinct evolutionary origins, with the former believed to evolve through module duplication and domain diversification while the latter has evolved via horizontal gene transfer.[43,44]

1.2.1.4 Iterative module use in type I modular PKSs

In type I modular PKSs, a molecular ratchet mechanism is typically employed to govern the unidirectionally of the assembly line.[45] However, iterative usage of individual modules has been observed as either an aberrant or a programmed event in several modular PKS systems (Figure 1.9). The discovery of a ring expanded 6-dEB analogue **20** as a minor product of DEBSs resulting

from iteration in module 4 represents the first example of module 'stuttering' (Figure 1.9 A). A similar phenomenon was found during epothilone biosynthesis, which iteratively use module 5 and module 6 to generate 6 different ring expanded analogue.[46] On the other hand, programmed module iterations are demonstrated to be required for the biosynthesis of stigmatellin, aureothin,[47] (each two iterations) borrellidin,[48] and etnangien[31] (each three iteration) (Figure 1.9 B).

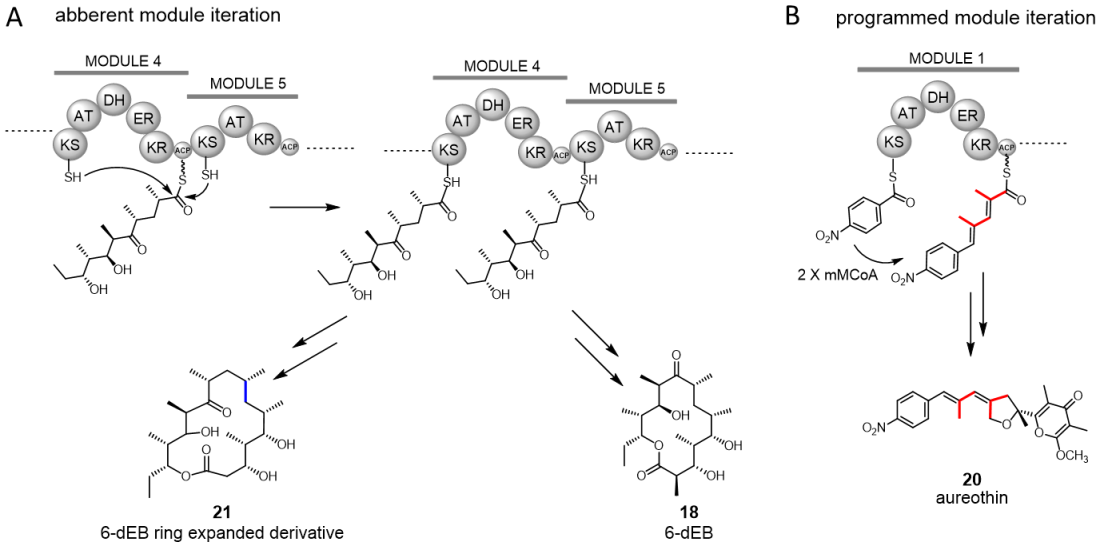


Figure1.9 Examples of module iteration in type I modular PKSs. (A) Module 'stuttering' occurs in module 4 of the 6-dEB PKS leading to a minor ring-expanded product is an example of aberrant module iteration. **(B)** Iterative usage of module 1 of the aureothin PKS is required for the biosynthesis of the final product, representing the programmed iteration.

Current understanding of programmed module iteration in modular PKSs is mainly based on investigations of aureothin biosynthesis. Since modular PKSs exist as homodimers, it has been proposed that the iteration occurs by retrotransfer of the intermediate from the ACP domain onto the KS domain located on the opposite polypeptide strand[49]. In addition evidence was provided to show that iteration is controlled by several factors, including a 'gate-keeper' function and downstream KS domains and the kinetics of chain acylation of and elongation by the KS domains.[50]

1.2.1.5 Catalytic domains of type I modular PKSs

Acyltransferase domains

The AT domains catalyze the transfer of the acyl unit from CoA to the downstream ACP domain via a 'ping-pong bi-bi' mechanism which involves two half reactions: acylation of the active site serine residue of the AT domain followed by transfer to the ACP domain[51,52] (Figure 1.10).

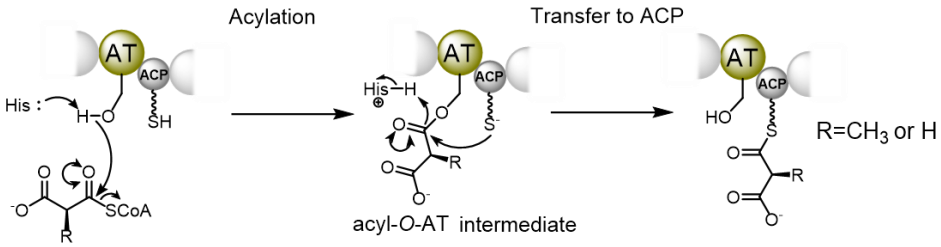


Figure1.10 Mechanism of acyl transfer reaction catalyzed by the AT domains from modular PKSs. The AT domains catalyse starter unit (e.g. malonyl- or 2S-methylmalonyl-CoA) loading via an acyl-*O*-A T intermediate

The AT domains from modular PKSs play a crucial role in conserving the fidelity of the assembly line through their specificity towards the substrates. The extender AT domains mainly select primary metabolism originated malonyl- and (*2S*)-methylmalonyl-CoA,[53] with *trans*-acting AT domains normally specific for malonyl-CoA.[38] However, some examples of extender AT domains incorporating unusual acyl building blocks, such as hydroxyl- and aminomalonyl-CoA have been observed, which contribute to the structural diversity of polyketides.[53] In contrast, the AT domains associated with the loading module utilize a variety of acyl-CoAs as starter units, such as acetyl-, propionyl-, isobutyryl-, crotonyl- and isopentyl-CoA.[54,55] Structural and bioinformatic studies on several AT domains have allowed the identification of the conserved motifs that correlates the with substrate specificity.[36]

Ketosynthase domains

KS domains catalyse carbon-carbon bond formation through two half reactions, including transthiolation and decarboxylative Claisen condensation. Reported structures of modular PKS KS domains showed they form homodimers with each monomer adopting a thiolase-like fold.[56,57] The substrate binding channel is split into a well-conserved PPant arm binding region and a more variable thioester substrate binding site. Bioinformatics analysis combined with structural studies

revealed a conserved Cys-His-His catalytic triad is harbored that all chain elongating KS domains contain.[56,57,58] Firstly, the transthiolation proceeds via nucleophilic attack of the thiol group in the active site Cys residue on the carbonyl group of the starter unit or polyketide thioester tethered to the upstream ACP domain, forming an acyl-*S*-KS complex (Figure1.11). Following release of the upstream ACP domain, the KS domain catalyses decarboxylation of the extender unit delivered by the PPant arm of the downstream ACP domain. This enables the resulting enolate to attack the carbonyl group of the acyl-*S*-KS intermediate, thereby generating a two-carbon elongated β-keto-acyl-ACP thioester product. One of the histidine residues within the catalytic Cys-His-His triad stabilizes the enolate intermediate while the other histidine residue is hypothesized to promote decarboxylation via activation of a water molecule to towards nucleophilic attack on the carbonyl group extender unit.[59,60]

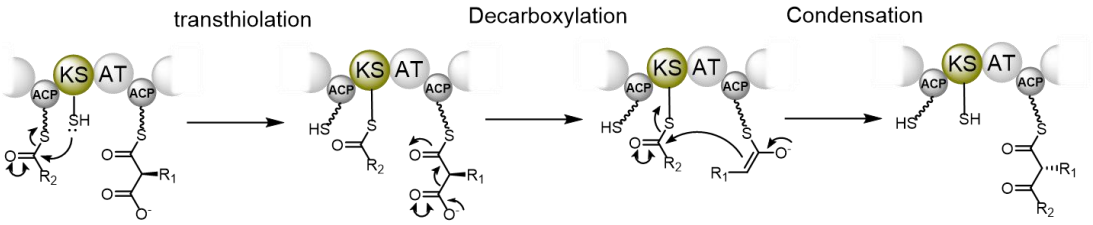


Figure 1.11 Mechanism of KS domain catalyzed chain elongation reaction in modular PKSs. The KS domain catalyzed chain elongation reaction involves a transthiolation and a decarboxylative condensation steps.

Biochemical studies of KS domains from *cis*-AT PKSs revealed they exhibit varying levels of substrate specificity[61] while KS domains from *trans*-AT PKSs are more selective towards the structure of the functional group at the β-position, especially in the chain elongation step.[62,63] The substrate specificity prediction of *trans*-AT PKS KS domains has been facilitated by phylogenetic analysis, which demonstrated that evolutionarily close *trans*-AT KS domains usually have similar substrate specificity.[44] These insights have greatly facilitated the biosynthetic assignment of *trans*-AT PKSs which are normally difficult using colinearity-based predictions.

In addition to chain elongating KS domains, other two variants of KS domains exist in modular PKSs. The first is located in the loading modules of *cis*-AT PKSs and catalyses the decarboxylation of a starter unit (malonyl or methyl-malonyl unit) due to a catalytic cysteine to glutamine mutation, and are thus termed KS^{Q} domain.[64] The other ones are found as a characteristic feature

of *trans*-AT PKSs. These lack an active site histidine residue which is essential for chain elongation, and are thus termed non-elongating KS (KS^{o}) domains.[38] KS^{o} domains have been postulated to pass the growing polyketide chain to the next module but the reason this is needed is unclear.

Ketoreductase and Dehydratase domains

Ketoreductase domains in type I modular PKSs belong to the short-chain dehydrogenase/reductase (SDR) family and catalyse stereospecific reduction of β-keto-acyl-ACP thioester intermediates to the corresponding β-hydroxy-acyl-ACP thioester using NAD(P)H as a cosubstrate.[65] They consist of a structural and a catalytic subdomain with a conserved Tyr-Ser-Lys triad in the active site participating in catalysis. [66,67] The reaction proceeds via the transfer of 4-*pro-S* hydride from NAD(P)H to the β-carbon of the substrate, followed by the protonation of the resulting alkoxide by the active site tyrosine residue. Based on the stereochemistry of the resulting β-hydroxy group, KR domains are classified into A-type (for *L*-configuration) and B-type (for *D*-configuration) (Figure 1.12 A) KR domains.[68] The stereochemistry of the product is thought to be determined by the binding orientation of the substrate relative to the NADPH, which is influenced by the relative position of amino acid residues in the active site amino acids.[68]

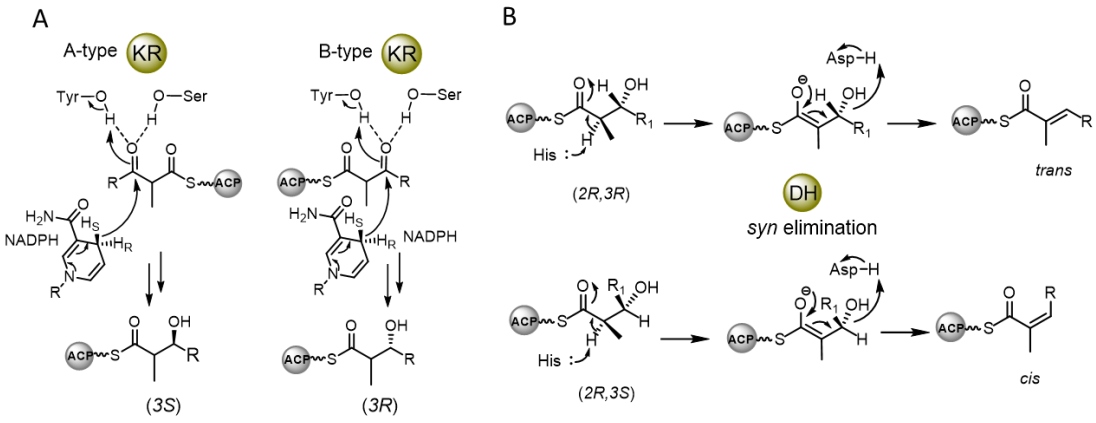


Figure 1.12 Catalytic and stereo control mechanisms of KR and DH domains from modular PKSs. (A) KR domains catalyzed ketoreduction proceeds via 4-pro-*S* hydride transfer from NADPH to the β-carbonyl carbon followed by α-position protonation by a catalytic Tyr residue. A type and B type KR domains lead to different stereo outcome at C3, determined by the binding orientation of the substrate relative to the NADPH. **(B)** A *Syn* elimination mechanism that involves the catalytic His-Asp diad is applied by modular PKS DH domains to achieve stereo

control with *3R*-hydroxy groups resulting in *trans* double bonds while *3S*-hydroxy groups resulting in *cis*.

Dehydratase domains from modular PKSs catalyse elimination of water from the β-hydroxy-acyl-ACP substrate and typically result in formation of a *trans* (*E*)-α, β-double bond, however, *cis*(*Z*)-α, β-double bond has been observed[69]. Structurally, they are organized as dimers with each monomer forming a double hot dog (DHD) fold and bearing a conserved His-Asp catalytic dyad at the active site.[70,71,72] The dehydration reaction has been proposed to proceed via deprotonation at the α-position by the His residue and elimination of the β-hydroxy group, promoted by protonation by the Asp residue (Figure 1.12 B).[70,71] Biochemical and evolutionary evidence have suggested most double bonds formation by modular PKS DH domains share a common *syn* elimination mechanism.[73,74,75] The geometry of the double bond is hypothesized to be mainly determined by the chirality of the β-hydroxy group, with *3R*-hydroxyl group generated by A-type ketoreduction leading to *trans* double bond, and *3S*-hydroxyl group generated by B-type ketoreduction giving *cis*-double bound.[76]

Cis-acting enoyl reductase domains

The modular PKS integrated ER domains catalyse *cis*-acting NADPH-dependent reduction of an enoyl-ACP to an α, β-saturated acyl-ACP.[77] Current understanding of the mechanism involves two processes: the transfer of the 4-*pro*-*R* hydride of NADPH to the C3 position of the enoyl acyl-ACP, followed by a stereospecific protonation of the C2 position by either a general acid or solvent (Figure 1.13 A). Multiple sequence alignment of *cis*-acting ER domains combined with mutagenesis studies revealed that a conserved tyrosine residue confers the *S* configured stereo specificity at the C2 position protonation process, whereas a valine (occasionally alanine or phenylalanine) at the corresponding position led to *2R* configured products (Figure 1.13 A).[78,79]

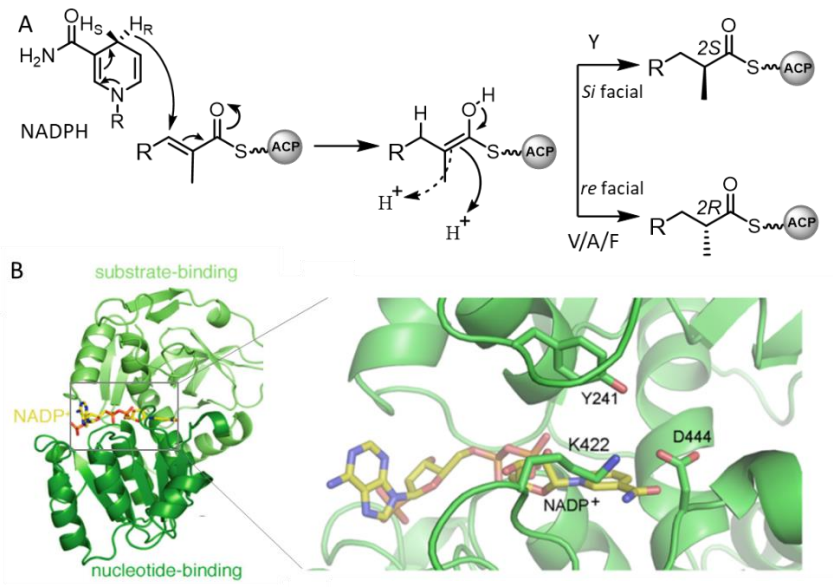


Figure 1.13 Catalytic mechanism, stereo control and structural overview of of *cis*-acting ER domains from type I modular PKSs. (A) *cis*-acting ER domains catalyse enoyl reduction via transfer of the 4-pro-*R* hydride from NADPH to the C3 position of the enoyl-ACP thioester, followed by a stereospecific protonation at C2, resulting in *2S*- (when conserved Tyr present) or *2R*-configurated products (when the conserved Tyr is replaced by Val/Ala/Phe). **(B)** Crystal structure of the SpnER2 domain shows the NADPH binding site located in a cleft between the two subdomains and the conserved catalytically important residues (K422, Y241 and D444) located close to the nicotinamide moiety (Figure adapted from ref.[80]).

Like their counterparts in mammalian FASs, *cis*-acting ER domains belong to the Medium-chain Dehydrogenases/Reductases (MDR) family that typically comprise of a C-terminal NADPH binding subdomain with a Rossman fold and a N-terminal substrate binding subdomain.[81] The only reported crystal structure of an ER domain from a modular PKS is SpnER2 (solved as a KR-ER didomain) from the spinosyn PKS. It showed, unlike other homodimeric members of the MDR family, that the SpnER2 domain exists as a monomer with the NADPH binding site located in a cleft between the two subunits (Figure 1.13 B).[80] Within its active site, three conserved residues Y241, K422, and D444 were observed with in ~6 Å of the nicotinamide 4-pro-R hydrogen atom, indicating they may be involved in substrate protonation.[80]However, both previous *in vivo* mutagenesis study on RapER13 domain (in rapamycin PKS)[82] and the structural guided *in vitro* mutagenesis study on SpnER2[79] concluded that no single residue is essential for catalytic activity.

***Trans*-acting enoyl reductase domains**

In addition to the embeded *cis*-acting ER domains, another phylogenetically distinct type of ER domain in *trans*-AT PKSs is the stand-alone *trans*-acting ERs.[38] In some cases these are fused to the N-terminus of AT or AH domains. *Trans*-acting ERs clade with stand-alone ER domains (PfaD) employed by polyunsaturated fatty acid synthases and close to non-canonical bacterial FAS ER (FabK), fungal FAS ER (fFAS ER) and *Corynebacteria*, *Mycobacteria*, and *Nocardia* (CMN)-FAS ERs (Figure 1.14 A).[83] While *cis*-acting ER domains directly reducing substrate tethered on the adjacent ACP domain using NADPH as cosubstrate, this type of ER performs enoyl reduction in a *trans*-acting manner using FMN to mediate hydride transferring from NAD(P)H to substrate (Figure 1.14.B). The *trans*-acting enoyl reduction activity of a PksE ER domain involved in dihydrobacillaene biosynthesis was biochemically characterized, showed it engaged in a specific interaction with its cognate ACP domain.[84] Other homologues are found in several *trans*-AT polyketide biosynthetic clusters, such as kalimantacin/batumin,[85] leinamycin[86] and myxovirescin[87].

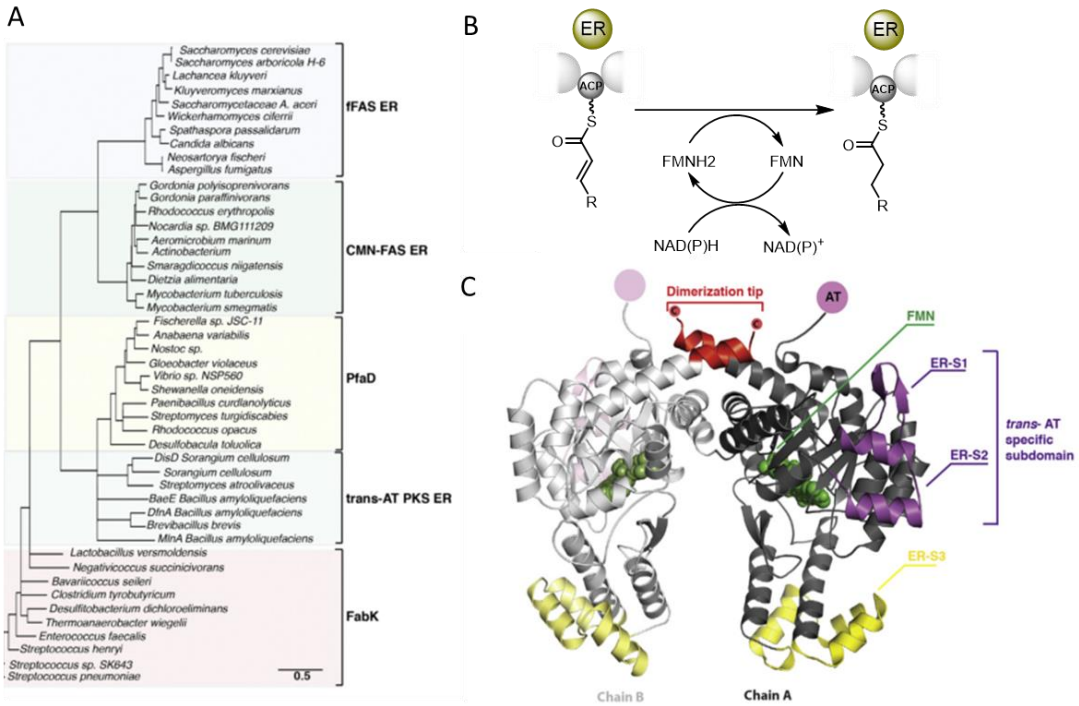


Figure 1.14 Phylogenetic, functional and structural overview of *trans*-acting ER domains from type I modular PKSs. (A) Phylogenetic tree and distribution of modular PKSs *trans*-acting ER related proteins among bacteria and

fungi showing *trans*-acting ERs form *trans*-AT PKS are phylogenetically close to PfaD, FabK, CMN-FAS ERs and fungal FAS ERs (Figure reproduced from ref.[83]) . **(B)** *Trans*-acting ERs catalyse enoyl reduction using FMN to mediate the transfer of hydride from NAD(P)H to the substrate. **(C)** Crystal structure of DifA ER domain shows it forms a homodimer with FMN (highlighted in green) bound to the TIM barrel subdomain and contains additional loops and helix insertions compared to FabK (highlighted in purple and yellow) which are proposed to serve as interdomain interaction interfaces (Figure reproduced from ref.[83]). The dimerization tip is highlighted in red.

The first structural insights into *trans*-acting ERs was provided by the crystal structure of the DifA ER domain (excised from an AT-ER didomain), which is involved in the biosynthesis of difficidin.[83] The DifA ER domain exists as a homodimer with each monomer consisting of a triosephosphate-isomerase (TIM) barrel domain with FMN bound and an inserted α-helical substrate-binding domain (Figure 1.14 C).[83] Though highly structurally homologous to FabK, the DifA ER domain contains additional loops and helix insertions (highlighted in purple and yellow), which were proposed to serve as interdomain interaction interfaces.[83] However, the detailed catalytic mechanism of *trans*-acing ERs and the molecular basis underlying their specific interactions with cognate ACP domains remain unclear.

Thioesterase domains

TE domains fused to the C-termini of modular PKSs catalyze polyketide chain release via a two-step reactions.[88] The highly conserved serine residue located within the active site initiates the reaction by nucleophilic attack on the C=O bond of the assembled polyketide chain tethered via a thioester linkage to the adjacent ACP domain, resulting in formation of an acyl-*O*-TE intermediate. This intermediate subsequently undergoes hydrolysis from the TE domain or a regio-specific macrocyclization via attack of internal nucleophilic.[89,90] In addition to the TE domains fused to the C-termini of modular PKSs (termed type I TE domains), standalone type II TE domains are also usually associated with modular PKSs.[91,92] This type of TE domain catalyses cleavage of aberrant acyl chains from ACP domains to ensure the assembly line remain catalytically active.[93]

Acyl carrier protein domains and their interactions with other domains

Acyl carrier protein domains are small (~80-100 aa) non-catalytic domains involved in all of the catalytic steps of polyketide chain assembly. Structurally, they consist of four-helix bundles, stabilized by inter-helical hydrophobic interactions.[94,36] Inactive *apo*-ACP domains are converted to their functional *holo* forms by a dedicated PPTase, which catalyses post-translational attachment of a PPant moiety of CoA to a conserved the serine residue located in the DSL motif (Figure 1.16 A).[95,96]

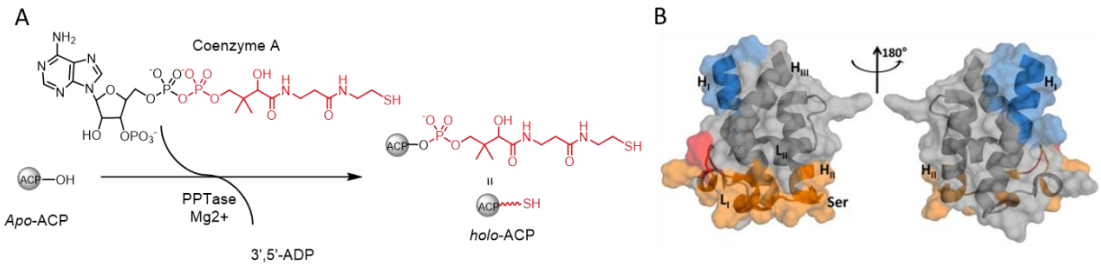


Figure 1.16 Schematic view of PPtase catalyzed post-translational modification of ACP domains (A) and the distinct interfaces that mediate KS-ACP interactions (B). Chain elongation (orange and red) and chain transfer (blue) epitopes lie on entirely different faces of the DEBS ACP2 domain (figure reproduced from ref.[97])

The PPant moiety of ACP domains serves as a ~20 Å long and flexible arm that allows the shuttling of covalently-bound building blocks or assembled intermediates to and between different catalytic domains.[57] ACP domains also engage in the specific protein-protein interactions with partner domains through distinct interface. The conserved DSL motif in helix II has been demonstrated be a universal recognition site for interacting with other catalytic domains.[98] Recent work has showed the docking sites for intramodular KS_n-ACP_n interaction (in chain elongation step) and intermodular ACP_n-KS_{n+1} interaction (in chain translocating step) are located at distinct regions of ACP domains, with the former located in Helix I and the latter located in Loop I (Figure 1.16 B).[97,45] These specific KS-ACP protein-protein interactions play an important role in conserving the unidirectionally of the assembly line.[36]

1.2.3 Biosynthesis of gladiolin and etnangien

Complete genome sequencing of *B. gladioli* BCC0238 and *S. cellulosum* So ce56, in combination with targeted gene inactivation experiments, have facilitated the identification of the

biosynthetic gene clusters for gladiolin and etnangien.[28,99] Both clusters harbor six large *trans*-AT PKS-encoding genes, associated with a conserved set of genes encoding *trans*-acting polyketide processing enzymes and export related proteins (Figures 1.17).[28,31]

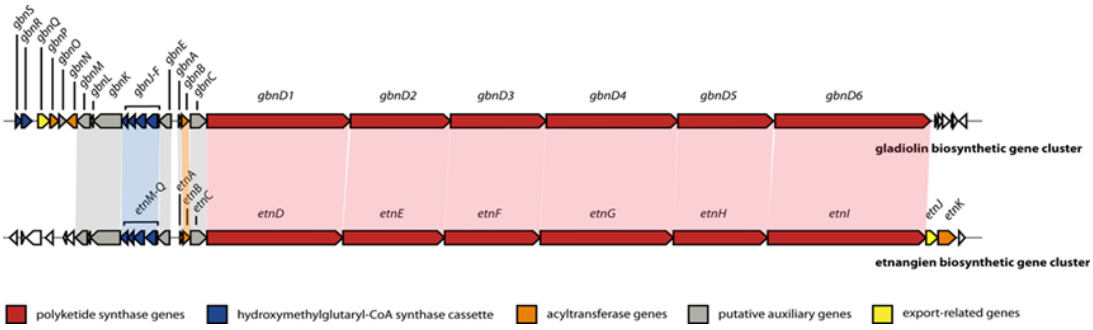


Figure 1.17 Comparison of the gladiolin and etnangien biosynthetic gene clusters.[28] The proposed functions of genes are indicated in color corresponding to the annotation at the bottom and the homology between the genes from the two clusters is indicated by the colored shading.

BLAST searches revealed that these two *trans*-AT PKSs share remarkably similar domain organizations, with the only difference identified as the substitution of an ACP domain in module 1 of the etnangien PKS with an ER (GbnD1 ER1) domain in the gladiolin PKS (Figure1.18). The proposed biosynthetic pathways for gladiolin and etnangien show that the main structural differences between the two compounds are incorporated by the first five modules of their respective assembly lines.[28,31] Both are proposed to utilize a succinyl-CoA starter unit, which subsequently undergoes chain elongation with a malonyl unit catalyzed by the KS domain in the first module (Figure 1.19). A β-branch is then incorporated. This is catalyzed by a cassette of *trans*-acting enzymes ('HCS cassette'), including a 3-hydroxymethylglutaryl synthase (HMGS)-like enzyme, standalone KS^{Q} and ACP domains, and two enoyl-CoA hydratase (ECH)-like enzymes, and results in a β-methyl, α, β-unsaturated intermediate (Figure 1.19). The α, β-double bond is predicted to be saturated by the ER (GbnD1 ER1) domain in module 1 of the gladiolin PKS. The GbnD1 ER1 domain is substituted with an ACP domain in the etnangien PKS. Thus, this α, β-double bond remains in the final product, accounting for one of the structural differences between the two compounds (Figure 1.19).

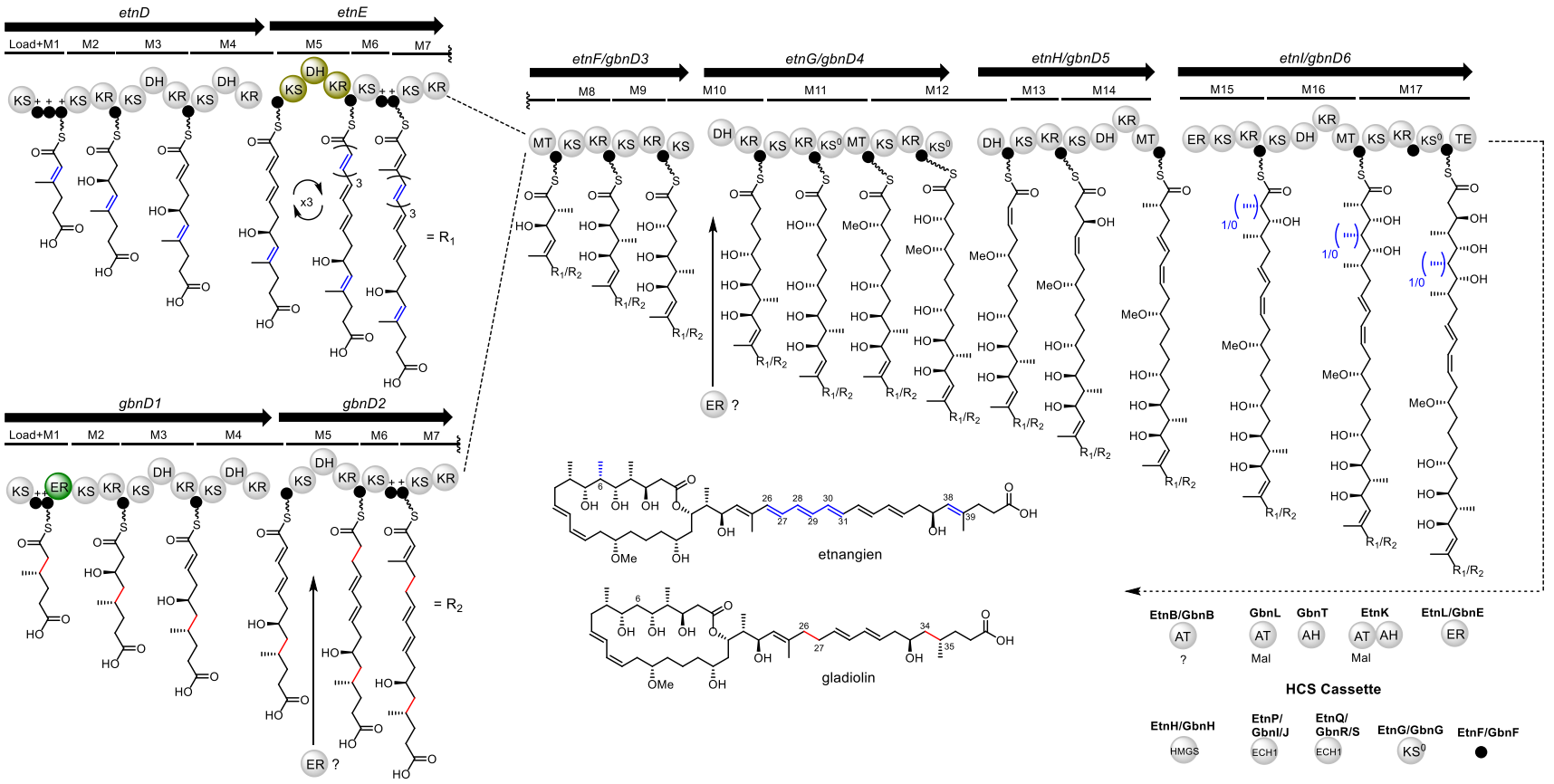


Figure1.18 Comparison of the *trans*-AT PKSs responsible for etnangien and gladiolin assembly[28,31,100]**.** Domain and module organization of the gladiolin and etnangien PKS, showing the proposed structure of chain-elongation intermediates attached to the ACP domain in each module. Structural differences between intermediates and final products are shown in red (gladiolin) and blue (etnangien). The GbnD1 ER1 domain is highlighted in green as the sole difference between the two PKSs. The predicted iterative EntE module 5 is highlighted in gold. The configurations of the stereocenters are predicted based on sequence analysis of the KR domains. Acyl transferase (AT)/hydrolase (AH) domains acting *in trans* with their predicted specificities are indicated. Putative *trans*-acting ERs, non-elongating KSs and HMGS enzymes cassette are indicated. 'HCS cassette' interacting ACPs are labeled with '+'.

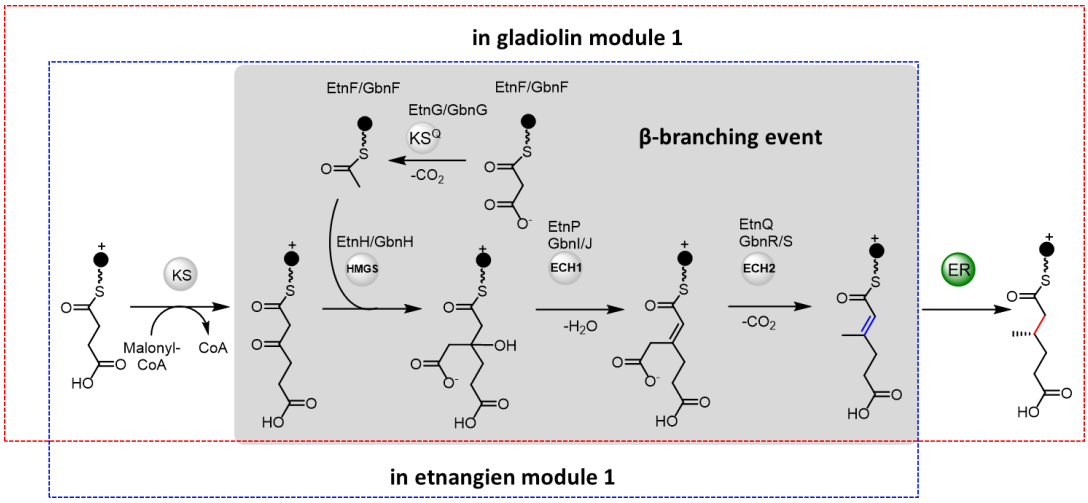


Figure 1.19 Comparison of the proposed biosynthesis in the first module of gladiolin (red box) and etnangien (blue box) PKSs. β-branching event is highlighted in gray box. ACP domains labeled with '+' denotes the integrated ACP domains, while without denotes the ACP domain belongs to the *trans*-acting 'HCS cassette' enzymes.

The resulting intermediate from module 1 of each assembly line is then transferred to module 2 and subsequently module 3 and 4 for identical chain elongation and modification reactions. According to the key structural differences between the two metabolites, module 5 of the etnangien PKS (EntE Module 5) is predicted to be iteratively used for three times while the corresponding iteration does not occur at this biosynthetic stage of gladiolin chain assembly. In addition, an enoyl reduction is proposed to take place in module 5 of the gladiolin PKS (GbnD2 Module 5) which does not occur in EntE Module 5 (Figure. 1.18). Interestingly, the corresponding ER domains is absent from GbnD2 Module 5. This ER activity has been postulated to be provided by the GbnD1 ER1 domain. However, according to the proposed biosynthetic pathways, a *trans*-acting ER activity is also required by module 10 of both assembly lines (Figure 1.18). The *trans*-acting ER activities of standalone GbnE and EntL in each cluster have been predicted based on their homologies to the putative *trans*-acting ER (BatK) involved in the Kalimantacin/Batumin biosynthesis[28,101]. Therefore, GbnE could be the other candidate to provide the ER activity in GbnD2 Module 5.

After the chain assembly process has passed module 5 in each PKS, β-branching is employed again in module 6 of both assembly lines. Interestingly, while the etnangien assembly line uses

only one set of 'HCS cassette' enzymes to carry out two β-branching events, two sets of ECH enzymes (GbnI/J and GbnR/S) are found in the gladiolin biosynthetic gene cluster. The subsequent biosynthetic steps appear to be identical for etnangien and gladiolin, except that an methyl group is predicted to be installed in module 15 of the former, even though no C-MT domain is present in the module.

As gladiolin and etnangien PKSs are striking similar, the catalytic origins of the structural differences between the two compounds is a very intriguing aspect of their biosynthesis. The shorter side chain of gladiolin indicates a different chain length control mechanism is employed by its assembly line in compare to etnangien, which is possibility linked to an additional enoyl reduction in the module 5.

1.3 Non-ribosomal peptides

1.3.1 Non-ribosomal peptide natural products

In addition to polyketides, non-ribosomal peptides (NRPs) represent a second major superfamily of natural products and have also been exploited in the development of numerous therapeutically important agents, such as antibiotics (e.g. vancomycin **22,** tyrocidine A **24** and daptomycin **26**), anticancer agents (e.g. bleomycin A2 **23**) and immunosuppressants (e.g. cyclosporin A **25**) as shown in Figure 1.20.[102] The broad range of the bioactivity of NRPs reflects their remarkable structural complexity and diversity, which is greatly expanded by various chemical modifications, including incorporation of fatty acid moieties, nonproteinogenic amino acids (including as *D*-amino acids), sugars and heterocycles as well as methylation and halogenation. The biological functions of NRPs are closely associated with their structural features, which often play an important role in constraining the peptide in its biologically active conformation. It is the unique conformation of each peptide that ensures its selective binding to a dedicated molecular target.[103,104]

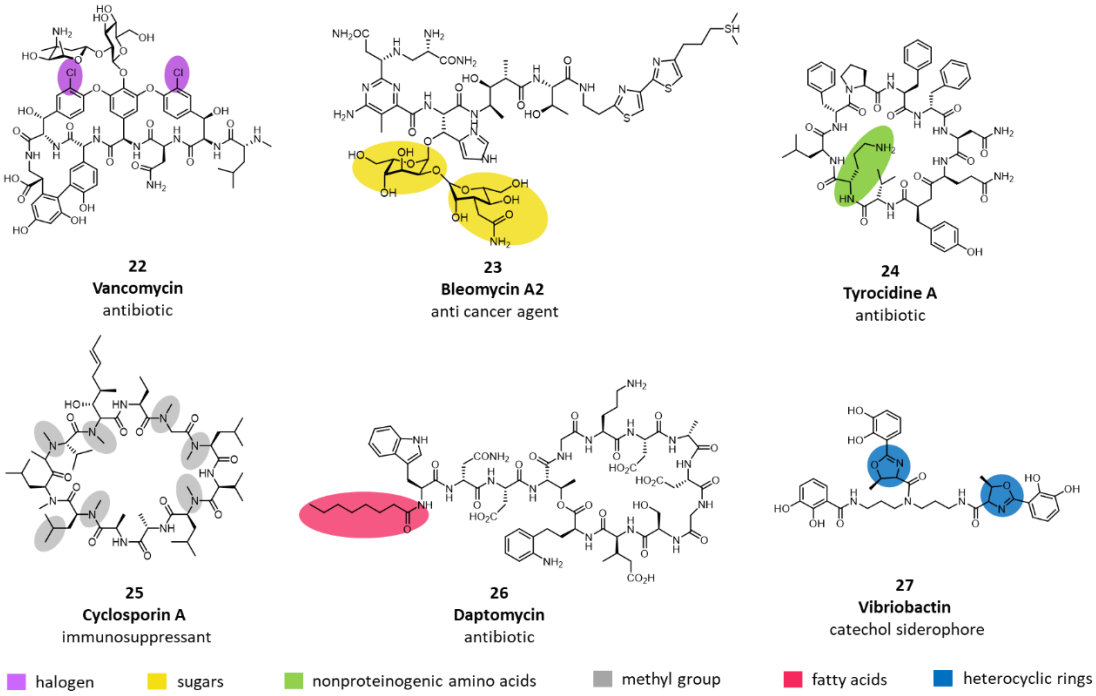


Figure 1.20 Representative bioactive non-ribosomal peptides. Examples of diverse structural modifications are highlighted with coloured ovals.

1.3.2 Non-ribosomal peptide biosynthesis

1.2.1.6 Non-ribosomal peptides synthetases

The majority of non-ribosomal peptides are assembled from simple amino acid building blocks by the modular non-ribosomal peptide synthetases (NRPSs), which are large multienzyme complexes containing catalytic domains organized in an assembly line fashion (Figure 1.21). NRPSs adopt a similar biosynthetic logic to type I modular PKSs to assemble the peptide through chain initiation, elongation and termination processes.

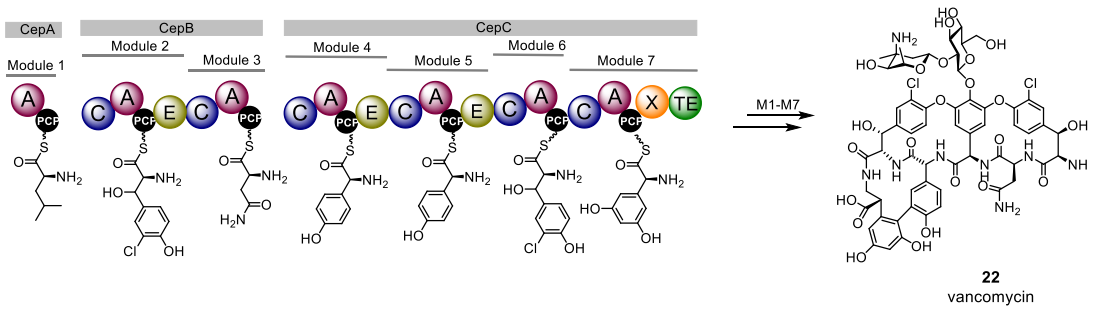


Figure 1.21 Schematic view of the representative modular NRPS that responsible for vancomycin biosynthesis.

The order and number of NRPS modules corresponds to the sequence and number of amino acids in the final product.

1.2.1.7 Chain elongation and core domains

The core domains required for peptide chain extension consist of an adenylation (A), peptidyl carrier protein (PCP) and condensation (C) domains. As shown in Figure 1.22, during each peptide chain extension cycle, the A domain first selects and binds the cognate amino acid building block and then activates it as the corresponding aminoacyl-AMP via reaction with ATP. The amino acyl adenylate intermediate is subsequently attacked by the thiol group of a 4'-phosphopantethiene arm that is post-translationally coupled to the adjacent PCP domain, resulting in a PCP domain-bond aminoacyl thioester. The aminoacyl substrate is then condensed with the peptidyl or aminoacyl thioester tethered to the upstream PCP domain by the C domain from the same module via a peptide bond formation. The elongated peptide intermediate is then provided as the substrate for further chain extension by the downstream module.[103,105]

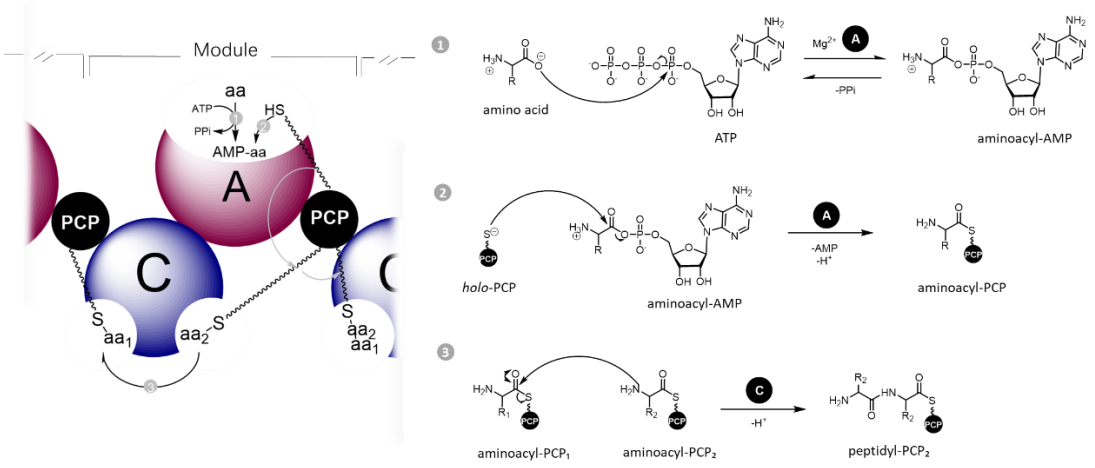


Figure 1.22 Schematic overview of non-ribosomal peptide chain elongation reactions catalyzed by the three core domains. (1) The A domains select and activate the amino acid (aa) using ATP to generate the corresponding aminoacyl-AMP. **(2)** The aminoacyl-AMP undergoes a nucleophilic attack from the thiol group of the PPant prosthetic group of the PCP domain forming an aminoacyl thioester. **(3)** The C domain subsequently catalyses a condensation reaction between the α-amino group of the PCP-bound acceptor amino acid with the thioester of the upstream PCP-bound donor amino acid or peptide.

Adenylation domains

A domains (~550 aa) function as the initial gate-keeper of the NRPS assembly lines as a consequence of their substrate specificity. Though A domains catalyse very similar chemistry as aminoacyl tRNA synthetases, which are part of the protein biosynthetic machinery, they are structurally and evolutionally distinct.[106] Several crystal structures of A domains have been determined, such as phenylalanine-activating PheA from the gramicidin S NRPS[107] and 2,3-dihydroxybenzoic acid specific DhbE from the bacillibactin NRPS[108]. They share significant structural similarity to enzymes from the adenylate-forming-superfamily, which includes acetyl CoA synthetases (ACS) [109] and acyl-CoA ligases. [110] The overall structure of A domains comprises a large N-terminal subdomain and a small flexible C-terminal subdomain with the active site located at the interface. Bioinformatic analysis combined with extensive structural and mutagenesis studies allowed the identification of 8-10 residues within the substrate binding pocket, which were shown to be crucial for substrate recognition.[111,112,113] Some of these residues interact with the carboxy and α-amino moiety of the substrate, whereas others interact with the substrate side chain. Aided by computational design, these insights have been exploited to rationally engineer the selectivity of A domains for novel NPRs biosynthesis.[113,114]

Peptidyl carrier protein domains

PCP domains are small (~100 aa) non-catalytic domains located downstream of A domains. Like ACP domains in PKSs, a conserved serine residue is found by all PCP domains. It is covalently modified via attachment of the a 4'-PPant moiety from CoA, a post-translational modification catalyzed by PPTases.[115] In doing so, the PCP domain is converted from the inactive *apo* form to the functional *holo* form with the thiol group of the flexible 4'-PPant arm serving as an anchor to covalently bind aminoacyl and peptidyl intermediate substrates. PCP domains are central in mediating NPR biosynthesis by shuttling biosynthetic intermediates between catalytic domains.[104] PCP domains are four-helix bundles with the conserved serine residue located in helix 2. Three distinct conformations of PCP domains were observed: A state (*apo*), H state (*holo*) and A/H state(*apo*/*holo*) as shown in Figure 1.23.[116] It was proposed that the transitions between the conformations contribute to further flexibility of the 4'-PPant arm of PCP domains in

interacting with different partner catalytic domains.[104]

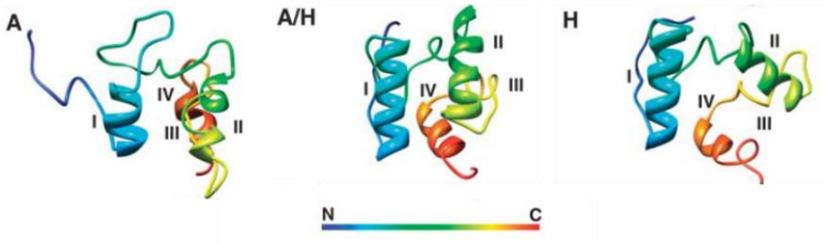


Figure 1.23 Ribbon diagrams of the solution NMR structures of the TycC3-PCP (from tyrocidine A synthetase) showing three conformers in the A, A/H, and H states (figure reproduced from ref.[116]).

Condensation domains

C domains normally contain ~450 aa and are located at the N-terminus of elongation modules. Bioinformatic analysis revealed C domains bear a conserved HHXXXDG motif and share sequence similarity with acyltransferases.[117] Comprehensive mutagenesis studies of the TycB1 C domain (from the tyrocidine A NRPS) and EntF C domain (from the enterobactin NRPS) highlighted the putative role of the second Histidine residue in the conserved H**H**XXXDG motif as a general base.[118,119] The catalytic mechanism of the C domains is proposed to involve deprotonation of the α-amino group of the acceptor aminoacyl-substrate thioester by the second histidine residue in the conserved motif to promote nucleophilic attack on the thioester of the acyl donor.[118,119]

Based on the crystal structures of the stand-alone C domain VibH (from the vibriobactin NRPS)[120] and the TycC PCP-C didomain (from the tyrocidine NRPS)[121] as well as several further structures,[122,123] C domains were found to be monomeric enzymes that are composed of N and C-terminal subdomains organized in a V-shape with the conserved HHXXDG active site located at the junction between them. A tunnel running through the active site was observed at the interface providing the entry to the donor and acceptor PCP-bond substrates (Figure 1.24 A).[120] A structure of the calcium-dependent antibiotic (CDA)-C1 domain with a substrate covalently tethered in the active site provided further insights to the role of the second His within the catalytic motif (H157).[124] In addition to the previously proposed catalytic role, H157 was observed to be critical for substrate positioning via forming a hydrogen bond with the α-amino group of acceptor aminoacyl substrate (Figure 1.24 B).[124]

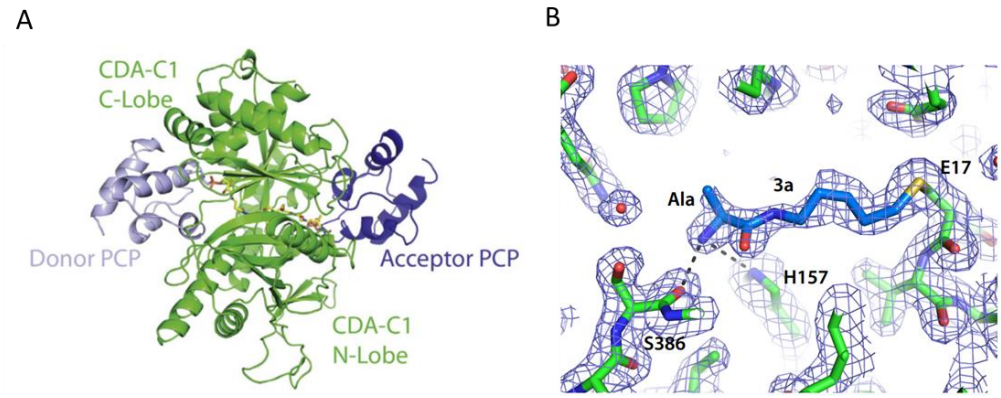


Figure 1.24 Crystal structure and detailed view of active site of CDA-C1 domain. (A) A Model of CDA-C1 domain with PCP domains bound showing the active site tunnel based on the crystal structures of CDA-C1 (highlighted in green, PDB:4JN3) with the position of acceptor PCP domain (shown in dark blue) modeled from PDB:2VSQ and the position of donor PCP domain (shown in light blue) modeled from biochemical data (figure reproduced from ref.[125]). **(B)** Active site of CDA-C1 domain, containing engineered E17C mutation with bromoalkylamine analogue covalently bound to mimic the substrate delivery by PCP domain, shows the important role of H157 in positioning α-amino group of the substrate for nucleophilic attack via forming a hydrogen bond (figure reproduced from ref.[124]).

Several bioinformatic and phylogenetic studies show C domains clade according to their biochemical activity.[126,127] Subtypes were found as shown in Figure 1.25 A : The subtype $^{L}C_{L}$ catalyses the condensation between L-configured amino acid building blocks tethered on PCP domains; $^{D}C_{L}$ domains link the acceptor L-aminoacyl substrate with a D-configured donor;[128] starter C domains, situated at the priming position of the NRPS, acylate the first loaded amino acid a the 3-hydroxyacyl unit hijacked from fatty acid biosynthesis;[129] dual E/C domains first epimerize the C-terminal amino acid in the donor peptide chain and subsequently condense it to with the acceptor aminoacyl substrate;[130] the Cy domain first performs the condensation between aminoacyl/peptidyl-PCP donors and serinyl-/threoninyl-/cysteinyl-PCP acceptors, and then catalyses the cyclodehydration between the thiol or hydroxyl side chains of the donor residues and the carbonyl group of the newly-formed peptide bond, resulting in incorporation of a heterocycle into the growing peptide chain.[131] Profile Hidden Markov Models (pHMMs) for the sequence motifs as well as for the entire sequences of C domain subtypes were provided as shown in Figure 1.25 B, the determined specificity conferring positions of which have greatly

facilitated C domain subtype identification and engineering.

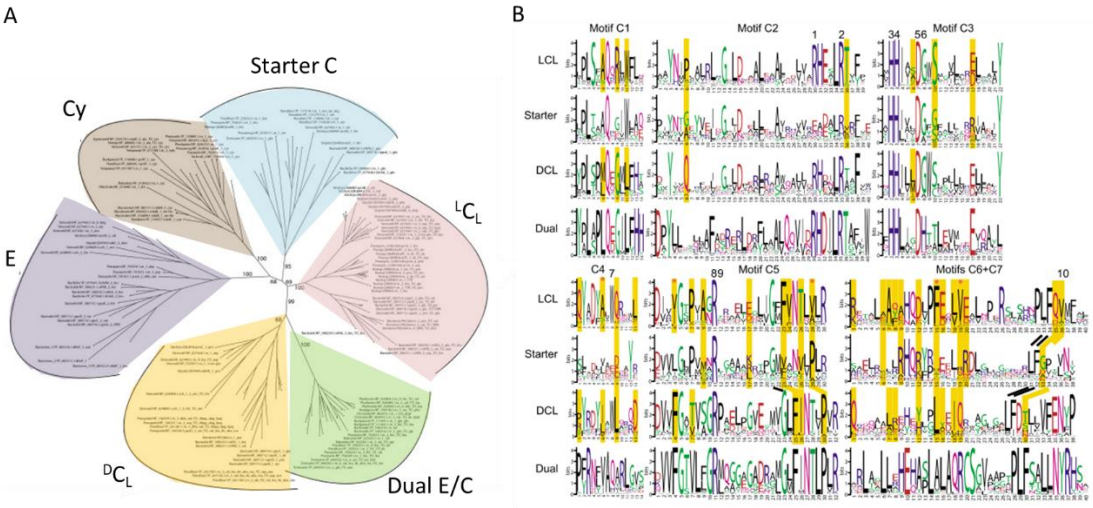


Figure 1.25 Classification of NRPS C domains based on phylogenetic and bioinformatic analysis (reproduced and adapted from ref.[127]) . (A) Phylogenetic tree of C domains showing the subtypes. **(B)** Hidden Markov Models (HMMs) of the C domain subtypes. Yellow bars indicate significant specificity determining positions in $^{L}C_{L}$, Starter and $^{D}C_{L}$ domains with the most significant residues highlighted with a red star.

1.2.1.8 Chain modification

In addition to the core domains catalysing chain elongation, NRPSs can contain a variety of chain modification domains to introduce other chemical features, which enrich the structural diversity and complexity of NRPs (Figure 1.20). As well as the dual E/C domains, introductino of D-configured amino acid residues into the peptide chain can be mediated by individual epimerization (E) domains, through epimerization of newly incorporated L-amino acid with the growing peptide chain.[132,133] The thiazolines and oxazolines incorporated by Cy domains can be further oxidized to the corresponding thiazoles and oxazoles by oxidation (Ox) domains using FMN as a cofactor, as in the biosynthesis of bleomycin and epothilone.[134,135] These heterocycles could be alternatively reduced to thiazolidines and oxazolidines respectively by a NADPH-dependent reductase (R) domains, for example during pyochelin assembly.[136] Both N-methyl and C-methyl groups are also commonly installed into the NRPs, such as cyclosporin,[137] actinomycin[138] and yersiniabactin.[139] The methylations are catalyzed by N-methyl transferase (N-MT) and C-methyl transferase (C-MT) domains using methyl donor, S-adenosyl methionine (SAM)

as methyl donor.

1.2.1.9 Chain initiation and release

Peptide assembly is usually initiated by the activity of an A domain located in the loading module, which normally lacks a C domain. However, in the biosynthesis of lipopeptides, such as surfactin, fengycin and gliadobactin, a starter C domain (discussed previously) is always found at the N-terminus of the loading module.[126] In this kind of chain initiation mechanism, the starter C domain catalyses N-acylation of the loaded amino acid with a 3-hydroxylacyl unit.[129] The loading modules of NRPS sometimes contain additional chain modification domains, for example a E domain is located in first module of the Gramicidin S assembly line.[140]

Once the peptide chain has been fully assembled, it is released from the assembly line either as a linear product, such as vancomycin, or more commonly as a macrocyclic product, such as surfactin. This process is normally accomplished by a C-terminal TE domain, which utilizes a similar catalytic mechanism to type I TE domains from modular PKSs (discussed in the previous section).[141,142] Alternatively, in some cases, the product is liberated as a linear aldehyde or alcohol via NADPH-dependent reduction catalyzed by a C-terminal thioester reductase (TR) domain.[143]

1.3.3 Biosynthesis of icosalides

Genome sequencing of *B. gladioli* BCC238 identified a single ~15 kb gene encoding an NRPS proposed to be responsible for the biosynthesis of icosalides(Figure 1.26 A). Within the NRPS, four adenylation domains were identified, with their predicted substrate specificities in agreement with amino acid residues incorporated into the icosalides (A1: Leucine, A2: Serine, A3: Leucine, A4: Serine). Five C domains and a cyclising TE domain were assigned to the six bond-forming reactions required for the assembly of the 20-membered cyclic core of the icosalides. Phylogenetic analysis of the five C domains revealed they clade into the following subtypes (Figure 1.26 B): C3 and C5 are $^{L}C_{L}$ domains, C2 is a dual E/C $^{D}C_{L}$ domain and C1 and C4 are starter C domains. The dual epimerization-condensation activity of the C2 domain accounts for the existence of the D-configuration of the Leu residue in the final product. The presence of the two

starter C domains, C1 and C4, is also in agreement with the incorporation of two 3-hydroxy acyl chains into the icosalides. Inactivation of *icoA* abolished production of the icosalides, confirming the involvement of the NRPS in the icosalides biosynthesis (Figure 1.26 C)The biosynthetic pathway proposed for the icosalides is shown in Figure 1.27.

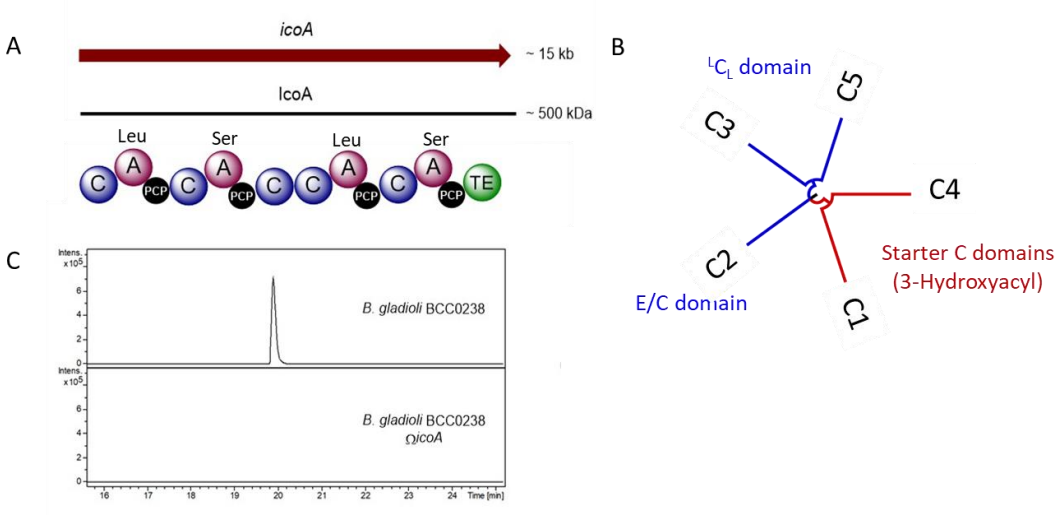


Figure 1.26 Identification of the gene encoding the icosalide NRPS biosynthetic gene and phylogenetic analysis of the C domains (unpublished data from Challis group). (A) Schematic of the ~15 kb *icoA* NRPS biosynthetic gene. The sequence of enzymatic domains within the NRPS is shown below. **(B)** phylogenetic analysis of the C domains of IcoA NRPS. 3-hydroxyacyl unit incorporating starter C1 and C4 domains are highlighted in red. **(C)** Insertional mutagenesis of *icoA* abolishes icosalide A1 production in *B. gladioli* BCC0238.

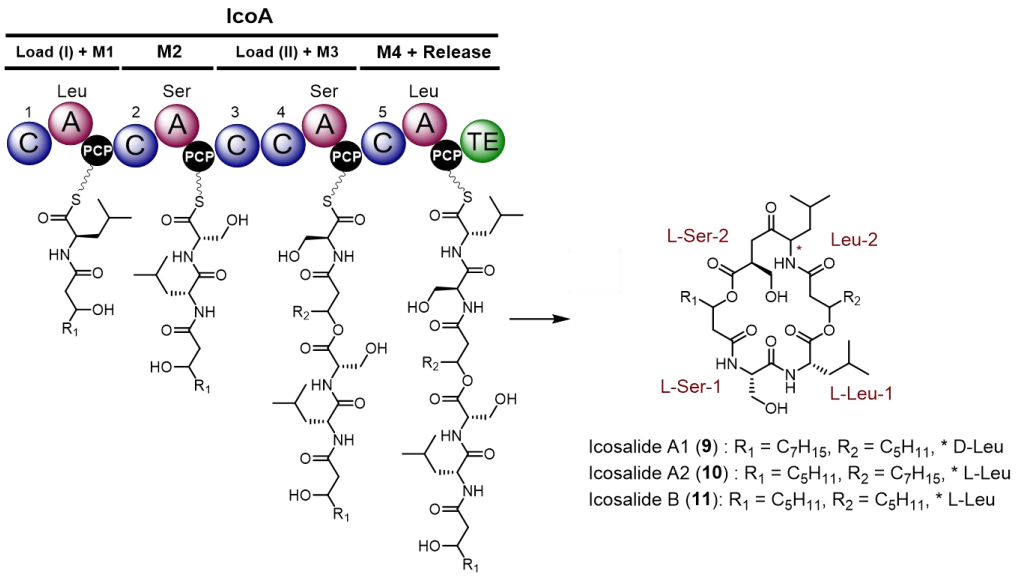


Figure 1.27 NRPS and proposed biosynthesis of icosalides. Domain and module organization of the icosalide

NRPS showing the proposed structure of chain elongation intermediates. The two starter condensation domains are highlighted by red numbers. The starter condensation domain C_4 is 'embedded' in the predicted NRPS. Substrate specificity of each adenylation domain is denoted above each domain.

The icosalide NPRS exhibits an unprecedented domain architecture, with the chain initiating C domain IcoA C4 situated in the middle of IcoA and directly adjacent to the IcoA C3 domain. The IcoA C4 domain is believed to be the first example of a starter condensation domain that is internal to an NRPS. The domain organization of module 3 also indicates IcoA C3 domain and IcoA C4 domain are required to co-operate for accomplishing a second round of chain initiating event as well as a chain elongation event in this module, the biosynthetic mechanism of which is of considerable interests.

1.4 Study aims

The aims of this study were to gain a better understanding of gladiolin and icosalides biosynthesis in *B. gladioli* with a long-term goal of employing the obtained insights to engineer their and related biosynthetic pathways.

The first aim was to uncover the biosynthetic origins for the key structural differences between gladiolin and etnangien. This included identification of the enoyl reductase required by module 5 of the gladiolin PKS and probing the molecular basis for the non-iteration of this module.

The second aim was to investigate the double chain initiation mechanism underpinned by the unprecedented domain organization of the icosalide NRPS. This involves characterization of the internal starter and the adjacent 'normal' condensation domains in module 3 of icosalide NRPS.

Chapter 2

Characterization of *cis*- and *trans*-acting enoyl reductases in the gladiolin polyketide synthase

As discussed in section 1.2.3, the gladiolin and etnangien PKSs possess strikingly similar domain architectures, yet produce quite different biosynthetic products. A key structural difference is the hexaene moiety of etnangien (absent in gladiolin), which is likely installed by the iterative activity of EtnE Module 5 resulting in the three double bonds between C26-C31, and contributes to the inherent instability of etnangien (Figure 2.1A). The corresponding region in gladiolin is a saturated centre between C26-C27, suggesting an enoyl reduction event has halted the formation of a polyene (Figure 2.1B). Intriguingly, the corresponding module 5 of the gladiolin PKS does not harbor an ER domain to catalyze the reduction of the α,β-unsaturated intermediate.

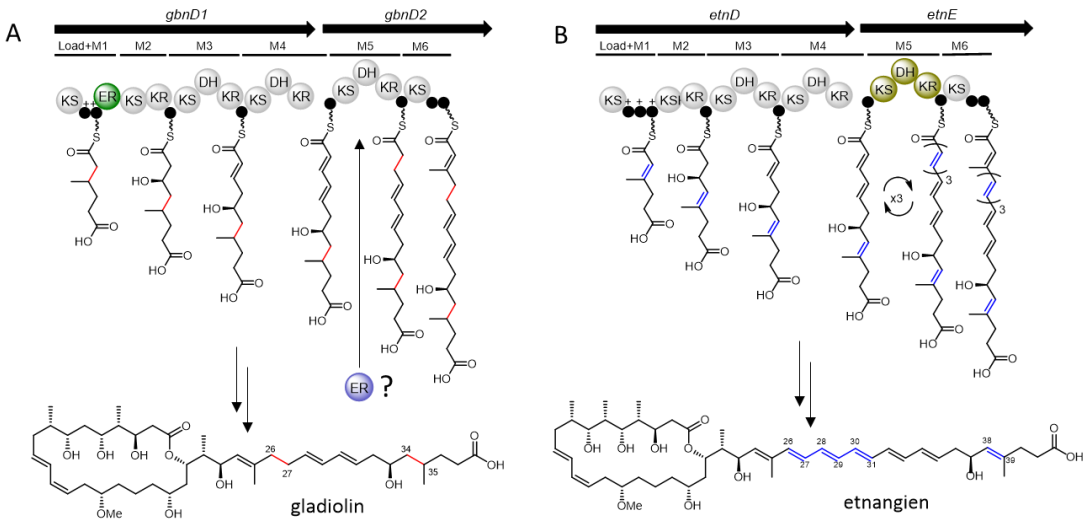


Figure 2.1 Partial biosynthetic pathways of the gladiolin (A) and etnangien (B) PKSs. The iterative module 5 of

the etnangien PKS is highlighted in gold. The missing ER required for the additional enoyl reduction event in module 5 of the gladiolin PKS is denoted. The substituted ER domain in module 1 of the gladiolin PKS is highlighted in green. Structural differences in the intermediates and final products resulted from the two additional events and iteration and non-iteration of module 5 are highlighted in red (gladiolin) and blue (etnangien).

In order to investigate this phenomenon, the ER required by module 5 of the gladiolin PKS is needed to be first identified. Building upon the previous detailed analyses of the gladiolin and etnangien biosynthetic gene clusters, the most likely candidate was assigned as the sole difference between the two PKSs architecture, the substituted ER (GbnD1 ER1) domain in module 1 of the gladiolin PKS (Figure 2.2). Here, the GbnD1 ER1 domain would be predicted to act dual-functionally that catalyzes the enoyl reduction in GbnD1 module1 in an *intra*-modular manner, and in GbnD1 Module 5 in an *inter*-modular manner. Another feature identified from bioinformatics analysis of both clusters is the presence of a putative *trans*-acting enoyl reductase GbnE/EntL are proposed to reduce the intermediate in module 10 of their PKSs according to the structure of the final products (Figure 2.2). However, with this knowledge, it is feasible that GbnE could be the second candidate supplying dual-functional *trans*-acting ER activity in module 5 apart from in module 10.

This chapter describes efforts to elucidate the 'cryptic' ER activity in the module 5 of the gladiolin PKS. Using a combination of *in vivo* genetic engineering and *in vitro* biochemical assays, the function of the aforementioned two candidate ERs in the gladiolin biosynthesis are investigated.

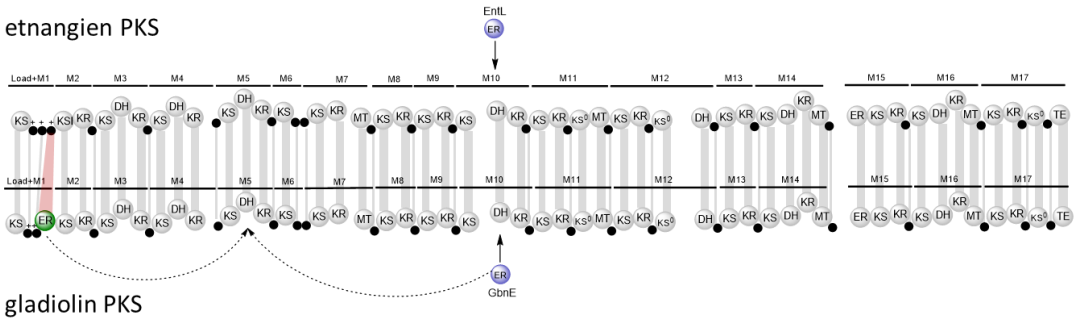


Figure 2.2 Comparison of catalytic domain architectures of gladiolin and etnangien PKSs showing the hypothetic catalytic origin of the additional enoyl reduction event in module 5 of the gladiolin PKS. 70 out 71 catalytic domains are identical between the two assembly lines. The two candidates for the ER required by

module 5 of the gladiolin PKS are denoted as: 1) the sole domain architecture difference, the substituted GbnD1 ER1 domain in the gladiolin PKS (highlighted in green), and 2) the putative *trans*-acting ER GbnE (highlighted in purple). The counterpart of GbnE in the etnangien PKS is shown as EntL.

2.1 Characterization of the GbnD1 ER1 domain

2.1.1 *In vivo* characterization of the GbnD1 ER1 domain

2.1.1.1 *In-frame* deletion of *gbnD1* ER1

An *in vivo* inactivation approach was first employed to investigate the function of the GbnD1 ER1 domain in gladiolin biosynthesis. *In-frame* deletion of the DNA-encoding region for the GbnD1 ER1 domain was considered to be the most direct strategy to inactivate this domain. This approach also aimed to mimic the catalytic domain organization of the etnangien PKS at the sequence level (minus an ACP domain), in attempt to elicit etnangien-like biosynthesis.

In-frame deletion of targeted chromosomal regions in *B. gladioli* were achieved by the homologous recombination mutagenesis system, based on the yeast homing endonuclease I-SceI[144] (Figure2.3). The system involves cloning the sequences flanking the chromosomal region targeted for deletion into the suicide plasmid pGPI-SceI that cannot replicate in *Burkholderia* strain and carries the I-SceI recognition site. This plasmid is then transferred to *Burkholderia* by tri-parental mating, resulting in its targeted insertion into the chromosome via homologous recombination and the introduction of a single crossover mutant allele. The second plasmid pDAI-SceI that constitutively expresses the I-SceI nuclease was then introduced into the single crossover mutant allele. I-SceI causes a double strand break into the inserted plasmid sequence, which stimulates intramolecular homologous recombination between the flanking region on plasmid and chromosome. The resolution of this cointegrate can either restore the parental allele or cause a gene deletion, depending on the site of the crossover.

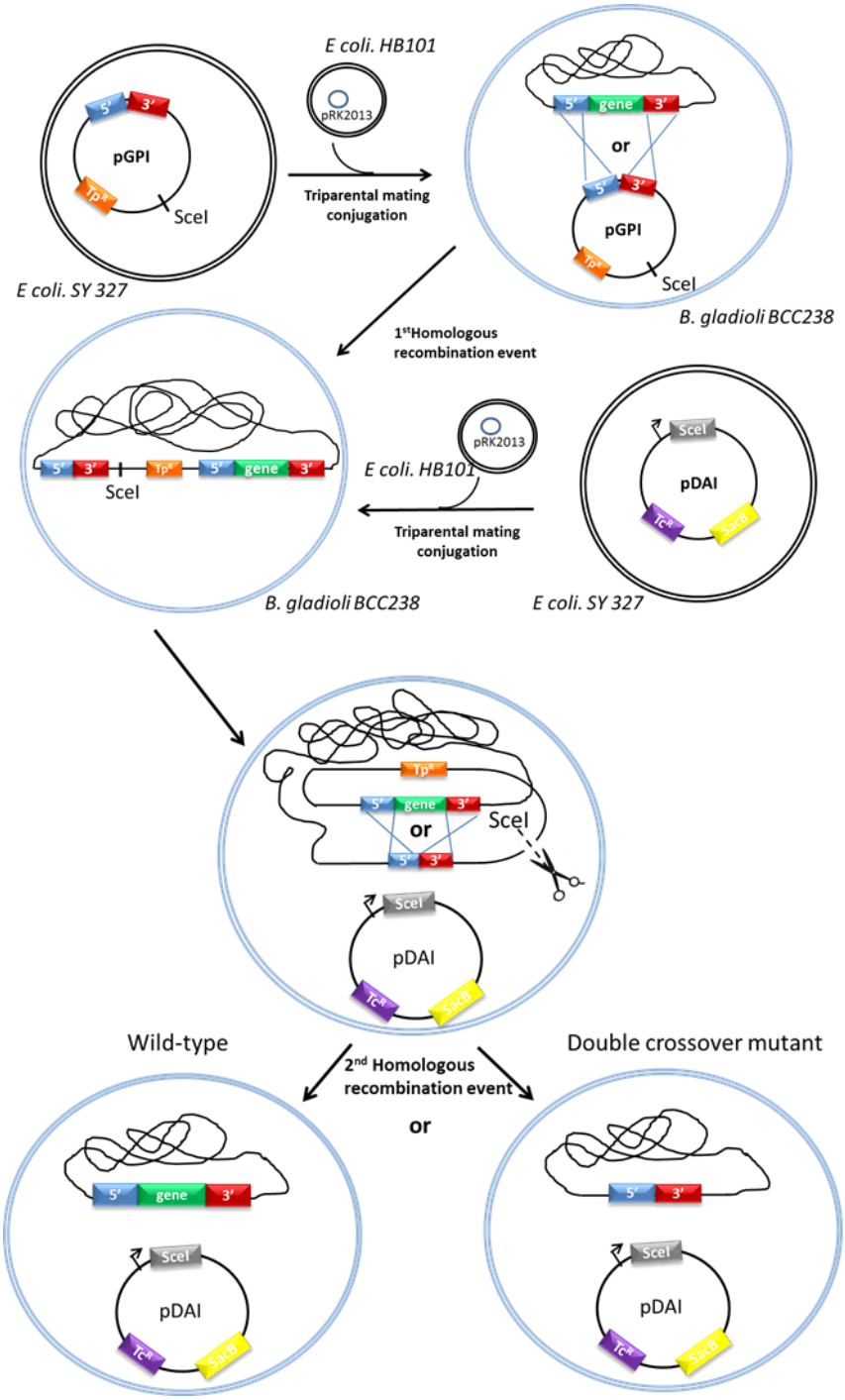


Figure 2.3 Work flow of the mutagenesis system applied in *in-frame* deletion in *Burkholderia*. Blue and red box: 5' and 3'flanking regions, green box: targeted gene, grey box: SceI nuclease encoded in pDAI-SceI, orange and purple box: trimethylprime and tetracycline resistance gene, yellow box: counter selection marker SacB, arrow on pDAI: constitutive promoter, blue cross marks are referred to homologous recombination events.

As the previous trial of gene deletion in the original gladiolin producer *B. gladioli* BCC0238 resulted in failure due to its spontaneous resistance to the antibiotic for the mutant selection,

the *in vivo* investigation was then carried out in another gladiolin producing isolate *B. gladioli* BCC1622, which harbors the identical biosynthetic gene cluster (99.5% identity of DNA sequences). Metabolite profiling of wild type *B. gladioli* BCC1622 by LC-MS also observed another two gladiolin related compounds which present in the extract of *B. gladioli* BCC0238. Structural elucidation (by Dr. Lijiang Song) confirmed one of them is the iso-gladiolin **28**, generated via rearrangement, and the other is the linear gladiolin **29** due to hydration of gladiolin[28] (Figure 2.4).

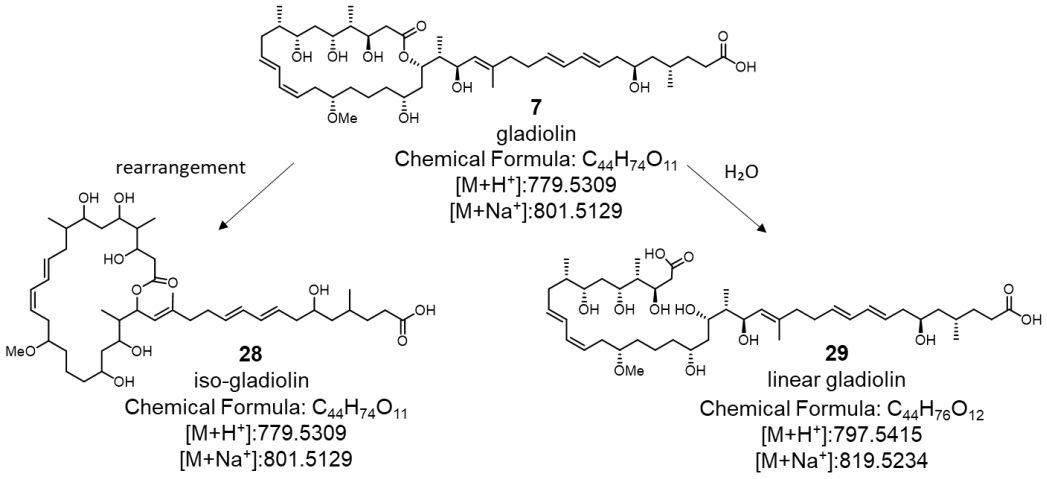


Figure 2.4 Structure comparison of gladiolin and its analogues. Iso-gladiolin **28** and linear-gladiolin **29** derive from rearrangement and hydration of gladiolin respectively.

To *in-frame* delete *gbnD1* ER1 following the strategy described above, the deletion construct pGPI-*gbnD1*_ER1 was generated and mobilized into *B. gladioli* BCC1622 (detailed in 6.3.1). The *gbnD1* ER1 deletion mutant *B. gladioli* BCC1622_Δ*gbnD1*_ER1 was successfully obtained and confirmed by PCR. (Figure 2.5).

Having generated the *B. gladioli* BCC1622_Δ*gbnD1*_ER1 mutant, the effect of the mutation on gladiolin production could then be examined. Comparative metabolite profiling of *B. gladioli* BCC1622_Δ*gbnD1*_ER1 and WT *B. gladioli* BCC1622 was conducted by LC-MS (detailed in 6.4). The *B. gladioli* BCC1622_Δ*gbnD1*_ER1 mutant base peak chromatogram (BPS) showed the three gladiolin related peaks were absent, with no obvious new peak being formed (Figure 2.6 A). Furthermore, the extracted ion chromatogram (EIC) confirmed the production of gladiolin was abolished in BCC1622_Δ*gbnD1*_ER1 (Figure2.6 B).

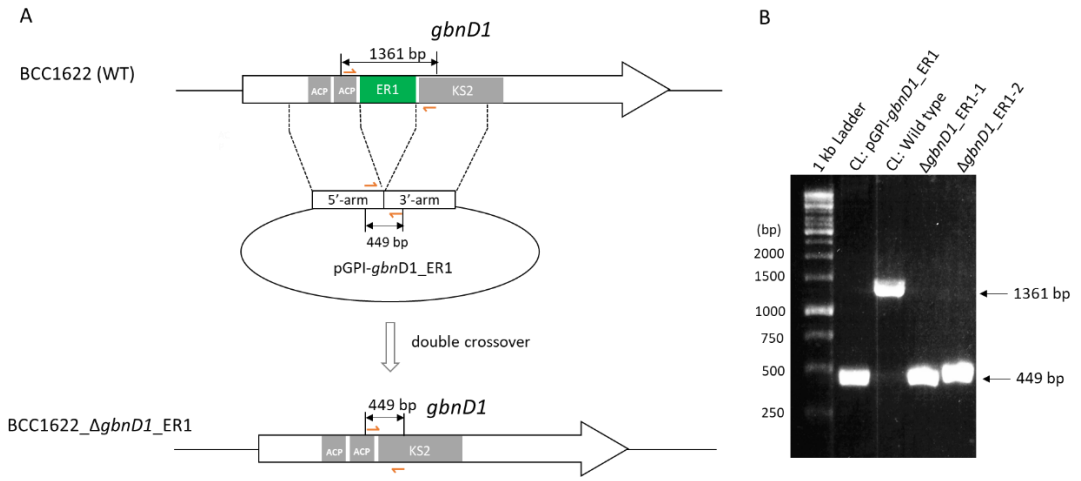


Figure 2.5 Schematic representation of generation and PCR confirmation of *B. gladioli* BCC1622_Δ*gbnD1*_ER1. **(A)** Targeted *in-frame* deletion of the *gbnD1* ER1 region of wild type chromosome is highlighted in green. Loss of the green region on the mutant chromosome after double crossover denotes successful deletion. Checking PCR primers are indicated by orange arrows and their binding sites on the mutant and WT chromosome and deletion plasmid DNA are indicated, respectively. The length of corresponding predicted PCR products is shown. **(B)** DNA electrophoresis gel of the PCR products confirms the genome type of the deletion mutant, resulting in a 449 bp product as observed in the positive control (deletion plasmid), and the negative control (WT allele) resulted in a 1361 bp product. 'CL' refers to control, 'WT' refers to *B. gladioli* BCC1622 wild type.

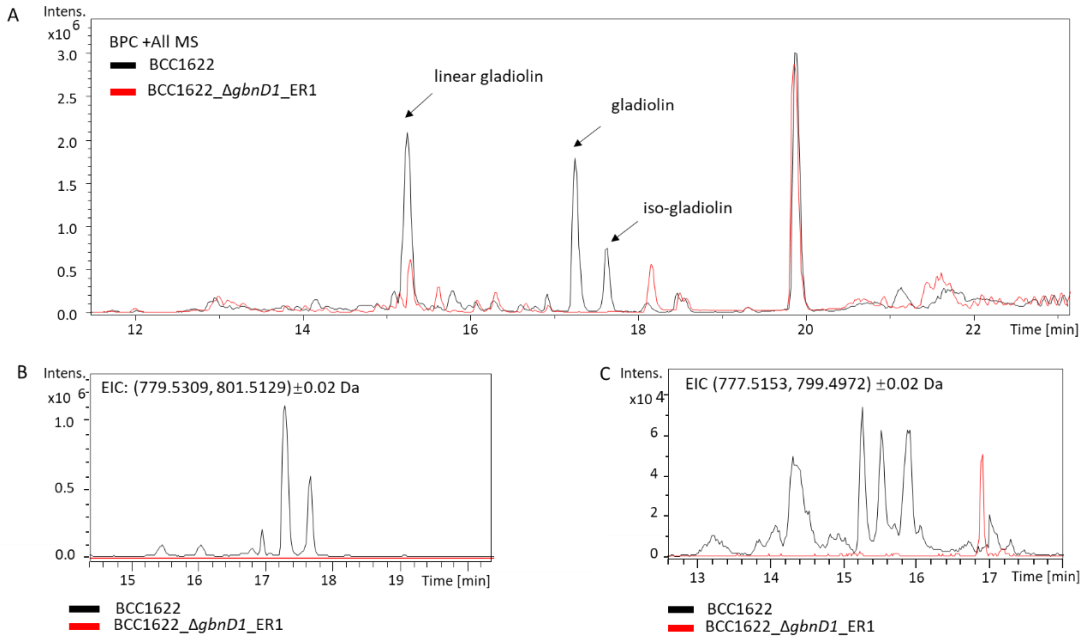


Figure 2.6 LC-MS analysis of metabolite extracts from BCC1622 wild type and BCC1622_Δ*gbnD1*_ER1. All chromatograms follow the layout of; WT *B. gladioli* BCC1622 indicated in black and *B. gladioli*

BCC1622_Δ*gbnD1*_ER1 mutant indicated in red. **(A)** Base peak chromatograms of metabolite extracts. **(B**) EICs at *m/z*=779.5309 ± 0.02 Da and 801.5129 ± 0.02 Da, corresponding to the $[M+H]^+$ and $[M+Na]^+$ ions of gladiolin **7** and iso-gladiolin **28**, **(C)** EICs at *m/z*= 777.5153 ± 0.02 Da and 799.4972 ± 0.02 Da corresponding to $[M+H]^+$ and $[M+Na]^+$ ions of unsaturated gladiolin **32**. See appendix for mass spectra of the EICs.

The possible products produced by the GbnD1 ER1 domain inactivated mutant are expected as shown in Figure 2.7, based on three scenarios: 1) The GbnD1 ER1 domain is of dual function, performing enoyl reduction in both module 1 and module 5, and the activity of GbnD1 ER1 in module 5 halts the inherent iterative nature of module 5, thus inactivation of GbnD1 ER1 would resulted in generation of structure **30**; 2) The GbnD1 ER1 domain is of dual function but GbnD2 Module 5 has no underlying ability to perform iteration, thus inactivation of GbnD1 ER1 would generate structure **31**; 3) The GbnD1 ER1 domain only reduces the α, β-unsaturated intermediate arising from module 1 and has no effect on module 5, thus inactivation of GbnD1 ER1 would generate structure **32**.

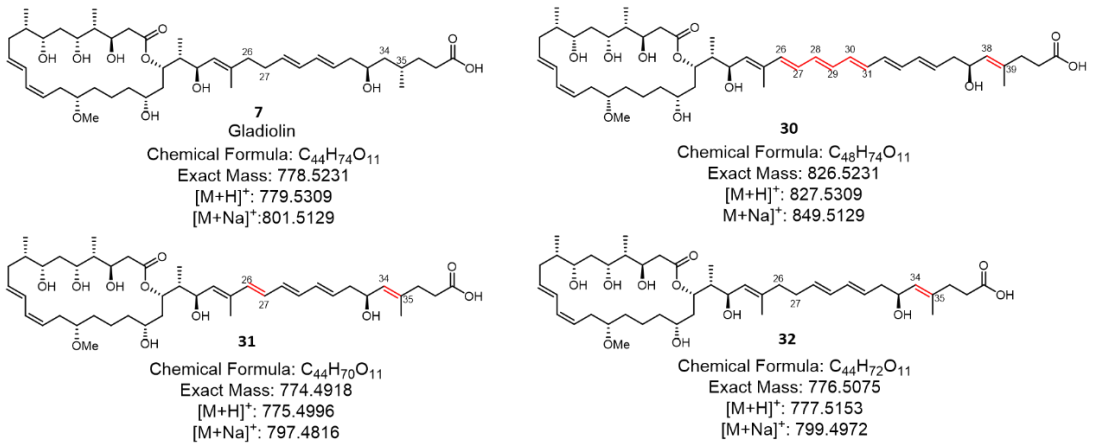


Figure 2.7 Structural comparison of gladiolin and proposed derivatives arising from inactivation of GbnD1 ER1 *in vivo*. Structural differences from gladiolin are highlighted in red, and m/z values for $[M+H]^+$ and $[M+Na]^+$ are shown. The proposed derivatives are comprised of the structural differences resulting from a combination of possible consequences of GbnD1 ER1 inactivation. The desaturation of C34-C35 (in all three derivatives) originates from elimination of GbnD1 ER1 activity in module 1. The desaturation of C26-C27 (in **30** and **31**) originates from elimination of the proposed GbnD1 ER1 additional activity in module 5. The installation of a polyene region at C26-C31 (in **30**) is the predicted result of restoring the iteration of module 5 by removing GbnD1

ER1 function towards this module.

To probe whether gladiolin derivatives **30**-**32** (Figure 2.7) were produced after deletion of *gbnD1* ER1, the corresponding $[M+H]^+$ and $[M+Na]^+$ ions were searched in both mutant and WT metabolite extracts. The ions with molecular weight corresponding **30** and **31** were not detected in both the WT and mutant extract (data not shown). This result indicates the GbnD1 ER1 domain does not catalyse the *inter*-modular enoyl reduction in module 5. Interestingly, multiple peaks with *m/z* value corresponding to **32** were observed in the WT, whilst a new single peak with a different retention time (16.9 min) appeared in the mutant (Figure 2.6 C). Due to the very low intensities of all the peaks being noted, it was not possible to purify them for structural elucidation. However, none of these peaks could be observed in metabolite extraction from the single crossover mutant (data not shown), which strongly suggested they were all related to gladiolin biosynthesis.

The new peak (16.9 min) in the mutant likely corresponds to **32** as this would be consistent with the predicted GbnD1 ER1 domain's *cis*-acting enoyl reduction activity in module 1. The multiple peaks in the WT EIC at *m/z*= 777.5153 ± 0.02 Da and 799.4972 ± 0.02 Da (the same of the one of **32**) are predicted to be **33** and **34** (Figure 2.8), as well as their iso-gladiolin analogous. They would be formed by the *trans*-acting enoyl reduction events in module 5 or module 10 not being fully efficient during polyketide assembly. The low yield of **32** from the mutant strain could be the result of the excision of the GbnD1 ER1 domain from GbnD1 affecting the overall folding of GbnD1 which slowed down the polyketide assembly. This could also be an alternative explanation for inability to detect the derivative **30** and **31**.

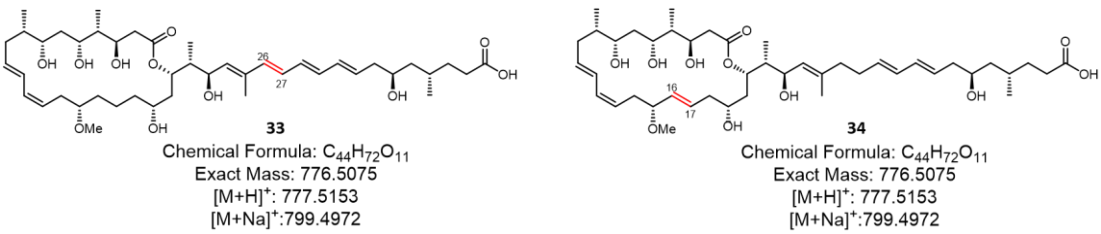


Figure 2.8 Proposed structure of unreduced gladiolin derivatives. Structures were proposed based on the hypothesis of *trans*-acting enoyl reduction events in module 5 (generates **33**) or module 10 (generates **34**) not being fully efficient during the polyketide assembly.

2.1.1.2 Site-directed mutagenesis of GbnD1 ER1 domain in the gladiolin PKS

In order to avoid the disturbance to the overall structure of GbnD1, inactivation of the GbnD1 ER1 domain was alternatively pursued via site-directed mutagenesis. This strategy aims complete domain inactivation without significantly affecting the conformation or folding of the GbnD1 protein.

The *cis*-acting ER domains of modular PKSs belong to the NAD(P)H-dependent medium-chain dehydrogenase/reductase (MDR) superfamily of enzymes, which typically comprise of two subunits: a N-terminal catalytic subdomain and a C-terminal Rossman-fold NADPH bind subdomain [81]. Unfortunately, unlike other domains from modular PKSs, there is very limited structural and catalytical information on ER domains. Previous mutagenesis studies of the proposed catalytically important residues in the ER domain from rapamycin PKS (RapER13) and spinosyn PKS (SpnER2) all resulted in only lowering but not abolishing the activity (as introduced in section 1.2.1.5 '*cis*-acting Enoyl reductases')[82,80]. However, site-directed mutagenesis of the NADPH binding motif of ER domains from nystatin (NysER5) [145] and erythromycin (EryER3)[146] PKS have been reported to successfully inactivate their enoyl reduction activity. Changing the two adjacent glycine residues in the conserved NADPH binding motif HAAA**GG**VGMA of NysER5 and EryER3 to a serine and a proline resulted in non-production of parental products and relatively high production of unreduced analogous[145,146], suggesting an completely inactive ER domain and a fully functioning assembly line in each instance.

Further inspection of the NADPH binding pocket in SpnER2, the only reported crystal structure of ER domain from modular PKSs, suggested the two adjacent glycine residues (G138 and G139) located at the active site and involved in the binding of the NADPH (Figure 2.9). Replacing G138 with serine and G139 with proline would disturb conformation of the α-helix thus impede accommodation of NADPH. This observation further supported the logic of mutagenesis study in NysER5 and EryER3. Thus, one possible strategy to inactivate the GbnD1 ER1 domain via mutagenesis, in the absence of structural information, would be following the mutagenesis approach outlined above.

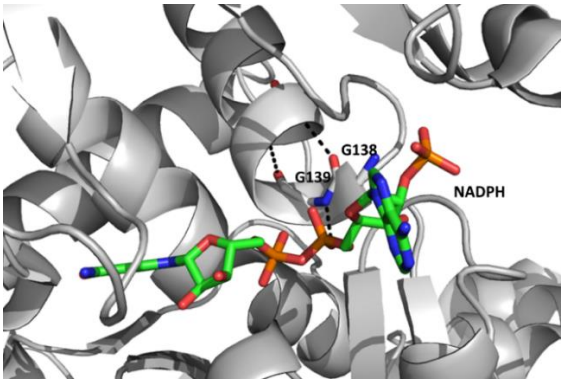


Figure 2.9 Structure of NADPH binding site in SpnER2 domain from spinosyn PKS. Cofactor binding site of SpnER2 shows G138 and G139 (shown as sticks) involved in the binding of NADPH. Structure created in Pymol from PDB: *3S*LK[80].

The cofactor binding motif of the GbnD1 ER1 domain was identified by sequence alignment with other ER domains from modular PKSs (Figure 2.10, see appendix for full result). The two adjacent glycines embedded in the NADPH binding motif of the GbnD1 ER1 domain were located at positions 388 and 389. The subsequent double mutation (G388S/G389P) was achieved by *in vitro* site-directed mutagenesis of the *gbnD1* ER1 DNA encoding region and then an *in-frame* 'knock-in' of the mutated *gbnD1* ER1(G388S/G389P) segment back into *B. gladioli* BCC1622_Δ*gbnD1*_ER1 by double crossover homologous recombination (detailed in 6.3.2) (Figure 2.11 A). This was achieved based on the same yeast homing endonuclease I-SceI[144] system used for *in frame* deletion as previous described (Figure2.3). Both PCR analysis (Figure 2.11 B) and sequencing of the designed mutation region confirmed the successful mutagenesis.

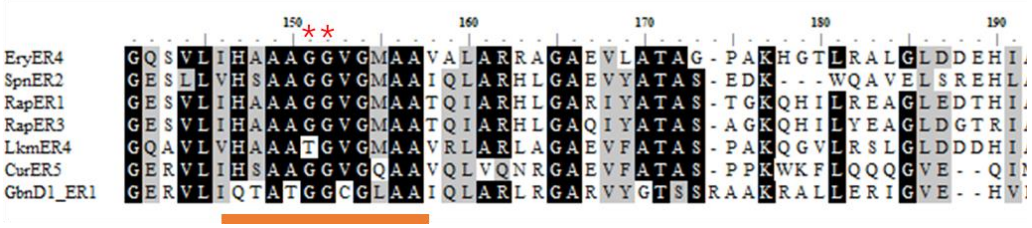


Figure 2.10 Partial result of multiple sequence alignment between GbnD1 ER1 domain and reported ER domains from type I modular PKSs. The conserved NADPH binding motif is indicated as an orange bar and the conserved two adjacent glycine residues are indicated by a red star. The numbers after ER domains refer to the module number within the respective PKS. Ery, erythromycin; Spn, spinosyn; Rap, rapamycin; Lkm, lankamycin; Cur, curacin.

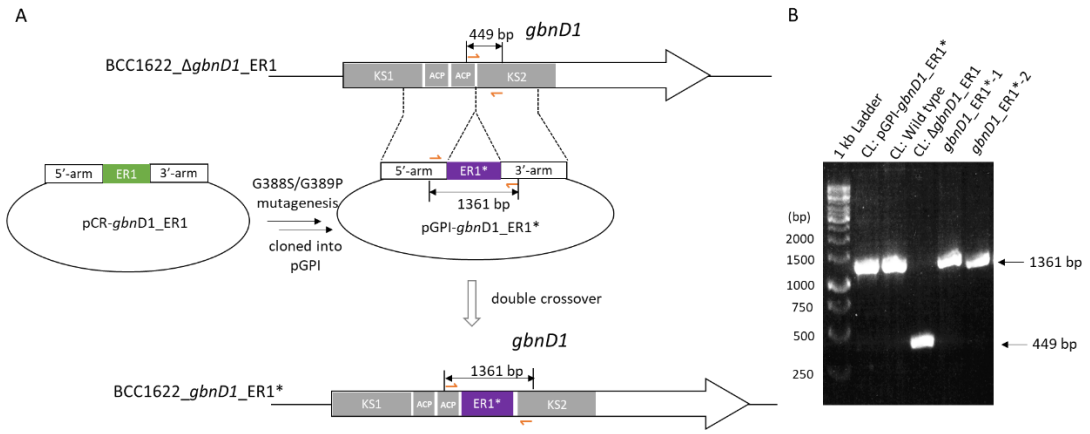


Figure 2.11 Schematic representation of generation and PCR confirmation of *B. gladioli* BCC1622_ *gbnD1*_ER1 (G388S/G389P). (A) Wild type *gbnD1* ER1 cloned with flanking arms is denoted in green. Mutated *gbnD1* ER1* fragments on pGPI plasmid and the *in-frame* knock-in mutant chromosome are highlighted by a purple. Insertion of the purple region on the mutant chromosome after double crossover denotes as a successful knock-in. Checking PCR primers are indicated by orange arrows and their binding sites on the mutant chromosome, parental allele BCC1622_Δ*gbnD1*_ER1 chromosome and knock-in plasmid pGPI-*gbnD1*_ER1* are indicated together with the length of corresponding predicted PCR products. (**B)** DNA electrophoresis gel of the PCR products confirmed the genotype of mutants, resulting in a 1136 bp product as observed in positive controls (WT and knock-in plasmid), while negative control (parental allele) resulted in 449 bp product. 'CL' refers to control, '*' refers to G388S/G389P mutation.

Comparative metabolite profiling of the mutants and wild type was then conducted by LC-MS analysis. Again, gladiolin production was abolished in BCC1622_*gbnD1*_ER1*(G388S/G389SP) as shown in BPCs with no obvious new peak observed (Figure2.12 A). The abolished production of gladiolin was confirmed by EIC (Figure2.12 B). All predicted derivatives (Figure 2.7) were searched in the mutant and wild type metabolite extracts. A similar gladiolin derivatives production pheno type was observed in BCC1622_*gbnD1*_ER1*(G388S/G389P) mutant with the *gbnD1* ER1 *in-frame* deletion mutant (Figure 2.12 C). Only ions with molecular weight corresponding to unreduced gladiolin **32** could be detected (16.9 min) and the yield did not significantly increase compared to in the *gbnD1* ER1 deletion mutant. This suggested that the low production of gladiolin derivatives in both *gbnD1* ER1 deleted and mutated mutants was the result of a restricted substrate preference of the downstream assembly line.

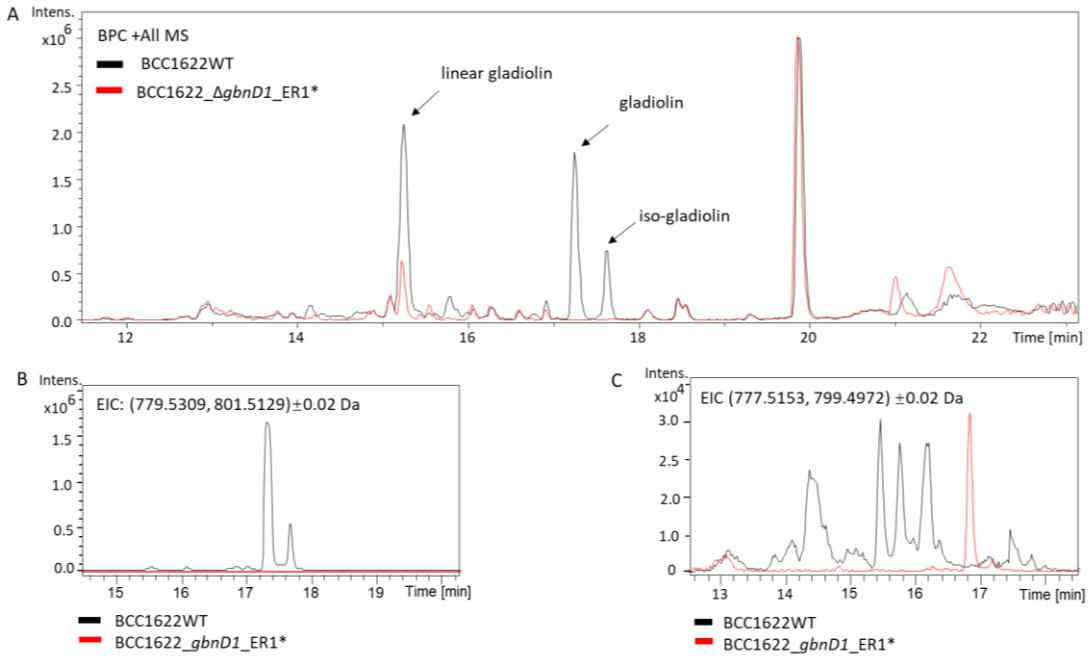


Figure 2.12 LC-MS analysis of metabolite extract from BCC1622 wild type and BCC1622_*gbnD1*_ER1*(GG388-399SP). All chromatograms follow the layout of: wild type indicated in black and BCC1622_*gbnD1*_ER1*(G388S-G389P) indicated in red. **(A)** Base peak chromatograms. **(B**) Extracted ion chromatograms at *m/z*=779.5309±0.02 Da and 801.5129±0.02 Da corresponding to the $[M+H]^+$ and $[M+Na]^+$ ion of gladiolin **7** and iso-gladiolin **28**, **C**) Extracted ion chromatogram at *m/z*= 777.5153±0.02 Da and 799.4972±0.02 Da corresponding to $[M+H]^+$ and $[M+Na]^+$ ions of unsaturated gladiolin **33** and **34** in WT and **32** in BCC1622_*gbnD1*_ER1*mutant. See appendix for mass spectra of the EICs.

Although no structural elucidation of the new metabolite (16.9 min) was possible, its presence in both the *gbnD1* ER1 deletion and site-directed mutation mutants but not in WT and single crossover indicated the structure of this new metabolite as **32**. It thus suggested the GbnD1 ER1 domain has *cis*-acting enoyl reduction activity in module 1.

2.1.2 *In vitro* characterization of the GbnD1 ER1 domain

In an effort to complement the genetic data shown previously, detailed *in vitro* biochemical assays were also employed to provide additional information regarding the ability of the GbnD1 ER1 domain to conduct both *intra*- and *inter*-modular enoyl reductions. This hypothesis requires that the GbnD1 ER domain must interact with ACP domains from module 1 and module 5,

catalyzing reduction of the α, β-unsaturated intermediate in each instance.

In order to investigate this, the following protein constructs were designed: GbnD1 ACP1-ER1 didomain, isolated GbnD1 ER1 domain and isolated GbnD2 ACP5 domain. pET-151-GbnD1_ACP1-ER1, pET-151-GbnD1_ER1 and pET-151-GbnD2_ACP5 were created as N-terminal pHis$_6$ constructs, which were overexpressed in *E. coli* BL21 (DE3) and the corresponding N-pHis$_6$ recombinant proteins were purified as described in 6.5. The mass and purity of the purified recombinant protein were confirmed by SDS-PAGE and LC-MS analysis (Figure 2.13).

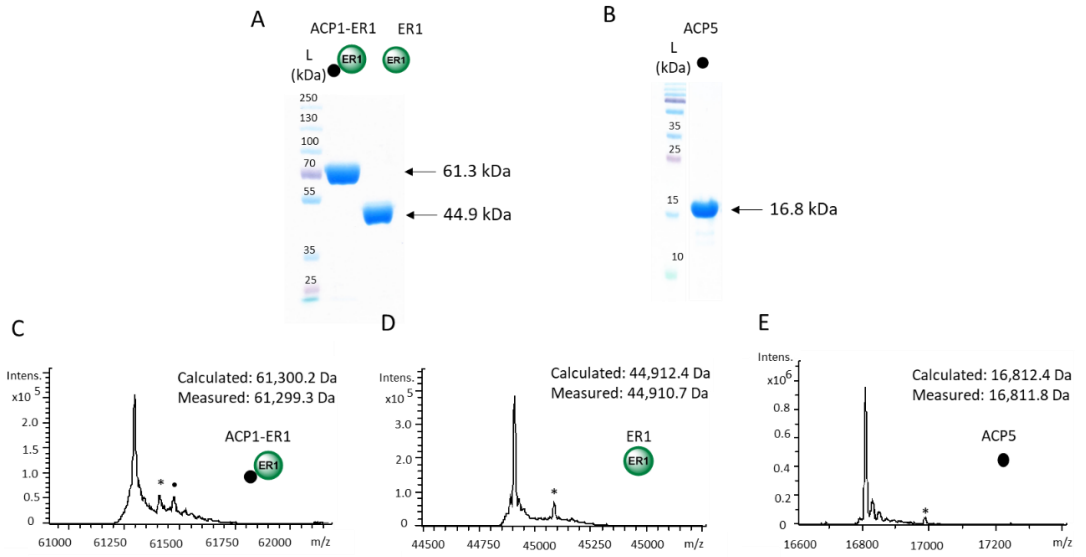
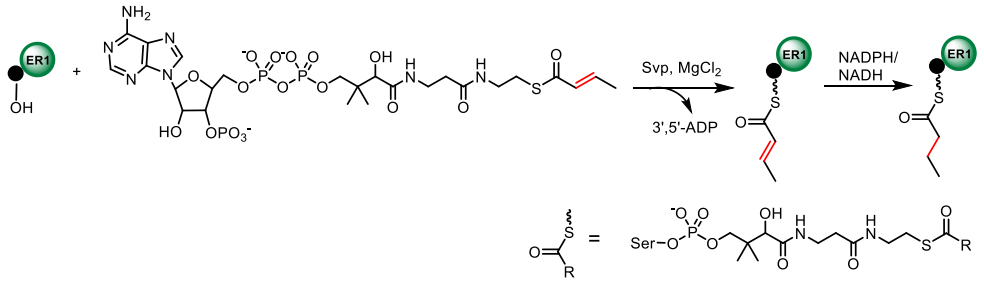


Figure 2.13 SDS-PAGE and LC-MS analysis of purified N-His$_6$-GbnD1 ACP1-ER1, N-His$_6$-GbnD1 ER1, N-His$_6$-GbnD2 ACP5. 10% SDS-PAGE gel showed the size and purity of N-His$_6$-GbnD1 ACP1-ER1 and N-His$_6$-GbnD1 ER1 **(A)**, 15% SDS-PAGE gel showed the size and purity of N-His$_6$-GbnD2 ACP5 **(B)**. Deconvoluted mass spectra of N-His$_6$-GbnD1 ACP1-ER1 **(C)**, N-His$_6$-GbnD1 ER1 **(D)** and N-His$_6$-GbnD2 ACP5 **(E)**. Peaks labeled as '*' refer to the known gluconoylation of His-tag fusion proteins with +178 Da being observed[147]. The second '●' peak in N-His$_6$-GbnD1 ER1 deconvoluted spectrum referred to phosphogluconoylation modification of His-tag with mass addition of 258 Da[147]. For See appendix for raw mass spectra.

2.1.2.1 Characterization of GbnD1 ER1 *intra*-modular enoyl reduction activity in module 1

Firstly, the ability of the GbnD1 ER1 domain to catalyse enoyl reduction of a 2-butenoyl substrate attached to the GbnD1 ACP1 domain was examined. An *in vitro* enoyl reduction assay (outlined in Scheme2.1) was designed, using the commercially available 2-butenoyl-CoA ester as a

substrate mimic of the intermediate in module 1. The 2-butenoyl-phosphopantethienyl (Pant) chain from the CoA ester was first loaded onto the ACP domain using the reported PPtase catalyzed loading reaction (described in 6.6.6.1). The successful attachment of the 2-butenoyl-PPant chain was monitored by LC-MS analysis of intact GbnD1 ACP1-ER1 before and after the loading reaction, which showed the expected 409 Da mass shift (Figure 2.14 A). The following reduction of the loaded 2-butenoyl unit by the GbnD1 ER1 domain was conducted by adding either NADH/NADPH as cofactor to the reaction. A control reaction was conducted with the same condition, without adding any cofactor.



Scheme 2.1 Design of *in vitro* assay for investigating the GbnD1 ER1 domain's *intra*-modular enoyl reduction activity. GbnD1 ACP1-ER1 didomain was first loaded with 2- butenoyl-PPant unit from substrate mimic crotonyl-CoA via PPtase, Svp catalyzed loading reaction, followed by addition of cofactor NADPH/NADH

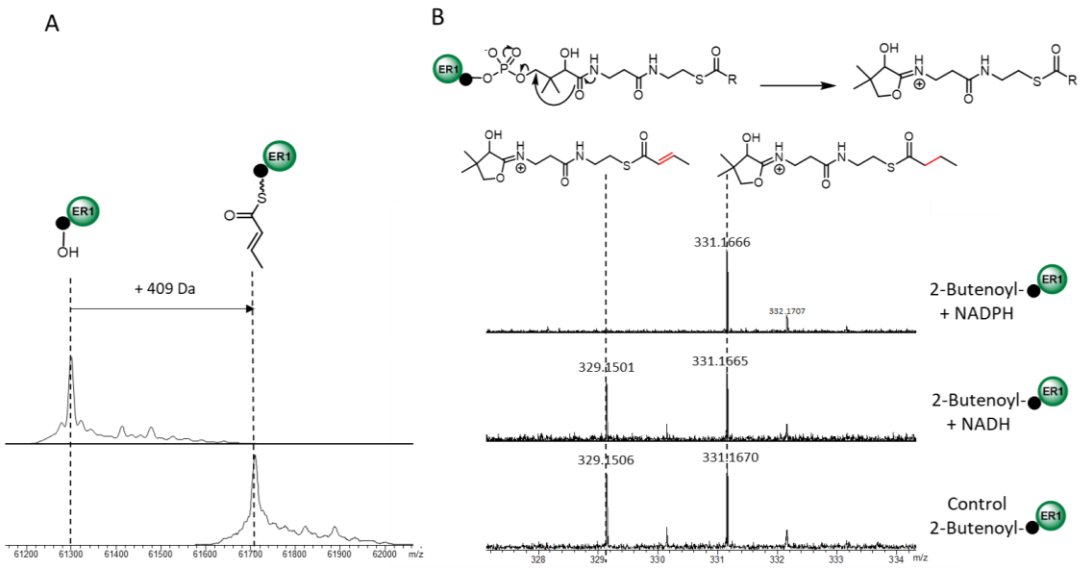


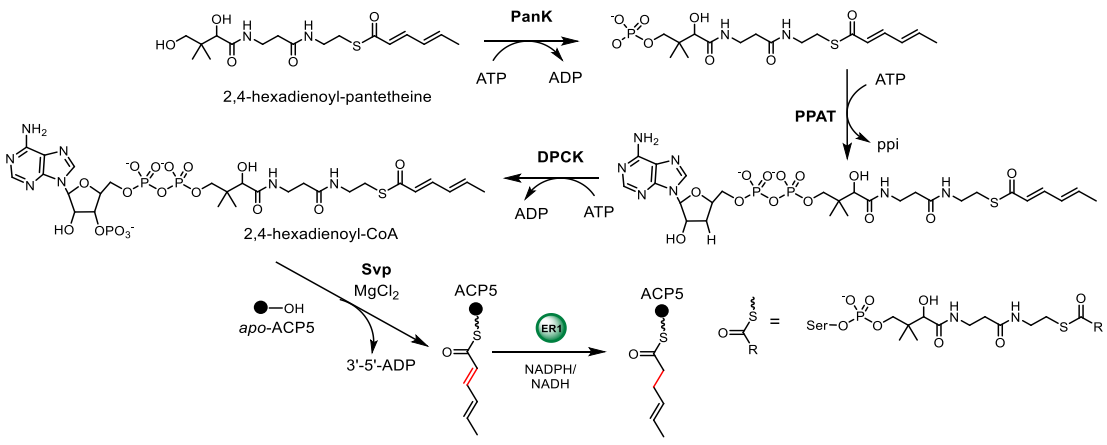
Figure 2.14 LC-MS analysis of *in vitro* GbnD1 ER1 domain *intra*-module enoyl reduction activity. (A) Deconvoluted mass spectra of *apo*-GbnD1 ACP1-ER1 (top) and following incubation with 2-butenoyl-CoA Svp and $MgCl_2$ (bottom). **(B)** Mass spectra of PPant ejected ions activated from 2-butenoyl-GbnD1 ACP1-ER1 following

incubating with NADPH (top), NADH (middle) or with no co-factor (bottom).

As the mass shift from the reduced product to the substrate was only 2 Da, and therefore difficult to detect by intact protein MS, the enoyl reduction activity was instead monitored using the PPant ejection procedure[148]. This involves collision-induced activation of the phosphopantetheinylated proteins within the mass spectrometer to liberate the pantetheine (Pant) moiety from the ACP as a small molecule, allowing small mass shifts resulting from acyl chain modifications to be examined (Figure 2.14 B). Interestingly, the control reaction (lacking the required NADPH/NADH cofactor) resulted in the presence of both 2-butenoyl-PPant and butyryl-PPant ions at a ratio of around 1:1. This was probably due to GbnD1 ACP1-ER1 co-purifying with the cofactor bound. However, complete conversion from 2-butenoyl-GbnD1 ACP1-ER1 to butyryl-GbnD1 ACP1-ER1 was only observed following incubation with additional NADPH (Figure 2.14 B). There was only 5-10% conversion with NADH addition compared to the control reaction, indicating NADPH is the true cofactor of this ER domain. This result confirmed the GbnD1 ER1 domain's *intra*-modular enoyl reduction activity in module 1.

2.1.2.2 Probing the GbnD1 ER1 *inter*-modular enoyl reduction activity in module 5

To investigate the proposed *inter*-modular activity of the GbnD1 ER1 domain with the ACP domain of module 5, a 2,4-hexadienoyl pantetheine (Pant) (synthesized by PhD student Christian Hobson), was used as a closer substrate mimic of the intermediate generated by module 5 to perform the designed *in vitro* assay shown in Scheme 2.2.



Scheme 2.2 Design of *in vitro* assay for investigating the putative *inter*-modular enoyl reduction activity of the GbnD1 ER1 domain. The GbnD2 ACP5 domain was first loaded with 2,4-hexadienoyl-PPant unit using the corresponding pantetheine form substrate mimic and the phosphopantetheinylation enzyme cassette catalyzed loading reaction followed by addition of GbnD1 ER and cofactor NADPH.

Loading of the 2,4-hexadienoyl PPant unit onto the GbnD2 ACP5 domain was achieved using the phosphopantetheinylation enzyme cassette (detailed in 6.6.1.2). The resulting 2,4-hexadienoyl-GbnD2 ACP5 was confirmed by a +435 Da mass shift on from the *apo*-GbnD2 ACP5 when monitored by LC-MS (Figure 2.15 A). The 2,4-hexadienoyl-GbnD2 ACP5 was then incubated with the GbnD1 ER1 domain and additional NADPH using the same conditions used in 2.1.2.1 (detailed in 6.2.2.2). A control reaction was conducted using the same conditions, but lacking GbnD1 ER1. Only 2,4-hexadienoyl-PPant ion could be detected in both of the control and experimental reaction (Figure 2.15 B) showing there was no ER-catalyzed reduction. This suggested that GbnD1 ER1 does not interact with GbnD2 ACP5 in an *inter*-modular fashion.

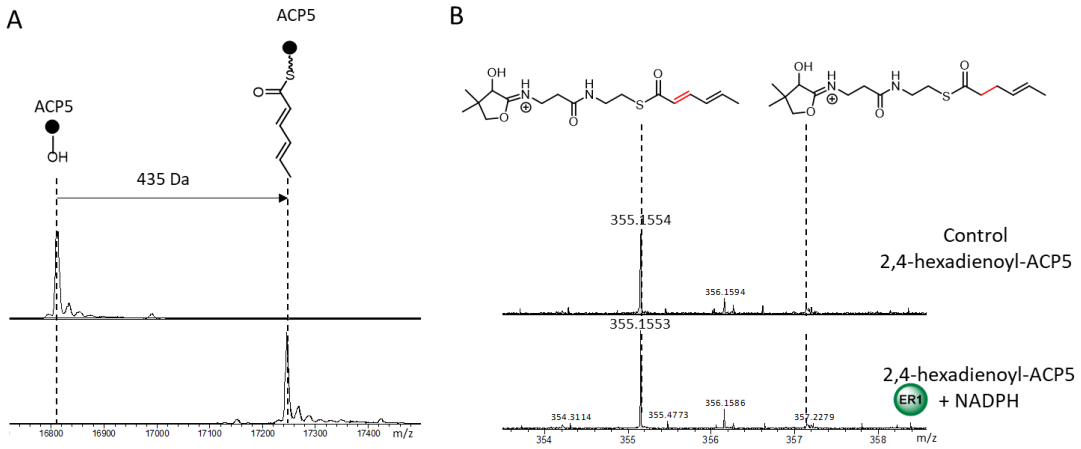
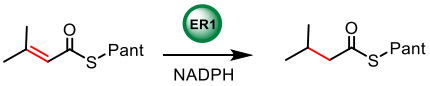


Figure 2.15 LC-MS analysis of *in vitro* GbnD1 ER1 domain *inter*-modular enoyl reduction activity in module 5. **(A)** Deconvoluted mass spectra of *apo*-GbnD2 ACP5 (top) and following incubation with 2,4-hexadienoyl-Pant and loading enzymes cassette in the presence of $MgCl_2$ and ATP (bottom). **(B)** Mass spectra of PPant ejected ion activated from 2,4-hexadienoyl-GbnD2 ACP5 incubated with GbnD1 ER1 and NADPH (bottom) or only with NADPH (top).

However, this negative result could also be caused by the isolated GbnD1 ER1 being purified in an inactive form. Therefore, in order to test this possibility, another *in vitro* assay (Scheme 2.3)

was conducted, in which the GbnD1 ER domain was incubated with 3-methyl-butenoyl-Pant (synthesized by Dr. Douglas Roberts) and monitored for reduction of the enoyl functionality (detailed in 6.2.2.2). LC-MS analysis of extracted reaction mixture showed 60-70% substrate was converted to 3-methyl-butyryl-Pant while there was no conversion observed in the control reaction when using denatured GbnD1 ER1 domain (Figure 2.16). This result confirmed the isolated GbnD1 ER1 domain was purified in active form.



Scheme 2.3 Design of *in vitro* assay for monitoring activity of isolated GbnD1 ER1 domain. The isolated GbnD1 ER1 domain was incubated with 3-methyl-butenoyl-Pant and cofactor NADPH.

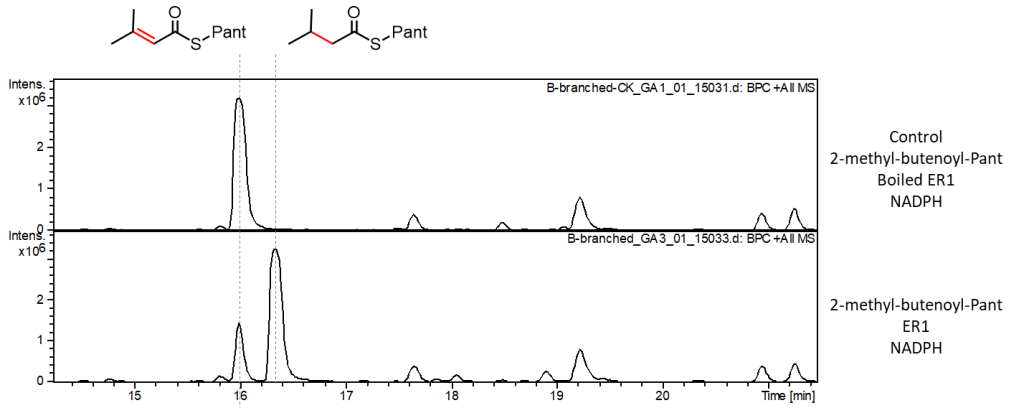


Figure 2.16 LC-MS analysis of *in vitro* assay demonstrating enoyl reduction activity of the isolated GbnD1 ER1 domain. Base peak chromatogram from LC-MS analysis of extracted reaction mixture of the GbnD1 ER1 domain incubated with 3-methyl-butenoyl-Pant and NADPH (bottom) and control group using denatured GbnD1 ER1 (top).

All the above results in this section provided key evidence to support the idea that GbnD1 ER1 domain is unable to engage in productive interactions with the ACP domain from module 5, suggesting that the GbnD1 ER1 domain is not able to catalyze the proposed '*inter*-module' reduction in module 5.

2.2 Characterization of GbnE

The findings detailed in Section 2.1 essentially ruled out the ability of the GbnD1 ER1 domain to catalyse enoyl reduction activity on a module 5 intermediate in an *inter*-modular manner. This suggested that the putative *trans*-acting enoyl reductase, GbnE, is a more likely candidate to perform enoyl reduction in both module 10 and module 5. To probe this dual functional activity of GbnE as well as the potential effect of GbnE to the non-iterative usage of module 5, GbnE was first characterized via *in vivo* approach.

2.2.1 *In vivo* characterization of GbnE

2.2.1.1 *In-frame* deletion of *gbnE*

Since GbnE is encoded by a stand along gene, *in-frame* deletion was still considered as useful approach to investigate its function in the biosynthesis of gladiolin. Utilizing the same methodology as employed for deletion of *gbnD1*_ER1, a deletion construct, pGPI-*gbnE,* was first generated and mobilized into *B. gladioli* BCC1622 to obtain *B. gladioli* BCC1622_Δ*gbnE* (Figure 2.17 A). Successful deletion of *gbnE* was confirmed by PCR and sequencing (Figure 2.17 B).

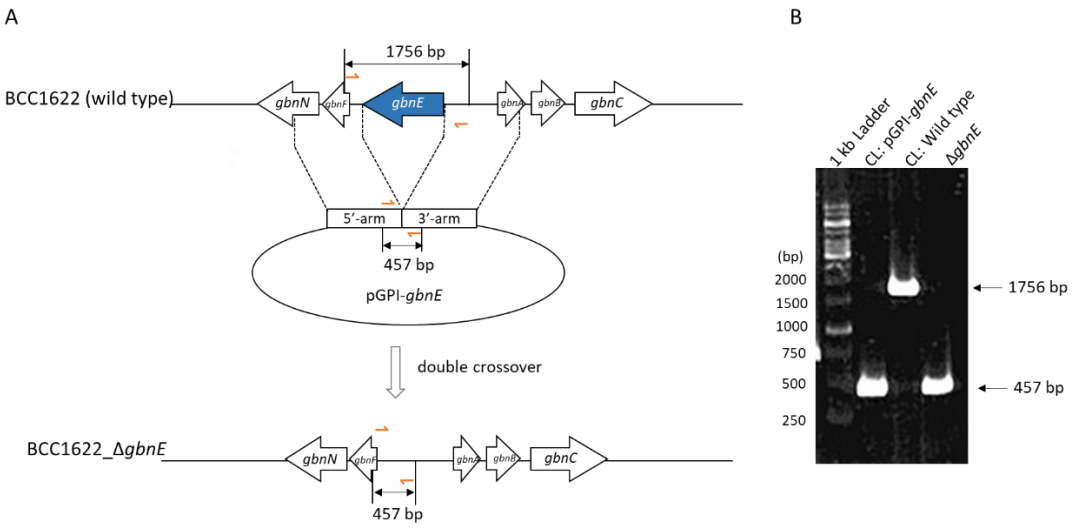


Figure 2.17 Schematic representation of generation and PCR confirmation of *B. gladioli* BCC1622_Δ*gbnE*. (A) Targeted *in-frame* deletion region on WT *B. gladioli* BCC1622 chromosome is highlighted as a blue. Loss of the blue region on the mutant chromosome after double crossover denotes successful deletion. Checking PCR

primers are indicated by orange arrows and their binding sites on the mutant chromosome, WT chromosome and deletion plasmid are indicated together with the length of corresponding predicted PCR products. **(B)** DNA electrophoresis gel of the PCR products confirmed the genome type of the *gbnE* deletion mutant, resulting in a 457 bp product as observed in the positive control (deletion plasmid), and the negative control (WT allele) resulted in a 1756 bp product. 'CL' refers to control, 'WT' refers to *B. gladioli* BCC1622 wild type.

Comparative metabolite profiling of *B. gladioli* BCC1622_Δ *gbnE* and WT *B. gladioli* BCC1622 was conducted by LC-MS analysis. Interestingly, the *B. gladioli* BCC1622_Δ *gbnE* metabolite extract (Figure 2.18 A and B) showed that although the gladiolin production significantly reduced, it was maintained at low level, around 5-10% of wild type production. This seemed to suggest GbnE is intimately involved in gladiolin biosynthesis, and could be partially complemented by homologous enzyme(s) in the cell.

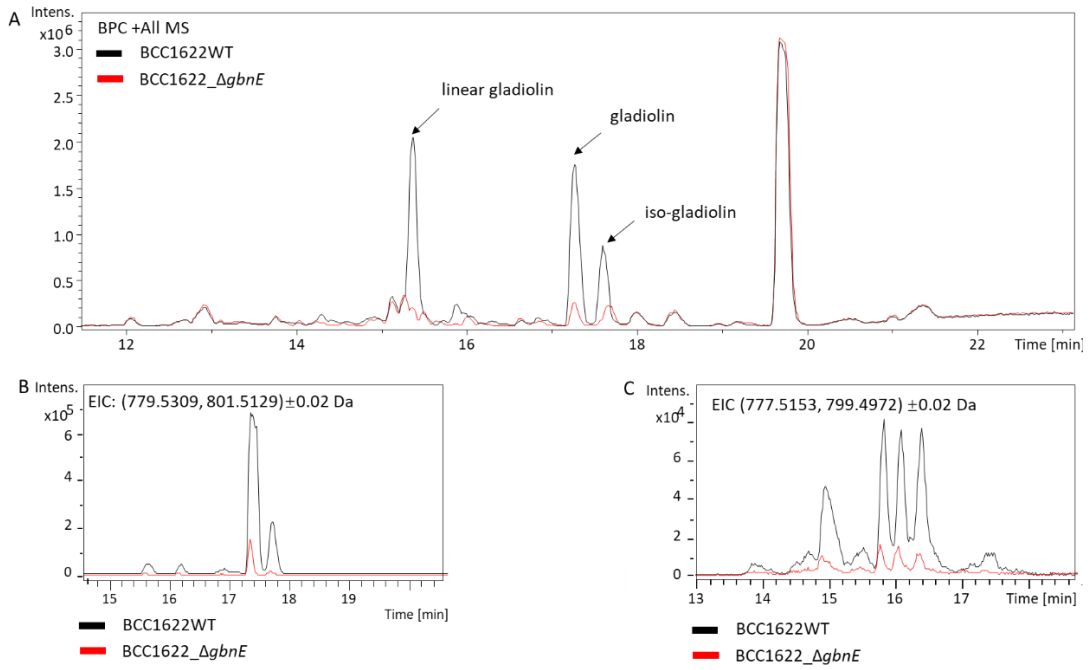


Figure 2.18 LC-MS analysis of metabolite extracts from *B. gladioli* BCC1622 wild type and BCC1622_Δ*gbnE*. All chromatograms follow the layout of: WT *B. gladioli* BCC1622 indicated in black and *B. gladioli* BCC1622_Δ*gbnE* indicated in red. **(A)** Base peak chromatograms. **(B**) Extracted ion chromatograms at *m/z*=779.5309 ± 0.02 Da and 801.5129 ± 0.02 Da corresponding to the $[M+H]^+$ and $[M+Na]^+$ ions of gladiolin **7** and iso-gladiolin **28**. **C**) Extracted ion chromatogram at *m/z* = 777.5153 ± 0.02 Da and 799.4972 ± 0.02 Da corresponding to $[M+H]^+$ and $[M+Na]^+$ ions of unsaturated gladiolin **33** and **34**. See appendix for mass spectra of the EICs.

Possible derivatives produced by *gbnE* deletion mutant were proposed (Figure 2.15) based on three possible scenarios:1) GbnE domain is of dual function, performing enoyl reduction in both module 5 and module 10, and the activity of GbnE in module 5 halts the inherent iterative nature of this module, thus inactivation of GbnE would resulted in generation of structure **35**; 2) GbnE is of dual function but GbnD2 Module 5 has no underlying ability to perform iteration, thus inactivation of GbnE would generate structure **36**; 3) Like its counterpart EntL in etnangien biosynthesis, GbnE only reduces the α, β-unsaturated intermediate arising from GbnD4 module 10 and has no effect on module 5, thus inactivation of GbnE would generate structure **34**.

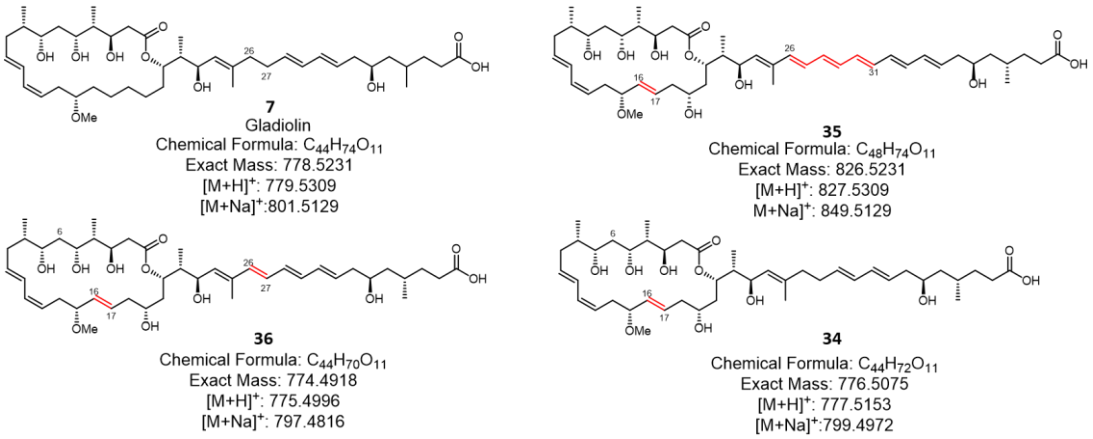


Figure 2.19 Structural comparison of gladiolin and proposed derivatives produced by the *gbnE* deletion mutant. Structural differences from gladiolin are highlighted in red, and m/z values for $[M+H]^+$ and $[M+Na]^+$ are shown. The proposed derivatives are comprised of the structural differences resulting from a combination of possible consequences of *gbnE* deletion. The desaturation of C16-C17 (in all three derivatives) originate from elimination of GbnE activity in module 10. The desaturation of C26-C27 (in **35** and **36**) originate from elimination of GbnE activity in module 5. The installation of a polyene region at C26-C31 (in **35**) is the result of restoring the iteration of module 5 by removing GbnE function towards this module.

To probe whether derivatives **34**-**36** (Figure 2.19) were produced after deletion of *gbnE*, the ions corresponding to their molecular weight were searched for in metabolite extracts of *B. gladioli* BCC1622_ Δ*gbn*E. Neither **35** or **36** was observed (data not shown) and **34** (and **33**) was detected at significantly lower level than in the wild type (Figure 2.18 C).

The decreased production of gladiolin and related derivatives in the *B. gladioli* BCC162_Δ*gbnE* mutant strain suggested a general slowing down of the PKS assembly line in the absence of GbnE,

which further indicated the crucial role of the enoyl reduction events performed by GbnE to the assembly line. This could be because the reduced intermediates are preferred by the downstream assembly line or the assembly line requires the *trans*-acting protein-protein interaction from GbnE for efficient processing.

It is, however, possible that the polar effect introduced by deletion of *gbnE* was responsible for the overall low production of gladiolin related products. In order to rule out this possibility and to thus validate the crucial function of GbnE to the assembly line, *gbnE* complementation to *B. gladioli* BCC1622_Δ*gbnE* was performed.

2.2.1.2 Complementation of *gbnE*

Firstly, *gbnE* was cloned onto an expression vector pMLBAD (detailed in 6.3.3), which was reported to be used for performing successful regulated gene expression under the control of an arabinose-inducible P_{BAD} promoter in *Burkholderia cepacian* species[149]. The confirmed construct pMLBAD-*gbnE* was then mobilized into BCC1622_Δ*gbnE* by tri-parental mating (detailed in 6.3.4). Mutant BCC1622_Δ*gbnE*::*gbnE* was screened by antibiotic selection and confirmed by PCR (detailed in 6.3.3) (Figure 2.20).

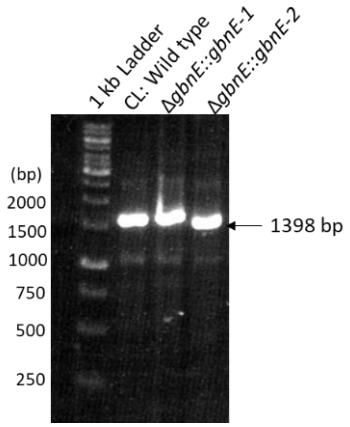


Figure 2.20 PCR confirmation of *B. gladioli* BCC1622_Δ*gbnE*::*gbnE*. DNA electrophoresis gel of the PCR products confirmed the genome type of *gbnE* complemented mutant, resulting in a 1398 bp product like the positive control wild type 'CL' refers to control.

0.2%-2% of L-arabinose was reported to be the optimal final concentration used for the induction

of gene expression in *Burkholderia* species[149]. For comparison, 0.2% or 2% of L-arabinose was added to the medium for mutant metabolite production (detailed in 6.4.1). Comparative metabolite profiling between BCC1622 wild type, BCC1622_Δ*gbnE*, BCC1622_Δ*gbnE*::*gbnE* and empty vector control BCC1622_Δ*gbnE*::pMLBAD induced at different concentration of L-arabinose were performed by LC-MS analysis (Figure 2.21). Gladiolin EICs showed that the complemented mutant BCC1622_Δ*gbnE*::*gbnE* induced at 0.2% L-arabinose restored 90% gladiolin production. The empty vector control group BCC1622_Δ*gbnE*::pMLBAD induced at the same concentration of L-arabinose seemed to maintain the similar level of the gladiolin production as BCC1622_Δ*gbnE*. This confirms the successful complementation. Nevertheless, 2% L-arabinose was found to abolish gladiolin production in both the complemented mutant BCC1622_Δ*gbnE*::*gbnE* and control BCC1622_Δ*gbnE*::pMLBAD. Since the mutant with 2% L-arabinose added in production medium showed no growth defect, this was predicted to result from other competitive pathway(s) being switched on or greatly up regulated by high concentration of L-arabinose.

This result provided evidence that no polar effect was brought into the *gbnE* deletion mutant, and the decreased gladiolin production in *gbnE* deletion mutant related to the loss of GbnE's function in the assembly line.

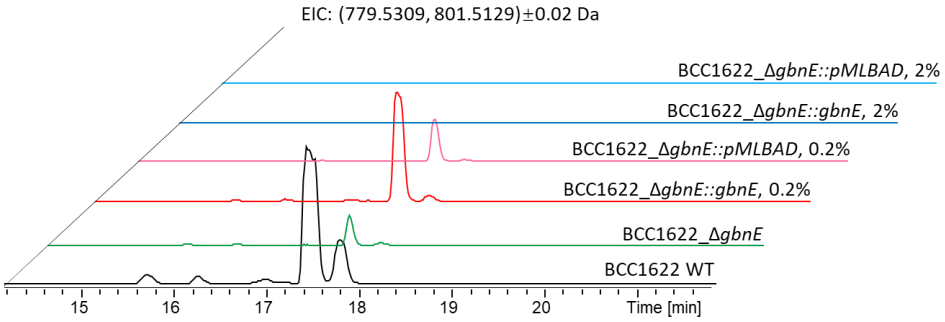


Figure 2.21 LC-MS analysis of gladiolin production in *B. gladioli* BCC1622 wild type, BCC1622_Δ*gbnE*, BCC1622_Δ*gbnE*::*gbnE* and BCC1622_Δ*gbnE*::pMLBAD. Extracted ion chromatograms at *m/z*=799,5309±0.02 Da, 801.5129±0.02 Da corresponding to $[M+H]^+$ and $[M+Na]^+$ ion of gladiolin (17.4 min) and iso-gladiolin (17.8 min). Percentage refers to the concentration of L-arabinose in media.

2.2.1.3 *In-frame* deletion of putative complementary *trans*-acting ER

In order to probe whether GbnE's function in the gladiolin biosynthesis could be complemented by homologous enzyme(s) in *B. gladioli* BCC1622, the genome of *B. gladioli* BCC1622 was searched for *gbnE* homologue(s). A standalone AT-AT-ER tri-domain, GdsB is located in a *trans*-AT PKS gene cluster and GdsB ER domain was found to have 57.7 % sequence identity to GbnE. The metabolic product of this products was revealed by the Challis group to be gladiosatin, a novel secondary metabolite. From the structure and the associated gene cluster, a biosynthetic pathway was proposed, in which several *trans*-acting enoyl reduction events are predicted to be required (data not published). It was thus considered as a very good candidate for complementing the function of GbnE in the gladiolin biosynthesis.

In an attempt to investigate the putative complementary function of the GdsB ER domain, as well as to probe whether iteration of module 5 could be restored when the enoyl reduction event in module 5 is completely removed, *in-frame* deletion of GdsB ER domain encoded DNA region was then performed in both *B. gladioli* BCC1622 WT and BCC1622_Δ*gbnE*. Deletion of *gdsB*_ER followed the same methodology used for deletion of *gbnD1* ER1 and *gbnE*. Single knock-out mutant *B. gladioli* BCC1622_Δ*gdsB*_ER and double knock-out mutant *B. gladioli* BCC1622_Δ*gbnE*Δ*gdsB*_ER were successfully obtained and confirmed by PCR (Figure 2.22 B). A weak band of size close to 1851 bp was noted in the double knockout mutant PCR products, which seemed to suggest a single crossover genotype as a PCR product of 606 bp was also observed. Attempts to amplify the SceI recognition region from the BCC1622_Δ*gbnE*Δ*gdsB*_ER genome, which could verify whether the genome has the pGPI-*gdsB*_ER integrated, resulted in no PCR products while the control group using pGPI-*gdsB*_ER as a template resulted in positive PCR products (data not shown). This confirmed the loss of pGPI-*gdsB*_ER and thus validated it was double crossover mutant.

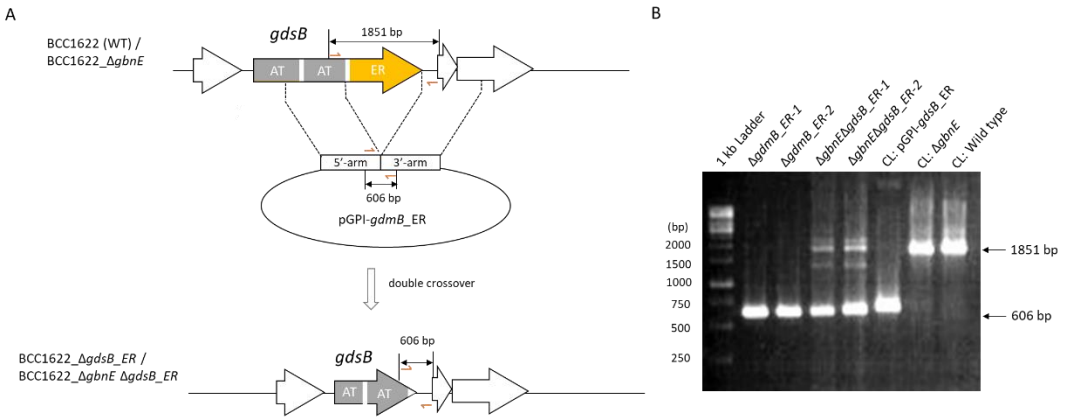


Figure 2.22 Schematic representation of generation and PCR confirmation of *B. gladioli* BCC1622_Δ*gdsB*_ER and BCC1622_Δ*gbnE*Δ*gdsB*_ER. (A) Targeted *in-frame* deletion region on wild type chromosome and BCC1622_Δ*gbnE* is highlighted in orange. Loss of the orange region on the mutant chromosome after double crossover denotes successful deletion. Checking PCR primers are indicated by orange arrows and their binding sites on the mutant chromosome, parental chromosome and deletion plasmid respectively are indicated together with the length of corresponding predicted PCR products. **(B)** DNA electrophoresis gel of the PCR products confirmed the genotype of *gdsB*_ER deletion mutants, resulting in a 606 bp product as observed in the positive control (deletion plasmid), and the negative control (WT and BCC1622_Δ*gbnE* allele) resulted in a 1851 bp product. 'CL' refers to control, 'WT' refers to *B. gladioli* BCC1622 wild type.

LC-MS analysis of the metabolites produced by both the single and double knock-out mutants showed that knocking out *gdsB* ER did not affect the production of gladiolin. Knocking out both *gdsB*_ER and *gbnE* seemed to promote gladiolin production compared to *B. gladioli BCC1622_ΔgbnE* (Figure 2.23). This indicated the GdsB ER domain was not the (only) homologue that could complement GbnE's function. The increased production of gladiolin could because it shares common precursors with gladiostatin, so obstruction of gladiostatin biosynthesis by deletion of GdsB ER domain shifts metabolite flux towards gladiolin production. These results suggested other enzymes must be encoded by the genome that can complement GbnE.

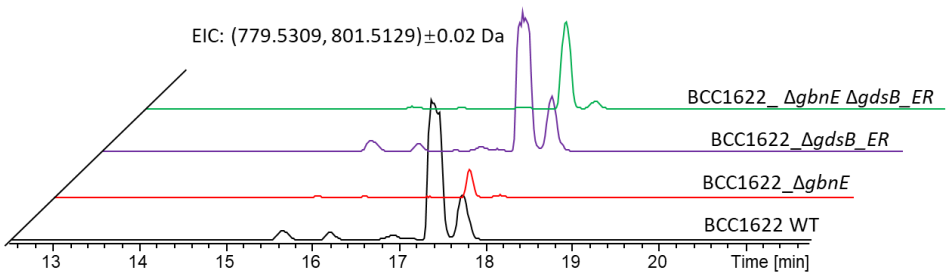


Figure 2.23 LC-MS analysis of gladiolin production in *B. gladioli* BCC1622 wild type, BCC1622_Δ*gbnE*, BCC1622_Δ*GdsB*_ER and BCC1622_Δ*gbnE*Δ*GdsB*_ER. Extracted ion chromatograms at *m/z*=799.5309±0.02 Da, 801.5129±0.02 Da corresponding to $[M+H]^+$ and $[M+Na]^+$ ion of gladiolin (17.4 min) and iso-gladiolin (17.8 min).

2.2.2 *In vitro* characterization of GbnE

To complement the *in vivo* data, the proposed activity of GbnE in module 5 and module 10 was further characterized via *in vitro* biochemical assays.

2.2.2.1 Enoyl reduction activity of GbnE

To demonstrate the FMN-dependent enoyl reduction activity of GbnE, pET151-based constructs for recombinant N-His_6-GbnE expression were first generated (described in 6.5.1). Unfortunately, overproduction and purification trials resulted in only trace amounts of recombinant N-His_6-GbnE; whilst most of the material remained in the insoluble fraction. Optimization of overproduction conditions, including lowering IPTG concentration, changing expression cell lines (such as *E. coli* C43 (DE3) for avoiding possible toxicity, *E. coli* Rosetta (DE3) for possible rare codon usage), did not significantly increase the soluble N-His_6-GbnE yield.

Glutathione S-transferase (GST) and small ubiquitin-related modifier (SUMO) tags are known to successfully promote the solubility of many fusion proteins. Therefore, constructs of pGEX-4T1-GbnE and pET-SUMO-GbnE were generated for to obtain N-GST-GbnE and N-His_6-SUMO-GbnE constructs (described in 6.5.1). Overproduction and purification trials showed that, although the N-GST tag did not significantly promote the solubility of GbnE, a dramatic increase in the amout of GbnE was purified for the N-His_6-SUMO-GbnE fusion protein (detailed in 6.5.3 and 6.5.4). It is was noted that the purified GbnE exhibited a strong yellow coloration, indicating GbnE was

purified with the flavin co-factor bound, consistent with its predicted FMN-dependent enoyl reduction activity. The purity and mass of the purified N-His_6-SUMO-GbnE were confirmed by SDS-PAGE analysis and LC-MS analysis (Figure 2.24).

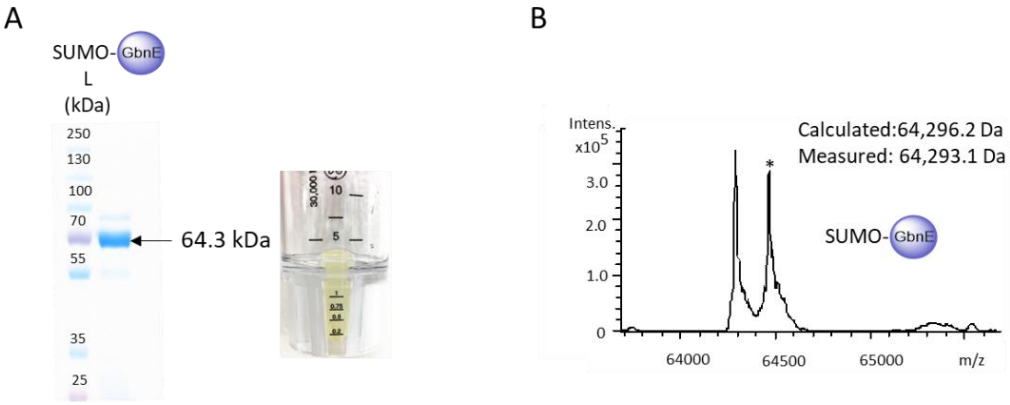
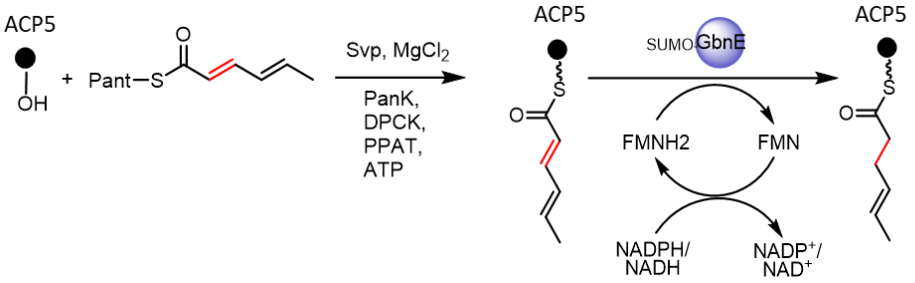


Figure 2.24 SDS-PAGE and LC-MS analysis of purified N-His_6-SUMO-GbnE. (A) 10% SDS-PAGE gel showing size and purity of N-His_6-SUMO-GbnE, the yellow color of the protein indicated incorporation of a flavin cofactor. **(B)** Deconvoluted mass spectrum confirming the exact mass of N-His_6-SUMO-GbnE (-N-Met). Peaks labeled as '*' refer to the spontaneous gluconoylation of His-tag on corresponding fusion proteins with additional 178 Da being observed[147]. See appendix for raw mass spectra.

To characterize the enoyl reduction capability of GbnE, an *in vitro* assay was designed as shown in Scheme 2.4, using the GbnD2 ACP5 domain loaded with 2,4-hexadienoyl PPant as the substrate mimic of intermediate in module 5. Similar to the assays for testing the GbnD1 ER1 domain activity (section 2.1.2.2), the GbnD2 ACP5 domain was first loaded with the 2,4-hexadienoyl PPant unit using the phosphopantetheinylation enzyme cassette. The enoyl reduction was then initiated by addition of GbnE and cofactor. Both NADPH and NADH were tested as the hydride donating cofactor for FMN. As GbnE was purified with flavin co-factor bound no additional $FMNH_2$ or FMN was added into the reaction. A negative control reaction was conducted using the same conditions, but lacking GbnE. The enoyl reduction activity was monitored by LC-MS analysis of the intact GbnD2 ACP5 domain.



Scheme 2.4 Design of *in vitro* assay for examination of enoyl reduction activity of GbnE. The GbnD2 ACP5 domain was first loaded with 2,4-hexadienoyl PPant unit using the corresponding pantetheine form substrate mimic and phosphopantetheinylation enzymes cassette, followed by addition of GbnE and cofactor NAD(P)H.

By comparison of the ejected PPant ions activated from GbnD2 ACP5-bound species (Figure 2.25), a ratio of 95:5 of the 4-hexenoyl-PPant ion: 2,4-hexadienoyl-PPant ion was observed in both reaction groups with NADPH or NADH added. In contrast, only the 2,4-hexadienoyl-PPant ion could be detected in the control reaction. This result showed the FMN-NAD(P)H-dependent enoyl reduction activity of GbnE and also provided evidence of *trans*-acting activity towards the ACP domain of module 5.

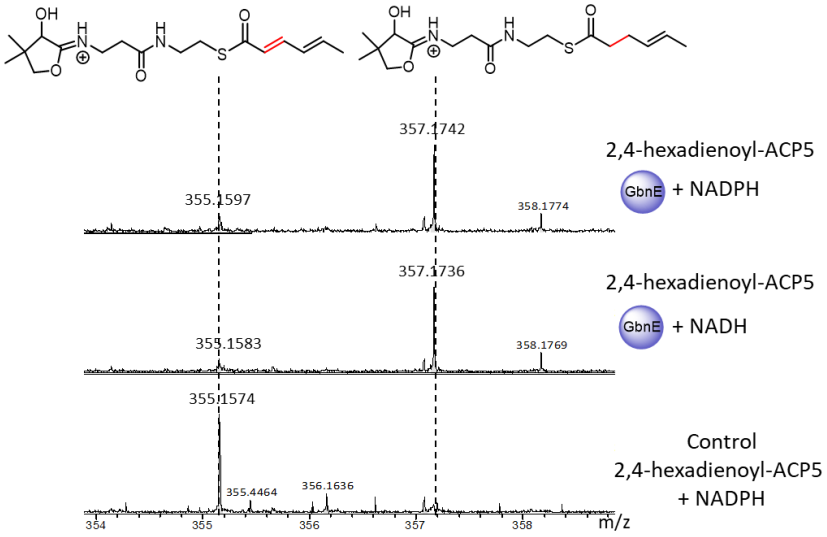


Figure 2.25 LC-MS analysis of in vitro enoyl reduction activity of GbnE. Mass spectra of PPant ejected ion activated from 2,4-hexadienoyl-ACP5 incubated with (from top to bottom) GbnE and NADPH, GbnE and NADH, only NADPH.

2.2.2.2 GbnE exhibits specific interactions with the gladiolin PKS

Having established the *trans*-acting activity of GbnE towards the ACP domain in module 5, the specific interaction between GbnE and the assembly line was then interrogated. In addition to GbnD2 ACP5 domain, GbnE is hypothesized to interact with ACP domain in module 10 (Figure 2.2). As α, β-double bonds of the intermediate formed in module 3, 4 and 12 are retain in the final product, the ACP domains from these modules should not interact with GbnE. To examine these specific interactions, module 3, 4,10 and 12 ACP domains, were selected to be overproduced as recombinant proteins for enoyl reduction assays with GbnE. The corresponding pET28a (+) based constructs for all ACPs were generated and overexpressed in *E. coli* BL21 (DE3). All ACP domains, except for the GbnD1 ACP4 domain with no overproduction being observed, were purified as N-His_6 fusion protein successfully (procedures detailed in 6.5.3 and 6.5.4, plasmid construction and purification of the GbnD5 ACP12 domain were conducted by Dr. Matthew Jenner). The purity and mass of the purified ACP domains was confirmed by SDS-PAGE and LC-MS analysis (Figure 2.26).

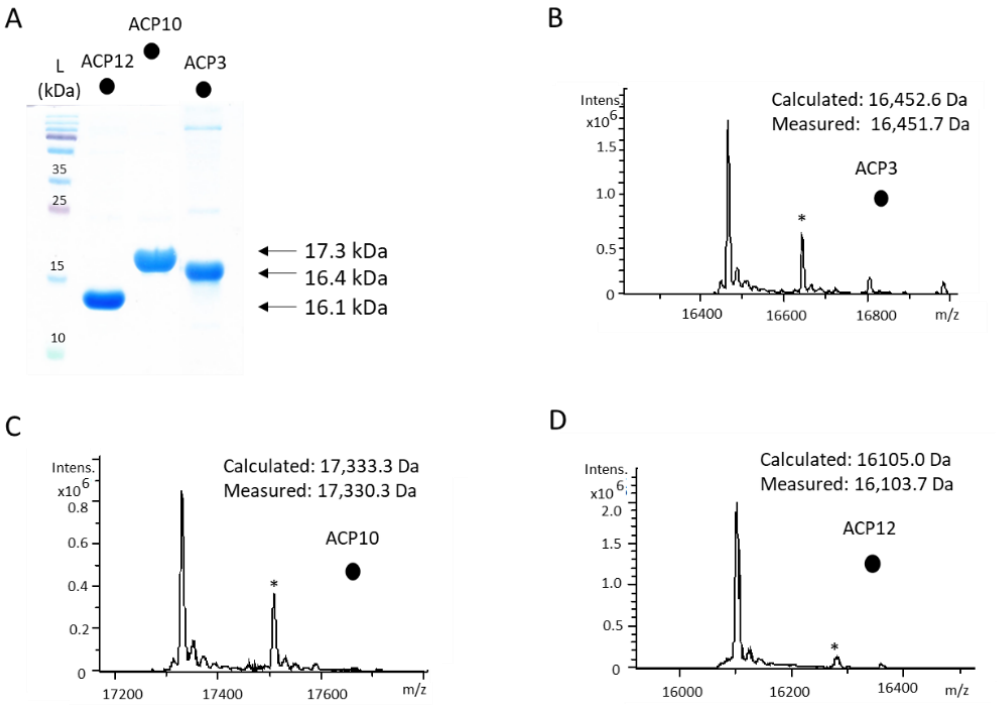


Figure 2.26 SDS-PAGE and LC-MS analysis of N-His_6-GbnD1 ACP3, N-His_6-GbnD4 ACP10 and N-His_6-GbnD5 ACP12. (A) 15% SDS-PAGE gel showing the size and purity of N-His_6-GbnD1 ACP3, N-His_6-GbnD4 ACP10 and N-His_6-GbnD5 ACP12. Deconvoluted mass spectra confirming the exact mass of **(B)** N-His_6-GbnD1 ACP3(-N-Met), **(C)** N-His_6-GbnD4 ACP10 (-N-Met) and **(D)** N-His_6-GbnD5 ACP12(-N-Met). Peaks labeled as '*' refer to the

spontaneous gluconoylation of His-tag on corresponding fusion proteins with additional 178 Da being observed[147]. See appendix for raw Mass spectra.

All ACPs were then subjected for the same *in vitro* enoyl reduction assays with GbnE (detailed in 6.6.3) (scheme 2.4). Successful loading of 2,4-hexadienoyl-PPant unit onto all ACP domains were first confirmed by LC-MS analysis (Figure 2.27). The enoyl reduction assays were initiated by addition of GbnE and NADPH. GbnE was omitted from all control reactions. The reduced product, 4-hexenoyl-PPant ion, could only be detected when the substrate was loaded on the GbnD2 ACP5 and GbnD4 ACP10 domain, with only 2-5% 2,4-hexadienoyl PPant ion observed (Figure 2.28 left). Meanwhile no 4-hexenoyl-PPant ion was found in all control groups (Figure 2.28 right). These data provided evidence that only the GbnD2 ACP5 and GbnD4 ACP10 domain can effectively deliver the enoyl substrate to GbnE for reduction, whereas the GbnD1 ACP3 and GbnD5 ACP12 domain cannot. It thus confirmed the specific interactions between GbnE and ACP domains from module 5 and 10 as proposed.

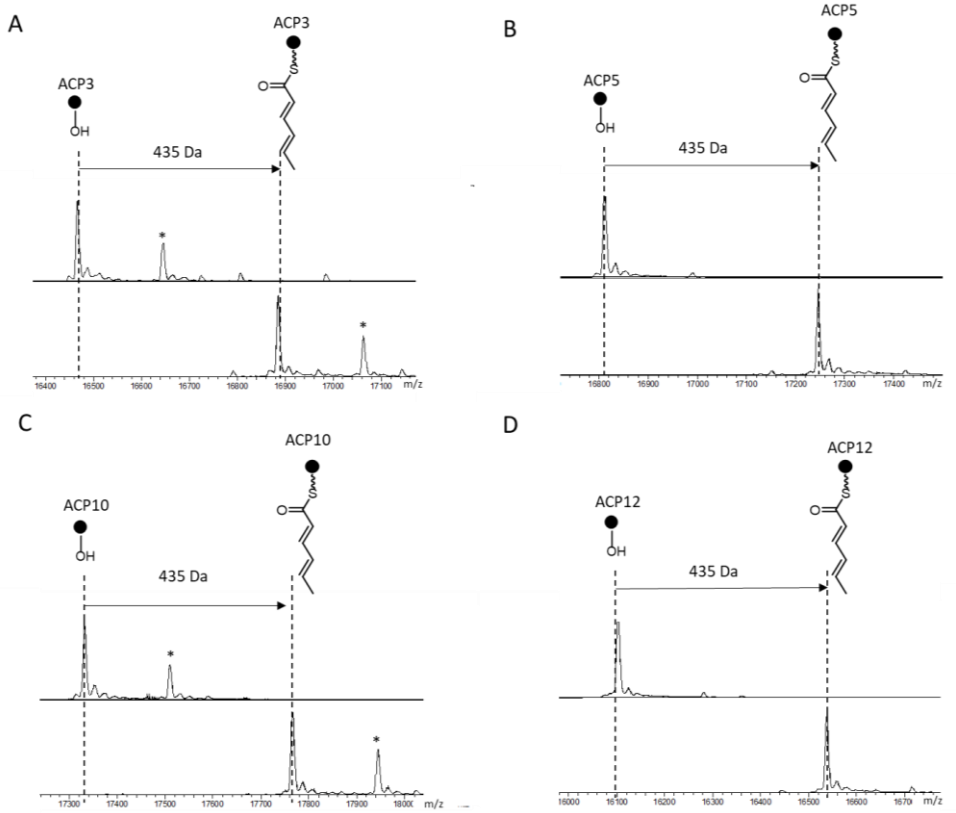


Figure 2.27 LC-MS analysis of 2,4-hexadienoyl-PPant loading onto the GbnD1 ACP3 (A), GbnD2 ACP5 (B), GbnD4 ACP10 (C) and GbnD5 ACP12 domain(D). Deconvoluted mass spectra of apo-ACP (top) and following incubation

with 2,4-hexadienoyl-Pant and loading enzymes cassette in the presence of $MgCl_2$ and ATP (bottom).

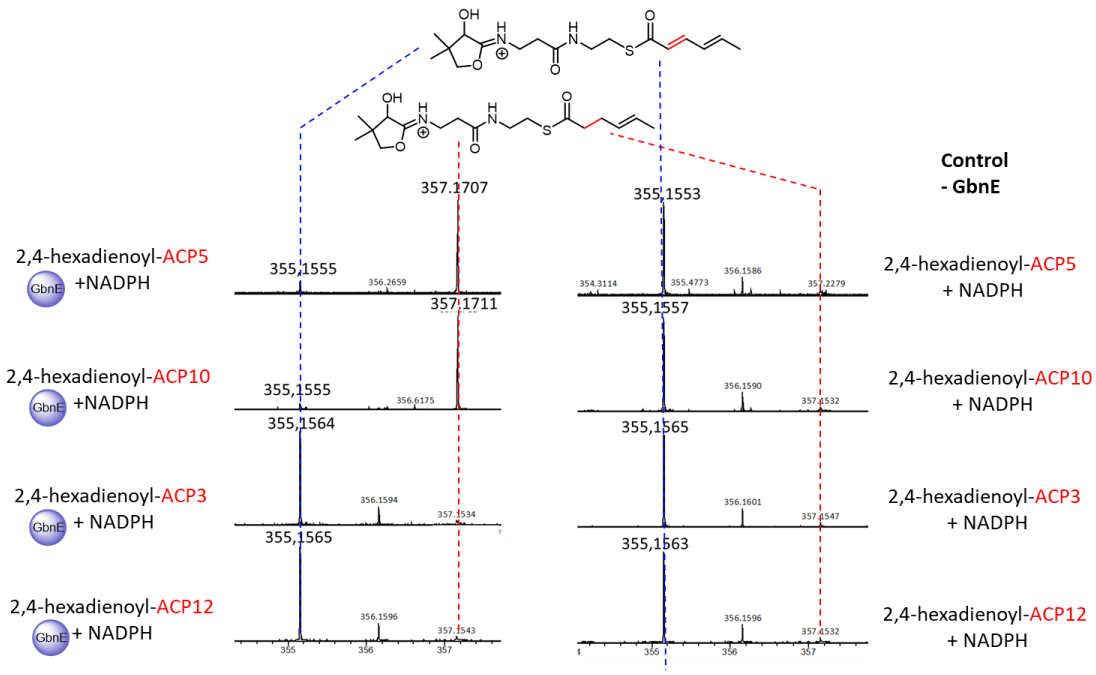


Figure 2.28 LC-MS analysis of GbnE *in vitro* enoyl reduction with GbnD2 ACP5, GbnD4 ACP10, GbnD1 ACP3, and GbnD5 ACP12 domain. Mass spectra of PPant ejected ion activated from 2,4-hexadienoyl-ACPs incubated with both GbnE and NADPH (left), or with only NADPH (right). Spectra from top to bottom of both groups were from assays conducted with the GbnD2 ACP5, GbnD4 ACP10, GbnD1 ACP3 and GbnD5 ACP12 domain respectively.

2.2.2.3 Probing catalytic mechanism of GbnE

As discussed in section 1.2.2.5 '*trans*-acting enoyl reductases', the catalytic mechanism of *trans*-acting ERs from *trans*-AT PKS remains unclear and requires further investigation. However, the structure and catalytic mechanism of their phylogenetically and structurally-related homologue FabK, the FMN-dependent enoyl-ACP reductase involved in the bacterial fatty acid biosynthesis, has been well studied[150,151,152]. The enoyl reduction catalyzed by FabK proceeds via a Bi-Bi double displacement mechanism, with the active site first bond with NAD(P)H cofactor to reduce the FMN group, followed by the resulting $NAD(P)^+$ being displaced by the enoyl substrate, which is then reduced by the $FMNH_2$ group, thereby returning the enzyme to its original state[151,152]. The X-crystal strucutre of FabK from *Streptococcus pneumoniae* suggested that a histidine residue (His144) (Figure 2.29), which is located in the active site and also near the FMN cofactor, is the key catalytic residue[150]. This was also supported by the observation that the side chain of His144

was found to be involved in a conformational change with and without substrate bound, which is consistent with the Bi-Bi double displacement mechanism.

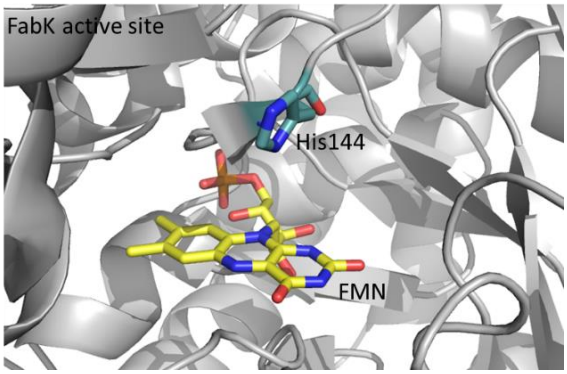


Figure 2.29 Structure of active site in FabK from *Streptococcus pneumoniae*. The X-crystal structure of FabK shows His144 located in the active site to the FMN cofactor. Generated using Pymol from PDB: 2Z6I[150].

In order to probe the catalytic mechanism of GbnE, a site-directed mutagenesis study for identification of the key catalytic residue(s) of GbnE was also pursued. Multiple sequence alignment between GbnE, other homologous *trans*-acting ERs from *trans*-AT PKSs, the PUFA biosynthesis ER PfaD and enoyl-ACP reductase FabK were performed (see appendix for full sequence alignment results). The proposed catalytic histidine residue (His144 in the FabK) was found to be conserved in the GGHTD motif in most of the sequences, and was located at position His198 in GbnE (Figure 2.30).

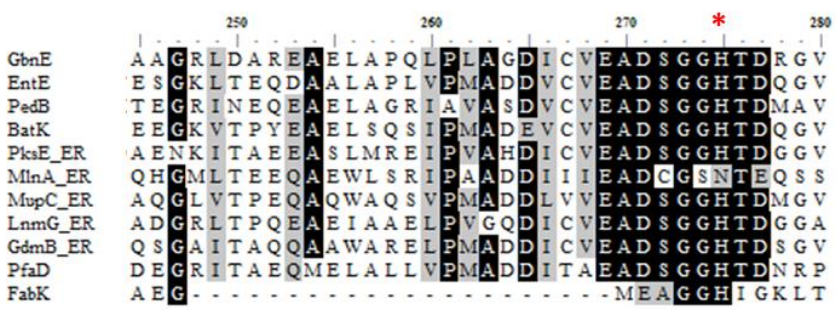


Figure 2.30 Partial result of multiple sequences alignment between GbnE and homologous enzymes. The conserved Histidine in the GGHTD motif is highlighted with a red star. Full sequence alignment result is shown in the appendix.

A mutant construct pETSUMO-GbnE(H198V) was generated by site-directed mutagenesis (detailed in 6.5.2), and GbnE(H198V) was successfully overproduced and purified as a soluble N-

His_6-SUMO fusion protein through the same process as for the wild type GbnE (detailed in 6.5.3 and 6.5.4). Notably, the solubility and yellow color of the mutant GbnE suggested it was correctly folded, and structurally unaffected by the H198V mutation. The purity and mass of the N-His_6-SOMO-GbnE-H198V were then confirmed by SDS-PAGE analysis and LC-MS analysis (Figure 2.31).

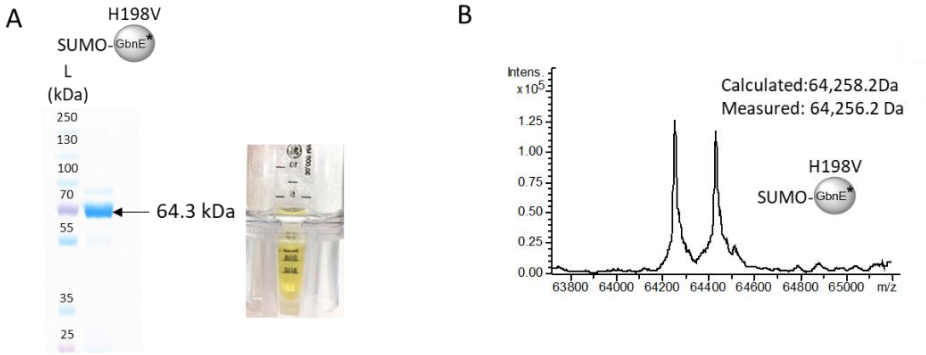


Figure 2.31 LC-MS analysis of purified N-His_6-SUMO-GbnE, N-His_6-SUMO-GbnE(H198V). Deconvoluted spectrum confirming the exact mass of (A) N-His_6-SUMO-GbnE (-N-Met), and (B) N-His_6-SUMO-GbnE(H198V) (-methionine). Peaks labeled as '*' refer to the spontaneous gluconoylation of His-tag on corresponding fusion proteins with additional 178 Da being observed[147]. See appendix for raw Mass spectra.

To evaluate the enoyl reduction activity of GbnE(198V), the same *in vitro* assay (Scheme 2.4) was performed with GbnE(H198V), together with WT protein as control. The LC-MS analysis of the GbnD2 ACP5 domain with 2,4-hexadinoyl PPant bound following incubation of GbnE(H198V) and cofactor showed (Figure 2.32) the enoyl reduction activity of GbnE was completely abolished in the mutant protein. This confirmed the critical catalytic role of His198 in GbnE catalysing enoyl reduction.

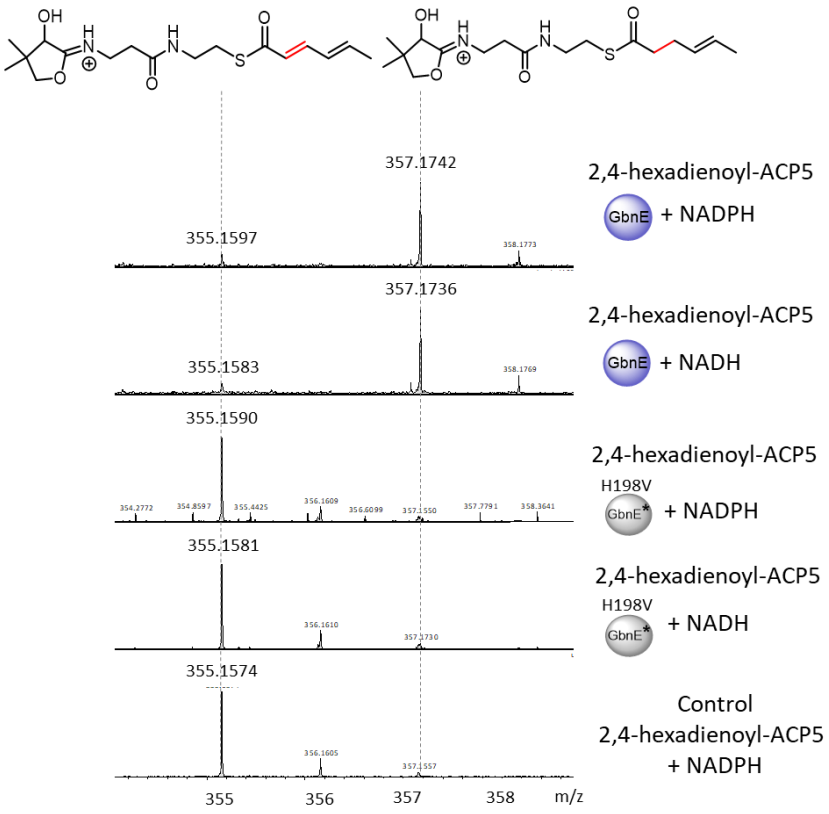


Figure 2.32 LC-MS analysis of *in vitro* enoyl reduction activity of GbnE (H198V). Mass spectra of PPant ejected ions activated from 2,4-hexadienoyl-ACP5 incubated with (from top to bottom) GbnE and NADPH, GbnE and NADH, GbnE(H198V) and NADPH, GbnE(H198V) and NADH, only NADPH.

2.3 Conclusions and future work

In this chapter, for identification of the ER required for module 5 of the gladiolin PKS, as well as to probe the potential impact of this enoyl reduction event to the non-iterative usage of this module, *in vivo* and *in vitro* approaches were combined to characterize the candidate ERs, *cis*-acting enoyl reductase GbnD1 ER1 domain and *trans*-acting enoyl reductase GbnE,

In-frame deletion and site-directed mutagenesis to inactivate the GbnD1 ER1 domain in the gladiolin PKS, followed by the metabolite profiling of the mutants, identified a new derivative to be the proposed unreduced gladiolin **32**. This suggests the GbnD1 ER1 domain has *intra*-modular enoyl reduction activity in module 1. This was further confirmed *in vitro* by incubating the purified recombinant GbnD1 ACP1-ER1 di-domain, loaded with a 2,4-hexadienoyl unit, with NADPH which resulted in successful reduction of the substrate. No *in vivo* data supported the

inter-modular enoyl reduction activity in module 5 of the GbnD1 ER1 domain, which was also confirmed by the *in vitro* assay that isolated GbnD1 ER1 domain was not capable of reducing the 2,4-hexadienoyl unit tethered on GbnD2 ACP5 domain.

In-frame deletion of *gbnE* and metabolite profiling indicated GbnE plays an important role in supporting efficient processing of the gladiolin assembly line as a decreased yield of gladiolin and related derivatives was observed when *gbnE* was absent. The FMN-dependent *trans*-acting enoyl reduction activity of GbnE was biochemically characterized. Incubation of GbnE with NADPH and four different ACP domains (GbnD1 ACP3, GbnD2 ACP5, GbnD4 ACP 10 and GbnD5 ACP12) showed GbnE is capable of catalysing reduction of substrates tethered on ACP domains from module 5 and 10 but not from module 3 and 12. This suggested the gladiolin assembly line recruits the *trans*-acting GbnE specifically in module 5 and module 10. Besides, the catalytic mechanism of *trans*-acting ERs was probed by site-directed mutagenesis and biochemical assays, which identified a conserved catalytic essential histidine residue (H198).

As *trans*-acting enoyl reduction is a widely employed strategy to install the saturated C-C bond in *trans*-AT polyketide biosynthesis, the molecular mechanism underlying the specific interaction between the ERs and the ACP domains is of great interest and could guide future engineering efforts. Bioinformatic and biophysical studies need to be employed to better understand the nature of this specific protein-protein interaction.

Chapter 3

Probing the molecular mechanism of chain length control in gladiolin biosynthesis

The work described in chapter 2 identifies a *trans*-acting ER, GbnE, as the catalytic entity responsible for installing a saturated center at C26-C27 in gladiolin biosynthesis (Figure 3.1), and accounts for one of the principle structural differences between etnangien and gladiolin. In place of a saturated C-C bond, etnangien possesses three double bonds, which are hypothesized to originate from iterative use of EtnE Module 5. Although the etnangien biosynthetic gene cluster also harbors a *trans*-acting ER, EtnL, it is possible that it does not engage in a productive interaction with the ACP domain of EtnE Module 5, and thus the module undergoes three rounds of chain extension, keto-reduction and dehydration to generate the triene moiety. However, given the high degree of similarity between the gladiolin and etnangien PKSs, it is possible that the equivalent module, GbnD2 Module 5, can also catalyse iterative chain elongation, ketoreduction and dehydration, which is prevented by the action of GbnE (Figure 3.1).

This details efforts to examine the iterative nature of GbnD2 Module 5, and the effect of GbnE on this process using *in vitro* biochemical methods.

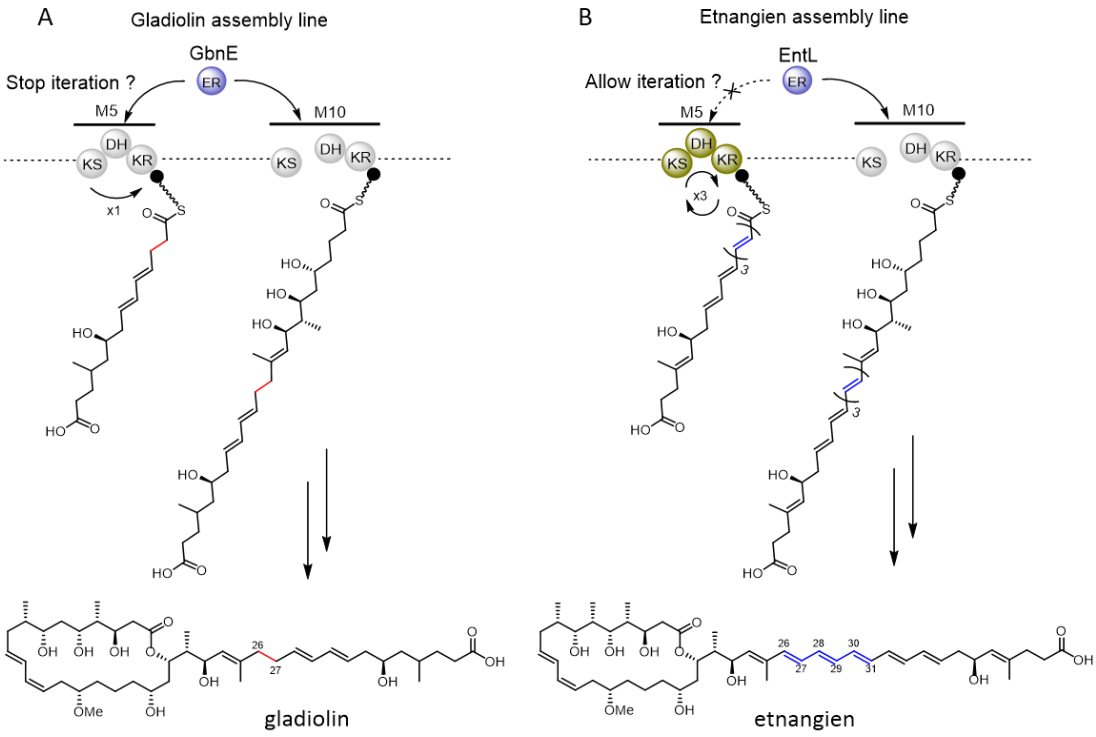


Figure 3.1 Comparison of partial biosynthetic pathways of gladiolin (A) and etnangien (B). Schematic view shows the iterative use of EntE Module 5 and non-iterative use of GbnD2 Module 5, and the proposed roles of the *trans*-acting ERs in controlling these behavious.

3.1 Iterative nature of GbnD2 Module 5

3.1.1 Production of recombinant GbnD2 Module 5

To examine whether GbnD2 Module 5 exhibits iterative activity, an *in vitro* biochemistry approach was employed. Using standard cloning procedures, pET28a-GbnD2_Module 5 (KS-DH-KR-ACP) was generated for overproduction and purification as an N-His_6 construct (detailed in 6.5.1). Large scale overproduction followed by nickel affinity chromatography and size exclusion purification (detailed in 6.5.3 and 6.5.4), afforded the purified GbnD2 Module 5 tetra-domain protein, with the size and purity of the construct confirmed by SDS-PAGE and LC-MS analysis (Figure 3.2).

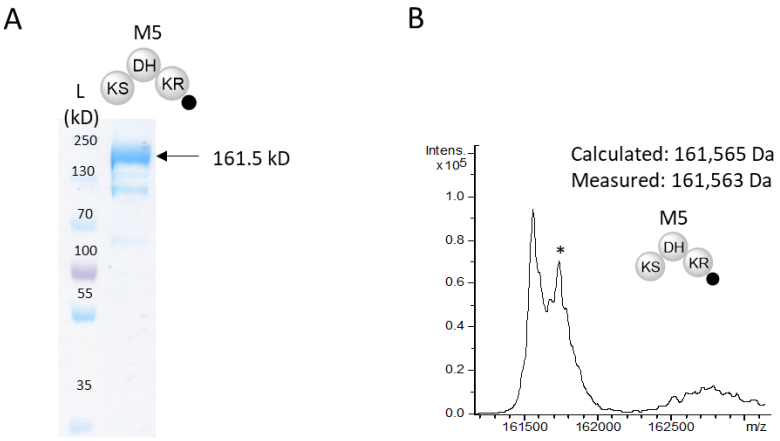
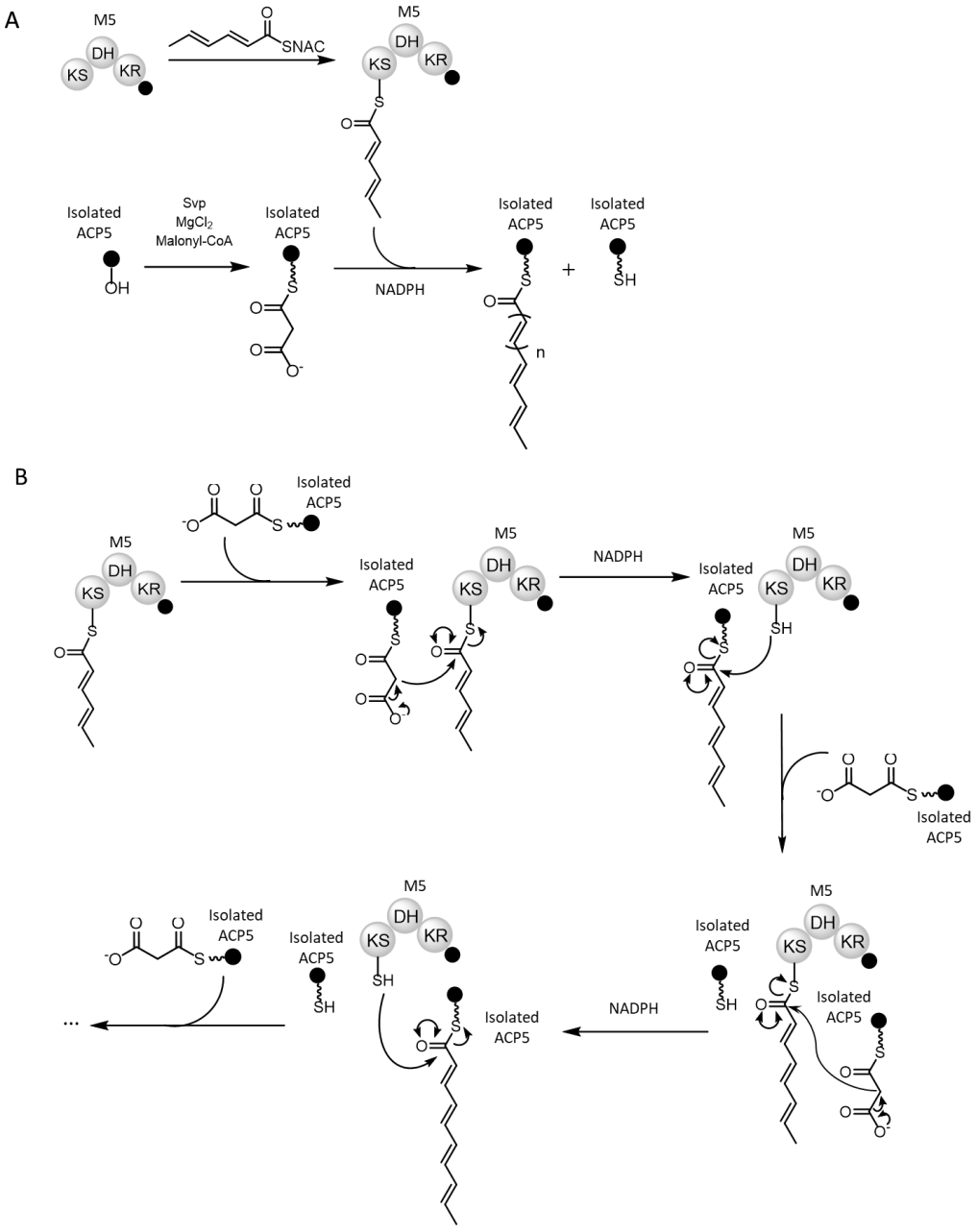


Figure 3.2 SDS-PAGE and LC-MS analysis of purified N-His$_6$-GbnD2 Module 5. (A) 6% SDS-PAGE gel showing the size and purity of N-His$_6$-GbnD2 Module 5. **(B)** Deconvoluted mass spectrum confirming the mass of N-His$_6$-GbnD2 Module 5 (-N-Met). Peak labeled '*' corresponds to the spontaneous gluconoylation of His-tag on fusion proteins with additional 178 Da being observed[147]. See appendix for raw Mass spectra.

3.1.2 Investigating the activity of GbnD2 Module 5 *in vitro*

Initial attempts to probe chain elongation in GbnD2 Module 5 centered on monitoring the covalently tethered growing acyl chain, using intact protein mass spectrometry. However, due to the large size of GbnD2 Module 5, the acquired mass spectra of the intact module resulted in a broad peak (Figure 3.2 B), which made resolution of small mass increments due to elongation extremely difficult. Therefore, an alternative strategy based on monitoring the isolated GbnD2 ACP5 domain, which is significantly smaller in size and thus allowing discrimination of small mass shifts on the acyl chain, was devised (Scheme 3.1 A). In this assay, instead of loading malonyl-PPant onto the integral ACP domain, the preacylated-KS domain of GbnD2 Module 5 was supplied with the malonylated GbnD2 ACP5 domain as a stand-alone protein. Upon interaction of malonylated ACP5 domain with the acylated-module 5, chain elongation, ketoreduction and dehydration can then occur, with the extended acyl chain attached to the isolated GbnD2 ACP5 domain (Scheme 3.1 B). If GbnD2 Module 5 is capable of catalyzing multiple rounds of chain extension, then the isolated ACP5 domain can transfer the attached extended acyl chain back to the KS domain and itself become *holo* form, priming GbnD2 Module 5 for a second round of elongation (Scheme 3.1 B). This process can occur multiple times resulting in iterative activity. In this way, the isolated malonyl-GbnD2 ACP5 domain would be continually consumed to form

species of the GbnD2 ACP5 domain with iteratively extended products tethered, and the *holo*-GbnD2 ACP5 species would be accumulated accordingly as every back transfer occurs.



Scheme 3.1 Design of *in vitro* assay (A) and proposed mechanism (B) for examining the iterative nature of GbnD2 Module 5. (A) The GbnD2 ACP5 domain is loaded with a malonyl unit by Svp catalyzed malonyl-CoA loading reaction, followed by incubation with GbnD2 Module 5 pre-acylated using 2,4-hexadienoyl-SNAC. **(B)** The pre-acylated GbnD2 Module 5 is supplied with the isolated manoyl-GbnD2 ACP5 domain. The chain extension reaction yields a product tethered to the isolated GbnD2 ACP5 domain which is transferred back to the KS domain and the *holo*-GbnD2 ACP5 domain is formed. The second round of chain extension proceeds via condensation

with another molecule of the malonylated GbnD2 ACP5 domain and further rounds of these process to produce the iteratively-extended product.

The *in vitro* assay was conducted as described, and was allowed to proceed for up to 20 hours to allow maximum possible conversion to product (detailed in 6.6.4.1). Intact protein mass spectra of the GbnD2 ACP5 domain were collected at three time points, 2, 5 and 20 hours. The deconvoluted mass spectra of the intact GbnD2 ACP5 domain at these time points displayed multiple additional peaks when compared to the malonyl-ACP5 starting material (Figure 3.3 A). Calculation of the mass shifts for the newly formed peaks in the 2 and 5 hours reactions revealed three new species, corresponding to one to three rounds of chain elongation, reduction and dehydration, ultimately generating a pentaene acyl chain (Figure 3.3 A). Allowing the reaction to proceed for 20 hours resulted in a fourth chain extension and reduction, yielding a hexaene as well as a small amount of β-keto unmodified species (Figure 3.3 A). The formation of all iteratively extended products increased with time. Moreover, an increasing formation of the *holo* GbnD2 ACP5 domain and consumption of malonylated GbnD2 ACP5 domain over time was observed as expected. Peaks corresponding to the acetylated GbnD2 ACP5 domain were also observed, which are likely the result of decarboxylation of the malonyl unit (either in the MS, spontaneous or KS-catalysed). As an additional observation, during the course of these reactions, the solution turned yellow indicating polyene formation in the test tube (Figure 3.3 B).

These observations strongly suggest that GbnD2 Module 5 is able to catalyze iterative chain elongation using the isolated GbnD2 ACP5 domain. The ability of the GbnD2 ACP5 domain to re-acylate the KS domain of module 5, and therefore re-prime the module for elongation, is an interesting observation. However, whether this activity is an artifact of the *in vitro* set-up of the reaction, or indeed a feature exclusive to iterative modules requires further study, as detailed in section 3.1.3.

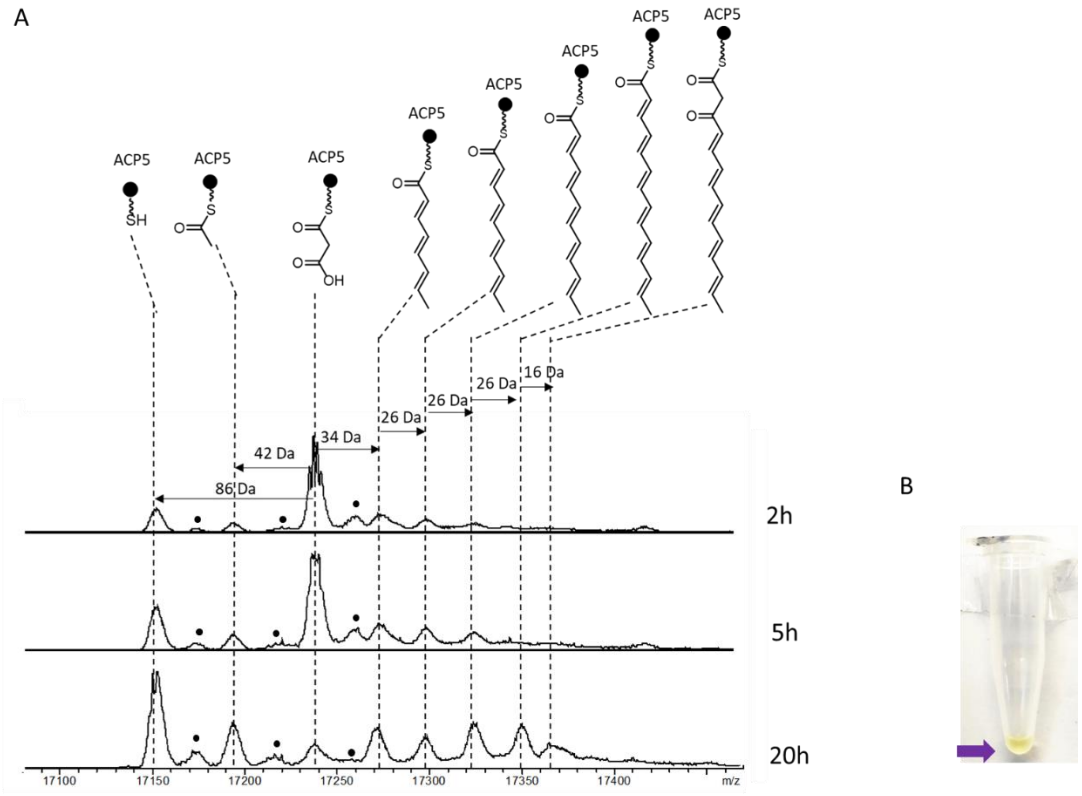


Figure 3.3 LC-MS analysis of *in vitro* reconstitution of iterative activity of GbnD2 Module 5. (A) Deconvoluted mass spectra of the isolated malonyl-GbnD2 ACP5 domain, following incubation with 2,4-hexadienoyl-acylated-GbnD2 Module 5 and NADPH for 2 hours (top), 5 hours (middle) and 20 hours (bottom). The peaks labelled with a black dot were found to be the sodium adduct of the corresponding protonated ion as their intensities showed the same peak intensity pattern. **(B)** Yellow colour of the reaction solution indicates polyene formation.

3.1.3 Probing the ACP-selectivity of GbnD2 Module 5 iteration

The results described in 3.1.2 demonstrate the ability of GbnD2 Module 5 to catalyse iterative chain elongation using a stand-alone GbnD2 ACP5 domain. The key step which promotes iterative activity is the reacylation of the KS domain by the chain-extended ACP5 domain. In these experiments, the GbnD2 ACP5 domain is supplied as an isolated entity, which is different to the integrated situation of the 'native' module, and therefore the reacylation activity may be an artifact of 'dissected' nature of the assay. However, equally, the ability of GbnD2 ACP5 domain to reacylate the KS domain could be a result of specific protein-protein interaction that facilitates the chain translocation between the GbnD2 ACP5 domain and the KS domain of module 5 (as introduced in section 1.2.2.5 'Acyl carrier protein domains and their interactions with other

domains'). This would mean that other ACP domains may not possess this ability. To test this hypothesis, ACP domains from the gladiolin PKS modules containing the same set of reductive domains (i.e. KS-DH-KR) were selected, cloned, overexpressed and purified. The GbnD1 ACP3 and the GbnD4 ACP10 domains, used in the Chapter 2, were selected as candidates to test this theory.

It was noted that the purified recombinant N-His_6-GbnD1 ACP3 and N-His_6-GbnD4 ACP10 domains underwent considerable gluconoylation (Figure 2.26). The mass addition of the gluconoylated species (+ 178 Da) is similar to an ACP species with two rounds of iteratively extended product tethered (+ 172 Da). In addition to this, another set of peaks corresponding to the gluconoylated species would also be formed during the reaction, which would bring complexity to the data analysis. Thus, the His-tag cleavage of both N-His_6-GbnD1 ACP3 and N-His_6-GbnD4 ACP10 domains was cleaved using thrombin. The cleavage was confirmed by SDS-PAGE analysis and LC-MS (Figure 3.4).

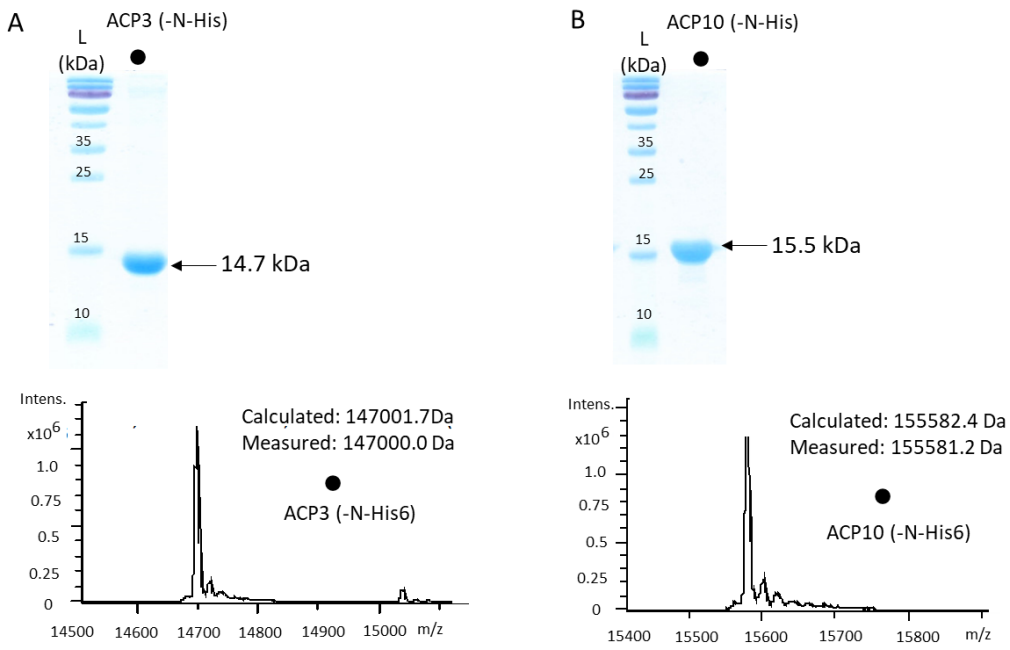


Figure 3.4 SDS-PAGE analysis and LC-MS analysis of N-His_6-tag cleaved GbnD1 ACP3 and GbnD4 ACP10 domains. 15% SDS-PAGE gel (top) and deconvoluted mass spectra (bottom) confirming the identity of His-tag cleaved GbnD1-ACP3 **(A)** and GbnD4 ACP10 domains **(B)**. See appendix for raw mass spectra.

The *in vitro* assay in Scheme 3.1 A was then repeated using the purified GbnD1 ACP3 (-His-tag) and GbnD4 ACP10 (-His-tag) respectively under the same conditions (detailed in 6.4.4.1). The

deconvoluted mass spectra of the intact ACP species from the corresponding reactions showed that, unlike in the reaction using the GbnD2 ACP5 domain (Figure 3.5 a), no chain extension products were observed for the GbnD1 ACP3 domain (Figure 3.5 b). Intrestingly, additional to first round chain extension, the GbnD2 ACP10 domain was able to undergo the second and third chain extension reaction. (Figure 3.5 c). However, the obvious difference in intensity of the iteratively extended products between reactions using the GbnD2 ACP5 and GbnD4 ACP 10 domains could still supports the conlcusion that there is a specific interaction between the GbnD2 ACP5 domain and the GbnD2 KS5 domain during the intermediate back-transfer, or reacylation process for iteration.

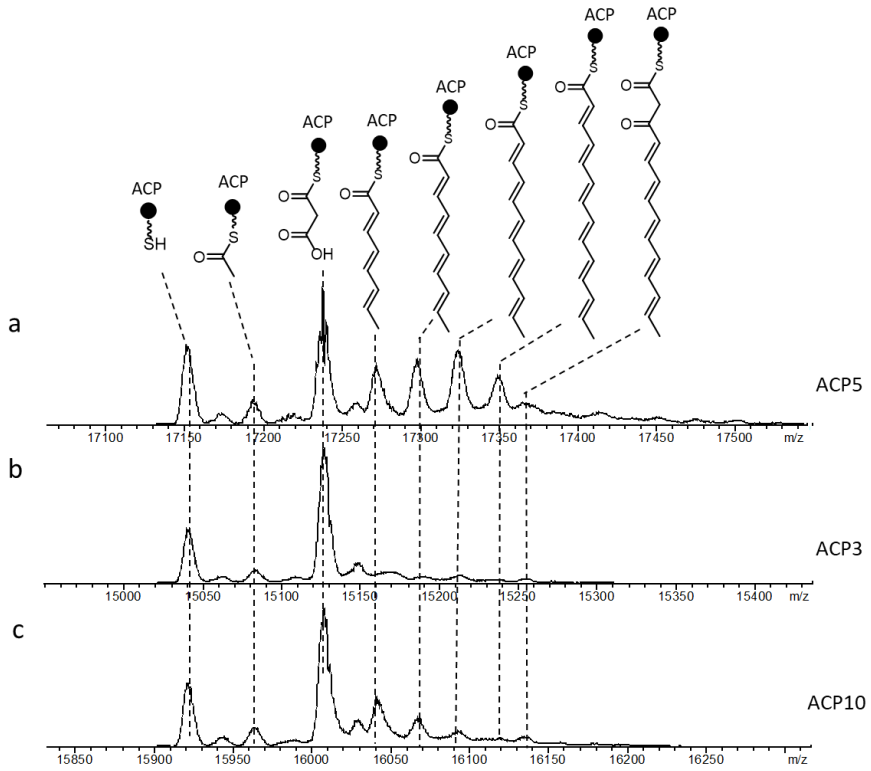


Figure 3.5 LC-MS analysis of *in vitro* iteration control assays for GbnD2 Module 5. Deconvoluted mass spectra of the isolated malonyl-GbnD2 ACP5 (a), malonyl-GbnD1 ACP3 (b) and malonyl-GbnD4 ACP10 (c) following respective incubation with 2,4-hexadienoyl-acylated-GbnD2 Module 5 and NADPH.

The minor interaction between the GbnD4 ACP10 domain and the GbnD2 KS5 domain was tentatively attributed to the structural similarity between the GbnD2 ACP5 and the GbnD4 ACP10 domain, as this was consistent with the findings that they are the only ACP domains from the gladiolin assembly line that are capable of interacting with GbnE. It was therefore plausible to

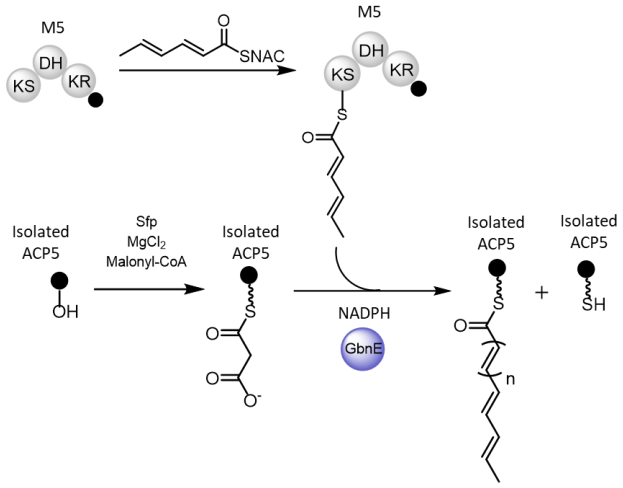
speculate that the GbnD2 ACP5 domain interacts with the GbnD2 KS5 domain for *intra*-modular translocation in a similar way that it interacts with GbnE for enoyl reduction. Thus, in a competitive situation, GbnE could prevent the GbnD2 ACP5 domain from transfering the one-round extended product back to the GbnD2 KS5 domain.

3.2 The effect of GbnE on GbnD2 Module 5 iteration

Having established the iterative nature of GbnD2 Module 5, and that a specific interaction exists with its cognate ACP domain, the next step was to examine the effect of the *trans*-acting ER domain upon the iterative activity. The hypothesis outlined in Figure 3.1 A dictates that, following enoyl reduction of the α,β-unsaturated intermediate, iterative chain elongation is halted.

3.2.1 GbnE abolishes iteration of GbnD2 Module 5

To probe whether the GbnE has an impact on the iterative activity of this module5, the *in vitro* iteration assay was repeated in the presence of GbnE. In the same manner as in section 3.1.2, the GbnD2 KS5 domain was pre-acylated with a 2,4-hexadienoyl unit, followed by incubation with isolated malonylated GbnD2 ACP5, NADPH and GbnE (Scheme 3.2). The ratio of pre-acylated GbnD2 Module 5: malonyl-GbnD2 ACP5: GbnE was set as 1:2:1.



Scheme 3.2 Design of *in vitro* iteration assays for monitoring GbnD2 Module 5 iteration in the presence of GbnE .

The reaction in the presence of GbnE appeared to abolish the production of any iteratively-extended products (Figure 3.6 b). The formation of the *holo*-GbnD2 ACP5 domain, resulting from the back transfer process (Figure3.6 b) was consistently found to be much lower compared to the reaction without GbnE (Figure 3.6 a), indicating no back transfer occurred.

Interestingly, although the one-round elongated product was observed (Figure3.6 b), it was found to be the un-reduced species. This indicated that, with GbnE present, the isolated malonyl-GbnD2 ACP5 domain could still interact with the GbnD2 KS5 domain for the chain elongation, but the interaction of the resulting the elongated β-keto-GbnD2 ACP5 with the KR and DH domains for the chain modification was blocked in the *in vitro* reconstituted condition. This could be due to the high affinity between GbnE and the discrete GbnD2 ACP5 domain.

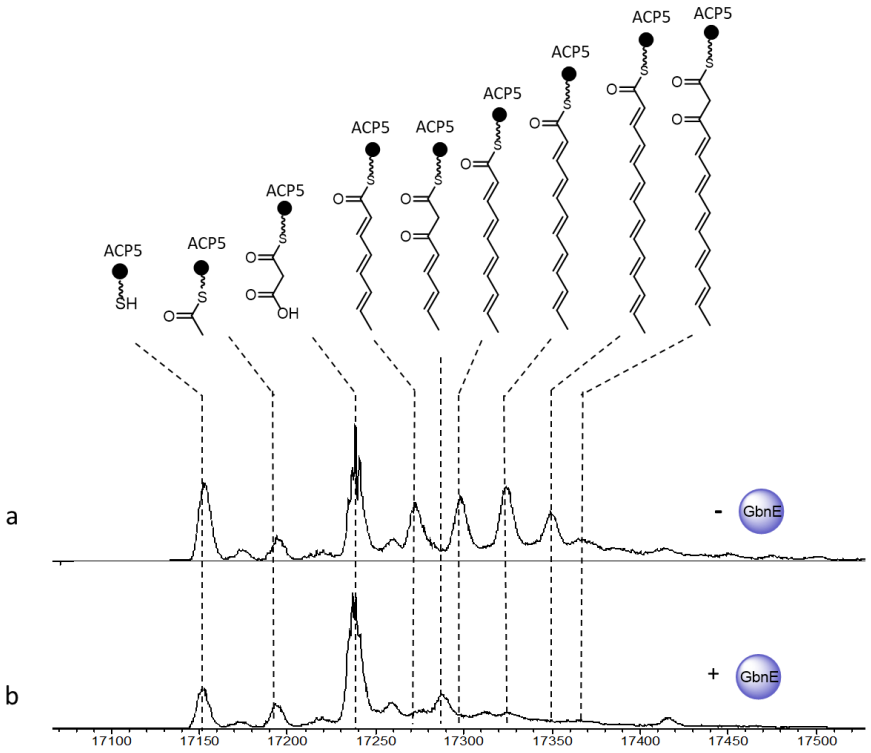


Figure 3.6 LC-MS analysis of *in vitro* GbnD2 Module 5 iteration assays with and without addition of GbnE. Deconvoluted mass spectra of isolated malonyl-GbnD2 ACP5 domain following incubation with 2,4-hexadienoyl-acylated-GbnD2 Module 5 and NADPH without (a) and with (b)addition of GbnE.

These data strongly support that GbnE is able to prevent the iterative activity of GbnD2 Module 5 during the gladiolin biosynthesis. However, the mechanism by which this occurs requires

further investigation. In principle, two routes for this phenomenon exist: a) GbnE engages in protein-protein interactions with the GbnD2 ACP5 domain preventing intermediate back. b) The catalytic activity of GbnE results in a saturated intermediate, which prevents the back-transfer process due to a substrate specificity of the GbnD2 KS5 domain or the kinetic difference in acylation of the upstream KS5 domain and the downstream GbnD2 KS6 domain.

3.2.2 Protein-protein interactions between GbnE and the GbnD2 ACP5 domain prevent re-acylation of the GbnD2 KS5 domain

In order to investigate whether protein-protein interactions between GbnE and GbnD2 ACP5 domain have an effect on the iteration, the catalytically inactivated GbnE mutant, GbnE(H198V), was employed. It was previously shown to be well folded in Chapter 2, and therefore can still engage in protein-protein interactions with the GbnD2 ACP5 domain.

The assay outlined in Scheme 3.2 was repeated, but this time in the presence of GbnE(H198V). Interestingly, the subsequent LC-MS analysis of the GbnD2 ACP5 domain revealed that iteration of GbnD2 Module 5 was inhibited by the presence of GbnE(H198V), yielding an almost identical spectrum to WT GbnE (Figure 3.7). These observations strongly suggest that merely the presence of GbnE, and the protein-protein interactions between itself and GbnD2 ACP5 domain, are sufficient to halt iterative activity.

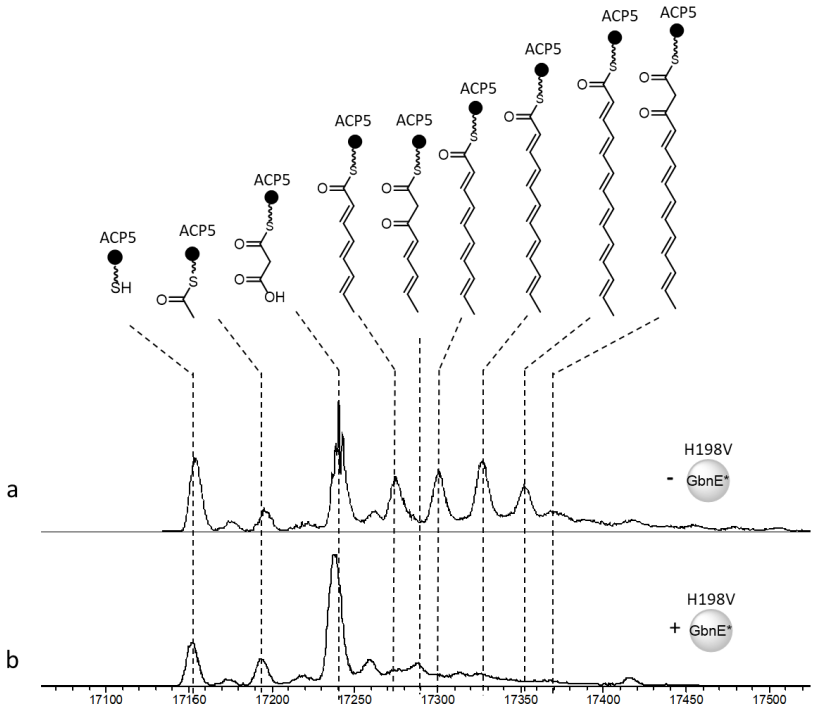


Figure 3.7 LC-MS analysis of *in vitro* GbnD2 Module 5 iteration assays with and without addition of GbnE (H198V). Deconvoluted mass spectrum of the isolated malonyl-GbnD2 ACP5 domain following incubation with 2,4-hexadienoyl-acylated-GbnD2 Module 5 and NADPH without (a) and with (b)addition of GbnE (H198V).

3.2.3 The 'gate-keeping' role of the GbnD2 KS6 domain

The data presented in 3.2.2 convincingly shows how the mere presence of GbnE, and not necessarily its catalytic activity, is all that is required to abolish iteration in GbnD2 module 5. However, whether formation of a saturated acyl intermediate, resulting from the GbnE catalysis, contributes to the directionality or flux of the assembly line at this biosynthetic stage also required investigation. Genetic data from chapter 2 showed that deletion of GbnE in the gladiolin producing strain resulted in a general 'slowing down' of the assembly line (section 2.1.1.1) This suggested that the fidelity of the assembly line is preserved by the downstream KS domains in a 'gate-keeping' fashion to preferentially accept the GbnE processed α,β-saturated intermediate.

A KS domain conducts two catalytic processes during the chain elongation cycle: chain translocation and Claisen-like chain elongation. Substrate specificity of KSs from modular PKS, especially from *trans*-AT PKS, have been revealed in both processes.[62,153] Thus, to fully examine the proposed 'gate-keeping' function of the GbnD2 KS6 domain, the substrate specificity of it in

both processes was inspected.

3.2.3.1 Construct design and protein overproduction

GbnD2 Module 6 consists of a KS domain and two tandem ACP domains (Figure 2.2). This type of enzymatic domain architecture is typically associated with β-branching, with the latter ACP domain aiding in the interaction with the *trans*-acting β-branching enzymes.[154] Therefore, the GbnD2 ACP6b domain was considered not needed for both the chain translocation and elongation process. Unfortunately, initial construct of GbnD2 KS6-ACP6a yielded insoluble protein. In an effort to obtain a soluble construct, instead of further isolation of GbnD2 KS6 and GbnD2 ACP6a as individual domains, the entire module 6 with the second ACP domain inactivated was explored as an alternative.

Expression construct pET28a-GbnD2 Module 6 was initially generated, followed by mutation of the conserved serine residue (S941) for phosphopantetheinylation of the second ACP domain, to alanine via site-directed mutagenesis, to achieve the final construct pET28-GbnD2 Module 6 (S941A). N-His_6-GbnD2 Module6 (S941A) was then overproduced and purified as described in 6.5.3 and 6.5.4. The identity of the purified N-His_6-GbnD2 Module6 (S941A) was confirmed by SDS-PAGE and LC-MS analysis (Figure 3.8).

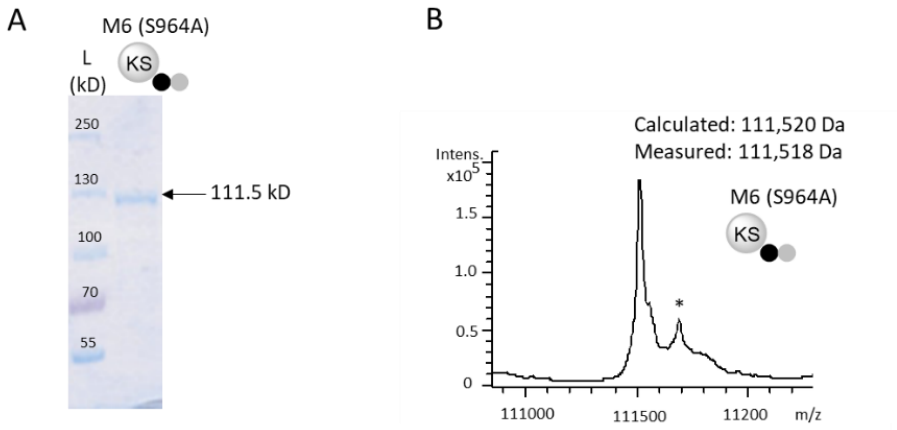
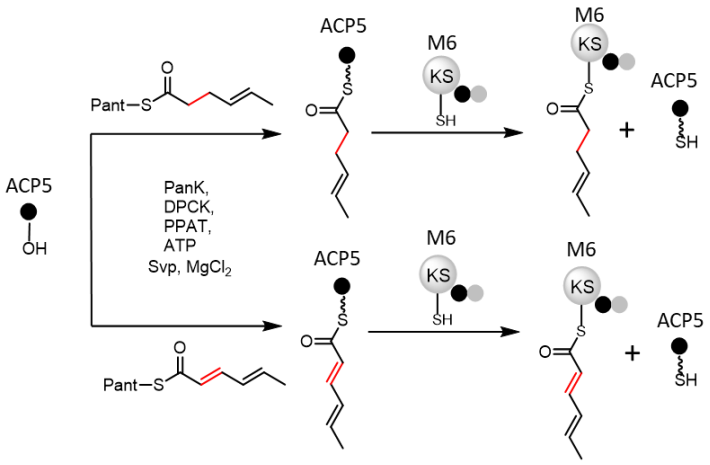


Figure 3.8 SDS-PAGE and LC-MS analysis of purified N-His_6-GbnD2 Module 6(S941A). (A) 6% SDS-PAGE gel of N-His_6-GbnD2 Module 6 (S941A). **(B)** Deconvoluted mass spectrum confirming the exact mass of N-His_6-GbnD2 Module 6 (S941A). The Peak labeled '*' corresponds to the spontaneous gluconoylation of His-tag on corresponding fusion proteins with additional 178 Da being observed[147]. See appendix for raw mass spectra.

3.2.3.2 Substrate specificity of the GbnD2 KS6 domain in chain translocation

In order to examine the substrate specificity of the GbnD2 KS6 domain towards α,β-saturated and α,β-unsaturated substrates in the chain translocation step, a comparative *in vitro* assay was conducted (Scheme 3.3). Utilizing GbnD2 Module 6 and the upstream GbnD2 ACP5 domain loaded with either a 4-hexenoyl or a 2,4-hexadienoyl acyl chain, the ability of the GbnD2 ACP5 domain to transfer these intermediates to the GbnD2 KS6 domain could be measured. This could be achieved by monitoring the ratio of acyl-GbnD2 ACP5 to the newly formed *holo*-GbnD2 ACP5 species.

Both 2,4-hexadienoyl-GbnD2 ACP5 and 4-hexenoyl-GbnD2 ACP5 were first synthesized via phosphopantetheinylation enzymes cassette catalyzed acyl-PPant loading assays, using the corresponding pantetheine form substrates synthesized by PhD student Christian Hobson. The full conversion from *apo* form to respective acyl-PPant-loaded form were first confirmed (Figure 3.9), to ensure the GbnD2 KS6 domain is provided with equal concentration of the acyl transfer partner.



Scheme 3.3 Design of *in vitro* assays for monitoring substrate specificity of GbnD2 KS6 domain in the translocation process. The GbnD2 ACP5 domain is respectively loaded with 4-hexenoyl-Pant (top) and 2,4-hexadienoyl-Pant (bottom) via phosphopantetheinylation enzymes cassette, followed by incubation with GbnD2 Module 6(S941A) for inter-modular chain translocation.

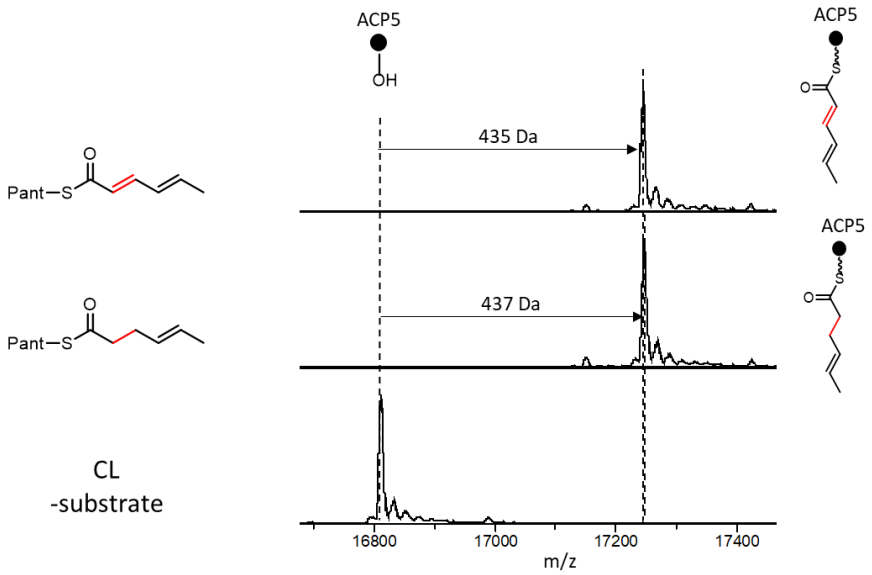


Figure 3.9 LC-MS analysis of *in vitro* PPtase catalyzed 2,4-hexadienoyl-PPant and 4,-hexenoyl-PPant loading of GbnD2 ACP5 domain. Deconvoluted mass spectrum of *apo*-GbnD2 ACP5 (bottom), following incubation with phosphopantetheinylation catalytic enzymes cassette, Svp, $MgCl_2$, and 4-hexenoyl-Pant (bottom), or 2,4-hexadienoyl-Pant (top). Successful loading results in mass shift of 435 Da for 2,4-hexadienoyl-PPant unit and 437 Da for 4-hexenoyl unit-PPant.

The acyl transfer reaction was initiated by incubating the saturated and unsaturated acyl-GbnD2 ACP5 domain with GbnD2 Module 6(S941A) in a ratio of 1:3 (as detailed in 6.6.5.1). Control reactions were conducted lacking the GbnD2 Module 6(S941A). The assays were allowed to proceed for up to 8 hours and data were collected at 4 time points. The deconvoluted mass spectrum of the GbnD2 ACP5 domain from each assay was compared. This showed the increasing formation of the *holo*-GbnD2 ACP5 domain from the 4-hexenoyl-GbnD2 ACP5 was incubated with GbnD2 Module 6 (Figure 3.10 A). However, very little *holo*-GbnD2 ACP5 when the 2,4-hexadienoyl-GbnD2 ACP5 was used compared to the control reaction (Figure 3.10 B). These observations clearly demonstrated the preference of the GbnD2 KS6 domain for α, β-saturated acyl units over α, β-saturated acyl units in the chain translocation step.

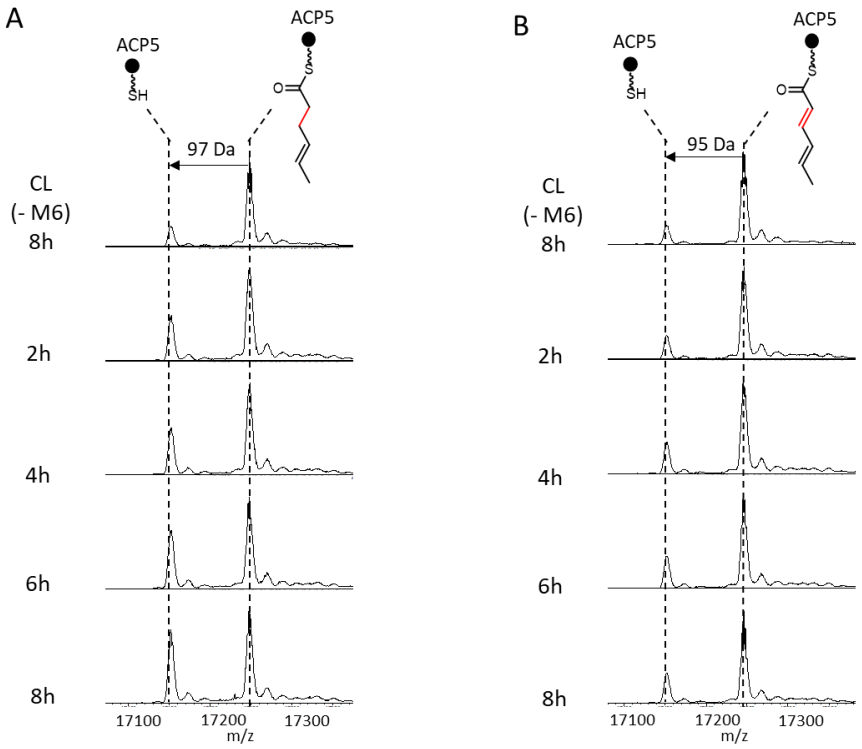
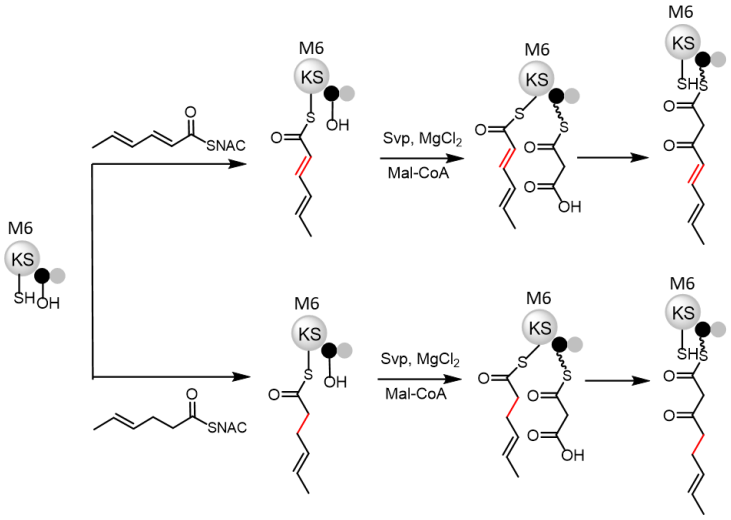


Figure 3.10 LC-MS analysis of *in vitro* chain translocation between the GbnD2 ACP5 domain and GbnD2 KS6 domain. (A) Deconvoluted mass spectra of 4-hexenoyl-GbnD2 ACP5 (top) following incubation with GbnD2 Module 6(S941A) for 2 hours, 4 hours, 6 hours, 8 hours. **(B)** Deconvoluted mass spectra of 2,4-hexadienoyl-GbnD2 ACP5 (top) following incubation with GbnD2 Module 6(S941A) for 2 hours, 4 hours, 6 hours, 8 hours. 'CL' refers to control reaction lacking GbnD2 Module 6(S941A) after 8 hours incubation.

3.2.3.3 Substrate specificity of the GbnD2 KS6 domain in chain elongation

In addition to the chain translocation step, the GbnD2 KS6 domain was also tested for substrate preference at the chain elongation stage, utilizing the *in vitro* assays outlined in scheme 3.4. In the first instance, acylation of GbnD2 KS6 domain active site with a α, β-saturated or a α,β-unsaturated acyl units was required, in order to provide the same starting point for the comparison of the elongation reactions. Although substrate preference can be observed for the ACP-bound acyl chains, it is possible to obtain reasonable levels of KS acylation using acyl-SNAC substrates. Therefore, in an attempt to equally supply GbnD2 KS6 domain with saturated and unsaturated acyl substrates, GbnD2 Module 6 (S941A) was incubated with the same concentration of 2,4-hexadienoyl-SNAC and 4-hexenoyl-SNAC (detailed in 6.6.5.2). The acylation

reactions were analysed by intact protein mass spectrometry, showing a expected similar acylation percentage (around 30%) (Figure 3.11).



Scheme 3.4 Design of *in vitro* assays for monitoring substrate specificity of GbnD2 KS6 domain during chain elongation. The KS domain of GbnD2 Module 6(S941A) is respectively acylated by 2,4-hexadienoyl-SNAC (top route) and 4-hexenoyl-SNAC (bottom route) to reach a similar acylation level, followed by supply with extender unit via Svp catalyzed malonyl-CoA loading reaction for the KS domain catalyzed chain elongation reaction.

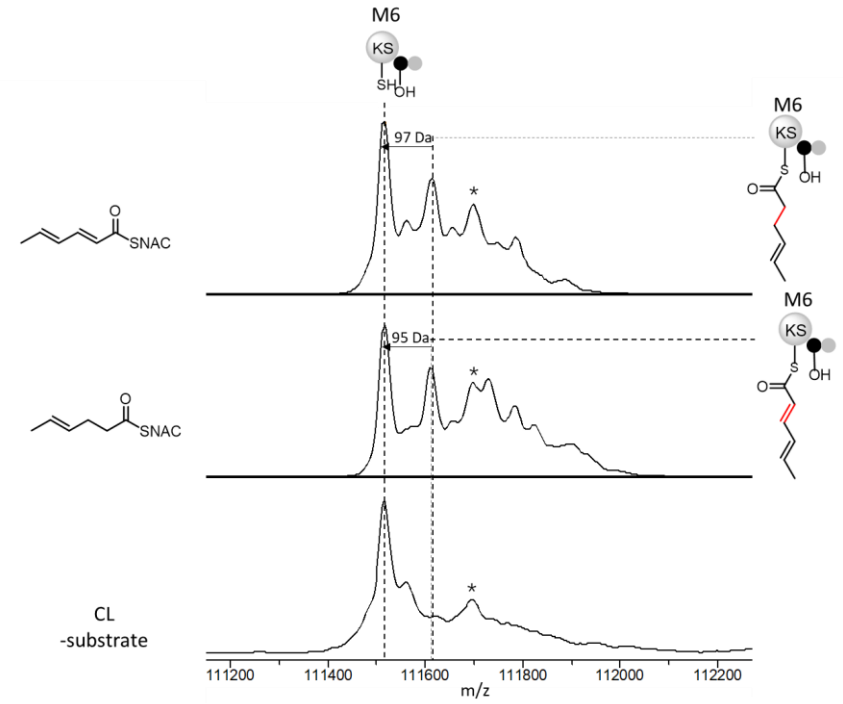


Figure 3.11 LC-MS analysis of acylation of the KS domain in GbnD2 Module 6(S941A). Deconvoluted mass spectra of *apo*-GbnD2 Module 6(S941A) (bottom) following incubation with 2,4-hexadienoyl-SNAC (middle) and

4-hexenoyl-SNAC (top) respectively. Acylation by 2,4-hexadienoyl and 4-hexenoyl unit was observed as 95 Da and 97 Da mass shift respectively.

With acylated GbnD2 Module 6 (S941A) in hand, the elongation assays were initiated by providing the GbnD2 ACP6a domain with a malonyl extender unit via the PPtase catalyzed malonyl-PPant loading reaction (detailed in 6.6.5.2). The deconvoluted mass spectra of intact GbnD2 Module 6 (S941A) species from both assays showed that both 4-hexenoyl- and 2,4-hexadienoyl-elongated products are observed with similar turnover ratio (Figure 3.12). This indicated that the GbnD2 KS6 domain does not have strict substrate specificity in the chain elongation process.

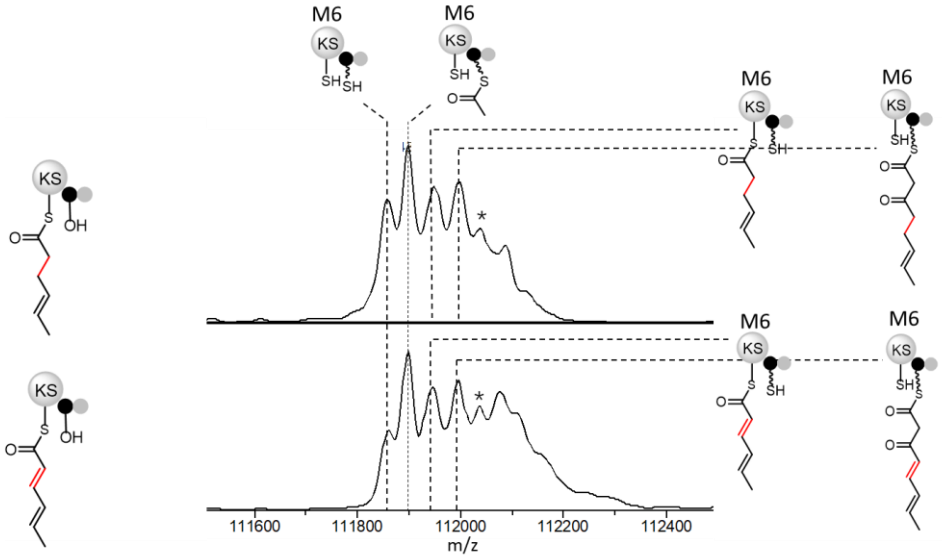


Figure 3.12 LC-MS analysis of GbnD2 Module6 (S941A) *in vitro* chain elongation assays. Deconvoluted mass spectra of 4-hexenoyl-acylated-GbnD2 Module 6(S941A) (top) and 2,4-hexadienoyl-GbnD2 Module 6 (S941A) (bottom) respectively following loading with malonyl-PPant by PPtase.

3.3 Conclusions and future work.

In this chapter, insights into the key structural difference between the construction of etnangien and gladiolin were obtained via investigation of the molecular mechanism underlying the non-iterative usage of module 5 in the gladiolin PKS. Firstly, an optimized *in vitro* iterative chain extension assay was developed for examining the ability of GbnD2 Module 5 to act iteratively in which the pre-acylated GbnD2 Module 5 was supplied with isolated malonyl-ACP5 'in *trans*'. This

allowed multiple turnovers to be achieved in addition to facilitating convenient reaction monitoring. With this strategy, GbnD2 Module 5 was successfully demonstrated to be capable of catalyzing up to 4 cycles of chain elongation and modification *in vitro*.

Furthermore, the addition of the *trans*-acting ER, GbnE abolished the iterative activity of GbnD2 Module 5. Utilizing the catalytically inactivated GbnE, it was possible to show that interactions between GbnE and the GbnD2 ACP5 domain are all that is required to shut-down iteration. This could be due to a multitude of complex factors, likely centered around protein-protein interactions preventing intermediate back transfer and higher order protein structural rearrangements, and requires further investigation. In addition, GbnD2 KS6 domain appeared to show a preference towards α,β-saturated acyl chains over an α,β-unsaturated ones. This suggests that although GbnE can abolish the module iteration process, the gatekeeping activity of GbnD2 KS6 domain also prevents chain translocation before enoyl reduction has occurred.

Using the aforementioned data, a molecular mechanism for chain length control in gladiolin biosynthesis can be proposed, and is outlined in (Figure 3.12). Following GbnD2 Module 5-catalysed chain elongation and subsequent ketoreduction and dehydration; the α,β-unsaturated intermediate is rejected by the downstream 'gate-keeping' GbnD2 KS6 domain for chain translocation (route b). However, the back-transfer of this α,β-unsaturated intermediate from the GbnD2 ACP5 domain to the GbnD2 KS5 domain (route c) is blocked by the efficient protein-protein interaction between GbnE and the GbnD2 ACP5 domain during the enoyl reduction event (route a). The GbnE-catalysed enoyl reduction event generates the α, β-saturated intermediate which is then preferentially accepted by the downstream GbnD2 KS6 domain (route d). Therefore, the single chain extended, α, β-unsaturated intermediate, is provided with a 'fast-track' to enoyl reduction and subsequent passage to the downstream GbnD2 KS6 domain, thereby maintaining the fidelity of gladiolin biosynthesis.

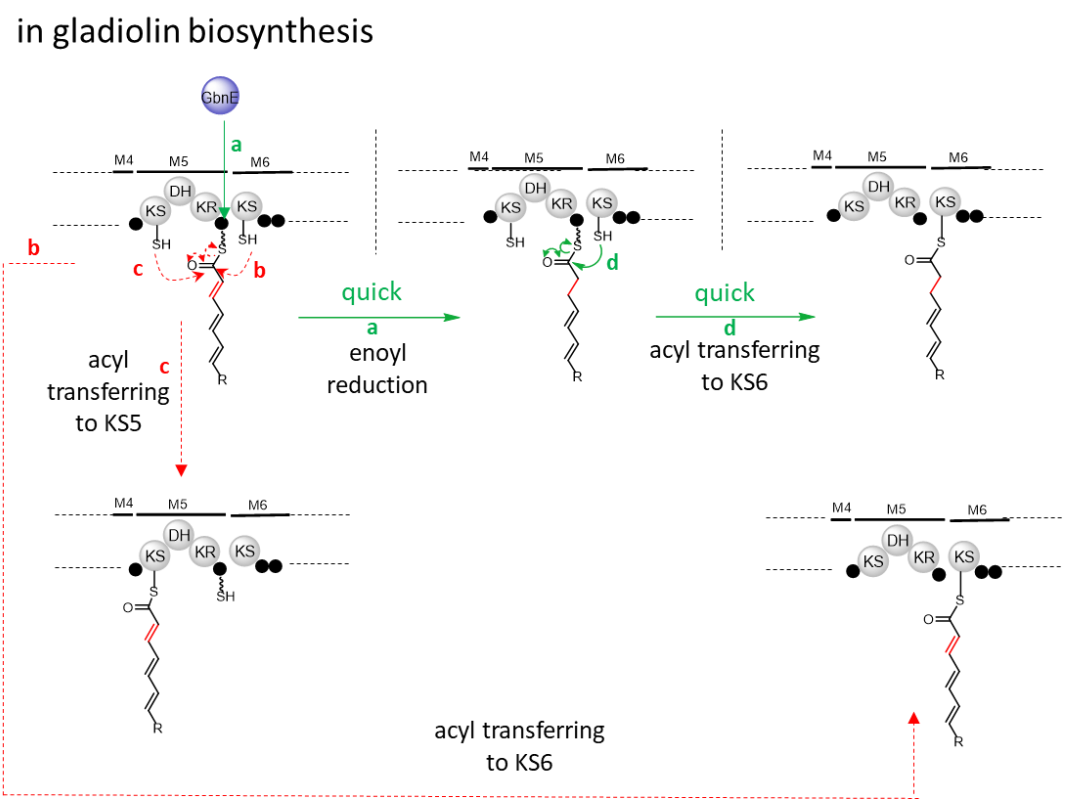


Figure 3.12 Schematic diagram of the proposed mechanism of the chain length control in the gladiolin biosynthesis. The iteration of module 5 is shut down via the recruitment of the *trans*-acting GbnE to this module. The direction and fidelity of chain assembly at this biosynthetic stage is preserved by the protein-protein interaction between GbnE and the GbnD2 ACP5 and 'gate-keeping' GbnD2 KS6 domains. The routes labeled by green are denoted as the free routes, the routes labeled by red are denoted as obstructed routes.

It is possible to extend this hypothesis to the etnangien system. Here, logic would dictate that the equivalent *trans*-acting ER domain (EtnL), located within the etnangien BGC, is unable to interact with the EntE ACP5 domain. Assuming an analogous sequence of events, the α, β-unsaturated intermediate formed by one cycle of chain extension and modification is not accepted by the downstream EtnE KS6 domain. Therefore, the chain is back-transferred onto the EntE KS5 domain to undergo another cycle of chain extension and reduction. This sequence would have to happen once more before being passed onto the EtnE KS6 domain. Whether the three cycles of chain extension and reduction is programmed, or merely an artifact of kinetics remains to be elucidated.

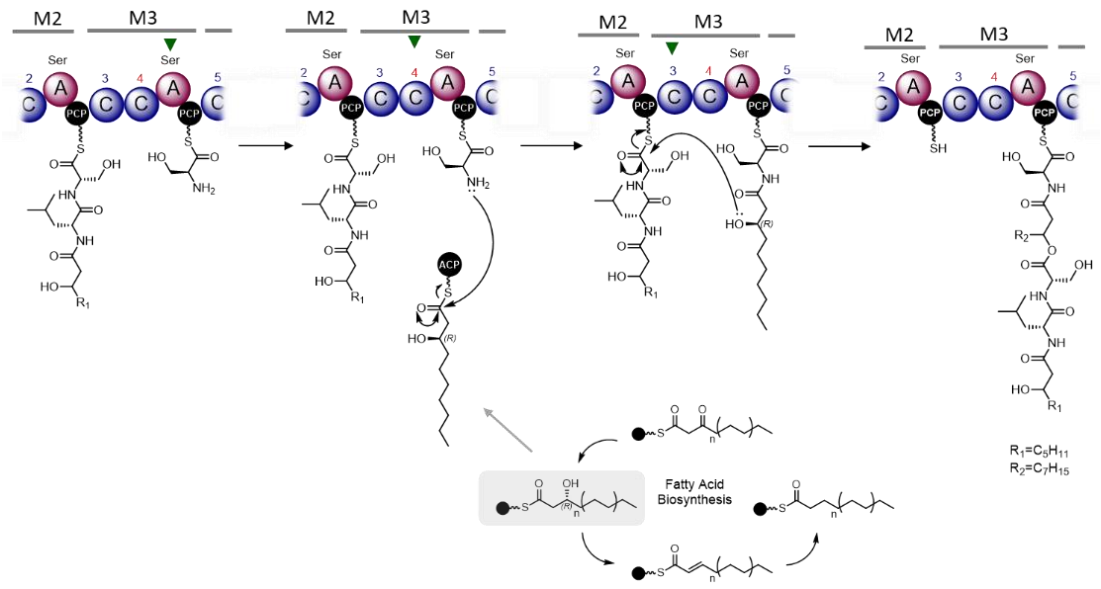
For future work, efforts could be made to further pin down the mechanism that GbnE uses to shut down the iteration of gladiolin module 5. This could include kinetic analysis of the GbnE

catalyzed enoyl-GbnD2 ACP5 enoyl reduction, and the GbnD2 KS6 domain catalyzed intermodular transfer and the GbnD2 KS5 domain catalyzed intramodular transfer. In order to obtain a complete picture, initially the analogous set of biochemical assays should be conducted using the etnangien EtnE Module 5 and the corresponding EtnL ER domain. In addition, a 'cross-talk' approach to *in vivo* engineering of both biosynthetic pathways could be a investigated, e.g. either swapping the GbnD2 ACP5 domain with the EntE ACP5 domain or GbnE with EntL in attempts to elicit iteration in gladiolin biosynthesis, or shut down the iteration in etnangien biosynthesis.

Chapter 4

Investigation of a double chain initiation mechanism in icosalides biosynthesis

As discussed in section 1.3.3, IcoA, responsible for the biosynthesis of icosalides, contains an unusual domain architecture due to the presence of an internal starter C domain (IcoA C4 domain) in module 3, which is directly adjacent to an $^{L}C_{L}$ domain (IcoA C3 domain). The domain architecture suggests the IcoA C3 domain and the IcoA C4 domain are required to co-operate for accomplishing a second round of chain initiation and an elongation event (Scheme 4.1). Firstly, the IcoA A3 domain is proposed to load the IcoA PCP3 domain with an L-serine residue. This would be followed by the IcoA C4 domain initiating a second round of chain assembly via condensation of the serine residue with a 3-hydroxyacyl unit, presumably hijacked from 3-hydroxyacyl-ACP in fatty acid biosynthesis[126,127]. Lastly, the IcoA C3 domain is proposed to catalyse a second condensation event, forming an ester bond between the β-hydroxyl acyl unit and the intermediate tethered to upstream IcoA PCP2 domain. If so, the IcoA C4 domain is believed to be the first example of an embedded starter C domain (Scheme 4.1).



Scheme 4.1 Proposed biosynthetic events in IcoA Module 3. Green arrows indicate which domains are proposed to catalyse each reaction. The black dots in the fatty acid biosynthesis represent ACP domains. IcoA A3 domain first load the downstream PCP3 domain with L-serine followed by C4 domain initiate a second round of chain assembly by incorporation of a 3-hydroxycayl unit hijacked from fatty acid biosynthesis, the C3 domain finally accomplish the biosynthesis in this module to link the newly formed Ser-hydroxydectonyl unit with the upstream assembled peptide chain by forming an ester bond.

4.1 Characterization of the IcoA C4 domain

The catalytic ability of the IcoA C4 domain to acylate the serine residue tethered onto the IcoA PCP3 domain with a 3-hydroxyacyl chain was first examined *in vitro*.

4.1.1 Production of recombinant IcoA C4 domain

The constructs pET151-IcoA_PCP3 and pET28a-IcoA_C4 were generated for the corresponding N-His_6-tag recombinant protein over production and purification (detailed in 6.5.1). The N-His_6-IcoA PCP3 domain and N-His_6-IcoA C4 domain were overproduced and purified as described in 6.5.3 and 6.5.4 The mass and purity of purified recombinant protein was confirmed by SDS-PAGE and LC-MS analysis (Figure 4.1)

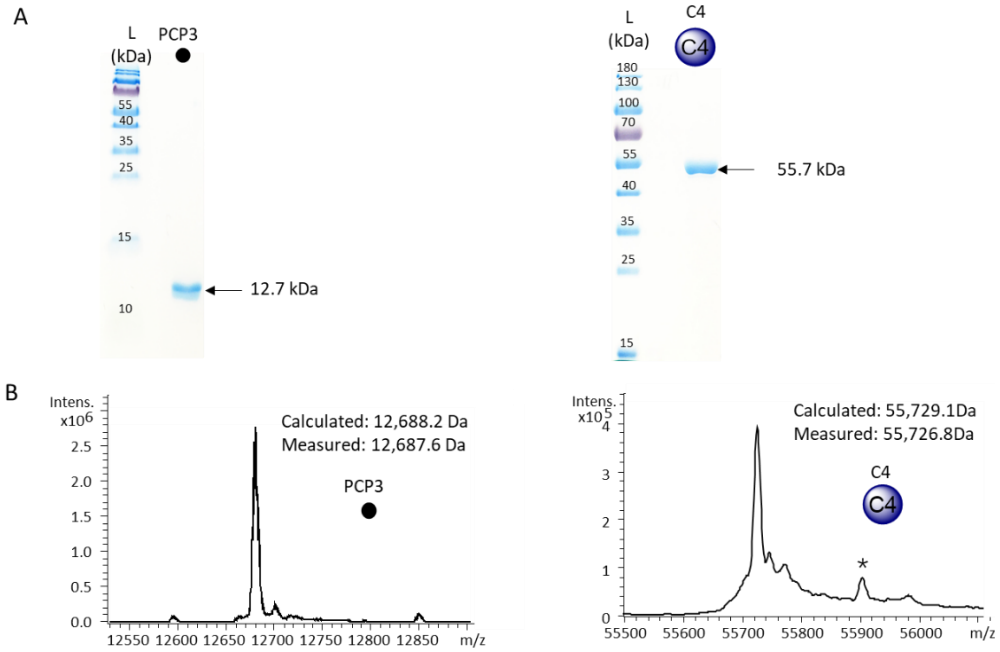
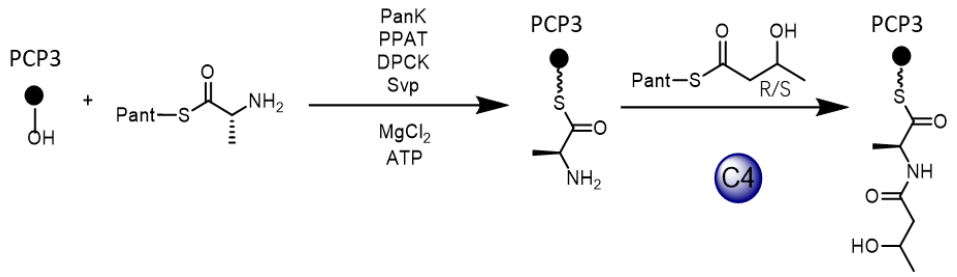


Figure 4.1 SDS-PAGE and LC-MS analysis of purified N-His_6-IcoA PCP3 and N-His_6-IcoA C4. (A) 12% SDS-PAGE gel and 10%SDS-PAGE gel showing the size and purity of N-His_6-IcoA PCP3(left) and N-His_6-IcoA C4(right) respectively, **(B)** Deconvoluted mass spectra confirming the exact mass of N-His_6-IcoA PCP3 (-N-Met) (left) and N-His_6-IcoA C4 (-N-Met) (right) respectively. Peak labeled as '*' corresponds to the spontaneous gluconoylation of the His-tag on fusion proteins, with an additional 178 Da observed.[147] See appendix for raw mass spectra.

4.1.2 Characterization of the N-acylation activity of the IcoA C4 domain

An *in vitro* assay outlined in Scheme 4.2 was utilized to characterize the N-acylation activity of the IcoA C4 domain. Due to the intrinsic instability of L-serine-Pant (found by Dr. Douglas Roberts during the synthesis trials), L-alanine-Pant (synthesised by Dr. Douglas Roberts) was alternatively used as a substrate mimic to first generate L-Ala-PPant-PCP3 as a substrate for the IcoA C4 domain. Loading of the L-Ala-PPant unit onto the IcoA PCP3 domain was achieved using the phosphopantetheinylation enzymes cassette as detailed in 6.6.1.2. The successful attachment of the L-alanine-PPant unit was confirmed by LC-MS analysis of the intact IcoA PCP3 domain following the loading reaction, with the expected 412 Da mass addition being observed (Figure 4.2 A).



Scheme 4.2 Design of *in vitro* assays for testing the N-acylation activity of the IcoA C4 domain. IcoA PCP3 domain. IcoA PCP3 domain is first loaded with L-Ala by phosphopantetheinylation enzyme cassette, followed by incubation with the IcoA C4 domain and *3R* or *3S*-hydroxybutyryl-Pant

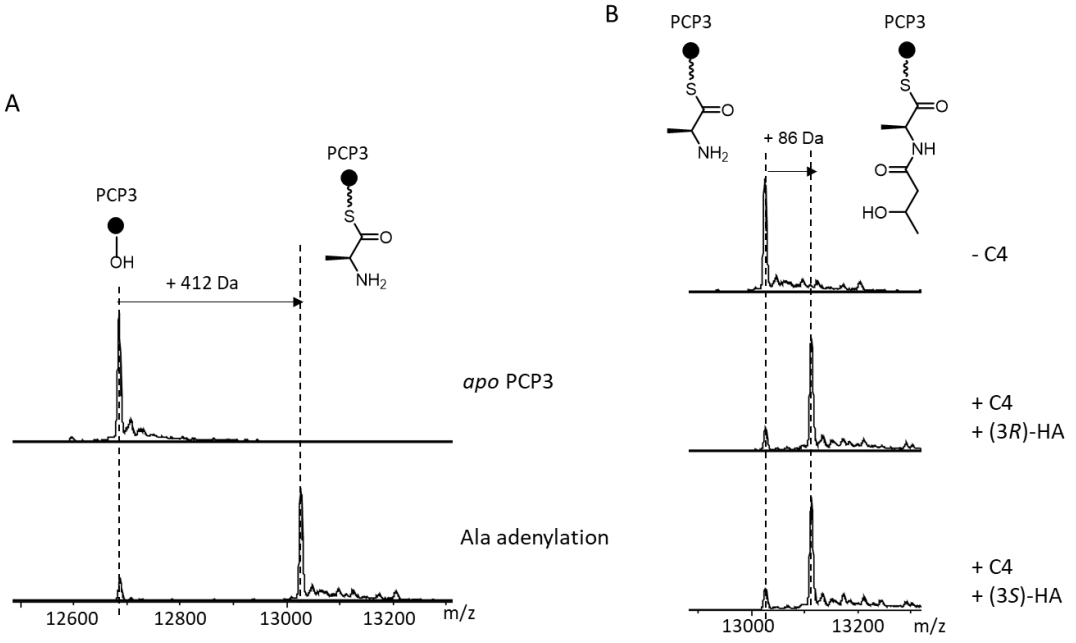


Figure 4.2 LC-MS analysis of the N-acylation activity of the IcoA C4 domain. (**A**) Deconvoluted mass spectra of the *apo* IcoA PCP3 domain (top) and following incubation with Svp, CoA, $MgCl_2$, ATP and L-Ala (bottom). L-Ala-PPant loading was observed as a 412 Da mass shift from *apo* protein. (**B**) deconvoluted mass spectra of the L-Ala-PPant-IcoA PCP3 species (top) and following incubation with the IcoA C4 domain and 3-hydroxylacyl-Pant with *3R* configuration (middle) or *3S* configuration (bottom).

The subsequent condensation reaction was then conducted by incubation of L-Ala-PPant-IcoA PCP3 with the IcoA C4 domain and 3-hydroxybutyryl-pantetheine, which mimics the predicted natural substrate 3-hydroxyoctanoyl-ACP (see 6.6.6). The 3-hydroxyacyl-ACPs are produced exclusively in the *3R*-configuration during fatty acid biosynthesis. To examine whether the IcoA

C4 domain has this proposed stereo specificity towards the fatty acid substrate, both *R* and *S* configured 3-hydroxybutyryl-pantetheine (synthesised by Dr. Matias Rey) were tested in the assay. LC-MS analysis of the intact L-Ala-PPant-IcoA PCP3 following condensation reactions revealed a peak corresponding to the condensed L-Ala-hydroxybutyryl product in both reactions, while no equivalent peak was detected in the control reaction lacking the IcoA C4 domain (Figure 4.2 B).

This result demonstrated the catalytic ability of the IcoA C4 domain to utilise a 3-hydroxyacyl moiety to re-prime the NRPS chain assembly but without any of the predicted stereospecificity. It is possible the promiscuous stereo specificity of the IcoA C4 domain was due to the substrate mimics not being close enough to the natural substrate.

4.2 Characterization of the IcoA C4-A3-PCP3 tri-domain

To probe whether using more accurate substrate mimics would increase the specificity of the IcoA C4 domain, an L-serine and a fatty acid substrate with a longer alkyl chain were required. Due to the instability of L-serine-pantetheine noted in the previous study, direct loading of L-serine by the IcoA A3 adenylation activity was attempted. Instead of separately including the isolated IcoA A3 domain into the assay, the IcoA C4-A3-PCP3 tri-domain was used.

4.2.1 Production of recombinant IcoA C4-A3-PCP3 tri-domain

Construct pET28-IcoA_C4-A3-PCP3 was generated for producing the N-His_6 recombinant IcoA C4-A3-PCP3 tri-domain. As the activity of the IcoA C4 domain within the tri-domain construct was also planned to be monitored via intact protein mass spectrometry, an appropriate negative control was also designed. As introduced in section1.3.2.2, C domains from modular NRPSs contain a conserved catalytic core motif HHxxxDG, and the catalytic role of the second histidine residue as a general base in several systems has been confirmed by mutagenesis studies.[119,118,129] This putative catalytic histidine was therefore located within the conserved HHxxxDG motif of the C4 domain and was mutated to an alanine on the pET28-IcoA_C4-A3-PCP3 construct, via site-directed mutagenesis as detailed in 6.5.2. The N-His_6 recombinant IcoA C4-A3-PCP3 and IcoA C4*-A3-PCP3 (H143A) tri-domains were overproduced and purified as detailed in 6.5.3 and 6.5.4. The

mass and purity of the purified recombinant proteins were confirmed by SDS-PAGE and LC-MS analysis (Figure 4.3). Surprisingly, a certain percentage of the *holo* form of both constructs was detected, suggested the PPase from primary metabolism in *E. coli* BL21(DE3) could recognize and modify the IcoA PCP3 domain. Although this is unusual it was also observed following reconstitution of the valinomycin NRPS in *E. coli.*[155]

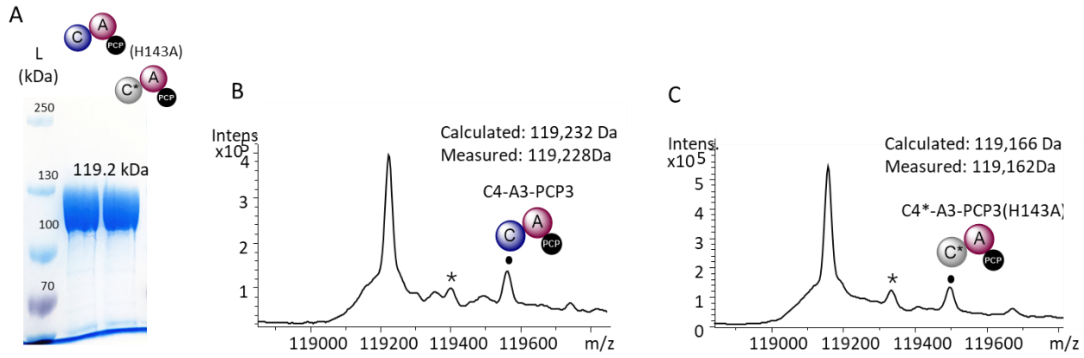
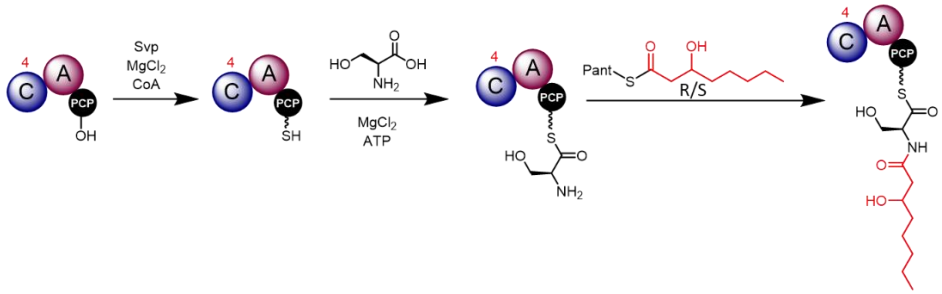


Figure 4.3 SDS-PAGE and LC-MS analysis of purified N-His$_6$-IcoA C4-A3-PCP3 and N-His$_6$-IcoA C4*-A3-PCP3(H143A). (A) 6% SDS-PAGE gel showing the size and purity of N-His$_6$-IcoA_C4-A3-PCP3 (left) and N-His$_6$-IcoA C4*-A3-PCP3(H143A) (right). Deconvoluted mass spectra confirming the exact mass of N-His$_6$-IcoA_C4-A3-PCP3 (-N-Met) **(B)** and N-His$_6$-IcoA_C4*-A3-PCP3(H143A) (-N-Met) **(C)**. Peak labeled as '*' corresponds to spontaneous gluconoylation of the His-tag on fusion proteins with additional 178 Da observed.[147] The peak labeled as '●' refers to the *holo* species with mass addition of 341 Da. For raw mass spectra see appendix.

4.2.2 Characterization of the chain initiation activity of the IcoA C4-A3-PCP3 tri-domain

An *in vitro* assay outlined in Scheme 4.3 using *3R* or *3S*-hydroxyoctanoyl-pantetheine was planned to reconstitute the chain initiating activity of the IcoA C4-A3-PCP3 tri-domain. Both wild type and mutant *apo*-C4-A3-PCP3 tri-domain were converted to their *holo* form by treatment with the PPant transferase, Svp, and coenzyme A, and subsequently incubated with ATP and L-Ser for the adenylation reaction. Intact protein mass spectrometry showed the expected mass shift of 428 Da from the *apo* form to the L-Ser loaded form (Figure 4.4). This demonstrated the proposed serine adenylation activity of the IcoA A3 domain, however a significant amount of the *holo* species was also detected in each instance. This could be the result of the hydroxy group in the

serine residue, without following condensed to fatty acid chain, attacking the thioester bond and thus causing the hydrolysis of serine residue from the IcoA PCP3 domain.



Scheme 4.3 Design of *in vitro* assay for examination of the chain priming activity of IcoA C4-A3-PCP3. The tri-domain is converted to *holo* form by Svp catalyzed PPant unit loading, followed by incubation with L-Ser, $MgCl_2$ and ATP for the A domain catalyzed L-Ser loading reaction and subsequently incubated with *3R* or *3S*-hydroxyoctanoyl-Pant for the C domain catalyzed condensation reaction.

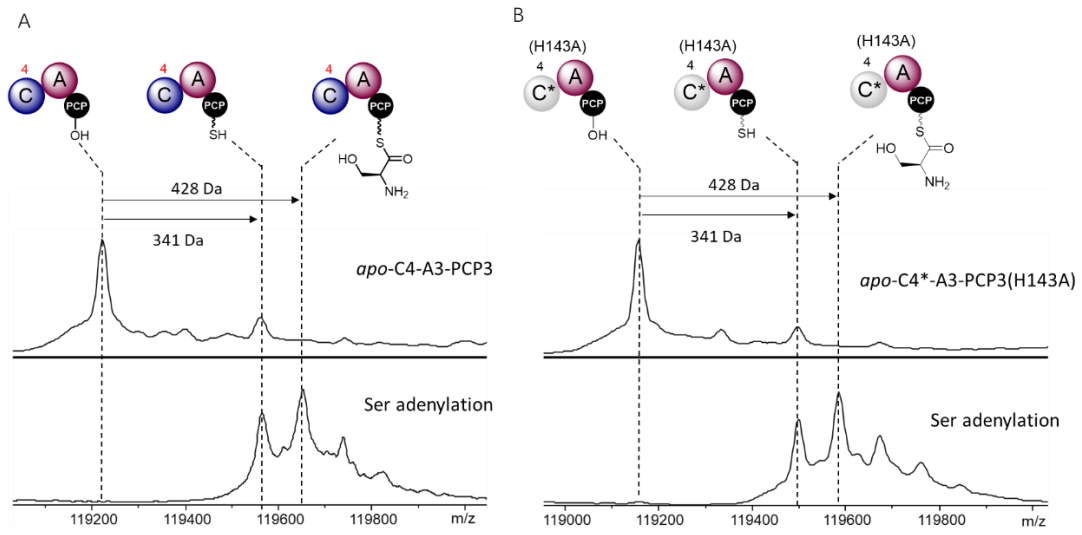
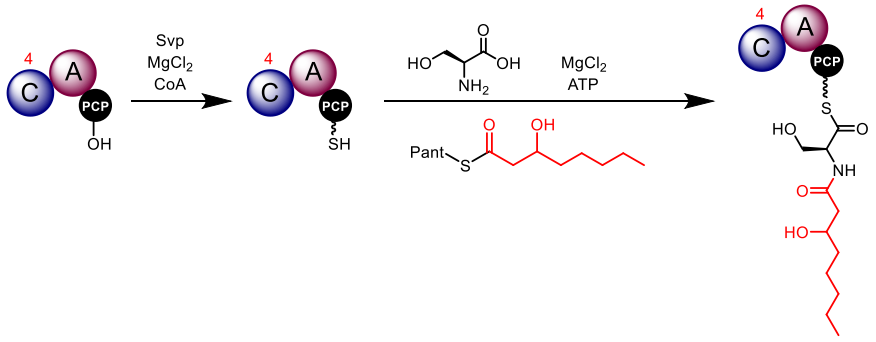


Figure 4.4 LC-MS analysis of the L-serine adenylation activity of A domain in IcoA C4-A3-PCP3 and IcoA C4*-A3-PCP3 (H143A). Deconvoluted mass spectra of the *apo* (top) IcoA C4-A-PCP tri-domain (**A**) and the IcoA C4-A-PCP(H143A) tri-domain (**B**) following incubation with Svp, CoA, $MgCl_2$, ATP and L-Ser(bottom). *Holo* modification and Ser-PPant loading were observed as 341 Da and 428 Da mass shift from *apo* protein respectively

To avoid hydrolysis of the serine residue from the IcoA PCP3 domain to allow for maximum turnover for the following condensation reaction, the activity of the IcoA C4-A3-PCP3 tri-domain was monitored using a one-pot adenylation and condensation reaction outlined in Scheme 4.4. In this assay the *apo*-IcoA C4-A3-PCP3 tri-domain was first converted to its *holo* form, and

subsequently incubated with L-Ser, ATP and either *3R*- or *3S*-hydroxyoctanoyl-pantetheine (see 6.6.7). The control reaction was conducted under the same conditions using the IcoA C4*-A3-PCP3 (H143A) tri-domain.



Scheme 4.4. Optimized *in vitro* one-pot assay for reconstitution of the chain initiating activity of IcoA C4-A3-PCP3. The tri-domain is first converted to *holo* form by Svp catalyzed PPant unit loading, followed by incubation with L-Ser and $MgCl_2$ and ATP *3R* or *3S*-hydroxyoctanoyl-Pant for the one-pot L-Ser adenylation and subsequent condensation reaction.

LC-MS analysis of the intact IcoA C4-A3-PCP3 tri-domain following incubation with *3R*-hydroxyoctanoyl-Pant resulted in a peak corresponding to condensation of L-Ser with 3-hydroxyoctanoyl, with peaks for the *holo*- and L-Ser-loaded forms also observed (Figure 4.5 A). However, no equivalent peak was observed when *3S*-hydroxyoctanoyl-Pant was used as a substrate or in the control reactions using the IcoA C4-A3-PCP3(H143A) tri-domain mutant using either *3R* or *3S* configured substrates (Figure 4.5 B). These results suggested the IcoA C4 domain is responsible for the embedded hydroxyacyl initiating activity and it does harbor the proposed stereospecificity for the *3R*-configured hydroxyoctonyl group.

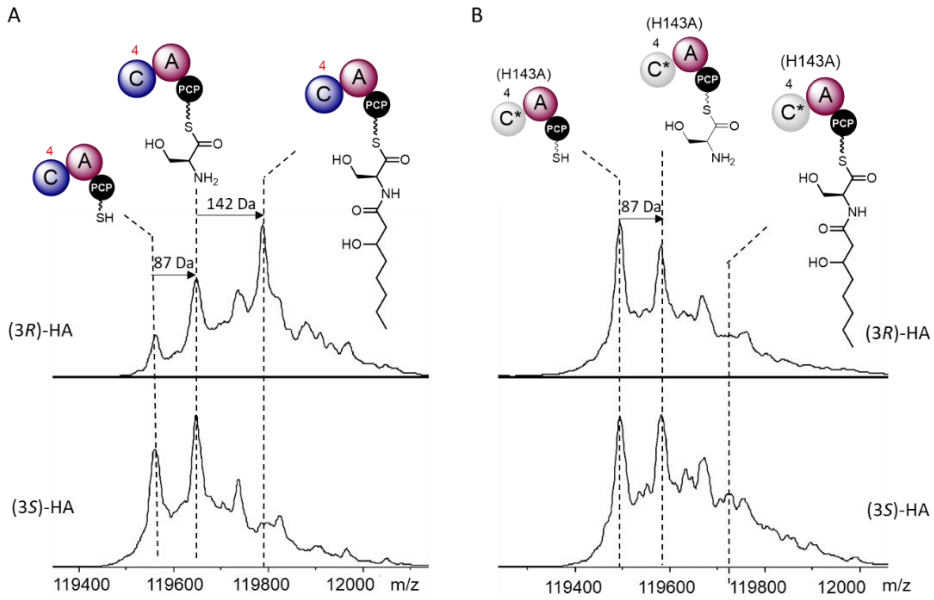
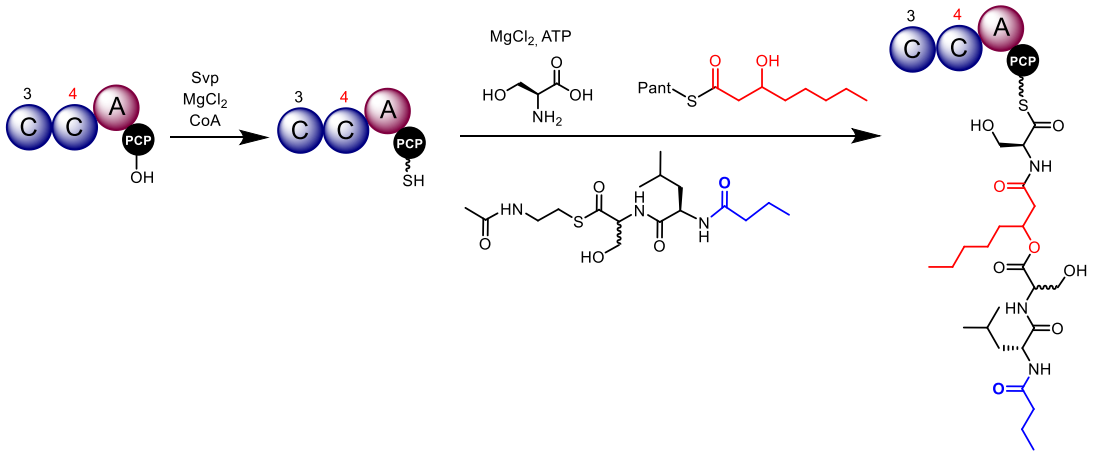


Figure 4.5 LC-MS analysis of *in vitro* reconstitution of chain initiating activity of IcoA C4-A3-PCP3. Deconvoluted mass spectra of (**A**) IcoA C4-A-PCP tri-domain and (**B**) IcoA C4-A-PCP(H143A) tri-domain following incubation with ATP, L-Ser and *3R*-hydroxyoctanoyl-Pant (top) or *3S*-hydroxyoctanoyl-Pant (bottom). L-Ser-PPant loading and subsequent condensation with 3-hydroxyoctanoyl unit were observed as an 87 Da and further 142Da mass shift from the *holo* proteins respectively.

4.3 Characterization of IcoA Module 3

The second condensation event in IcoA Module 3 is proposed to be the IcoA C3 domain catalyzed ester bond formation between the free hydroxyl group of the β-hydroxyl acyl unit and the intermediate tethered to the upstream IcoA PCP2 domain. An *in vitro* one-pot assay (Scheme 4.6.) using the full length IcoA Module 3 and a substrate mimic of the upstream assembled peptide chain, Ser-(D)-Leu-butyryl-SNAC was designed to reconstitute the activity of the whole module *in vitro.*



Scheme 4.5 Design of on-pot *in vitro* assay for reconstitution of the activity of IcoA Module 3. The IcoA Module 3 is first converted to *holo* form by Svp catalyzed PPant unit loading, followed by the one-pot incubation with L-Ser, $MgCl_2$, ATP, *3R* or *3S*-hydroxyoctanoyl-Pant and Ser-(D)-Leu-butyryl-SNAC.

The construct pET28-IcoA_module 3 was generated and the corresponding N-His_6 fusion protein was overproduced and purified as detailed in 6.5.1, 6.5.3 and 6.5.4. SDS-PAGE analysis of the purified N-His_6-IcoA Module 3 showed good purity (Figure 4.6 A) however intact protein mass spectrometry analysis identified 3 peaks in the deconvoluted chromatogram (Figure 4.6 B). The peak with m/z of 169,584 Da, corresponding to the expected N-His_6-IcoA Module 3 (calculated m/z=169,584 Da), was the lowest intensity species, with the major species containing an increase in mass of 594 Da. A further peak was also observed with an additional 178 Da mass compared to the major species, which corresponds to the spontaneous gluconoylation of a His-tag fusion protein. This suggested the major species was also an overproduced N-His_6 fusion protein and was therefore a derivative of N-His_6-IcoA Module 3.

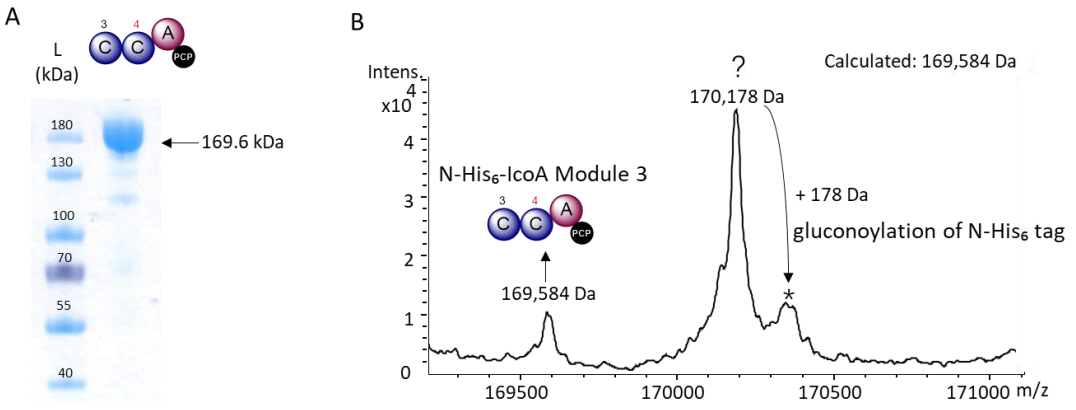


Figure 4.6 SDS-PAGE and LC-MS analysis of purified N-His_6-IcoA Module 3. (A) 6% SDS-PAGE gel showing the size

and purity of N-His_6-IcoA Module 3. **(B)** Deconvoluted mass spectra confirming the existence and the exact mass of N-His_6-IcoA Module 3 (-N-Met) as well as displayed another two peaks. The peak labeled as '?' refers to dominated species with observed m/z=170.178 Da (with additional 594 Da than expected mass of IcoA Module 3). The peak labeled '*' corresponds to the spontaneous gluconoylation of His-tag on the 'major species' with additional 178 Da observed.[147] Row MS spectrum see appendix.

To further characterize the identity of the purified IcoA Module 3 derivative, one-pot phosphopantetheinylation and L-serine adenylation of all three species was attempted using conditions previously described (see 6.6.8). Intact protein mass spectrometry analysis of the reaction showed the expected mass shift could only be observed on the authentic *apo*-IcoA Module 3 species but not the major species (Figure 4.7). This strongly indicated that for the major species, the serine residue of the PCP3 domain which undergoes phosphopantetheinylation was occupied. Given that the IcoA C4-A3-PCP3 tri-domain purified largely in its *holo* form, the detection of this major species (+594 Da) was tentatively explained as follows; the PCP domain of IcoA module 3 was converted to its *holo* form by the PPtase from the *E. coli* primary metabolism and the IcoA A3 domain and IcoA C4 domain subsequently performed their functions utilizing the available intracellular substrate. Addition of the proposed natural substrates of module 3, a serine residue and 3-hydroxydectonoyl unit, would account for a mass shift of 598 Da. It is unclear why this reaction did not occur during overproduction of the IcoA C4-A3-PCP3 tri-domain.

To probe this possibility, an analogue of the proposed substrate of the IcoA C3 domain, Ser-(D)-Leu-butyryl-SNAC, was directly incubated with the purified module 3 to see if a mass shift corresponding to fully assembled product of module 3 could be observed (6.6.8). Unfortunately, a dramatically decreased intensity and resolution of peaks corresponding to module 3 and its derivatives was observed by intact mass spectrometry, likely due to the high concentration of the small molecule SNAC substrate greatly depressing the signal of the ~170 kDa IcoA Module 3.

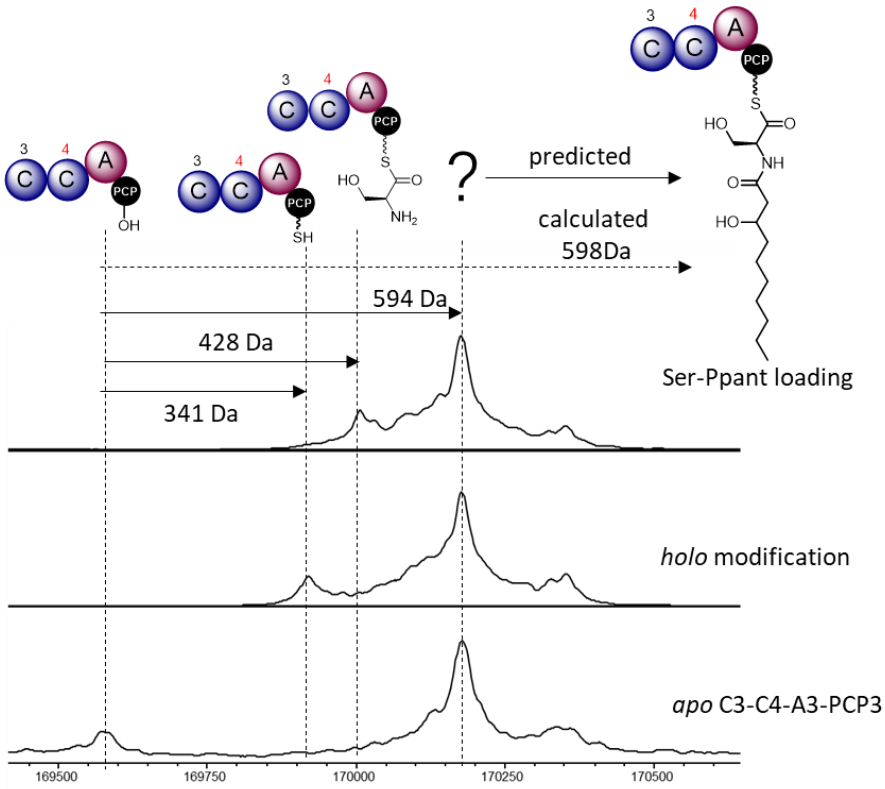


Figure 4.7 LC-MS analysis following *apo* to *holo* modification and L-Serine loading of IcoA module 3 and proposed identity of the unknown derivative of IcoA module 3. Deconvoluted mass spectra of IcoA Module 3 (bottom) following incubation with Svp, CoA, $MgCl_2$, (middle) and additional L-Ser, ATP (Top). *Apo* to *holo* conversion and subsequent loading of L-Ser were observed as a 341 Da and 428 Da mass shift from *apo* IcoA Module 3 respectively. The peak labelled as '?' denotes the 'major species discussed above and was predicted to be IcoA Module 3 with an analogue of the 3-hydroxyldectonyl-serinyl-PPant unit attached on the PCP domain.

4.4 Trials for *in vitro* reconstitution of icosalides derivatives biosynthesis

Since the IcoA C4 domain showed some substrate promiscuity with respect to the chain length of the fatty acid substrate, biosynthesis of novel icosalides derivatives via incorporation of new fatty acid moieties was attempted.

As icosalides are assembled from simple acyl chain and amino acid building blocks, and by a relatively small assembly line, *in vitro* reconstitution of the entire NRPS with unnatural substrate is an attractive approach for the production of novel icosalides derivatives (Figure 4.8). This

approach would also avoid any interference from the cellular natural fatty acid substrates. To first reconstitute the functional icosalide NRPS two strategies have been attempted.

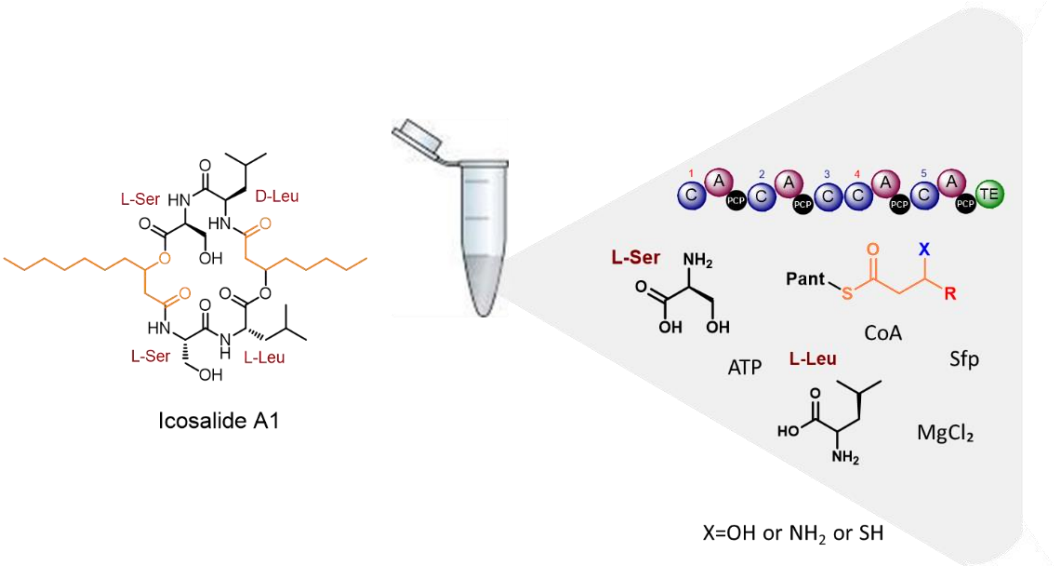


Figure 4.8 Schematic view of *in vitro* reconstituted biosynthesis of novel icosalides derivatives. The required components for *in vitro* biosynthesis of novel icosalides derivatives are shown. The naturally incorporated fatty acid moieties in Icosalide A1 and proposed substrate analogues are highlighted in orange.

4.4.1 Strategy 1: two-piece reconstitution

Successful *in vitro* reconstitution of small or partial assembly lines have been reported,[155,156,157] with a maximal protein molecular weight of ~350 kDa. The icosalide NRPS is a single protein with a molecular weight of ~550 kDa, so two constructs were planned to reconstitute the entire NRPS with the excision site between the IcoA C3 domain and the IcoA C4 domain (Figure 4.9 A). This cut site was chosen because both domains have previously been successfully isolated in soluble form (data for the IcoA C3 domain not shown).

The DNA encoding for IcoA-Part I and IcoA-part II was separately cloned into pET28a (+) via Gibson assembly (Figure 4.9 A) and the corresponding recombinant N-His_6 fusion proteins were overproduced and purified (detailed in 6.5.1, 6.5.3 and 6.5.4). The purified proteins were analysed by SDS-PAGE, showing bands of approximately the correct molecular weight, although a protein standard of sufficiently large molecular weight to confirm this was not available (Figure 4.9 B).

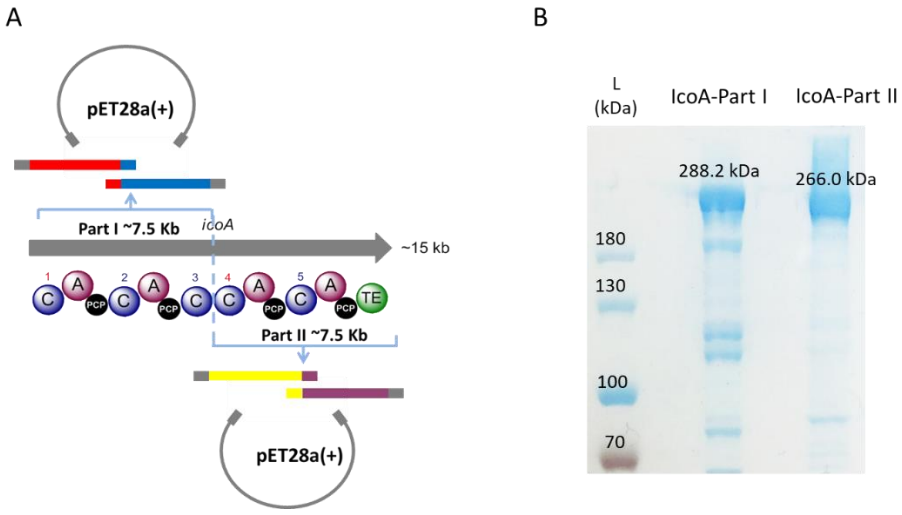


Figure 4.9 *In vitro* reconstitution of IcoA by a two-part strategy. (A) Schematic diagram of the DNA cloning strategy used to create *icoA* expression constructs (B) 6% SDS-PAGE gel analysis of purified recombinant N-His_6-IcoA-Part I and N-His_6-IcoA-Part II.

The two purified proteins were incubated with L-serine, L-leucine, *3R*-hydroxyoctanoyl-PPant, ATP, $MgCl_2$, CoA and SvP (detailed in 6.6.9.). The reaction was then extracted and the extracts we analysed by LC-MS analysis. Unfortunately, the expected enzymatic product was not observed.

4.4.2 Strategy 2: one-piece reconstitution

The failure of the first strategy was very likely due to the linker region between the IcoA C3 and IcoA C4 domains being important for the function of the entire NRPS. Therefore, cloning and expression of the entire IcoA NRPS as a single protein was attempted. Due to the challenges associated with purification of a protein of ~550 kDa molecular weight, the novel icosalides derivatives production was then planned to be attempted by cell free assays following confirming the functionality of the IcoA NRPS.

First, *icoA* (~15 kb) was captured from the genomic DNA by transformation associated recombination (TAR) cloning (detailed in 6.8). The *icoA* fragment was then cloned into pET28a(+) to generate pET28a(+)-*icoA* and the sequence was confirmed by restriction digestion (Figure 4.10)

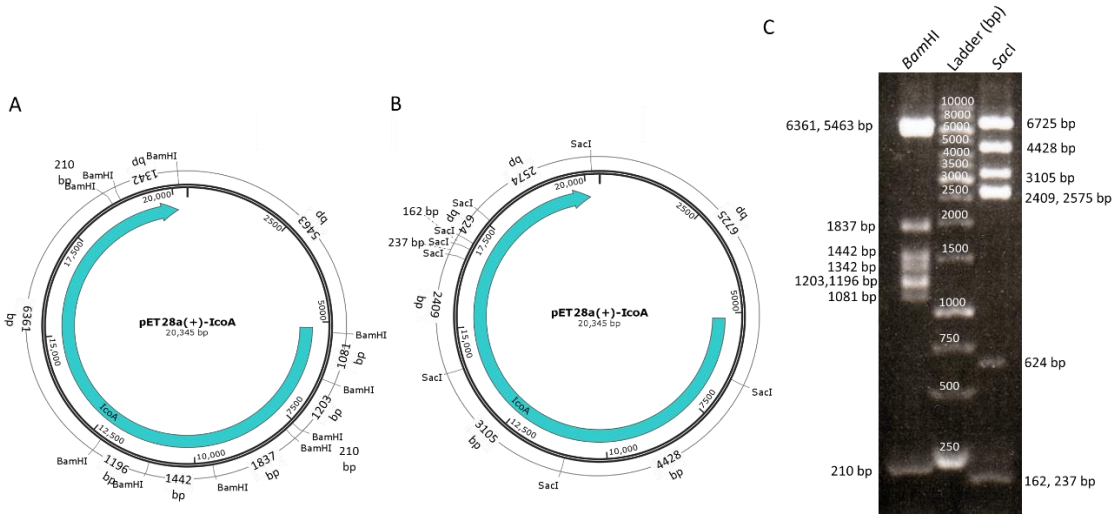


Figure 4.10 Restriction digestion map of the pET28a(+)-*icoA*. Expected restriction map for pET28a(+)-*icoA* digested with *Bam*HI (A) *or Sca*I (B). (C) DNA electrophoresis gel of the restriction digestion results confirming the pET28a(+)-*icoA* construct.

To first confirm the overproduction and functionality of IcoA, pET28a(+)-*icoA* was expressed in *E. coli* BAP1, an engineered *E. coli* strain with NRPS and PKS specific PPtase encoded gene integrated into the genome[158], for monitoring the icosalides production (detailed in 6.9). After 20 hours induction the cells were extracted and LC-MS analysis was used to search for icosalides production. Unfortunately, no icosalides were detected in the extracts.

4.5 Conclusions and future work

In this chapter, *in vitro* biochemical assays were employed to elucidate the biosynthetic reactions catalyzed by module 3 of the icosalide NRPS. The N-acylation activity of the IcoA C4 domain was first demonstrated by incubating the IcoA C4 domain with the PCP3 domain loaded with L-Ala and 3-hydroxybutyryl-pantetheine, which resulted in production of a 3-hydroxybutyryl-alaninyl-IcoA PCP3 species. This represents the first example of an embeded chain initiating C domain. The adenylation and stereospecific chain initiating activity of the IcoA A3-C4-PCP3 tri-domain was reconstituted in *vitro* using natural substrate L-serine and closer substrate mimic 3-hydroxyoctonyl-pantetheine. The observed *3R*- stereoselectivity of IcoA C4 domain towards the 3-hydroxyacyl moiety suggested the proposed origin of this moiety from fatty acid biosynthesis.

The full length IcoA Module 3 was overproduced and purified to investigate the ability of the IcoA C3 domain to catalyse the second condensation reaction of this module. Intact protein mass spectrometry analysis of the recombinant module 3, showed IcoA Module 3 had a higher than expected molecular weight that was close to the additional mass, which would be expected if a serine and a 3-hydroxydectonyl moiety were assembled on the *holo* PCP domain by the activity of the C and A domain. The PCP domain must be loaded since IcoA Module 3 could not be converted to *holo* form or loaded with serine by A domain. Since minor *holo* modification was also observed in the purified IcoA C4-A3-PCP3 tri-domain, this indicated the promiscuity of the PPtase from *E.* coli primary metabolism towards the icosalide NRPS.

In vitro reconstitution of the entire icosalide NRPS was also attempted in order to develop a system that could produce novel icosalides derivatives. A two-piece reconstitution strategy was firstly employed, in which IcoA was purified as two separate recombinant proteins which were then incubated with the required amino acid and fatty acid building blocks and cofactors. However, none of the expected final product was detected. The IcoA encoding gene was then captured by TAR cloning and the entire NRPS was overproduced in *E. coli*. This also did not result in success as no icosalide production was observed in the IcoA overproducing cells.

In the future, the mechanism for the stereoselectivity of the IcoA C4 domain towards the *3R*-hydroxyacyl unit could be probed via a structural study. The substrate scope of the IcoA C4 domain could also be investigated to allow for rational engineering of icosalides derivative. As the proposed promiscuity of PPtase from *E. coli* primary metabolism towards IcoA is quite intriguing, proteomics should be used to confirm the identity of the purified IcoA Module 3 derivative. Optimization of IcoA overproduction conditions and confirmation of the corresponding gene expression should also be performed to complete the in vitro reconstitution of icosalides production experiments. Successful generation of such a system would lead to the ability to access novel icosalides derivatives through feeding precursor analogues.

Chapter 5

Conclusions and future perspectives

Antimicrobial activity-guided and genomics-driven natural product discovery in *Burkholderia gladioli* BCC0238 has resulted in the identification of a polyketide macrolide antibiotic, gladiolin, that exhibits potent activity against *Mycobacterium tuberculosis*, and a set of cyclic lipopeptodilide compounds, icosalides, originally believed to be fungal metabolites. The biosynthetic pathways for both of these metabolites have interesting features that have been explored using various *in vivo* and *in vitro* methodologies in this thesis.

Gladiolin is structurally related to etnangien, isolated from the myxobacterium *Sorangium cellulosum*. However, etnangien suffers from significant instability due to the presence of a conjugated polyene in the side chain, originating from programmed iteration of EtnE Module 5. In comparison, though it has the same domain organization, GbnD2 Module 5 does not iterate, instead recruiting a *trans*-acting enoyl reductase to reduce the double bond generated from a single cycle of chain elongation. Despite their critical structural differences, the *trans*-AT polyketide synthases (PKSs) responsible for the biosynthesis of these compounds are remarkably similar, with 70 out of 71 enzymatic domains identical in the assembly line. This study aimed to identify the ER activity required by GbnD2 Module 5 and to probe the chain length control mechanism in gladiolin biosynthesis.

The presence of an ER domain in module 1 of the gladiolin PKS (the GbnD1 ER1 domain) in place of an ACP domain at the corresponding position in the etnangien PKS, represents the sole difference between the architecture of gladiolin and etnangien PKSs. The GbnD1 ER1 domain was therefore assigned as the first candidate to provide the required enoyl reduction activity in GbnD2 Module 5, in an *inter*-modular manner. Using a combination of *in vivo* in-frame deletion,

point-mutations and *in vitro* biochemical assays, the GbnD1 ER1 domain was shown to only perform enoyl reduction in module 1 but not module 5. An alternative candidate was a *trans*-acting ER, GbnE which is present in gladiolin biosynthetic cluster with its counterpart EntL presenting in the one of etnangien, was originally proposed to provide a *trans*-acting enoyl reduction in module 10 of each assembly line. By applying similar approaches, GbnE was shown to catalyse enoyl reduction specifically in both module 5 and module 10 of the gladiolin assembly line. Furthermore, biochemical evidence was provided for the inherent iterative nature of module 5 of the gladiolin PKS and the ability of GbnE to abolish the iteration of this module. The results also suggested the abolishment of the iteration is facilitated by the protein-protein interaction between GbnE and the ACP domain in this module. In addition, the 'gate-keeping' role of the downstream GbnD2 KS6 domain by preferentially accepting the GbnE processed α,β-saturated intermediate was also biochemically characterized. The results suggest GbnE and the GbnD2 KS6 domain play an important role in conserving the biosynthetic fidelity of the gladiolin biosynthesis, however, the kinetics that facilitate this process require further investigations (as detailed in 3.3).

In order to obtain a complete picture, initially the analogous set of biochemical assays should be conducted using the etnangien EtnE module 5 and the corresponding *trans*-acting ER EtnL. If our hypothesis is correct, the presence of EtnL should have no effect on the ability of EtnE module 5 to iterate (Figure 5.1 A). An even more intriguing study would be to examine whether there is 'cross-talk' between the *trans*-acting ERs (EtnL/GbnE) and the ACP domain in module 5 in etnangien and gladiolin pathways, for example as shown in Figure 5.1 B. This has the potential to shed light on the evolutionary origin of the intricate differences between these two PKS pathways. This work also forms the basis for further studies to uncover the molecular mechanism underlying the specific recruitment of *trans*-acting ERs in *trans*-AT PKS assembly lines, which could ultimately guide future bioengineering efforts. GbnE can be used as a model system to better understand the nature of the specific protein-protein interactions between *trans*-acting ERs and the corresponding ACP domains. This could be achieved using a suite of biochemical and biophysical techniques (Figure 5.1 C), including: 1) bio-layer interferometry analysis of the binding affinity between interacting ACPs and GbnE to provide further kinetic information of this specific

interaction. 2) sequence alignment and phylogenetic analysis of interacting ACP domains and *trans*-acting ERs, coupled with saturation mutagenesis to identify the conserved residue(s) that important for this specific interaction. 3) structural studies of GbnE and interacting ACP domains by crystallography, NMR or carbene footprinting to identify the structural features required for this specific interaction.

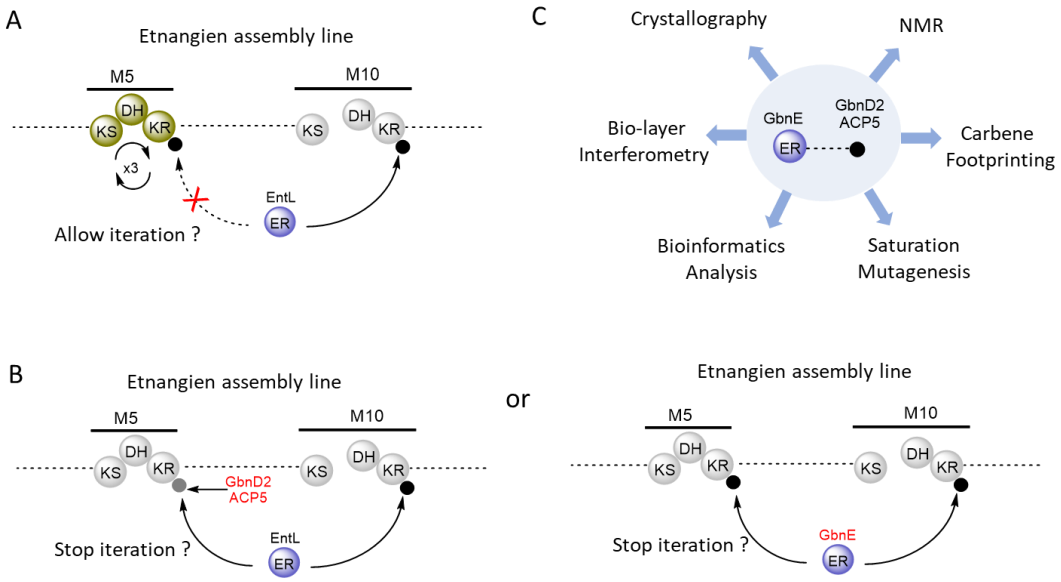


Figure 5.1 Future perspectives of different chain length control mechanisms employed by the gladiolin and etnangien assembly lines. (A) *In vitro* reconstitution of the *trans*-acting ER activity of EntL and iterative activity of module 5 of the etnangien PKS, showing non-interaction between EntL and the EntE ACP5 domain allows the iteration. (B) *In vivo* 'cross-talk' engineering of etnangien biosynthetic pathway using the chain length control logic from gladiolin assembly line. (C) Bioinformatics and biophysics studies of the nature of the specific interaction between GbnE and GbnD2 ACP5 domain.

The NRPS responsible for the biosynthesis of icosalides exhibits an unprecedented domain architecture revealing a starter C domain (IcoA C4 domain) situated directly adjacent to a $^{L}C_{L}$ domain (IcoA C3 domain) suggesting an unusual double chain initiation mechanism for icosalides assembly. The ability of the isolated IcoA C4 domain to load a 3-hydroxyacyl unit by *N*-acylation activity to re-prime the peptide chain assembly was biochemically characterized. The stereoselectivity of the IcoA C4 domain towards the *3R*-configured hydroxyl-acyl moiety was also demonstrated using the IcoA C4-A3-PCP3 tri-domain, confirming the proposed origin of this

moiety as being from primary fatty acid biosynthesis. This represents the first example of an experimentally examined internal starter C domain. Given the relatively modest size of the NRPS *icoA* gene (15 kb), cloning of this gene into a heterologous expression vector was also achieved, with the aim to reconstitute the entire icosalide NRPS *in vitro* which could be used to produce novel icosalides derivatives by feeding unnatural fatty acid substrates. IcoA overproduction in *E. coli* requires further optimization of the expression conditions to confirm the functionality of the reconstituted NRPS as well as for the production of novel derivatives.

Recent studies have shown that it is possible to create 'hybrid' NRPSs via assembly of exchange units (XUs).[159] The XU region of an NRPS is comprised of an A-PCP-C mobile unit, which can be swapped in and out of NRPS pathways. Using this approach in conjunction with a functional *in vitro* IcoA NRPS, in theory every position of the icosalides structure could be altered using this combinatorial biosynthesis approach (Figure 5.2).

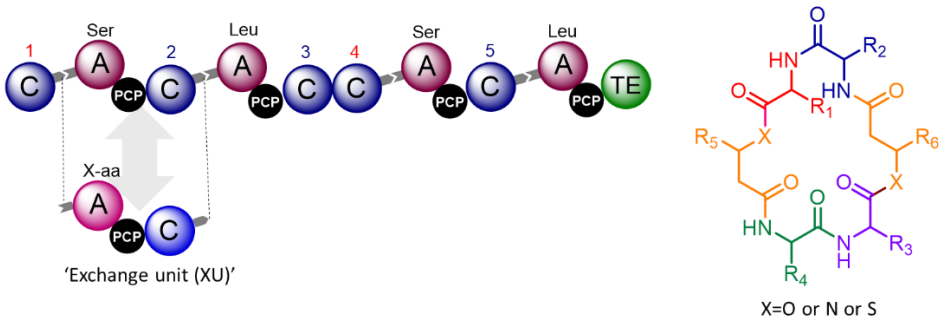


Figure 5.2 Schematic view of rational engineering of novel icosalides derivatives via a combinatorial biosynthesis strategy. Structurally diverse novel icosalides derivatives could be achieved by swapping the 'exchangeable unit' (A-PCP-C tri-domain) to incorporate different amino acid building blocks (X-aa) and feeding novel fatty acid moieties to the starter C domains (number highlighted in red).

Chapter 6

Materials and methods

6.1 Materials

6.1.1 Growth Media

Luria-Bertani (LB) medium was purchased from Sigma Aldrich. In 1 L H_2O, 25 g LB powder was added. For agar plate 15 g bacto agar (Sigma Aldrich) was added.

Tryptic Soy Broth (TSB) was purchased from Sigma Aldrich. In 1 L H_2O, 12 g TSB powder was added. For agar plate 15 g bacto agar (Sigma Aldrich) was added.

Basal salts medium (BSM): In 1 L H_2O, 4.25 g $K_2HPO_4{\cdot}3H_2O$, 1 g $NaH_2PO_4{\cdot}H_2O$, 2 g NH_4Cl, 0.2 g MgSO4·7H2O, 0.012 g $FeSO_4{\cdot}7H_2O$, 0.003 g $MnSO_4{\cdot}H_2O$, 0.003 g $ZnSO_4{\cdot}7H_2O$, 0.003g $CoSO_4{\cdot}7H_2O$, 0.10 g nitrilotriacetic acid, 0.2 g yeast extract and 4 g glycerol were added. pH was adjusted to 7.2. For agar plate 15 g bacto agar (Sigma Aldrich) was added.

M9 minimal medium: In 1 L H_2O, 6 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 1 g NH_4Cl and 10 g sucrose, 1mL 1M $MgSO_4$ and 100ul 1M $CaCl_2$ were added and pH was adjusted to 7.4. For agar plate 15 g bacto agar (Sigma Aldrich) was added.

6.1.2 Buffers

10X Tris/Borate/EDTA (TBE) buffer: 890 mM Tris, 890 mM boric acid, 20 mM pH 8.0 EDTA. Stored at room temperature and diluted to 1X before use.

Loading buffer： 20 mM Tris-HCl, 300 mM NaCl, 20 mM Imidazole, pH 8.0. Sterile filtered and

stored at 4 °C.

Imidazole elution buffer： 20 mM Tris-HCl, 300 mM NaCl, with serial imidazole concentration (50 mM, 100 mM, 200 mM, 300 mM, 500 mM) pH 8.0. Sterile filtered and stored at 4 °C

Storage buffer: 20 mM Tris-HCl, 300 mM NaCl, pH 8.0. Sterile filtered and stored at 4 °C.

5X SDS-PAGE sample buffer: 10% w/v SDS, 10 mM 2-mercaptoethanol, 20% w/v glycerol, 0.2 M Tris-HCl (pH 6.8), 0.05 % w/v bromophenol blue. Stored at -20 °C and diluted to 1X with protein sample when used.

10X SDS-PAGE running buffer: 25 mM Tris, 192 mM glycine, 0.1% w/v SDS. Stored at room temperature and diluted to 1X before use.

Chemically competent cell preparation buffer I (CCC I): 30 mM CH_3CO_2K, 80 mM $CaCl_2$, pH 5.8. Autoclaved and stored at 4 °C.

Chemically competent cell preparation buffer II (CCC II): 30 mM CH_3CO_2K, 80 mM $CaCl_2$, 15% w/v glycerol, pH 5.8. Autoclaved and stored at 4 °C.

Yeast plasmid isolation solution 1: 10% w/v sucrose, 50 mM pH 8.0 Tris-HCl, 10 mM EDTA

Yeast plasmid isolation solution 2: 0.2 M NaOH, 1% w/v SDS

Yeast plasmid isolation solution 3: 3 M CH_3CO_2K, pH 8.0

6.1.3 Reagents and Kits

Antibiotics:

Table 6.1 List of antibiotics

Antibiotic*	Working Conc.	Strain to be applied
Ampicillin	100 μg/mL	*E. coli* carrying recombinant pET151
Kanamycin	50 μg/mL	*E. coli* carrying recombinant pET28a(+) or pRK2013
Trimethoprim	50 μg/mL	*E. coli* carrying recombinant pGPI-SceI
	150 μg/mL	*Burkholderia* carrying recombinant pGPI-SceI

Tetracycline	50 µg/mL	*E. coli* carrying pDAI-SceI
	200 ug/mL	*Burkholderia* carrying pDAI-SceI
Polymyxins	600 U/mL	For triparental mating selection

***** All antibiotics were purchased from Sigma Aldrich

Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs.

MagJET Genomic DNA Kit was purchased from Thermo Fisher Scientific.

GeneJET Gel Extraction Kit was purchased from Thermo Fisher Scientific.

GeneJET Plasmid Miniprep Kit was purchased from Thermo Fisher Scientific.

Champion™ pET151 Directional TOPO® Expression Kit was purchased from Thermo Fisher Scientific.

Champion™ pET SUMO Protein Expression System Kit was purchased from Thermo Fisher Scientific.

Gibson Assembly® Master Mix was purchased from New England Biolabs.

GeneArt™ Seamless Cloning and Assembly Kit was purchased from Thermo Fisher Scientific.

Q5® Site-Directed Mutagenesis Kit was purchased from New England Biolabs.

6.1.4 Bacterial strains

Table 6.2 List of bacterial strains

Name	**Description**	**Source or ref.**
Escherichia coli		
E. coli TOP10	*mcrA*, Δ(*mrr-hsd*RMS-*mcrBC*), Phi80*lacZ(del)M15*, Δ*lacX74*, *deoR*, *recA1*, *araD139*, Δ(*ara-leu*) 7697, *galU*, *galK*, *rpsL(SmR)*, *endA1*, *nupG,* host for general cloning	Invitrogen

E. coli BL21 (DE3)	*E. coli B dcm ompT hsdS*($r_B^-m_B^-$) *gal,* host for recombinant protein expression	Invitrogen
E. coli BAP1	*E. coli* BL21 (DE3) *ΔprpRBCD (sfp)*, host for recombinant protein expression	reference[158]
E. coli SY327	*araD, Δ(lac pro) argE(Am) recA56 rif^R nalA λpir* , donor strain for conjugation between *E. coli* and *Burkholderia*	reference[160]
E. coli HB101/ pRK2013	*F$^-$ Lambda$^-$ araC14 leuB6(Am) DE(gpt-proA)62 lacY1 glnX44(AS) galK2(Oc) recA13 rpsL20(strR) xylA5 mtl-1 thiE1 hsdS20(r_B^-, m_B^-),* carries pRK2013, helper strain for conjugation between *E. coli* and *Burkholderia*	reference[161]
Burkholderia		
B. gladioli BCC1622	Wild type, gladiolin and icosalides producing strain	Mahenthiralingam/Challis labs
B. gladioli BCC0238	Wild type, gladiolin and icosalides producing strain	Mahenthiralingam/Challis labs
Saccharomyces		
S. cerevisiae VL6-48N	*MATα trp1-Δ1 ura3-Δ1 ade2-101 his3-Δ200 lys2 met14 cir°*, host for TAR cloning	ATCC MYA-3666

6.1.5 Vectors

Table 6.3 List vectors

Name	Description	Source or ref.
pGPI-SceI	*ori*$_{R6K}$ *Tp*^{R} *mob*$^+$ carries I-SceI cut site, for gene deletion in *Burkholderia*	reference[144]
pDAI-SceI	*ori*pBBR1 *Tet*^{R} *mob*$^+$ P_{dhfr} FLAG epitope, carries the *I-SceI*,	reference[144]
pMLBAD	*pBBR1 ori araC-PBAD Tp*^{R} *mob*$^+$, for gene expression in *Burkholderia*	reference[149]
pET28a(+)	*pBR322 ori Kan*^{R} *T7lac f1 N-His*$_6$ *C-His*$_6$, for recombinant protein expression in *E. coli*	Novagen
pET151-TOPO®	*pBR322 ori Amp*^{R} *T7lac N-His*$_6$*-V5,* carries TOPO-cloning site, for N-His$_6$-V5 recombinant protein expression in *E. coli*	Thermo Fisher Scientific
pET-SUMO	*pBR322 ori Kan*^{R} *T7lac N-His*$_6$*-SUMO,* carries TA-cloning site, for N-His$_6$-SUMO recombinant protein expression in *E. coli*	Thermo Fisher Scientific
pCAP1000	*pUC ori ARSH4/CEN6 Kan/Neo*^{R} *TRP1 φC31int-attP oriT (RP4) pAHD-URA3,* for gene cluster capturing	Challis lab

6.2 General DNA manipulation

6.2.1 Genomic DNA isolation

Preparation of genomic DNA for PCR was achieved using MagJET Genomic DNA Kit.

Preparation of genomic DNA for TAR cloning was achieved on a large scale using the following method. Cells were harvested from a 5 mL *Burkholderia* overnight culture by centrifugation (4000 xg, 10 mins) and the pellet was washed with 5 mL SET buffer. The cell pellet was then re-

suspended in 500 µL SET buffer and 60 µL 10% w/v SDS and 14 µL proteinase K (20 mg/mL) was added. The cell resuspension was incubated at 55 °C for 1 h followed by addition of 200 µL NaCl (5 M) and thorough mixing of the solution. The mixture was then supplemented with 500 µL PCI (Phenol/chloroform/isoamylol=25:24:1) solution, mixed thoroughly and centrifuged at 12,000 rpm for 10 min. The supernatant was transferred to a fresh Eppendorf tube and 500 µL chloroform was added, mixed thoroughly and centrifuged at 12,000 rpm for 10 min. The supernatant was transferred to a fresh Eppendorf tube and 0.6 CV isopropanol was added to allow the genomic DNA precipitation. The precipitate DNA was centrifuged at 12,000 rpm for 10 min and washed with 70 % v/v ethanol. The pellet was dry in air and finally dissolved in 100 µL sterile water.

6.2.2 Polymerase chain reaction (PCR)

Q5® Hot Start High-Fidelity DNA polymerase was used for DNA fragment PCR and OneTaq® Hot Start DNA polymerase was used for colony PCR screening and TA cloning PCR. A typical reaction was set up is detailed in Table 6.4.

Table 6.4 Components of typical PCR reactions

Component	**Volume**
Q5® Hot Start High-Fidelity 2X Master Mix/ OneTaq® Hot Start Quick-Load 2X Master Mix	12.5 µL
10 µM Forward primer	1.25 µL
10 µM Reverse primer	1.25 µL
DNA template	variable
DMSO	1-2µL*
Nuclease-Free Water	To 25µL

*4% or 8% DMSO was added depending on the GC content of target region of DNA template.

Thermocycling conditions was set up as in Table 6.5 using Master Cycler nexus GX2 (Eppendorf):

Table 6.5 Thermocycling conditions for a routine PCR

Step		**Temperature**	**Time**
Initial denaturation		98/95 °C*	3 min
	Denaturation	98/95 °C*	30 s
30 cycles	annealing	58-70 °C	30 s
	Extension	72 °C	30 s/kb or 1 min/kb**
Final extension		72 °C	10 min
Hold		4 °C	-

* 98 °C or 95 °C was applied at denaturation step when using Q5® Hot Start High-Fidelity DNA polymerase or OneTaq® Hot Start DNA polymerase respectively.

** 30 s/Kb or 1 min/Kb speed was applied at extension step when using Q5® Hot Start High-Fidelity DNA polymerase or OneTaq® Hot Start DNA polymerase respectively.

6.2.3 Fragment DNA purification

DNA products from PCR reaction or plasmid DNA digestion were separated on 0.7% agarose gel electrophoresis in TBE buffer and gel slices containing desired DNA fragments were excised and purified using GeneJET Gel Extraction Kit. Purified DNA fragments were dissolved in 30-100 µL sterile deionized water and the resulting concentration was measured on a NanoDrop™ Lite Spectrophotometer.

6.2.4 DNA restriction digestion and ligation

Typical DNA restriction enzyme digestions were set up as detailed in Table 6.6.

Table 6.6 Components of DNA restriction digestion reactions

Component	**Volume/Amount**
DNA	1 μg
10X CutSmart buffer	5 μL
Restriction enzyme(s)*	1μL (10 units)
Sterile water	To 50 μL

* 1μL for each enzyme was added when double digestion applied.

The reaction was mixed gently and incubated at 37 °C for 1.5 - 3 h followed by incubation at 65 °C for 5 min to inactivate the enzymes.

Purified digested DNA fragments and linearized vector plasmid DNA were ligated using T4 DNA ligase. A typical DNA ligation was set up as Table 6.7.

Table 6.7 Components of typical DNA ligation reactions

Component	**Volume/Amount**
Linearized vector DNA	50 ng
Fragment DNA	3 x of linearized vector molar concentration
10X T4 DNA ligae buffer	2 μL
T4 DNA ligase	1μL
Sterile water	To 50 μL

The reaction was mixed gently and incubated at room temperature for overnight and stored at -20 °C before transformation or electroporation.

6.2.5 TOPO® cloning

TOPO cloning within this work was achieved using Champion™ pET151 Directional TOPO®

Expression Kits. A 4 nucleotide CACC- overhang was added to the 5' of forward primer for primer design to allow directional cloning into pET151. Insert fragment DNA was amplified using Q5® Hot Start High-Fidelity DNA polymerase and purified by gel extraction before the cloning. A typical TOPO cloning reaction is detailed in Table 6.8.

Table 6.8 Components of TOPO® cloning reactions

Component	**Volume/Amount**
Purified fragment DNA	0.5-4 µL*
Salt solution	1 µL
TOPO® vector	1 µL
Sterile water	To 6 µL

* Variable volume of fragment DNA was added to reach the fragment DNA: TOPO® vector molar ratio as 0.5:1 to 2:1.

The reaction was mixed gently and incubate for 1 h at room temperature and then transformed into E. *coli* TOP10 chemical competent cells.

6.2.6 TA® cloning

TA cloning within this work was achieved using Champion™ pET SUMO Protein Expression System Kit. No special primer design was required for TA cloning. Insert fragment DNA was amplified using OneTaq® Hot Start DNA polymerase to give the both 5' and 3' A overhang and purified by gel extraction before the cloning. The typical TA cloning reaction was set as in Table 6.9.

Table 6.9 Components of typical TA® cloning reactions

Component	**Volume/Amount**
Purified fragment DNA	1-7 µL *
10X Ligation buffer	1 µL

pET-SUMO vector	1 µL (25 ng)
T4 DNA ligase	1µL
Sterile water	To 10 µL

* Variable volume of fragment DNA was added to reach the fragment DNA: TOPO® vector molar ratio as 1:1.

The reaction was mixed gently and incubate for 1 h at room temperature and then transformed into E. *coli* TOP10 chemical competent cells.

6.2.7 Gibson assembly and GeneArt assembly

Multiple DNA fragment assembly into a vector within this study were achieved using Gibson Assembly® Master Mix or GeneArt™ Seamless Cloning and Assembly Kit. For primers design, 20 bp was set as length of overlapping region. Before assembly reaction, the insert fragment DNA was amplified using Q5® Hot Start High-Fidelity DNA polymerase and purified by gel extraction and vector was linearized by restriction digestion and purified by PCR purification. The typical assembly cloning reaction was set as in Table 6.10 and Table 6.11.

Table 6.10 Components of typical Gibson assembly reactions

Component	**Volume/Amount**	
Linearized vector	(10 µL in total)	100 ng
Purified Fragments		Variable*
2X Gibson Assembly Master Mix	10 µL	
Sterile water	To 20 µL	

*Variable amount of each fragment was added to reach each fragment: vector molar ratio as 3:1, and total amount of fragments and vector are need to be between 0.03-0.2 pmol to give the best cloning efficiency.

The reaction was mixed gently and incubate for 2 h at 50 °C and stored at -20 °C before

transformed into E. *coli* TOP10 chemical competent cells.

Table 6.11 Components of typical GeneArt assembly reactions

Component	**Volume/Amount**
Linearized vector	100 ng
Purified Fragments	Variable*
5X Reaction buffer	4
Sterile water	to 18 µL
10X enzyme mix	2 µL**

*Variable amount of each fragment was added to reach each fragment: vector molar ratio as 2:1.

**10X enzyme mix must be added at last step and needed be returned back to -80 °C immediately.

The reaction was mix gently and incubate for 30 min at temperature and placed on ice less than 5 min and immediately proceed to the transformation step.

6.2.8 DNA purification for electroporation

All the above cloning reactions (described in 6.2.4 - 6.2.7) were purified by ethanol precipitation before electroporation. The ethanol precipitation was as follows; 1/10 reaction volume of 3 M pH 5.2 sodium acetate was added to the reaction mixture followed by addition of 2.5X reaction volume of ethanol and mixed thorough. The mixture was frozen at -20 °C for 20 min, and was centrifuged at 13,000 rpm, 4 °C for 30 min. The supernatant was decanted carefully and the pellet was washed with 70% ethanol and dried. 15 µL of water was added to dissolved the purified DNA.

6.2.9 Transformation and electroporation

Preparation of competent cells

For chemical transformation of *E. coli* TOP10 or BL21(DE2), chemically competent cells were first

made. A single colony of *E. coli* TOP10 or BL21(DE2) was inoculated in 5 mL LB and grown at 37 °C, 180 rpm overnight. Subculture the overnight culture 0.5 mL/50 mL LB in 250 mL flask at 37 °C, 180 rpm shaker for around 3.5 h to reach the OD_{600} value of 0.4-0.6. The culture was then cooled on ice for 30 min to stop cells growing and cells were harvested by centrifuging for 10 min at 3000 rpm and 4°C. The supernatant was removed and the cell pellet was resuspended in 30 mL CCC I buffer and kept on ice for 1 h. Cells were then harvested by centrifuging for 10 min at 3000 rpm and 4°C and supernatant was removed. The cell pellet was washed with 30 mL CCC I buffer again and resuspended in 1.5 mL CCC II buffer and aliquoted into 80 µL. Aliquots were flash froze and stored at -80 °C.

Similar cell growing and harvest process were applied when preparing the electrocompetent cells except the cell pellet was washed and resuspended in 15% glycerol, and aliquoted into 80 µL.

Chemical transformation

Frozen chemical competent cells were thawed on ice, and 50-100 ng of plasmid DNA or 8 µL of cloning reaction mixture was added and gently mixed. Mixture was left on ice for 30 min and heat shocked at 42 °C for 60 s and cooled on ice for 2 min. 500 µL was added and the cell resuspension was incubated at 37 °C for 1 h. 100 µL culture was spread on appropriate selective LB agar plate and incubated at 37 °C for overnight.

Electroporation

Frozen electrocompetent cells were thawed on ice, and 50-100 ng of plasmid DNA or 10 µL of purified cloning reaction (described in 6.2.8) was added and gently mixed. The mixture was left on ice for 5 min before transferred to a pre-cooled 1 mm gap electroporation cuvette (Bio-Rad). The electroporation was performed with the Gene Pulser II (Bio-Rad) set at 2.5 kV, 25 µF and 200 Ω. Immediately after electroporation 500 µL LB was added to the cuvette and cell resuspension was incubated at 37 °C for 1 h. 100 µL culture was spread on appropriate selective LB agar plate and incubated at 37 °C for overnight.

6.2.10 Plasmid DNA mini-prep

5 mL *E. coli* overnight culture was used for a typical plasmid DNA mini-prep using GeneJET Plasmid Miniprep Kit. Purified plasmid DNA was dissolved in 50 µL sterile deionized water and concentration was measured by NanoDrop™ Lite Spectrophotometer.

6.3 Genetic manipulation of *Burkholderia*

6.3.1 Targeted *in-frame* gene deletion in *Burkholderia*

The *in-frame* deletion of *genes* from *Burkholderia gladioli* BCC1622 was conducted using the following methodology.[144] First, the deletion construct was generated. Here, ~1 kb homologous region that sit upstream and downstream of the gene of interest were amplified, purified and cloned into pGPI-SceI vector (using the DNA restriction digest and ligation or Gibson Assembly method (see 6.2.4 and 6.2.7)) and transferred into *E. coli* SY327 by electroporation. The resulting plasmid was isolated from candidate *E. coli* SY327 colonies and confirmed by restriction digestion and sequencing. The plasmid was mobilized into *Burkholderia gladioli* BCC1622 via tri-parental mating (described in 6.3.3). Single trimethoprim resistance colonies were selected as candidates of the single cross-over mutant followed by colony PCR screening. pDAI-SceI was introduced into the single crossover mutant strain by tri-parental conjugation and exconjugants were selected for with tetracycline resistance (to select for pDAI-SceI) and trimethoprim sensitive (indicating the loss of the integrated plasmid). Candidates with $Tc^{R}Tp^{S}$ phenotype were either wild-type or double crossover mutants which can be discriminated by PCR screening. The *in-frame* deletion was confirmed by sequencing. The mutant was grown on minimal medium M9 to remove pDAI-SceI and confirmed by the loss of tetracycline resistance. Related primers and expected PCR products size for the deletion mutants: *B. gladioli* BCC1622_Δ*gbnD1*_ER1, *B. gladioli* BCC1622_Δ*gbnE, B. gladioli* BCC1622_Δ*gdsB*_ER, *B. gladioli* BCC1622_Δ*gbnE*Δ*gdsB*_ER are listed in Table 6.12.

6.3.2 Site-directed mutagenesis in *Burkholderia*

Site-directed mutagenesis of *gbnD1*_ER1 on *Burkholderia* genomic DNA was achieved via the *in-*

frame knocking in of the mutated *gbnD1_*ER1 encoded segment into BCC1622_*ΔgbnD1_*ER1. The *gbnD1_*ER1 fragment, flanked by the same homologous arms used for *gbnD1_*ER1 *in-frame* deletion, was first subcloned onto pCR-Blunt vector to generate pCR-Blunt-*gbnD1_*ER1. Mutagenesis of G388S/G399P was then conducted using the Q5 site directed mutagenesis kit following the protocol in section 6.5.2. to generate pCR-Blunt-*gbnD1_*ER1*. The mutated *gbnD1_*ER1 segment with flanking arms were then cloned into pGPI-SceI vector using restriction site of *Xba*I and *Eco*RI to generate pGPI-*gbnD1_*ER1*and transferred into *E. coli* SY327 by electroporation. In-frame knocking-in *gbnD1_*ER1* followed the same procedure as in-frame deletion in section 6.3.1.1, via mobilizing pGPI-*gbnD1_*ER1* into BCC1622_*ΔgbnD1_*ER1 and selection for double crossover mutants. Related primers for pCR-Blunt-*gbnD1_*ER1* construction, double crossover mutant PCR screening and expecting PCR products size are listed in Table 6.12.

6.3.3 Gene complementation in *Burkholderia*

For *gbnE* complementation in *Burkholderia*, the arabinose-inducible P_{BAD} promoter carrying expression vector pMLBAD[149] was used to construct pMLBAD-*gbnE*. *GbnE* was amplified with the primers listed in Table 6.12 below, purified and cloned into pMLBAD via Gibson assembly (described in 6.2.7). The confirmed pMLBAD-*gbnE* was electroporated into *E. coli* SY327 (described in 6.2.9) and mobilized into *B. gladioli* BCC1622_Δ*gbnE* via tri-parental mating (described in 6.3.3). Exconjugants with Tp^R phenotype were picked as *B. gladioli* BCC1622_Δ*gbnE*::*gbnE* candidates and were confirmed by PCR with primers listed in Table 6.12.

Table 6.12 Overview of constructs used in gene *in-frame* deletion, site-directed mutagenesis and complementation in *Burkholderia*

Construct (use)	**Primer (restriction site)**	**Product size***
pGPI-*gbnD1_*ER1 (*gbnD1_ER1*	5'-For: CTGTCTAGATTGTCGTCCTCGTCGC (*Xba*I) 5'-Rev: CAGAAGCTTGAGCGCGTGCACGGTA (*Hin*dIII)	1008 bp

in-frame deletion)	3’-For: GCAAAGCTTGTGGTGATCCGCCATCG (*Hin*dIII)	975 bp
	3’-Rev: GCCGAATTCCACGCCATGCACGGGA (*Eco*RI)	
	Screening-For: CGCGGCTCTACGATCATGTGACGAT	MT: 449 bp
	Screening--Rev: GCCCGACATGCCCACCACGGCGAT	WT:1361 bp
pGPI-*gbnE*	5’-For: ATATCTAGAAGCAAGCGGTCCGACAGG (*Xba*I)	928 bp
(*gbnE*	5’-Rev: ATAAAGCTTATCGGCGTTGCCAAGCGA (*Hin*dIII)	
in-frame deletion)	3’-For: TATAAGCTTTCGATGGCCGATTCGGGC (*Hin*dIII)	941 bp
	3’-Rev: ATAGAATTCTGGCGGGTCGAGCCCATC (*Eco*RI)	
	Screening-For: CTGTTGCAAGCGCACGATTAGC	MT: 457 bp
	Screening-Rev: GACCAGGGGCAGTCGCAGCGA	WT: 1756 bp
pGPI-*gdsB*_ER	5’-For: *gatcccaagcttcttctag*AGCACTGTTCGAGAAGTAC	1020 bp
(*gdsB*_ER	5’-Rev: *gttgggtgaa*GTAGGCGTACTTGAGCTT	
in-frame deletion)	3’-For: *gtacgcctac*TTCACCCAACTGCTCCG	1023 bp
	3’-Rev: *aacggctgacatgggaattc*GAAGATCGACGAGAG	
	Screening-For: CAGGGCGTCGAGACATTCC	MT: 606 bp
	Screening-Rev: CCTTGTCGCGCACGGATT	WT: 1851 bp
pCR-blunt-*gbnD1*_ER1*	For: CTGTCTAGATTGTCGTCCTCGTCGC (*Xba*I)	2865 bp
	Rev: GCCGAATTCCACGCCATGCACGGGA (*Eco*RI)	
(*gbnD1*_ER1 mutagenesis)	G388S/G399P-For: GACCGCCACC*agtcct*TGCGGTCTCGCCG	6389 bp
	G388S/G399P-Rev: GCCCGACATGCCCACCACGGCGAT	
pGPI-*gbnD1*_ER1*	Screening-For: CTGTTGCAAGCGCACGATTAGC	WT: 449 bp

(*gbnD1*_ER1 mutagenesis)	Screening-Rev: GACCAGGGGCAGTCGCAGCGA	MT: 1361 bp
pMLBAD-*gbnE* (gbnE complementation)	For: *tgggctagcaggaggaattc*ATGGCAATGATTACCGCA Rev: *tgggctagcaggaggaattc*ATGGCAATGATTACCGCA Screening primers are the same as For and Rev primers	1398 bp

*MT denotes mutant, WT denotes wild type. Italic letters denote the overlapping regions used for Gibson assembly. Underlying letters denote the restriction sites. Italic and underlying letters denote region for site-directed mutagenesis.

6.3.4 General tri-parental mating procedures

Transferring plasmids into *Burkholderia* strain from *E. coli* was achieved via the tri-parental mating process. Donor strains were *E. coli* SY327 carrying recombinant pGPI-SceI or pMLBAD or pDAI-SceI and helper strain was *E. coli* HB101 carrying pRK2013. In each tri-parental mating experiment, both donor strain and helper strain were inoculated into 5 mL LB medium containing appropriate antibiotics respectively and incubated overnight in a 37 °C shaking incubator. The host *Burkholderia* strain was grown in 5 mL TSB medium overnight in a 30 °C shaking incubator. Cells from the above cultures were harvested and resuspended in 5 mL LB medium containing 10 mM $MgCl_2$ and this procedure was repeated. 100 µL of each culture was mixed, resulting in a 300 µL mixture, and 100 µL of the mixture was plated onto sterile nitrocellulose disks which were placed onto LB agar containing 10 mM $MgCl_2$. The plate was allowed to dry and incubated overnight at 30 °C in duplicate. The cells were then removed from the disc and resuspended in 0.9% NaCl. A serial dilution from 0 to -5x was carried out and 100 µL from each dilution were plated onto selective agar plate, TSA (Tp 150 µg/mL, PMX 600 U/mL) was applied when donor strain was *E. coli* SY327 carrying recombinant pGPI-SceI or pMLBAD, TSA (Tc200 µg/mL, PMX 600 U/mL) was applied when donor strain was *E. coli* SY327 carrying pDAI-SceI. This selects against *E. coli* and wild type *Burkholderia* cells. The plates were incubated for 48-72 h at 30 °C to allow the exconjugants to grow. 100 µL of helper strain, donor strain and *Burkholderia* culture were plated

onto corresponding selective plate respectively as negative controls and these should remain clear with no growth.

6.4 *Burkholderia* metabolite extraction and LC-MS analysis

6.4.1 Metabolite production and extraction

Production of gladiolin, icosalides and their respective derivatives was achieved by growth of *B. gladioli* BCC1622 wild type or related gene deletion and mutagenesis mutants at 30 °C on solid minimal medium containing glycerol as a sole carbon source (BSM). For restoring gladiolin production in the gene complementation mutant, L-arabinose with final concentration of 0.2% or 2% was used as inducer and added to the BSM medium. Following incubation for 3 days the growth was removed by scraping with a spatula from the agar plate, any remaining bacteria were killed with chloroform vapor, and the agar was cut into ~0.5 cm cubes. Approximately 4 mL of acetonitrile was added per 1 plate of agar and the resulting mixture was left for 1 h. The extract was filtered and 100 µL of the resulting supernatant was then subjected for LC-MS analysis.

6.4.2 Metabolite LC-MS analysis

LC-MS analysis of Burkholderia metabolite was performed using a Dionex UltiMate 3000 UHPLC connected to a Zorbax Eclipse Plus column (C18, 100 × 2.1 mm, 1.8 µm) coupled to a Bruker MaXis IMPACT ESI-QTOF mass spectrometer. The UHPLC elution profile is detailed in Table 6.13. The mass spectrometer was operated in positive ion mode with a scan range of 50-3000 *m/z*. Source conditions were: end plate offset at −500 V; capillary at −4500 V; nebulizer gas (N_2) at 1.6 bar; dry gas (N_2) at 8 L min^{-1}; dry temperature at 180 °C. Ion transfer conditions were: ion funnel RF at 200 Vpp; multiple RF at 200 Vpp; quadrupole low mass at 55 *m/z*; collision energy at 5.0 eV; collision RF at 600 Vpp; ion cooler RF at 50–350 Vpp; transfer time at 121 s; pre-pulse storage time at 1 s. Calibration was performed with 1 mM sodium formate through a loop injection of 20 µL at the start of each run.

Table 6.13 UHPLC elution profile metabolite LC-MS analysis

Time (min)	Water / 0.1% FA (%)	MeCN / 0.1% FA (%)	Flow Rate (mL/min)
0.0	80	20	0.2
5.0	80	20	0.2
35.0	0	100	0.2
40.0	0	100	0.2
45.0	80	20	0.2

6.5 Recombinant protein overproduction and purification

6.5.1 Cloning of recombinant protein expression constructs:

The constructs used for recombinant protein expression in *E. coli* are cloned via different methods, and are listed in Table 6.14. All insert DNA fragments were amplified from *B. gladioli* BCC0238 gDNA and purified as described in 6.2.2 and 6.2.3. The detailed cloning procedures are described in 6.2.4 -6.2.7.

Table 6.14 Overview of primers and cloning methods for constructs used for recombinant protein expression

Construct (Inset size)	Primer*	Cloning method
pET151-*gbnD1*_ER1-ACP1 (1639 bp)	For: *CACC*CCGGAACCGGCTGAGGCG Rev: *TCA*CGGCGCATGCGTCTCGCT	TOPO cloning®
pET151-*gbnD1*_ER1 (1183 bp)	For: *CACC*GCTGCGCCTTCGCCTTCC Rev: *TCA*CGGCGCATGCGTCTCGCT	TOPO cloning®

pET151-*gbnD2*_ACP5	For: *CACC*CCGAATCCAGCGTCGCTTTCG	TOPO
(376 bp)	Rev: *TCA*GACCTCCGCCTTGGCCGTCTC	cloning®
pET28a(+)-*gbnD1*_ACP3	For: *ATAT*CATATGACCGCCACGGTTCCGACC	Digestion &
(419 bp)	Rev: *ATAT*GAATTCTCAGATCGCGGGCCCGGAGG	ligation
pET28a(+)-*gbnD2*_ACP4	For: *ATAT*CATATGCCGGCAACGCAACGCGCC	Digestion &
(566 bp)	Rev: *ATAT*GAATTCTCACGCGCTCGCTGTCGCAGG	ligation
pET28a(+)-*gbnD4*_ACP10	For: *ATAT*CATATGGCCCAGGCGAAGCCGGTT	Digestion &
(467 bp)	Rev: *ATAT*AAGCTTTCAGCCCTCGGGCAGGCCAA	ligation
pET151-*gbnE*	For: *CACC*ATGGCAATGATTACCGCATCA	TOPO
(1402 bp)	Rev: *CTA*TTGCCTCTGGAGTTGAAGTGA	cloning®
pET-SUMO-*gbnE*	For: ATGGCAATGATTACCGCA	TA cloning
(1398 bp)	Rev: CTATTGCCTCTGGAGTTGAAGTGA	
pET28a(+)-*gbnD2*_Module 5	For: *ATAT*GCTAGCGCGGGCGCGAGGACGATG	Digestion &
(4643 bp)	Rev: *ATAT*AAGCTTTCAGACCTCCGCCTTGGCCGT	ligation
pET28a(+)-*gbnD2*_KS6-ACP6a	For: *ATT*CATATGACGGCCAAGGCGGAGGTC	Digestion &
(2652 bp)	Rev: *ATA*AAGCTTTCAGACATCCTGGTC	ligation
pET28a(+)-*gbnD2*_Module 6	For: *ATA*GCTAGCGACGGCGAGACGGCCAAG	Digestion &
(3141 bp)	Rev: *ATA*AAGCTTTCACGGGGCGTTTCCGGTCTT	ligation
pET151-*icoA*_C4	For: *CACC*GATGTCTCGCGTCTGGTCGA	TOPO
(1461 bp)	Rev: *TCA*GGTAGCCTCGGGCGTGCGCTT	cloning®
pET151-*icoA*_PCP3	For: *CACC*AGATACGAAGCACCGCAAGGC	TOPO
(260 bp)	Rev: *TCA*GGCGCGACGCTTCGGCTT	cloning®

pET28a(+)-*icoA*_C4-A3-PCP3 (3222 bp)		For: *ATA*CATATGGATGTCTCGCGTCTGGT Rev: *ATA*AAGCTTTCACGCGACGCTTCGGCT	Digestion & ligation
pET28a(+)-*icoA*_Module3 (4593 bp)		For: *GCC*CATATGCGCGCTCATGATCTCGTT Rev: *ATA*AAGCTTTCACGCGACGCTTCGGCT	Digestion & ligation
pET28a(+)-IcoA-partI	IcoA-partI-a (3831 bp)	For: *gccgcgcggcagcca*ATGTGGTTTTCGCAGCAACG Rev: tggaagcgTGGGCGTATCGACGTCGC	Gibson assembly
	IcoA-partI-b (3971 bp)	For: *tacgccca*CGCTTCCATTCGGCCTCG Rev: *ccagtcatgctagcca*CAGACGCGAGACATCCGG	
pET28a(+)-IcoA-partII	IcoA-partII-a (3571 bp)	For: *gccgcgcggcagcca*GATGTCTCGCGTCTGGTC Rev: *cttcctcctcg*CATAGACGCAACCATTCG	Gibson assembly
	IcoA-partII-b (3713 bp)	For: *cgtctatg*CGAGGAGGAAGCCACGCA Rev: *ccagtcatgctagcca*TCATGCGTTCTCGGTAGCC	

*upper italic denotes un-annealing region, lower italic denotes overlapping regions used in Gibson assembly, underlined denotes restriction sites.

6.5.2 Site-directed mutagenesis

Mutant constructs listed in Table 6.16 were generated using Q5 Site-directed Mutagenesis Kit. Firstly, the linear mutant plasmids were amplified using corresponding primers, noting the amount of DNA template was limited to 1-25 ng/μL. The PCR products were analyzed by taking 5 μL of crude PCR product to DNA electrophoresis for confirming product yield and purity. Following Kinase, Ligase and DpnI (KLD) treatment was set up as in Table 6.15.

Table 6.15 Components of Q5® Site-Directed Mutagenesis reactions

Component	**volume**
PCR product	1 µL
2X KLD reaction buffer	5 µL
10X KLD Enzymes Mix	1 µL
Nuclease free water	3µL

The reaction was left at room temperature for 5 min and immediately transformed to *E. coli* TOP10 chemically competent cells.

Table 6.16 Overview of the primers and mutations for mutant constructs

Mutant Construct	**Primers***	**Mutation**
pET-SUMO-*gbnE*	For: TTCGGGCGGT*gtT*ACCGATCGC	CAT→GTT
(H198V)	Rev: TCGGCCTCCACGCAGATGTC	(His→Val)
pET28a(+)-*gbnD2*_Module 6	For: ACTGGGCATCGAT*g*CGGTGGTCG	TCG→GCG
(S941A)	Rev: TCGGCGAACGGCGTGTCG	(Ser→Ala)
pET28a(+)-*icoA*_C4-A3-PCP3	For: GCGTTCGCCT*gc*tGTGCAGAACGAC	CAT→GCT
(H143A)	Rev: AACGCCACCAGCACGCTA	(His→Ala)

* The lower case and italic letters indicate un-annealing nucleotide(s).

6.5.3 Recombinant protein overproduction

For general recombinant protein overproduction procedures, single transformants of *E. coli* BL21(DE3) carrying the recombinant plasmid were inoculated in 10 mL LB with the appropriate antibiotic and grown at 37 °C in a 180 rpm shaker for overnight as seed culture. A 1L scale LB flask with appropriate antibiotic was sub-cultured from overnight seed cultures at 37 °C in a 180 rpm shaker for ~3.5 h to reach an approximate OD_{600} value of 0.6-0.8, and then induced with 0.5 mM

IPTG, and was left shaking at 15°C for 16 - 20 h. For large scale protein production, followed by FPLC purification and size exclusion purification, the expression volume was scaled up to 3 - 5 L. For large size proteins (≥100 kDa, e.g. N-His_6-IcoA_C4-A3-PCP3 and N-His_6-IcoA Module 3) or proteins with solubility issues (e.g N-His_6-SUMO-GbnE), or proteins with both features (e.g. N-His_6-GbnD2 Module 5), the culture was induced with 0.1-0.25 mM IPTG to slow the overproduction process. Following expression, the cells were harvested by centrifuging 5000 rpm, 4 °C for 15 min and resuspended in 10 mL of loading buffer per litre of culture.

6.5.4 Recombinant protein purification

The harvested cell was lysed using a Constant Systems Cell disruptor. The lysate was clarified by centrifugation at 17,000 rpm, 4 °C for 30 min and the supernatant was filtered with 0.22 µM syringe filter. The resulting supernatant was subjected to different purification procedures (as listed in Table 6.17) depending on the soluble recombinant protein yield and purity. The following purification procedures are described in 6.5.4.1-6.5.4.3.

Table 6.17 Overview of tag type, size and purification method for recombinant protein

Recombinant protein	**Fusion tag**	**Size (kDa)**	**Purification method***
GbnD1 ER1-ACP1	N-His_6-V5	61.3	Bench-top IMAC
GbnD1 ER1	N-His_6-V5	44.9	Bench-top IMAC
GbnD2 ACP5	N-His_6-V5	16.8	Bench-top IMAC
GbnD1 ACP3	N-His_6	16.6	Bench-top IMAC
GbnD2 ACP4	N-His_6	21.2	Bench-top IMAC
GbnD4 ACP10	N-His_6	17.5	Bench-top IMAC
GbnE	N-His_6-V5	54.8	FPLC IMAC followed by size exclusion
	N-His_6-SUMO	64.4	FPLC IMAC followed by size exclusion
GbnE(H198V)	N-His_6-SUMO	64.3	FPLC IMAC followed by size exclusion

GbnD2 Module 5	N-His_6	161.7	FPLC IMAC followed by size exclusion
GbnD2 KS6-ACP6a	N-His_6	95	Bench-top IMAC
GbnD2 Module 6 (S941A)	N-His_6	111.6	FPLC IMAC followed by size exclusion
IcoA C4	N-His_6-V5	57.4	Bench-top IMAC
IcoA PCP3	N-His_6-V5	12.7	Bench-top IMAC
IcoA C4-A3-PCP3	N-His_6	119.4	Bench-top IMAC
IcoA C4-A3-PCP3 (H143A)	N-His_6	119.3	Bench-top IMAC
IcoA Module3	N-His_6	169.7	Bench-top IMAC
IcoA_PartI	N-His_6	288.2	Bench-top IMAC
IcoA_PartII	N-His_6	266.0	Bench-top IMAC

* IMAC referred to immobilized metal-ion affinity chromatography

6.5.4.1 Bench-top immobilized metal-ion affinity chromatography

The filtered supernatant was loaded on a 1 mL HiTrap® HP Chelating Column (GE Healthcare), pre-charged with 100 mM $NiSO_4$ and equilibrated with loading buffer. The loaded column was washed with 15 mL loading buffer to remove any non-specifically binding protein impurities, followed by serial elution with 5 mL 50 mM, 3mL 100 mM, 3 mL 200 mM, 3 mL 300 mM imidazole elution buffer. Fractions were analyzed by SDS-PAGE (described in 7.5.5) and target protein containing fractions were collected and concentrated by Viva-spin concentrator (GE Healthcare) with an appropriate MWCO according to the recombinant protein size (listed in Table 6.17). The concentrated fractions were washed with 10 CV storage buffer and aliquoted and stored at -80 °C.

6.5.4.2 Fast protein liquid chromatography (FPLC)

FPLC purification was performed with ÄKTA pure (GE Healthcare) connected with 5mL HiTrap® HP pre-charged column and Fraction Collector (GE Healthcare) F9-C. After loading with filtered

lysate, the column was washed with loading buffer until the UV_{280} absorption trace was flat followed by elution over a gradient of 0 - 100% 500 mM imidazole elution buffer in 100 mL. 2mL fractions were collected and analyzed by SDS-PAGE. Target protein-containing fractions were collected and concentrated to 2 mL by Viva-spin concentrator (GE Healthcare) with an appropriate MWCO according to the recombinant protein size (listed in Table 6.17) and subjected to the following size exclusion purification.

6.5.4.3 Size exclusion chromatography

Size exclusion purification was performed on ÄKTA pure (GE Healthcare) connected with HiLoad 16/600 Superdex 200 pg column (GE Healthcare) and Fraction Collector (GE Healthcare) F9-C. 2 mL nickel affinity chromatography purified protein was loaded on pre-equilibrated column with storage buffer. The column was then eluted over 1.2 CV with storage buffer and 2 mL fractions were collected and analyzed by SDS-PAGE. Pure target protein containing fractions were collected and concentrated by Viva-spin concentrator (GE Healthcare) with an appropriate MWCO according to the recombinant protein size (listed in Table 6.17) and aliquoted and stored at -80 °C.

6.5.5 SDS-PAGE analysis of protein

SDS-PAGE was used to analyse the size of recombinant protein overproduction, yield and purity of purification fractions. Protein samples were mixed with 5X SDS-PAGE loading buffer and loaded onto different percentage Tris-glycine SDS-polyacrylamide gels (recipe in Table 6.18) depending on the size of protein, 15% for ≤20 kDa, 10% for 20-80 kDa, 6% for ≥ 80 kDa. The gel was run in SDS-PAGE running buffer using Bio-Rad mini gel electrophoresis system. PageRuler Prestained Protein Ladder (Thermo Sientific) was used as reference to estimate the size of protein. Gel was stained in InstantBlue™ Protein Stain (Expedeon).

Table 6.18 Recipe of Tris-glycine SDS-polyacrylamide gels

Component	Volume of component (mL)			
	5 mL Resolving gel			2 mL
	6% gel	10% gel	15% gel	Stacking gel
H_2O	2.6	1.9	1.1	1.4
30% acrylamides mix	1.0	1.7	2.5	0.33
1.5 M Tris*	1.3	1.3	1.3	0.25
10% SDS	0.05	0.05	0.05	0.02
10% ammonium persulfate	0.05	0.05	0.05	0.02
TEMED	0.004	0.002	0.0002	0.002

*pH 8.8, 1.5 M Tris was used for resolving gel and pH 6.8, 1.5 M Tris was used for stacking gel.

6.6 *In vitro* biochemical assays

6.6.1 General PPTase-catalysed phosphopantetheinyl loading ACP assays

6.6.1.1 CoA substrate loading assays[162]

100 µM ACP was incubated with 2 µM Svp (PPtase from *Streptomyces verticillus*[162]), 10 mM $MgCl_2$ and 500 µM CoA or CoA thioester in storage buffer in a total volume of 50 µL. The reaction was allowed to proceeded at room temperature for 30 min. The completion of the loading reaction was monitored by LC-MS (section 6.7).

6.6.1.2 Pantetheine substrate loading assays[163]

100 µM ACP was incubated with 2 µM Svp, 2 µM pantethenate kinase (PanK), 2 µM phosphopantethein adenylytranferase (PPAT) and 2 µM dephosphocoenzyme A kinase (DPCK) in

the presence of 10 mM $MgCl_2$, 5 mM ATP and 1 mM acyl-pantetheine substrate in storage buffer in total volume of 50 µL. The reaction was allowed to proceeded at room temperature for 1h. The completion of loading reaction was monitored by LC-MS (section 6.7).

6.6.2 Characterization of GbnD1 ER1 domain

6.6.2.1 GbnD1 ACP-ER1 di-domain enoyl reduction assay

GbnD1 ACP-ER1 was first loaded with a 2-butenoyl-PPant unit using 2-butenoyl-CoA according to procedures described in 6.6.1.1. 50 µM crotonyl-GbnD1 ACP-ER1 was then incubated with 500 µM NADPH/NADH in storage buffer in a final volume of 50 µL. The reaction was allowed to proceeded at room temperature for 5h. The control reaction was conducted under the same conditions without adding cofactor NADPH/NADH. The enoyl reaction was monitored LC-MS analysis (section 6.7).

6.6.2.2 Stand-alone GbnD1 ER1 enoyl reduction assay

From an ACP bound substrate

GbnD1 ACP5 was loaded with 2,4-hexadienoyl-PPant unit using 2,4-hexadienoyl pantetheine as described in 6.6.1.2. Then, 50 µM 2,4-hexadienoyl-GbnD2 ACP5 was incubated with 50 µM GbnD1 ER1 and 500 µM NADPH or NADH in storage buffer in a final volume of 50 µL. The reaction was allowed to proceed at room temperature for 5 h. A control reaction was conducted under the same conditions without adding GbnD1 ER1. The enoyl reaction was monitored by injection ion generated from intact protein species in LC-MS analysis (section 6.7).

From a pantetheine substrate

50 µM GbnD1 ER1 was incubated with 1 mM 3-methyl-butenoyl-Pant in storage buffer in a final volume of 50 µL. The reaction was allowed to proceeded at room temperature for 5 h. Control reaction was conducted at same condition with boiled GbnD1 ER1. The completed reaction was extracted with (2x 100 µL) of acetonitrile and the extracted products was dissolved in 100 µL

methanol and subjected for LC-MS analysis (section 6.4.2).

6.6.3 Characterization of GbnE

GbnD2 ACP5 (or GbnD1 ACP3, GbnD4 ACP10 and GbnD5 ACP12 in corresponding assays) was first loaded with 2,4-hexadienoyl-PPant unit using 2,4-hexadienoyl pantetheine using procedures described in 6.6.1.2. 50 μM 2,4-hexadienoyl-ACP was than incubated with 50 μM GbnE or GbnE(H198V) and 500 μM NADPH or NADH in storage buffer in a final volume of 50 μL. The reaction was allowed to proceed at room temperature for 5 h. Control reaction was conducted under the same conditions but lacking GbnE. The enoyl reaction was monitored by injection ion generated from intact ACP species in LC-MS analysis (section 6.7).

6.6.4 Characterization of GbnD2 Module5

6.6.4.1 Characterization of the iterative nature of GbnD2 Module5

50 μM GbnD2 Module 5 was incubated with 1 mM 2,4-hexadienoyl-NAC in storage buffer in a final volume of 50 μL at room temperature for 3h followed by concentration to 150 μM using a Viva-spin concentrator (GE Healthcare). Meanwhile, GbnD2 ACP5 domain was loaded with a malonyl unit using malonyl-CoA following the procedures in 6.6.1.1. 50 μM acylated-GbnD2 Module 5 was incubated with 100 μM malonyl-GbnD2 ACP5, and 800 μM of NADPH in storage buffer to a final volume of 50 μL. The reaction was allowed to proceed at room temperature for 20h and monitored over time (2 h and 5 h and 20 h). Control reactions were conducted under the same conditions using GbnD1 ACP3 and GbnD4 ACP10 respectively instead of GbnD2 ACP5. For detection of iteration, intact ACP species were monitored by LC-MS analysis (section 6.7).

6.6.4.2 Probing the ability of GbnE to stop the iteration of GbnD2 Module5

The pre-acylation of GbnD2 Module 5 and malonylation of GbnD2 ACP5 followed the same conditions and procedures as in 7.6.4.1. 50 μM GbnE or GbnE(H198V) was then mixed with 50 μM acylated-GbnD2 Module 5, 100 μM malonyl-GbnD2 ACP5 and 800 μM of NADPH in storage

buffer to a final volume of 50 µL. The reaction was allowed to proceeded at room temperature for 5 h and monitored by LC-MS analysis of intact GbnD2 ACP5 species (section 6.7).

6.6.5 Characterization of GbnD2 Module 6 (S941A)

6.6.5.1 GbnD2 Module 6 (S941A) and GbnD2 ACP5 domain acyl transferring assays

GbnD2 ACP5 was first loaded with 4-hexenoyl-PPant or 2,4-hexadienoyl-PPant units using the corresponding pantetheine substrates using procedures described in 6.6.1.2. 100 µM GbnD2 ACP5 with substrate loaded was then incubated with 300 µM GbnD2 Module 6(S941A) in storage buffer in a final volume of 50 µL. The reaction was allowed to proceed at room temperature for 8h. To detect the acyl transferring from GbnD2 ACP5 to KS domain in GbnD2 Module 6(S941A), intact GbnD2 ACP5 species were monitored over time course (2 h ,4 h, 6 h and 8 h) by LC-MS analysis (section 6.7).

6.6.5.2 GbnD2 Module6 (S941A) elongation assay

50 µM GbnD2 Module 6(S941A) was first incubated with 1 mM 4-hexenoyl- or 2,4-hexadienoyl-NAC respectively in a final volume of 50 µL. The reaction was allowed to proceed at room temperature for 3h. For confirmation of acylation, intact GbnD2 Module 6 (S941A) species was monitored by LC-MS analysis. 50 µM acylated GbnD2 Module 6 (S941A) was then incubated with 1 µM Svp, 500 µM Malonyl-CoA, and 10 mM $MgCl_2$ in storage buffer in a final volume of 50 µL. The reaction was allowed to proceeded at room temperature for 5h. To detect the elongation the intact GbnD2 Module 6 (S941A) species was monitored by LC-MS analysis (section 6.7).

6.6.6 Characterization of IcoA C4 domain

IcoA PCP3 was first loaded with L-Ala PPant unit using *L*-Ala pantetheine as described in 6.6.1.2. 100 µM L-Ala-IcoA PCP3 was incubated with 100 µM IcoA C4 and 1 mM of *3R*-hydroxyoctanoyl pantetheine or *3S*-hydroxyoctanoyl pantetheine in storage buffer in a final volume of 50 µL. The reaction was allowed to proceed at room temperature overnight. For detection of condensation,

intact IcoA C4 species was monitored by LC-MS analysis (section 6.7).

6.6.7 Characterization of IcoA C4-A3-PCP3 tri-domain

IcoA C4-A3-PCP3 tri-domain was converted to its *holo* form as described in 6.6.1.1. 100 µM *holo*-IcoA C4-A3-PCP3 was then incubated with 500 µM L-Ser, 500 µM ATP, 10 mM $MgCl_2$, 1 mM of (*3R*)-hydroxyoctanoyl pantetheine or *3S*-hydroxyoctanoyl pantetheine in storage buffer in a final volume of 50 µL and the reaction was allowed to proceed for 6h at room temperature. The control reaction was conducted using the same conditions while using IcoA C4-A3-PCP3(H143A) instead of IcoA_C4-A3-PCP3. For detection of condensation, intact wild type and mutant IcoA C4-A3-PCP3 species was monitored by LC-MS analysis (section 6.7).

6.6.8 Characterization of IcoA Module 3

Holo modification of IcoA Module 3 was following the procedures decribed in 6.6.1.1. Serine adenylation of IcoA Module 3 was conducted by incubation of 100 µM *holo* IcoA Module 3 with 500 µM L-Ser, 500 µM ATP, 10 mM $MgCl_2$ in storage buffer in a final volume of 50 µL and the reaction was allowed to proceed for 3h at room temperature. For detection of holo modification and serine adenylation, intact IcoA Module 3 species was monitored by LC-MS analysis (section 6.7).

6.6.9 Two-piece *In vitro* reconstitution of icosalides biosynthesis

IcoA_Part I and IcoA_Part II were first converted to *holo* form by incubation of 40 µM recombinant protein with 2 µM Svp, 200 µM CoA, 5 mM $MgCl_2$ in the total volume of 50 µL at room temperature for 30 min. 20 µM of each holo form protein were than incubated with 5 mM L-serine, 5 mL L-leucine, 1 mM *3R*-hydroxyoctonyl-Pant and 1mM ATP in a total volume of 100 µL at room temperature for 5 h. The completed reaction was extracted with ethyl acetate (2 x 100 µL) and the solvent was evaporated. The extracted product was dissolved in 100 µL acetonitrile and analysed by LC-MS analysis (section 6.4.2).

6.7 LC-MS analysis of proteins and *in vitro* intact protein assays

All proteins and *in vitro* assays were analysed on a Bruker MaXis II electrospray ionisation time-of-flight mass spectrometer (ESI-TOF-MS), connected to a Dionex 3000 RS UHPLC fitted with an ACE C4-300 RP column (100 x 2.1 mm, 5 µm, 30 °C). The UHPLC elution profile is detailed in Table 6.19.

The mass spectrometer was operated in positive ion mode with a scan range of 200–3000 *m/z*. Source conditions were: end plate offset at −500 V; capillary at −4500 V; nebulizer gas (N_2) at 1.8 bar; dry gas (N_2) at 9.0 L min^{-1}; dry temperature at 200 °C. Ion transfer conditions were: ion funnel RF at 400 Vpp; multiple RF at 200 Vpp; quadrupole low mass at 300 *m/z*; collision energy at 8.0 eV; collision RF at 2000 Vpp; transfer time at 110.0 µs; pre-pulse storage time at 10.0 µs.

Table 6.19 UHPLC elution profile for intact protein LC-MS analysis

Time (min)	Water / 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

6.8 Full length icosalide NRPS encoded gene cloning

6.8.1 Capturing plasmid construction

The capturing plasmid pCAP1000-Icos was constructed from the optimized capturing vector pCAP1000 (generated by Dr. Chuang Huang) with the homologous flanking arms and counter selective marker pADH-URA3 encoded fragments assembled via GeneArt assembly as described

in 6.2.7 (primers listed in Table 6.20).

Table 6.20 Overview of primers for construct pCAP1000-*icoA*

PCR product	**Primer ***	**Product size**
icosA-5'-arm	For: *ttctacaaagatcgactagtcatatg*TGGTTTTCGCAGCAACG	1109 bp
	Rev: *gttgttgtag*GTGACCAGGCCATTGGCGAG	
icosA-3'-arm	For: *gcaaaactaa*AATACGAAGCGCCGCAAGGC	1121 bp
	Rev: *caggtacctcaagtctcgaggctagc*TCATGCGTTCTCGGTAGCCA	
pADH-URA3	For: *ggcctggtcac*CTACAACAACTAAGAAAAT	1543 bp
	Rev: *gcttcgtatt*TTAGTTTTGCTGGCCGCATC	

* Italic letters denote overlapping region used in GeneArt assembly.

6.8.2 Yeast transformation

Burkholderia gladioli BCC238 genomic DNA was isolated as described in 6.2.1. Purified genomic DNA was digested with *Apa*LI and capturing plasmid pCAP1000-Icos was digested with *Pme*I. Both the digestions were purified by ethanol precipitation method described in 7.2.8. 2 µg of digested genomic DNA and 50 ng of linearized pCAP1000-Icos were used for yeast spheroplast transformation. The yeast spheroplast preparation and transformation were conducted according to previously reported protocol.[164]

6.8.3 Candidate yeast colonies screening

To identify candidate yeast colonies harboring cluster captured pCAP1000 plasmid (pCAP1000-*icoA)* colonies were picked and grown on a selective plate at 30 °C overnight and then screened by colony PCR. Specifically, the cells were scooped and resuspended in 200 µl yeast plasmid isolation 1 buffer containing 0.1% v/v 2-mercaptoethanol and 0.2 mg/mL zymolyase 20T. The suspension was incubated at 37 °C for 2 h and then mixed with 400 µl solution 2 by inversion.

300 µl solution 3 was added and the mixture was centrifuged for 10 min at 13000 rpm. The supernatant transferred was then mixed with equivalent volume isopropanol to allow the DNA precipitation. The precipitated DNA was collected by centrifuging for 10 min at 13000 rpm and the pellet was dried in air and dissolved in 100 µl water. 1 µl of the isolated DNA was used for PCR screening using IcoA_C4-For and IcoA_C4-Rev (listed in Table 6.14).

6.8.4 *IcoA* containing plasmid DNA extraction and confirmation

The PCR screened positive candidate colony was grown in 10 mL selective liquid medium at 30 °C shaker for 20-30 h. The cells were collected and plasmid DNA was isolated by similar procedures as described in section 6.8.3, except after addition of solution 3 and centrifuging the supernatant was transferred and extracted with 500 µL PCI(Phenol/chloroform/isoamylol=25:24:1) solution and centrifuged at 13000 rpm for 10 min. The resulting supernatant was then mixed with 500 µL isopropanol to allow the DNA precipitation and centrifuged at 13000 rpm for 10 min. The pellet was then washed with 70 % ethanol and dried in air and finally dissolved in 20 µl water. 10 µl of isolated plasmid DNA was transferred into *E. coli* TOP10 by electroporation for replication (described in 6.2.9). The transformants were grown in 5 mL LB and plasmid DNA was isolated by mini-prep kit. The cluster containing plasmid pCAP1000-Icos-cluster was finally confirmed by *Sac*I and *BamH*I restriction enzyme digestion. Expecting fragments size were 4784, 4428, 3105, 2490, 237, 162, 642, 8174 bp for *Sac*I digestion and 1081, 1203, 210, 1837, 1442, 1196, 6361, 210, 10464 bp for *BamH*I digestion.

6.9 Icosalides full length NRPS heterologous expression in *E. coli*

For heterologous expression *in E. coli, icoA* was first cloned onto pET28a (+) from pCAP1000-Icos-cluster using *Nde*I and *Eco*RI site. The plasmid pET28a(+)-*icoA* was confirmed by restriction digestion and transformed to *E. coli* BAP1[158] cells were transformed. A single colony was picked for over production of the recombinant IcoA. Specifically, the overnight culture was sub-cultured in 50 mL LB containing 50 µg/mL kanamycin and induced with 0.1 mM IPTG at 15 °C for 20 h after cell density reaching OD_{600} 0.6-0.8. For confirmation of the functionality of the over produced

NRPS, 10 mL of the induced and non-induced culture was extracted with 20 mL ethyl estate. The extract was dried and dissolved in 500 µl acetonitrile for LC-MS analysis to monitor icosalides production.

References

(1) Coenye, T.; Vandamme, P. Diversity and Significance of Burkholderia Species Occupying Diverse Ecological Niches. *Environ. Microbiol.* **2003**, *5* (9), 719–729.

(2) Vial, L.; Marie-Christine, G.; Valérie, D.; Eric, D. Burkholderia Diversity and Versatility: An Inventory of the Extracellular Products. *J. Microbiol. Biotechnol.* **2007**, *17* (9), 1407–1429.

(3) Vial, L.; Chapalain, A.; Groleau, M. C.; Déziel, E. The Various Lifestyles of the Burkholderia Cepacia Complex Species: A Tribute to Adaptation. *Environ. Microbiol.* **2011**, *13* (1), 1–12.

(4) Burkholder, W. H. Sour Skin, a Bacterial Rot of Onion Bulbs. *Phytopathology* **1950**, *40* (1), 115--117 .

(5) Yabuuchi, E.; Kosako, Y.; Oyaizu, H.; Yano, I.; Hotta, H.; Hashimoto, Y.; Ezaki, T.; Arakawa, M. Proposal of Burkholderia Gen. Nov. and Transfer of Seven Species of the Genus Pseudomonas Homology Group II to the New Genus, with the Type Species Burkholderia Cepacia (Palleroni and Holmes 1981) Comb. Nov. *Microbiol. Immunol.* **1992**, *36* (12), 1251–1275.

(6) Depoorter, E.; Bull, M. J.; Peeters, C.; Coenye, T.; Vandamme, P.; Mahenthiralingam, E. Burkholderia: An Update on Taxonomy and Biotechnological Potential as Antibiotic Producers. *J. Appl. Microbiol.* **2016**, *100* (12), 5215–5229.

(7) Mahenthiralingam, E.; Baldwin, A.; Dowson, C. G. Burkholderia Cepacia Complex Bacteria: Opportunistic Pathogens with Important Natural Biology. *J. Appl. Microbiol.* **2008**, *104* (6), 1539–1551.

(8) Azegami, K.; Nishiyama, K.; Watanabe, Y.; Kadota, I.; Ohuchi, A. Pseudornonas Plantarii Sp . Nov . the Causal Agent of Rice Seedling Blight. *Int. J. Syst. Bacteriol.* **1987**, *37* (2), 144–152.

(9) Murai, Y.; Goto, M. Bacterial Leaf Spot of Amaranthus Cruentus L. Caused by Burkholderia Andropogonis (Stapp) Gills et Al. *Japanese J. Phytopathol.* **1996**, *62* (2), 181–183.

(10) Cheng, A. C.; Currie, B. J. Melioidosis: Epidemiology, Pathophysiology, and Management. *Send to Clin. Microbiol. Rev*. **2005**, *18* (2), 383–416.

(11) Silva, E. B.; Dow, S. W. Development of Burkholderia Mallei and Pseudomallei Vaccines. *Front. Cell. Infect. Microbiol.* **2013**, *3*, 10.

(12) Segonds, C.; Clavel-Batut, P.; Thouverez, M.; Grenet, D.; Le Coustumier, A.; Plésiat, P.; Chabanon, G. Microbiological and Epidemiological Features of Clinical Respiratory Isolates of Burkholderia Gladioli. *J. Clin. Microbiol.* **2009**, *47* (5), 1510–1516.

(13) Coenye, T.; Vandamme, P.; Govan, J. R. W.; Lipuma, J. J. Taxonomy and Identification of the Burkholderia Cepacia Complex. *J. Clin. Microbiol.* **2001**, *39* (10), 3427–3436.

(14) Loong, S. K.; Soh, Y. H.; Mahfodz, N. H.; Johari, J.; Abu Bakar, S. Synovial Tissue Infection with Burkholderia Fungorum. *Emerg. Infect. Dis*. **2016**, *22* (10), 1834–1835.

(15) Estrada-De Los Santos, P.; Bustillos-Cristales, R.; Caballero-Mellado, J. Burkholderia, a Genus Rich in Plant-Associated Nitrogen Fixers with Wide Environmental and Geographic Distribution. *Appl. Environ. Microbiol.* **2001**, *67* (6), 2790–2798.

(16) Reis, V. M.; Estrada-de los Santos, P.; Tenorio-Salgado, S.; Vogel, J.; Stoffels, M.; Guyon, S.; Mavingui, P.; Baldani, V. L. D.; Schmid, M.; Baldani, J. I.; et al. Burkholderia Tropica Sp. Nov., a Novel Nitrogen-Fixing, Plant-Associated Bacterium. *Int. J. Syst. Evol. Microbiol.* **2004**, *54* (6), 2155–2162.

(17) Parke, J. L.; Gurian-Sherman, D. Diversity of the Burkholderia Cepacia Complex and Implications for Risk Assessment of B Iological Control Strains. *Annu. Rev. Phytopathol.* **2001**, *39* (1), 225–258.

(18) Cain, C. C.; Henry, A. T.; Waldo, R. H.; Casida, J.; Falkinham, J. O. Identification and Characteristics of a Novel Burkholderia Strain with Broad-Spectrum Antimicrobial Activity. *Appl. Environ. Microbiol.* **2000**, *66* (9), 4139–4141.

(19) Gerhardson, B. Biological Substitutes for Pesticides. *Trends. Biotechnol*. **2002**, *20* (8), 338–343.

(20) Kilbane, J.; Chatterjee, D.; Karns, J.; Kellogg, S.; Chakrabarty, A. Biodegradation of 2, 4, 5-Trichlorophenoxyacetic Acid by a Pure Culture of Pseudomonas Cepacia. *Appl. Environ. Microbiol.* **1982**, *44* (1), 72–78.

(21) Krumme, M. L.; Timmis, K. N.; Dwyer, D. F. Degradation of Trichloroethylene by Pseudomonas Cepacia G4 and the Constitutive Mutant Strain G4 5223 PR1 in Aquifer Microcosms. *Appl. Environ. Microbiol.* **1993**, *59* (8), 2746–2749.

(22) Mahenthiralingam, E.; Urban, T. A.; Goldberg, J. B. The Multifarious, Multireplicon Burkholderia Cepacia Complex. *Nat. Rev. Microbiol.* **2005**, *3* (2), 144–156.

(23) Parker, W. L.; Rathnum, M. L.; Seiner, V.; Trejo, W. H.; Principe, P. A.; Sykes, R. B. Cepacin A and Cepacin B, Two New Antibiotics Produced by Pseudomonas Cepacia. *J. Antibiot.* **1984**, *37* (5), 431–440.

(24) Mahenthiralingam, E.; Song, L.; Sass, A.; White, J.; Wilmot, C.; Marchbank, A.; Boaisha, O.; Paine, J.; Knight, D.; Challis, G. L. Enacyloxins Are Products of an Unusual Hybrid Modular Polyketide Synthase Encoded by a Cryptic Burkholderia Ambifaria Genomic Island. *Chem. Biol.* **2011**, *18* (5), 665–677.

(25) Schmidt, S.; Blom, J. F.; Pernthaler, J.; Berg, G.; Baldwin, A.; Mahenthiralingam, E.; Eberl, L. Production of the Antifungal Compound Pyrrolnitrin Is Quorum Sensing-Regulated in Members of the *Burkholderia Cepacia* Complex. *Environ. Microbiol.* **2009**, *11* (6), 1422–1437.

(26) Lu, S. E.; Novak, J.; Austin, F. W.; Gu, G.; Ellis, D.; Kirk, M.; Wilson-Stanford, S.; Tonelli, M.; Smith, L. Occidiofungin, a Unique Antifungal Glycopeptide Produced by a Strain of Burkholderia Contaminans. *Biochemistry* **2009**, *48* (35), 8312–8321.

(27) Coenye, T.; Mahenthiralingam, E. *Burkholderia: From Genomes to Function*; Caister Academic Press, **2014**.

(28) Song, L.; Jenner, M.; Masschelein, J.; Jones, C.; Bull, M. J.; Harris, S. R.; Hartkoorn, R. C.;

Vocat, A.; Romero-Canelon, I.; Coupland, P.; et al. Discovery and Biosynthesis of Gladiolin: A *Burkholderia Gladioli* Antibiotic with Promising Activity against *Mycobacterium Tuberculosis*. *J. Am. Chem. Soc.* **2017**, *139* (23), 7974-7981..

(29) Irschik, H.; Schummer, D.; Höfle, G.; Reichenbach, H.; Steinmetz, H.; Jansen, R. Etnangien, a Macrolide-Polyene Antibiotic from Sorangium Cellulosum That Inhibits Nucleic Acid Polymerases. *J. Nat. Prod.* **2007**, *70* (6), 1060–1063.

(30) Menche, D.; Arikan, F.; Perlova, O.; Horstmann, N.; Ahlbrecht, W.; Wenzel, S. C.; Jansen, R.; Irschik, H.; Mü, R. Stereochemical Determination and Complex Biosynthetic Assembly of Etnangien, a Highly Potent RNA Polymerase Inhibitor from the Myxobacterium Sorangium Cellulosum. *J. Am. Chem. Soc*. **2008**, *130* (43),14234-14243.

(31) Menche, D.; Arikan, F.; Perlova, O.; Horstmann, N.; Ahlbrecht, W.; Wenzel, S. C.; Jansen, R.; Irschik, H.; Müller, R. Stereochemical Determination and Complex Biosynthetic Assembly of Etnangien, a Highly Potent RNA Polymerase Inhibitor from the Myxobacterium Sorangium Cellulosum. *J. Am. Chem. Soc.* **2008**, *130* (43), 14234–14243.

(32) Boros, C.; Smith, C. J.; Vasina, Y.; Che, Y.; Dix, A. B.; Darveaux, B.; Pearce, C. Isolation and Identification of the Icosalides--Cyclic Peptolides with Selective Antibiotic and Cytotoxic Activities. *J. Antibiot. (Tokyo).* **2006**, *59* (8), 486–494.

(33) Staunton, J.; Weissman, K. J. Polyketide Biosynthesis: A Millennium Review. *Nat. Prod. Rep.* **2001**, *18* (4), 380–416.

(34) Khosla, C. Structures and Mechanisms of Polyketide Synthases. *J. Org. Chem.* **2009**, *74* (17), 6416–6420.

(35) Weissman, K. J. Uncovering the Structures of Modular Polyketide Synthases. *Nat. Prod. Rep.* **2015**, *32* (3), 436–453.

(36) Xu, W.; Qiao, K.; Tang, Y. Structural Analysis of Protein–protein Interactions in Type I Polyketide Synthases. *Crit. Rev. Biochem. Mol. Biol.* **2013**, *48* (2), 98–122.

(37) Hopwood, D. A. Genetic Contributions to Understanding Polyketide Synthases. *Chem. Rev.* **1997**, *97* (7), 2465–2498.

(38) Helfrich, E. J. N.; Piel, J. Biosynthesis of Polyketides by Trans-AT Polyketide Synthases. *Nat. Prod. Rep.* **2016**, *33* (2), 231–316.

(39) Piel, J. A Polyketide Synthase-Peptide Synthetase Gene Cluster from an Uncultured Bacterial Symbiont of Paederus Beetles. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (22), 14002–14007.

(40) Moldenhauer, J.; Chen, X. H.; Borriss, R.; Piel, J. Biosynthesis of the Antibiotic Bacillaene, the Product of a Giant Polyketide Synthase Complex of the Trans-AT Family. *Angew. Chemie - Int. Ed.* **2007**, *46* (43), 8195–8197.

(41) Pulsawat, N.; Kitani, S.; Nihira, T. Characterization of Biosynthetic Gene Cluster for the Production of Virginiamycin M, a Streptogramin Type A Antibiotic, in Streptomyces Virginiae. *Gene.* **2007**, *393* (1–2), 31–42.

(42) Cheng, Y.-Q.; Tang, G.-L.; Shen, B. Type I Polyketide Synthase Requiring a Discrete Acyltransferase for Polyketide Biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (6), 3149–3154.

(43) Jenke-Kodama, H.; Sandmann, A.; Müller, R.; Dittmann, E. Evolutionary Implications of Bacterial Polyketide Synthases. *Mol. Biol. Evol.* **2005**, *22* (10), 2027–2039.

(44) Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J. Exploiting the Mosaic Structure of Trans-Acyltransferase Polyketide Synthases for Natural Product Discovery and Pathway Dissection. *Nat. Biotechnol.* **2008**, *26* (2), 225–233.

(45) Kapur, S.; Lowry, B.; Yuzawa, S.; Kenthirapalan, S.; Chen, A. Y.; Cane, D. E.; Khosla, C. Reprogramming a Module of the 6-Deoxyerythronolide B Synthase for Iterative Chain Elongation. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (11), 1–6.

(46) Hardt, I. H.; Steinmetz, H.; Gerth, K.; Sasse, F.; Reichenbach, H.; Höfle, G. New Natural Epothilones from Sorangium Cellulosum, Strains So ce90/B2 and So ce90/D13: Isolation,

Structure Elucidation, and SAR Studies. *J. Nat. Prod.* **2001**, *64* (7), 847–856.

(47) He, J.; Hertweck, C. Iteration as Programmed Event during Polyketide Assembly; Molecular Analysis of the Aureothin Biosynthesis Gene Cluster. *Chem. Biol.* **2003**, *10* (12), 1225–1232.

(48) Olano, C.; Wilkinson, B.; Moss, S. J.; Braña, A. F.; Méndez, C.; Leadlay, P. F.; Salas, J. A. Evidence from Engineered Gene Fusions for the Repeated Use of a Module in a Modular Polyketide Synthase. *Chem. Commu.n (Camb)*. **2003**, (22), 2780-2782.

(49) Busch, B.; Ueberschaar, N.; Sugimoto, Y.; Hertweck, C. Interchenar Retrotransfer of Aureothin Intermediates in an Iterative Polyketide Synthase Module. *J. Am. Chem. Soc.* **2012**, *134* (30), 12382–12385.

(50) Busch, B.; Ueberschaar, N.; Behnken, S.; Sugimoto, Y.; Werneburg, M.; Traitcheva, N.; He, J.; Hertweck, C. Multifactorial Control of Iteration Events in a Modular Polyketide Assembly Line. *Angew. Chem. Int. Ed. Engl*. **2013**, *52* (20), 5285–5289.

(51) Dunn, B. J.; Cane, D. E.; Khosla, C. Mechanism and Specificity of an Acyltransferase Domain from a Modular Polyketide Synthase. *Biochemistry.* **2013**, *52* (11), 1839–1841.

(52) Khosla, C.; Tang, Y.; Chen, A. Y.; Schnarr, N. a; Cane, D. E. Structure and Mechanism of the 6-Deoxyerythronolide B Synthase. *Annu. Rev. Biochem.* **2007**, *76,* 195–221.

(53) Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. Biosynthesis of Polyketide Synthase Extender Units. *Nat. Prod. Rep.* **2009**, *26* (1), 90–114.

(54) Liou, G. F.; Lau, J.; Cane, D. E.; Khosla, C. Quantitative Analysis of Loading and Extender Acyltransferases of Modular Polyketide Synthases. *Biochemistry* **2003**, *42* (1), 200–207.

(55) Lau, J.; Cane, D. E.; Khosla, C. Substrate Specificity of the Loading Didomain of the Erythromycin Polyketide Synthase. *Biochemistry* **2000**, *39* (34), 10514–10520.

(56) Tang, Y.; Kim, C.-Y.; Mathews, I. I.; Cane, D. E.; Khosla, C. The 2.7-Angstrom Crystal Structure of a 194-kDa Homodimeric Fragment of the 6-Deoxyerythronolide B Synthase. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (30), 11124–11129.

(57) Keatinge-Clay, A. T. The Structures of Type I Polyketide Synthases. *Nat. Prod. Rep*. **2012**, *29* (10), 1050–1073.

(58) Aparicio, J. F.; Molnár, I.; Schwecke, T.; König, A.; Haydock, S. F.; Khaw, L. E.; Staunton, J.; Leadlay, P. F. Organization of the Biosynthetic Gene Cluster for Rapamycin in Streptomyces Hygroscopicus: Analysis of the Enzymatic Domains in the Modular Polyketide Synthase. *Gene* **1996**, *169* (1), 9–16.

(59) Robbins, T.; Kapilivsky, J.; Cane, D. E.; Khosla, C. Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. *Biochemistry*. **2016**, 55 (32), 4476–4484.

(60) Zhang, Y. M.; Hurlbert, J.; White, S. W.; Rock, C. O. Roles of the Active Site Water, Histidine 303, and Phenylalanine 396 in the Catalytic Mechanism of the Elongation Condensing Enzyme of Streptococcus Pneumoniae. *J. Biol. Chem.* **2006**, *281* (25), 17390–17399.

(61) Khosla, C.; Gokhale, R. S.; Jacobsen, J. R.; Cane, D. E. Tolerance and Specificity of Polyketide Synthases. *Annu. Rev. Biochem.* **1999**, *68* (1), 219–253.

(62) Jenner, M.; Frank, S.; Kampa, A.; Kohlhaas, C.; Pöplau, P.; Briggs, G. S.; Piel, J.; Oldham, N. J. Substrate Specificity in Ketosynthase Domains from Trans-AT Polyketide Synthases. *Angew. Chemie - Int. Ed.* **2013**, *52* (4), 1143–1147.

(63) Gay, D. C.; Gay, G.; Axelrod, A. J.; Jenner, M.; Kohlhaas, C.; Kampa, A.; Oldham, N. J.; Piel, J.; Keatinge-clay, A. T. A Close Look at a Ketosynthase from a Trans-Acyltransferase Modular Polyketide Synthase. **2015**, *22* (3), 444–451.

(64) Long, P. F.; Wilkinson, C. J.; Bisang, C. P.; Cortés, J.; Dunster, N.; Oliynyk, M.; McCormick, E.; McArthur, H.; Mendez, C.; Salas, J. A.; et al. Engineering Specificity of Starter Unit Selection by the Erythromycin-Producing Polyketide Synthase. *Mol. Microbiol.* **2002**, *43* (5), 1215–1225.

(65) Kavanagh, K. L.; Jörnvall, H.; Persson, B.; Oppermann, U. The SDR Superfamily: Functional and Structural Diversity within a Family of Metabolic and Regulatory Enzymes. *Cell. Mol. Life Sci.* **2008**, *65* (24), 3895–3906.

(66) Zheng, J.; Taylor, C. A.; Piasecki, S. K.; Keatinge-Clay, A. T. Structural and Functional Analysis of A-Type Ketoreductases from the Amphotericin Modular Polyketide Synthase. *Structure* **2010**, *18* (8), 913–922.

(67) Keatinge-Clay, A. T.; Stroud, R. M. The Structure of a Ketoreductase Determines the Organization of the β-Carbon Processing Enzymes of Modular Polyketide Synthases. *Structure* **2006**, *14* (4), 737–748.

(68) Keatinge-Clay, A. T. A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. *Chem. Biol.* **2007**, *14* (8), 898–908.

(69) Weissman, K. J. Polyketide Stereocontrol: A Study in Chemical Biology. *Beilstein J. Org. Chem.* **2017**, *13*, 348–371.

(70) Keatinge-Clay, A. Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *J. Mol. Biol.* **2008**, *384* (4), 941–953.

(71) Akey, D. L.; Razelun, J. R.; Tehranisa, J.; Sherman, D. H.; Gerwick, W. H.; Smith, J. L. Crystal Structures of Dehydratase Domains from the Curacin Polyketide Biosynthetic Pathway. *Structure* **2010**, *18* (1), 94–105.

(72) Gay, D.; You, Y. O.; Keatinge-Clay, A.; Cane, D. E. Structure and Stereospecificity of the Dehydratase Domain from the Terminal Module of the Rifamycin Polyketide Synthase. *Biochemistry* **2013**, *52* (49), 8916–8928.

(73) Guo, X.; Liu, T.; Valenzano, C. R.; Deng, Z.; Cane, D. E. Mechanism and Stereospecificity of a Fully Saturating Polyketide Synthase Module: Nanchangmycin Synthase Module 2 and Its Dehydratase Domain. *J. Am. Chem. Soc.* **2010**, *132* (42), 14694–14696.

(74) Valenzano, C. R.; You, Y. O.; Garg, A.; Keatinge-Clay, A.; Khosla, C.; Cane, D. E. Stereospecificity of the Dehydratase Domain of the Erythromycin Polyketide Synthase. *J. Am. Chem. Soc.* **2010**, *132* (42), 14697–14699.

(75) Sedgwick, B.; Morris, C.; French, S. J. Stereochemical Course of Dehydration Catalysed by the Yeast Fatty Acid Synthetase. *J. Chem. Soc. Chem. Commun.* **1978**, *0* (5), 193–194.

(76) Kwan, D. H.; Schulz, F. The Stereochemistry of Complex Polyketide Biosynthesis by Modular Polyketide Synthases. *Molecules* **2011**, *16* (12), 6092–6115.

(77) Khosla, C.; Herschlag, D.; Cane, D. E.; Walsh, C. T. Assembly Line Polyketide Synthases: Mechanistic Insights and Unsolved Problems. *Biochemistry*, **2014**, *53* (18), 2875–2883.

(78) Kwan, D. H.; Sun, Y.; Schulz, F.; Hong, H.; Popovic, B.; Sim-Stark, J. C. C.; Haydock, S. F.; Leadlay, P. F. Prediction and Manipulation of the Stereochemistry of Enoylreduction in Modular Polyketide Synthases. *Chem. Biol.* **2008**, *15* (11), 1231–1240.

(79) Zhang, L.; Ji, J.; Yuan, M.; Feng, Y.; Wang, L.; Deng, Z.; Bai, L.; Zheng, J. Stereospecificity of Enoylreductase Domains from Modular Polyketide Synthases. *ACS Chem. Biol.* **2018**, *13* (4), 871–875.

(80) Zheng, J.; Gay, D. C.; Demeler, B.; White, M. A.; Keatinge-Clay, A. T.; Chem Biol, N. Divergence of Multimodular Polyketide Synthases Revealed by a Didomain Structure HHS Public Access Author Manuscript. *Nat. Chem. Biol.* **2012**, *8* (7), 615–621.

(81) Persson, B.; Zigler, J. S.; Jörnvall, H. A Super-Family of Medium-Chain Dehydrogenases/reductases (MDR). Sub-Lines Including Zeta-Crystallin, Alcohol and Polyol Dehydrogenases, Quinone Oxidoreductase Enoyl Reductases, VAT-1 and Other Proteins. *Eur. J. Biochem.* **1994**, *226* (1), 15–22.

(82) Kwan, D. H.; Leadlay, P. F. Mutagenesis of a Modular Polyketide Synthase Enoylreductase Domain Reveals Insights into Catalysis and Stereospecificity. *ACS Chem. Biol.* **2010**, *5* (9), 829–838.

(83) Bukhari, H. S. T.; Jakob, R. P.; Maier, T. Evolutionary Origins of the Multienzyme Architecture of Giant Fungal Fatty Acid Synthase. *Struct. Des.* **2014**, *22*, 1775–1785.

(84) Bumpus, S. B.; Magarvey, N. A.; Kelleher, N. L.; Walsh, C. T.; Calderone, C. T. Polyunsaturated Fatty-Acid-Like Trans -Enoyl Reductases Utilized in Polyketide Biosynthesis. *J. Am. Chem. Soc.* **2008**, *130* (35), 11614–11616.

(85) Uytterhoeven, B.; Lathouwers, T.; Voet, M.; Michiels, C. W.; Lavigne, R. A Protein Interaction Map of the Kalimantacin Biosynthesis Assembly Line. *Front. Microbiol.* **2016**,

7, 1726.

(86) Tang, G.-L.; Cheng, Y.-Q.; Shen, B. Leinamycin Biosynthesis Revealing Unprecedented Architectural Complexity for a Hybrid Polyketide Synthase and Nonribosomal Peptide Synthetase. *Chem. Biol.* **2004**, *11* (1), 33–45.

(87) Simunovic, V.; Zapp, J.; Rachid, S.; Krug, D.; Meiser, P.; Müller, R. Myxovirescin A Biosynthesis Is Directed by Hybrid Polyketide Synthases/Nonribosomal Peptide Synthetase, 3-Hydroxy-3-Methylglutaryl-CoA Synthases, and Trans-Acting Acyltransferases. *ChemBioChem* **2006**, *7* (8), 1206–1220.

(88) Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C. Mechanism and Specificity of the Terminal Thioesterase Domain from the Erythromycin Polyketide Synthase. *Chem. Biol.* **1999**, *6* (2), 117–125.

(89) Chakravarty, B.; Gu, Z.; Chirala, S. S.; Wakil, S. J.; Quiocho, F. A. Human Fatty Acid Synthase: Structure and Substrate Selectivity of the Thioesterase Domain. *Proc. Natl. Acad. Sci.* **2004**, *101* (44), 15567–15572.

(90) Tsai, S.-C.; Miercke, L. J. W.; Krucinski, J.; Gokhale, R.; Chen, J. C.-H.; Foster, P. G.; Cane, D. E.; Khosla, C.; Stroud, R. M. Crystal Structure of the Macrocycle-Forming Thioesterase Domain of the Erythromycin Polyketide Synthase: Versatility from a Unique Substrate Channel. *Proc. Natl. Acad. Sci.* **2001**, *98* (26), 14808–14813.

(91) Heathcote, M. L.; Staunton, J.; Leadlay, P. F. Role of Type II Thioesterases: Evidence for Removal of Short Acyl Chains Produced by Aberrant Decarboxylation of Chain Extender Units. *Chem. Biol.* **2001**, *8* (2), 207–220.

(92) Kim, B. S.; Cropp, T. A.; Beck, B. J.; Sherman, D. H.; Reynolds, K. A. Biochemical Evidence for an Editing Role of Thioesterase II in the Biosynthesis of the Polyketide Pikromycin. *J. Biol. Chem.* **2002**, *277* (50), 48028–48034.

(93) Kotowska, M.; Pawlik, K. Roles of Type II Thioesterases and Their Application for Secondary Metabolite Yield Improvement. *Appl. Microbiol. Biotechnol.* **2014**, *98* (18), 7735–7746.

(94) Alekseyev, V. Y.; Liu, C. W.; Cane, D. E.; Puglisi, J. D.; Khosla, C. Solution Structure and Proposed Domain Domain Recognition Interface of an Acyl Carrier Protein Domain from a Modular Polyketide Synthase. *Protein. Sci.* **2007**, *16* (10), 2093–2107.

(95) Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. A New Enzyme Superfamily - The Phosphopantetheinyl Transferases. *Chem. Biol.* **1996**, *3* (11), 923–936.

(96) Zhang, Y. M.; Wu, B.; Zheng, J.; Rock, C. O. Key Residues Responsible for Acyl Carrier Protein and β-Ketoacyl-Acyl Carrier Protein Reductase (FabG) Interaction. *J. Biol. Chem.* **2003**, *278* (52), 52935–52943.

(97) Kapur, S.; Chen, A. Y.; Cane, D. E.; Khosla, C. Molecular Recognition between Ketosynthase and Acyl Carrier Protein Domains of the 6-Deoxyerythronolide B Synthase. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (51), 22066–22071.

(98) Weissman, K. J.; Hong, H.; Popovic, B.; Meersman, F. Evidence for a Protein-Protein Interaction Motif on an Acyl Carrier Protein Domain from a Modular Polyketide Synthase. *Chem. Biol.* **2006**, *13* (6), 625–636.

(99) Schneiker, S.; Perlova, O.; Kaiser, O.; Gerth, K.; Alici, A.; Altmeyer, M. O.; Bartels, D.; Bekel, T.; Beyer, S.; Bode, E.; et al. Complete Genome Sequence of the Myxobacterium Sorangium Cellulosum. *Nat. Biotechnol.* **2007**, *25* (11), 1281–1289.

(100) Challis, G. L. Antibiotics from Gram-Negative Bacteria: A Comprehensive Overview and Selected Biosynthetic Highlights. *Nat. Prod. Rep.* **2017**, *34* (7), 712–783.

(101) Mattheus, W.; Gao, L.; Herdewijn, P.; Landuyt, B.; Verhaegen, J.; Masschelein, J.; Volckaert, G.; Lavigne, R. Article Isolation and Purification of a New Kalimantacin / Batumin-Related Polyketide Antibiotic and Elucidation of Its Biosynthesis Gene Cluster. *Chem. Biol***.** **2010**, *17* (2), 149-159.

(102) Winn, M.; Fyans, J. K.; Zhuo, Y.; Micklefield, J. Recent Advances in Engineering Nonribosomal Peptide Assembly Lines. *Nat. Prod. Rep.* **2016**, *33* (2), 317–347.

(103) Sieber, S. A.; Marahiel, M. A. Molecular Mechanisms Underlying Nonribosomal Peptide

Synthesis: Approaches to New Antibiotics. *Chem. Rev.* **2005**, *105* (2), 715-738.

(104) Hur, G. H.; Vickery, C. R.; Burkart, M. D. Explorations of Catalytic Domains in Non-Ribosomal Peptide Synthetase Enzymology. *Nat. Prod. Rep.* **2012**, *29* (10), 1074.

(105) Hur, G. H.; Vickery, C. R.; Burkart, M. D. Explorations of Catalytic Domains in Non-Ribosomal Peptide Synthetase Enzymology. *Nat. Prod. Rep.* **2012**, *29* (10), 1074.

(106) Mena-Segovia, J. Structural and Functional Considerations of the Cholinergic Brainstem. *J. Neural. Transm. (Vienna).* **2016**, *123* (7):731-736.

(107) Conti, E.; Stachelhaus, T.; Marahiel, M. A.; Brick, P. Structural Basis for the Activation of Phenylalanine in the Non-Ribosomal Biosynthesis of Gramicidin S. *EMBO. J.* **1997**, *16* (14), 4174–4183.

(108) May, J. J.; Kessler, N.; Marahiel, M. A.; Stubbs, M. T. Crystal Structure of DhbE, an Archetype for Aryl Acid Activating Domains of Modular Nonribosomal Peptide Synthetases. *Proc. Natl. Acad. Sci.* **2002**, *99* (19), 12120–12125.

(109) Gulick, A. M.; Lu, X.; Dunaway-Mariano, D. Crystal Structure of 4-chlorobenzoate:CoA Ligase/synthetase in the Unliganded and Aryl Substrate-Bound Statest. *Biochemistry*, **2004**, *43* (27), 8670–8679.

(110) Turgay, K.; Krause, M.; Marahiel, M. A. Four Homologous Domains in the Primary Structure of GrsB Are Related to Domains in a Superfamily of Adenylate-forming Enzymes. *Mol. Microbiol.* **1992**, *6* (4), 529–546.

(111) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. The Specificity-Conferring Code of Adenylation Domains in Nonribosomal Peptide Synthetases. *Chem. Biol.* **1999**, *6* (8), 493–505.

(112) Challis, G. L.; Ravel, J.; Townsend, C. A. Predictive, Structure-Based Model of Amino Acid Recognition by Nonribosomal Peptide Synthetase Adenylation Domains. *Chem. Biol.* **2000**, *7* (3), 211–224.

(113) Eppelmann, K.; Stachelhaus, T.; Marahiel, M. A. Exploitation of the Selectivity-Conferring

Code of Nonribosomal Peptide Synthetases for the Rational Design of Novel Peptide Antibiotics. *Biochemistry.* **2002**, *41* (30), 9718–9726.

(114) Chen, C.-Y.; Georgiev, I.; Anderson, A. C.; Donald, B. R. Computational Structure-Based Redesign of Enzyme Activity. *Proc. Natl. Acad. Sci.* **2009**, *106* (10), 3764–3769.

(115) Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D. The Phosphopantetheinyl Transferases: Catalysis of a Post-Translational Modification Crucial for Life. *Nat. Prod. Rep.* **2014**, *31* (1), 61–108.

(116) Koglin, A.; Mofid, M. R.; Löhr, F.; Schäfer, B.; Rogov, V. V; Blum, M. M.; Mittag, T.; Marahiel, M. A.; Bernhard, F.; Dötsch, V. Conformational Switches Modulate Protein Interactions in Peptide Antibiotic Synthetases. *Science.* **2006**, *312* (5771), 273–276.

(117) De Crécy-Lagard, V.; Marlière, P.; Saurin, W. Multienzymatic Non Ribosomal Peptide Biosynthesis: Identification of the Functional Domains Catalysing Peptide Elongation and Epimerisation. *C. R. Acad. Sci. III.* **1995**, *318* (9), 927–936.

(118) Roche, E. D.; Walsh, C. T. Dissection of the EntF Condensation Domain Boundary and Active Site Residues in Nonribosomal Peptide Synthesis. *Biochemistry.* **2003**, *42* (5), 1334–1344.

(119) Bergendahl, V.; Linne, U.; Marahiel, M. A. Mutational Analysis of the C-Domain in Nonribosomal Peptide Synthesis. *Eur. J. Biochem.* **2002**, *269* (2), 620–629.

(120) Keating, T. A.; Marshall, C. G.; Walsh, C. T.; Keating, A. E. The Structure of Vibh Represents Nonribosomal Peptide Synthetase Condensation, Cyclization and Epimerization Domains. *Nat. Struct. Biol.* **2002**, *9* (7), 522–526.

(121) Samel, S. A.; Schoenafinger, G.; Knappe, T. A.; Marahiel, M. A.; Essen, L. O. Structural and Functional Insights into a Peptide Bond-Forming Bidomain from a Nonribosomal Peptide Synthetase. *Structure.* **2007**, *15* (7), 781–792.

(122) Miller, B. R.; Drake, E. J.; Shi, C.; Aldrich, C. C.; Gulick, A. M. Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture. *J. Biol. Chem.* **2016**, *291* (43), 22559–22571.

(123) Bloudoff, K.; Rodionov, D.; Schmeing, T. M. Crystal Structures of the First Condensation Domain of CDA Synthetase Suggest Conformational Changes during the Synthetic Cycle of Nonribosomal Peptide Synthetases. *J. Mol. Biol.* **2013**, *425* (17), 3137–3150.

(124) Bloudoff, K.; Alonzo, D. A.; Schmeing, T. M. Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases. *Cell Chem. Biol.* **2016**, *23* (3), 331–339.

(125) Bloudoff, K.; Schmeing, T. M. Structural and Functional Aspects of the Nonribosomal Peptide Synthetase Condensation Domain Superfamily: Discovery, Dissection and Diversity. *Biochim. Biophys. Acta - Proteins Proteomics.* **2017**, *1865* (11), 1587–1604.

(126) Roongsawang, N.; Siew, P. L.; Washio, K.; Takano, K.; Kanaya, S.; Morikawa, M. Phylogenetic Analysis of Condensation Domains in the Nonribosomal Peptide Synthetases. *FEMS. Microbiol. Lett.* **2005**, *252* (1), 143–151.

(127) Rausch, C.; Hoof, I.; Weber, T.; Wohlleben, W.; Huson, D. H. Phylogenetic Analysis of Condensation Domains in NRPS Sheds Light on Their Functional Evolution. *BMC. Evol. Biol.* **2007**, *7*, 78.

(128) Luo, L.; Kohli, R. M.; Onishi, M.; Linne, U.; Marahiel, M. A.; Walsh, C. T. Timing of Epimerization and Condensation Reactions in Nonribosomal Peptide Assembly Lines: Kinetic Analysis of Phenylalanine Activating Elongation Modules of Tyrocidine Synthetase B. *Biochemistry* **2002**, *41* (29), 9184–9196.

(129) Imker, H. J.; Krahn, D.; Clerc, J.; Kaiser, M.; Walsh, C. T. N-Acylation during Glidobactin Biosynthesis by the Tridomain Nonribosomal Peptide Synthetase Module GlbF. *Chem. Biol.* **2010**, *17* (10), 1077–1083.

(130) Balibar, C. J.; Vaillancourt, F. H.; Walsh, C. T. Generation of D Amino Acid Residues in Assembly of Arthrofactin by Dual Condensation/epimerization Domains. *Chem. Biol.* **2005**, *12* (11), 1189–1200.

(131) Konz, D.; Klens, A.; Schörgendorfer, K.; Marahiel, M. A. The Bacitracin Biosynthesis Operon of Bacillus Licheniformis ATCC 10716: Molecular Characterization of Three Multi-

Modular Peptide Synthetases. *Chem. Biol.* **1997**, *4* (12), 927–937.

(132) Linne, U.; Marahiel, M. A. Control of Directionality in Nonribosomal Peptide Synthesis: Role of the Condensation Domain in Preventing Misinitiation and Timing of Epimerization. *Biochemistry* **2000**, *39* (34), 10439–10447.

(133) Stachelhaus, T.; Walsh, C. T. Mutational Analysis of the Epimerization Domain in the Initiation Module PheATE of Gramicidin S Synthetase. *Biochemistry* **2000**, *39* (19), 5775–5787.

(134) Du, L.; Chen, M.; Sánchez, C.; Shen, B. An Oxidation Domain in the BlmIII Non-Ribosomal Peptide Synthetase Probably Catalyzing Thiazole Formation in the Biosynthesis of the Anti-Tumor Drug Bleomycin in Streptomyces Verticillus ATCC15003. *FEMS. Microbiol. Lett.* **2000**, *189* (2), 171–175.

(135) Schneider, T. L.; Shen, B.; Walsh, C. T. Oxidase Domains in Epothilone and Bleomycin Biosynthesis: Thiazoline to Thiazole Oxidation during Chain Elongation. *Biochemistry* **2003**, *42* (32), 9722–9730.

(136) Reimmann, C.; Patel, H. M.; Serino, L.; Barone, M.; Walsh, C. T.; Haas, D. Essential pchG-Dependent Reduction in Pyochelin Biosynthesis of Pseudomonas Aeruginosa. *J. Bacteriol.* **2001**, *183* (3), 813–820.

(137) Velkov, T.; Lawen, A. Non-Ribosomal Peptide Synthetases as Technological Platforms for the Synthesis of Highly Modified Peptide Bioeffectors - Cyclosporin Synthetase as a Complex Example. *Biotechnology Annual Review*. **2003**, *9*, 151–197.

(138) Schauwecker, F.; Pfennig, F.; Grammel, N.; Keller, U. Construction and in Vitro Analysis of a New Bi-Modular Polypeptide Synthetase for Synthesis of N-Methylated Acyl Peptides. *Chem. Biol.* **2000**, *7* (4), 287–297.

(139) Miller, D. A.; Walsh, C. T.; Luo, L. C-Methyltransferase and Cyclization Domain Activity at the Intraprotein PK/NRP Switch Point of Yersiniabactin Synthetase. *J. Am. Chem. Soc.* **2001**, *123* (34), 8434–8435

(140) Miller, L. M.; Mazur, M. T.; McLoughlin, S. M.; Kelleher, N. L. Parallel Interrogation of

Covalent Intermediates in the Biosynthesis of Gramicidin S Using High-Resolution Mass Spectrometry. *Protein Sci.* **2005**, *14* (10), 2702–2712.

(141) Kohli, R. M.; Walsh, C. T. Enzymology of Acyl Chain Macrocyclization in Natural Product Biosynthesis. *Chem. Commun.* **2003**, *3* (3), 297–307.

(142) Trauger, J. W.; Kohli, R. M.; Mootz, H. D.; Marahiel, M. A.; Walsh, C. T. Peptide Cyclization Catalysed by the Thioesterase Domain of Tyrocidine Synthetase. *Nature* **2000**, *407* (6801), 215–218.

(143) Du, L.; Lou, L. PKS and NRPS Release Mechanisms. *Nat. Prod. Rep.* 2010, *27* (2), 255-278.

(144) Flannagan, R. S.; Linn, T.; Valvano, M. A. A System for the Construction of Targeted Unmarked Gene Deletions in the Genus Burkholderia. *Environ. Microbiol.* **2008**, *10* (6), 1652–1660.

(145) Borges, S. E. F.; Sletta, H.; Fjærvik, E.; Brautaset, T.; Ellingsen, T. E.; Gulliksen, O. M.; Zotchev, S. B. Effect of Glucose Limitation and Specific Mutations in the Module 5 Enoyl Reductase Domains in the Nystatin and Amphotericin Polyketide Synthases on Polyene Macrolide Biosynthesis. *Arch. Microbiol.* **2006**, *185* (3), 165–171.

(146) Donadio, S.; McAlpine, J. B.; Sheldon, P. J.; Jackson, M.; Katz, L. An Erythromycin Analog Produced by Reprogramming of Polyketide Synthesis. *Proc. Natl. Acad. Sci.* **1993**, *90* (15), 7119–7123.

(147) Geoghegan, K. F.; Dixon, H. B. F.; Rosner, P. J.; Hoth, L. R.; Lanzetti, A. J.; Borzilleri, K. A.; Marr, E. S.; Pezzullo, L. H.; Martin, L. B.; LeMotte, P. K.; et al. Spontaneous α-N-6-Phosphogluconoylation of a “His Tag” inEscherichia coli:The Cause of Extra Mass of 258 or 178 Da in Fusion Proteins. *Anal. Biochem.* **1999**, *267* (1), 169–184.

(148) Dorrestein, P. C.; Bumpus, S. B.; Calderone, C. T.; Garneau-Tsodikova, S.; Aron, Z. D.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Kelleher, N. L. Facile Detection of Acyl and Peptidyl Intermediates on Thiotemplate Carrier Domains via Phosphopantetheinyl Elimination Reactions during Tandem Mass Spectrometry. *Biochemistry* **2006**, *45* (42), 12756–12766.

(149) Lefebre, M. D.; Valvano, M. a. Construction and Evaluation of Plasmid Vectors Optimized

for Constitutive and Regulated Gene Expression in. *Society* **2002**, *68* (12), 5956–5964.

(150) Saito, J.; Yamada, M.; Watanabe, T.; Iida, M.; Kitagawa, H.; Takahata, S.; Ozawa, T.; Takeuchi, Y.; Ohsawa, F. Crystal Structure of Enoyl-Acyl Carrier Protein Reductase (FabK) from Streptococcus Pneumoniae Reveals the Binding Mode of an Inhibitor. *Protein Sci.* **2008**, *17* (4), 691–699.

(151) Marrakchi, H.; Dewolf, W. E.; Quinn, C.; West, J.; Polizzi, B. J.; So, C. Y.; Holmes, D. J.; Reed, S. L.; Heath, R. J.; Payne, D. J.; et al. Characterization of Streptococcus Pneumoniae Enoyl-(Acyl-Carrier Protein) Reductase (FabK). *Biochem. J.* **2003**, *370* (Pt 3), 1055–1062.

(152) Hevener, K. E.; Mehboob, S.; Boci, T.; Truong, K.; Santarsiero, B. D.; Johnson, M. E. Expression, Purification and Characterization of Enoyl-ACP Reductase II, FabK, from Porphyromonas Gingivalis. *Protein Expr. Purif.* **2012**, *85* (1), 100–108.

(153) Jenner, M.; Afonso, J. P.; Bailey, H. R.; Frank, S.; Kampa, A.; Piel, J.; Oldham, N. J. Acyl-Chain Elongation Drives Ketosynthase Substrate Selectivity in Trans-Acyltransferase Polyketide Synthases. *Angew. Chemie - Int. Ed.* **2015**, *54* (6), 1817–1821.

(154) Davison, J.; Dorival, J.; Rabeharindranto, H.; Mazon, H.; Chagot, B.; Gruez, A.; Weissman, K. J. Insights into the Function of Trans-Acyl Transferase Polyketide Synthases from the SAXS Structure of a Complete Module. *Chem. Sci*. **2014**, *5* (8), 3081.

(155) Jaitzig, J.; Li, J.; Süssmuth, R. D.; Neubauer, P. Reconstituted Biosynthesis of the Nonribosomal Macrolactone Antibiotic Valinomycin in Escherichia Coli. *ACS Synth. Biol.* **2014**, *3* (7), 432–438.

(156) Lowry, B.; Robbins, T.; Weng, C.; O'Brien, R. V; Cane, D. E. In Vitro Reconstitution and Analysis of the 6- Deoxyerythronolide B Synthase. *J. Am. Chem. Soc.* **2013**, *135*, 16809–16812.

(157) Kuo, J.; Lynch, S. R.; Liu, C. W.; Xiao, X.; Khosla, C. Partial in Vitro Reconstitution of an Orphan Polyketide Synthase Associated with Clinical Cases of Nocardiosis. *ACS Chem. Biol.* **2016**, *11* (9), 2636–2641.

(158) Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C. Biosynthesis of Complex

Polyketides in a Metabolically Engineered Strain of E. Coli. *Science.* **2001**, *291* (5509), 1790–1792.

(159) Bozhüyük, K. A. J.; Fleischhacker, F.; Linck, A.; Wesche, F.; Tietze, A.; Niesert, C.-P.; Bode, H. B. De Novo Design and Engineering of Non-Ribosomal Peptide Synthetases. *Nat. Chem.* **2017**, *10* (3), 275–281.

(160) Miller, V. L.; Mekalanos, J. J. A Novel Suicide Vector and Its Use in Construction of Insertion Mutations: Osmoregulation of Outer Membrane Proteins and Virulence Determinants in Vibrio Cholerae Requires toxR. *J. Bacteriol.* **1988**, *170* (6), 2575–2583.

(161) Figurski, D. H.; Helinski, D. R. Replication of an Origin-Containing Derivative of Plasmid RK2 Dependent on a Plasmid Function Provided in Trans. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76* (4), 1648–1652.

(162) Sánchez, C.; Du, L.; Edwards, D. J.; Toney, M. D.; Shen, B. Cloning and Characterization of a Phosphopantetheinyl Transferase from Streptomyces Verticillus ATCC15003, the Producer of the Hybrid Peptide-Polyketide Antitumor Drug Bleomycin. *Chem. Biol.* **2001**, *8* (7), 725–738.

(163) Nazi, I.; Koteva, K. P.; Wright, G. D. One-Pot Chemoenzymatic Preparation of Coenzyme A Analogues. *Anal. Biochem.* **2004**, *324* (1), 100–105.

(164) Kouprina, N.; Larionov, V. Selective Isolation of Genomic Loci from Complex Genomes by Transformation-Associated Recombination Cloning in the Yeast Saccharomyces Cerevisiae. *Nat. Protoc.* **2008**, *3* (3), 371–377.

Appendix

Protein sequences

N-His_6-GbnD1 ER1

MHHHHHHGKP IPNPLLGLDS TENLYFQGID PFT

```
       10         20         30         40         50         60
AAPSPSGEAS TPLPAGDYGL VVRTVHALDE LSVEPWTPGE PGDDEVLIEV RASALNFPDV

       70         80         90        100        110        120
MCVQGLYPTQ PAYPFVPGFE VAGVVAAVGR AVVGIRVGEA VLALTGERMG GLASRVVVPA

      130        140        150        160        170        180
ANVLPKPSRL SFEEAASLPV AFLTAHHAFE TGRLAAGERV LIQTATGGCG LAAIQLARLR

      190        200        210        220        230        240
GARVYGTSSR AAKRALLERI GVEHVLDYRA AFDRELAGLT DGRGVDVVLN MLSGDAIQRG

      250        260        270        280        290        300
LDSLAPAGRY VEIAVHGLRT SGTLDLSRLV DNQSFHSIDL RRARARGLDW GETLAGLLPL

      310        320        330        340        350        360
FETGALVPIV SRVYPAEQLG EALRYVATGE HVGKVVIRHR GGALDDCVEA CVSRLVAQRE

      370        380        390
IAAREAARPA SLPVVPRRAS AAAQPASETH AP
```

N-His_6-GbnD1 ACP1-ER1

MHHHHHHGKP IPNPLLGLDS TENLYFQGID PFT

```
10         20         30         40         50         60
PEPAEAGNVA VQATDADSFD LNISHDEIIE RSVAPDAATS RRRTIDSGPD LILERELAES
```

```
        70         80         90        100        110        120
LGQSLYLPAD EIDPERPFVE LGLDSIVGVE WARAINKAYG IALPATRLYD HVTIRLLAAH

       130        140        150        160        170        180
LVEQWGVARR DSARPAEPPV GLVAPVTRAP IAAAPSPSGE ASTPLPAGDY GLVVRTVHAL

       190        200        210        220        230        240
DELSVEPWTP GEPGDDEVLI EVRASALNFP DVMCVQGLYP TQPAYPFVPG FEVAGVVAAV

       250        260        270        280        290        300
GRAVVGIRVG EAVLALTGER MGGLASRVVV PAANVLPKPS RLSFEEAASL PVAFLTAHHA

       310        320        330        340        350        360
FETGRLAAGE RVLIQTATGG CGLAAIQLAR LRGARVYGTS SRAAKRALLE RIGVEHVLDY

       370        380        390        400        410        420
RAAFDRELAG LTDGRGVDVV LNMLSGDAIQ RGLDSLAPAG RYVEIAVHGL RTSGTLDLSR

       430        440        450        460        470        480
LVDNQSFHSI DLRRARARGL DWGETLAGLL PLFETGALVP IVSRVYPAEQ LGEALRYVAT

       490        500        510        520        530        540
GEHVGKVVIR HRGGALDDCV EACVSRLVAQ REIAAREAAR PASLPVVPRR ASAAAQPASE

THAP
```

N-His$_6$-SUMO-GbnE

```
MGSSHHHHHH GSGLVPRGSA SMSDSEVNQE AKPEVKPEVK PETHINLKVS DGSSEIFFKI

KKTTPLRRLM EAFAKRQGKE MDSLRFLYDG IRIQADQTPE DLDMEDNDII EAHREQIGG

        10         20         30         40         50         60
MAMITASSLG NADFRNDYRI KYAYLAGAMY RAIASKELVV ALGKAGLMGF LGTGGLRLDD

        70         80         90        100        110        120
IEAAILQIQA ELGVDGVYGM NLLADLERPE REEATVDLYL RHGVRHVEAA AYMRVTPALV

       130        140        150        160        170        180
RYRLRGASIG ADGRAVARHH VVAKVSRPEV ASQFMQPAPP ALVQQLVAAG RLDAREAELA

       190        200        210        220        230        240
```

```
PQLPLAGDIC VEADSGGHTD RGVAYVLIPA MQSLRDELMT RHRYAKRIRI GAAGGIGTPQ

       250        260        270        280        290        300
AAAAAFVMGA DFIVTGSINQ CTREAGTSDV VKALLQELDV QDTAYAPAGD MFELGAKVQV

       310        320        330        340        350        360
ARRGLFFAAR ASRLHELYQQ HASLDEIDPA MRSQIEQKFF RRGFDEIWDE TRRHYASSRP

       370        380        390        400        410        420
DQIAEIERSP KKKMAAVFRW YFAHSTQLAL SGDDTRLTDF QVHCGPALGA FNRWVRGTPL

       430        440        450        460
ESWQNRHVAD LAERIMQGTA AWLEARLASM ADSGGQRPSL QLQRQ
```

N-His$_6$-GbnD1 ACP3

```
MGSSHHHHHH SSGLVPRGSH

       10         20         30         40         50         60
TATVPTTVTT PTATENDALA ARAIEYFRKW LAAQLKVGAD QLDDDAPLDR YGIDSVRVMQ

       70         80         90        100        110        120
LTSALESRFG PLSKTLFFEY RNVAELSRHF VQAHRERMLD LLGLASPSAA PLPTSRPAGP

      130
APRPVLSSGP AI
```

N-His$_6$-GbnD2 ACP4

```
MGSSHHHHHH SSGLVPRGSH

       10         20         30         40         50         60
PATQRADVRD EITALYRQIQ AGSIERGEAA ARIARLRALA RDGGSDGSSG TDDGPAVPLA

       70         80         90        100        110        120
AIVEAVKTRV AAVLEVPLAR IEVDAPLDRY GIDSVTVMRL TGEFERELGP LSKTLLFEYR

      130        140        150        160        170        180
TVGEVSRYFS VSHAASVARW LGVAASAPSA PPGAYQAGAR TMRAMAVATR GLAVPPATAS A
```

N-His$_6$-GbnD2 ACP5

MHHHHHHGKP IPNPLLGLDS TENLYFQGID PFT

```
        10         20         30         40         50         60
PNPASLSSVP GGQDATRAAS AALLADEALR HVKRQLAAVI RLPVERIDED ASFEEYGIDS

        70         80         90        100        110        120
VMAVELTDRL ERACGPLSKT LLFEYQSVRD LTAFLVRHHA EGLGAALGLG EASSDGETAK

AEV
```

N-His$_6$-GbnD4 ACP10

MGSSHHHHHH SSGLVPRGSH

```
        10         20         30         40         50         60
AQAKPVAGAA PAAAGVPAPA PIADAAELGA RVDAALARAV CEILRIGESD VDAETDFSAY

        70         80         90        100        110        120
GFDSISLTEF SNRIGERLGV ELLPTIFYEY PSLLALRGFL LAEHGAALAA ALQVQPAAGR

       130        140
QVEPDPEPHR ESNRESNPGA AAVGLPEG
```

N-His$_6$-GbnD5 ACP12

MHHHHHHGKP IPNPLLGLDS TENLYFQGID PFT

```
        10         20         30         40         50         60
VAAGYDGARA AALAAGESTR ASFDEALRRF VTDQLAAQGV ALAGRLGDDT PFFDAGLDST

        70         80         90        100        110
HLLALVRALE THCGRTFYPT LLFEHQTLRE LAAHLHRETP AAFGQAVPVW SESVAA
```

N-His$_6$-GbnD2 Module5

MGSSHHHHHH SSGLVPRGSH

```
        10         20         30         40         50         60
AGARTMRAMA VATRGLAVPP ATASASLPQD RPEATATGTQ AVAVIGLAGR YPQAADLDAF

        70         80         90        100        110        120
WENLSTGRDC ITEIPSTRWD HEVYFDARKG QPGKSYSKWG GFLDGVDEFD PSFFSISPRE

       130        140        150        160        170        180
AQLMDPQERL FLQCAYHALE DAGHTRASLG AARVGVFVGV MYEEYPYHGS PSQGTTQPQA

       190        200        210        220        230        240
LGGSSASIAN RVSYAFNLNG PSIAMDTMCS SSLTALHLAC QSLRLGECEL ALAGGVNVSI

       250        260        270        280        290        300
HPNKYLGLSQ GQFASSEGRC RSFGAGGDGY VPSEGVGCVL LRPLAAAEAA GDRILGVIRA

       310        320        330        340        350        360
SAINHGGRTN GYTVPNPNAQ GELIAEALRA SGVDARAISY LEAHGTGTAL GDPIEIAGLV

       370        380        390        400        410        420
KAYGAWEGEP GEPGEPGDAR LEPCAIGSVK SNIGHCESAA GIAGLTKVLL QMRHGKLAPS

       430        440        450        460        470        480
LHAQTLNPLI DFGRTPFRVQ RELAPWRRPR VRVDGVEREM PRLAGISSFG AGGANAHVIV

       490        500        510        520        530        540
EEYVARAVEA ADARREGQPA IVVLSARSEA QVLIQARNLQ AAIEREAYGE GELAALAHTL

       550        560        570        580        590        600
QAGREAFEVR LATTVTSMAM LVERLASLAG EAPDYSAWMR GETRRDAADL PAPETVEGWL

       610        620        630        640        650        660
AQGRVEPLLR AWLDGLAFDW RRLRAADPPI RPLRLPGYPF ARQRYWADPP AAAAPRVARL

       670        680        690        700        710        720
HPLLHENRST AFEPQFVSHF DGREACLADH RVAHARVLPG AAQLEIARAA AALSLGEAAA

       730        740        750        760        770        780
LELTEVAWLQ PACFDEQGGS LRIALFVDSE TEAEFDIGSL DGDTVYSQGR LRLREAASPV

       790        800        810        820        830        840
VLELASLRAA CGQVPLAPEA LYASFAAAGL QYGPAHRAIA ELRTGLDTAG RPQVLARLEV

       850        860        870        880        890        900
PDGAADEALV LHPSLVDGAI QATIGLLLVP GGARRAMLPA RLGRVGVAAA TPARGWAWLR
```

```
       910        920        930        940        950        960
FAEGSGPEDV APRIDLSICD EHGAVALEIE ALTLMPAPVA SQTGVETLWL APAWAEAEAP

       970        980        990       1000       1010       1020
LDAPRGAAQP TREVFVAGQL PEGVLAALAR RYGESRVHVA PGADAMPLAA GYVAAADALF

      1030       1040       1050       1060       1070       1080
AIVRERLADV ASGECLFQLL RVEDGTAGAA CLDGLGALLR TASIENPRFA TQSIRIDAAD

      1090       1100       1110       1120       1130       1140
ATDGETLLAL LDANARERDS REIAYLDGRR RQRRHLELSQ PGGDVPAWPA GTVALITGGL

      1150       1160       1170       1180       1190       1200
GGIGYRVAES IAASGPGTTL LLCGRSEPAD AAARLGALRA AGANAEFVRT DITDAAATEA

      1210       1220       1230       1240       1250       1260
LVAGIVARHG RLDTVIHSAG IVRDNFVIRK NASELHAVLA PKVAGLVNLD AATRSLPLRE

      1270       1280       1290       1300       1310       1320
LIVFSSVSGA FGNAGQADYA CANAFMDAFA AWRATRVAAG ERSGRTLSIG WPLWAEGGMR

      1330       1340       1350       1360       1370       1380
VDAAVEAKLR RDGLAPLDTA SGLAVLARCR QLPPELTQVV VLAGDAARLR ARHGLAAAAI

      1390       1400       1410       1420       1430       1440
GAAASAPNRA PNPAPNPASL SSVPGGQDAT RAASAALLAD EALRHVKRQL AAVIRLPVER

      1450       1460       1470       1480       1490       1500
IDEDASFEEY GIDSVMAVEL TDRLERACGP LSKTLLFEYQ SVRDLTAFLV RHHAEGLGAA

      1510
LGLGEASSDG ETAKAEV
```

N-His$_6$-GbnD2 KS6-ACP6a and N-His$_6$-GbnD2 Module6(S941A)

```
MGSSHHHHHH SSGLVPRGSH

        10         20         30         40         50         60
DGETAKAEVA VAALAAGAVA TARAAGPALA PVAPRTRRRP RGRLPGRESE PPRITAIAVI

        70         80         90        100        110        120
GLAGRYPQAA DLDAFWENLS TGRDCITEIP STRWDHEAYF DARKGQPGKS YSKWGGFLDG
```

```
        130         140         150         160         170         180
VDEFDPFFFN ISPREAQLMD PQERLFLQCA YHALEDAGHT RASLGAVRVG VFVGVMYEEY

        190         200         210         220         230         240
HLLSDPAGGD LAQTYIPGGY LSSVANRVSY FGNFRGPSFG VDTMCSSSLT ALHLACQSLR

        250         260         270         280         290         300
LGECELALAG GVNVSIHPNK YLGLSQGQFA SSEGRCRSFG AGGDGYVPSE GVGCVLLRPL

        310         320         330         340         350         360
AAAEAAGDRI LGVIRASAIN HGGRTNGYTV PNPNAQGELI AEALRASGVD ARAISYLEAH

        370         380         390         400         410         420
GTGTALGDPI EIAGLVKAYG AWEGEPGEPG DARLEPCAIG SVKSNIGHCE SAAGIAGLTK

        430         440         450         460         470         480
VLLQMRHGKL APSLHAQTLN PLIDFGRTPF RVQRELAPWR RPRVRVDGVE REMPRLAGIS

        490         500         510         520         530         540
SFGAGGANAH LIVEEYVARA VEAADARREG QPAIVVLSAR SEAQVLIQAR NLQAAIAREA

        550         560         570         580         590         600
YGEGELAALA HTLQAGREAF EVRLATTVTS MAMLVERLAS LAGETPDYSA WMRGEARRDG

        610         620         630         640         650         660
NDALAHFAKD PELNEVLRKW LRAGNPVRIA ELWTRGLDID WAAMQEDGPA PARLRLPGYP

        670         680         690         700         710         720
FATKRYWMAR ADGSPALAAR AAVPARPEPP TRAAVPPAAV KVLAEVLADA PRVAIVLGEP

        730         740         750         760         770         780
GGFEARTEGA ARAASIRLAV LDPLDEFEPP AAPAAARQDD GQTAAAARQV PAGIEVSIER

        790         800         810         820         830         840
LRLSLAQMLC AEAADLGADV PFGELGLDSV VGVEWTRAIG REYGVTLAAA TLYDHPTLTT

        850         860         870         880         890         900
LAAWLAATVE AGAAIATAGR DVPPRPDAHP VRLVEPDQDV ATPLAAAAAV EGGNEAEVAA

        910         920         930         940         950         960
GMAMPIEALV ERLAGTLAQA LCAEREEIDP DTPFAELGID SVVGVEWVRD INRAFGTELK

        970         980         990        1000        1010        1020
```

```
ATVLFDHATV KLLAAHLAPL ACPAPHSVRV DPPAAPVEQL AAASAVPARE LADTARHAGS

      1030       1040
RGHVQAVEPA APVAKTGNAP
```

N-His$_6$-IcoA PCP3

```
MHHHHHHGKP IPNPLLGLDS TENLYFQGID PFT

        10         20         30         40         50         60
RYEAPQGETE QAVAALWAEL LGVERIGRHD NFFALGGHSL LAVRMLNRLR AMQAGDLSLS

        70         80
SLFDHPTVSA VAQAIDSGSR SVA
```

N-His$_6$-IcoA Module 3 (C3-C4-A3-PCP3), N-His$_6$-IcoA C4-A3-PCP3 and N-His$_6$-IcoA C4

```
MGSSHHHHHH SSGLVPRGSH

        10         20         30         40         50         60
MRAHDLVRID RNVTMPLSFA EQRVLAIEAG GTRGAMLNSS RLFILSGRVR EDVLDRALRA

        70         80         90        100        110        120
VVERHEILRT HYAGDETTGS FVPVIGRSAL FHLERREVAE DDAMPIAQSE AARHLDCFRG

       130        140        150        160        170        180
PVFGATLFVE SPRRAILMIA IHHIAMDAAS WNIVWRDFRQ AYAALAADEP PALAPLALQY

       190        200        210        220        230        240
RDYAAWTTKQ VDAERLAQLR QTWRERLAGA PVYLNLPADR PRPDKLSDRA GRIERMLAPE

       250        260        270        280        290        300
LVQGIRQIAT AYKVTPFIVL ETTLAIACAQ LARSRDIVIG TVTEGRTHAA TEDMVGLFVN

       310        320        330        340        350        360
TLPLRHRLDR TASVASHLLA TSRELLSALE ASELPFADVV AAVNPPRASS HDPIFQVFCQ

       370        380        390        400        410        420
FQQGPGKTRE TIADMTIEAI PRTRMARGAD LAVVFLDAGE YLSADVSYSA DLFDAESIDA

       430        440        450        460        470        480
```

```
LIELFLRFLA ESVQRPDAVV DELWDAGVAS ARAVASGPDV SRLVDRPGSP AETWYALSPA

       490        500        510        520        530        540
QQAVWLQEQA ARPGTSFFSV AAMRCAPEID RMRLVAASRA LIAQNQAFWI QVSDAGLQCE

       550        560        570        580        590        600
ASTPTTRFQH FVEPTTMTGE AMRRAVIEWH EKLNEDPRDK SGAVAVFDSP GSVLVALRSP

       610        620        630        640        650        660
HVQNDGWSAL RYFERIGRNY AALESDPARA FDMDRIFLDT LSLDERYLCS PTYERDAAFW

       670        680        690        700        710        720
QSACARIDGQ PLVTLVADHA HPVDARGVVR SLRKVFPQTL QERVLNAARK LSLSPAECLT

       730        740        750        760        770        780
ALTALYLMRV TGERSTVLGV SFLNRTREAL DIPGQFAKVI PLPVSIGQGD IPLSSTLNGI

       790        800        810        820        830        840
RDAFKDVMQH GRFPFGEMVR RYGFDPRHIE ISVNTLFLRH PVEVGGQPAH VQWLSGPEHG

       850        860        870        880        890        900
LSFLFTQFGR SAPIDIELRY NGNAFDSESV ERHARRLLDF IERACEDDSV SARGIELVSS

       910        920        930        940        950        960
EERALLIDAL NATDAPYDRN QYLHGLFEAQ AKRTPEATAL IAGDERLSYA ELDARANRFA

       970        980        990       1000       1010       1020
RYLVDLGVGP DALVAVCLER STAMVVSLIG ILKAGGAYVP IDPAYPGPRI AHIVSDSAPA

      1030       1040       1050       1060       1070       1080
VVLVDATGRE ALVDALGDEK LAEYGLIDVS AASTPWNKLS SDSLSSNALG LNPRHLAYVI

      1090       1100       1110       1120       1130       1140
YTSGSTGMPK GVQNEHDALV NRLTWMQEAY RLGGQDVVLQ KTPFSFDVSV WEFFWTLANG

      1150       1160       1170       1180       1190       1200
ATLVIAEPGA HRDADYLTEI IAKHGVTTLH FVPSMLAGFL EAQDLTRCKT LSRIICSGEA

      1210       1220       1230       1240       1250       1260
LPAPIARRCL ERLPHAQLHN LYGPTESAID VTAFTCPPDF DAQAVPIGKP IANTRIYLLD

      1270       1280       1290       1300       1310       1320
ERQAPVPLGA IGELYIGGVG VARGYLNRAD LTAQRFLADP FARAAGHPEA RMYRTGDLAR
```

```
      1330        1340        1350        1360        1370        1380
YLPDGNIVFL GRNDDQVKIR GFRIELGEIE VQLAKHEAVR DAIVIARQDS TGNARLLAYV

      1390        1400        1410        1420        1430        1440
TPQESASREE LARSLREHLT ARLPEYMVPA AFVVLETLPL TPNGKLDRRA LPEPADDAFV

      1450        1460        1470        1480        1490        1500
QSRYEAPQGE TEQAVAALWA ELLGVERIGR HDNFFALGGH SLLAVRMLNR LRAMQAGDLS

      1510        1520
LSSLFDHPTV SAVAQAIDSG SRSVA
```

Protein Mass Spectra

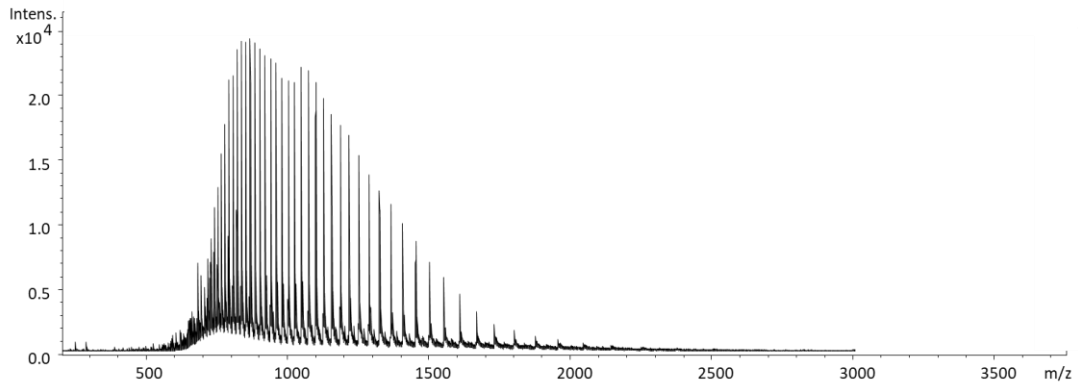


Figure S1 High resolution mass spectrum of N-His_6-GbnD1 ER1.

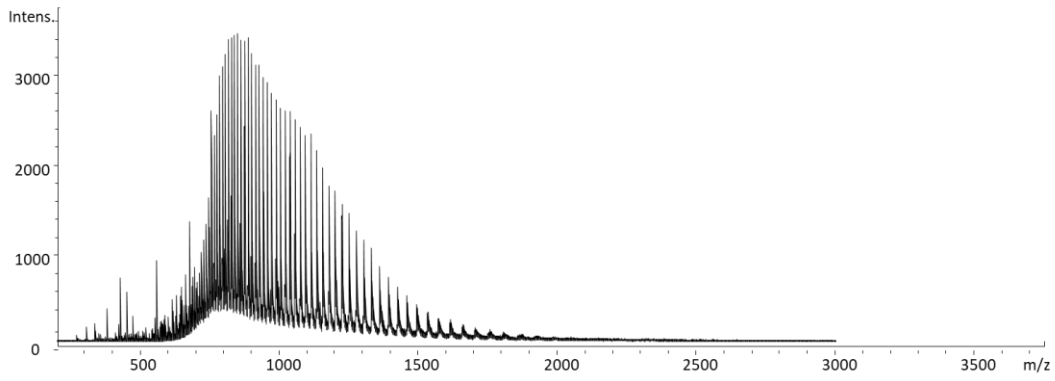


Figure S2 High resolution mass spectrum of N-His_6-GbnD1 ACP1-ER1.

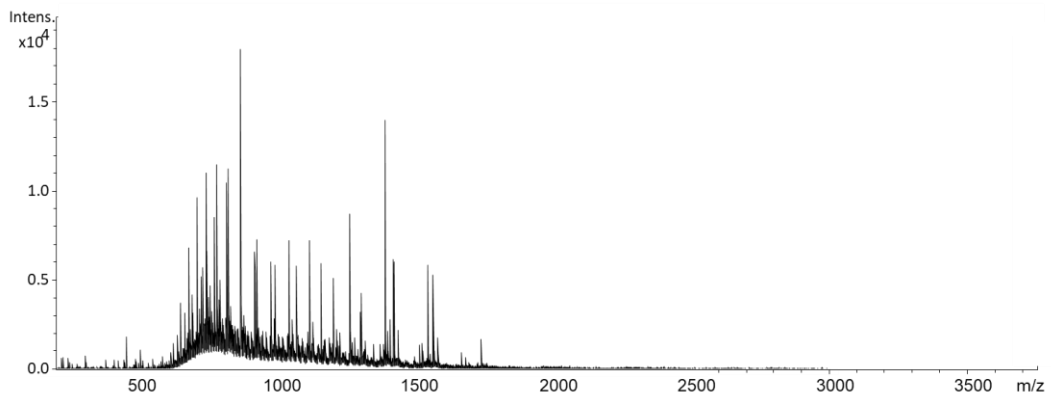


Figure S3 High resolution mass spectrum of N-His_6-SUMO-GbnE

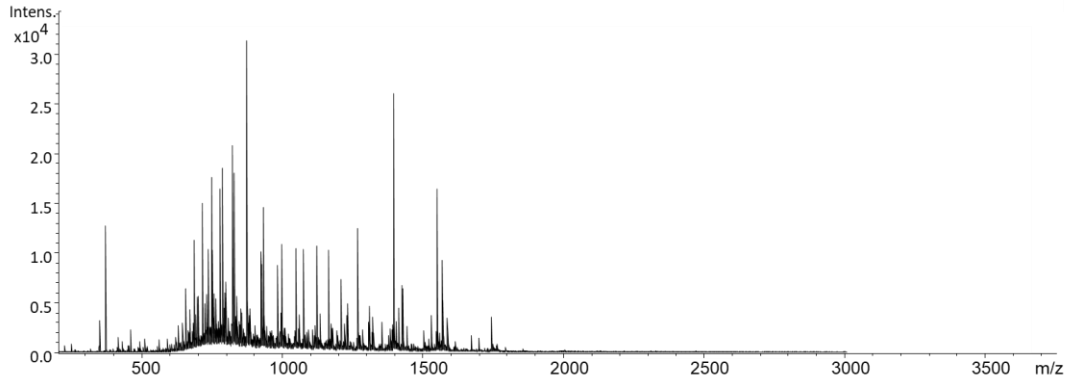


Figure S4 High resolution mass spectrum of N-His_6-SUMO-GbnE(H198V)

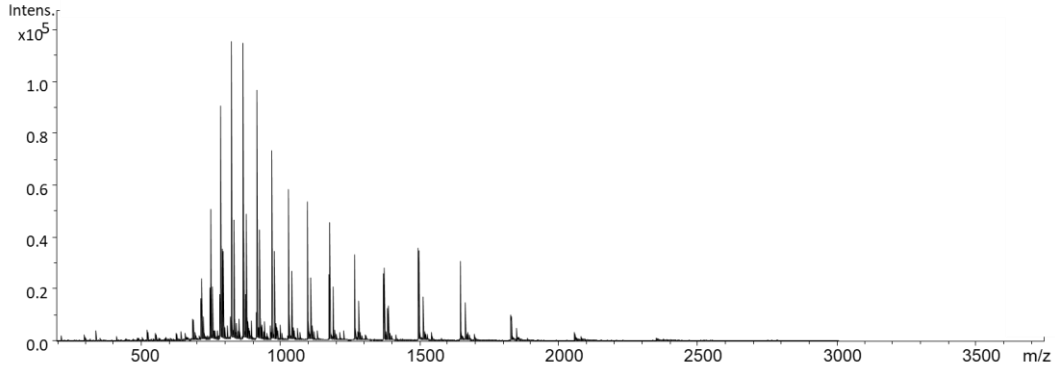


Figure S5 High resolution mass spectrum of N-His_6-GbnD1 ACP3

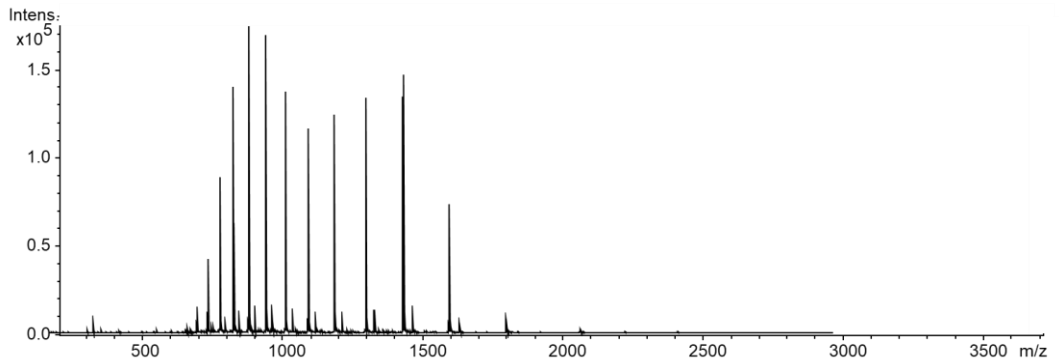


Figure S6 High resolution mass spectrum of GbnD1 ACP3 (His-tag cleaved)

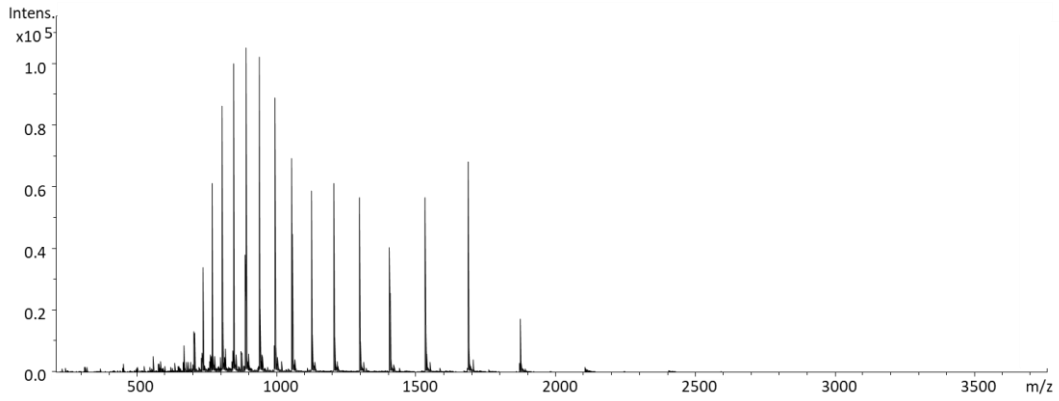


Figure S7 High resolution mass spectrum of N-His_6GbnD2 ACP5.

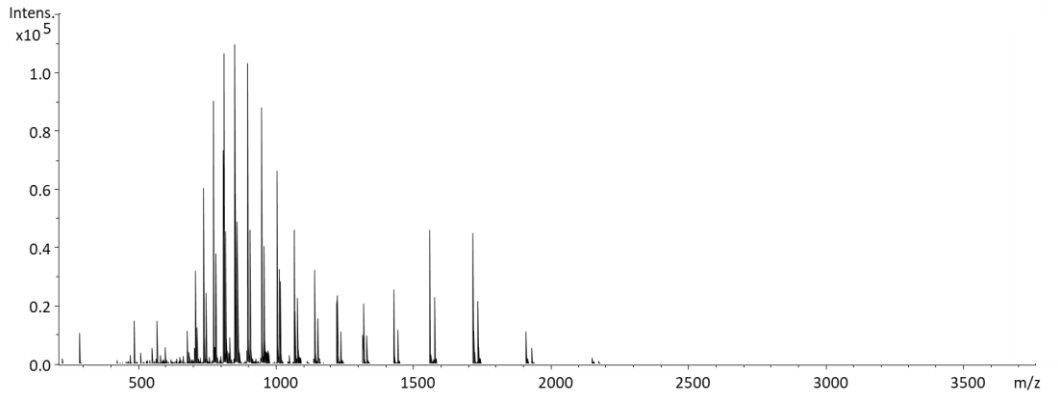


Figure S8 High resolution mass spectrum of N-His_6-GbnD4 ACP10

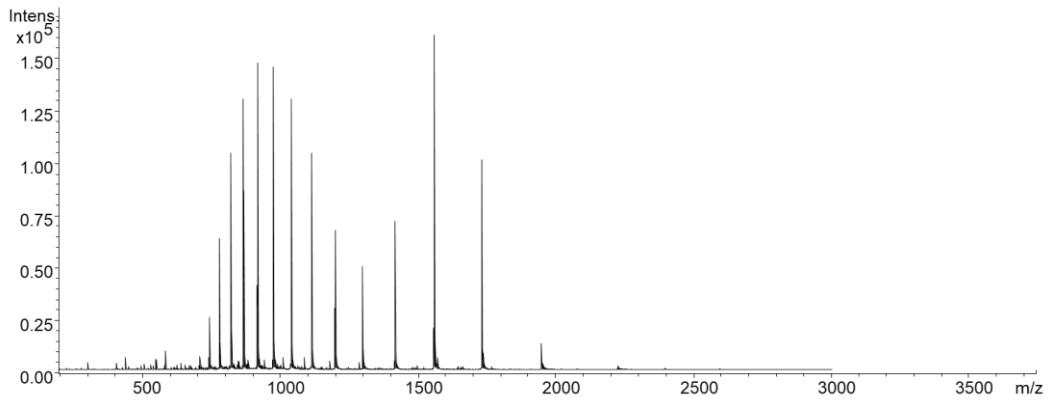


Figure S9 High resolution mass spectrum of GbnD4 ACP10 (His-tag cleaved)

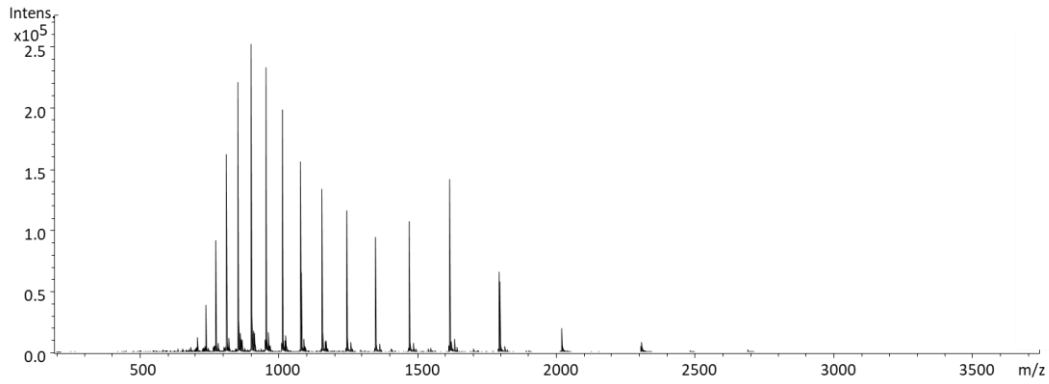


Figure S10 High resolution mass spectrum of N-His_6-GbnD5 ACP12

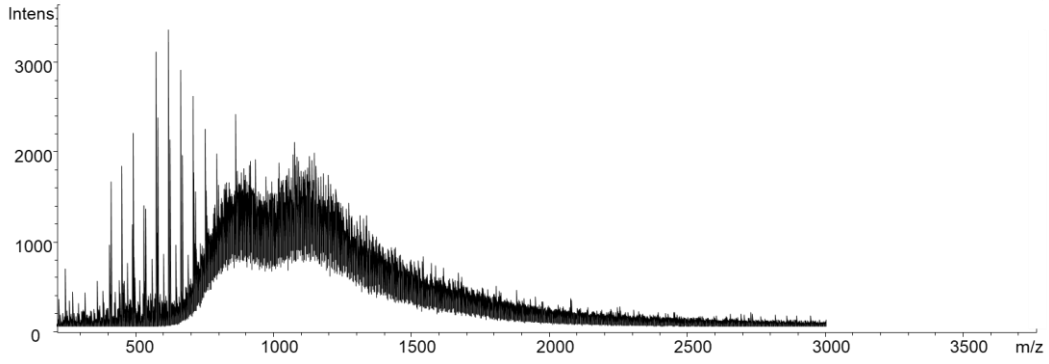


Figure S11 High resolution mass spectrum of N-His_6-GbnD2 Module 5

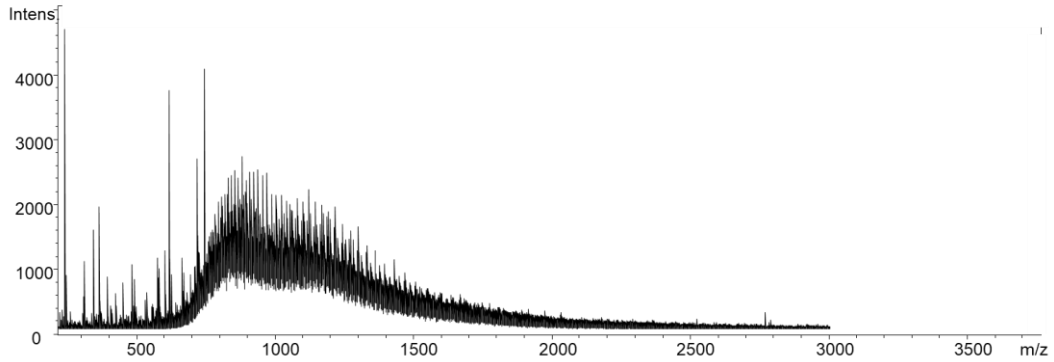


Figure S12 High resolution mass spectrum of N-His$_6$-GbnD2 Module 6

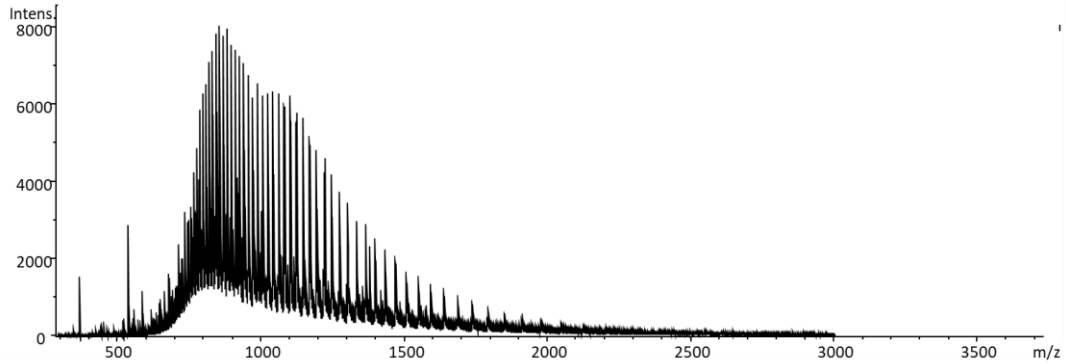


Figure S13 High resolution mass spectrum of N-His$_6$-IcoA C4

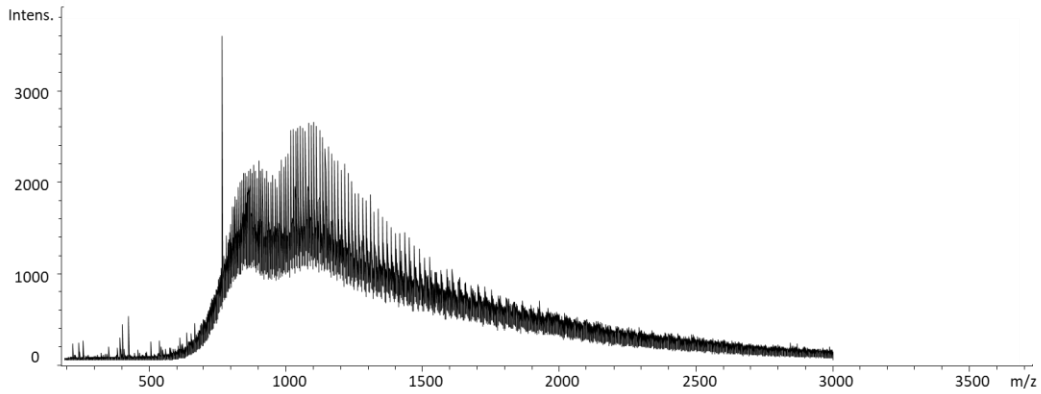


Figure S14 High resolution mass spectrum of N-His$_6$-IcoA C4-A3-PCP3

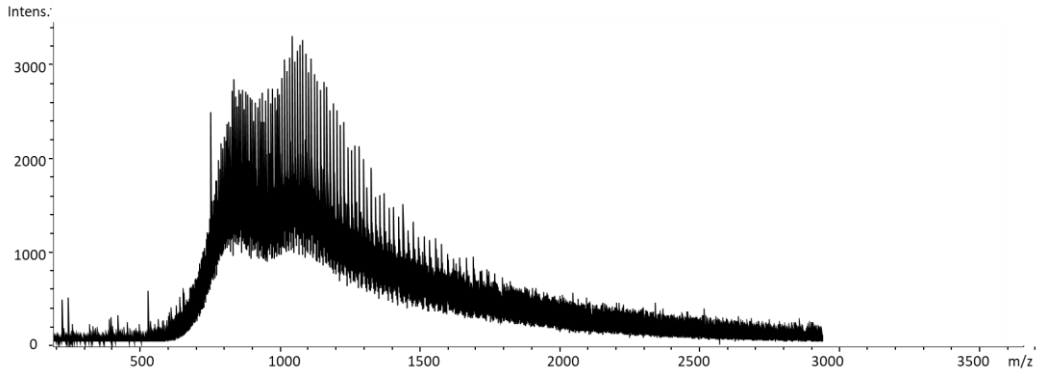


Figure S15 High resolution mass spectrum of N-His$_6$-IcoA C4-A3-PCP3 (H143A)

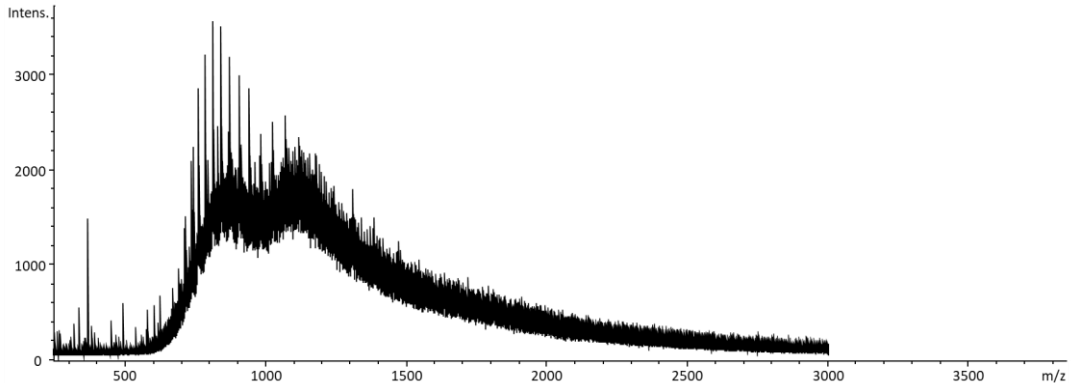


Figure S16 High resolution mass spectrum of N-His$_6$-IcoA Module 3 (C3-C4-A3-PCP3)

Sequence alignments

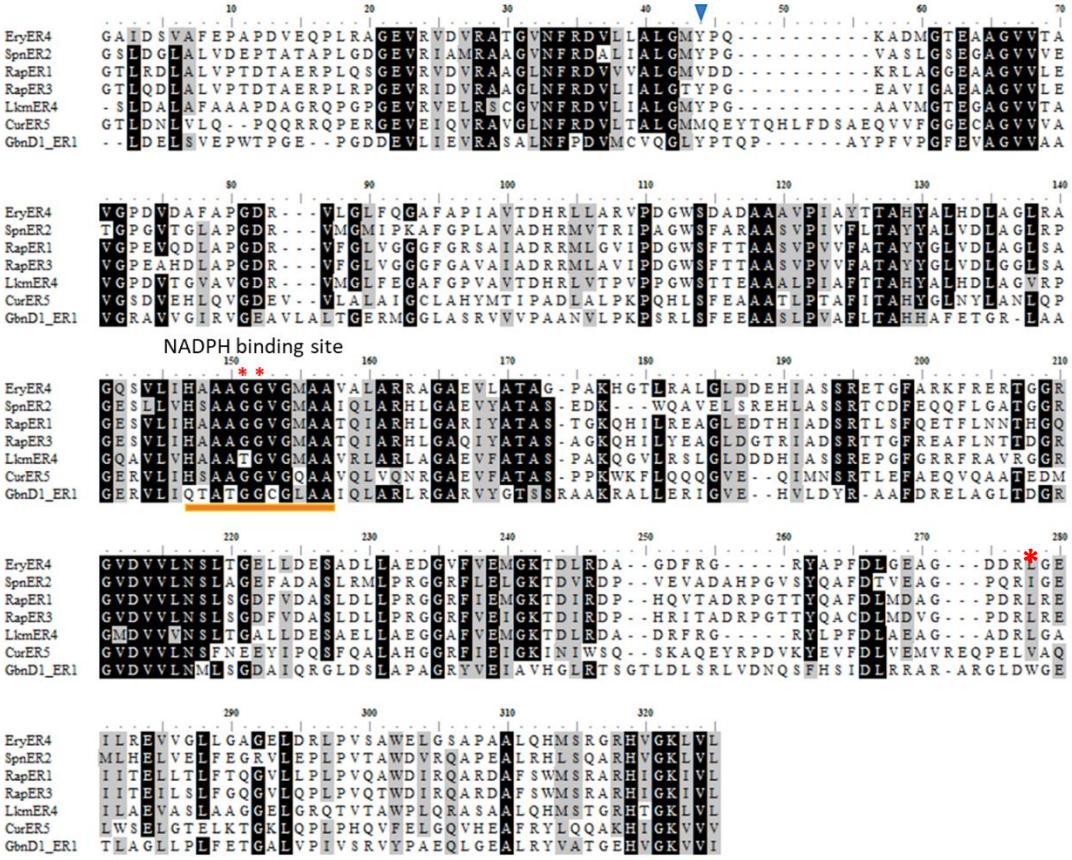


Figure S17 Multiple sequences alignment of GbnD1 ER1 domain with integrated ER domains from modular PKSs. Ery: erythromycin, Spn: spinosyn, Rap: rapamycin, Lkm: lankamycin, Cur:curicin. 'ERx' refer to the ER domain in the module x.

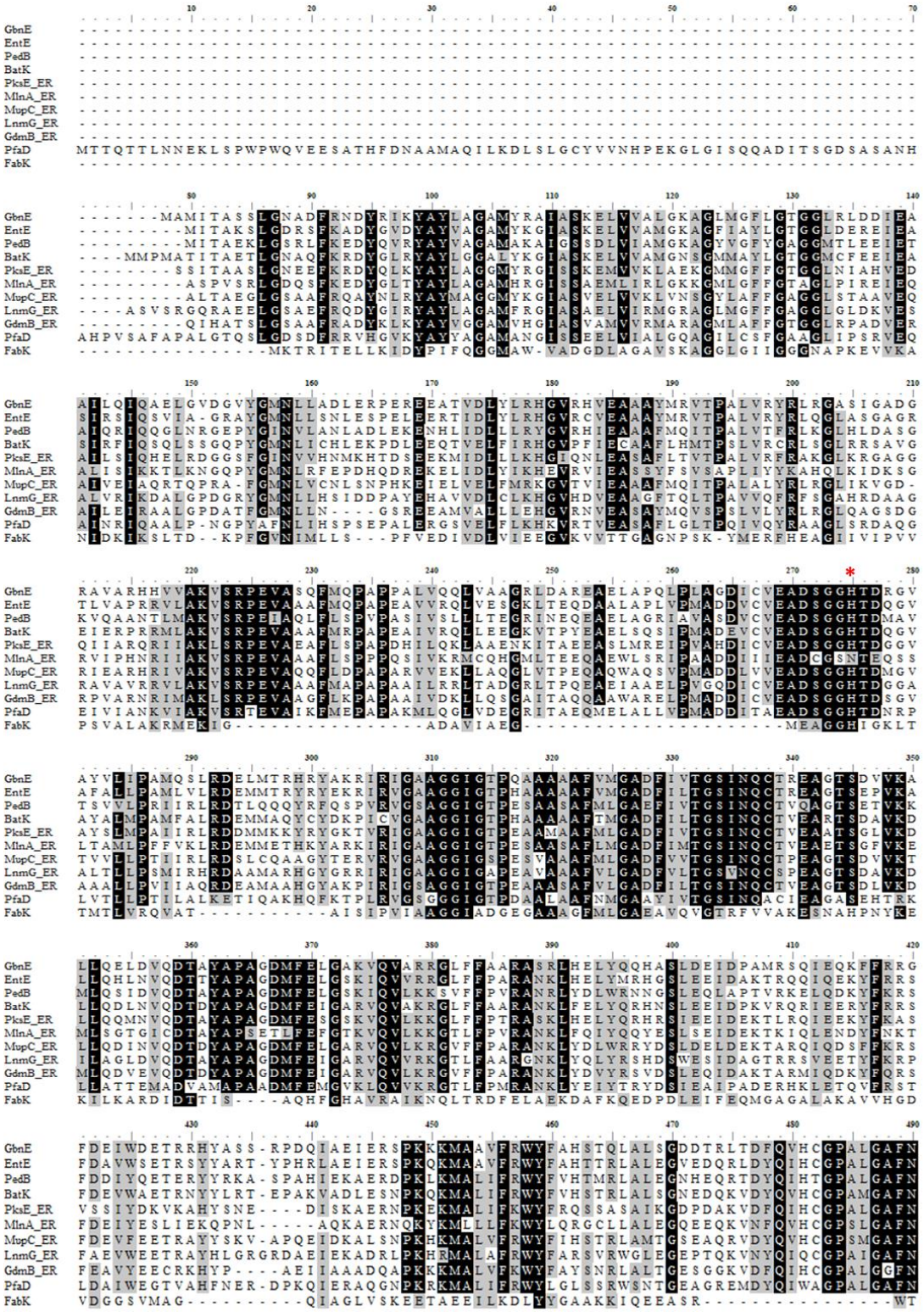


Figure S18 Multiple sequence alignment of GbnE and *trans*-acting ERs from *trans*-AT PKS. Ent: etnangien, Ped: pederin, Bat: Batumin/kalimantacin, Pks: dihydrobacillaene, Mln: macrolactin, Mup: mupirocin, Lnm: leinamycin, Gdm: gladiostatin.

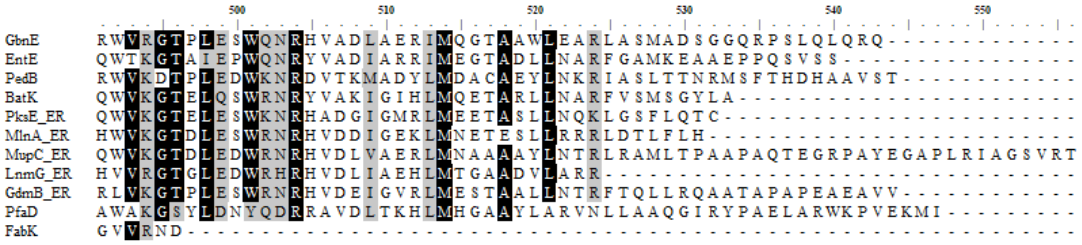


Figure S18 (continue) Multiple sequence alignment of GbnE and *trans*-acting ERs from *trans*-AT PKS. Ent: etnangien, Ped: pederin, Bat: Batumin/kalimantacin, Pks: dihydrobacillaene, Mln: macrolactin, Mup: mupirocin, Lnm: leinamycin, Gdm: gladiostatin.

Mass spectra of extracted ion chromatograms of gladiolin and related derivatives

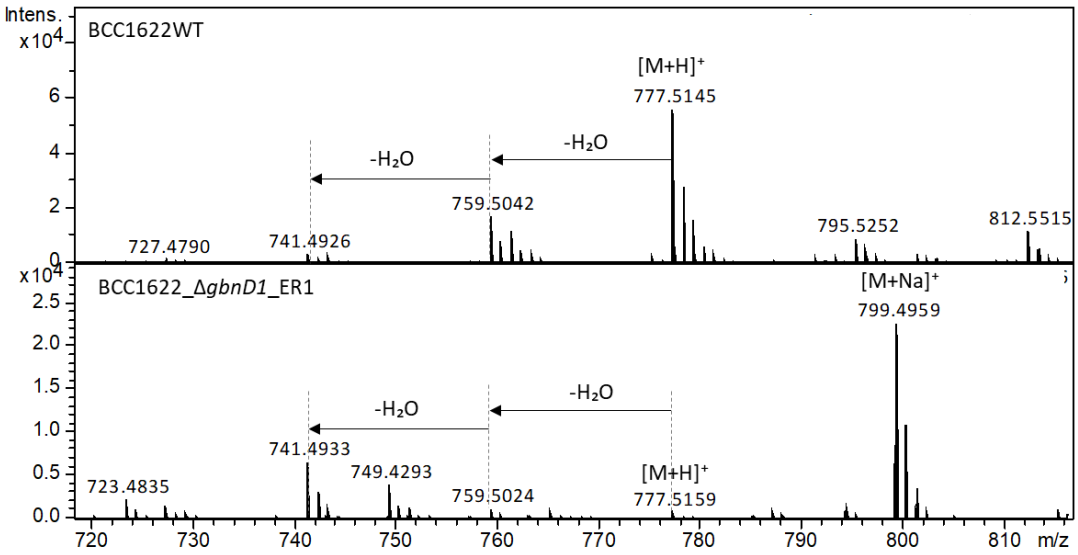


Figure S19 Comparison of high-resolution mass spectra of EICs at *m/z*= 777.5153 ± 0.02 Da and 799.4972 ± 0.02 Da, corresponding to $[M+H]^+$ and $[M+Na]^+$ ions of gladiolin derivatives **33** and **34** in *B.gladioli* BCC1622 WT and **32** in *B.gladioli* BCC1622_Δ*gbnD1*_ER1

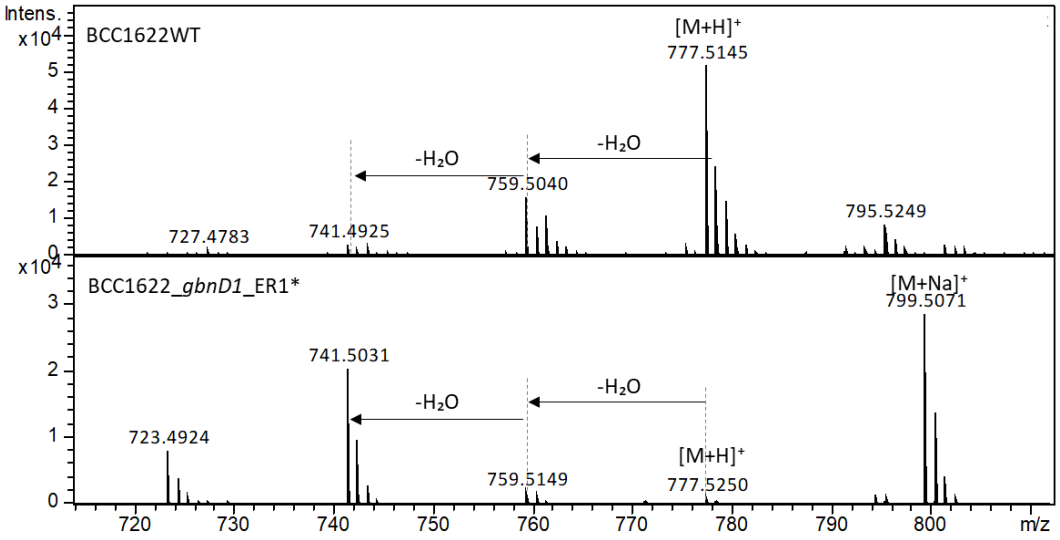


Figure S20 Comparison of high-resolution mass spectra of EICs at *m/z*= 777.5153 ± 0.02 Da and 799.4972 ± 0.02 Da, corresponding to $[M+H]^+$ and $[M+Na]^+$ ions of gladiolin derivatives **33** and **34** in *B.gladioli* BCC1622 WT and **32** in *B. gladioli* BCC1622_*gbnD1*_ER1*(G388S/G388P).

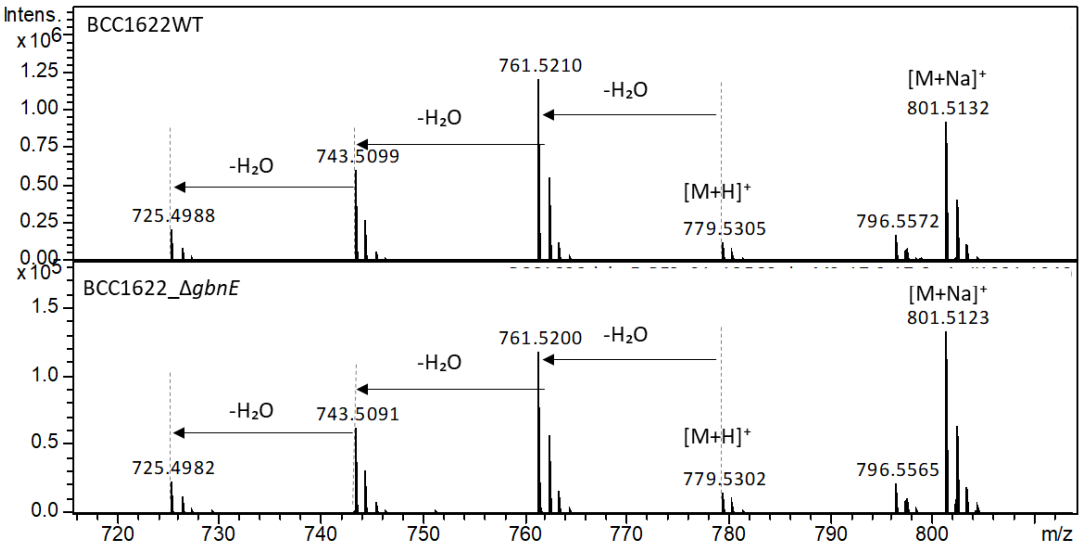


Figure S20 Comparison of high-resolution mass spectra of EIC at *m/z*=779.5309 ± 0.02 Da and 801.5129 ± 0.02 Da, corresponding to the $[M+H]^+$ and $[M+Na]^+$ ions of gladiolin **7** and iso-gladiolin **28** in *B.gladioli* BCC1622 WT and *B.gladioli* BCC1622_Δ*gbnE*

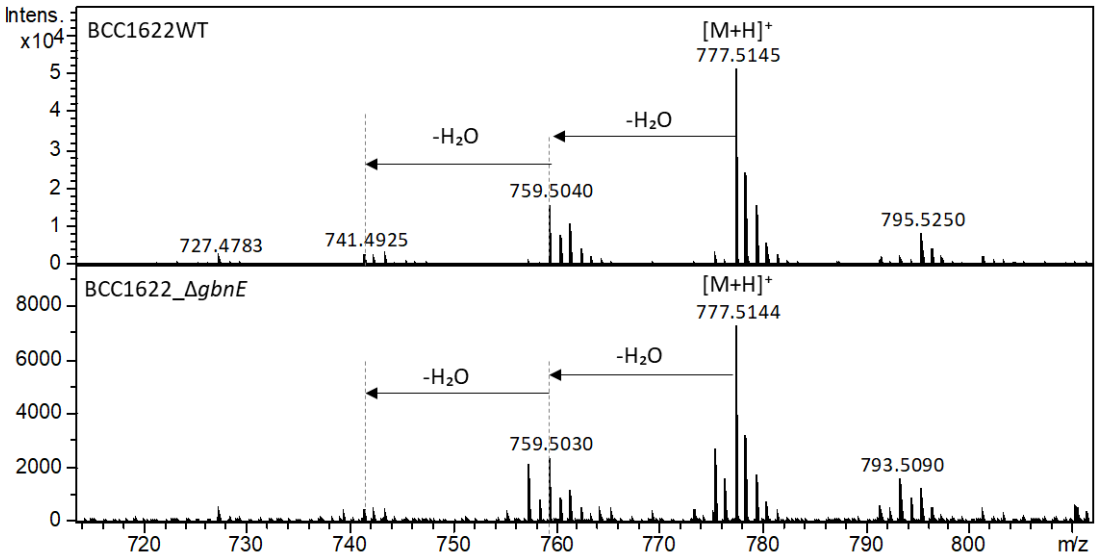


Figure S 21 Comparison of high-resolution mass spectra of EICs at *m/z*= 777.5153 ± 0.02 Da and 799.4972 ± 0.02 Da, corresponding to $[M+H]^+$ and $[M+Na]^+$ ions of gladiolin derivatives **33** and **34** in *B.gladioli* BCC1622 WT and *B.gladioli* BCC1622_Δ*gbnE*