

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

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/167175>

Copyright and reuse:

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

Please scroll down to view the document itself.

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

Our policy information is available from the repository home page.

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



Discovery and biosynthesis of gladiostatin:
A novel glutarimide antibiotic from *Burkholderia*
gladioli

Ioanna Theodora Nakou

A thesis submitted in partial fulfilment of the requirements of the degree of

Doctor of Philosophy in Chemistry

Department of Chemistry

University of Warwick

October 2021

Table of Contents

Acknowledgements	xi
Declaration	1
List of abbreviations	2
Abstract	4
Chapter 1 Introduction	5
1.1 Natural products	6
1.2 Genome mining to exploit the hidden biosynthetic potential of biologically active compounds	8
1.3 <i>Burkholderia</i> as an untapped source of natural products	9
1.4 Polyketide natural products	12
1.5 Polyketide synthases	13
1.5.1 Modular type I polyketide synthases	14
1.5.2 Post-PKS tailoring	29
1.5.3 <i>Trans</i> -AT PKSs	29
1.6 Study objectives	33
Chapter 2 Discovery, biosynthesis and biological evaluation of gladiostatin	34
2.1 Introduction	35
2.1.1 Biosynthetic potential of <i>B. gladioli</i> BCC0238 strain	35
2.1.2 Glutarimide-containing polyketides	36
2.2 Results and Discussion	43
2.2.1 Isolation and <i>de novo</i> structural elucidation of metabolic product derived from <i>gds</i> cryptic cluster	43
2.2.2 Biological activity of gladiostatin	48
2.2.3 Studies on the biosynthesis of gladiostatin	52
2.3 Conclusions	59
Chapter 3 Characterisation of <i>trans</i>-acting enoyl reduction steps in gladiostatin biosynthesis	61
3.1 Introduction	62
3.2 Results and Discussion	66

3.2.1 Overproduction and purification of putative proteins involved in enoyl reduction events during gladiostatin biosynthesis	66
3.2.2 <i>In vitro</i> biochemical characterisation of putative <i>trans</i> -acting enoyl reductases in GdsF modules 5-8.....	69
3.2.3 <i>In vivo</i> characterisation of <i>gdsH</i>	76
3.2.4 Isolation and structure elucidation of compound 55.....	81
3.2.5 <i>In silico</i> analysis of ACPs from <i>gds</i> BGC	83
3.3 Conclusions.....	86
Chapter 4 Investigating the mechanism of 2-acyl-4-hydroxy-3-methyl butenolide formation in gladiostatin biosynthesis	88
4.1 Introduction.....	89
4.1.1 Butyrolactones/butenolide substructures found in <i>Streptomyces</i> autoregulators	89
4.2 Previous studies on the polyketide chain release in gladiostatin biosynthesis	93
4.3 Results and Discussion	96
4.3.1 Identification of genes in <i>gds</i> cluster encoding putative phosphatases and investigation of their involvement in dephosphorylation of biosynthetic intermediates	96
4.3.2 <i>In vitro</i> reconstitution of the AHMB formation steps during gladiostatin biosynthesis	106
4.3.3 Gene complementation studies on $\Delta gdsG$ mutant using <i>sabP</i> and <i>p022</i> orthologs ..	116
4.3.4 <i>In vitro</i> characterisation of SabP and P022.....	118
4.4 Conclusions.....	121
Chapter 5 Comparison of the gladiostatin biosynthetic gene cluster in numerous <i>B. gladioli</i> isolates reveals a novel mechanism for trans-AT polyketide synthase evolution	124
5.1 Introduction.....	125
5.1.1 Engineering of modular PKSs inspired by evolution	125
5.2 Results and Discussion	129
5.2.1 Distribution of <i>gds</i> locus among <i>B. gladioli</i>	129
5.2.2 Secondary metabolite analysis reveals the presence of a gladiostatin variant	131
5.2.3 Domain-level analysis of the recombination event in the <i>gds</i> BGC	134
5.2.4 Evolution-guided engineering of the <i>gds</i> PKS	138
5.3 Conclusions.....	141
Chapter 6 Conclusions and future perspectives	142
6.1 Discovery of a polyketide bearing unusual functionalities (I)	145

6.2 Unravelling an unprecedented number of enoyl reductions performed by a sole <i>trans</i> -acting ER (II)	146
6.3 An unusual chain off-loading mechanism (III)	147
6.4 Evolution-guided PKS engineering (IV)	148
Chapter 7 Experimental	145
7.1 Instruments and Equipment	150
7.1.1 Semi-preparative HPLC	150
7.1.2 Mass Spectrometry	150
7.1.3 NMR spectroscopy	151
7.2 Materials	151
7.2.1 General	151
7.2.2 Culture Media	151
7.2.3 Antibiotics	152
7.2.4 Buffers	153
7.2.5 Strains and vectors	154
7.3 General DNA manipulation	157
7.3.3 Isolation of genomic DNA/plasmid DNA	157
7.3.4 Polymerase chain reaction (PCR)	157
7.4 Agarose gel electrophoresis	158
7.5 Construction of recombinant plasmids	159
7.5.1 Restriction digestion and ligation into pET28a(+)/pMLBAD	159
7.5.2 Restriction digestion and ligation into pGPI	159
7.5.3 Preparation of competent cells	160
7.5.4 Chemical transformation	160
7.5.5 Electroporation	160
7.6 Site-directed mutagenesis	161
7.7 Recombinant protein overproduction and purification	161
7.7.1 Recombinant protein expression constructs	161
7.7.2 Recombinant protein overproduction	163
7.7.3 Recombinant protein purification	163
7.7.4 SDS-PAGE gel electrophoresis	164
7.8 <i>In vitro</i> biochemical assays	165

7.8.1	<i>p</i> -Nitrophenyl (pNPP) colorimetric assay	165
7.8.2	Enoyl Reduction assay.....	165
7.8.3	Butenolide formation assay.....	166
7.9	Genetic Manipulation of <i>Burkholderia</i>	166
7.9.1	Insertional Mutagenesis in <i>Burkholderia</i>	166
7.9.2	In-frame gene deletions in <i>Burkholderia</i>	168
7.9.3	Gene complementation in <i>Burkholderia</i>	170
7.10	<i>Burkholderia</i> metabolite extraction and analysis.....	171
7.10.1	Analysis of metabolite profile of <i>B. gladioli</i> strains/knockout mutants and time-course analysis.....	171
7.10.2	Isolation and structure elucidation of gladiostatin and derivatives and time-course analysis.....	172
7.11	<i>In vivo</i> biological activity assays.....	177
7.12	Chemical synthesis of authentic standard	178
7.13	Computational methods	182
7.13.1	Sequence alignments.....	182
7.13.2	Phylogenetic analyses	182
7.13.3	Cladogram construction	182
	References	184
	Appendix	213

List of Figures

Figure 1.1.1 Overview of structures of antibiotics discovered during the golden era of antibiotics.	7
Figure 1.3.1 Examples of natural products derive from <i>Burkholderia sp.</i>	11
Figure 1.4.1 Examples of diverse polyketide scaffolds of therapeutic potential.....	12
Figure 1.5.1 Post-translational modification of the ACP domain. Attachment of phosphopantetheine from co-enzyme A to a conserved Ser residue onto the apo-ACP domain catalysed by phosphopantetheinyl transferase.	16
Figure 1.5.2 Structural analysis of modular type I ACP from the second module of DEBS and ACP domains from type II FAS and type I PKS with phosphopantetheine chain attached to the conserved Ser residue of the ACP. (A) The ACP comprises a bundle of four helices and an additional four residue helix ("helix 0") which is proposed to be involved in the docking to the proceeding module of the PKS (PDB: 2JU2). (B) Type II ACPs sequestering their substrates and type I ACPs does not, possibly because, as part of a multi-domain modular system, they can transfer between processing domains fast enough to avoid significant rates of hydrolysis.(PDB: 2FAE for FAS II ACP;PDB: 2LIW for type I PKS ACP).	17
Figure 1.5.3 The crystal structure of the ketoreductase domain from DEBS module 1 (PDB: 2FR0) and mechanism of reduction of the β -keto group catalysed by KR domain. (A) KR domain is comprised of a structural (shown in grey) and catalytic (shown in light blue). The catalytic subdomain binds NADPH, which is shown in spheres. (B) A ₁ -type and B ₁ -type KR domains furnish a different stereo outcome at C-3 position.....	21
Figure 1.5.4 Structure of cis-acting ER domain of spinosyn modular PKS (PDB: 3SLK). (A) Stereogram of SpnER2 (PDB: 3SLK) comprising of the substrate-binding (shown in cyan) and nucleotide-binding subdomain. NADPH binding site is placed in a cleft between the two subdomains. (B) Conserved Tyr-Asp-Lys catalytic residues located close to the nicotinamide moiety.	23
Figure 1.5.5 Comparison of a cis-AT PKS module harbouring integrated AT and ER domains to trans-AT PKSs module which encode standalone AT and ER domains.....	30
Figure 2.1.1 Biosynthetic potential of <i>B. gladioli</i> BCC0238. Comparison of chromosome 2 of the strain BCC0238 to the complete genome sequence of <i>B. gladioli</i> BSR3 (green, 99% identity) and the draft genome of NBRC 13700 (blue, 99% identity) and 3848s-5 (red, 99% identity). ¹³¹ The putative BGCs are shown in various colours. The orphan BGC studied here is indicated with a question mark.	36
Figure 2.1.2 Structures of known glutarimide antibiotics derived from Streptomyces species. The family of glutarimide-containing polyketides is subdivided here in: (A) glutarimides containing cyclohexanone/phenol moiety in C-4, (B) glutarimide-containing polyketides with a 12/14-membered	

macrolide, (C) acyclic members of the glutarimide antibiotics and (D) glutarimides bearing a modified furan moiety. The common 2, 6-piperidinedione moiety presented in all metabolites is highlighted in black..... 37

Figure 2.1.3 Comparative analysis of chx, smd, mgs, ltm BGCs that direct glutarimide biosynthesis in Streptomyces species to the cryptic gds BGC. (A) Comparison of biosynthetic gene clusters (BGCs) that direct glutarimide biosynthesis in Streptomyces species (lactimidomycin (ltm), migrastatin (mgs), cycloheximide (chx) and 9-methylstreptimidone (smd)) with the cryptic *B. gladioli* BCC0238/BCC1622 gene cluster (highlighted by the dashed box). The genes (or regions thereof) encoding the machinery responsible for assembling the common 2-(2,6-dioxopiperidin-4-yl)acetyl thioester biosynthetic intermediate are highlighted in blue. Figure was generated using Clinker (identity threshold: 0.3). (B) Comparison of the 9-methylstreptimidone, cycloheximide, lactimidomycin and migrastatin PKS domain architectures with the PKS encoded by the cryptic *B. gladioli* BGC. Shadowed domains are involved in constructing the common 2-(2,6-dioxopiperidin-4-yl)acetyl thioester intermediate. Domains responsible for chain release (TE and PBS) are highlighted in turquoise. Abbreviations for protein domains are displayed in the dashed box..... 40

Figure 2.1.4 Glutarimide-containing compounds target E-site of the eukaryotic ribosome. (A) Crystal structure of 80S ribosomes from *Saccharomyces cerevisiae* where A and P binding sites are occupied with aminoacyl-tRNAs.¹⁶⁰ (B) Close-up of lactimidomycin bound to the tRNA CCA-end in the E-site of the large subunit. Glutarimide moiety of lactimidomycin is highlighted in yellow. 42

Figure 2.2.1 Glutarimide-containing compounds target E-site of the eukaryotic ribosome. (A) Crystal structure of 80S ribosomes from *Saccharomyces cerevisiae* where A and P binding sites are occupied with aminoacyl-tRNAs.¹⁶⁰ (B) Close-up of lactimidomycin bound to the tRNA CCA-end in the E-site of the large subunit. Glutarimide moiety of lactimidomycin is highlighted in yellow. 43

Figure 2.2.2 Identification of the true metabolic product of the cryptic trans-AT PKS in *B. gladioli* BCC1622. (A) High resolution mass spectrometry analysis of gladiostatin. Measured spectra for the $[M+H]^+$ species, $m/z = 506.2760$, and $[M+Na]^+$, $m/z = 528.2573$, species. Calculated molecular formulae are displayed and a simulated spectrum is displayed for comparison. (B) Extracted ion chromatograms at $m/z = 506.27 \pm 0.02$, (corresponding to the $[M+H]^+$ ion of gladiostatin) from UHPLC-ESI-Q-TOF-MS analysis of ethyl acetate extracts from *B. gladioli* BCC1622 grown on a minimal medium containing different carbon sources. (C) Time-course analysis of gladiostatin degradation. Extracted ion chromatograms from UHPLC-ESI-Q-TOF-MS analyses, monitoring gladiostatin (EIC: 506.27 ± 0.02 , blue circles) and keto-degradation product (EIC: 408.27 ± 0.02 , red circles). Time points for each chromatogram are denoted on the right. 45

Figure 2.2.3 Structure elucidation of gladiostatin (49). Key COSY (A) and HMBC (B) correlations established the planar structure of the compound. Detailed NMR assignments can be found in Appendix.

..... 47

Figure 2.2.4 Inhibition of A2780 ovarian tumour cell migration by 49. (A) The bar chart compares wound-widths after 0 and 24 h in the absence of gladiostatin with the wound width after 24 h in the presence of 240 nM gladiostatin. (B) Microscope images showing wound-width after 24 hours of wound induction for untreated negative control cells and cells treated with 240 nM gladiostatin. 51

Figure 2.2.5 Generation of *B. gladioli* BCC1622 Ω *gdsE*_Bdomain using pGP Ω TP^R. (A) Overview of construction of the Ω *gdsE*_Bdomain mutant. 700 bp DNA containing the target gene fragment is cloned into pGP Ω TP^R vector and then integrated into the *B. gladioli* BCC1622 genome through tri-parental mating. Oligoprimers used for colony PCR and sequencing are shown in black arrows. The length of corresponding predicted PCR product is shown (796 bp). (B) Agarose gel electrophoresis of the PCR product provide genetic evidence for the successful insertional mutagenesis. The lanes correspond to the PCR product of pGP Ω *gdsE* construct (796 bp product) and *B. gladioli* BCC1622 Ω *gdsE* mutant (796 bp product). Wild-type allele *B. gladioli* BCC162 was used as a negative control. 53

Figure 2.2.6 Insertional inactivation of *gdsE* abolishes gladiostatin (49) production in *B. gladioli* BCC1622. (A) Base peak chromatograms from LC-MS analyses of *B. gladioli* BCC1622 wild type and *gdsE* mutant extracts. (B) Extracted ion chromatograms at m/z 506.27 \pm 0.02 (corresponding to [M+H]⁺ for 49 from UHPLC-ESI-Q-TOF-MS analysis of ethyl acetate extracts of *B. gladioli* BCC1622 (top) and the *gdsE* mutant (bottom). 54

Figure 4.1.1 Proposed signalling cascade with MmfR involved in methylenomycin A biosynthesis. 90

Figure 4.1.2 Proposed pathway for the biosynthesis of GBLs (including A-factor, SCBs and VBs), 2-hydroxyalkyl-4-hydroxy-3-methyl butenolides (SRBs and SABs) and AHFCAs (represented by MMFs), involving intermediates with a common phosphorylated butenolide core structure (highlighted in orange). 91

Figure 4.3.1 Comparison of biosynthetic gene clusters responsible for the production of and 2-hydroxyalkyl-4-hydroxy-3-methyl butenolides moieties in *Streptomyces* sp. with that encoded by the cryptic *B. gladioli* gene cluster. 97

Figure 4.3.2 Catalytic mechanism of the histidine phosphatase superfamily. The residues of the catalytic core are shown as numbered in *E. coli* SixA²²², a well-characterised protein of the family. His8 is phosphorylated during the course of the reaction. The rest of the residues interact electrostatically with the phosphate group before, during and after its transfer. Additional neutral or positive residues (represented as PP) contribute to the phosphate pocket by hydrogen-bonding to the phosphor group. Asp or Glu proton donors are denoted as PD.²²⁰ 98

Figure 4.3.3 Construction and verification of <i>B. gladioli</i> BCC1622ΔgdsG mutants. (A) Schematic diagram depicting the in frame gene deletion procedure of gdsG in <i>B. gladioli</i> . The primers used for colony PCR are indicated in the scheme (Table 7.3). (B) Agarose gel electrophoresis of the PCR product provide genetic evidence for the successful in-frame gene deletion. The lanes correspond to the PCR product of the <i>B. gladioli</i> BCC1622 wild type strain (1496 bp product), pGPI vector containing the flanking regions of gdsG (1008 bp product), first crossover allele <i>B. gladioli</i> BCC1622ΩgdsG mutant and the <i>B. gladioli</i> BCC1622ΔgdsG mutant (1008 bp product).....	99
Figure 4.3.4 LC-MS analysis of metabolite extracts derived from <i>B. gladioli</i> BCC1622ΔgdsG and <i>B. gladioli</i> BCC1622 wild type. (A) Base peak chromatogram of ΔgdsG mutant strain (shown in red) in comparison to wild type metabolic profile (shown in black). (B) Extracted ion chromatogram of $m/z = 506.27 \pm 0.02$, corresponding to the [M+H] ⁺ species observed for gladiostatin D (red) gladiostatin (black), showing a reduced retention time for the former.	100
Figure 4.3.5 DNA electrophoresis gel of the PCR products derived from complementation studies. The colony PCR confirmed successful complementation of the mutant (lane 3), resulting in a 714 bp product, similar to positive controls used (pMLBAD containing the gdsG and wild type strain).	101
Figure 4.3.6 LC-MS analysis for gladiostatin production after complementing <i>gdsG</i> back to the mutant. (A) Base peak chromatograms of <i>B. gladioli</i> BCC1622ΔgdsG (red), <i>B. gladioli</i> BCC1622ΔgdsG::gdsG (blue) and <i>B. gladioli</i> BCC1622 wild type (black). (B) Extracted ion chromatograms of $m/z = 506.27 \pm 0.02$, corresponding to [M+H] ⁺ for gladiostatin, indicated the restoration of gladiostatin production after complementation.	101
Figure 5.1.1 Phylogenetic analysis of KS domains from cis-AT and trans-AT PKS systems. Reproduced from ref.123.....	126
Figure 5.1.2 Evolution-inspired PKS engineering. (A) Module and co-evolved unit boundaries. Boundaries for both models have been successfully used for PKS engineering. (B) Approaches include the use of splice points of co-evolved units or inter- or intra-PKS recombination.....	128
Figure 5.2.1 Unrooted core-gene phylogeny of 229 <i>Burkholderia gladioli</i> genomes highlighting the distribution of gds-like BGCs. gds-like BGCs were restricted to a single clade within the phylogeny and were found in 112 of the 113 genomes in this clade. The maximum-likelihood phylogenetic tree was constructed with 100 bootstraps using a general time reversible model and a gamma model of rate heterogeneity. Bootstrap support is indicated for major clade splits (node circles). The distribution of several other BGCs of known function in <i>B. gladioli</i> genomes are also shown. Scale bar represents number of base substitutions per site. The figure was generated by Dr. Alex Mullins.....	130
Figure 5.2.2 Read depth graph indicating the 4440 bp deletion region compared to the reference genome of <i>B. gladioli</i> BCC1622.	131

Figure 5.2.3 Production of gladiostatin and the derivative, gladiostatin C (78), by genome sequenced *B. gladioli* strains. (A) Extracted ion chromatograms at $m/z = 478.27 \pm 0.02$ (corresponding to the $[M+H]^+$ for gladiostatin C) and $m/z = 506.27 \pm 0.02$ (corresponding to the $[M+H]^+$ for gladiostatin) from LC-MS analyses of extracts of *B. gladioli* strains where the gds cluster was present (in modified and unmodified form) or absent based on read mapping against the *B. gladioli* BCC1622 reference strain containing the gds BGC. (B) High-resolution mass spectrometry analysis of 78. Comparison of measured spectra with simulated spectra for $C_{25}H_{36}NO_8^+$ and $C_{25}H_{35}NO_8Na^+$, corresponding to the $[M+H]^+$ and $[M+Na]^+$ ions, respectively. (C) Key COSY and HMBC correlations observed for 78 (d-chloroform, 1H 700 MHz, ^{13}C 175 MHz). 133

Figure 5.2.4 Organisation of PKSs involved in the biosynthesis of gladiostatin C (78) based on comparison to 49 biosynthetic pathway. (A) Module organisation of the PKS responsible for the gladiostatin (top) and for gladiostatin C (bottom) assembly. Region missing from the GdsF' subunit in 78 assembly is highlighted in blue. (B) Partial result of sequence alignment between the GdsF/GdsF' PKS found in *B. gladioli* BCC1622 and *B. gladioli* BCC1726, respectively. The alignment was generated using Clustal Omega²⁴². 135

Figure 5.2.5 Homology model of chimeric KR_{M7} responsible for the biosynthesis of gladiostatin C. (A) Homology model of GdsF' module 7 KR domain showing cartoon and surface representations. The model is colour coded depending on the origin of the peptide sequence, grey for the N-terminal region of the GdsF' module 7 KR domain and blue for the C-terminal region of the GdsF module 8 KR domain. NADPH is represented as spheres. The point of sub-domain fusion is highlighted by a red box. (B) Structure of the GdsF' module 7 KR domain active site showing NADP⁺ and residues that bind it, indicating that both halves contain essential elements for the function of the domain. 136

Figure 5.2.6 Structure based sequence alignment. Structure-based multiple sequence alignment of GdsCKR₇, EryKR₁ (24% identical to GdsF' KR_{M7}, accession AAB84070.1), PksJKR₁ (36% identical to GdsF₂KR_{M7}, accession NP_389598.3)¹⁷⁴, SpnKR₃ (27% identical to GdsF' KR_{M7}, accession AAG23265)²⁴³, AmpKR₁ (24% identical to GdsF' KR_{M7}, accession AAK73513)²⁴⁴ and PlmKR₁ (29% identical to GdsF' KR_{M7}, accession AAQ84156)²⁴⁵. Elements are coloured grey and blue for the structural and catalytic halves of the KR domains, respectively. The splice point is indicated with a red arrow. Green boxes indicate the W motif for A-type KRs (PmKR₁ and AmpKR₁) and "LDD" motif for B-type KRs. Orange stars indicate the location of the active site residues. The blue box and the dashed pink line show the GxGxxG dinucleotide binding motif and the end of the structural domain and start of the catalytic domain, respectively. The sequence alignment was generated with ClustalW²⁴⁶ and ESPript²⁴⁷ was used to generate secondary structure annotations. 137

Figure 5.2.7 Construction and verification of *B. gladioli* BCC1622 Δ *gdsF*[S₇₂₉₆-D₈₇₇₄] mutants. (A) Schematic diagram depicting the in-frame gene deletion procedure for region of *gdsF* targeted in *B. gladioli* 1622. (B) Agarose gel electrophoresis of the PCR product provided evidence for successful in-frame gene deletion. The lanes correspond to the PCR product of the *B. gladioli* BCC1622 wild type strain (5204 bp product), single crossover *B. gladioli* BCC1622 Ω *gdsF*[S₇₂₉₆-D₈₇₇₄] mutant and the *B. gladioli* BCC1622 Δ *gdsF*[S₇₂₉₆-D₈₇₇₄] double crossover mutant (770 bp product). 138

Figure 5.2.8 Engineering gladiostatin C production in the gladiostatin producing strain *B. gladioli* BCC1622. Extracted ion chromatograms at $m/z = 478.27 \pm 0.02$ (corresponding to the [M+H]⁺ for gladiostatin C) from LC-MS analyses of the *B. gladioli* BCC1622 Δ [S₇₂₉₅-D₈₇₇₄] mutant, wild type *B. gladioli* BCC1622 strain and *B. gladioli* BCC1726. The chimeric KR domain created by the deletion is highlighted with an asterisk. 139

List of Schemes

Scheme 1.5.1 Studies of historical significance to polyketide research. (A) Collie's proposal for orsinol formation through the triketide intermediate. (B) Birch's ^{14}C -labelled acetate studies demonstrating the polyketide basis of 6-MSA (16).	13
Scheme 1.5.2 Assembly line Assembly line of 6-deoxyerythronolide (19) catalysed by DEBS. Post-PKS tailoring steps of the released polyketide produce the antibiotic erythromycin A (12).	15
Scheme 1.5.3 Catalytic cycle of AT domain. The conserved Ser of the AT attacks the extender unit to generate an AT-acyl intermediate followed by nucleophilic attack by the thiol moiety of the PPant arm of the intramodular holo-ACP, furnishing the ACP-acyl complex.	18
Scheme 1.5.4 KS-catalysed chain elongation in modular PKSs. The catalysis involves a trans-thiolation and a decarboxylative condensation that results in chain elongation by two carbon units.	19
Scheme 1.5.5 DH-catalysed dehydration of (3R)/(3S)-hydroxy thioester intermediate. Syn elimination mechanism with His-Asp to achieve stereocontrol.	22
Scheme 1.5.6 Cis-acting ER-catalysed reaction mechanisms. (A) Mechanism of the reduction of the α,β -double bond catalysed by the ER to afford stereodefined saturated acyl thioesters. (B) The ene mechanism proceeds through a C2-intermediate which is converted to its corresponding enolate intermediate and subsequently protonated.	24
Scheme 1.5.7 General mechanism for PKS released by type I TE domains. Depending on the TE, the acyl chain of acyl-TE can be hydrolysed (shown in red) or macrocyclised (shown in blue) to yield a macrolactone or macrolactam if an amino group is present.	26
Scheme 1.5.8 Examples of chain release mechanisms employed by modular type I PKSs to enhance structural diversity. (A) Type II chain release in polyether biosynthesis. (B) In gregatin (21) biosynthesis, the type II TE, GrgF, catalyses Claisen condensation to join to polyketide chains and hydrolysis to yield the linear dimer. (C) Mechanism of reductase domains involves a two electron reduction to release the corresponding polyketide as aldehyde or four electron reduction to release a primary alcohol. (D) Chain release catalysed by OAS proceeds through decarboxylative condensation of an amino acid and the ACP-bound polyketide thioester. (E) Bamb_5915 catalyses the chain release via transesterification with DHCCA. (F) Ketosynthase-like tetronate synthases (TRs) catalyse offloading through condensation of the polyketide chain with an ACP-bound glyceryl thioester to furnish a 2-acyl-4-hydroxymethyl-tetronic acid.....	27
Scheme 2.1.1 Proposed model for the formation of the glutarimide head. The mechanism proceeds through a decarboxylative Michael addition of the ACP-bound malonate unit to the KS-bound α,β -	

unsaturated thioester. The branched polyketide intermediate is then covalently linked to the ACP and the KS domain. Subsequent nucleophilic attack of the carboxamide into the KS-bound thioester yields the glutarimide moiety in the polyketide chain which is then passed to the downstream assembly line.

.....	41
Scheme 2.2.1 Proposed mechanism of gladiostatin conversion to the degradation product.	48
Scheme 2.2.2 Proposed pathway for 49 biosynthesis. Domain and module organization of the gds trans-AT PKS showing the proposed structure of each ACP-bound thioester intermediate. KS domains have been numbered sequentially, and the TransATor predictions for their acyl-ACP substrates are shown in the dashed box. The A / B subscripts denote the predicted stereospecificity of KR domains. The trans-acting ER domain (gold) of GdsB is predicted to reduce enoyl thioesters attached to the ACP domains in modules 5, 6, 7 and 8. Domain abbreviations are shown in the dashed box. The putative functions of the proteins encoded by the gds biosynthetic gene cluster are listed in appendix Table A1.....	55
Scheme 3.2.1 Overview of the MS-based assay used to probe the putative function GdsB and GdsH as trans-acting enoyl reductases. (A) ACP domains are first converted to their holo forms by a phosphopantetheinylation reaction with 2-butenoyl-CoA that attaches a 2-butenoyl-phosphopantetheine unit onto the conserved Ser residue of the domain. (B) Proposed mechanism of phosphopantetheine elimination reaction. During collisional activation, the phosphate anion remains attached to the Ser residue of the ACP and the amide carbonyl displaces the pantetheinylate moiety forming a 5-membered ring with a protonated imine.	69
Scheme 4.2.1 In vitro reconstitution of chain release by the PBS domain. Condensation of DHAP and the NAC thioester of 3-ketooctanoate by the purified recombinant GdsF ACP-PBS di-domain and dephosphorylation of the resulting phosphorylated butenolide 57 using shrimp alkaline phosphatase to yield the butenolide product.	94
Scheme 4.2.2 Overview of the scope of the study described in this chapter.....	95
Scheme 4.3.1 HAD-like phosphatase structure and catalysis. The general catalytic mechanism of HAD phosphatases. Catalysis proceeds through an aspartyl-phosphate intermediate.....	98
Scheme 4.3.2 Overview of the enzymatic assay used to investigate the formation of the 2-acyl-4-hydroxy-3-methyl butenolide in vitro.....	109
Scheme 4.3.3 Overview of the route used to synthesise 71 authentic standard. Reagents and conditions: (i) ^{3,4} DMBOH (1.1 eq.), amberlyst-15 (10% w/w), CH ₂ Cl ₂ , reflux, 48 h; (ii) i-Pr ₂ NH (1.2 eq.), n-BuLi (1.2 eq.) in THF 0 °C to -78 °C, 73 then hexanal (1.2 eq.) 5 h; (iii) Dess-Martin periodinane (1.1 eq.), CH ₂ Cl ₂ , rt, 2 h; (iv) PIFA (2.0 eq., CH ₂ Cl ₂ , rt, 3 h.....	110
Scheme 4.4.1 Proposed pathway for AMHB unit biosynthesis demonstrated in this study. (A) Condensation of DHAP with the fully-assembled polyketide intermediate is catalyzed by AfsA-like	

enzyme PBS (previous work). After this condensation step, nonenzymatic intramolecular aldol condensation yields the 3-hydroxymethyl intermediate. GdsG-catalysed phosphate rearrangement in the butenolide unit and subsequent dephosphorylation by GdsA generates the tri-substituted butenolide.

(B) Proposed mechanism for the 1,3 phosphate rearrangement catalysed by GdsG. 123

List of Tables

Table 2.2.1 Antimicrobial activity of gladiostatin.....	49
Table 2.2.2 Anticancer activity of 49.	50
Table 2.2.3 Predicted configurations of stereogenic carbon centres bearing hydroxyl groups in 49. Sequence alignments and comparisons with KR domains from trans-AT PKSs with known stereospecificities were performed to determine the type of KR domain.....	56
Table 2.2.4 Predicted specificities of KS domains from the gds PKS. All predictions were conducted using the TransATor webserver. ¹⁷⁵	58
Table 4.3.1 ¹ H and ¹³ C chemical shifts establishing the 2-acyl-4-hydroxy-3-methyl butenolide core in gladiostatin D. Key HMBC correlations are shown with red arrows. Complete chemical shift assignment can be found in Appendix.	103
Table 7.2.1 List of antibiotics used for these studies.	152
Table 7.2.2 List of microbial strains used within this project.	154
Table 7.2.3 List of vectors used within this project.	156
Table 7.3.1 PCR components per 25 µL total volume.	157
Table 7.3.2 Conditions used in the thermocycler.	158
Table 7.6.1 Overview of the primers and mutations for mutant constructs.	161
Table 7.7.1 Recombinant constructs generated using pET28a(+) vector. Size of recombinant protein, insert size and restriction sites are shown.	162
Table 7.9.1 Overview of constructs used in in-frame gene deletion within this project.....	169
Table 7.9.2 Overview of constructs used in in-frame gene deletion within this project.....	171
Table 7.10.1 ¹ H and ¹³ C NMR data for gladiostatin (49) in CDCl ₃	173
Table 7.10.2 ¹ H and ¹³ C NMR data for gladiostatin B (55) in CDCl ₃	174
Table 7.10.3 ¹ H and ¹³ C NMR data for gladiostatin D (59) in acetonitrile-d ₃	175
Table 7.10.4 ¹ H and ¹³ C NMR data for gladiostatin C (78) in CDCl ₃	176

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Prof Greg Challis for giving me this amazing opportunity to be part of his team. He has always supported me with his scientific knowledge and his supervision. Most importantly, I would like to thank him for his belief in me and for being a great mentor and genuinely caring for me, my progress and my decisions. I would also like to thank Dr Matthew Jenner for being a great mentor and always helping me to evolve not only as a researcher but as a person as well. His patience, motivation and positive thinking helped me a lot during my PhD. I am grateful for conducting research under their supervision.

I would like to thank Dr Lona Alkhalaf for her guidance and advice throughout my years in the lab. A special thanks to Dr Xinyun Jian for her insightful comments and suggestions. A special thanks to Dr Yousef Dashti for helping me settling in the lab and for the useful discussions. I would also like to thank past and present members of the Challis and Jenner groups since I was fortunate enough to make great friends and create beautiful memories which I will never forget.

I would like to extend my sincere thanks to Prof. Eshwar Mahenthiralingam and Dr Alex Mullins for the valuable collaboration and the interesting discussions, and Dr Isolda Romero-Canelón for conducting the cytotoxicity experiments for gladiostatin. Thanks are also owned to BBSRC MIBTP for funding my PhD.

I would like to thank Prof Dominic Campopiano and Dr Christophe Corre for an enjoyable discussion we had about my research during my viva. Their comments and suggestions were really helpful. I hope there will be more opportunities to discuss about research in NPs in the future.

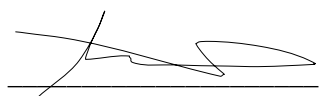
I am grateful to my boyfriend George for his love, immense support and for always keeping me smiling throughout my studies. These years were so special to me because of him. I would also like to thank my family for their encouragement and support during my PhD years. A big thanks goes to Stavroula who had to put up with my stresses and worries for the past four years of study and for spending so many hours on the phone hearing me talking about

enzymes. To Andreanna, Aggeliki, Dimitra and Eva for making these years much more memorable with spontaneous trips and virtual parties during COVID.

Most importantly, I would like to thank my dad who has always been my role model and for always inspiring me to be a better version of myself.

Declaration

The experimental work reported in this thesis is original research conducted by the author in the Department of Chemistry, University of Warwick, between September 2018 and September 2021. No material has been submitted in any previous application for any degree. Contributions from collaborators are references throughout the text.

A handwritten signature in black ink, consisting of a series of loops and a long horizontal stroke at the end, positioned above a solid horizontal line.

Date: 02/10/21

Ioanna Theodora Nakou

List of abbreviations

ACP	Acyl carrier protein (domain)
AH	Acyl hydrolase (domain)
AHFCAs	2-alkyl-4-hydroxymethylfuran-3-carboxylic acids
AT	Acyltransferase (domain)
BGC	Biosynthetic gene cluster
bp	Base pairs
CoA	Coenzyme A
Cys	Cysteine
DHAP	Dihydroxyacetone phosphate
DEBS	6-deoxyerythronolide B synthase
DH	Dehydratase (domain)
dH ₂ O	Deionized water
DNA	Deoxyribonucleic acid
EIC	Extracted ion chromatogram
ER	Enoylreductase (domain)
ESI	Electrospray ionization
FAS	Fatty acid synthase
HPLC	High performance liquid chromatography
GBLs	γ-butyrolactones
IPTG	Isopropyl-β-D-thiogalactoside
Kan	Kanamycin
KR	Ketoreductase (domain)
KS	Ketosynthase (domain)
KS ⁰	Non-elongating ketosynthase (domain)
Lys	Lysine
Met	Methionine
MIC	Minimum inhibitory concentration
MMFs	Methylenomycin furans
MS	Mass spectrometry

NADPH	b-nicotinamide adenine dinucleotide 2'-phosphate
NMR	Nuclear magnetic resonance
NRPS	Non-ribosomal peptide synthase
OD ₆₀₀	Optical density at 600 nm
PAGE	Polyacrylamide gel electrophoresis
PPant	Phosphopantetheinyl
PBS	Phosphorylated butenolide synthase
PCR	Polymerase chain reaction
PDB	Protein data bank
Phe	Phenylalanine
PKS	Polyketide synthase
Pmx	Polymyxin
Ppm	Parts per million
RNA	Ribonucleic acid
SAP	Shrimp alkaline phosphatase
Ser	Serine
SDS	SDS Sodium dodecyl sulphate
Tris	Tris(hydroxymethyl)aminomethane
Tyr	Tyrosine
VB	Virginiae butanolide
WHO	World health organisation

Abstract

Bacteria belonging to the *Burkholderia* genus thrive in various ecological niches, ranging from the plant rhizosphere to the human lung, and play critical roles in ecological interactions often through the secretion of bioactive secondary metabolites. Genome mining and gene knockout experiments identified a *trans*-AT PKS in the *B. gladioli* BCC0238 and BCC1622 which is responsible for the biosynthesis of gladiostatin, a novel natural product that contains a glutarimide pharmacophore and a butenolide motif, moieties mainly appended in secondary metabolites of *Streptomyces* sp.. The biological activity of gladiostatin was investigated, showing a promising activity against several human cancer cell lines and inhibition towards tumor cell migration. The peculiarities found in PKS responsible for gladiostatin assembly were investigated through functional characterisation of key enzymes through *in vitro* biochemical assays and genetic engineering, indicating an unprecedented number of enoyl reduction events occurring in *trans* as well as a novel catalytic mechanism involved in the installation of the butenolide head.

The thesis describes a novel mechanism for megasynthase evolution which involves a recombination event in the middle of a catalytic domain. Using evolution's point of recombination, a fully functional PKS was engineered in the lab as an optimally functioning PKS. The study provides new insights on evolutionary pathways of *trans*-AT PKS systems as well as a direction for efficient future diversity oriented biosynthesis of numerous polyketide compounds.

Chapter 1

Introduction

1.1 Natural products

Natural products represent a large family of diverse chemical entities with a wide variety of biological activities that find application in human and veterinary medicine as well as in agriculture.¹ They derive from bacterial, fungal and plant sources. The bacterial and fungal natural products represent molecules that are not directly involved in primary metabolic processes of the organisms, however they provide some advantage to the producer in its native environment.^{2,3} This class of natural products is called secondary metabolites. Due to their structural complexity and diversity, natural products serve as a continuing source of architectural and synthetic inspiration to the scientific community.

The discovery of penicillin from *Penicillium notatum* in 1928 by Alexander Fleming, with notable contributions from Florey, Chain and Dorothy Hodgkin, marked a significant shift from plants to microorganisms as a source of natural products.⁴ The period between the 1940s and 1960s is characterised as the golden era of antibiotic discovery and it is accredited to Selman Waksman who developed a screening platform for the discovery of antimicrobial natural products.⁵ The “Waksman platform” was used to screen soil-derived actinomycetes for antimicrobial activity against a susceptible test organism by detecting zones of growth inhibition. This led to the discovery of streptomycin (**1**) which was effectively used against *Mycobacterium tuberculosis*.^{6,7} This was followed by discovery of other well-known antibiotics including chloramphenicol⁸ (**2**) from *Streptomyces venezuelae*, cephalosporin C⁹ (**3**) from *Cephalosporium acremonium* and vancomycin¹⁰ from *Amiclatopsis orientalis* (**4**) (**Figure 1.1.1**).

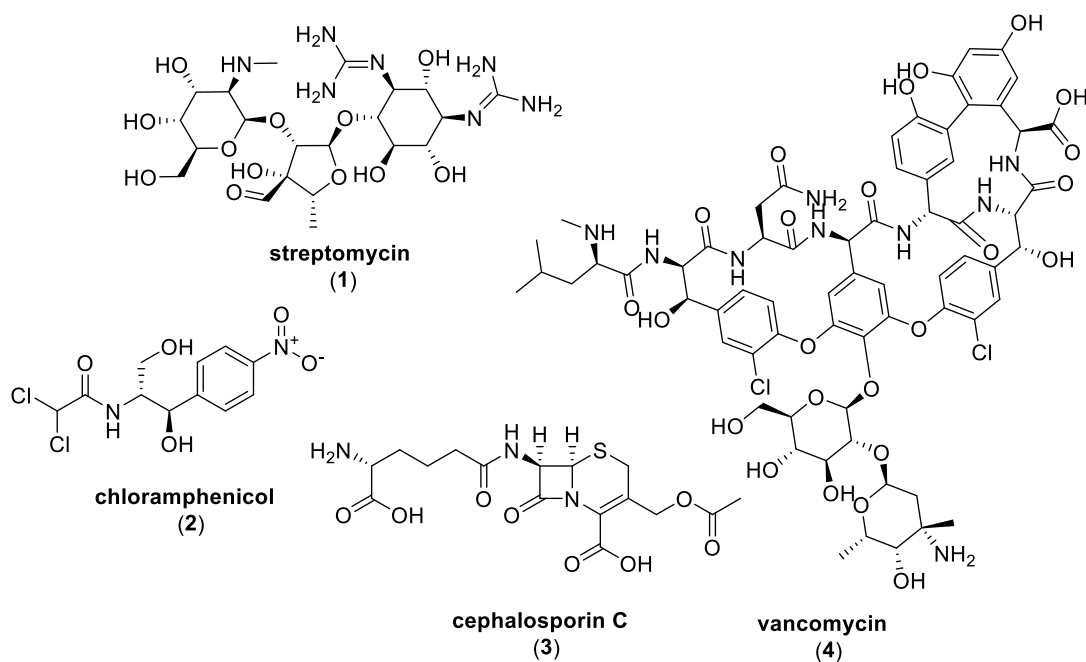


Figure 1.1.1 Overview of structures of antibiotics discovered during the golden era of antibiotics.

Over the years, bioprospecting for antibiotics became less popular due to obstacles such as redetection of known compound and low compound yields of the isolated compounds.¹¹ This, in conjunction with the emergence of antibiotic resistance, diverted efforts towards modifying the existing antibiotic scaffolds and hindered efforts for the discovery of new antibiotics. However, recent advances in metagenomic/ metabolomic analyses and reducing costs of high-throughput sequencing, has allowed the discovery of novel bioactive metabolites from unexplored bacterial sources.¹² In addition, the focus of many studies is on understanding the architecture and dynamics of natural product biosynthetic machineries in order to exploit these systems for production of novel compounds with a variety of bioactivities.

1.2 Genome mining to exploit the hidden biosynthetic potential of biologically active compounds

The identification of novel natural products from bacterial sources has mainly relied on bioactivity-guided approaches, where once activity is identified, the metabolites are structurally characterised. However, traditional screening approaches have become unproductive due to an increasing rediscovery rate.

In the past decade, genomic sciences have undeniably changed the field of life sciences. Mapping of a genome enables scientists to address questions about physiology and metabolic traits of prokaryotic and eukaryotic organisms. Genomic data have also been utilised in the field of drug discovery for the identification of potential drug targets¹³ or for unravelling biosynthetic pathways from previously overlooked metabolites.^{14–16}

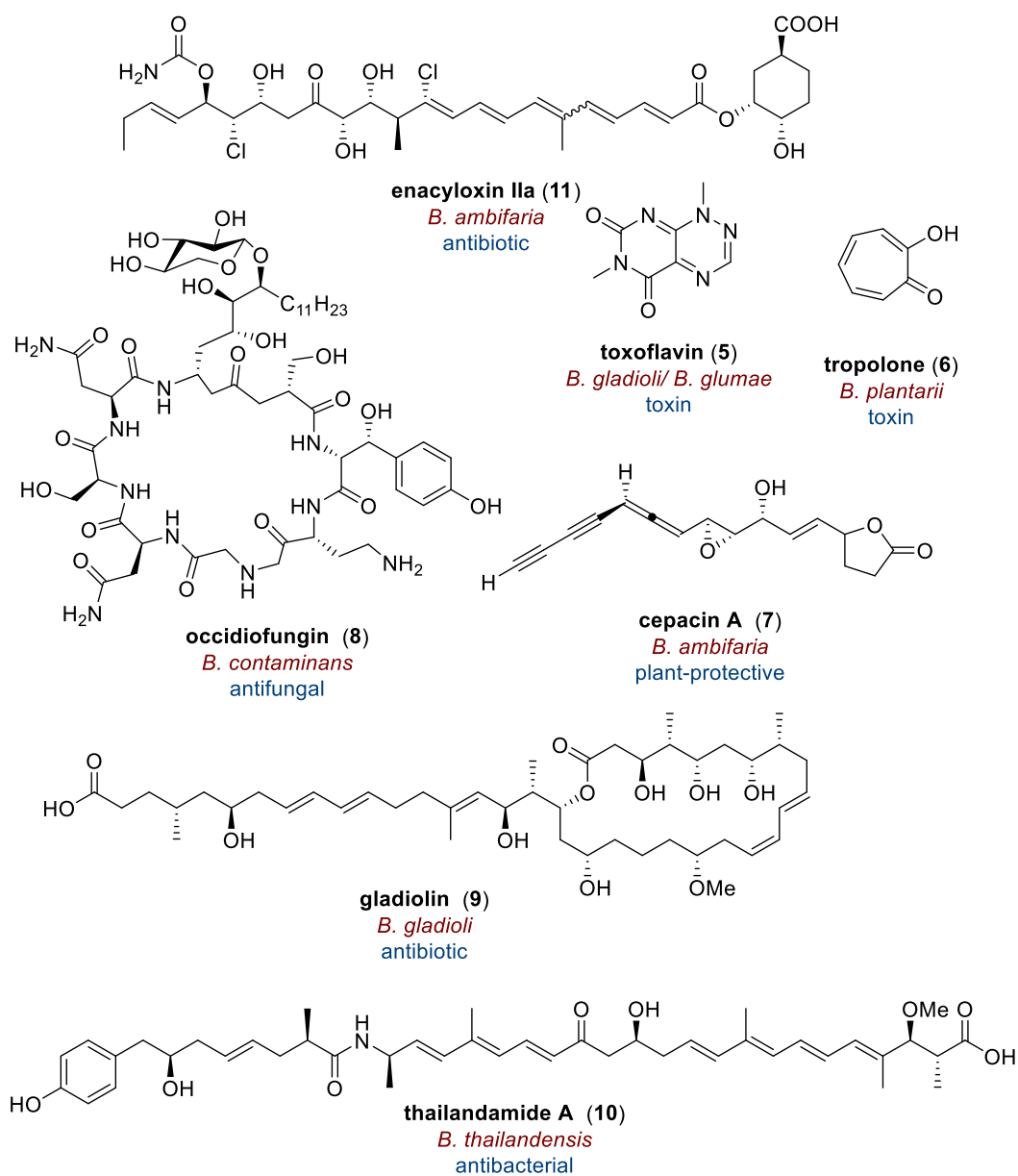
The term “genome mining” applies to every bioinformatics approach used to detect not only the biosynthetic pathway of bioactive natural products, but also their possible function and chemical interactions. This is conducted by searching genomes for conserved motifs. In cases where the gene cluster is silent under laboratory conditions, activation of pathways under the control of a constitutive promoter or inactivating repressors has been pursued.^{17–19} This can be typified by the discovery of stambomycins.²⁰ Genome mining of *Streptomyces ambofaciens*, revealed a cryptic biosynthetic cluster, however transcriptional analysis suggested that this is not expressed under laboratory growth conditions. This was circumvented by constitutive expression of a regulatory gene within the cluster which induced the expression of the biosynthetic genes and subsequently led to the identification of 51-membered glycosylated macrolides.

Molecular-based approaches have shown that only 1% of the total microbial population can be cultured under lab growth conditions.²¹ This presents an untapped source of putative natural products. Although Gram-positive Actinobacteria are the most explored source for the majority of clinically used compounds, Gram-negative bacteria such as β -proteobacteria are starting to gain ground as a valuable source of novel secondary metabolites with pharmacological properties.²²

1.3 *Burkholderia* as an untapped source of natural products

Burkholderia, a genus of Gram-negative β -proteobacteria, are a well-known source of versatile metabolites that allow their quick adaptation to a variety of ecological niches (**Figure 1.2**).^{23,24} The heterogeneity of this genus led to several taxonomical revisions.^{25–27} The first *Burkholderia* sp. was described in 1949 by Walter H. Burkholder as a phytopathogenic bacterium, which caused rot in onion bulbs. The species was named 'cepacia', after the Latin word for onion, which was later known as *Pseudomonas cepacia*.²⁸ Due to its broad and vague phenotypic definition, *Burkholderia* was included in the genus *Pseudomonas*, until the 1970s where rRNA-DNA hybridization analysis indicated considerable genetic diversity among members of this genus.^{29–31} Currently, *Burkholderia* encompasses 125 well-established members. The genomes of *Burkholderia* presents multi-chromosomal organisation, typically containing 2 to 3 chromosomes.³² The organisms are frequently isolated from terrestrial and aquatic ecosystems, and can occur both free-living or as symbionts with fungi, protozoans, plants, animals and humans.^{33–35} These interactions can either be beneficial or harmful. Such examples include members of the *Burkholderia cepacia* complex (Bcc), consisting of at least 22 closely related species, which act as opportunistic pathogens towards immunocompromised individuals, such as cystic fibrosis (CF) patients.^{35,36} *B. mallei* and *B. pseudomallei* are causative agents of glanders and melioidosis and were both used as bioweapons during World War I and II on livestock, prisoners and civilians.^{37,38} In addition, plant pathogens from these species such as *Burkholderia glumae* and *Burkholderia plantarii* produce several toxic compounds, such as toxoflavin (**5**) and tropolone (**6**), which affect economically important crops (**Figure 1.3.1**).^{39,40} In contrast to the pathogenic traits, various *Burkholderia* can be utilised in agriculture, due to their plant growth-promoting and biopesticidal characteristics. *Burkholderia* strains function as biocontrol agents by producing specialised secondary metabolites like cepacin A **7** and occidiofungin **8**.^{41,42} Recent interest has also been focused on this species with regard to their potential for producing antibiotics. Gladiolin (**9**), a polyketide produced from *B. gladioli* CF clinical isolate as well as thailandamide A (**10**) isolated from *Burkholderia thailandensis*, provided promising activity against *Mycobacterium tuberculosis* and *S. aureus*, respectively.^{43,44} Enacyloxin IIa (**11**), a metabolite produced from *B. ambifaria* was also found

have narrow-spectrum activity against *Acinetobacter baumannii*, a pathogen which is high in the global priority pathogen list for R&D of new antibiotics, published by World Health Organisation (WHO).^{45,46} These recent discoveries, in combination to the high number of predicted 'cryptic' biosynthetic gene clusters found in the genomes of *Burkholderia* bacteria, indicate that these organisms present an untapped source for drug discovery.⁴⁷ A large proportion of biocontrol and biomedically important natural products produced from *Burkholderia* species are biosynthesised multidomain megaenzymes called polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs).^{47,48}

Figure 1.3.1 Examples of natural products derive from *Burkholderia* sp.

1.4 Polyketide natural products

Polyketides (PKs) constitute a large family of natural products with diverse scaffolds that present impressive structural complexity, and possess great therapeutic utility. Representative examples of this class of natural products include: the macrolactone antibiotic erythromycin A (**12**), the immunosuppressive drug rapamycin (**13**), the antifungal nystatin (**14**) and the first statin to reach market, lovastatin (**15**) (Figure 1.4.1).

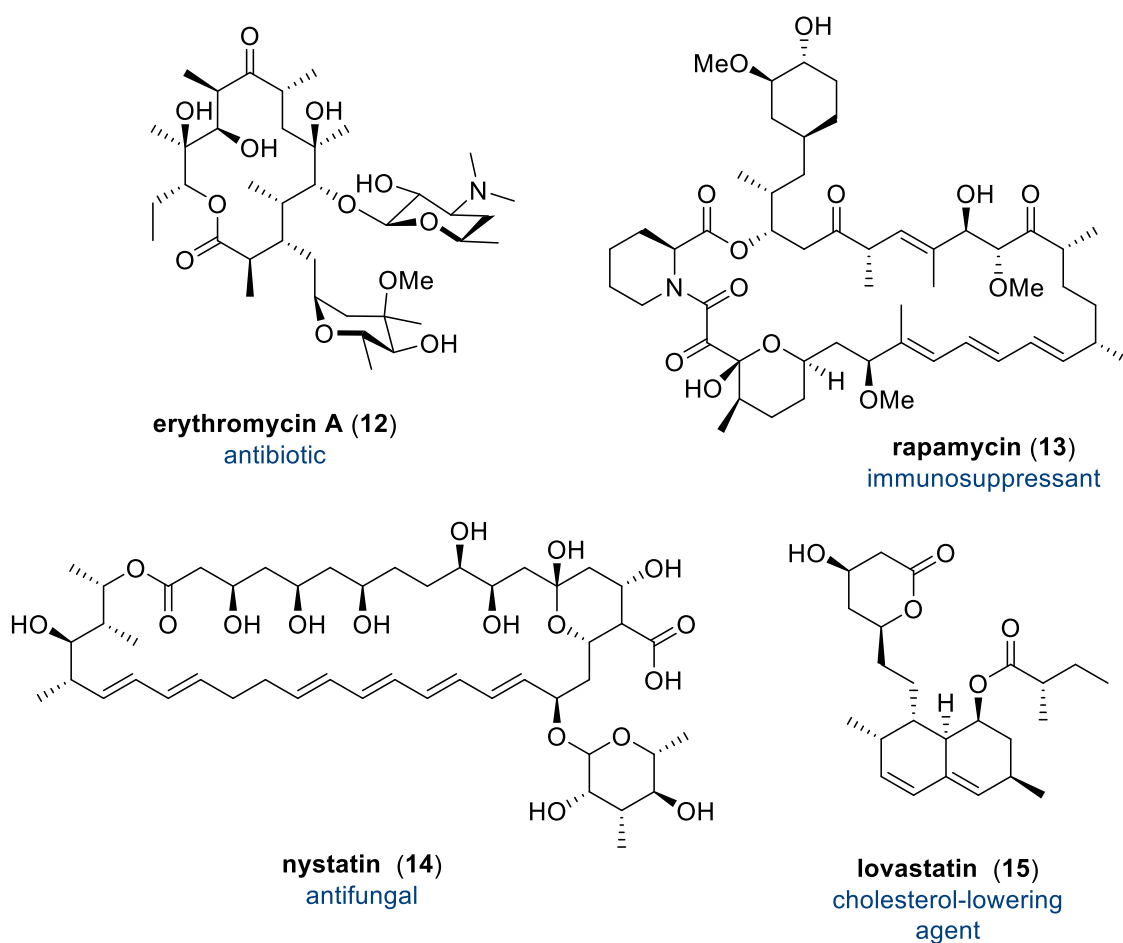


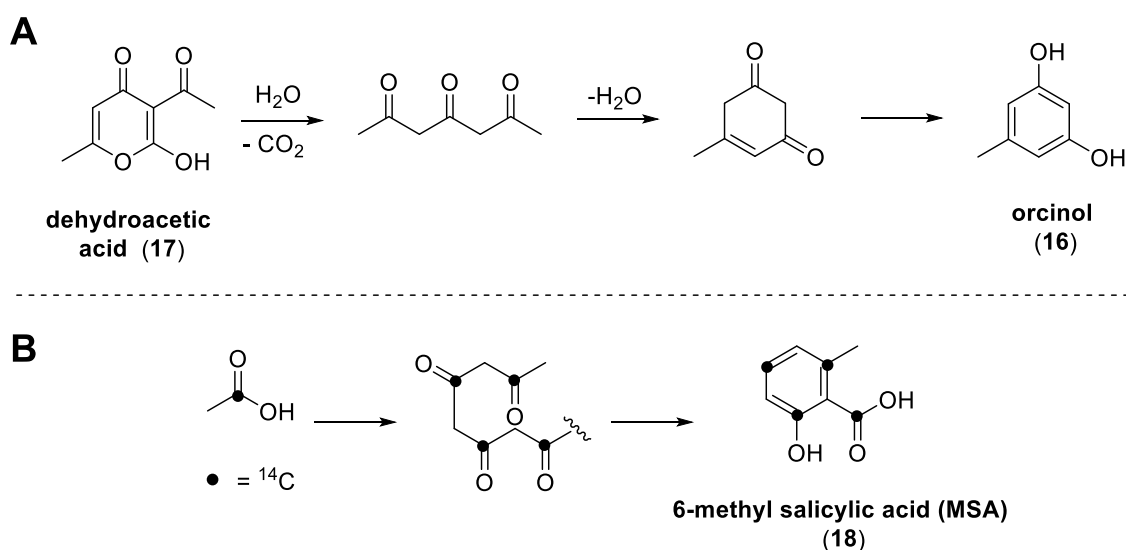
Figure 1.4.1 Examples of diverse polyketide scaffolds of therapeutic potential.

Although Nature has gifted us with an apparently limitless source of bioactive compounds, unravelling their therapeutic potential is difficult due to their low isolation yields. In addition, the remarkable array of functionality and stereochemistry adds to the structural challenges faced by synthetic chemists to replicate their scaffolds using synthetic chemistry. The therapeutic and economic value of these compounds have attracted vast interest in investigating the

biosynthetic mechanisms involved in their assembly that facilitate the production of novel, biochemically active compounds.

1.5 Polyketide synthases

The first notion that certain classes of aromatic natural products might derive from simple “CH₂-CO” building blocks through linear poly- β -keto intermediates that undergo cyclisation was serendipitously conceived by Collie in 1893, after a degradation reaction of dehydroacetic acid (17) (**Scheme 1.5.1**).⁴⁹ It was nearly half a century later that Birch administered ¹⁴C-labelled acetate to a culture of *Penicillium griseofulvum* which produces the polyketide 6-methyl salicylic acid (6-MSA) (17) (**Scheme 1.5.1**).⁵⁰ The pattern of radioactivity distribution allowed him to propose that polyketides originate from head to tail condensation of acetate-derived units. Six years later the groups of Leadlay and Katz revealed the gene organization of a modular polyketide synthase responsible for the production of erythromycin.^{51–54} These two discoveries became the cornerstone for the contemporary field of polyketide synthase research.



Scheme 1.5.1 Studies of historical significance to polyketide research. (A) Collie's proposal for orsinol formation through the triketide intermediate. (B) Birch's ¹⁴C-labelled acetate studies demonstrating the polyketide basis of 6-MSA (16).

Polyketide synthases (PKSs) are giant biosynthetic machineries that construct complex hydrocarbon skeletons from a series of acyl thioester building blocks.⁵⁵ PKSs are evolutionarily

related to fatty acid synthases (FAS), which play a pivotal role in primary metabolism across all organisms. In analogy to FAS type categorisation, PKSs are subdivided into three types based on their protein architecture and catalytic mechanism.^{55,56} Type I PKSs, similar to type I fatty acid synthase (FAS), comprises of gigantic multifunctional polypeptides, each consisting of a series of catalytic domains. Type II PKSs are large multienzyme complexes of several discrete, iteratively monofunctioning enzymes. Type III PKSs employ a single condensing enzyme functioning in an iterative manner.^{56,57}

Type I PKSs can be further subdivided depending on whether the multifunctional polypeptides are functioning in an iterative or modular manner. Iterative type I PKSs use a single set of active domains repetitively to biosynthesise polyketides, similar to FAS pathways. In contrast, modular type I PKSs consist of multiple modules where each of the catalytic domains is typically used once during chain assembly. Diversions from this classification are exemplified in stigmatellin and borrelidin biosynthesis which possess modular organisation but one module is assumed to be used iteratively during the biosynthetic process.^{58,59}

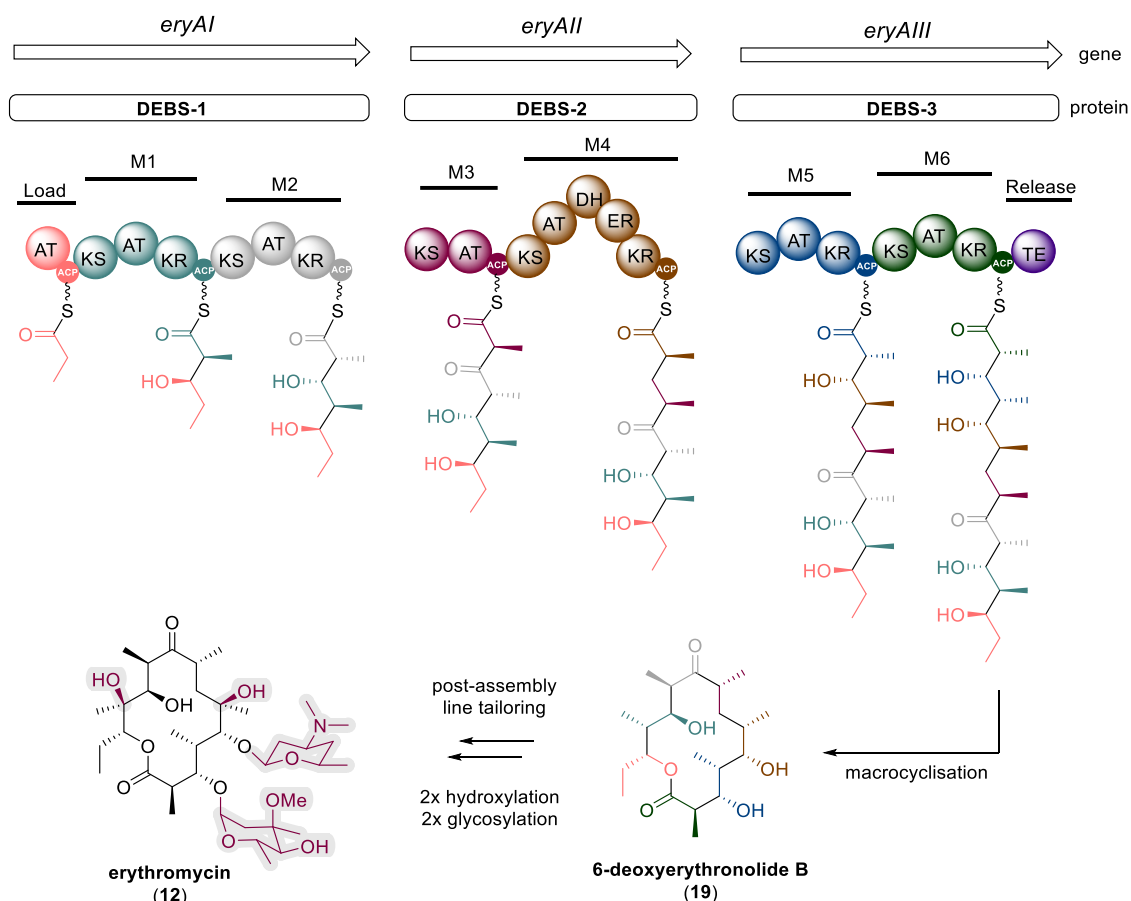
1.5.1 Modular type I polyketide synthases

The minimal module required for chain elongation incorporate three catalytic domains which extend the polyketide backbone by one ketide unit. These include: an acyl carrier protein (ACP) domain, an acyltransferase (AT) domain and a ketosynthase (KS) domain.

The archetypal modular type I PKS is 6-deoxyerythronolide B synthase (DEBS) from *Saccharopolyspora erythraea* (**Scheme 1.5.2**). Sequence analysis of the *S. erythraea* genome identified three large open frames (ORFs) encoding three giant polypeptides at ~ 350 kDa each. The regions of the polypeptides were assigned biosynthetic functions based on sequence comparison to FAS domains.^{53,60} The polypeptide units can be split into seven modules, each responsible for condensation of propionate unit from 2-(S)-methylmalonyl-CoA. The final assembled heptaketide acyl chain is offloaded *via* lactonisation to afford 6-deoxyerythronolide B (**19**). Post-PKS tailoring enzymes modify the released product to furnish erythromycin A (**12**).

Herein, the module is defined as the domains involved in the catalysis of one chain extension, starting from an elongating KS until the first downstream ACP domain.^{61,62} The same definition

of the module is referred to the domain architecture of mammalian FAS (KS-AT-DH-ER-KR-ACP-TE). The conventional module boundaries have been recently questioned, after a study indicating that during the evolution of PKS, the accessory domains appear to co-migrate with the downstream KS, which has implications for engineering efforts.^{63,64} However, for purposes of identifying the functional role of the module, this thesis uses the conventional boundaries to describe a module.



Scheme 1.5.2 Assembly line Assembly line of 6-deoxyerythronolide (19) catalysed by DEBS. Post-PKS tailoring steps of the released polyketide produce the antibiotic erythromycin A (12).

1.5.1.1 Core domains

Acyl carrier protein (ACP) domain

ACP domains are four-helical bundles which tether the growing polyketide chain *via* a thioester bond. These domains play a central role in biosynthesis by shuttling malonyl-based building blocks and polyketide intermediates between catalytic domains of a module. They are post-translationally converted from their inactivated *apo*-form to their *holo*-form by phosphopantetheinyl transferases.⁶⁵ This modification occurs after addition of the 4'-phosphopantetheine moiety of coenzyme A (CoA) onto a conserved Ser residue located in the (D/E)xGxDSL motif of the ACP domain, appended in the N-terminal side of helix II of the domain (**Figure 1.5.1**).⁶⁶ This generates a phosphopantetheinyl arm onto which intermediates can be covalently attached to the PKS through a thioester linkage, during biosynthesis. This also creates a flexible arm required to reach active sites of accessory domains within a given module.^{62,67} The negatively charged helix II also acts as the 'recognition helix' and based on mutagenesis studies is believed to mediate docking to positively charged regions on AT and KS domains.⁶⁸

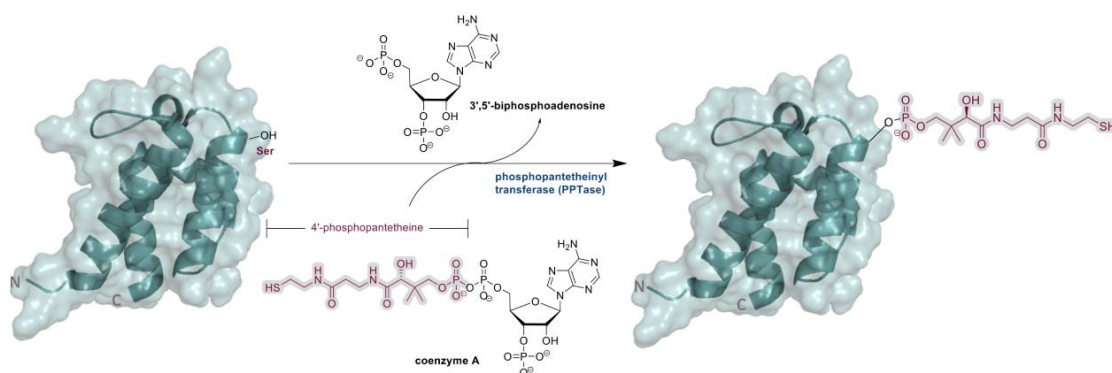


Figure 1.5.1 Post-translational modification of the ACP domain. Attachment of phosphopantetheine from co-enzyme A to a conserved Ser residue onto the *apo*-ACP domain catalysed by phosphopantetheinyl transferase.

Type I PKS ACP domains have been shown to contain an additional helix, termed “helix 0”, that is thought to play a role in the docking to the downstream KS module, facilitating a forward transfer of polyketide intermediates within the assembly line.^{69,70} Crystal structures of ACPs

from type II FAS indicate that the acyl chain of fatty acid is sequestered into a narrow internal tunnel between the helices of ACP domains leading to minor alterations in ACP structure due to the perturbation (**Figure 1.5.2**).⁷¹ Unlike type II ACPs, type I FAS and PKS ACP domains do not appear to bury their substrates as shown in NMR studies of acyl-ACP domains derive from the curacin cluster.⁷² The ability to form a substrate cavity is seen in molecular dynamics simulations as one mode by which ACPs can be distinguished from one another, although the functional importance of this difference is unclear.⁷³

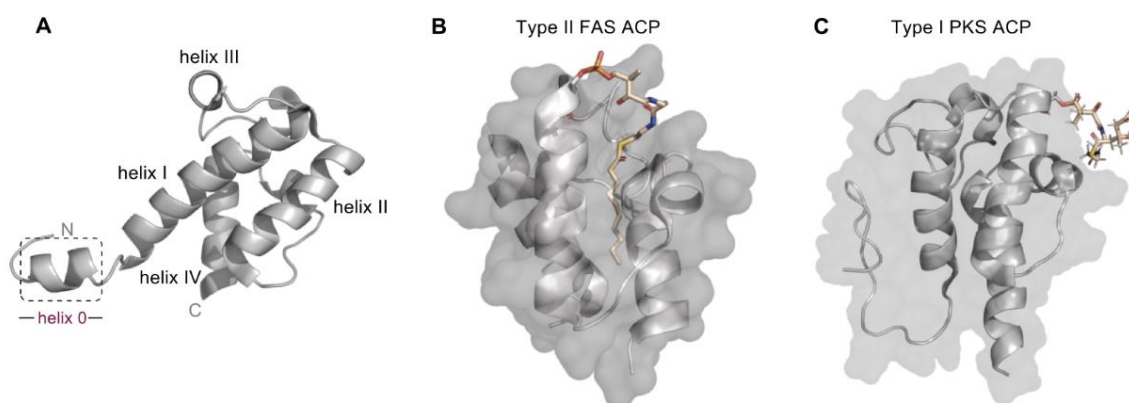


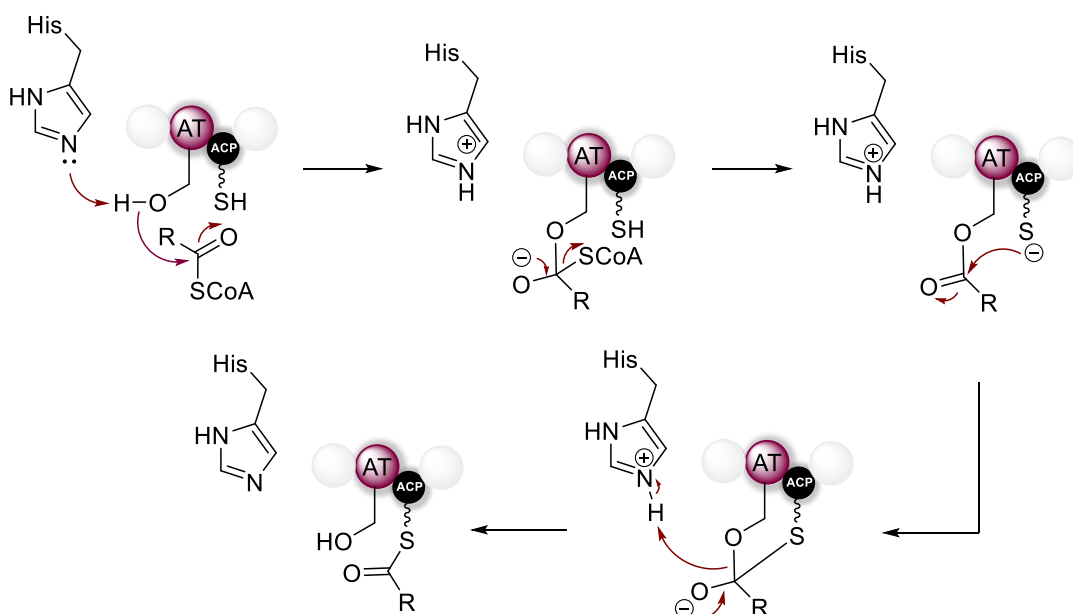
Figure 1.5.2 Structural analysis of modular type I ACP from the second module of DEBS and ACP domains from type II FAS and type I PKS with phosphopantetheine chain attached to the conserved Ser residue of the ACP. (A) The ACP comprises a bundle of four helices and an additional four residue helix (“helix 0”) which is proposed to be involved in the docking to the preceding module of the PKS (PDB: 2JU2). (B) Type II ACPs sequestering their substrates and type I ACPs does not, possibly because, as part of a multi-domain modular system, they can transfer between processing domains fast enough to avoid significant rates of hydrolysis.(PDB: 2FAE for FAS II ACP;PDB: 2LIW for type I PKS ACP).

Acyl transferase (AT) domain

The AT domains are comprised of approx. 300 residues and form two subdomains: the catalytic domain with an α/β -hydrolase fold and a ~ 60 residue subdomain that is structured in a ferredoxin-like ($\beta/\alpha/\beta/\alpha/\beta$) fold.^{74,75} The latter subdomain is believed to mediate interactions with the ACP.⁷⁶

The role of the AT domain is to prime a starter/extender unit from CoA substrates onto the phosphopantetheine arm of the ACP domain. Loading and extender AT domains present

different substrate tolerance. AT domains associated with the loading module can accept a wider variety of substrates such as acetyl-, isopropionyl-, propionyl- and crotonyl-CoA, whereas AT domains employed for chain extension present a more strict specificity, directed by “fingerprint” motifs.^{77,78} Commonly employed extender units are malonyl- and (2S)-methylmalonyl-CoA. The reaction mechanism occurs through a ping-pong mechanism forming an AT-acyl intermediate.⁷⁹ Initially, the starter/extender unit is first loaded onto a catalytic Ser residue of the AT domain, which is activated as a nucleophile by a conserved His residue. The AT-acyl group is then subjected to nucleophilic attack by the thiol residue of the phosphopantetheinyl prosthetic group of the *holo*-ACP domain (**Scheme 1.5.3**).

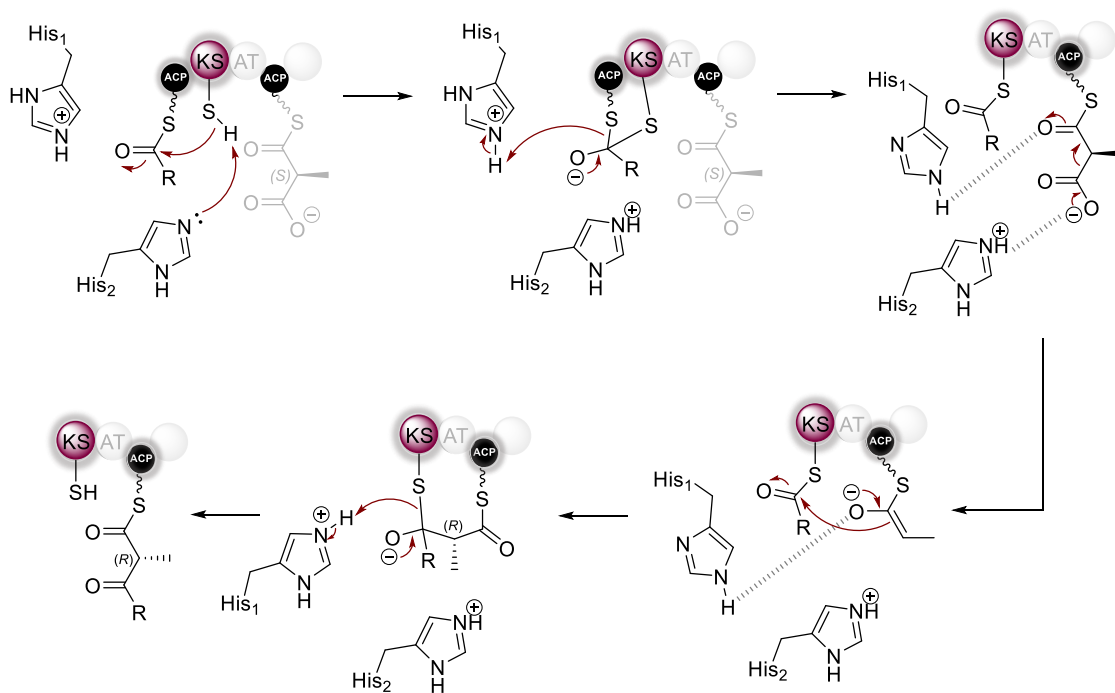


Scheme 1.5.3 Catalytic cycle of AT domain. The conserved Ser of the AT attacks the extender unit to generate an AT-acyl intermediate followed by nucleophilic attack by the thiol moiety of the PPant arm of the intramodular *holo*-ACP, furnishing the ACP-acyl complex.

Ketosynthase (KS) domain

Ketosynthase domains are homo-dimeric enzymes which are highly conserved within FASs and PKSs, and are involved in constructing the carbon skeletons of all complex polyketides through a decarboxylative Claisen condensation. The active site of KS contains a well-characterised catalytic triad of Cys-His-His. The catalysis initiates with a *trans*-thioesterification,

during which the Cys acts as a nucleophile and attacks the thioester of the polyketide intermediate from the upstream module (**Scheme 1.5.4**). Subsequently, the extender unit, covalently bound to the intramodular ACP is decarboxylated to yield a carbanion. This intermediate then attacks the polyketide intermediate bound to the KS, resulting the elongated polyketide chain. Site-directed mutagenesis studies have shown two His residues present in the catalytic triad are essential for facilitating the translocation and Claisen condensation.⁸⁰ Of these two His residues, one stabilises the enolate intermediate while the other is proposed to promote decarboxylation by activating water molecule towards nucleophilic attack on the carbonyl group extender unit.^{80,81} The chain extension reaction has been demonstrated to invert the configuration at C2 of the extender unit (**Scheme 1.5.4**).⁸²



Scheme 1.5.4 KS-catalysed chain elongation in modular PKSs. The catalysis involves a trans-thiolation and a decarboxylative condensation that results in chain elongation by two carbon units.

1.5.1.2 Processing domains

Apart from the minimal catalytic domains described above, modules may also contain additional domains which are responsible for processing the C-2 and/or C-3 atoms of the 3-keto thioester intermediate.

Ketoreductase (KR) domain

Ketoreductase (KR) domains belong to the short-chain dehydrogenase/reductase (SDR) family and its function depends on NAD(P)H to stereoselectively reduce the β -keto group of the β -ketoacyl intermediate to furnish a hydroxyl group.⁸³ KR domains consist of two subdomains: the N-terminal structural domain and the C-terminal catalytic domain, both of which present a Rossmannoid fold (**Figure 1.5.3**).⁸⁴ The catalytic subdomain contains a conserved catalytic triad of Tyr-Ser-Lys. The catalysis proceeds *via* transfer of *pro*-4S hydride from NAD(P)H to the β -carbon of the substrate and deprotonation of the generated alkoxide by Tyr residue (**Figure 1.5.3**). The catalysis proceeds stereospecifically, with KR domains referred to as A₁-type if they produce an (*S*)-configured hydroxyl group or B₁-type if they generate an (*R*)-configuration.^{82,85} Studies have shown that this can be predicted by sequence alignment of the domains.⁸⁶ The B₁-type KR domain encompass an Asp residue which is not present in the A₁-type KR domains. Structural studies of A₁- and B₁- KR domains demonstrated that the position of the catalytic residues and NADPH preserved in both structure, therefore it was suggested that stereoselectivity is determined by the binding orientation of the substrate towards the NADPH (**Figure 1.5.3**).⁸⁴ In some instances, KR domains are reported to catalyse epimerisation of α -substituents.^{82,87} Reductase-incompetent KR domains are referred to C₁-type KR domains which are missing the NADPH-binding motif as well as the catalytic Tyr.⁸⁷

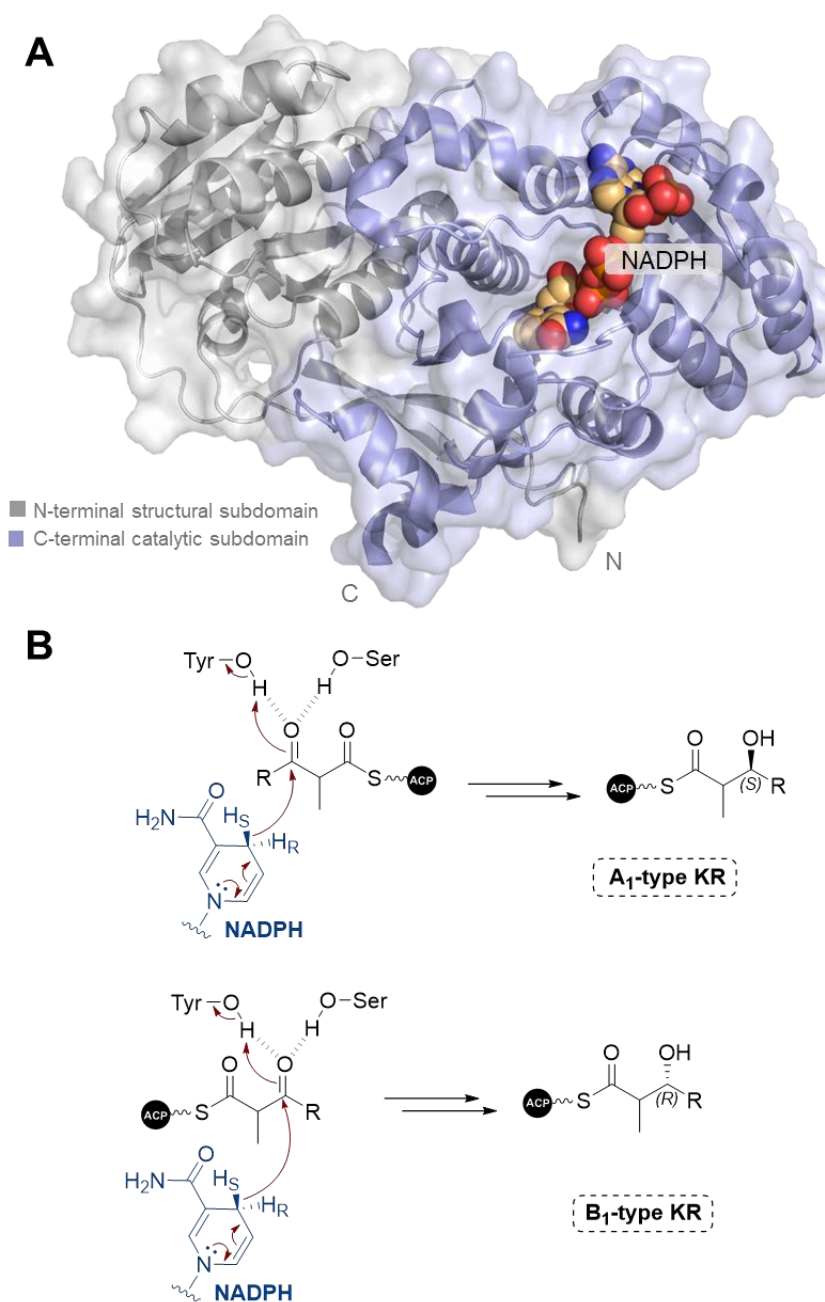
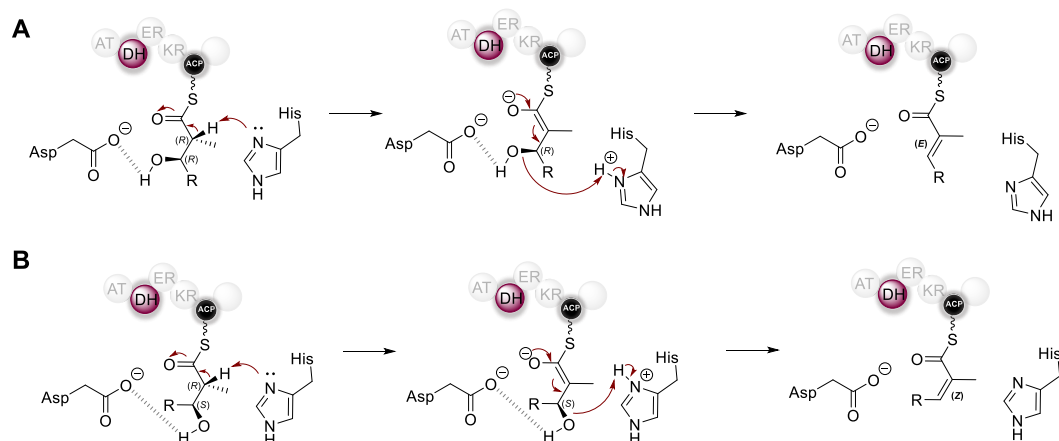


Figure 1.5.3 The crystal structure of the ketoreductase domain from DEBS module 1 (PDB: 2FR0) and mechanism of reduction of the β -keto group catalysed by KR domain. (A) KR domain is comprised of a structural (shown in grey) and catalytic (shown in light blue). The catalytic subdomain binds NADPH, which is shown in spheres. (B) A₁-type and B₁-type KR domains furnish a different stereo outcome at C-3 position.

Dehydratase (DH) domain

Following activity of a KR domain, a dehydratase (DH) domain can then be employed and process the 3-hydroxy thioester intermediate further. The majority of DH domains catalyse elimination of water from the β -hydroxy-acyl-ACP substrate yielding a *trans*-double bond between the α - and β - carbons; however, some DH domains have been shown to form *cis*-double bonds. DH domains exhibit a double hot dog fold containing a conserved His-Asp/Glu active site dyad, and form a dimer along the 25-residue stretch of the N-terminus.^{88–90} The dehydration of the 3-hydroxy thioester is proposed to occur via *syn*-elimination mechanism to generate a 2-enoyl thioester (**Scheme 1.5.5**).^{91,92} Therefore the chirality of the β -hydroxy group determine the geometry of the double bond, resulting in the dehydration of (*R*)-configured alcohols to yield (*E*)-configured alkenes and (*S*)-configured alcohols to (*Z*)-configured alkenes.⁹³ The dehydration is proposed to occur *via* an E1_{CB} mechanism (**Scheme 1.5.5**). In particular, deprotonation occurs at the C-2 carbon by the His to form an enolate from which the C-3 hydroxyl group is eliminated.^{88,89} The Asp/Glu residue is thought to bind and orient the hydroxyl group of the substrate in the stereoelectronically favored conformation.⁹³



Scheme 1.5.5 DH-catalysed dehydration of (3*R*)/(3*S*)-hydroxy thioester intermediate. *Syn* elimination mechanism with His-Asp to achieve stereocontrol.

Enoyl reductase (ER) domain (*cis*-acting)

The resulting double bond generated by the DH domain can be reduced by *cis*-acting enoyl reductases (ERs) to yield an α,β -saturated acyl-ACP. The ER domains are the least understood parts of the polyketide elongation. Unlike their functional analogs in bacterial FAS which are SDRs, they fold similar to medium-chain dehydrogenase/reductase (MDR) enzyme family, containing a C-terminal NADPH binding subdomain and a N-terminal substrate binding subdomain.⁹⁴ Unlike other members of the MDR family, crystal structure obtained for the excised KR-ER₂ didomain from the bacterial spinosyn (spn) pathway revealed that the ER exists as a monomer with the active site located in the cleft between core cofactor binding and substrate-binding subdomain (**Figure 1.5.4**).

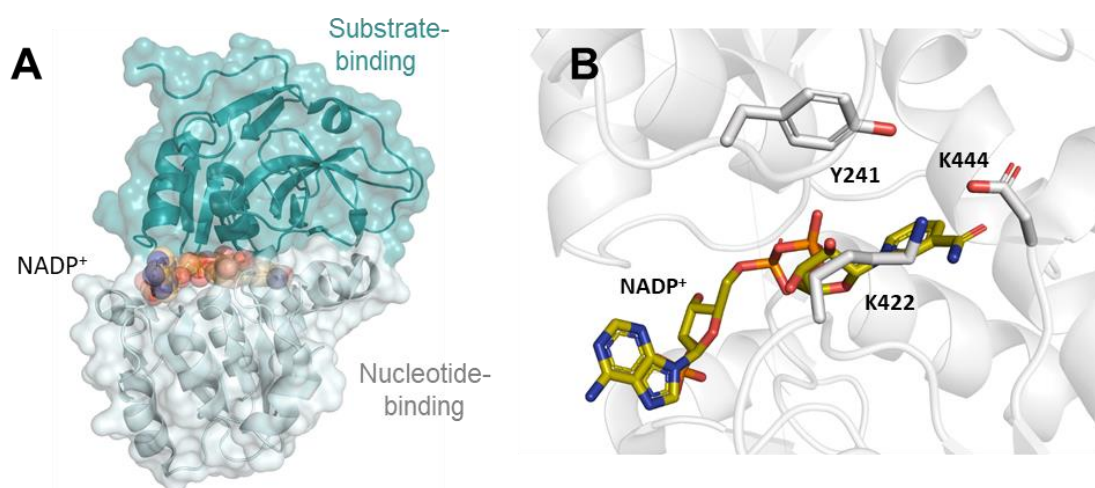
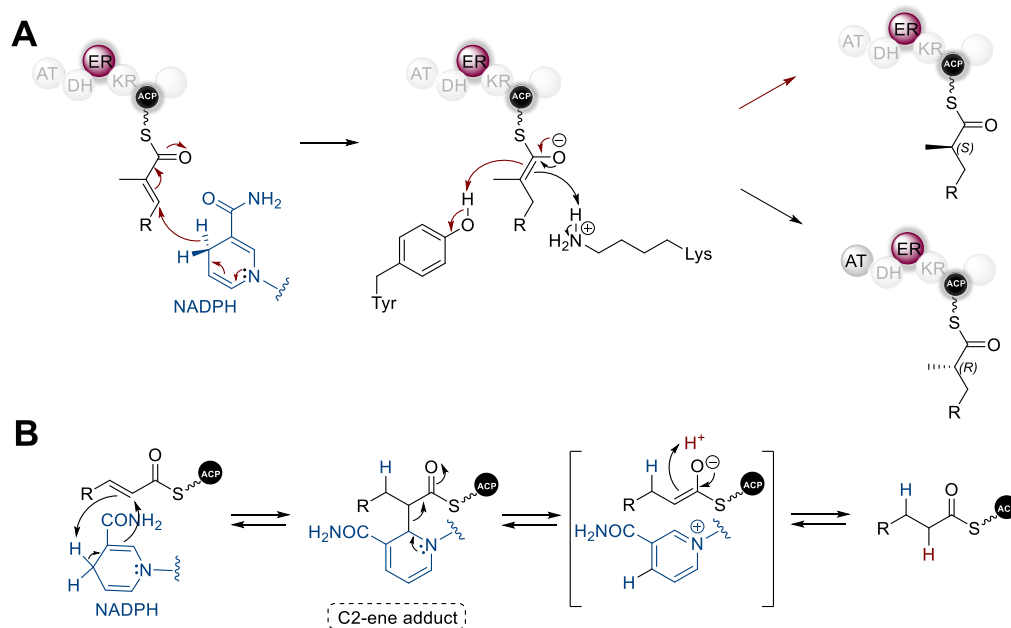


Figure 1.5.4 Structure of *cis*-acting ER domain of spinosyn modular PKS (PDB: 3SLK). (A) Stereogram of SpnER2 (PDB: 3SLK) comprising of the substrate-binding (shown in cyan) and nucleotide-binding subdomain. NADPH binding site is placed in a cleft between the two subdomains. (B) Conserved Tyr-Asp-Lys catalytic residues located close to the nicotinamide moiety.

The oligomeric state of many ERs is believed to have evolved from the dimeric form to enable intramodular ACPs to reach their cognate domains.^{62,95} The catalytic mechanism involves the transfer of the *pro*-4*R* hydride of NADPH to the C-3 position of the enoyl substrate followed by stereospecific protonation at the C-2 which directs the configuration of the methyl group appended in the generated saturated acyl thioester (**Scheme 1.5.6**).⁹⁶ A more recent suggested

ene mechanism encompasses the formation of a transient covalent NADPH-substrate adduct (“C2-ene adduct”) in the transfer of the hydride equivalent followed by conversion to the consensus enolate intermediate which is protonated to yield the reduced product (**Scheme 1.5.6**).⁹⁷ In both scenarios, a proton donor in the active site is required and is proposed to determine the configuration of the C-2 methyl branch during reduction catalysed by ER domains. Three conserved residues were identified in SpnER₂ around 6 Å of the nicotinamide 4-*pro-R* hydrogen atom which were proposed to mediate protonation; Asp, Tyr and Lys (**Figure 1.5.4**).^{96,98}



Scheme 1.5.6 *Cis*-acting ER-catalysed reaction mechanisms. (A) Mechanism of the reduction of the α,β-double bond catalysed by the ER to afford stereodefined saturated acyl thioesters. (B) The ene mechanism proceeds through a C2-intermediate which is converted to its corresponding enolate intermediate and subsequently protonated.

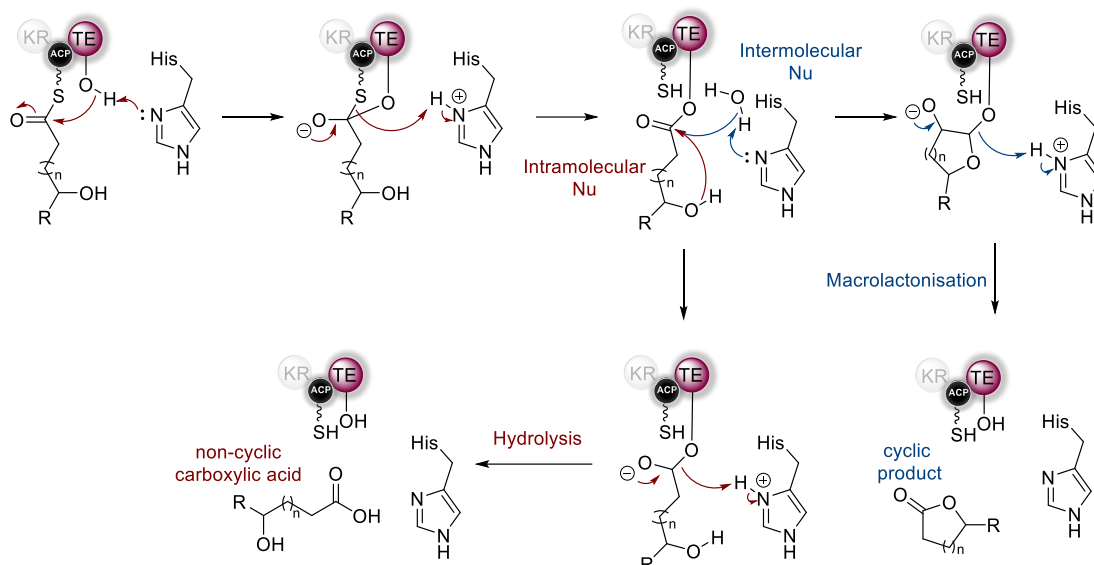
The Lys and Tyr lie on opposite sides of the active site cleft to protonate the C-2 carbon of a bound polyketide substrate. It is proposed that when Tyr is present, it acts as a proton donor, but in its absence the Lys delivers its proton from the opposite site, hence the reverse in stereochemistry observed with Tyr to Val mutants in EryER₄.⁹⁶ Mutagenesis experiments on putative proton-donating residues so far have failed to identify the proton donor that determines the stereochemistry of the reaction.^{95,99} The above indicate that rational manipulation of ER

stereochemistry is not straightforward and further stereochemical determinants in the ER active site needs to be identified.

Apart from the processing domains described above, some PKS modules employ methyltransferase (MT) domains which perform methylation either the 2-position of the 3-keto thioester (C-MTs) or the 3-hydroxy group resulting from ketoreduction (O-MTs) using *S*-adenosylmethionine (SAM) cofactor.^{100,101}

1.5.1.3 Chain release mechanisms in PKSs

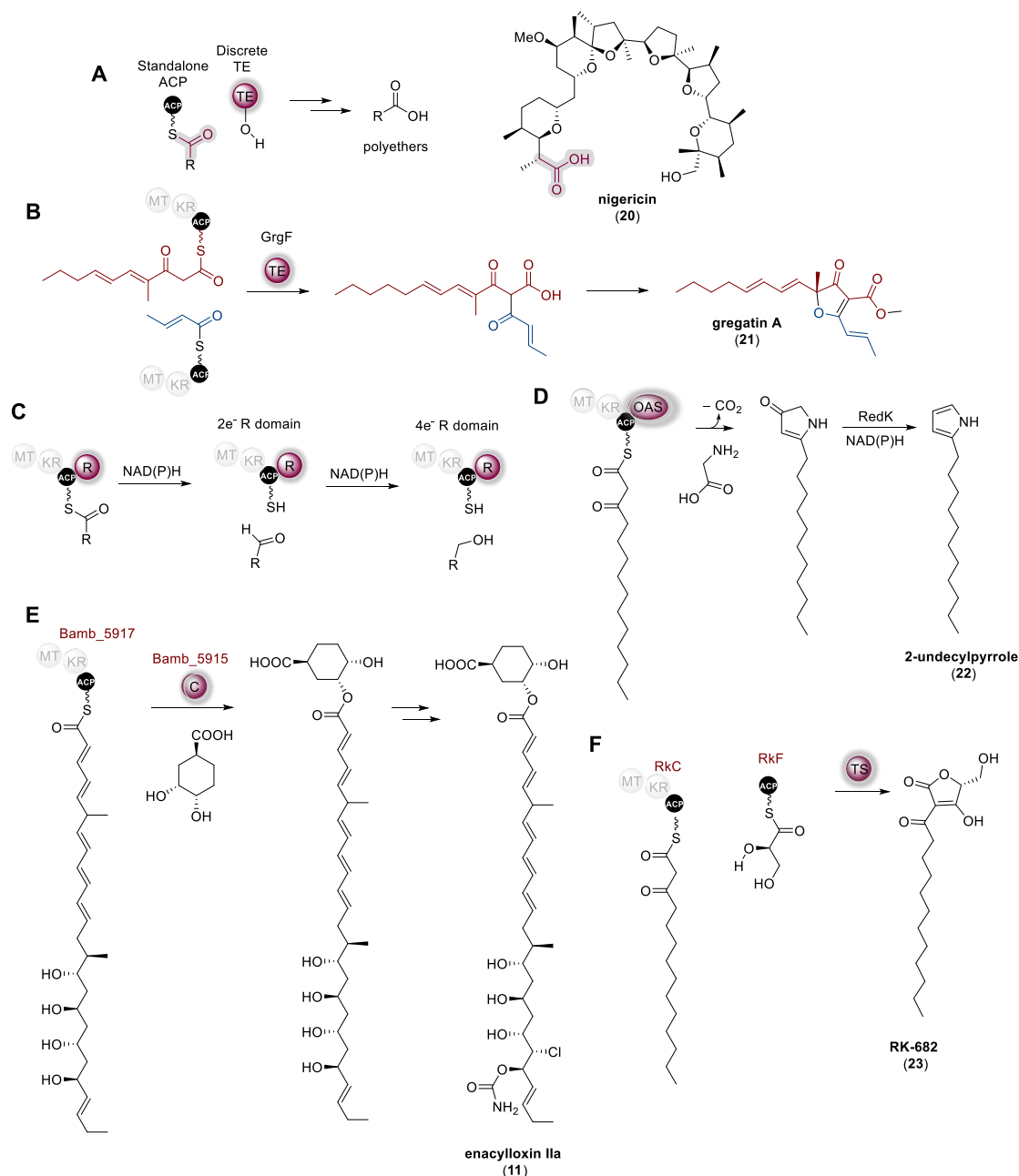
After the final chain extension, the fully assembled polyketide chain is left tethered to the ACP located on the C-terminus of the PKS. The release of the polyketide product is performed by numerous mechanisms, exploiting both enzymatic domains integrated into the PKS or standalone chain release enzymes. The most common domain for polyketide chain release are the α/β hydrolase fold thioesterases (TEs). Type I TEs are located in the C-terminus of the PKS. The offload catalysis is initiated by transesterification of the ACP-bound polyketide onto a conserved Ser residue of the TE domain, which is catalysed by a conserved His residue (**Scheme 1.5.7**).¹⁰² The chain is then offloaded through either intramolecular or intermolecular nucleophilic attack catalysed by the same conserved His residue.¹⁰³



Scheme 1.5.7 General mechanism for PKS released by type I TE domains. Depending on the TE, the acyl chain of acyl-TE can be hydrolysed (shown in red) or macrocyclised (shown in blue) to yield a macrolactone or macrolactam if an amino group is present.

Type II TE domains are found as standalone domains and like their domain counterparts, type II TE domains also contain an α/β hydrolase fold with the conserved Ser-Asp-His triad.¹⁰⁴ They are phylogenetically distinct from type I TE domains and contain a conserved Met next to the catalytic Ser residue. These act as proofreading enzymes and are responsible for hydrolysing small non-reactive thioester intermediates that stall the PKS. However, examples of nigericin¹⁰⁵ (**20**), maduramicin¹⁰⁶ and others illustrate that these domains are responsible for catalysing the product chain release by translocating the final polyketide into a standalone ACP protein followed by hydrolytic release by the type II TE enzyme. Apart from the hydrolytic activity, *in vivo* and *in vitro* reconstitution studies demonstrated that GrgF, a type II TE enzyme, catalyses an unusual chain-fusion of a C₁₁ and C₄ carbon polyketides *via* Claisen condensation followed by hydrolytic chain release of the linear dimer gregatin (**22**).¹⁰⁷

Aside from TE-catalysed chain release other chain offloading mechanisms have become apparent. A reductase (R) domain can be located in the C-termini of PKS enzymes which are proposed to catalyse reductive release of polyketide chains as aldehydes via a two-electron reduction or primary alcohols via a four-electron reduction (**Scheme 1.5.8**).¹⁰⁸



Scheme 1.5.8 Examples of chain release mechanisms employed by modular type I PKSs to enhance structural diversity. **(A)** Type II chain release in polyether biosynthesis. **(B)** In gregatin (21) biosynthesis, the type II TE, GrgF, catalyses Claisen condensation to join to polyketide chains and hydrolysis to yield the linear dimer. **(C)** Mechanism of reductase domains involves a two electron reduction to release the corresponding polyketide as aldehyde or four electron reduction to release a primary alcohol. **(D)** Chain release catalysed by OAS proceeds through decarboxylative condensation of an amino acid and the ACP-bound polyketide thioester. **(E)** Bamb_5915 catalyses the chain release via transesterification with DHCCA. **(F)** Ketosynthase-like tetronate synthases (TRs) catalyse offloading

through condensation of the polyketide chain with an ACP-bound glyceryl thioester to furnish a 2-acyl-4-hydroxymethyl-tetronic acid.

Pyridoxal 5'-phosphate (PLP)-dependent α -oxoamine synthase (OAS) domain is another class of discrete enzymes to catalyse chain release. The catalysis proceeds *via* a decarboxylative Claisen condensation with the α -carbon of an amino acid. The first characterised example of this type of catalysis encompass OAS enzymes from the prodiginine biosynthetic pathway where condensation occurs between ACP-bound β -keto- β -pyrrolyl-propanoyl and L-glycine to release the 4-methoxy-2,2'-bipyrrole-5-carbaldehyde intermediate which is subsequently condensed with 2-undecylpyrrole to yield undecyl prodiginine (**22**) (**Scheme 1.5.8**).^{109,110}

Another example of an unusual mechanism of offloading comprises the chain release catalysed by condensation-like enzymes (**Scheme 1.5.8**). Condensation (C) domains are one of the core domains appended in non-ribosomal peptide synthetase (NRPS) modules. C domains catalyse C-N bond formation between a peptidyl carrier protein (PCP) bound NRP chain and an amino acid tethered to the downstream PCP domain of the module. An example of free-standing C domain, Bamb_5915, is observed for modular PKS chain release in the enacyloxin IIa (**11**) biosynthesis.¹¹¹ The offloading proceeds through intermolecular esterification reaction between the ACP-bound polyketide chain and the C3 hydroxyl of the (1*S*,3*R*,4*S*)-3,4-dihydroxycyclohexane carboxylic acid (DHCCA). Various DHCCA synthetic analogues were successfully condensed by Bamb_5915, illustrating the potential of the enzyme in producing novel enacyloxin compounds.

KS domains are also reported to facilitate chain release, apart from their established role in chain extension (**Scheme 1.5.8**). KS-like domains initiate condensation of the polyketide chain with a glyceryl-ACP to furnish a 4-hydroxymethyl-2-acyltetronic acid in the biosynthesis of the tetronate RK-682 (**23**).¹¹²

The mechanism of chain release affect greatly the structure of the final biosynthetic product, and therefore it can be utilised for bioengineering purposes to alter pre-existing assembly lines.

1.5.2 Post-PKS tailoring

Either following the offloading from the synthase, or while still tethered to one of the ACP domains, polyketide intermediates can be subjected to a variety of modifications to further enhance their structure diversity.¹¹³ These modifications are often required for polyketide's biological activity and are performed by enzymes encoded in the same BGC. Common post-PKS polyketide tailoring reactions include oxidation to furnish hydroxyl or carbonyl groups, methylation at carbon, nitrogen or oxygen centres, cyclisation, acylation, alkylation and glycosylation. Utilising these processing enzymes provide great potential to produce novel natural product scaffolds.

1.5.3 *Trans*-AT PKSs

The DEBS PKS represents a textbook example of the collinearity principle. This term is used to describe the phenomenon that the order and function of catalytic domains within the modules of a PKS generally correlate to the structure of the generated polyketide. This allows the prediction of the polyketide structure based upon domain organisation, and therefore genetic information alone.⁶¹ Due to the discovery of novel polyketide synthases with non-canonical modules which do not fit the principle of co-linearity, type I modular PKSs were further subdivided into two classes. The predominant feature that these non-canonical PKSs possess is the presence of an AT domain as a free-standing enzyme and not embedded in within each module (**Figure 1.5.5**).^{114,115}

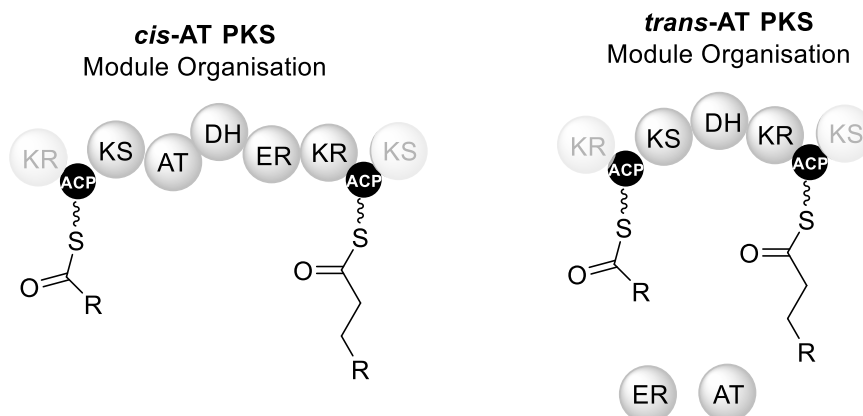


Figure 1.5.5 Comparison of a *cis*-AT PKS module harbouring integrated AT and ER domains to *trans*-AT PKSs module which encode standalone AT and ER domains.

Therefore, this new type is referred as *trans*-AT PKS. Examples of *trans*-AT PKSs include the PKSs responsible for the biosynthesis of bacillaene¹¹⁶, pederin and leinamycin¹¹⁷. Usually a single malonyl-specific AT encoded by one gene in the cluster is responsible for interacting with most of the PKS modules in the systems.¹¹⁵ A few notable exceptions include the ethyl-malonyl specific AT in kirromycin biosynthesis¹¹⁸, the AT accepting methoxymalonyl units in oxazolomycin¹¹⁹ and succinyl units in etnangien biosynthesis¹²⁰. The AT domains from *trans*-AT PKSs can be also found fused with an additional ER and/or acyl hydrolases (AHs) that shows hydrolytic activity in various acyl thioesters except malonylated, indicating a proof-reading function.¹²¹ Standalone ER domains recruited in *trans*-AT PKS are phylogenetically distinct from integrated ER domains found in *cis*-AT PKSs and they closer resemble more polyunsaturated fatty acid synthases.^{122,123} A more detailed analysis on *trans*-acting ERs is presented in **Chapter 3**.

Another characteristic of a *trans*-AT PKS is the non-elongating ketosynthases (KS⁰). These domains are structurally identical to KS domains (section 1.5.1.1) apart from the His conserved residue required for the decarboxylative Claisen condensation, therefore they are not capable of catalysing chain elongation. However, they still harbour the conserved Cys that allow them to mediate transfer of the acyl chain onto the downstream ACP domain as shown in enacyloxin biosynthesis. Additional uncommon features that distinguish *trans*-AT PKSs from their *cis*-AT counterparts include modules of unusual domain organisations that are split across two

subunits and incorporation of trans-acting β -branching cassettes that are responsible for the introduction of β -branches into the polyketide chain

Although both systems seems to be similar, phylogenetic analyses revealed that they have distinct evolutionary origins. *Cis*-AT PKSs is proposed to evolve through module duplication whereas *trans*-AT PKS evolution is directed through horizontal gene transfer.^{124,125}

1.5.4 Overall architecture of PKS modules

Due to the assumption that FAS and PKSs are likely to have derived from a common ancestor⁶⁷, the architecture of PKS modules was until recently expected to emulate the type I iterative porcine FAS which was structurally characterised by X-ray crystallography (**Figure 1.5.6**).¹²⁶ The first structural information for the full-length type I PKS module was obtained from pikromycin PKS module 5 (PikAIII) from *Streptomyces venezuelae*, using single-particle cryo-electron microscopy (cryo-EM) and highlighted the difference in architecture compared to the FAS (**Figure 1.5.6**).^{127,128} The architecture possess a single-reaction chamber in comparison to a more open architecture of mammalian FAS which contains two ACP chambers. The single chamber provides access to all catalytic sites for the intramodule ACP whereas the ACP from preceding module utilises an outside reaction chamber to deliver the upstream polyketide intermediate which reduces the risk of unwanted substrates delivered by neighbouring ACPs.

Recently, a 3.1 angstrom-resolution cryo-EM structure of the Lsd14 polyketide synthase revealed an asymmetrical architecture that contradicts previously reported structures that supported a twofold symmetry.¹²⁹ Similar to mammalian FAS, Lsd14 contains two reaction chambers though only one of them contains the catalytic domains, indicating that only one chamber generates the polyketide product. In Lsd14

structure, ACP domain is only found in one of the two reaction chamber and it is positioned either docked to the AT or to the KS domain. The second ACP is disordered indicating that the domain is not docked to any domain. This asymmetrical arrangement of domains was further supported by Cogan *et al.* by visualising in action DEBS module 1.¹³⁰

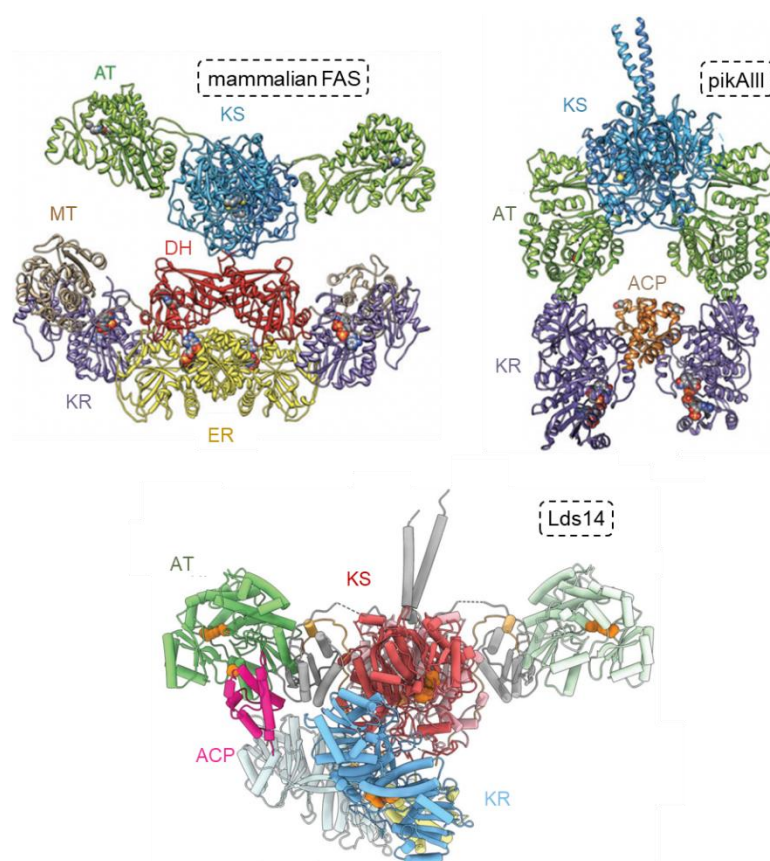


Figure 1.5.6 Structural architecture of mammalian FAS and PKS modules. Active sides are highlighted in orange. Structures reproduced from ref. 127 and 129.

PKS engineering offers the opportunity to generate numerous polyketides for drug discovery. Efforts towards domain substitution, insertion and deletion resulted in low yields of the new natural product analog which can be attributed to possible protein misfolding and deficient domain-domain interactions due to incorporation of non-native domains¹³¹. Future work on understanding PKS modular dynamics could provide insights for conducting minimally invasive domain manipulation for PKS engineering purposes.

1.6 Study objectives

Genome mining of the clinical isolate *B. gladioli* BCC0238 revealed that the strain contains numerous cryptic clusters in which their metabolic products are currently unknown. The main aim of this research project was to identify the metabolic product of an orphan *trans*-acyl transferase PKS-encoding BGC as well as identify its biological activity.

A second aim was to investigate the biosynthetic pathway of this metabolite which encompasses moieties mainly found in NPs from *Streptomyces* sp. A key feature in the gladiostatin biosynthesis presented by *in silico* analysis is an AsfA-like domain appended to the C-terminus of the PKS that is hypothesised to catalyse polyketide chain release. The AfsA enzymology, which has been characterised for the biosynthesis of γ -butyrolactones in *Streptomyces*, has never been found appended to a polyketide synthase (PKS) pathway as a release mechanism. Another interesting feature of gladiostatin biosynthesis is an unprecedented number of *trans*-acting enoyl reduction events, which are predicted to act at various points during polyketide chain assembly. A detailed understanding of the mechanisms underpinning the interaction of these enzymes with the PKS will facilitate future biosynthetic engineering efforts.

Finally, understanding the evolution of PKS systems within the same species was pursued in an effort to enhance our understanding in manipulating these systems to produce a diverse arsenal of natural products.

Chapter 2

*Discovery, biosynthesis and biological evaluation of gladiostatin**

[·] This Chapter is based on:

I. T. Nakou, M. Jenner, Y. Dashti, I. Romero-Canelón, J. Masschelein, E.
Mahenthiralingam and G. L. Challis. *Angew.Chem. Int. Ed.* **2020**, 59, 23145–23153

2.1 Introduction

2.1.1 Biosynthetic potential of *B. gladioli* BCC0238 strain

As described in **section 1.3**, *Burkholderia* species are found in a range of ecological niches and produce a vast array of specialised metabolites. Despite this, a large number of silent biosynthetic gene clusters (BGCs) in the *Burkholderia* genome remain uncharacterized and represent a promising resource for new natural product discovery.¹³² *B. gladioli* BCC0238 is a clinical isolate obtained from the sputum of a cystic fibrosis patient in 1996 (Genome accession number: GCA_900631635). Genome sequencing previously conducted in our group identified the promising biosynthetic potential of this strain. As shown in (**Figure 2.1.1**), antiSMASH¹³³ analysis predicted the presence of putative BGCs located on the second chromosomal replicon of the strain with a potential of producing novel metabolites. Although *in silico* analysis revealed a vast biosynthetic capability of this clinical isolate, initial LC-MS-based analyses of secondary metabolite production identified two compounds that were successfully connected to their BGC using insertional mutagenesis strategies: gladiolin (**9**), a novel macrolide antibiotic with promising activity against *Mycobacterium tuberculosis*, and icosalide A1 (**11**), an asymmetric lipopeptidolide antibiotic which was originally believed to be of fungal origin (**Figure 2.1.1**).^{134,135} This inconsistency between the number of BGCs and the amount of secondary metabolites present in the metabolic profile could be attributed to the controlled culture conditions, using glycerol as the sole carbon source. Indeed, secondary metabolite production usually occurs during the late growth phase of the producing microorganism, and although the nature of their formation is translated genetically, expression can be greatly influenced by changes in their habitat (nutrients, microorganism competition, pH, temperature, etc.).^{136,137} Previous work on this strain also revealed that altering the carbon source resulted in a different metabolic profile. This strategy was used for the identification of bolagladins (**24**).¹³⁸ It was observed that after switching to a mixture of glycerol and ribose as carbon sources, induced

the production of bolagladins A and B, novel lipodepsipeptides containing a unique citrate-primed fatty acid and an unusual dehydro-b-alanine residue.

In efforts to activate the production of other specialised metabolites produced by the *B. gladioli* BCC0238 strain, we targeted a ~50 kb cryptic BGC (*gds*) encoding a *trans*-AT PKS. Bioinformatics analysis indicated that this cluster encodes for a cassette of enzymatic domains with close homology to those responsible for the biosynthesis of the glutarimide class of polyketide antibiotics in *Streptomyces* species.

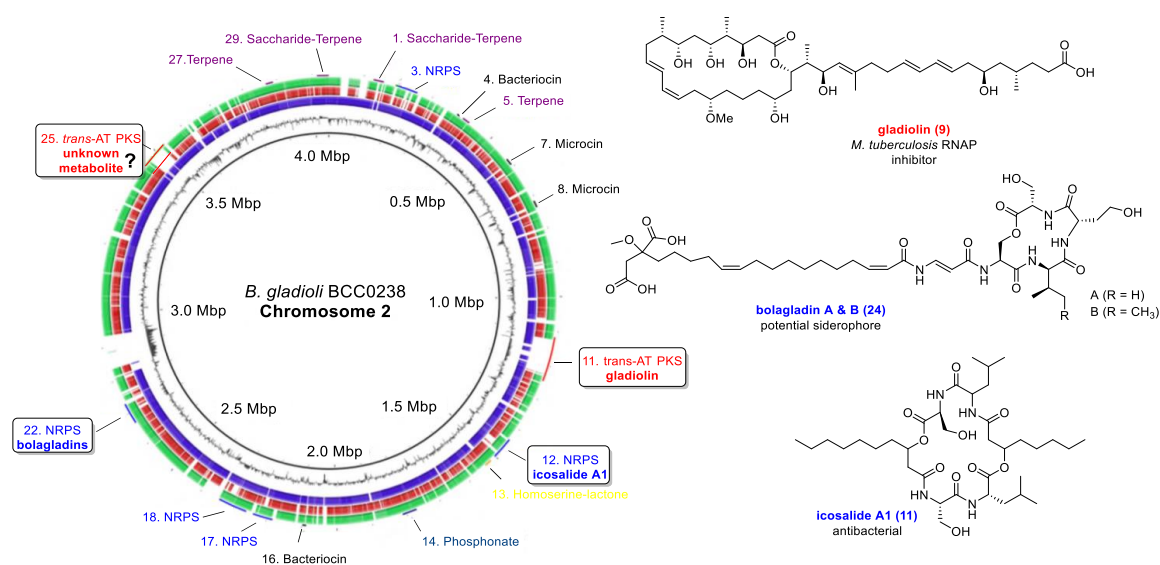


Figure 2.1.1 Biosynthetic potential of *B. gladioli* BCC0238. Comparison of chromosome 2 of the strain BCC0238 to the complete genome sequence of *B. gladioli* BSR3 (green, 99% identity) and the draft genome of NBRC 13700 (blue, 99% identity) and 3848s-5 (red, 99% identity).¹³⁹ The putative BGCs are shown in various colours. The orphan BGC studied here is indicated with a question mark.

2.1.2 Glutarimide-containing polyketides

Glutarimide-containing polyketides are a diverse class of polyketides that share a characteristic 2,6-piperidinedione motif in their structure, also known as a glutarimide (**Figure 2.1.2**). The associated gene clusters have been identified mainly in Actinomycetes from the *Streptomyces* genus.

Cycloheximide (**25**) is one of the most well-known members of the glutarimide-containing polyketide family, which was originally isolated from *S. griseus* and subsequently rediscovered

in *Streptomyces* sp. YIM56141, YIM56132, *S. albulus* and *S. noursei*.^{140–142} Many of the strains reported to produce **25** also produce actiphenol (**26**) which shares the same carbon skeleton as **25** but the cyclohexanone moiety is replaced by a phenol. Streptimidone (**27**) and its 9-methyl derivative, 9-methylstreptimidone (**28**) represents non-carbocyclic members of the glutarimide group of polyketides.^{143–145} The 14-membered macrolide migrastatin (**29**), was first isolated from *Streptomyces* sp. MK929-43F1 and represents one of the subclasses of glutarimide-containing polyketides.^{146–148} Linearized migrastatin congeners, dorrigin A (**30**), 13-*epi*-dorrigin A (**31**) and dorrigin B (**32**) were first isolated from *Streptomyces plantensis* and was later hypothesised that these acyclic migrastatin congeners are biosynthesised by the same pathway as **29**.^{149,150}

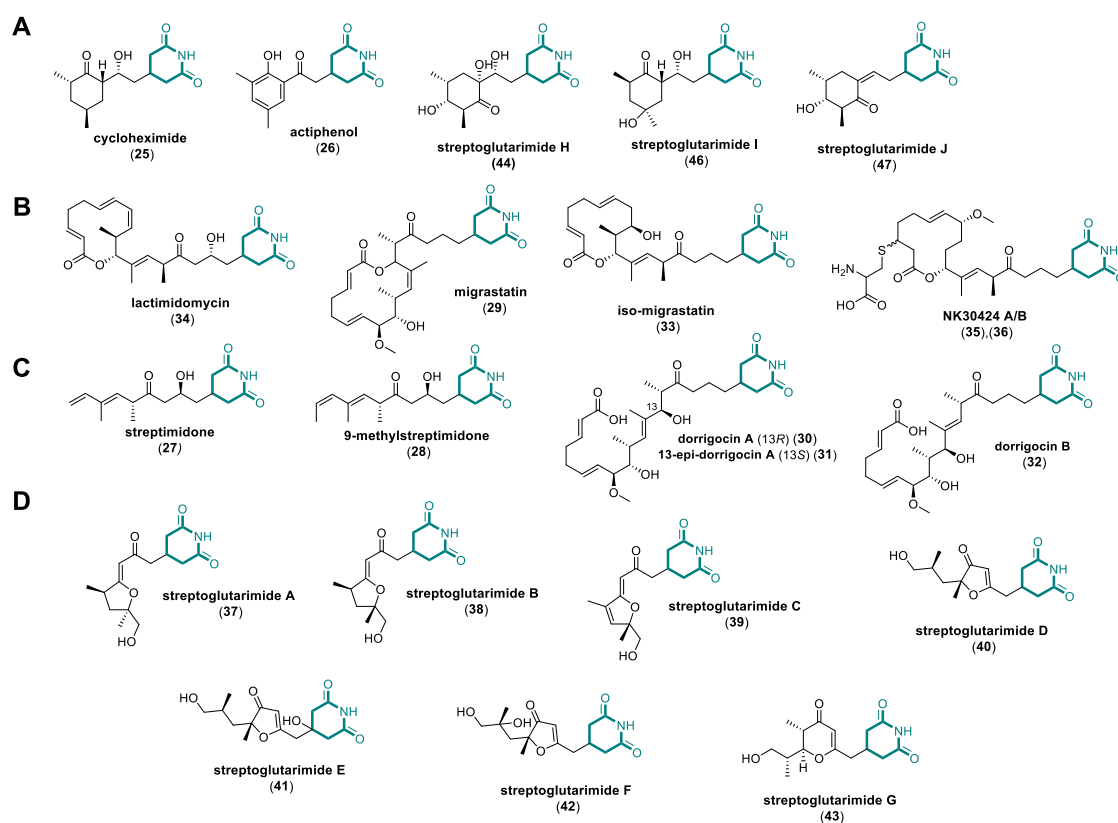


Figure 2.1.2 Structures of known glutarimide antibiotics derived from *Streptomyces* species. The family of glutarimide-containing polyketides is subdivided here in: **(A)** glutarimides containing cyclohexanone/phenol moiety in C-4, **(B)** glutarimide-containing polyketides with a 12/14-membered macrolide, **(C)** acyclic members of the glutarimide antibiotics and **(D)** glutarimides bearing a modified furan moiety. The common 2, 6-piperidinedione moiety presented in all metabolites is highlighted in turquoise.

Careful analysis of the metabolic profile of re-cultured *S. plantensis* extracts proved this hypothesis and also revealed the presence of a new 12-membered macrolide class of glutarimide-containing polyketide, iso-migrastatin (**33**).¹⁵⁰ Three additional representatives of the 12-membered macrolide class of glutarimide-containing polyketides are lactimidomycin (**34**) which was isolated in 1992 from *Streptomyces amphibiosporus* and NK30424 A (**35**) and B (**36**) which were isolated from *Streptomyces* sp. NA30424.^{151,152} Recently, a study reported 10 new streptoglutarimides (A-J) (**37-47**) which consist of a 15-carbon compounds with a 2,6-piperidinedione motif bearing a side-chain at the C-4 position.¹⁵³ The side chains of the streptoglutarimides A-F (**37-42**) possess a modified furan moiety, which has not been previously reported for other glutarimide-containing polyketides.

The domain organisation of all known glutarimide PKSs is almost identical over the first four modules (**Figure 2.3**). Three conserved genes (*ltmHCD*, *mgsHCD*, *chxBCD* and *smdFGH* in the lactimidomycin, migrastatin, cycloheximide and 9-methylstreptimidone BGCs, respectively) are proposed to generate a malonamyl thioester starter unit for the PKSs that assemble these metabolites (**Figure 2.1.3**).^{140,146,151,154–157} These genes are responsible for encoding an acyl transferase (AT; LtmH/MgsH/ChxB/SmdF) that malonylates an acyl carrier protein (ACP; LtmC/MgsC/ChxDC/SmdG) and an asparagine synthetase homologue (LtmD/MgsD/ChxX/SmdH), which is proposed to convert the resulting malonyl thioester to a malonamyl thioester. The first module in each PKS (located in the LtmE, MgsE, ChxE and SmdI subunits) contains seven conserved domains that are proposed to elaborate the malonamyl unit into a common 2-(2,6-dioxopiperidin-4-yl)-acetyl thioester intermediate (**Figure 2.1.3**).

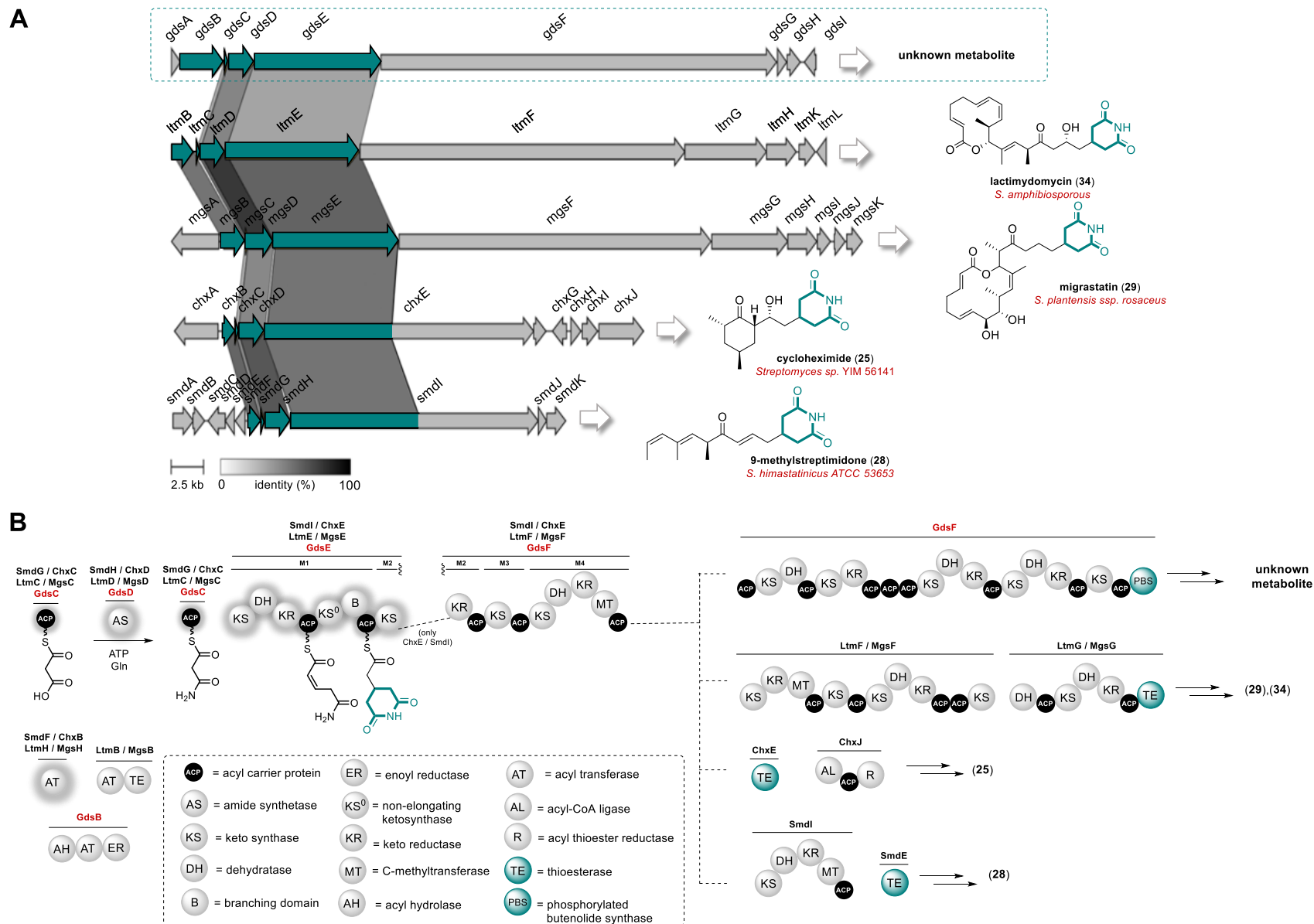
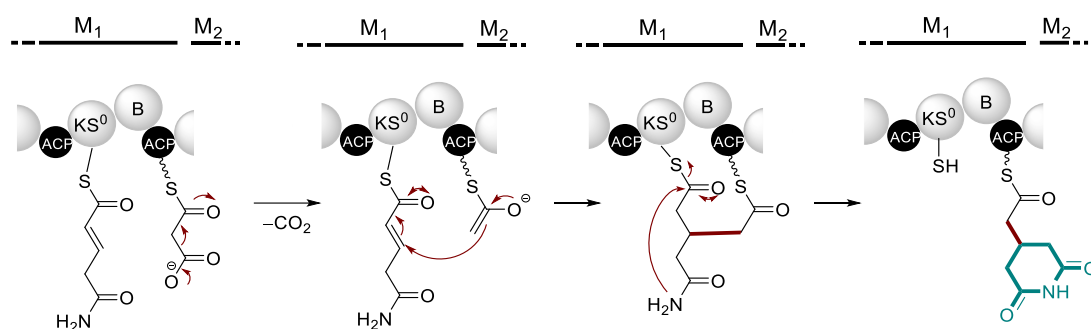


Figure 2.1.3 Comparative analysis of *chx*, *smd*, *mgs*, *ltm* BGCs that direct glutarimide biosynthesis in *Streptomyces* species to the cryptic *gds* BGC.

(A) Comparison of biosynthetic gene clusters (BGCs) that direct glutarimide biosynthesis in *Streptomyces* species (lactimidomycin (*ltm*), migrastatin (*mgs*), cycloheximide (*chx*) and 9-methylstreptimidone (*smd*)) with the cryptic *B. gladioli* BCC0238/BCC1622 gene cluster (highlighted by the dashed box). The genes (or regions thereof) encoding the machinery responsible for assembling the common 2-(2,6-dioxopiperidin-4-yl)acetyl thioester biosynthetic intermediate are highlighted in blue. Figure was generated using Clinker (identity threshold: 0.3). **(B)** Comparison of the 9-methylstreptimidone, cycloheximide, lactimidomycin and migrastatin PKS domain architectures with the PKS encoded by the cryptic *B. gladioli* BGC. Shadowed domains are involved in constructing the common 2-(2,6-dioxopiperidin-4-yl)acetyl thioester intermediate. Domains responsible for chain release (TE and PBS) are highlighted in turquoise. Abbreviations for protein domains are displayed in the dashed box.

The KS-B (sometimes denoted as X)-ACP tri-domain is present in rhizoxin (*rhi*) PKS, and is responsible for the δ -lactone installation in rhizoxin biosynthesis.^{158,159} A crystal structure of RhiE KS-B di-domain showed that the B domain follows a double hot dog (DHD) fold that is present in dehydratase domains.¹⁵⁸ A range of chimeras were constructed to demonstrate complementation between B/X-domains and the structural rather than catalytic role they play in the formation of β -branched polyketides.¹⁶⁰ In analogy to δ -lactone formation in rhizoxin pathway, the installation of the glutarimide functionality is proposed to start after decarboxylation of an ACP-bound malonyl group which then attacks the β -position of a KS-bound α,β -unsaturated thioester, introduced earlier in the pathway (**Scheme 2.1.1**). Subsequent cyclisation liberates the glutarimide functionality. Biochemical, kinetic and mutational analyses have identified that the KS domain alone is in charge of mediating the entire catalytic sequence, which leads to lactone or glutarimide rings.^{158,160}



Scheme 2.1.1 Proposed model for the formation of the glutarimide head. The mechanism proceeds through a decarboxylative Michael addition of the ACP-bound malonate unit to the KS-bound α,β -unsaturated thioester. The branched polyketide intermediate is then covalently linked to the ACP and the KS domain. Subsequent nucleophilic attack of the carboxamide into the KS-bound thioester yields the glutarimide moiety in the polyketide chain which is then passed to the downstream assembly line.

Similarly, *gdsB*, *gdsC* and *gdsD* genes in the cryptic *B. gladioli* BCC0238 BGC encode homologues of the three proteins hypothesised to create the malonamyl thioester starter unit (**Figure 2.1.3**). The first module of the PKS encoded by this cluster (situated in the GdsE subunit) has an identical seven-domain architecture to the corresponding modules of the PKSs that assemble the *Streptomyces* glutarimides. The domain architecture of the subsequent modules in the *B. gladioli* BCC0238 PKS diverges from the *Streptomyces* glutarimide assembly lines. In most members, the polyketide chain terminates with a thioesterase (TE) domain appended to the C-terminus of the last module, which catalyses release of the fully assembled polyketide chain via hydrolysis or macrocyclisation, except for **28** for which the side chain is acyclic. In this case, Ju and co-workers propose that the fully assembled polyketide backbone is released by the type II TE SmdE encoded in the cluster.¹⁵⁴ In contrast to these other glutarimide-producing assembly lines, an AfsA-like domain (pfam03756-identified by a conserved domain search)¹⁶¹ is found appended to the C-terminus of the last PKS module of the *gds* pathway. Studies on this mechanism of polyketide chain termination are described in **section 1.5.1.3**.

Glutarimide-containing polyketides are well-studied for their antitumor properties. Previous work has shown that **34** and **25** are able to inhibit protein synthesis by blocking the elongation step in eukaryotic protein translation (**Figure 2.1.4**).^{162,163} Biochemical and structural analyses

have shown that the glutarimide moiety of **25** and **34** compete with the binding of the tRNA trinucleotide CCA-end in the E site of the large subunit (**Figure 2.1.4**).¹⁶³ In addition, **34**, **29**, **33** have been shown to possess anti-metastatic properties, inhibiting tumour cell migration, however the mechanism for this is currently unclear.¹⁶⁴ Members of this family of polyketides also present antiviral activities; lactimidomycin is active against dengue and other RNA viruses from the *Flaviviridae* family but also vesicular stomatitis virus and poliovirus 1, suggesting that the elongation step in protein translation, an obligate step in the viral replication cycle, may provide a general antiviral strategy against fast-replicating RNA viruses.¹⁶⁵ Furthermore, **28** was also found exerts prophylactic effect against influenza A2 (H2N2) virus in mice through interferon-inducing activities.^{166,167}

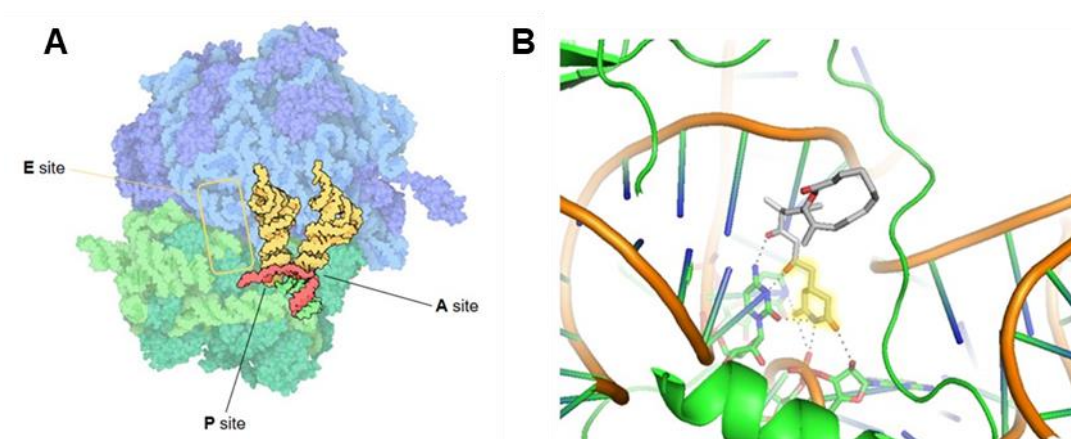


Figure 2.1.4 Glutarimide-containing compounds target E-site of the eukaryotic ribosome. (A) Crystal structure of 80S ribosomes from *Saccharomyces cerevisiae* where A and P binding sites are occupied with aminoacyl-tRNAs (PDB: 4U3N).¹⁶⁸ **(B)** Close-up of lactimidomycin bound to the tRNA CCA-end in the E-site of the large subunit (PDB file: 4U4U). Glutarimide moiety of lactimidomycin is highlighted in yellow.

2.2 Results and Discussion

2.2.1 Isolation and *de novo* structural elucidation of metabolic product derived from *gds* cryptic cluster

The cryptic *trans*-AT PKS BGC was originally identified in *B. gladioli* BCC0238, however, due to previously known genetic manipulation difficulties associated with this strain we continued our studies using a different CF isolate (*B. gladioli* BCC1622), which is more amenable to genetic modifications, and also contains the cluster of interest. Using standard culture conditions, *B. gladioli* BCC1622 was grown on a 2L scale on a minimal agar medium containing glycerol as carbon source for 3 days and its ethyl acetate extract was analysed using UHPLC-ESI-Q-TOF-MS. This analysis identified a metabolite with a base peak at $m/z = 408.2733$, corresponding to $[M+H]^+$ signal and a molecular formula of $C_{23}H_{37}NO_5$. The planar structure of this molecule **48** was elucidated by Dr. Yousef Dashti using 1H , ^{13}C , COSY, HSQC and HMBC NMR experiments (**Figure 2.2.1**). Although HMBC correlations between H-3 and C-1/C-1' as well as the exchangeable NH proton and C-2/C-2' confirmed the presence of a 2,6-piperidinedione, the expected butenolide functionality installed by the PBS domain was absent. Therefore, we postulated that this compound is a degradation product of the true metabolic product.

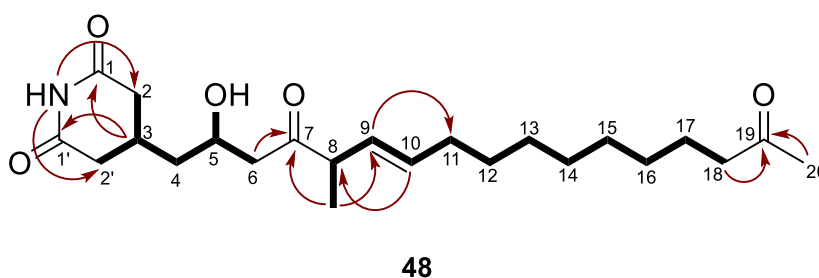


Figure 2.2.1 Glutarimide-containing compounds target E-site of the eukaryotic ribosome. (A) Crystal structure of 80S ribosomes from *Saccharomyces cerevisiae* where A and P binding sites are occupied with aminoacyl-tRNAs.¹⁶⁸ (B) Close-up of lactimidomycin bound to the tRNA CCA-end in the E-site of the large subunit. Glutarimide moiety of lactimidomycin is highlighted in yellow.

In an effort to capture the true metabolite of the cluster, different carbon sources were supplemented into the media and a time course analysis was conducted. UHPLC-ESI-Q-TOF-MS analysis of ethyl acetate extracts retrieved from a culture where glucose was used as the sole carbon source, showed an $[M+H]^+$ and $[M+Na]^+$ signals at $m/z = 506.2760$ (calc. 506.2748) and 528.2568 (calc. 528.2573) respectively, consistent with the neutral sum formula of $C_{27}H_{39}NO_8$ (**Figure 2.2.2**). In addition, based on a time-course analysis that was performed on extracts derived from media containing different carbon sources, it was found that the metabolite, which we named gladiostatin (**49**), is produced prior to the rest of the metabolites present in the metabolic profile of this strain (**Figure 2.2.2**). The metabolite production of this strain reaches its maximum at 20 hours and decreases rapidly over more extended growth periods with a gradual conversion to the degraded product **48** (**Figure 2.2.2**).

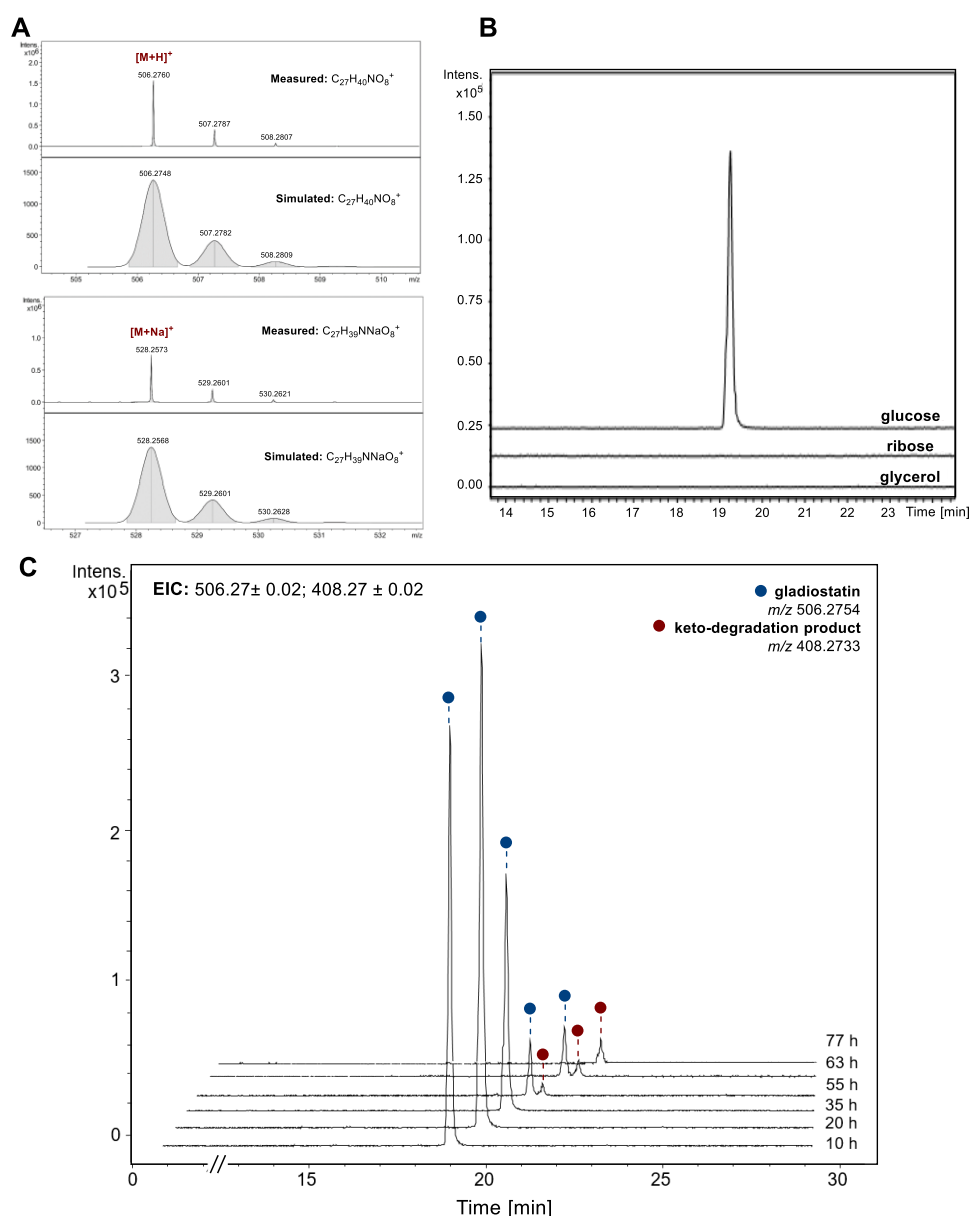


Figure 2.2.2 Identification of the true metabolic product of the cryptic *trans*-AT PKS in *B. gladioli* BCC1622. (A) High resolution mass spectrometry analysis of gladiostatin. Measured spectra for the $[M+H]^+$ species, $m/z = 506.2760$, and $[M+Na]^+$, $m/z = 528.2573$, species. Calculated molecular formulae are displayed and a simulated spectrum is displayed for comparison. (B) Extracted ion chromatograms at $m/z = 506.27 \pm 0.02$, (corresponding to the $[M+H]^+$ ion of gladiostatin) from UHPLC-ESI-Q-TOF-MS analysis of ethyl acetate extracts from *B. gladioli* BCC1622 grown on a minimal medium containing different carbon sources. (C) Time-course analysis of gladiostatin degradation. Extracted ion chromatograms from UHPLC-ESI-Q-TOF-MS analyses, monitoring gladiostatin (EIC: 506.27 ± 0.02 , blue circles) and keto-degradation product (EIC: 408.27 ± 0.02 , red circles). Time points for each chromatogram are denoted on the right.

Comparison of the molecular formula of this newly identified metabolite **49** to compound **48** showed that the former contains four additional carbon atoms, and three additional hydrogen and oxygen atoms.

To isolate this metabolite for characterisation, *B. gladioli* BCC1622 was grown on a large scale using the glucose-supplemented BSM agar for one day. The crude ethyl acetate extract was fractionated using preparative reverse-phase high-performance liquid chromatography (HPLC), to yield 8 mg/L of compound as a white solid. Structure elucidation of the molecule **49** was performed using a combination of ^1H , ^{13}C , COSY, HSQC and HMBC NMR experiments (**Figure 2.2.3**, **Table 7.10.1**). DEPT-135 analysis showed the presence of two methyl groups, six methines, twelve methylenes and seven quaternary carbons, from which five were assigned to carbonyls. HMBC correlations between H-3 and C-1/C-1' and the exchangeable NH proton as well as C-2/C-2' confirmed the presence of a 2,6-piperidinedione. These chemical shift values for this moiety were in good agreement with the literature.^{143,151,155,169} Two networks of COSY correlations established the structures of the C-2/C-2' to C-6 and C-8 to C-18 regions of the molecule, and HMBC analysis further confirmed association of the C-8 methyl group and the C-9/C-10 unsaturated bond. Furthermore, HMBC correlations displayed that C-6 and C-8 are linked via a keto group and that the methyl ketone is attached to C-18. Based on a $^3J_{\text{HH}}$ coupling constant of 15 Hz, the C-9/C-10 double bond was assigned to *E* configuration. Importantly, the methyl ketone signal was absent from the ^{13}C NMR spectrum of **49** and the spectrum contained additional signals assigned to a carbonyl group (C-21), two fully substituted alkene carbons (C-20 and C-2''), a methyl group (C-3'') and a hemiacetal (C-1''). HMBC correlations between the C-3'' protons and C-1''/C-20 and the C-1'' proton as well as C-2''/C-21 led us to propose the presence of a disubstituted 4-hydroxy-3-methylbutenolide functionality in the compound. The NMR data are in concordance to those reported in 2-alkyl-4-hydroxy-3-methylbutenolides (AHMBs) structures isolated from *S. rochei* and *S. ansachromogenes*.^{170,171} Additional HMBC correlation depicted the connectivity of this butenolide to the aliphatic region of the molecule. The existence of a 2-acyl-4-hydroxy-3-methyl-butenolide and the 2,6-piperidinedione in this molecule, coupled with the time course data for its production, led us to conclude that **49** is the true metabolic product of the cryptic-like BGC in the *B. gladioli* BCC1622 and BCC0238. The presence of compound **48** could result from possible conjugate addition of water to the

butenolide followed by ring opening. The intermediate **51** likely undergoes decarboxylation and retro-Aldol cleavage to yield **48** (Scheme 2.2.1).

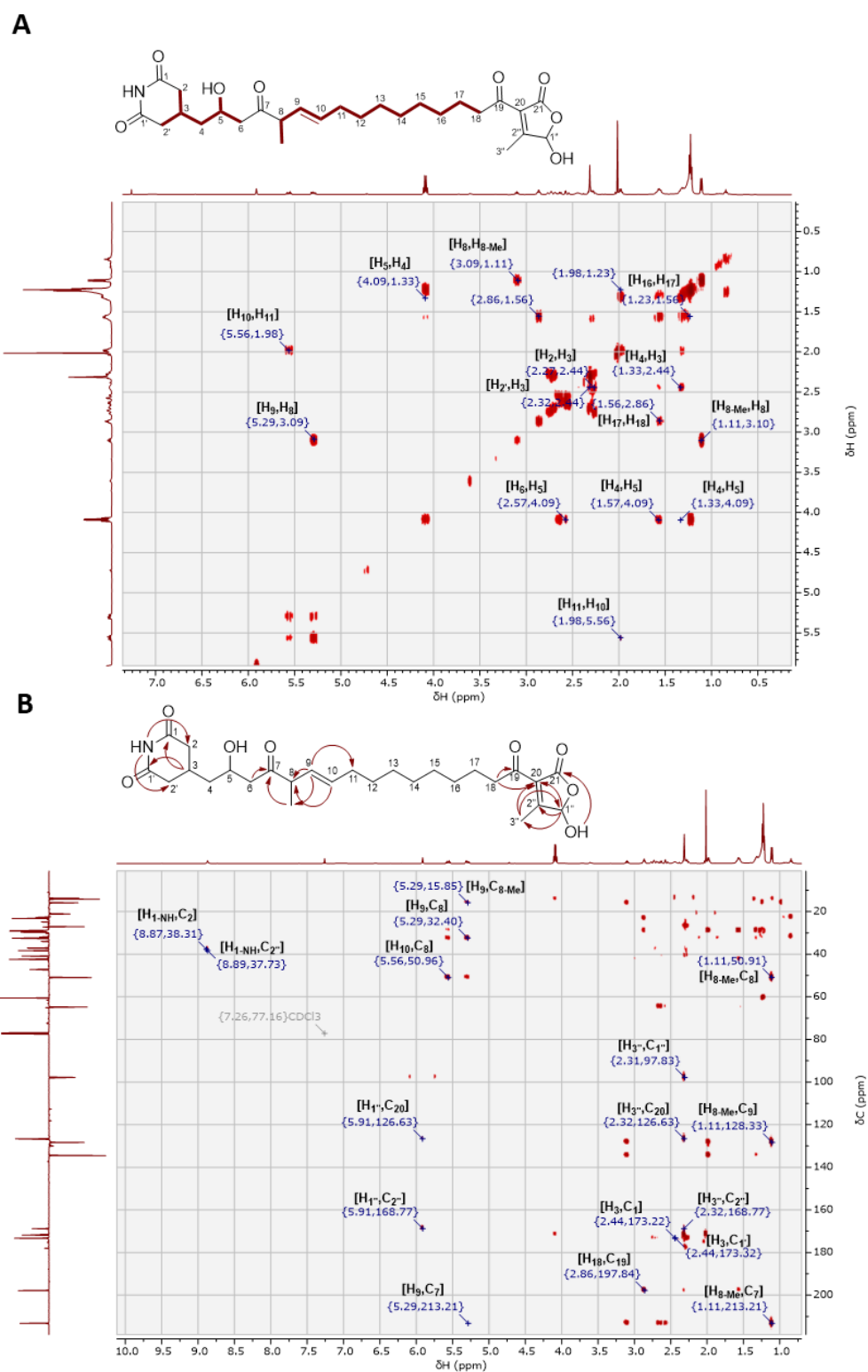
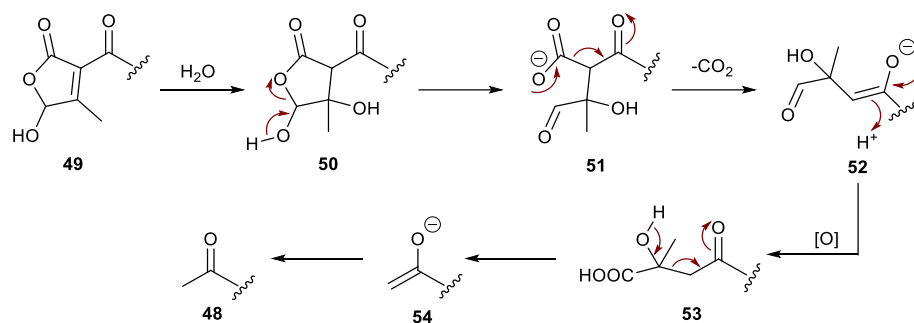


Figure 2.2.3 Structure elucidation of gladiostatin (49**).** Key COSY (**A**) and HMBC (**B**) correlations established the planar structure of the compound. Detailed NMR assignments can be found in Appendix.



Scheme 2.2.1 Proposed mechanism of gladiostatin conversion to the degradation product.

2.2.2 Biological activity of gladiostatin

2.2.2.1 Antimicrobial studies of gladiostatin against ESKAPE panel of pathogens, *Candida albicans* and *Saccharomyces cerevisiae*

As discussed in **Section 2.1.2**, previous studies on the family of glutarimide-containing polyketides have shown identified a plethora of biological activities. To verify whether **49** is active against a range of pathogens, we initiated our studies by performing broth microdilution susceptibility assays against the ESKAPE panel of pathogens. The ESKAPE panel consists of six pathogens with growing multidrug resistance (MDR) and virulence that are responsible for the majority of nosocomial infections worldwide. Recently, the World Health Organisation (WHO) has listed them in the global priority pathogens list of antibiotic-resistant bacteria to help in prioritizing the research and development (R&D) of new and effective antibiotic treatments.¹⁷² These include: *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter spp.* The activity of **49** against these pathogens was evaluated using Minimum Inhibitory Concentration (MIC) assays which determine the lowest concentration of an antimicrobial agent that prevents visible growth of a microorganism. The MIC assays were conducted according to CLSI microdilution guidelines.¹⁷³ The compound showed no activity against the ESKAPE panel of pathogens (**Table 2.2.1**). In addition, infections due to *Candida* species contribute greatly to morbidity and mortality of hospitalised patients and are related to a variety clinical manifestations. Gladiostatin (**49**) was also assessed for antifungal activity against *Candida albicans*, however no activity was observed against this fungal pathogen (**Table 2.2.1**). Prompted by reports which indicated

that lactimidomycin presented no activity against *C. albicans* but was active against other eukaryotes, we evaluated the bioactivity of **49** against a model yeast organism, *Saccharomyces cerevisiae*.¹⁵¹ Because of the similarity between *S. cerevisiae* and pathogenic fungi, compounds inhibiting the growth of the former are also potentially active against fungal pathogens.¹⁷⁴ Indeed, gladiostatin was found to be active against *S. cerevisiae* with an MIC value of 4 µg/ml.

Table 2.2.1 Antimicrobial activity of gladiostatin.

Strain	MIC (µg/mL)
Gram-negative Bacteria	
<i>Klebsiella pneumonia</i> DSM 26371	>64
<i>Acinetobacter baumannii</i> DSM25645	>64
<i>Pseudomonas aeruginosa</i> DSM29239	>64
<i>Enterobacter cloacae</i> DSM 16690	>64
Gram-positive Bacteria	
<i>Enterococcus faecium</i> DSM25390	>64
<i>Staphylococcus aureus</i> DSM21979	>64
Fungi	
<i>Candida albicans</i> SC 5314	>64
Yeast	
<i>S. cerevisiae</i> W303-1a	4

2.2.2.2 Cytotoxicity assays of gladiostatin against a range of cancer cell lines

An interesting observation that emerged from studies on the biological activity of the glutarimide-containing polyketides have been the reports of cell migration inhibitory activity of migrastatin and the extremely potent cytotoxicity of lactimidomycin. It is known that tumor cell migration is connected to tumor metastasis, and it is being considered the most important process dictating a tumor's ability to eventually kill its host.^{175,176} Encouraged by these findings in relation to glutarimide functionality, we elected to explore the activity of the compound against a range of cancer cell lines. Dr. Isolda Romero assisted in performing the cytotoxicity and wound healing assays to determine potential anticancer properties of gladiostatin against a range of cancer cell lines. Compound **49** was found to be active against ovarian, pancreatic and colon cancer cell lines (**Table 2.2.2**). These values are within a similar range as those reported for cycloheximide (**25**), migrastatin (**29**) and lactimidomycin (**34**) against various other cell lines (**Table A1 Appendix**).^{146,151,177} Interestingly, the secondary metabolite was found to be inactive against the A549 lung cancer cell line, indicating it may exhibit some selective cytotoxicity.

Table 2.2.2 Anticancer activity of 49.

gladiostatin (49)	
Cancer cell line	IC ₅₀ (μM)
A2780 (ovarian)	0.24 ± 0.03
A549 (lung)	no activity
MiaPaca2 (pancreatic)	0.57 ± 0.06
HCTT116 P53 -/- (colon)	0.82 ± 0.04
SKOV3 (ovarian)	0.6 ± 0.1
PEA1 (ovarian)	1.4 ± 0.1

After the activity of **49** against various cancer cell lines has been established, a wound-healing assay was conducted to test whether **49** can inhibit the migration of A2780 ovarian cancer cells.

Strong suppression of cell migration was observed after 24 h exposure to 240 nM **49** (Figure 2.2.4).

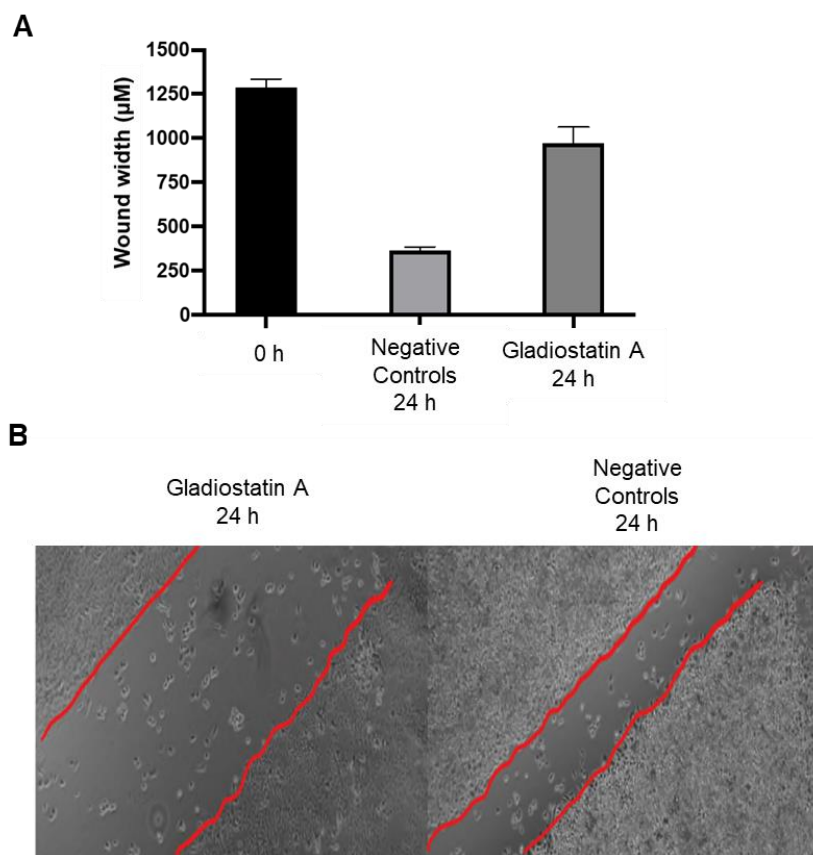


Figure 2.2.4 Inhibition of A2780 ovarian tumour cell migration by 49. (A) The bar chart compares wound-widths after 0 and 24 h in the absence of gladiostatin with the wound width after 24 h in the presence of 240 nM gladiostatin. (B) Microscope images showing wound-width after 24 hours of wound induction for untreated negative control cells and cells treated with 240 nM gladiostatin.

The above data outlined that gladiostatin can be developed as a potent anticancer agent. Additional structural studies of compound bound to the eukaryotic ribosome will reveal whether the mode of action is similar to this reported for lactimidomycin.¹⁶³

2.2.3 Studies on the biosynthesis of gladiostatin

2.2.3.1 Insertional inactivation of *gds* biosynthetic gene cluster

In order to confirm that the cryptic *trans*-AT PKS cluster directs the biosynthesis of gladiostatin (**49**), we performed an inactivation of *gdsE*, which encodes for the first PKS subunit. In particular, we targeted an internal fragment of *gdsE* which encodes module 1 B domain. This domain is proposed to provide structural role in the glutarimide-forming module, as discussed earlier in the Chapter.¹⁶⁰ The insertional inactivation was conducted using the suicide plasmid pGPΩTp (**Figure 2.2.5**)¹⁷⁸, with a 700 bp PCR-amplified internal fragment from the target gene cloned into the plasmid via EcoRI and XbaI restriction sites. The suicide plasmid carries a Pir-dependent origin of replication from the R6K replicon. In the absence of Pir, the plasmid cannot replicate and becomes integrated in the genome of *B. gladioli* BCC1622 by homologous recombination. The plasmid was integrated into *B. gladioli* BCC1622 wild-type via triparental conjugation using *E. coli* SY327 (pGPΩ*gdsE*_Bdomain) donor strain and *E. coli* HB101 (pRK2013) helper strain. The resulting transconjugants were selected for resistance to trimethoprim and polymyxin. Trimethoprim promotes the recombination event to proceed, since survival of the *B. gladioli* BCC1622 depends upon integration of the pGPI plasmid into its genomic DNA. Polymyxin provides the selection against both *E. coli* strains, enabling only the $Tp^R Pmx^R$ single crossover of the *B. gladioli* strain to survive. The single crossover of the mutagenic plasmid in *B. gladioli* BCC1622 was confirmed by colony PCR analysis and sequencing analysis (**Figure 2.2.5**).

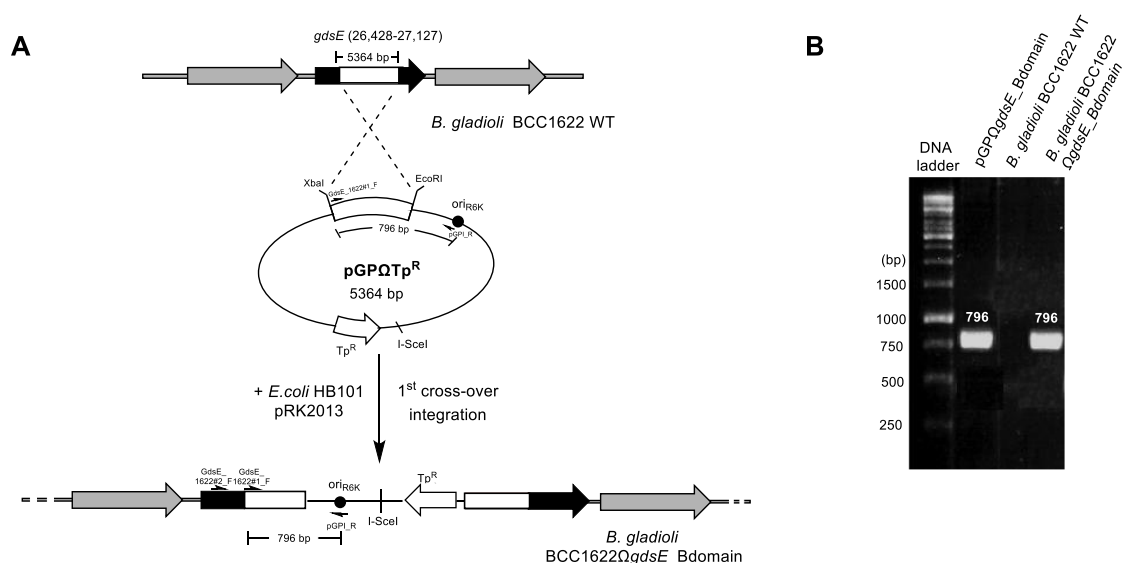


Figure 2.2.5 Generation of *B. gladioli* BCC1622 Ω *gdsE_Bdomain* using pGP Ω Tp^R. (A) Overview of construction of the Ω *gdsE_Bdomain* mutant. 700 bp DNA containing the target gene fragment is cloned into pGP Ω Tp^R vector and then integrated into the *B. gladioli* BCC1622 genome through tri-parental mating. Oligoprimers used for colony PCR and sequencing are shown in black arrows. The length of corresponding predicted PCR product is shown (796 bp). (B) Agarose gel electrophoresis of the PCR product provide genetic evidence for the successful insertional mutagenesis. The lanes correspond to the PCR product of pGP Ω *gdsE* construct (796 bp product) and *B. gladioli* BCC1622 Ω *gdsE* mutant (796 bp product). Wild-type allele *B. gladioli* BCC162 was used as a negative control.

To examine whether **49** is produced after gene inactivation, the metabolic profile of the *B. gladioli* BCC1622 Ω *gdsE_Bdomain* mutant was analysed by high-resolution LC-MS (**Figure 2.2.6**). Comparison of extracts between the wild type and the single cross over mutant showed that the [M+H]⁺ ion corresponding to **49** is only present in the wild type strain. This finding provides a direct link between the *gds trans*-AT PKS cluster identified from genome mining and the **49** metabolite (**Figure 2.2.6**).

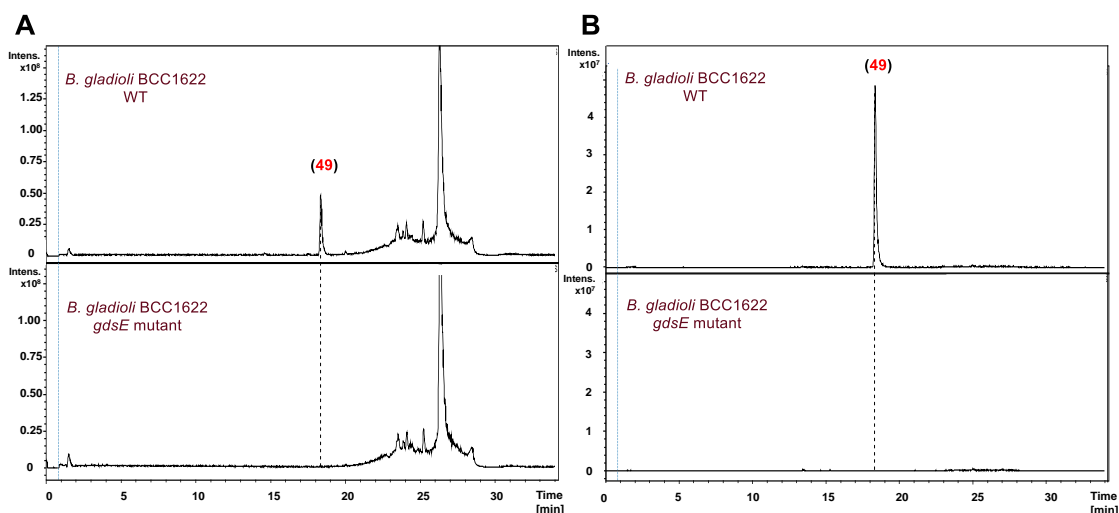
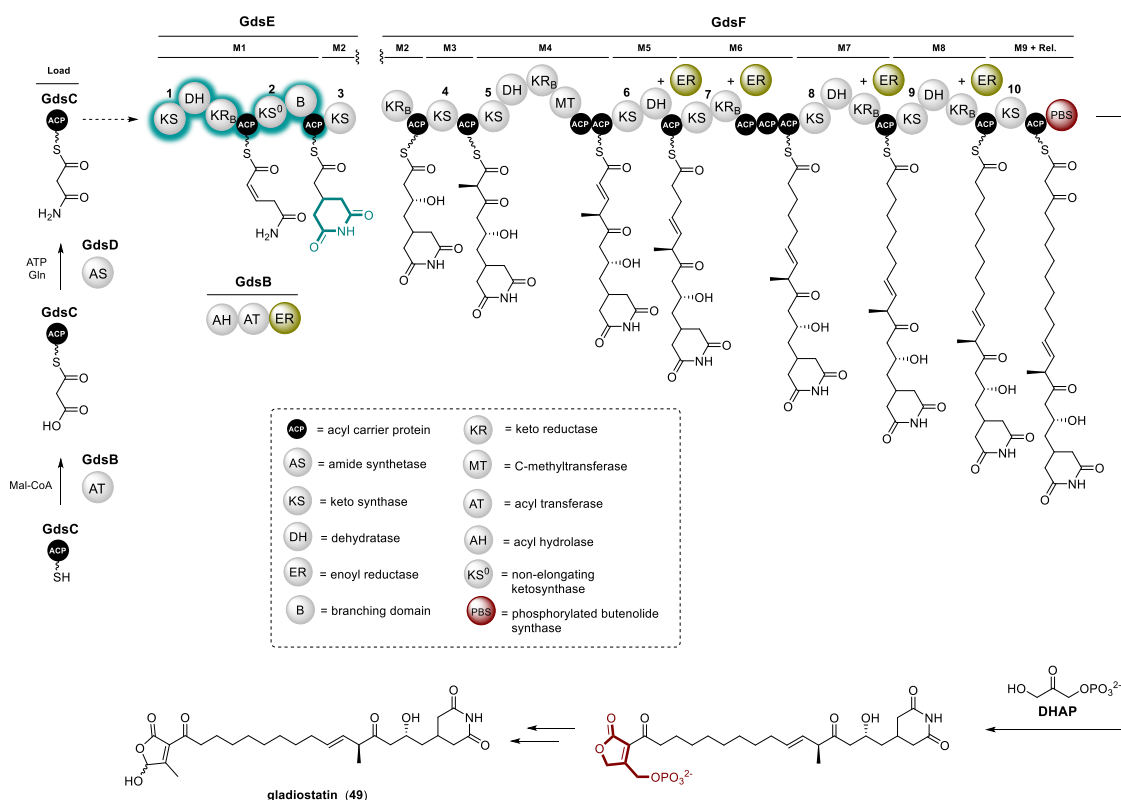


Figure 2.2.6 Insertional inactivation of *gdsE* abolishes gladiostatin (49) production in *B. gladioli* BCC1622. (A) Base peak chromatograms from LC-MS analyses of *B. gladioli* BCC1622 wild type and *gdsE* mutant extracts. **(B)** Extracted ion chromatograms at m/z 506.27 \pm 0.02 (corresponding to [M+H]⁺ for 49 from UHPLC-ESI-Q-TOF-MS analysis of ethyl acetate extracts of *B. gladioli* BCC1622 (top) and the *gdsE* mutant (bottom).

2.2.3.2 *In silico* analysis of the *gds* gene cluster and proposed assembly pathway for gladiostatin

After validating that the *gds* BGC directs the biosynthesis of 49, *in silico* analyses were employed in an effort to rationalise a plausible biosynthetic pathway for the *in vivo* formation of 49. Gene annotations using NCBI Conserved Domain search showed that the cluster harbours two large genes encoding *trans*-AT PKS subunits, designated *gdsE* and *gdsF*, which are flanked by few genes that encode putative polyketide processing enzymes. BLAST searches revealed the high degree of similarity between the first PKS subunit involved in the gladiostatin biosynthesis and the PKS subunits responsible for installing the glutarimide functionality in similar metabolites produced in *Streptomyces* sp.⁶¹ A detailed analysis of the catalytic domains within the PKS subunits allowed us to propose a plausible pathway for 49 biosynthesis, the start of which is identical to the assembly of glutarimides reported in *Streptomyces* species (detailed in **Section 2.1.2**). All KS domains in the PKS GdsE/GdsF subunits possess the catalytic triad Cys-His-His indicative for their chain elongation catalysis.^{62,179} All ACP domains harbor the the

canonical active-site Ser that is required for attachment of the phosphopantetheine moiety.^{62,180} KR domains feature the active site GxGxxG for NAD(P)H binding as well as the Ser-Lys-Tyr catalytic triad required.^{62,181} DH domains contain the catalytic Asp (HPALLD) which is proposed to donate a proton to the β -hydroxy group and the conserved HxxxGxxxxP motif containing the catalytic histidine which abstracts an α -proton to promote the *syn* elimination of water.^{88,89} Finally, phylogenetic analysis of GdsB-AT identified its specificity for malonyl-CoA. Based on *in silico* analysis and considering the elucidated chemical structure of **49**, we therefore postulated a biosynthesis scheme (**Scheme 2.2.2**).



Scheme 2.2.2 Proposed pathway for **49 biosynthesis.** Domain and module organization of the *gds* *trans*-AT PKS showing the proposed structure of each ACP-bound thioester intermediate. KS domains have been numbered sequentially, and the TransATor predictions for their acyl-ACP substrates are shown in the dashed box. The A / B subscripts denote the predicted stereospecificity of KR domains. The *trans*-acting ER domain (gold) of GdsB is predicted to reduce enoyl thioesters attached to the ACP domains in modules 5, 6, 7 and 8. Domain abbreviations are shown in the dashed box. The putative functions of the proteins encoded by the *gds* biosynthetic gene cluster are listed in appendix **Table A1**.

The *gdsB* gene encodes for an AT domain fused to an AH-ER di-domain which is proposed to orchestrate the loading of malonyl extender units *in trans* onto each of the ACP domains of the

assembly line. As described previously, three conserved genes (*ltmHCD*, *mgsHCD*, *chxBCD* and *smdFGH*) are proposed to generate the malonamyl thioester starter unit for the PKSs that assemble this class of metabolites (**Figure 2.1.3**). In a similar fashion, GdsD, an amide synthetase which exhibits close homology to LtmD/MgsD/ChxD/SmdH (**Figure 2.1.3**) is responsible for installing the amino group essential for the glutarimide moiety, resulting the GdsC-bound malonamyl thioester starter unit. This unit is then elaborated to a (2,6-dioxopiperidin-4-yl)acetyl thioester intermediate by the first PKS module. Modules 2, 3 and 4 of the *gds* PKS encoded by this cluster exhibit an identical domain architecture to the corresponding modules of the PKSs that assemble the *Streptomyces* glutarimides (**Figure 2.1.3**). Module 2 preserves a KS-KR-ACP domain architecture and is responsible for the installation of the hydroxyl group in C-5 of the molecule. Sequence comparisons with KR domains of known stereospecificity were performed to determine the type of KR domain and thus predict the configuration of the stereocenter bearing the hydroxyl group in C-5 (**Table 2.2.3**). KR_{M2} domain harbours a diagnostic aspartate motif (LRD) and thus is predicted to generate a hydroxyl group with (*R*) configuration (B-type KR domain).¹⁸²

Table 2.2.3 Predicted configurations of stereogenic carbon centres bearing hydroxyl groups in 49. Sequence alignments and comparisons with KR domains from *trans*-AT PKSs with known stereospecificities were performed to determine the type of KR domain.

KR domain	Diagnostic Asp Motif	KR type	Resulting stereochemistry
KR _{M1}	AGVL--Q <u>D</u> GLI	B type (<i>D</i>)	-
KR _{M2}	AGTL--R <u>D</u> GFA	B type (<i>D</i>)	(<i>R</i>)
KR _{M4}	AIVL--Q <u>D</u> RTL	B type (<i>D</i>)	-
KR _{M6}	AGFL--R <u>D</u> AYL	B type (<i>D</i>)	-
KR _{M7}	AGLL--R <u>D</u> AYL	B type (<i>D</i>)	-
KR _{M8}	AGLL--R <u>D</u> AYL	B type (<i>D</i>)	-

In comparison to the *Streptomyces* assembly lines, the MT domain in module 4 of the *gds* PKS appears to methylate only the β -keto thioester intermediate attached to the module 3 ACP domain (**Figure 2.1.3**, **Scheme 2.2.2**). The structures of known glutarimides, on the other hand,

suggest that MT domain in module 4 methylates the α -carbon of the β -keto-ACP thioester intermediate generated by the KS domain in both modules 3 and 4 of the PKS, installing methyl branches at C-8 and C-10 of the resulting glutarimides. Additional studies are needed to understand the different methylation pattern in the *gds* PKS that establishes a methyl branch at C-8 but not C-10 in the structure of **49**.

The rest of the domain architecture of the *gds* PKS diverges significantly from the corresponding modules in the *Streptomyces* glutarimide assembly lines. The C-terminus of module 4 appears to be the branch point in all glutarimide-producing PKSs (**Figure 2.1.3**), with the prospect to be a possible recombination hotspot. Comparative analysis of the domain architecture and **49** structure revealed that *gds* PKS is presenting a high degree of non-linear programming. In particular, module 5 is lacking a KR domain required for the installation of a β -hydroxy group and an ER domain which is required for the saturation of the double bond introduced by the DH domain present in the module. Similarly, module 6 lacks a DH domain and modules 6, 7 and 8 lack ER domains that are facilitate the production of α,β saturated thioester intermediates and consequently the aliphatic region in **49**. This hypothesis was also supported by an *in silico* analysis of KS substrate specificity conducted using the TransATor¹⁸³ software (**Table 2.2.4**). In particular, this analysis predicted that the KS domains in modules 7, 8 and 9 are selective for a fully reduced thioester intermediate at the α,β position. Although this data agree with our predictions based on the structure of the molecule, the software predicted that the module 6 KS domain prefers a α,β -unsaturated thioester as substrate (**Table 2.2.4**). While it is unclear which domains are responsible for the keto reduction in module 5 and dehydration in module 6, further studies on putative reductases present in the cluster revealed an unprecedented number of *trans*-acting enoyl reductase events within this cluster, as predicted (described in **Chapter 3**).

Table 2.2.4 Predicted specificities of KS domains from the *gds* PKS. All predictions were conducted using the TransATor webserver.¹⁸³

KS domain	Substrate specificity
GdsE_KS1	unusual starter (AMT/succinate)
GdsE_KS2	non-elongating (double bonds (mostly <i>Z</i> -configured)
GdsE_KS3	vinyllogous chain branching
GdsF_KS1	β D-OH (some with α L-Me)
GdsF_KS2	α Me reduced/keto/D-OH
GdsF_KS3	α Me double bonds
GdsF_KS4	double bonds (mostly <i>E</i> -configured)
GdsF_KS5	completely reduced
GdsF_KS6	completely reduced
GdsF_KS7	completely reduced

The most prominent variation between the *gds* PKS and the other glutarimide assembly lines is the off-loading mechanism of the polyketide chain. All of the *Streptomyces* glutarimide PKSs are proposed to use TE domains that either catalyse hydrolysis or macrolactonisation. In this case, the *gds* PKS employs a polyketide synthase domain (PBS) that catalyse condensation of the final β -keto thioester attached to the last module of the PKS with dihydroxyacetone phosphate (DHAP). The intermediate phosphorylated butenolide is proposed to undergo rearrangement and dephosphorylation to afford the 2-acyl-4-hydroxy-3-methylbutenolide (**Scheme 2.2.2**). This hypothesis is in agreement with the increase in **49** production levels observed when glucose is added as the sole carbon source, since DHAP is an intermediate in glycolysis. This mechanism of off-loading is found for the first time being appended in glutarimide-producing PKSs. Studies aim to underpin the mechanism of release are described on **Chapter 4**.

2.3 Conclusions

Following previous studies conducted in our group, this work focused on the discovery of a novel glutarimide-containing polyketide, gladiostatin (**49**). The planar structure of the new metabolite was elucidated using a combination of ^1H , ^{13}C , COSY, HSQC and HMBC NMR experiments. The product was successfully connected to its cryptic *trans*-AT PKS assembly line in *B. gladioli* BCC0238 and BCC1622 through insertional mutagenesis on the gene encoding the B domain, a structurally essential protein for the installation of the glutarimide core. We proposed a biosynthetic pathway for the assembly of this new glutarimide-containing polyketide which is in line with previously reported pathways concerning the construction of the 2,6-piperidinedione moiety. However, from this point onward **49** biosynthesis diverges significantly from other members of this family, all of which are produced *Streptomyces* species, making **49** the first glutarimide isolated from Gram-negative bacteria. *In silico* analysis of the BGC in conjunction with structural characterisation of the metabolite showed structural differences (such as the fully saturated C-11 to C-18 chain and the 2-acyl-4-hydroxy-3-methylbutenolide) translated to alternations in PKS architecture. Other differences such as the lack of C-10 methyl group does not directly correlate to differences between the PKS architecture. The discovery of gladiostatin (**49**) offers a great opportunity to explore further the horizontal gene transfer in *trans*-AT PKS evolution, which could assist biosynthetic engineering approaches to polyketide structural diversification. In contrast to most PKS assembly lines, an AfsA-like PBS domain is appended to the C-terminus of the *gds* and was proposed to play a role of releasing the 3-keto thioester from the upstream ACP domain by forming a butenolide. Our *in silico* predictions were in concordance with the structure of the molecule. This constitutes a new mechanism for polyketide chain release, which could be a valuable addition to the synthetic biology toolbox. Further studies on this topic are presented in **Chapter 4**.

Gladiostatin (**49**) is of interest owing to its cytotoxic bioactivity against several cancer cell lines, which is a characteristic of glutarimide-containing polyketides. Interestingly, the secondary metabolite presented no activity against the A549 lung cancer cell line, indicating it may exhibit some selectivity. The potent cytotoxic activity makes it a candidate for further investigations with regards to its mode of action and development as an antitumor agent. Future efforts will

be focused on obtaining structural data of this compound in complex with the *S. cerevisiae* 80S ribosome in order to verify whether (**49**) is a ribosome inhibitor, following a similar mode of action to this observed for cycloheximide and lactimidomycin.¹⁶³ Since the biosynthetic logic of the molecule has been elucidated, once the target will be identified, rational engineering can be employed in order to construct a library of (**49**) analogues that could serve as leads in structure-activity relationship (SAR) studies. In parallel, its ability to inhibit cell migration could be further investigated aiming its potent use as anti-metastatic agent.

Chapter 3

*Characterisation of trans-acting enoyl reduction steps in
gladiostatin biosynthesis*

3.1 Introduction

As discussed in **section 1.5.3**, *trans*-AT PKSs utilise standalone domains to facilitate modification of the nascent polyketide chain during biosynthesis. Comparisons between the domain architecture and the predicted biosynthetic intermediates showed that module 5 of the gladiostatin PKS lacks both a KR domain, which is responsible for installing the β -hydroxy group, and an ER domain, which is required to saturate the double bond introduced by the DH domain. In a similar manner, in module 6, a DH domain that is required to catalyse the formation of an α,β -double bond in the nascent polyketide intermediate is not present. Furthermore, modules 6, 7 and 8 lack integrated ER domains required to generate the saturated regions from C11-C18. This suggests a high degree of non-linear programming in the gladiostatin PKS, which is a characteristic feature of *trans*-AT PKS assembly lines (**Figure 3.1.1**).⁶¹

Although it is still unclear which domains catalyse the keto reduction and dehydration reactions in modules 5 and 6, *in silico* analysis of the gladiostatin BGC using protein BLAST (appendix **Table A1**) revealed two genes encoding putative reductases that could function as *trans*-acting ERs in modules 5-8 (**Figure 3.1.2**). The two reductase candidates were: i). a tri-domain protein with an AH and AT fused to an FMN-dependent ER domain which is encoded by *gdsB* (AH-AT-ER), and an NAD(P)H-dependent oxidoreductase which is encoded by *gdsH*.

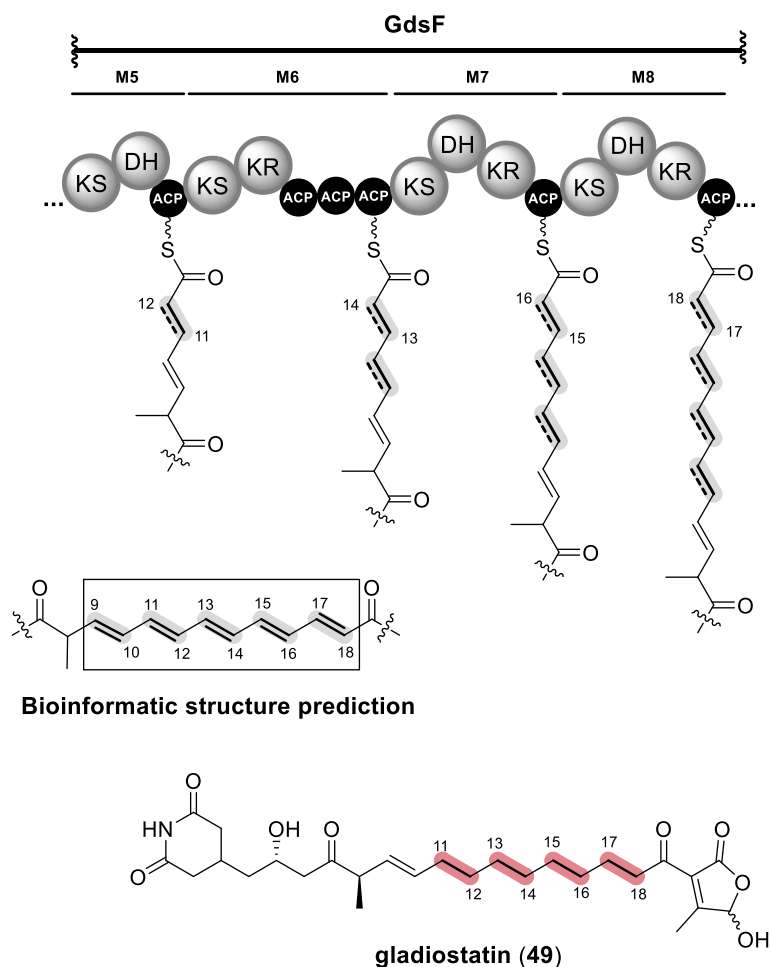


Figure 3.1.1 Partial domain architecture in the gladiostatin *trans*-AT PKS that diverges from PKS collinearity rules. The predicted structure of the product derived from *gds* PKS assembly line is shown in grey and compared to the true metabolite which presents saturated centers in C-11 to C-18 (shown in red).

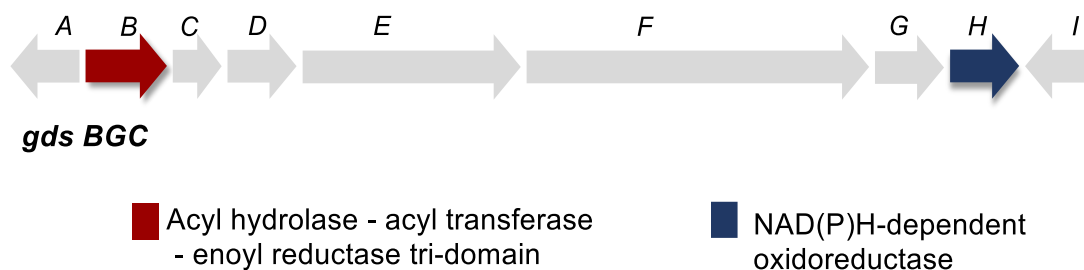


Figure 3.1.2 Genetic organisation of the gladiostatin biosynthetic gene cluster (*gdsA-I*). The two candidates for the enoyl reduction required in modules 5 to 9 are shown in red (*gdsB*) and blue (*gdsH*).

Trans-acting ER domains are common feature of the *trans*-AT PKS systems. Both types of phylogenetically distinct ER domains are present in *trans*-AT PKS, though NAD(P)H-dependent ER domains are normally integrated into the PKS modules and functions in *cis*. In particular, *trans*-acting ER domains clade with standalone ER domains (PfaD) found in polyunsaturated fatty acid synthases and are phylogenetically close to non-canonical bacterial FAS ER domains such as FabK, fungal FAS ER (fFAS ER) and Corynebacteria, Mycobacteria and Nocardia (CMN)-FAS ER domains (**Figure 3.1.3**).¹²³ A few *trans*-acting ER domains are fused to AT or AH domains,^{122,123} as observed for the GdsB-ER. This type of ER domains catalyse enoyl reduction using FMN to mediate hydride transferring from NAD(P)H to substrate (**Figure 3.1.3**). Functional characterisation of PksE ER domain from the dihydrobacillaene biosynthesis showed specific interactions with the associated ACP.¹²² A yeast two-hybrid interaction analysis screen on the kalimantacin assembly line revealed specific interactions between ERs and their cognate ACP¹⁸⁴, however a detailed catalytic mechanism utilised by *trans*-acting ERs as well as the molecular basis underlying their specific interactions with cognate ACPs still remain elusive. Efforts towards this direction were pursued using an overlay of the X-ray crystal structure of the excised ER from AT-ER didonain DifA with FabK.¹²³ Structural analysis shown an FMN bound TIM barrel domain and an α -helical substrate binding domain (**Figure 3.1.3**). When compared to the bacterial fatty acid ER FabK, the DifA ER was characterised by two *trans*-AT PKS subdomains and an interdomain that was proposed to act as interaction site.

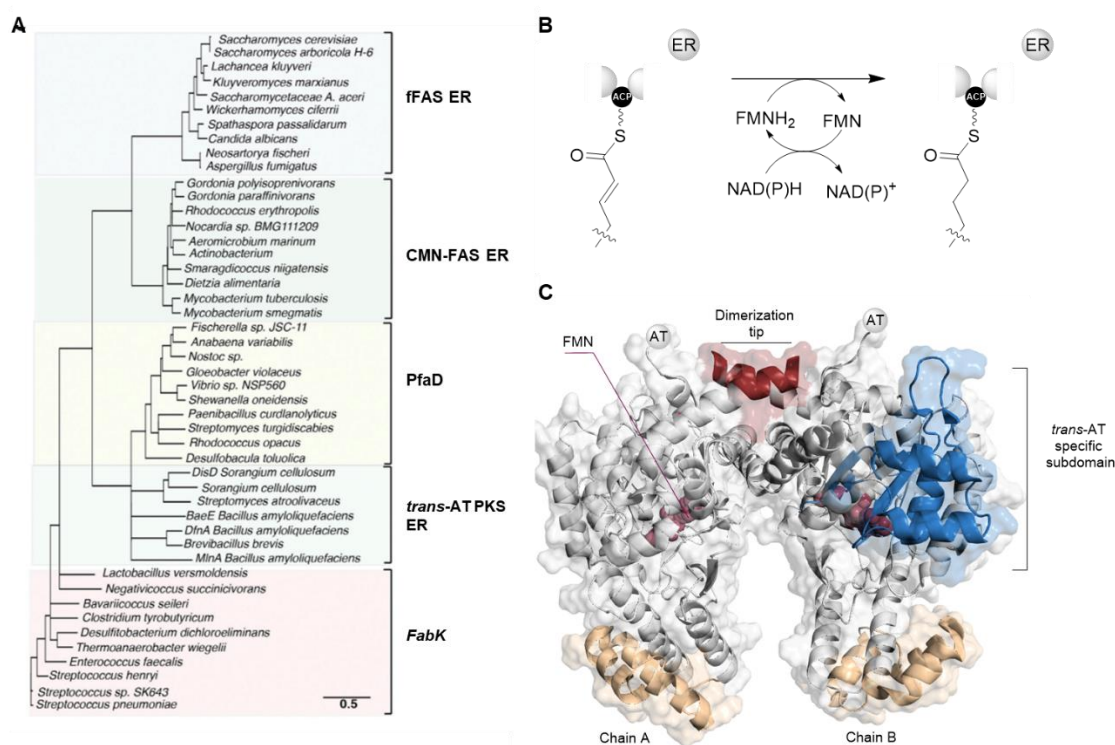


Figure 3.1.3 Overview of phylogenetic analysis, function and structure of *trans*-acting ER domains from type I modular PKSs. (A) Distribution of modular PKSs *trans*-acting ER among bacteria and fungi (figure reproduced from ref.¹²³). (B) Enoyl reduction catalysed by *trans*-acting ERs using FMN to mediate transfer from NAD(P)H into the substrate. (C) Cartoon representation of homodimer DifA-ER bound to FMN (PDB file: 4CW5). Additional loops and helix insertions proposed to serve as interdomain interaction interface are shown in blue and orange.

Herein we propose that one, or both, of these reductases catalyse the enoyl reduction in modules 5, 6, 7 and 8 of the gladiostatin PKS, resulting the saturated centres between C-11 to C-18. This study exemplifies an unprecedented number of specific interactions between ACP domains and a *trans*-acting ER. This was investigated using a combination of *in vitro* biochemical assays in conjunction with *in vivo* genetic engineering experiments and *in silico* analysis.

3.2 Results and Discussion

3.2.1 Overproduction and purification of putative proteins involved in enoyl reduction events during gladiostatin biosynthesis

As discussed above, based on the final structure of **49**, enoyl reduction events are proposed to occur in Modules 5, 6, 7 and 8 via a *trans*-acting ER domain. To examine this hypothesis, ACP domains from GdsF modules 5, 6, 7 and 8, as well as the two genes encoding putative reductases that could function as *trans*-acting ERs, *gdsB* and *gdsH*, were amplified by PCR and cloned into a pET28a overexpression vector, allowing heterologous production in *E. coli* BL-21 (DE3) as hexa-histidine tagged proteins, which could be purified to homogeneity using Ni-NTA affinity chromatography. Post-purification, the proteins were analysed by SDS-PAGE and the expected molecular weight was recorded by intact protein MS (**Figure 3.2.1**). Intriguingly, module 6 of the GdsF PKS contain three consecutive ACP domains. ACP tri-domains have been observed in a number of biosynthetic pathways including onnamide¹⁸⁵, etnangien¹²⁰, mupirocin¹⁸⁶ and kalimanticin¹⁸⁷. Modules that recruit *trans*-acting ER domains often encompass two or three ACP domains apparently because enoyl reduction is much faster than the subsequent chain elongation reaction.⁶¹ Notably, the purified GdsB exhibited a strong yellow colouration, suggesting that the protein was purified with the flavin mononucleotide (FMN) co-factor bound to it, which was expected.

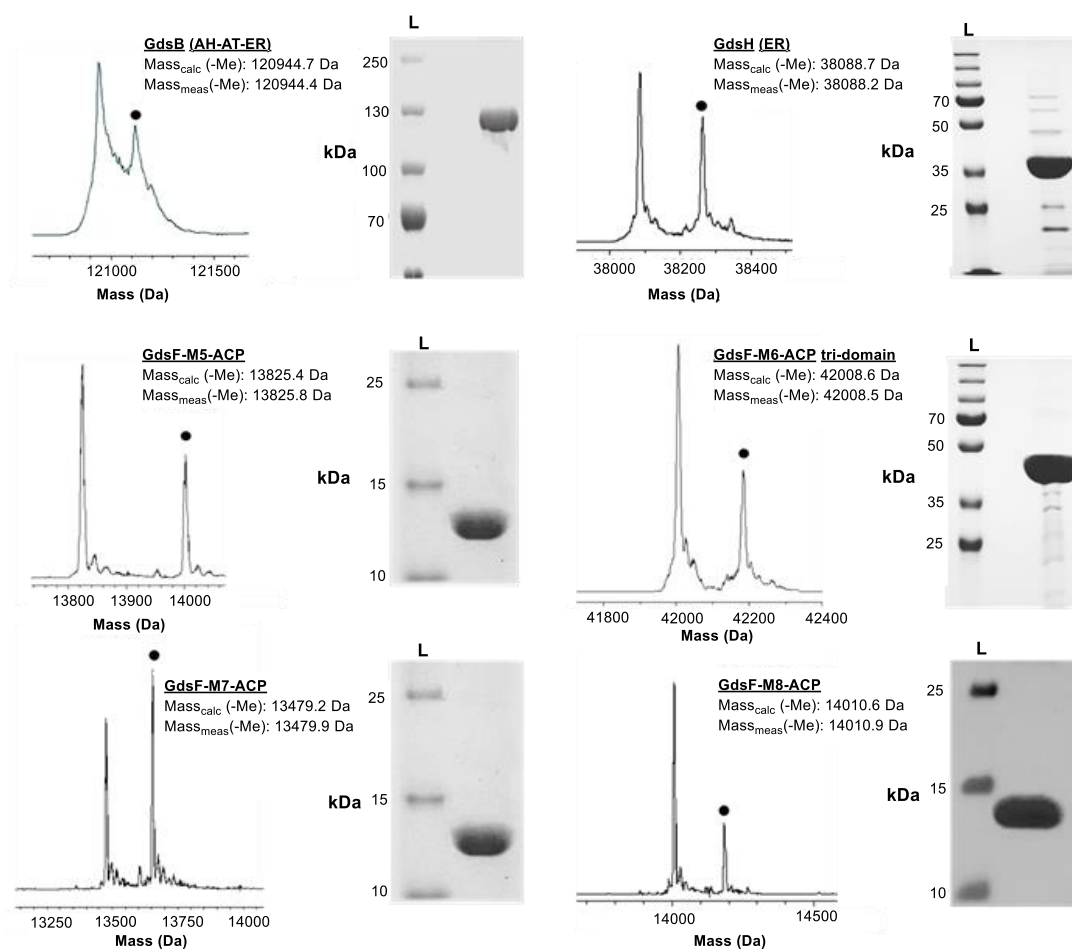


Figure 3.2.1 Analysis of recombinant proteins used in this study by SDS-PAGE and intact MS. SDS-PAGE gel indicate the purity and estimated molecular weight of the purified proteins, and deconvoluted ESI-MS spectra show the mass of each protein used for this study. Peaks annotated with a black circle indicate N-terminal gluconoylation of the His-tag with a mass shift of + 178 Da.¹⁸⁸ Raw MS data can be found in the Appendix.

To accurately test the specificity of the *trans*-acting enoyl reductase candidate towards the designated ACPs, a negative control was produced. The ACP di-domain from module 4 was selected, since these ACP domains are predicted not to interact with a *trans*-acting ER based on the proposed biosynthetic pathway and the structure of gladiostatatin. Intact MS analysis of the protein indicated a peak with a mass increase of + 680 Da to the theoretical mass (**Figure 3.2.2**). This intact M_r value of the ACP di-domain indicated that the ACP domains were purified in the *holo* form, with this mass increase consistent with the protein being phosphopantetheinylated (+ 340 Da, $2 \times 340 = 680$ Da) during overexpression in *E. coli*. This

is inconvenient, since the Ser residue of ACP domains will be phosphopantetheinylated and *in vitro* loading of desired CoA substrates cannot occur. The holo-form is likely to be observed because of the phosphopantetheinyl transferase EntD, which is known to phosphopantetheinylate ACP/PCP domains when they are expressed in *E. coli*.^{189,190} To overcome this issue, the pET28a_ACP-di-domain construct was cloned into a mutant *E. coli* BL21 (DE3) lacking the *entD* from its chromosome (cells kindly provided by Dr. Thierry Izoré). Expression of the GdsF module 4 ACP di-domain pET28a construct in the *entD* mutant showed no increase in mass, demonstrating that the di-domain is expressed in its *apo*-form, which could be further utilised to load selected acyl-CoA substrates *in vitro* to investigate the enoyl reduction chronicles in gladiostatin pathway (**Figure 3.2.2**).

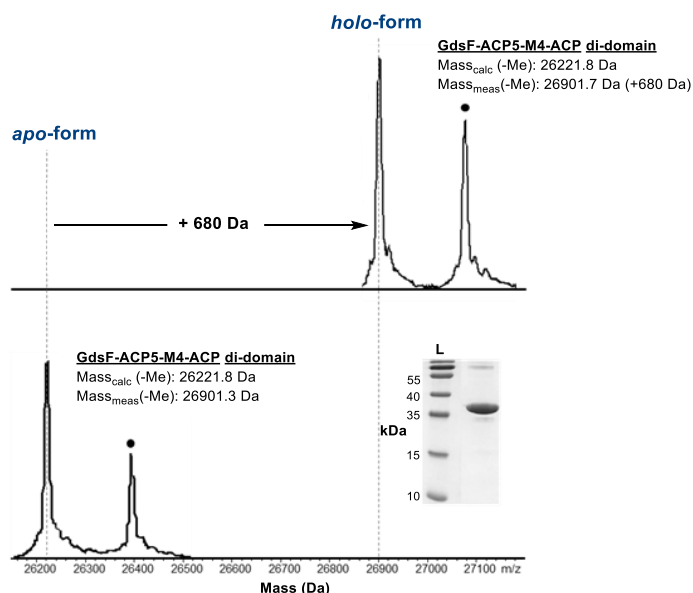
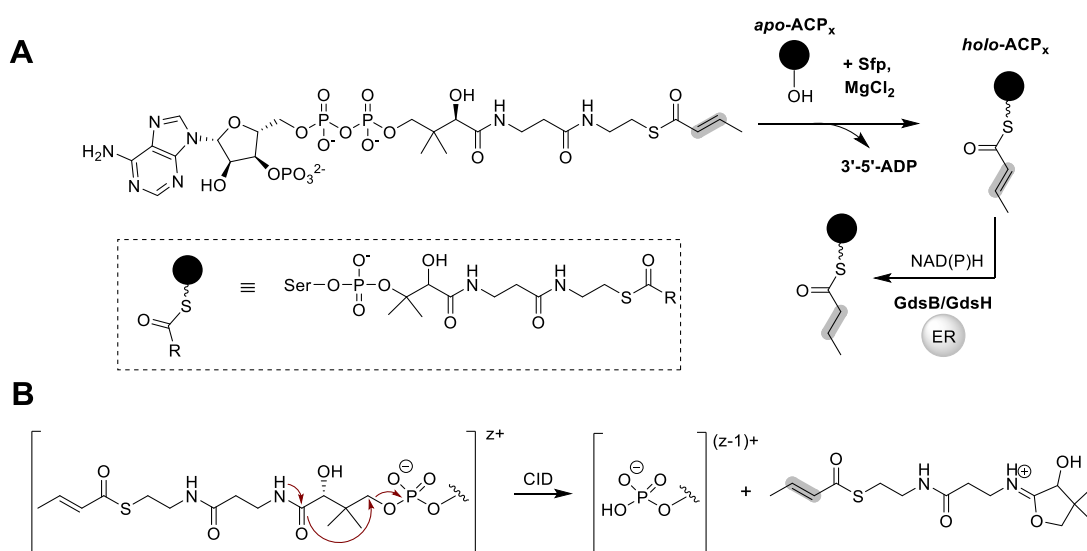


Figure 3.2.2 ESI-MS analysis and SDS PAGE of GdsF-Mod4-ACP di-domain before and after expression in *E. coli* BL21Δ*entD* mutant. Stacked deconvoluted ESI-MS spectra of N-His₆-GdsF-Mod4-ACP di-domain purified in its *holo* (top) and *apo*-form (bottom) and 10% SDS-PAGE gel showing the GdsF-Mod4-ACP di-domain used for this study.

3.2.2 *In vitro* biochemical characterisation of putative *trans*-acting enoyl reductases in GdsF modules 5-8

The ability of the putative reductases, GdsB and GdsH, to interact with the designated ACPs and reduce their α,β -unsaturated thioester intermediates was investigated using an *in vitro* biochemical assay, as described in **section 7.8.2 (Scheme 3.2.1)**. A commercially available 2-butenoyl-CoA thioester was utilised as substrate mimic, containing a double bond in the α,β -position. The 2-butenoyl-CoA was used together with a non-specific 4'-phosphopantetheinyl transferase, Sfp, to load the 2-butenoyl-phosphopantetheine arm onto the *apo*-ACP domains, yielding 2-butenoyl-*holo*-ACP domains that could be used to probe enoyl reductase activity.^{191–}

193



Scheme 3.2.1 Overview of the MS-based assay used to probe the putative function GdsB and GdsH as *trans*-acting enoyl reductases. (A) ACP domains are first converted to their *holo* forms by a phosphopantetheinylation reaction with 2-butenoyl-CoA that attaches a 2-butenoyl-phosphopantetheine unit onto the conserved Ser residue of the domain. (B) Proposed mechanism of phosphopantetheine elimination reaction. During collisional activation, the phosphate anion remains attached to the Ser residue of the ACP and the amide carbonyl displaces the pantetheinylate moiety forming a 5-membered ring with a protonated imine.

The successful conversion of the *apo*-ACP to its *holo*-form after attachment of the 2-butenoyl-PPant chain was monitored by intact protein MS (**Figure 3.2.3**). The conversion of each *apo*-ACP to its *holo* form after attachment of the 2-butenoyl-PPant arm was accompanied by a + 409 Da mass shift (**Figure 3.2.3**).

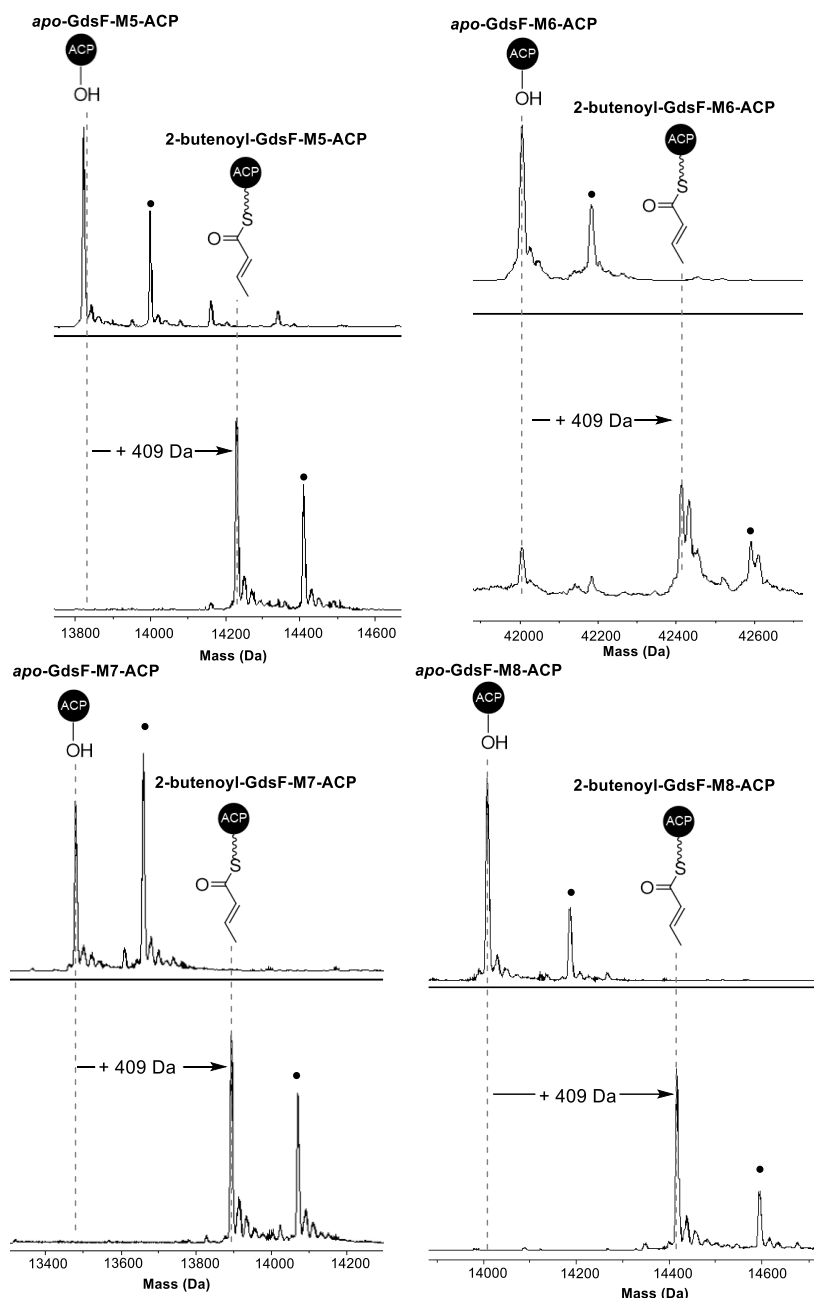


Figure 3.2.3 ESI-MS analysis of GdsF ACP domains from modules 5-8 after phosphopantetheinylation. Deconvoluted intact protein MS showing the successful attachment of the 2-butenoyl-PPant arm onto the ACP domains (+ 409 Da mass shift).

The enoyl reduction activity was then investigated by incubating the loading reaction with either GdsB or GdsH in 1:1 ratio. The cofactors incorporated in the reaction were selected based on the bioinformatics predictions for each reductase. For GdsB, the reduction of the loaded 2-butenoyl unit was proposed to occur after addition of a hydride donating cofactor for FMN. For this reaction, both NADPH and NADH were screened. The colour of the GdsB indicated the copurification of the protein bound to FMN, therefore no additional co-factor was added to the reaction. GdsH was screened using either NADPH or NADH. A control reaction was performed by adding the ER candidate to the loaded-ACP domains without the addition of the cofactor. A plausible enoyl reduction will result a +2 Da mass shift which is difficult to detect through intact MS, and therefore it was monitored through generated PPant ejection ions.¹⁹⁴ During collisional activation of the loaded-ACP domains, the pantetheine moiety is released as a protonated imine, keeping the acyl thioester intact. This allowed us to monitor the reaction on a small molecule level of m/z either 329.15, corresponding to the 2-butenoyl-Pant ejected ion or 331.16 for the butyryl-Pant ejected ion.

Purified GdsH was the first candidate to be investigated for its function as a putative enoyl reductase. The protein was incubated with each ACP, loaded with the 2-butenoyl substrate on a 1:1 ratio and NADPH/NADH as a cofactor. A control reaction was performed in parallel, lacking the required cofactors. The reaction before and after addition of GdsH domain was monitored using the PPant ejection procedure. Only fragment ions peaks at m/z 329 were observed, corresponding to the 2-butenoyl-PPant were observed after incubation of the ACP domains with GdsH, indicating that the double bond of the substrate has not been reduced (**Figure 3.2.4**).

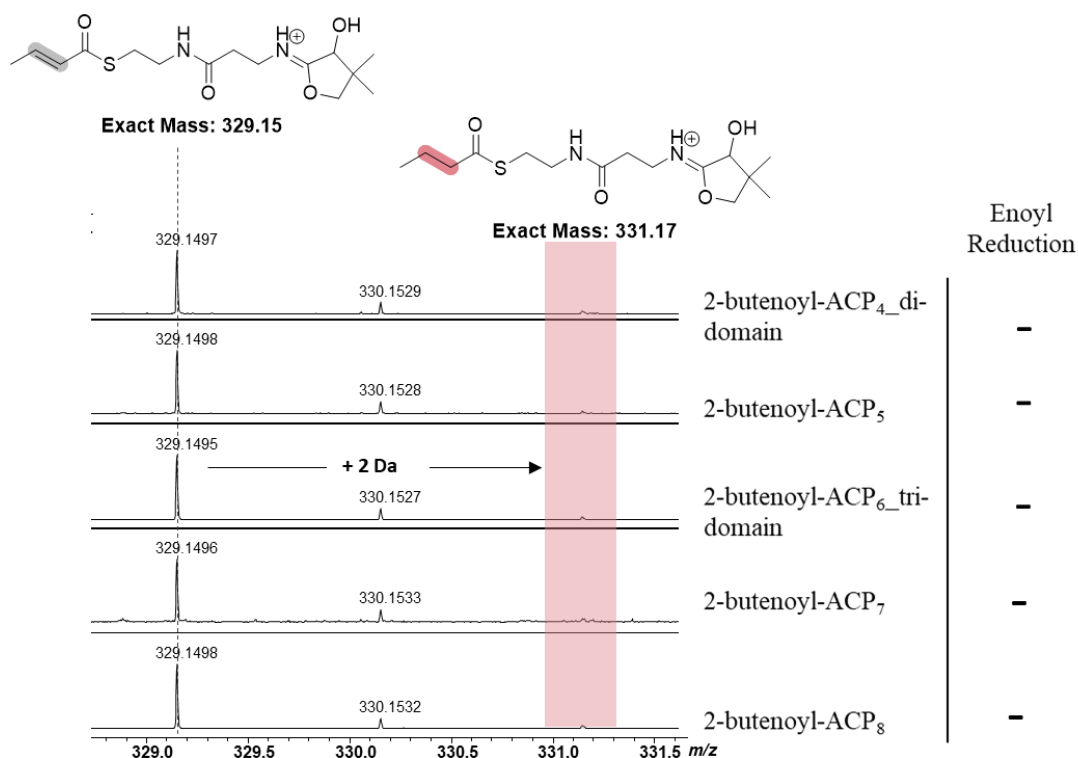


Figure 3.2.4 Stacked MS spectra of PPant ejected ions generated from 2-butenoyl-ACP domains after incubation with GdsH, in the presence of NADPH. Observed enoyl reduction is accompanied with a peak of m/z 331 corresponding to generated PPant ejected ions.

Increasing the reductase:ACP ratio to 2:1 and leaving the reaction for longer incubation periods did not promote enoyl reduction of the substrate. It was therefore speculated that GdsH is either not responsible for catalysing this type of reaction or it is purified in an inactive form.

The purified GdsB tri-domain was then screened for its ability to catalyse the enoyl reduction steps in modules 5 to 8. Using the same conditions, the protein was incubated with each ACP, loaded with the 2-butenoyl substrate on a 1:1 ratio and NADPH/NADH as a cofactor. A reaction lacking the required cofactors was run in parallel as a control. Interestingly, following incubation of the ACP domains with GdsB, ESI-MS analysis revealed a – 68 Da mass shift, suggesting that the acyl-ACP domains had undergone thioester hydrolysis, yielding the *holo*-ACP signal (**Figure 3.2.5**). This was further confirmed by observation of the characteristic peak of m/z 261.1 of the relevant fragment ion.

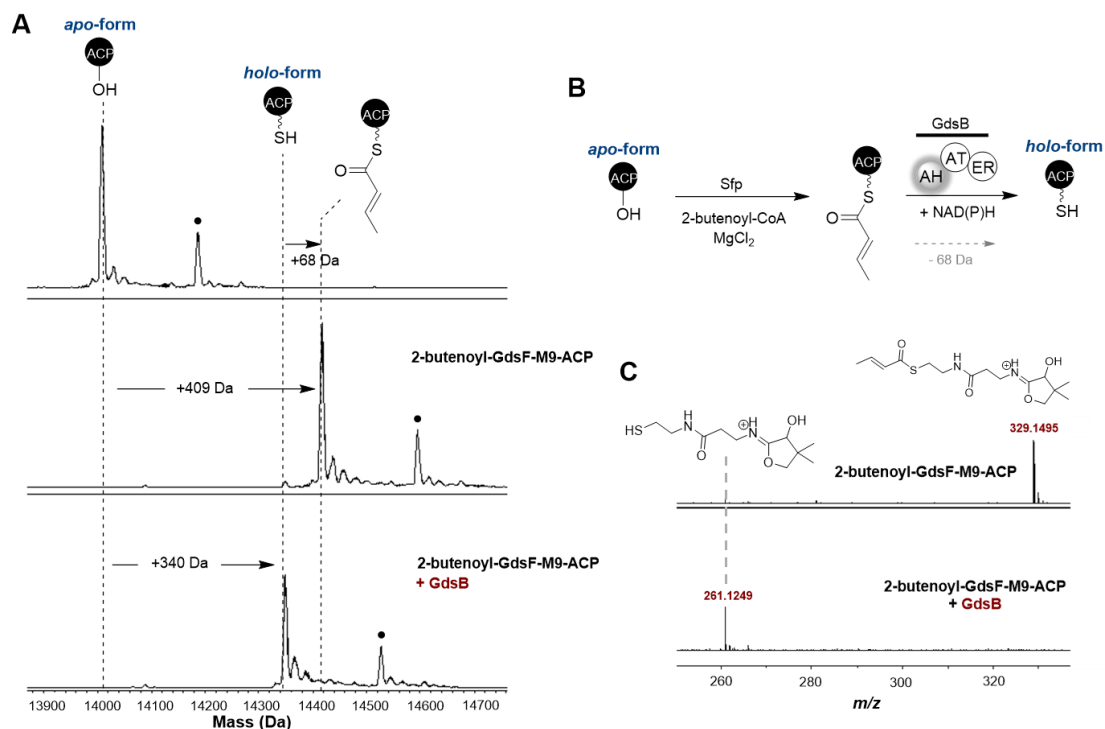


Figure 3.2.5 Representative example of ESI-MS analysis of the *in vitro* assay, highlighting hydrolysis of the 2-butenoyl-PPant attached to the ACP of module 9 after incubation with the GdsB AH-AT-ER tri-domain. **(A)** Deconvoluted MS spectra of GdsF-M9-ACP in its *apo*-form (top), loaded with the 2-butenoyl substrate (middle) and *holo*-form after addition of GdsB. **(B)** Overview of the steps involved in the conversion of 2-butenoyl-GdsF-M9-ACP to the *holo*-form. **(C)** Incubation of 2-butenoyl-GdsF-M9-ACP with GdsB AH-AT-ER tri-domain shows the detection of the PPant *m/z* 261.2 species.

This outcome suggested that the acyl hydrolase (AH) domain appended in the tri-domain proved to have a hydrolytic activity against the 2-butenoyl acyl group attached to the PPant arm of the ACP domains tested, interfering with the biochemical assay (**Figure 3.2.5**).^{75,121} AH domains are recruited during polyketide biosynthesis for its housekeeping role and studies have shown a strong preference of the enzyme towards hydrolysis of short acyl chains.^{75,121} Since previous studies have shown that excised ER domains that are fused to AT domains still retain their catalytic activity¹²², we performed a deletion of the AH-AT domains on the pET28a_GdsB construct, using site-directed mutagenesis to generate the GdsB-excised ER (detailed in **section 7.6**). A mutant construct was generated and overproduced and purified using the same protocol as described previously (**Figure 3.9**).

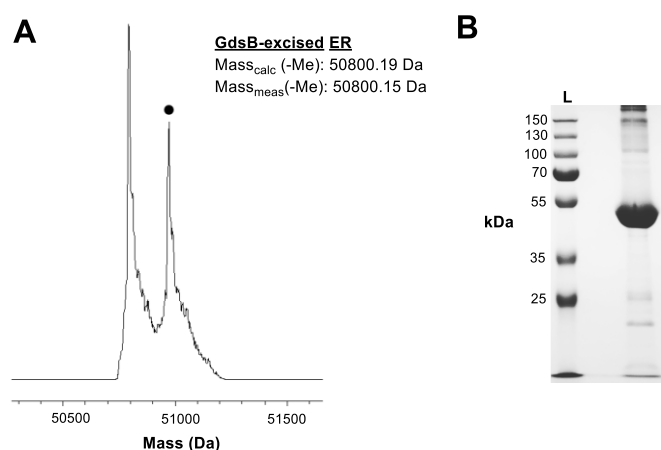


Figure 3.2.6 ESI-MS analysis and SDS-PAGE of purified N-His₆-GdsB-excised ER. (A) Deconvoluted ESI-MS spectra of GdsB-excised ER. Peak annotated with a black circle indicate N-terminal gluconoylation of the His-tag with a mass shift of + 178 Da.¹⁸⁸ Raw MS can be found in the Appendix. (B). 10% SDS-PAGE gel of purified N-His₆-GdsB-excised ER.

The biochemical assays were repeated using the GdsB-excised ER in 1:1 ratio to ACP domains from modules 4 to 8. ESI-MS analysis of the reactions showed that GdsB-excised ER domain successfully reduced α,β unsaturated substrate appended in ACP domains from modules 5 to 8, generating a fragment ion of m/z 331.16 that corresponds to the butyryl PPant ejected ion (**Figure 3.2.7**). In addition, GdsB-excised ER did not interact with the ACP di-domain from module 4. This is in agreement with our biosynthetic proposal (**Figure 3.2.7**). NADPH proved to be the true cofactor for the catalysis of this reaction, since addition of the NADH only produced low-level conversions. These initial findings demonstrated that the ER domain of GdsB is the sole enzyme responsible for reducing the α,β -unsaturated substrates bound to ACPs for modules 5, 6, 7, 8 and 9, ruling out the ability of GdsH oxidoreductase domain to catalyse enoyl reduction in the aforementioned modules.

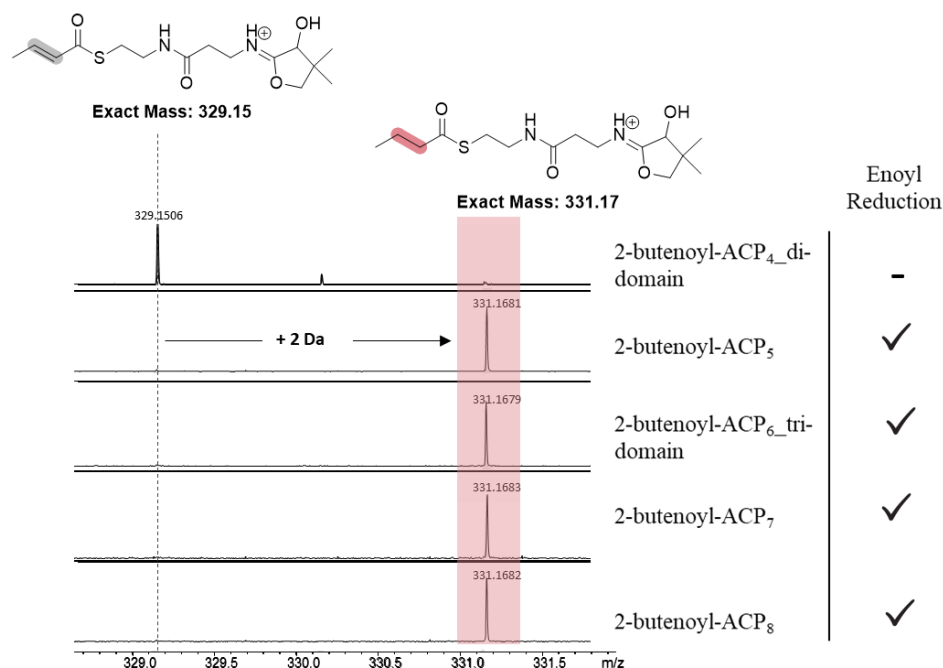


Figure 3.2.7 Stacked MS spectra of PPant ejected ions generated from 2-butenoyl-ACP domains after incubation with GdsB ER, in the presence of NADPH as hydride donor to FMN. Observed enoyl reduction is accompanied with a peak at m/z 331 corresponding to generated PPant ejection ion.

The selectivity of GdsB towards the ACP domains from modules 5 to 8 is perhaps unsurprising, as carefully programmed enoyl reduction events are required to maintain biosynthetic fidelity. However, the molecular basis for this selectivity remains unclear. A multiple sequence alignment of ACP domains from *trans*-AT PKSs that are known/believed to interact with *trans*-acting ER domains was constructed to examine conservation of residues that may direct ER recruitment (**Figure 3.2.8**). The alignment was compared to the two ACP domains from GdsF module 4, which did not interact with GdsB ER in the biochemical assays (**Figure 3.2.8**). Careful inspection revealed a conserved Phe residue appended in the domains present in all ACP domains that were shown *in vitro* to form specific interactions with *trans*-acting enoyl reductases. This Phe residue is replaced with a Leu residue in the 2 ACP domains tested in our system *in vitro* and showed no interaction with the GdsB-ER. Although in depth biophysical data to characterise the molecular basis for the interaction specificity between ACP domains and *trans*-acting ERs need to be obtained, these findings serve as preliminary evidence highlighting similarities between interacting ACP domains and key differences in the sequence of non-interacting ACP domains.

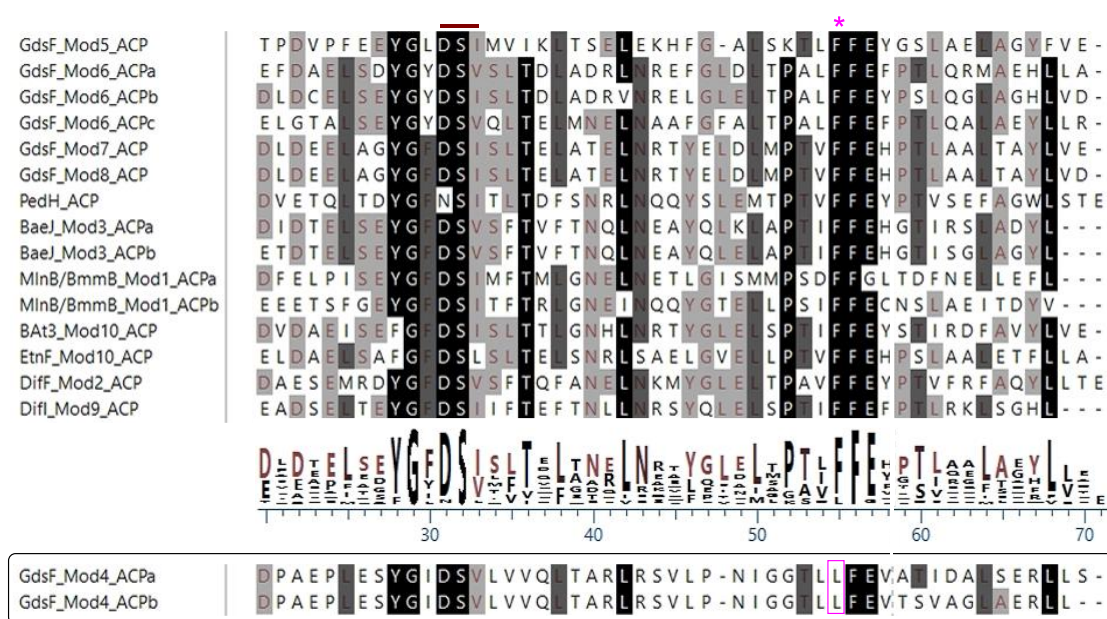


Figure 3.2.8 Multiple sequence alignment between gds ACP domains and homologues found to interact with *trans*-acting ERs and generated sequence logo. ACP domains used in this study that did not exhibited interactions with GdsB-ER are shown in a box. Phe to Leu conversion is indicated with a mangenta star/box. The helix II recognition triad is shown with a red line. Abbreviations are as follows: Gds, gladiostatin; Ent, etnangien; Ped, pederin; Bat, Batumin/kalimantacin; Bea, bacillaene; Dif, difficidin and Mln/Bmm, macrolactins.

3.2.3 *In vivo* characterisation of *gdsH*

While the biochemical assays discussed above demonstrate that the ER domain of GdsB is responsible for multiple enoyl reduction event on the PKS assembly line, the role of the second reductase, GdsH, remained obscure. To investigate the role of this NAD(P)H-dependent oxidoreductase in the biosynthesis of gladiostatin, a knock-out of the gene encoding this enzyme, *gdsH*, was generated. Gene deletions were created using a double homologous recombination mutagenesis approach (as detailed in **section 7.9.2**), that involved cloning of the flanking regions (800- 1000 bp) of the gene targeted for deletion into the suicide plasmid, pGPI (**Figure 3.2.9**). The suicide plasmid not only contains a trimethoprim resistance marker that enables selection of tranconjugants, but also contains a *SceI* endonuclease restriction site.

The plasmid is introduced into *E. coli* SY327 cells by electroporation. The transformants were selected in LB agar plates supplemented with trimethoprim and verified using colony restriction digestion and sequencing. The plasmid was introduced into the *B. gladioli* BCC1622 wild-type strain via triparental conjugation using *E. coli* SY327 containing the pGPI construct as a donor strain together with the helper strain *E. coli* HB101. Transconjugants were then selected for resistance in trimethoprim and polymyxin. The first recombination event is driven by trimethoprim, as described in the insertional mutagenesis procedure. Polymyxin is used to select against *E. coli* strains, allowing only the Tp^R/PMX^R *B. gladioli* BCC1622 Ω *gdsH* mutant allele to grow. The correct integration into the wild-type strain was verified by PCR and sequencing.

A second plasmid is then introduced to the single crossover mutant allele, pDAI-*SceI* through triparental mating. This plasmid constitutively expresses the I-*SceI* endonuclease which upon introduction, causes a double strand break into the inserted plasmid that propagates an intramolecular homologous recombination between the flanking region on the plasmid and chromosome. This results in either restoration of the parental allele, *B. gladioli* BCC1622 wild-type or the desired deletion, depending on the crossover site. Successful targeted deletion was monitored using colony PCR and sequencing (**Figure 3.2.9**). The PCR used to verify the successful deletion of *gdsH* was conducted using the forward primer that covers the middle of the 5' flanking arm and the reverse primer covering a region of the 3' flanking arm. The successful deletion produced a shorter amplified region (1000 bp) in comparison to wild-type strain which produces longer (1619 bp) amplified regions, which verified the absence of the targeted gene.

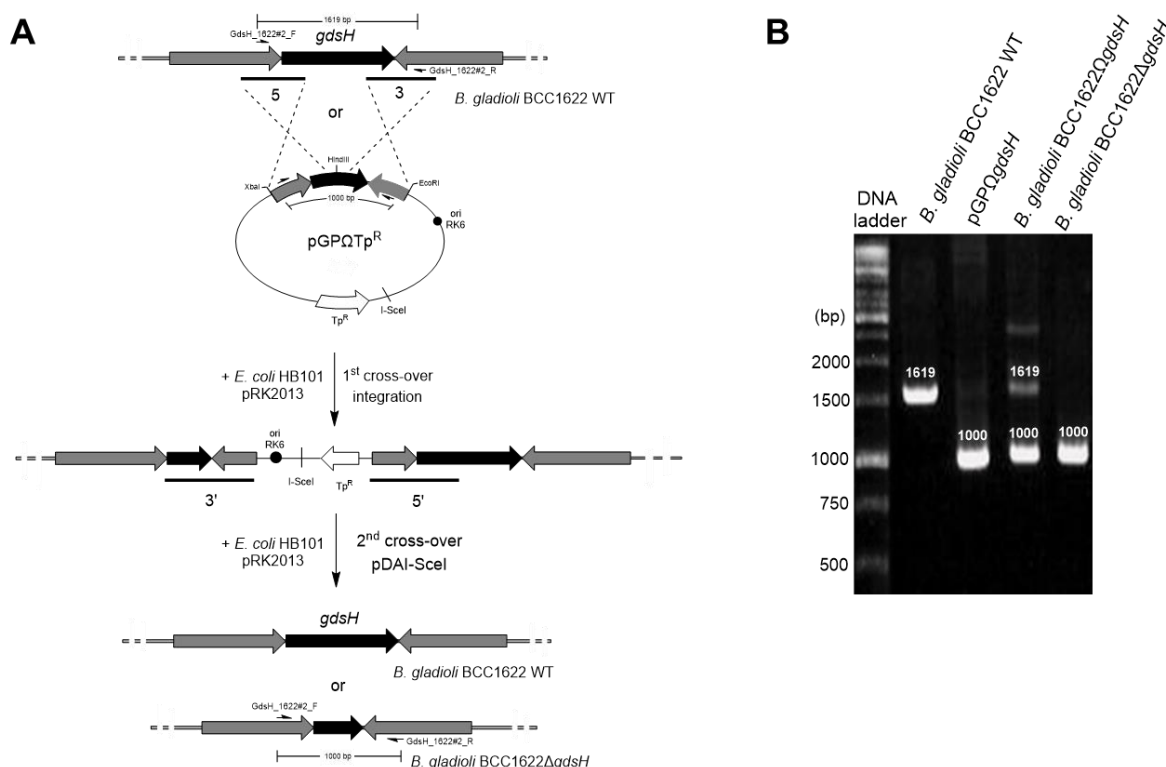


Figure 3.2.9 Construction and verification of *B. gladioli* BCC1622Δ*gdsH* mutants. (A) Schematic diagram depicting the in-frame gene deletion procedure of *gdsH* in *B. gladioli*. The suicide plasmid that contains regions flanking the targeted gene was transferred into *B. gladioli* by conjugation and integrated into the chromosome by the first homologous recombination event, yielding trimethoprim-resistant merodiploids. Subsequently, pDAI-SceI vector is introduced into the merodiploid which results on a double-stranded DNA break at the I-SceI recognition site on the chromosome. The DNA break stimulates a second homologous recombination event using the host DNA repair system and depending on the location of the recombination event, this generates either the desired gene deletion or restores the wild-type allele. The primers used for colony PCR are indicated in **Table 7.9.1**. (B) Agarose gel electrophoresis of the PCR product provide genetic evidence for the successful in-frame gene deletion. The lanes correspond to the PCR product of the *B. gladioli* BCC1622 wild type strain (1619 bp product), pGPI vector containing the flanking regions of *gdsH* (1000 bp product), first crossover allele *B. gladioli* BCC1622Ω*gdsH* mutant and the *B. gladioli* BCC1622Δ*gdsH* mutant (1000 bp product).

To detect the impact of *gdsH* in the biosynthesis of gladiostatin, the mutant strain was grown under the same conditions as the wild-type strain, and the ethyl acetate extracts were analysed by UHPLC-ESI-Q-TOF-MS (**Figure 3.2.10**). Analysis of the chromatograms showed that gladiostatin is abolished in the Δ*gdsH* mutant and a new metabolite is being produced, which

elutes at an earlier retention time with m/z 504.2592 $[M+H]^+$ and m/z 526.2410 $[M+Na]^+$ (deduced molecular formula of $C_{27}H_{36}NO_8$) (**Figure 3.2.10**). After careful analysis of the base peak chromatogram of the wild-type strain, it was found that this metabolite, which we named gladiostatin B (**55**), is also present in the wild-type strain metabolome but its production is significantly lower compared to gladiostatin (**49**) (**Figure 3.2.10**).

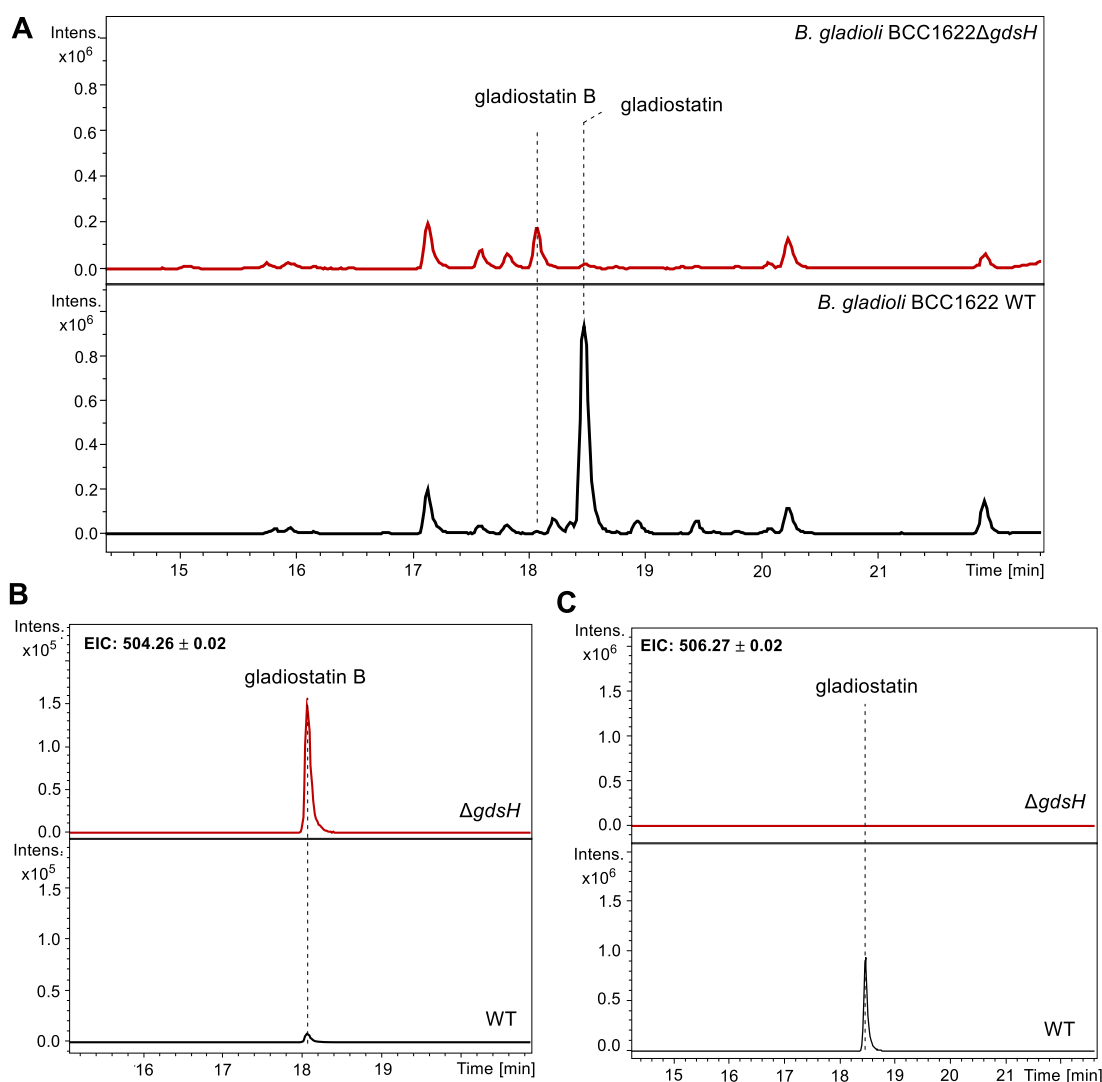


Figure 3.2.10 LC-MS analysis of metabolite extracts derived from *B. gladioli* BCC1622 $\Delta gdsH$ and *B. gladioli* BCC1622 wild type. (A) Base peak chromatogram of $\Delta gdsH$ mutant strain (shown in red) in comparison to wild type metabolic profile (shown in black). (B) Extracted ion chromatogram of m/z =

504.26 \pm 0.02, corresponding to the [M+H]⁺ species observed for gladiostatin B. (C) Extracted ion chromatogram of m/z = 506.27 \pm 0.02, corresponding to the [M+H]⁺ species observed for gladiostatin.

To rule out possible polar effects that could have been introduced in the biosynthetic pathway of **49** by the *gdsH* deletion that could potentially disrupt gladiostatin biosynthesis, we performed complementation of *gdsH* into the $\Delta gdsH$ mutant to validate the function of this enzyme. The complementation was achieved using the arabinose-inducible pMLBAD expression vector, which has been reported to perform regulation of gene expression under the control of the P_{BAD} promoter in *Burkholderia cepacia* species.¹⁹⁵ As detailed in **section 7.9.2**, the recombinant construct pMLBAD-*gdsH* was mobilised in the $\Delta gdsH$ mutant by triparental mating. The successful complementation mutants were screened by antibiotic selection and confirmed by PCR (**Figure 3.2.11**).

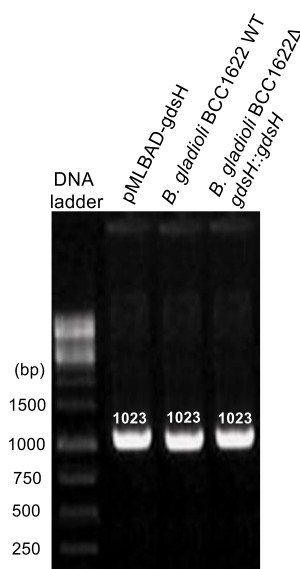


Figure 3.2.11 DNA electrophoresis gel of the PCR products derived from complementation studies. The colony PCR confirmed successful complementation of the mutant (lane 3), resulting in a 1023 bp product, similar to positive controls used (wild type strain and pMLBAD containing the *gdsH*).

A final concentration of 0.25% of L-arabinose was used as an optimal concentration for the induction of gene expression. Comparative LC-MS analysis retrieved from ethyl acetate extracts from the $\Delta gdsH$ mutant, $\Delta gdsH::gdsH$ and wild type strain showed that complementation of *gdsH* successfully restored production of **49**, validating that no polar effect

was brought into the *gdsH* deletion mutant and the absence of gladiostatin in *gdsH* mutant is an outcome after loss of GdsH's function in the assembly line (**Figure 3.2.12**).

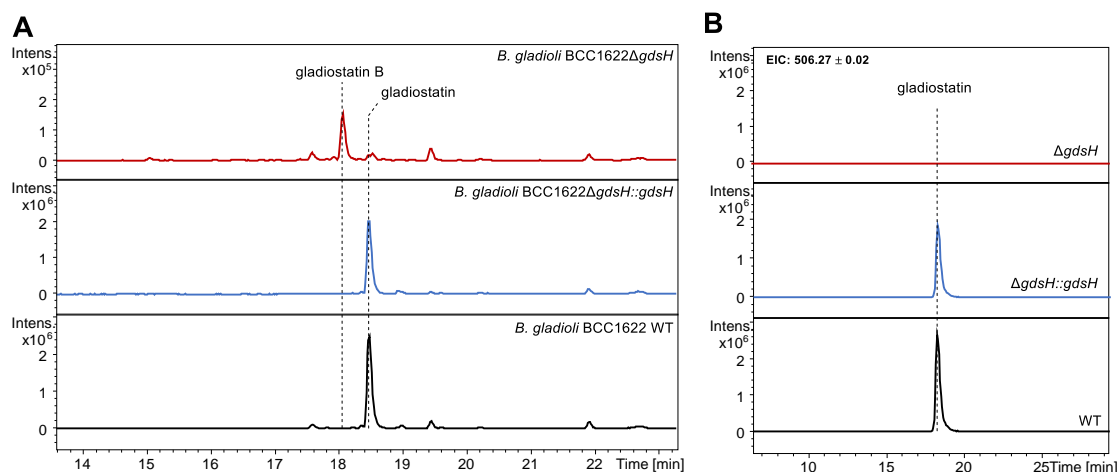


Figure 3.2.12 LC-MS analysis for gladiostatin production after complementation of BCC1622Δ*gdsH* with *gdsH*. (A) Base peak chromatograms of *B. gladioli* BCC1622Δ*gdsH* (red), *B. gladioli* BCC1622Δ*gdsH*::*gdsH* (blue) and *B. gladioli* BCC1622 wild type (black). (B) Extracted ion chromatograms of $m/z = 506.27 \pm 0.02$, corresponding to gladiostatin protonated species.

3.2.4 Isolation and structure elucidation of compound 55

In order to characterise the new metabolite **55** produced after the deletion of *gdsH*, a large-scale culture of *B. gladioli* BCC1622Δ*gdsH* was grown on BSM agar containing glucose as carbon source for one day. Preparative HPLC was used to fractionate the crude ethyl acetate extract and 6 mg of pure **55** was obtained (1.6 mg/L). The structure of **55** was elucidated by 1D- and 2D-NMR spectroscopy. Overall, the NMR assignments match the structure of **49** (**Table 7.10.2**).

Comparisons between ¹H NMR spectra of **49** and **55** indicated additional olefinic protons signals (**Figure 3.2.13**). A network of COSY correlations between C-8 to C-18 helped to establish the location of the new double bond in the structure, which was further confirmed by HMBC correlations between 8-Me protons to C-9 and C-12 protons to C-14. Based on a ³J_{HH} coupling constant of 15 Hz, the C-11/C-12 double bond was assigned the *E* configuration. The

rest of the chemical shift assignments were identical to that of gladiostatin, establishing that deletion of *gdsH* triggered no other structural alterations in the molecule.

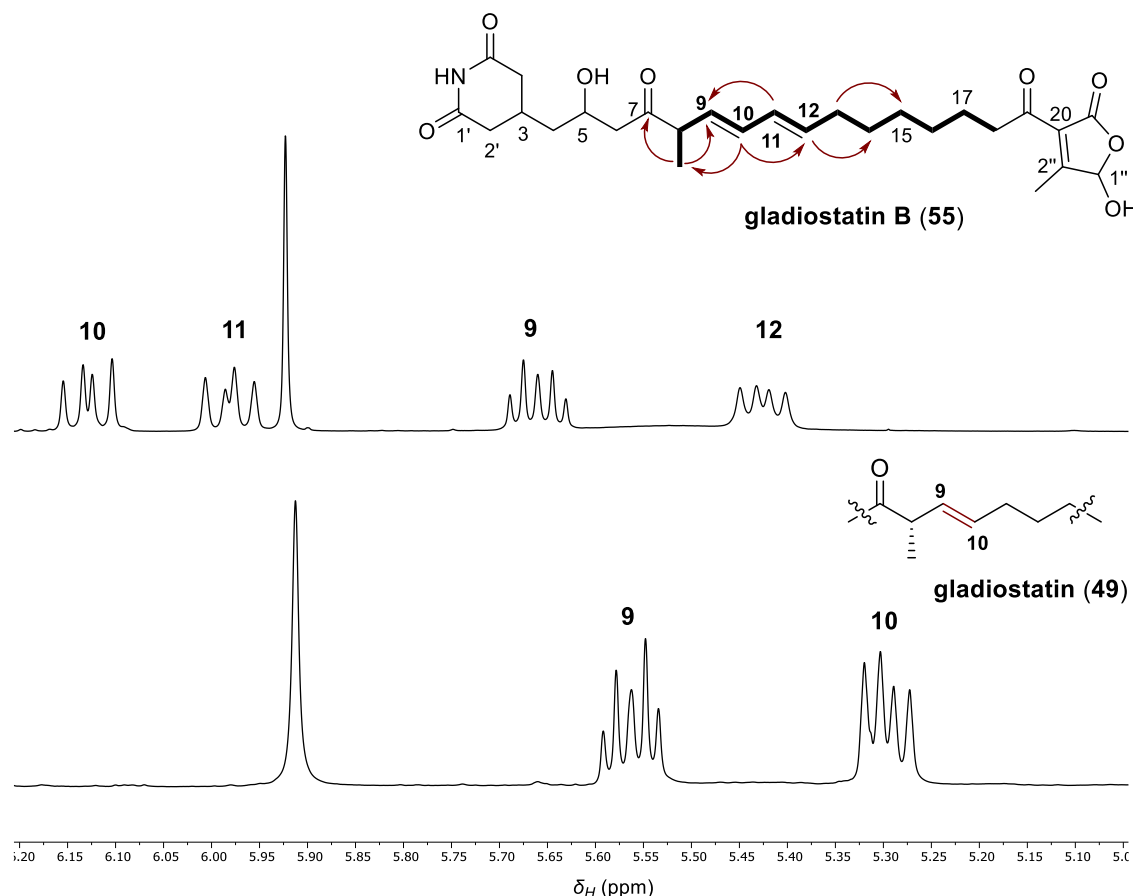


Figure 3.2.13 Comparative analysis of the signals in the olefinic proton region of the ¹H NMR spectra between gladiostatin B (top) and gladiostatin (bottom). Key COSY and HMBC correlations confirming the position of the additional double bond are shown with bold lines and red arrows, respectively.

The presence of a second double bond between C-11/C-12 in gladiostatin B suggests the direct involvement of GdsH in the enoyl reduction step of the α,β unsaturated thioester intermediate produced in module 5 of the *gds* PKS. Since the *in vitro* and *in vivo* characterisation of GdsH did not provide clear evidence with regard to its function, an *in silico* analysis of the ACPs encoded by *gdsE* and *gdsF* was conducted to obtain additional insights into the peculiarities found in module 5 of *gds* PKS.

3.2.5 *In silico* analysis of ACPs from *gds* BGC

In the first instance, a cladogram was constructed from a library of ACP domains derived from well-characterised *trans*-AT PKS assembly lines, following methods used from a previous study. In this work, the authors show that *trans*-AT PKSs are mosaics of evolutionary units of catalytic domains, with defined ACP domains at the core of each evolutionary unit (**Figure 3.2.14**).¹⁹⁶ The library of ACP domains used to construct the cladogram also included sequences of ACP domains from stand-alone ACPs involved in starter unit biosynthesis and branching sub-modules (KS-B-ACP). This resulted in two new clades that are identified here as family 1 and family 2, respectively (**Figure 3.2.14**).

From the cladogram it was observed that ACP domains from module 6,7 and 8 from the gladiostatin pathway clade with ACP domains in evolutionary units with a similar architecture, DH-KR-ACP-KS + *trans*-acting ER, found in different assembly lines (**Figure 3.2.14**).

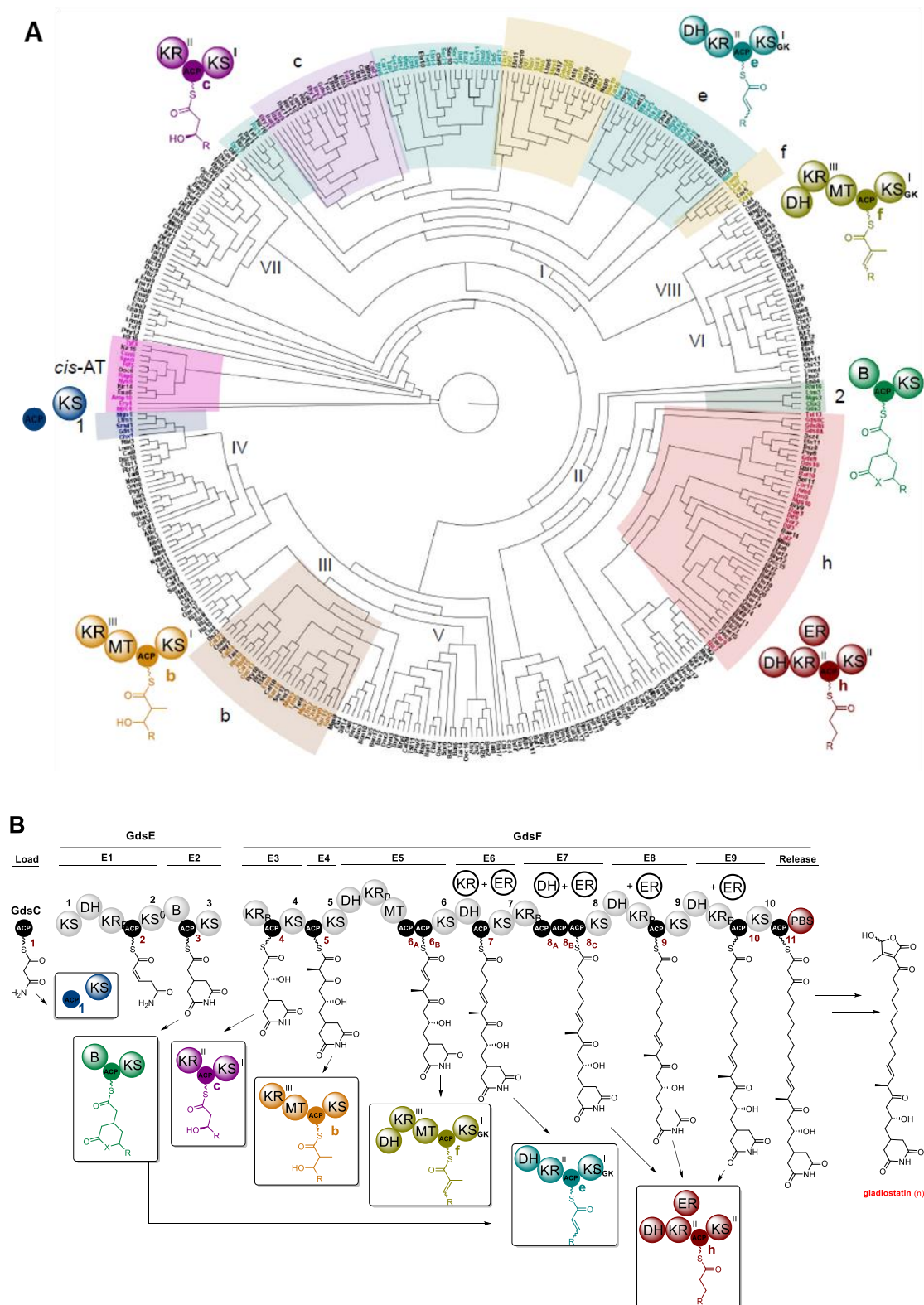


Figure 3.2.14 Cladogram analysis and module organisation of *gds* cluster using evolutionary units.

(A) Cladogram comparing the sequences of ACP domains from various *trans*-AT PKSs with the ACP

domains of the gladiostatin PKS. Domain architectures for the module types observed in the gladiostatin PKS are highlighted and illustrated. A selection of ACP domains from *cis*-AT PKSs were used to root the tree and clade notation as described by Vander Wood et al.¹⁹⁷ **(B)** Evolutionary units and ACPs from *gds* biosynthetic gene cluster are numbered from 1-9 and 1-11, respectively. Arrows indicate the predicted phylogenetic classifiers based on redefined module theory. Gate keeping (GK) enzymes are shown.

Surprisingly, the ACP domain from module 5 clades with ACP domains that belong to the ACP family (e) with DH-KR-ACP-KS evolutionary unit architecture. This module type contains a B-type KR that generates a D- β -hydroxyacyl substrate which upon function of the active DH results in the formation of *trans*- α,β double bond. In addition, as mentioned in **Chapter 2 (Section 2.2.3.)**, using the TransATor webserver it was predicted that KS from module 6 is specific for substrates containing an *E* configured double bond. Since both metabolites, **49** and **55** can be found in the wild type strain, it can be postulated that KS from module 6 can accept both α,β unsaturated and fully saturated intermediates.

These findings in combination to the secondary metabolite found to be produced from the $\Delta gdsH$ mutant led us to the hypothesis that GdsH is complementing the function of the GdsB-excised ER though interaction with module 5 ACP domain could not be observed using *in vitro* biochemical studies. Interestingly GdsH presents 41% identity to MgsI, an oxidoreductase that was established to saturate a double bond as a post-PKS modification in migrastatin (**29**) biosynthesis through gene inactivation.¹⁹⁸ The hypothesis of GdsH to catalyse a post-PKS enoyl reduction is highly unlikely to occur though in gladiostatin biosynthesis, since in the structure of **29**, the double bond is conjugated to a carbonyl group whereas in the case of gladiostatin, the double bond is conjugated to a double bond, and therefore this makes it a less reactive electrophile.

3.3 Conclusions

This study elucidates the enoyl reduction steps in the biosynthesis of gladiostatin that were proposed to occur in *trans* during the assembly of the polyketide. Using NCBI conserved search, two candidates were predicted to possibly saturate the α,β -unsaturated intermediates, GdsB and GdsH. *In vitro* biochemical assays indicated that the GdsB-excised flavin-dependent ER is the sole enzyme interacting with the ACP domains from modules where the enoyl reduction is suggested to occur. Sequence alignment of different ACP domains that interact with *trans*-acting ERs revealed a conserved Phe that is absent from the non-interacting ACPs. This difference in the sequence could be used as starting point in an effort to understand the specificity of the *trans*-acting ER towards ACP domains. The standalone oxidoreductase, GdsH, showed no reduction activity towards the 2-butenoyl unit tethered on GdsF ACP domains during the *in vitro* characterisation.

In vivo approaches were then employed to further understand the function of the aforementioned enzyme in the biosynthesis. In-frame deletion and analysis of the metabolite profile indicated that GdsH is indeed important for the production of gladiostatin since the deletion of the gene abolished the production of **49** and promoted the accumulation of a new metabolite, gladiostatin B (**55**), which is produced at low-levels in the wild type strain. Structure elucidation of the molecule revealed the presence of a second double bond. Based on the module organisation of gladiostatin it became apparent that the second consecutive double bond is a product of an α,β -unsaturated intermediate not being reduced in module 5 of the assembly line.

Construction of a cladogram, incorporating ACP domains from known *trans*-AT PKS system together with all ACP domains from the *gds* system outlined the divergence of module 5 ACP with regards to the rest of the ACP domains which show interaction with *trans*-acting ERs. The different clading of the ACP domain in module 5 and the predicted preference of the downstream KS domain for α,β -unsaturated substrates reveal that evolution drove KS to accept this substrate in a much lesser degree, favouring the formation of the main metabolite. This is accompanied by the proofreading role of GdsH on saturating substrates with cumulative double

bonds. This can be achieved by (i) interacting directly with the module 5 ACP, (ii) mediating the interaction of GdsB with the ACP or by (iii) performing an enoyl reduction as post-PKS modification, although this is unlikely mechanistically, as described above.

Future studies on the molecular basis underlying the specific interactions between module 5 ACP domain with GdsB-ER and/or GdsH as well as studies on the substrate specificity of module 6 KS will provide evidence on whether these two enzymes are collaborating for the reduction of the C-11/C-12 double bond that form during the assembly of the metabolite.

Chapter 4

*Investigating the mechanism of 2-acyl-4-hydroxy-3-methyl
butenolide formation in gladiostatin biosynthesis*

4.1 Introduction

4.1.1 Butyrolactones/butenolide substructures found in *Streptomyces* autoregulators

The vast accumulation of genome sequencing data and genome mining have revealed that Actinobacteria possess a great potential for secondary metabolite production.¹⁹⁹ Yet most of these biosynthetic gene clusters remain cryptic under laboratory growth conditions. These pathways are usually controlled by very complex regulatory networks, in which low molecular weight molecules, also known as autoregulators or bacterial hormones, play an important role.^{199,200} In the absence of specific signalling molecules some biosynthetic gene clusters remain untranscribed or 'silent', therefore their characterisation is essential to understand their role in secondary metabolite production and bacterial morphological differentiation.

The archetypal A-factor containing the chemical structural characteristic γ -butyrolactone (GBL), was the first *Streptomyces* hormone reported in 1967 and it was found to be responsible for streptomycin production in *S. griseus*.^{201,202} In following years, novel GBLs, including the virginiae butanolides (VBs)^{203,204} and SCBs^{205–208} were identified, which regulate antibiotic biosynthesis in *S. virginiae*, and *S. coelicolor* A3(2), respectively. The γ -butyrolactone-type signalling molecules all contain the 2,3 disubstituted γ -butyrolactone skeleton and differ only in the oxidation state, hydroxyl group stereochemistry, position of methyl branches and length of the alkyl chain (**Figure 4.1.1**). Furan-type autoregulators such as 2-alkyl-4-hydroxy-methylfuran-3-carboxylic acids (AHFCAs), typified by the methylenomycin furans were also found to regulate the production of methylenomycin in *S. coelicolor* (**Figure 4.1.1**).^{209,210} More recently, two more classes of *Streptomyces* butenolide signalling molecules, and 2-hydroxyalkyl-4-hydroxy-3-methyl butenolides, SABs¹⁷¹ and SRBs²¹¹ were identified to regulate production of nikkomycin and lankamycin/lankacidin, respectively (**Figure 4.1.1**). The butenolide class of signalling molecules contain a five-membered ring backbone with an unsaturated bond between C-3 and C-4, with diverse side chains at C-3, C-4 or C-5.

Autoregulators usually function via binding to cognate receptor proteins, and their effects can be transmitted in cascade fashion to specific pathways by a series of regulators. An example of autoregulatory system that has been studied is the AHFCA/MmfR/MmyR system responsible for the methylenomycin A biosynthesis in *S. coelicolor* (**Figure 4.1.1**).^{210,212} It is proposed that translation of *mmfLHP* results in the assembly of AHFCA signalling molecules. Once the AHFCA reaches a critical concentration, they bind to the MmfR repressor, which in turn triggers the derepression of MmfR binding to the operator within the *mmfL*-*mmfR* intergenic region, leading to the increase of production of AHFCAs. This is followed by derepression of MmfR on the *mmyB* promoter, resulting in the transcriptional activator MmyB to be produced. MmyB subsequently binds to its operator, propelling the transcription of the biosynthetic genes responsible for the production of methylenomycin A.

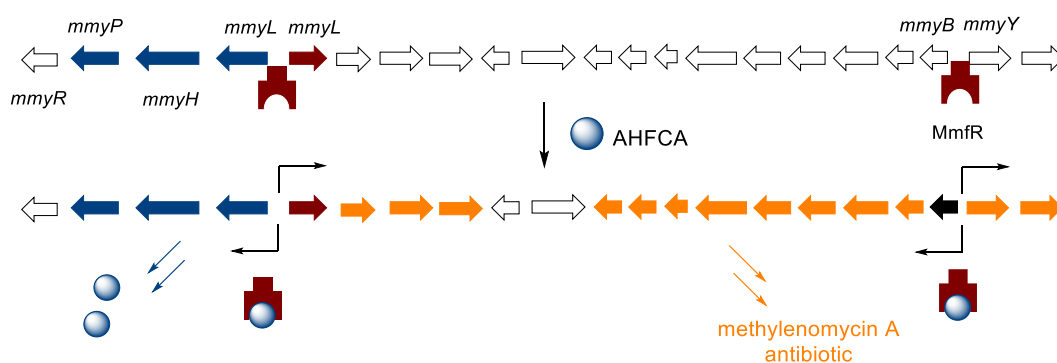


Figure 4.1.1 Proposed signalling cascade with MmfR involved in methylenomycin A biosynthesis.

Since signalling molecules are essential for triggering the signalling cascades, research towards their biosynthesis is of paramount importance. Studies on the biosynthesis of the signalling molecule A-factor, identified AfsA as the key enzyme. The precise role of AfsA was elucidated by Kato *et al.* who demonstrated that this enzyme catalyses the condensation of an ACP-bound 8-methyl-3-oxononanoate thioester with the hydroxyl group of dihydroxyacetone phosphate (DHAP) (**Figure 4.1.2**).¹⁶¹ The final core of A-factor is furnished after reduction of the phosphorylated butenolide to the corresponding butanolide by BprA, followed by dephosphorylation.

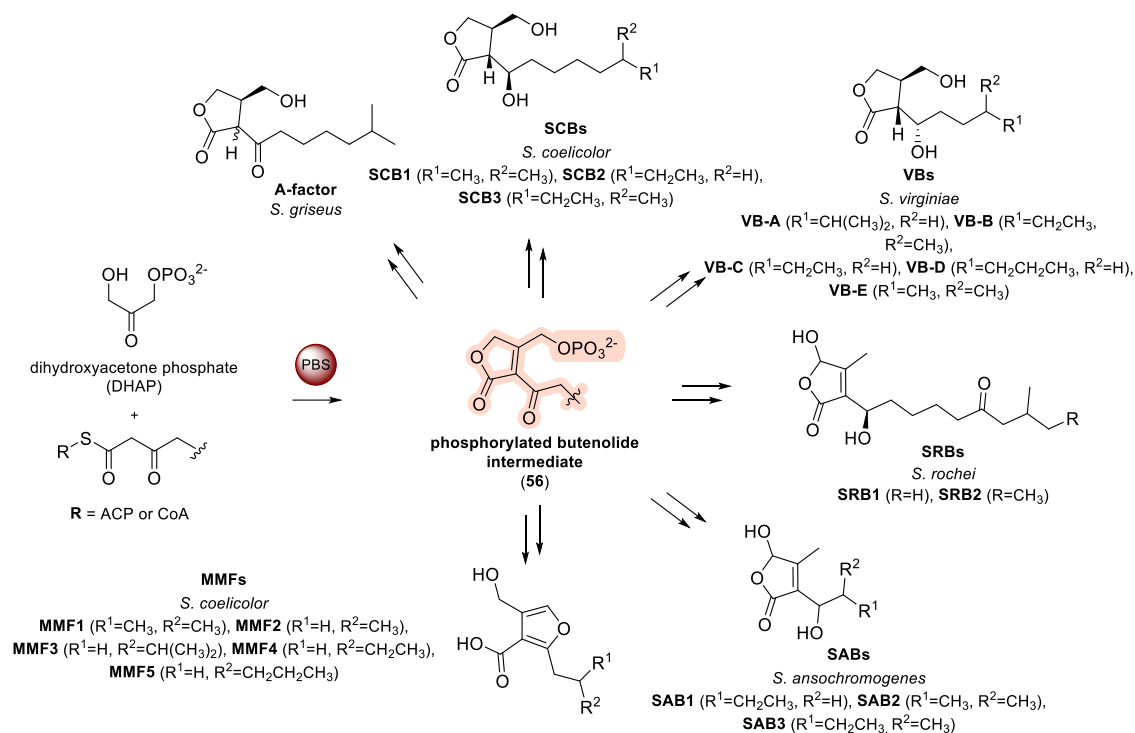


Figure 4.1.2 Proposed pathway for the biosynthesis of GBLs (including A-factor, SCBs and VBs), 2-hydroxyalkyl-4-hydroxy-3-methyl butenolides (SRBs and SABs) and AHFCAs (represented by MMFs), involving intermediates with a common phosphorylated butenolide core structure (highlighted in orange).

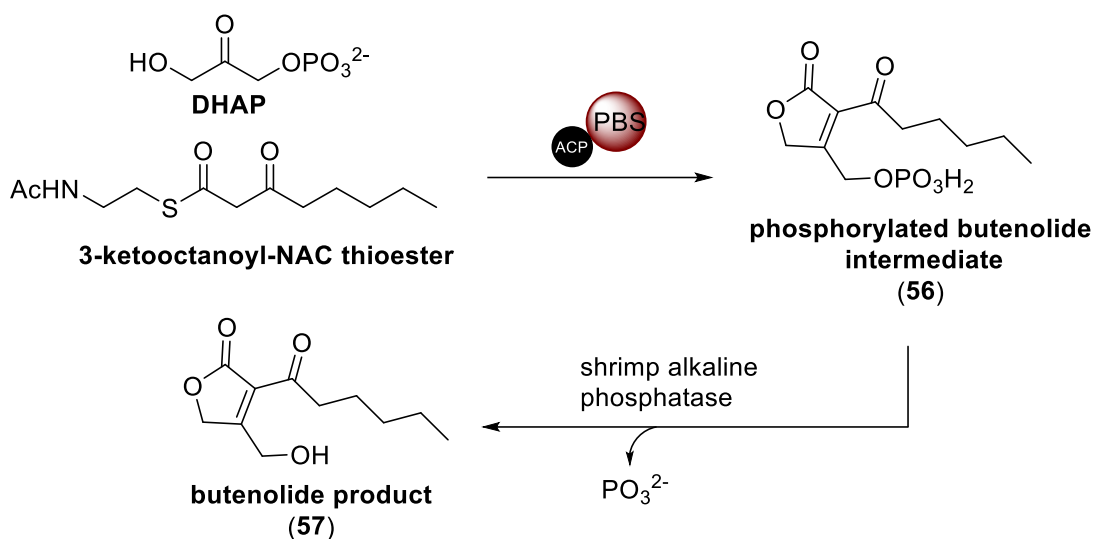
AfsA homologues have been identified in BGCs of other members GBL family including SCBs in *S. coelicolor* where ScbA was shown to direct their biosynthesis.²¹³ Interestingly, studies have demonstrated that this family of enzymes are also responsible for the biogenesis of 2-hydroxyalkyl-4-hydroxy-3-methyl butenolides and AHFCAs signalling molecules. AfsA homologues, SabA, SrrX and MmfL, contribute to regulating the biosynthesis of nikkomycin, lankacidin/lankamycin and methylenomycin A, respectively.^{171,210,211,214,215} It has been proposed that these enzymes perform analogous reactions to that catalysed by AfsA, and all families of *Streptomyces* autoregulators are biosynthesised through a common phosphorylated butenolide intermediate (**Figure 4.1.2**). Incorporation studies using isotopically-labeled precursors provided an initial indication that the chemical origin of MMFs derives from DHAP.²¹⁶ Recently, enzymatic characterisation of the MmfL established that MMF biosynthesis proceeds via the suggested phosphorylated butenolide which subsequently undergoes dephosphorylation and

MmfH-catalysed rearrangement to yield the AHFCA core.²¹⁴ Thus, it appears that phosphorylated 2-acyl-3-hydroxymethylbutenolide intermediates are present in the biosynthesis of structurally diverse natural products, including GBLs, AHFCAs and AHMBs. An exception being GBLs, further work is needed to understand the mechanisms by which these intermediates become elaborated into the final metabolic products.

4.2 Previous studies on the polyketide chain release in gladiostatin biosynthesis

The structure of gladiostatin (**49**) represents an amalgamation of substructures of high biological importance located solely in secondary metabolites from *Streptomyces* sp. As described previously, related butenolide/butyrolactone moieties are found in *Streptomyces* signalling molecules which are known to regulate secondary metabolism and morphogenesis. Whilst the core biosynthetic genes required for installing a glutarimide moiety in gladiostatin were evident in the putative *trans*-AT PKS cluster, the finding of an AfsA-like domain appended in the C-terminus of the multi-modular PKS, however, is highly unexpected. Typically, PKSs employ a thioesterase (TE) domain at their C-terminus to catalyse release of the nascent polyketide chain via hydrolysis or macrocyclisation, although other release mechanisms have been observed.²¹⁷

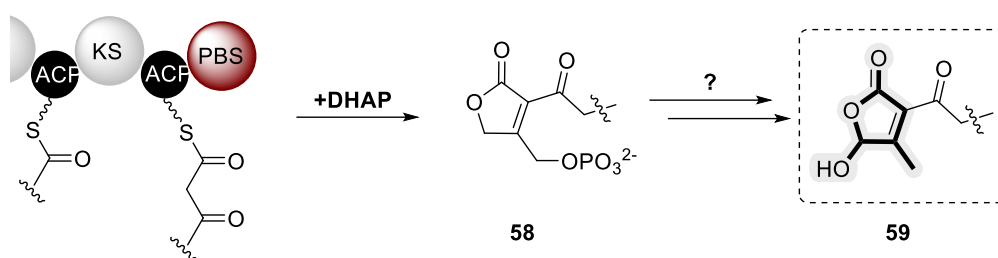
The presence of an AfsA-like domain at the C-terminus of GdsF suggested that chain release might proceed via condensation of a β -ketoacyl-ACP with DHAP to yield a phosphorylated butenolide polyketide intermediate (mentioned in **section 4.1.1**). Congruent with the proposed enzymatic product, the domain was assigned as a phosphorylated butenolide synthase (PBS). Previous studies conducted by Dr. Matthew Jenner have enabled the *in vitro* reconstitution of the chain release mechanism from the gladiostatin PKS with the use of simplified substrate mimics of the fully assembled polyketide chain and DHAP, elaborating the catalytic activity of the PBS domains (**Scheme 4.2.1**).



Scheme 4.2.1 *In vitro* reconstitution of chain release by the PBS domain. Condensation of DHAP and the NAC thioester of 3-ketooctanoate by the purified recombinant GdsF ACP-PBS di-domain and dephosphorylation of the resulting phosphorylated butenolide **57** using shrimp alkaline phosphatase to yield the butenolide product.

Similar to the biochemical studies on MMFs, dephosphorylation yielded the enzymatic product **57**, demonstrating that the phosphorylated intermediate **58** common to previously described systems is indeed being produced, but must be further modified enzymatically to yield the 2-acyl-4-hydroxy-3-methyl butenolide unit.²¹⁸ Independent studies also demonstrated that upon deletion of PBS domain appended in gladiofungin (*syn* gladiostatin) PKS abolished production of the metabolite, indicating its key role in the biosynthesis.²¹⁹

The overall aim of this project was to investigate the steps involved in biosynthetic construction of the 2-acyl-4-hydroxy-3-methyl butenolide unit appended in the structure of gladiostatin, following PBS catalysis (**Scheme 4.2.2**). Using gene comparisons to known clusters that produce similar molecules, key enzymes responsible for the biosynthesis of the butenolide **59** were identified. Characterisation of these enzyme was performed through *in vivo* gene deletion strategies and *in vitro* biochemical assays with a principal aim to reconstitute *in vitro* the enzymatic cascade responsible for the production of the tri-substituted butenolide **59**.



Scheme 4.2.2 Overview of the scope of the study described in this chapter.

4.3 Results and Discussion

4.3.1 Identification of genes in *gds* cluster encoding putative phosphatases and investigation of their involvement in dephosphorylation of biosynthetic intermediates

Previous studies on the *in vitro* characterisation of the AfsA-like PBS domain appended to the C-terminus of the gladiostatin PKS established that release of 3-keto-thioester intermediate from the upstream ACP domain occurs via PBS-catalysed condensation with DHAP.²²⁰ The phosphorylated butenolide product **58** is similar to intermediates produced by AfsA and the AfsA-homologue, MmfL.^{161,214} The resulting phosphorylated butenolide is proposed to undergo rearrangement and dephosphorylation to afford the AHMB found in gladiostatin. BLAST analysis identified two genes, *gdsA* and *gdsG*, in the BGC which encode for putative phosphatases that might catalyse the aforementioned steps. As described in **Section 4.1**, PBSs are proposed to condense β -ketothioesters with DHAP in the biosynthesis of several classes of *Streptomyces* signalling molecules, such as SABs and SRBs which contain a similar butenolide moiety to gladiostatin (**Figure 4.3.1**).^{171,211}

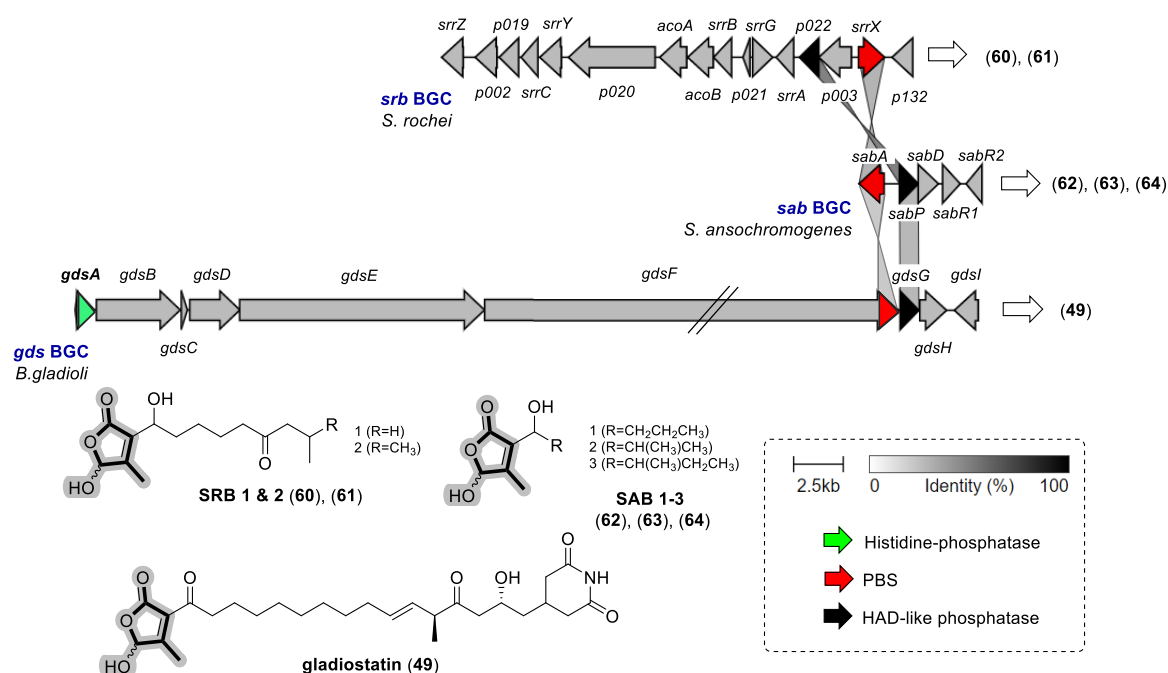
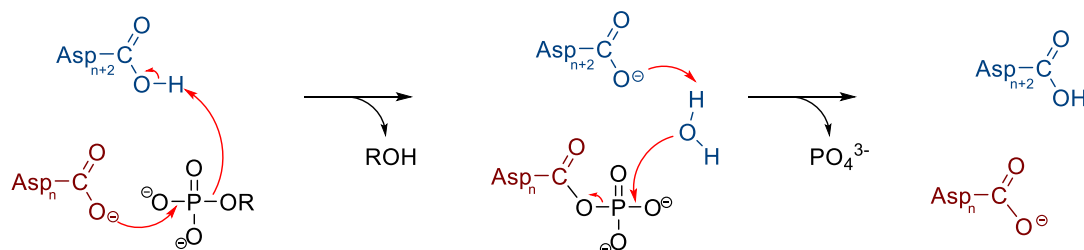


Figure 4.3.1 Comparison of biosynthetic gene clusters responsible for the production of and 2-hydroxyalkyl-4-hydroxy-3-methyl butenolides moieties in *Streptomyces* sp. with that encoded by the cryptic *B. gladioli* gene cluster.

Comparison of the *gds* BGC to the *srb* and *sab* BGCs from *S. rochei* and *S. ansochochromogenes*, respectively, revealed that besides the conserved PBS enzyme found in all BGCs, SabP and P022 presented ~ 40% identity in amino acid sequence to GdsG. These enzymes are annotated as HAD-like phosphatases. The HAD superfamily is named after the archetypal halogenated acid dehalogenase²²¹ in bacteria, which perform carbon or phosphoryl group transfer reactions on a diverse range of substrates. Specifically, the HAD-like phosphatases are a family of assorted enzymes that are responsible for the catalysis of the majority of phosphomonoester hydrolysis reactions in all kingdoms of life. Although the sequence similarity between HAD phosphatases is general really low, family members can be identified by the presence of a characteristic Rossmannoid fold and the active site sequence DxDx(V/T).^{222–225}

HAD phosphatases utilise an Asp residue as a nucleophile through a magnesium-dependent phosphoaspartyl transferase reaction (**Figure 4.3.1**). Phosphatase substrate specificity is

dictated by flexible loop structures on the α/β core domain and on the cap domain (**Scheme 4.3.1**).²²⁶ The spectrum of phosphoryl substrate hydrolysed is usually extensive, including sugar and non-sugar phosphates.²²⁷ While the HAD superfamily is a ubiquitous class of enzymes present in all superkingdoms of life, only a small fraction of them have been functionally characterised.



Scheme 4.3.1 The general catalytic mechanism of HAD phosphatases. Catalysis proceeds through an aspartyl-phosphate intermediate.

Based on the gene annotation, the catalysis performed by the corresponding enzyme and the conservation among *srb*, *sab* and *gds* BGCs, we postulated that GdsG is a possible candidate for the dephosphorylation of the intermediate **58**.

Interestingly, *gdsA*, another gene encoding an enzyme annotated as histidine phosphatase present in the *gds* BGC, was also considered as a candidate for the proposed dephosphorylation step in gladiostatin biosynthesis. This superfamily is identified by a conserved catalytic histidine residue in the motif 'RHG' present at the N terminus which becomes phosphorylated during the reaction (**Figure 4.3.2**).^{228,229}

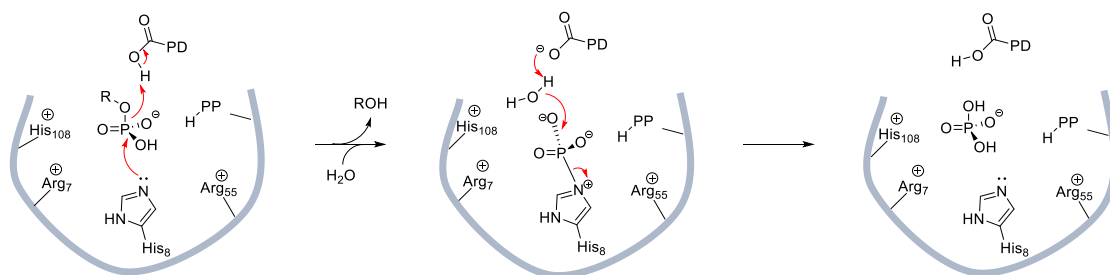


Figure 4.3.2 Catalytic mechanism of the histidine phosphatase superfamily. The residues of the catalytic core are shown as numbered in *E. coli* SixA²³⁰, a well-characterised protein of the family. His8 is

phosphorylated during the course of the reaction. The rest of the residues interact electrostatically with the phosphate group before, during and after its transfer. Additional neutral or positive residues (represented as PP) contribute to the phosphate pocket by hydrogen-bonding to the phosphor group. Asp or Glu proton donors are denoted as PD.²²⁸

In order to determine the specific biosynthetic role of *gdsA* and *gdsG* in the assembly of gladiostatin, functional characterisation of these enzymes were carried out using *in vivo* and *in vitro* approaches.

4.3.1.1 *In vivo* characterisation of *gdsG*

The identification of GdsG as a homologue to phosphatases that are involved in the biosynthesis of structurally identical AHMB moieties, allowed us to initiate our studies by inactivating the gene to verify its function. The in-frame deletion mutant was generated using the method discussed in **section 7.9.2**. The $\Delta gdsG$ mutant was obtained *via* selection with tetracycline and polymyxin antibiotics and confirmed by PCR (**Figure 4.3.3**).

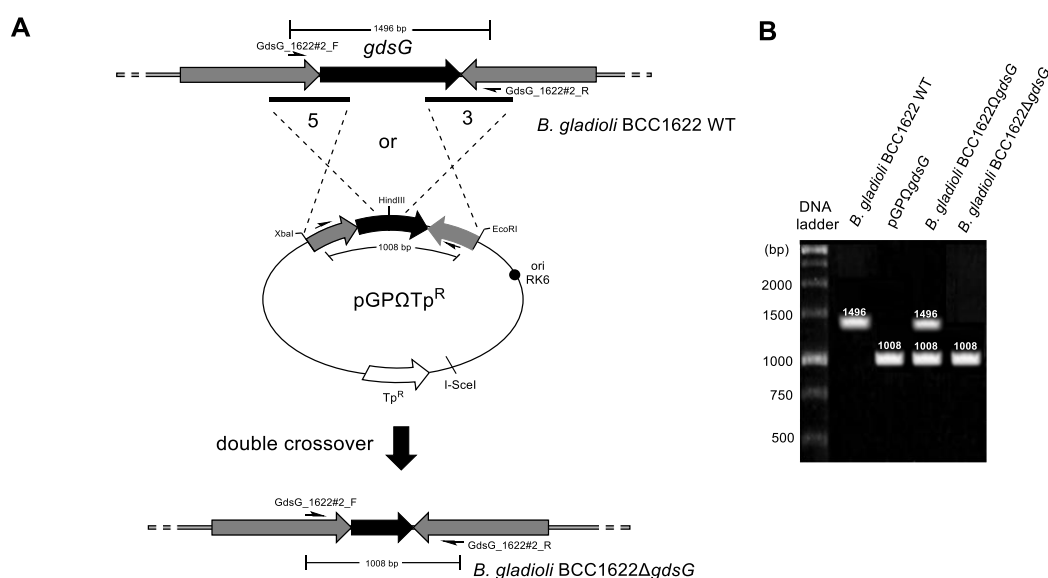


Figure 4.3.3 Construction and verification of *B. gladioli* BCC1622Δ*gdsG* mutants. (A) Schematic diagram depicting the in frame gene deletion procedure of *gdsG* in *B. gladioli*. The primers used for colony PCR are indicated in **Table 7.9.1**. (B) Agarose gel electrophoresis of the PCR product provide genetic evidence for the successful in-frame gene deletion. The lanes correspond to the PCR product of the *B.*

gladioli BCC1622 wild type strain (1496 bp product), pGPI vector containing the flanking regions of *gdsG* (1008 bp product), first crossover allele *B. gladioli* BCC1622 Δ *gdsG* mutant and the *B. gladioli* BCC1622 Δ *gdsG* mutant (1008 bp product).

The effect of the Δ *gdsG* mutation in gladiostatin biosynthesis was examined using high-resolution LC-MS (**Figure 4.3.4**). Comparative analysis of the metabolite profiling of *B. gladioli* BCC1622 Δ *gdsG* mutant and the wild-type strain revealed that gladiostatin production was abolished in the mutant strain and a new secondary metabolite, gladiostatin D (**59**), was produced at a reduced retention time (**Figure 4.3.4**). Surprisingly, the corresponding protonated species of $m/z = 506.27$ and deduced molecular formula of $C_{27}H_{38}NO_8$ were identical to that observed for gladiostatin.

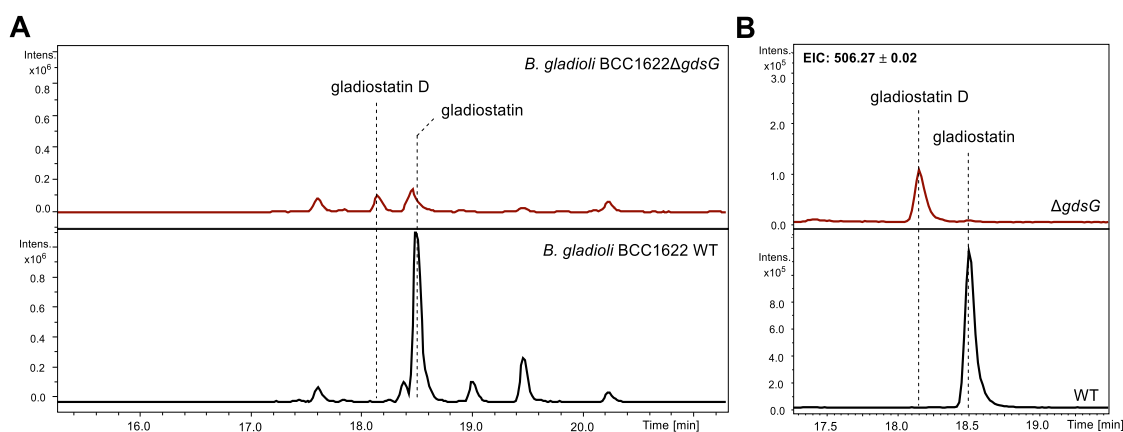


Figure 4.3.4 LC-MS analysis of metabolite extracts derived from *B. gladioli* BCC1622 Δ *gdsG* and *B. gladioli* BCC1622 wild type. **(A)** Base peak chromatogram of Δ *gdsG* mutant strain (shown in red) in comparison to wild type metabolic profile (shown in black). **(B)** Extracted ion chromatogram of $m/z = 506.27 \pm 0.02$, corresponding to the $[M+H]^+$ species observed for gladiostatin D (red) gladiostatin (black), showing a reduced retention time for the former.

To verify that the introduced mutation generated this new metabolite, the mutant was complemented by *in trans* expression of *gdsG*. The gene complementation was achieved by utilising the arabinose-inducible pMLBAD vector, which the *gdsG* gene was cloned into, as described in **section 7.3.3**. The pMLBAD-*gdsG* vector was then mobilised into the Δ *gdsG* mutant by triparental-mating. The mutant *B. gladioli* BCC1622 Δ *gdsG*::*gdsG* was screened for

antibiotic selection and PCR analysis confirmed its successful integration (**Figure 4.3.5**). LC-MS analysis showed that complementation of $\Delta gdsG$ with *gdsG* *in trans* recovered gladiostatin production to wild-type levels (**Figure 4.3.6**), indicating that the observed loss of gladiostatin production was due to deletion of the *gdsG* gene.

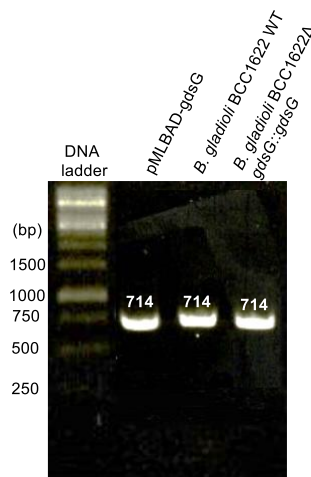


Figure 4.3.5 DNA electrophoresis gel of the PCR products derived from complementation studies.

The colony PCR confirmed successful complementation of the mutant (lane 3), resulting in a 714 bp product, similar to positive controls used (pMLBAD containing the *gdsG* and wild type strain).

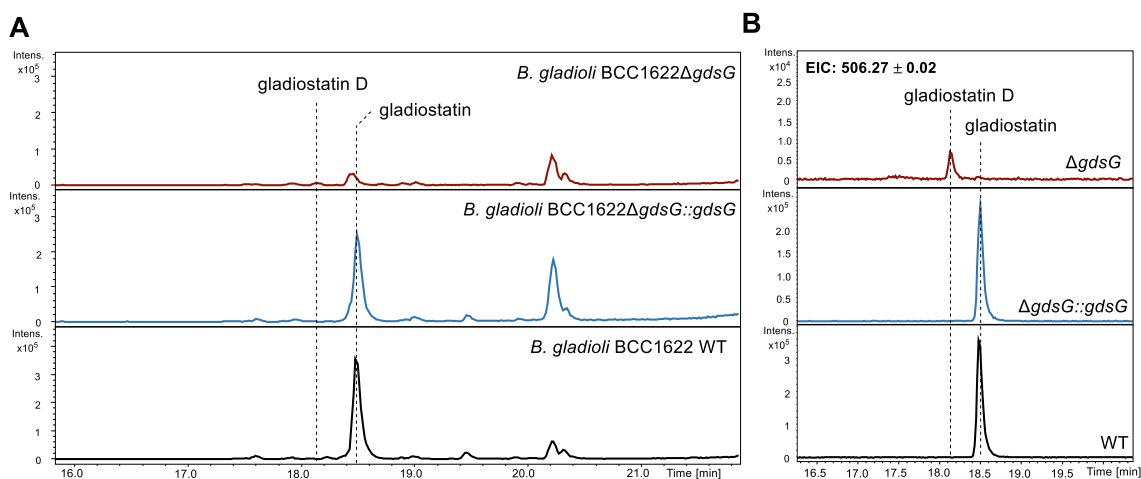


Figure 4.3.6 LC-MS analysis for gladiostatin production after complementing *gdsG* back to the mutant. (A) Base peak chromatograms of *B. gladioli* BCC1622 Δ *gdsG* (red), *B. gladioli* BCC1622 Δ *gdsG*::*gdsG* (blue) and *B. gladioli* BCC1622 wild type (black). (B) Extracted ion

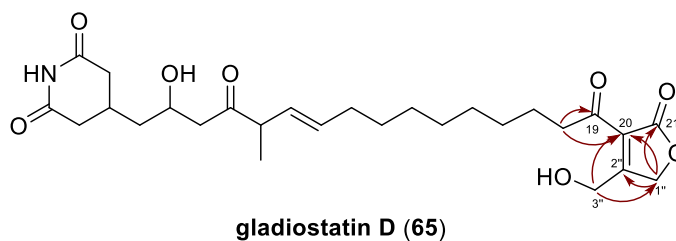
chromatograms of $m/z = 506.27 \pm 0.02$, corresponding to $[M+H]^+$ for gladiostatin, indicated the restoration of gladiostatin production after complementation.

4.3.1.1.1 Isolation and structure elucidation of secondary metabolite produced in $\Delta gdsG$ mutant

The planar structure of the new metabolite, named gladiostatin D (**65**), was elucidated using 1H , ^{13}C , COSY, HSQC and HMBC NMR experiments. Although the production levels of the metabolite were low, a sufficient amount of gladiostatin D was isolated for NMR spectroscopic analysis. Preparative-HPLC purification of the compound and subsequent NMR spectroscopic analysis proved to be difficult due to the instability of this compound. The ^{13}C spectrum of **65** lacked the signals assigned to the methyl group (appended in the C-2'') and the hemiacetal (appended in the C-3'') found in the gladiostatin butenolide. Two new peaks in the ^{13}C spectrum of 71.19 ppm and 60.90 ppm, allowed us to assign a 2-acyl-3-hydroxymethyl butenolide moiety in the molecule. HMBC correlations between the C-3'' protons and C-1''/C-20 and the C-1'' proton and C-2''/C-20 further confirmed the structure. In addition, an HMBC correlation between protons from C-18 to the carbonyl C-19 established the connectivity of the 2-acyl-3-hydroxymethyl butenolide moiety and the rest of the structure (**Table 4.3.1, 7.10.4**). The rest of the NMR assignments for gladiostatin D were analogous to that for gladiostatin²²⁰, establishing that deletion of *gdsG* resulted in no further structural alteration to the rest of the molecule.

2-acyl-3-hydroxymethyl butenolide intermediates have been previously reported to be a key intermediate in the biosynthesis of structurally diverse natural products, such as AHFCAs, and proposed for 2-hydroxyalkyl-4-hydroxy-3-methyl butenolides.^{171,210,211} The structure elucidation of gladiostatin D lead us to propose that *gdsG* is implicated in the 1,3-shift of the phosphate group and/or dephosphorylation to elaborate the AHMB found in the final metabolic product.

Table 4.3.1 ^1H and ^{13}C chemical shifts establishing the AHMB core in gladiostatin D. Key HMBC correlations are shown with red arrows. Complete chemical shift assignment can be found in Appendix.



Position	δ_{C} (ppm)	δ_{H} (ppm)	HMBC
18	42.1	2.99 (t, $J = 7.2$ Hz)	C-19, C-20
19	197.8	-	-
20	125.7	-	-
21	171.7	-	-
1''	71.2	5.15 (s)	C-20, C-21, C-2''
2''	170.2	-	-
3''	60.9	5.02 (s)	C-20, C-1'', C-2''

4.3.1.2 *In vivo* characterisation of *gdsA*

To rule out the possibility of a second phosphatase to be associated in the dephosphorylation step during the formation of the final butenolide in gladiostatin biosynthesis, a targeted in-frame deletion was conducted, targeting the histidine phosphatase appended in the cluster and encoded by *gdsA*.

Applying the same knockout mutagenesis strategies applied for *gdsG* deletion, regions flanking the *gdsA* gene were amplified and cloned into pGPI vector which was then mobilised into *B. gladioli* BCC1622 via tri-parental mating. The successful mutant was identified by PCR analysis (**Figure 4.3.7**). High-resolution LC-MS analysis of the ethyl acetate extracts of the *B. gladioli*

BCC1622 Δ *gdsA* revealed that the production of gladiostatin was maintained after deletion of the gene (**Figure 4.3.7**).

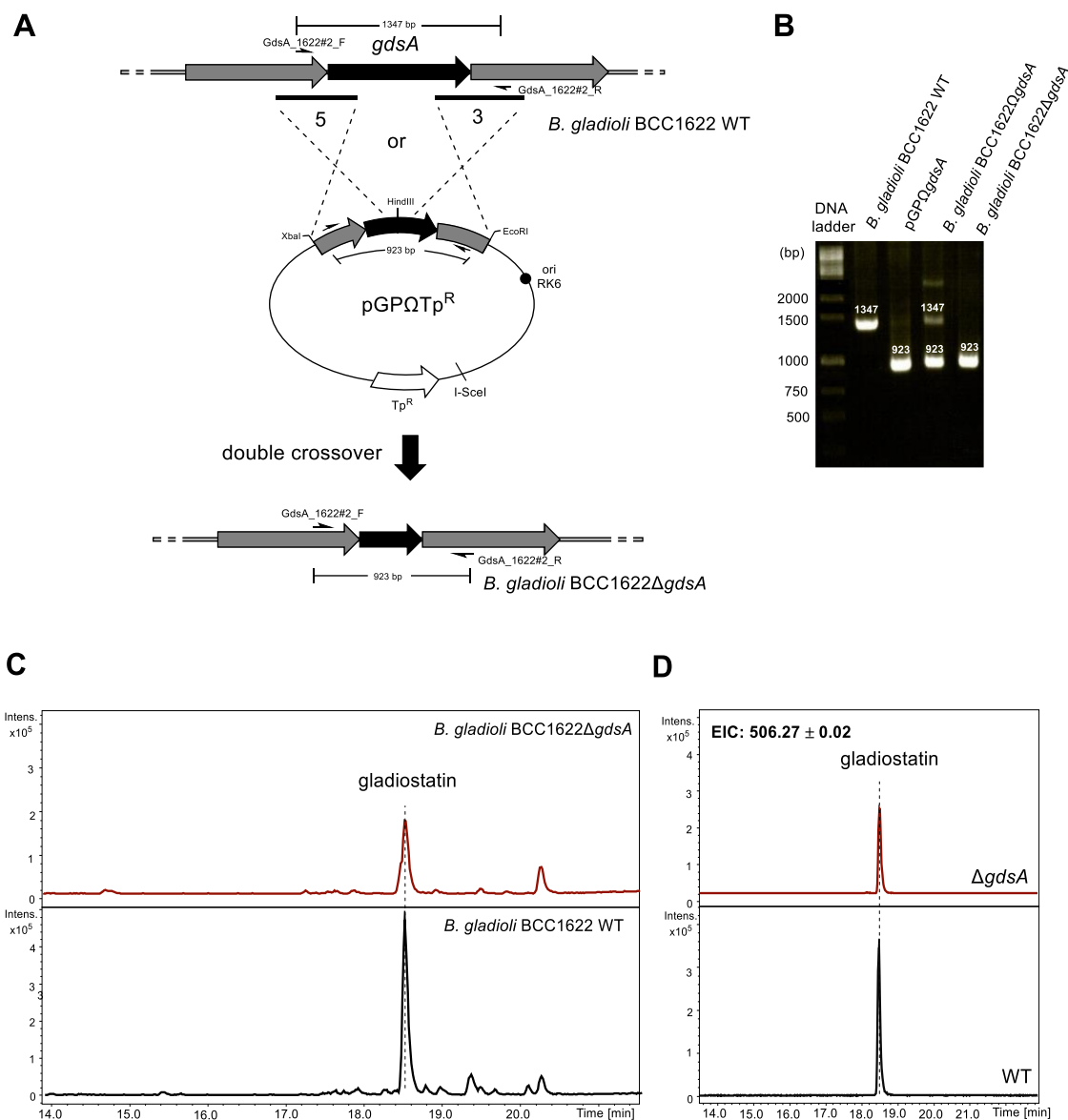


Figure 4.3.7 Construction of *gdsA* mutant and LC-MS analysis of metabolite extracts derived from *B. gladioli* BCC1622 Δ *gdsA* and *B. gladioli* BCC1622 wild type.. (A) Schematic diagram depicting the in-frame gene deletion procedure of *gdsA* in *B. gladioli*. The primers used for colony PCR are shown in **Table 7.9.1**. (B) Agarose gel electrophoresis of the PCR product provide genetic evidence for the successful in-frame gene deletion. The lanes correspond to the PCR product of the *B. gladioli* BCC1622 wild type strain (1347 bp product), pGPI vector containing the flanking regions of *gdsA* 923 bp product), first crossover allele *B. gladioli* BCC1622 Ω *gdsA* mutant and the *B. gladioli* BCC1622 Δ *gdsA* mutant (923 bp product). (C) Base peak chromatogram of Δ *gdsA* mutant strain (shown in red) in comparison to wild

type metabolic profile (shown in black). (D) Extracted ion chromatograms of $m/z = 506.27 \pm 0.02$, corresponding to the $[M+H]^+$ for gladiostatin, showed that gladiostatin was still present after deletion of *gdsA* at lower production levels compared to wild type.

The results shown in Fig. 4.3.7 suggested that GdsA is not essential for the biosynthesis of gladiostatin. This raises the logical question on whether GdsG is the sole enzyme responsible for the final steps involved in the conversion of 3-hydroxymethyl butenolide to the AHMB head. Since no phosphatase enzymes are flanking the biosynthetic gene clusters of SABs and SRBs, a likely scenario is that GdsG and the orthologs P022 and SabP all catalyse a 1,3-shift of the phosphate group and subsequent dephosphorylation during biosynthesis. Another hypothesis could be that a non-specific phosphatase found in cell is capable of performing the dephosphorylation step. The latter can be related to the lower production levels of the metabolite observed in the $\Delta gdsA$ mutant compared to the wild type strain, indicating that the function of GdsA could be partially complemented by other phosphatases located in the cells.

4.3.2 *In vitro* reconstitution of the AHMB formation steps during gladiostatin biosynthesis

4.3.2.1 Overproduction and purification of recombinant His₆-GdsA and His₆-GdsG proteins

In an effort to complement the genetic evidence shown previously, recombinant protein constructs of pET28a_GdsA and pET28a_GdsG containing a 6xHis tag at the N-terminus were created using traditional cloning procedures, as detailed in **section 7.5**. Recombinant protein expression constructs were used to transform *E. coli* BL21 (DE3) and the corresponding protein was overproduced using IPTG-induced overexpression. Purification of the recombinant proteins was accomplished by Ni-NTA column. SDS-PAGE analysis indicated a single protein band with the expected molecular mass of each protein and ESI-Q-TOF-MS measurement of GdsA and GdsG revealed masses of 27981.00 and 29341.69, respectively, which are in good agreement with the theoretical masses (**Figure 4.10**).

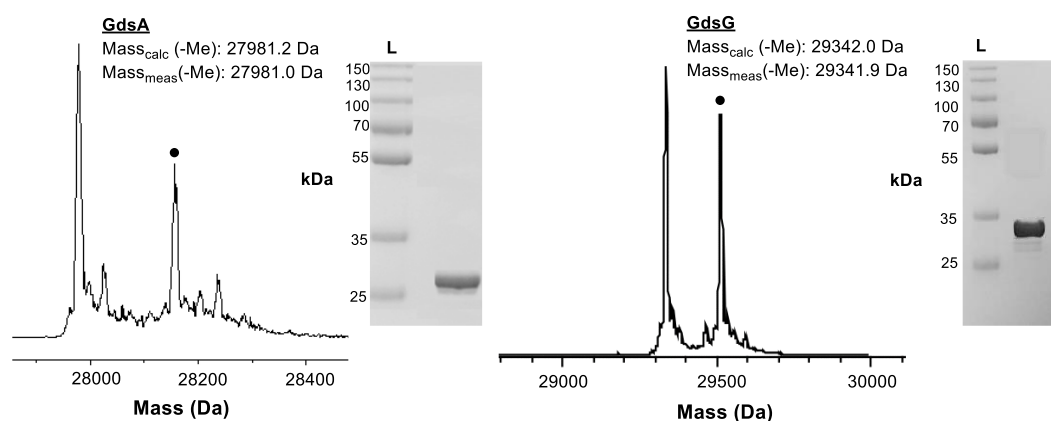


Figure 4.10 Deconvoluted ESI-MS spectra and 10% SDS-PAGE analysis of purified N-His₆-GdsA and N-His₆-GdsG. Peaks annotated with a black circle shows spontaneous N-terminal gluconoylation of the His-tag with a mass shift of + 178 Da.¹⁸⁸

4.3.2.2 Use of *p*-nitrophenyl phosphate for direct determination of phosphatase activity

The ability of GdsA and GdsG to catalyse dephosphorylation was examined using a colorimetric assay with *p*-nitrophenyl phosphate (*p*NPP) (**66**).²³¹ This assay employs the chromogenic *p*NPP as a non-specific substrate which upon dephosphorylation generates *p*-nitrophenol (**67**) (Figure 4.3.8). **67** is subsequently converted to its 4-nitrophenolate form, **68**, that produces an intense yellow colour.

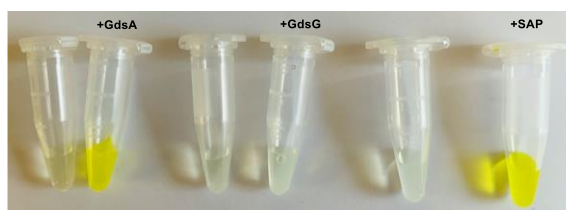
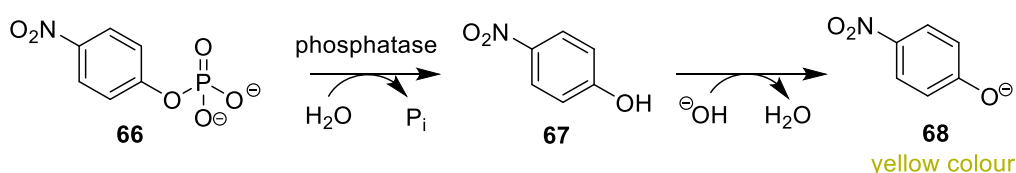


Figure 4.3.8 Colorimetric assay to detect phosphatase activity using *p*NPP. The phosphatases remove the phosphate group to generate *p*-nitrophenol, which is deprotonated under alkaline conditions to produce *p*-nitrophenolate that has strong absorption at 405 nm. Commercially available shrimp alkaline phosphatase (SAP) was used as a control.

Incubation of *p*NPP with GdsA yielded a colour change in the reaction solution (Figure 4.3.8). This colour change was also observed in the control reaction that was conducted in parallel, using a commercially available shrimp alkaline phosphatase (SAP). In contrast, upon incubation of *p*NPP with GdsG, no dephosphorylation was detected over the time-frame of the experiment. This pattern was repeated in all reactions tested in a range of pH values.

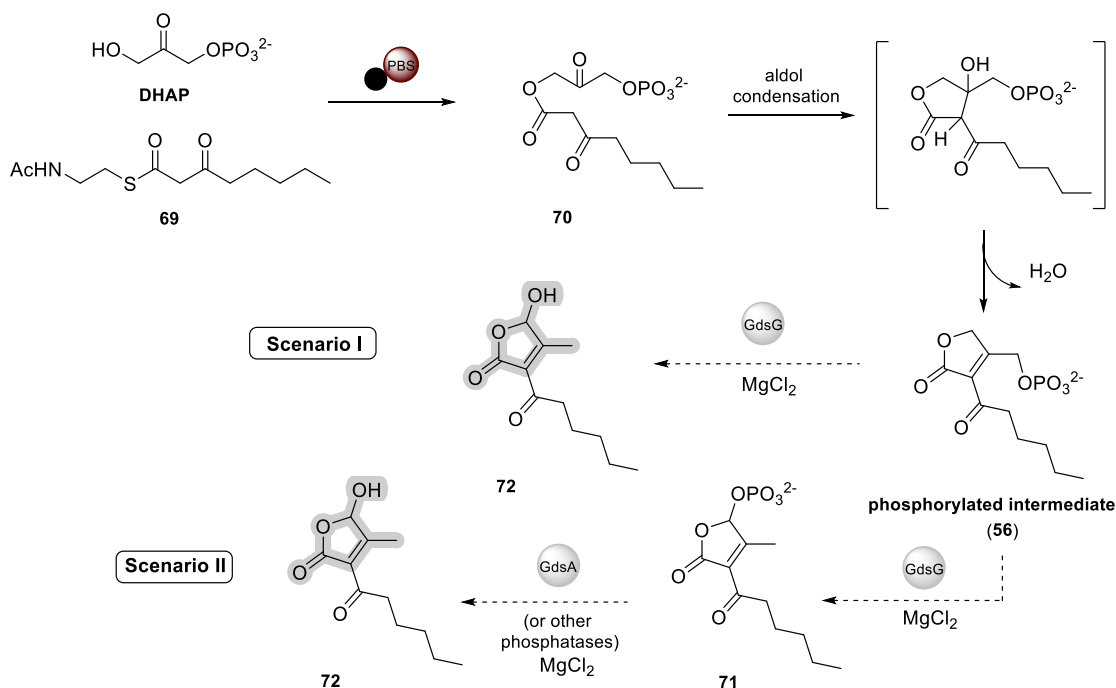
Although *p*NPP is not a physiological substrate, it is an efficient way to determine non-specific dephosphorylation activity. The use of a generic substitute which is not related to the natural *in vivo* substrate could therefore explain the absence of dephosphorylation activity observed in the reaction with GdsG. On the other hand, this assay provided additional information regarding

the activity of GdsA that is in agreement with our previous observations, indicating that GdsA could indeed play a role in the dephosphorylation of the phosphorylated butenolide intermediate, which can be complemented by phosphatases with a broad substrate specificity when the GdsA phosphatase is absent.

4.3.2.3 Design of an enzymatic assay to reconstitute formation of the 2-acyl-4-hydroxy-3-methyl butenolide *in vitro*

To reconstitute the enzymatic steps involved in the formation of the tri-substituted butenolide, a coupled assay was designed, utilising the PBS chain release mechanism reported previously²¹⁸, combined with the enzymes proposed to catalyse the rearrangement and dephosphorylation steps in gladiostatin biosynthesis (**Scheme 4.3.2**). The N-terminal His₈-fused Mod9_ACP-PBS di-domain (the purified construct was kindly provided by Dr. Matthew Jenner) was used to condense the NAC thioester of 3-ketooctanoate (**69**), a simple mimic of the presumed fully assembled polyketide (the synthetic substrate was kindly provided by Dr. Shanshan Zhou) with DHAP.^{214,218} Upon condensation, the 2-oxo-3-(phosphonoxy)propyl 3-oxooctanoate intermediate **70** is proposed to undergo spontaneous intramolecular aldol condensation that furnishes the phosphorylated intermediate **56**. The function of GdsA and GdsG was then tested by incorporating each of the enzymes to the reaction and monitoring it by HR-LCMS. The same reaction was conducted using boiled GdsA and GdsG as a control. Based on the biosynthetic proposal of gladiostatin and the *in vivo* experimental evidence obtained, two possible scenarios regarding the outcome of this enzymatic assays were proposed; (i) GdsG has a dual function, performing both 1,3 rearrangement of the phosphate group and dephosphorylation of the intermediate **56** to yield the final butenolide **70**; (ii) GdsG catalyses the 1,3-shift of the phosphate group which is then removed by GdsA or phosphatases encoded by genes appended nearby the cluster. Shrimp alkaline phosphatase has been shown to dephosphorylate the intermediate **56** yielding the butenolide product **57** and it was therefore included in the studies for direct comparison. Since most of phosphatase enzymes utilise a divalent cation as co-factor to coordinate the phosphate group within the catalytic pocket, Mg²⁺

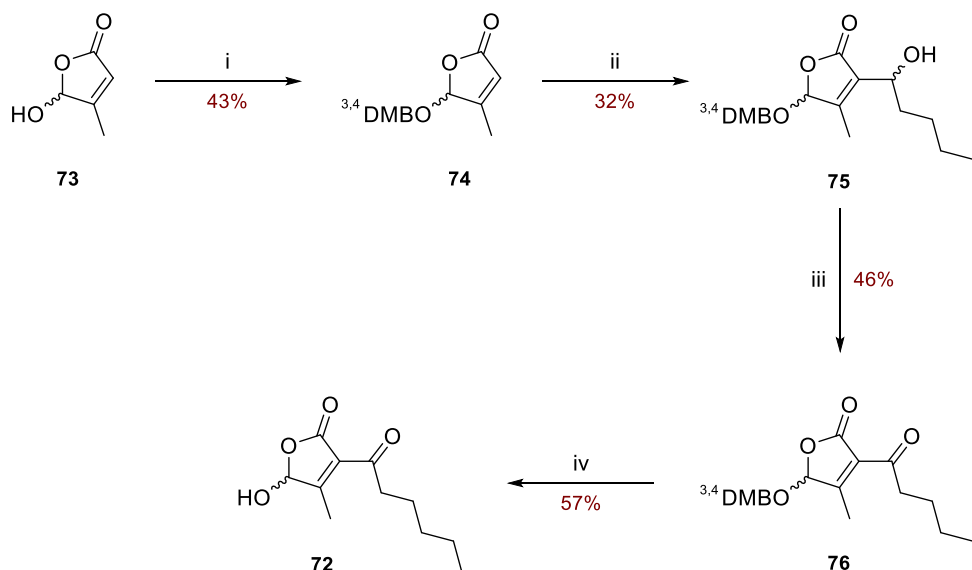
was used as a cofactor in the assays.²³² Two synthetic standards, 2-acyl-3-hydroxymethyl butenolide **57** (synthesised by Dr. Shanshan Zhou) and 2-acyl-4-hydroxy-3-methyl butenolide **72** were used as reference for the characterisation of the product GdsA/GdsG-catalysed reaction.



Scheme 4.3.2 Overview of the enzymatic assay used to investigate the formation of the 2-acyl-4-hydroxy-3-methyl butenolide *in vitro*.

4.3.2.4 Synthesis of 3-hexanoyl-5-hydroxy-4-methylfuran-2(5H)-one as an authentic standard (**72**)

In the course of validating the proposed structure of the dephosphorylated butenolide product, an authentic standard of gladiostatin was chemically synthesised to use as an authentic standard for our biochemical assays. Here, **72** was synthesised from a commercially available 5-hydroxy-4-methylfuran-2(5H)-one (**73**) using a synthetic route outlined in **Scheme 4.3.3**.



Scheme 4.3.3 Overview of the route used to synthesise 71 authentic standard. *Reagents and conditions:* (i) ^{3,4}DMBOH (1.1 eq.), amberlyst-15 (10% w/w), CH₂Cl₂, reflux, 48 h; (ii) *i*-Pr₂NH (1.2 eq.), *n*-BuLi (1.2 eq.) in THF 0 °C to -78 °C, **73** then hexanal (1.2 eq.) 5 h; (iii) Dess-Martin periodinane (1.1 eq.), CH₂Cl₂, rt, 2 h; (iv) PIFA (2.0 eq.), CH₂Cl₂, rt, 3 h.

The hydroxy group of **73** was protected using 3,4-dimethoxy benzyl alcohol (^{3,4}DMBOH) to yield product **74** under acidic conditions. The lithiated veratryl-protected methyl furanone generated *in situ* from the protected furanone was subsequently coupled with hexanal to afford **75** as a diastereomeric mixture. Although electron deficient furanones are prone to lithiation at the 5-position due the decrease in pK_a of H-5, it was suggested that the bulkiness of the 3,4-dimethoxybenzyl protecting group prevented the formation of undesired by-products. Oxidation of the hydroxy group appended in **75** proceeded using Dess-Martin periodinane to produce **76**. Completion of the synthesis was accomplished by mild oxidative cleavage of DMB-group using phenyliodine bis(trifluoroacetate) (PIFA) to give **72**.

4.3.2.5 *In vitro* characterisation of the 1,3-rearrangement of the phosphate group and dephosphorylation steps to form the trisubstituted butenolide

The catalytic activity of GdsA was first investigated by addition of the enzyme into the PBS-catalysed reaction which forms the phosphorylated intermediate **56**. Reactions were quenched using methanol and the supernatants were analysed using UHPLC-ESI-Q-TOF-MS in positive ion mode. First, the PBS domain-catalysed condensation reaction between DHAP and β -keto thioester NAC substrate **69** was conducted, and subsequent addition of GdsA resulted a peak with the same retention time as the dephosphorylated product **57** derived from PBS-catalysed condensation and dephosphorylation, catalysed by shrimp alkaline phosphatase (SAP) (**Figure 4.3.9**). This peak corresponds to ion species of $m/z = 235.0936$, which correlating to the $[M+Na]^+$ of the butenolide carbinol **57**. This peak was absent from the reaction using boiled GdsA, confirming that **57** is the outcome of GdsA-catalysed dephosphorylation of the intermediate **56** (**Figure 4.3.9**). This was further established with comparative analysis of the retention times observed for the synthetic standard **57** and **72** and the similar MS/MS fragmentation patterns observed (**Figure 4.3.9**). The *in vitro* data were in good agreement with the *in vivo* studies, establishing that GdsA catalyses a non-specific dephosphorylation of substrates bearing a phosphate group and absence of the function of this enzyme can be partially complemented by other phosphatases present in the cells.

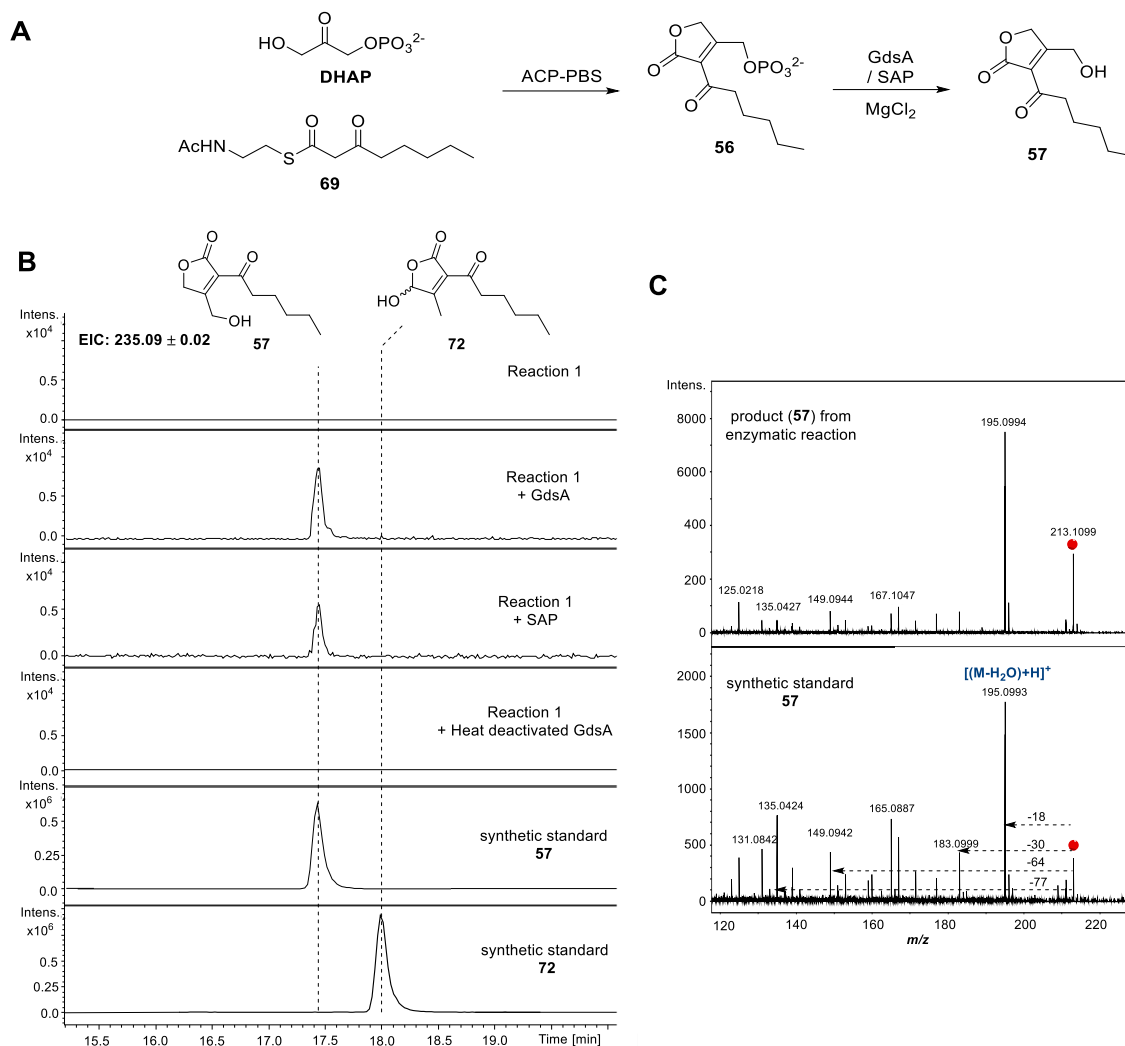


Figure 4.3.9 Enzymatic reaction cascade performed by PBS and GdsA. (A) Schematic representation of condensation of DHAP and the NAC thioester of 3-ketooctanoate by the purified recombinant GdsF ACP-PBS di-domain (denoted as Reaction 1) and dephosphorylation of the resulting phosphorylated butenolide **56** using GdsA or SAP. (B) Extracted-ion chromatograms at $m/z = 235.09 \pm 0.02$ (corresponding to $[M+Na]^+$) for butenolide **57** from UHPLC-ESI-Q-TOF-MS comparisons of the enzymatic reaction cascade and synthetic standards of **57** and **72**. The heat-deactivated GdsA was used as a control reaction. (C) Comparison of MS/MS spectra for the product of the enzymatic reaction cascade and the synthetic standard of butenolide **57** ($[M+H]^+ = 213.11$).

After establishing that GdsA is performing dephosphorylation of substrate **56**, the HAD-like phosphatase GdsG was studied *in vitro* for its role in the formation of the 2-acyl-4-hydroxy-3-methyl butenolide. Here, GdsG was added to the reaction with DHAP, β -keto NAC thioester substrate and GdsF ACP-PBS which is known to form the phosphorylated butenolide

intermediate **56** (**Figure 4.3.10**). Intriguingly, a peak of $m/z = 235.09 \pm 0.02$ consistent with the $[M+Na]^+$ ion for the dephosphorylated species of either **57** or **72** was not observed in the UHPLC-ESI-Q-TOF-MS analysis. Instead, a new peak was observed with an $m/z = 293.0788$ and 315.0602 , corresponding to the $[M+H]^+$ and $[M+Na]^+$ ion species with deduced molecular formula of $C_{11}H_{17}O_7P$ (**77**) (**Figure 4.3.10**). Peaks with the same mass to charge ratio of these ion species were also observed in the PBS-catalysed reaction before addition of the GdsG as well as in the reaction that was performed with the boiled GdsG as a control. However, in both cases the eluents presented earlier retention times (**Figure 4.3.10**). A negative control was also introduced to the study where the ACP-PBS di-domain was omitted from the reaction.

The data suggested that GdsG functions as a phosphato-mutase, transferring the phosphate group from the methyl group appended in C-3 to C-4 of the butenolide moiety, hence the different retention times observed between the phosphorylated species before and after incubation of GdsG in the reaction. The analysis also provided clear evidence that GdsG cannot perform dephosphorylation. Building on this point, the PBS/GdsG reaction was incubated with phosphatase GdsA and SAP to investigate potential formation of the dephosphorylated product **72**. Indeed, after incubation of intermediate **77** with GdsA or SAP generated a peak with the same retention time as authentic standard **72** (**Figure 4.3.11**).

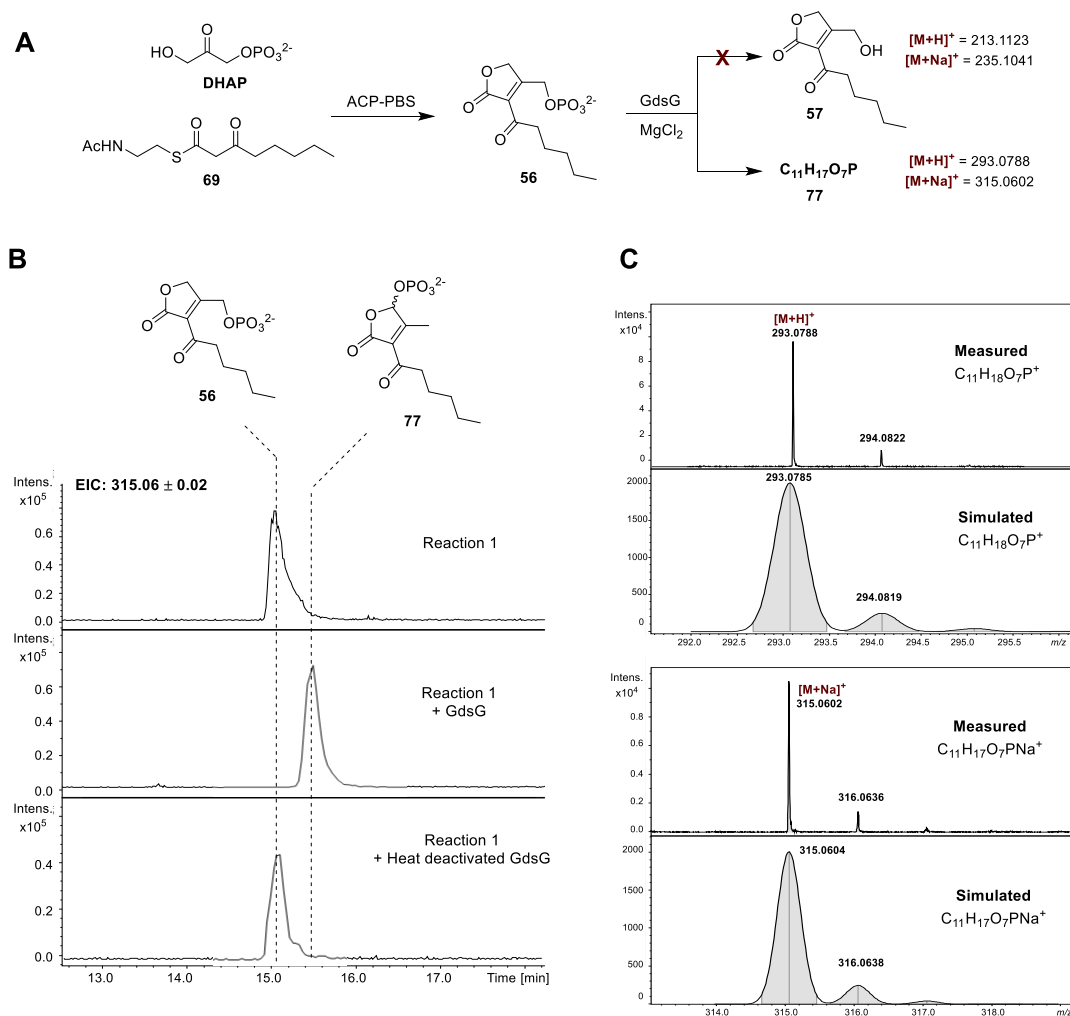


Figure 4.3.10 Enzymatic reaction cascade performed by PBS and GdsG. (A) Schematic representation of condensation reaction of DHAP and the NAC thioester of 3-ketooctanoate by the purified recombinant GdsF ACP-PBS di-domain (denoted as Reaction 1) and incubation of GdsG. (B) Extracted-ion chromatograms at $m/z = 315.06 \pm 0.02$ (corresponding to $[M+Na]^+$ for the proposed butenolide **77**). The heat-deactivated GdsG was used as a control reaction. (C) Measured (top) and simulated (bottom) spectra for the $[M+H]^+$ and $[M+Na]^+$ ion corresponding to molecular formula of $C_{11}H_{17}O_7P$.

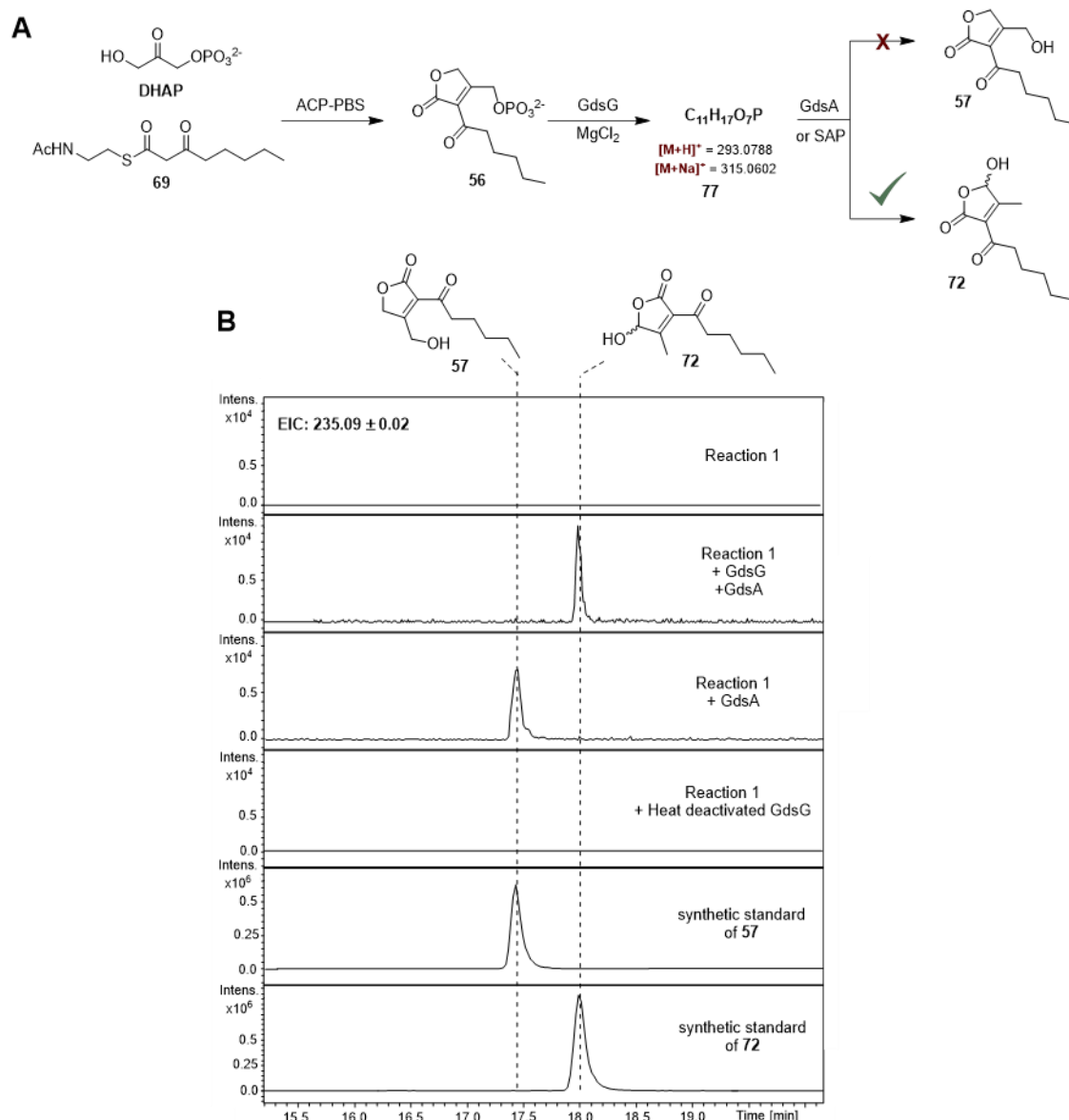


Figure 4.3.11 *In vitro* reconstitution of the 2-acyl-4-hydroxy-3-methyl butenolide assembly. PBS-catalysed condensation of DHAP and NAC β -ketothioester (**69**), followed by an unprecedented 1,3-shift of a phosphate group by GdsG, which after dephosphorylation yields the unusual trisubstituted butenolide of **72**.

The products derive from the enzymatic reaction cascades performed in this study provided insights with regards to the sequence by which these enzymes are operating during the biosynthesis of gladiostatin. The proposed scenario II was in concordance to the findings of this study, indicating that GdsG is not bifunctional and precedes the dephosphorylation step catalysed by GdsA.

4.3.3 Gene complementation studies on $\Delta gdsG$ mutant using *sabP* and *p022* orthologs

After the successful *in vitro* reconstitution of the enzymatic steps involved in furnishing the 2-acyl-4-hydroxy-3-methyl butenolide during the biosynthesis of gladiostatin, we aimed to investigate whether this type of catalysis is applied to systems from *Streptomyces* species found to assemble secondary metabolites with the same butenolide moiety.

Since the *gdsG* homologs, *p022* and *sabP*, are proposed to encode enzymes that catalyse the analogous reaction to deliver the structural identical tri-substituted butenolide, we postulated that complementation of these homologous enzymes into *B. gladioli* BCC1622 $\Delta gdsG$ mutant might restore production of the original metabolite, gladiostatin. Utilising the L-arabinose-inducible vector pMLBAD as described in **section 7.9.3**, the deletion mutant was complemented by *in trans* expression of *sabP* and *p022*. The successful tri-parental mating was using PCR analysis and the production levels of the secondary metabolite was investigated through HPLC-MS analysis conducted to ethyl acetate extracts of BSM media supplemented with 0.2% L-arabinose. EICs eluting earlier at $m/z = 506.27$, corresponding to the protonated adduct of gladiostatin (**49**) revealed the partial restoration of gladiostatin production in $\Delta gdsG$ extracts after *in trans* expression of the orthologs *sabP* and *p022* (**Figure 4.15**).

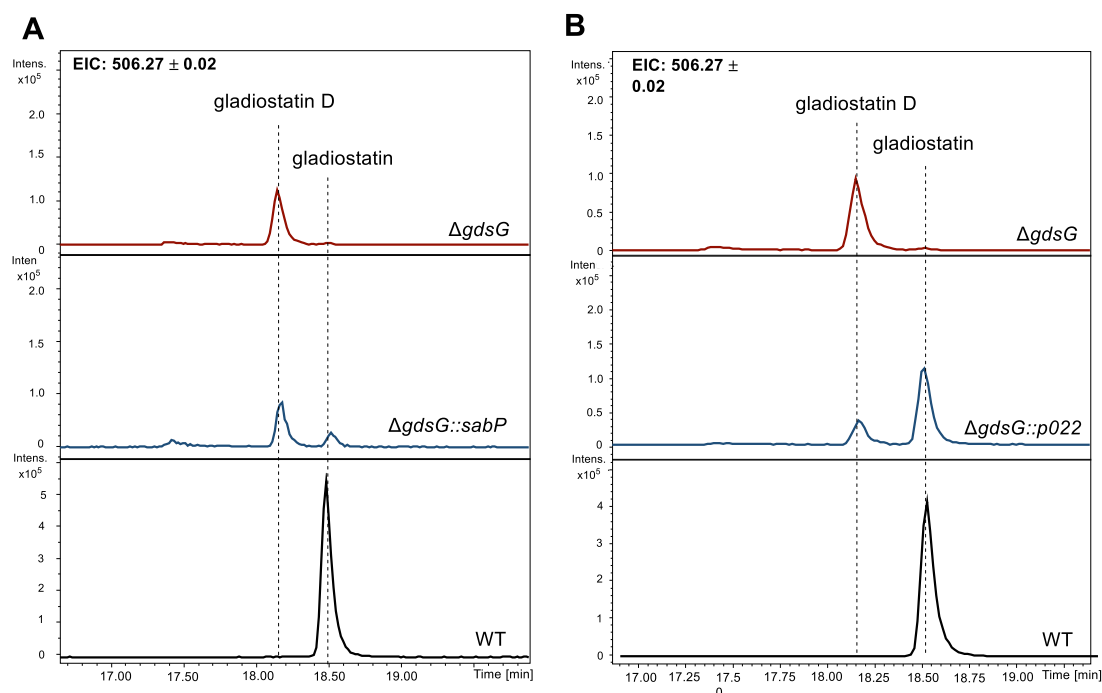


Figure 4.3.12 LC-HRMS analyses of metabolites produced from *B. gladioli* BCC1622 Δ gdsG after complementation with the orthologs SabP and p022. (A) EICs of *B. gladioli* BCC1622 Δ gdsG (red), *B. gladioli* BCC1622 Δ gdsG::sabP (blue), *B. gladioli* BCC1622 wild type (black), and (B) EICs of *B. gladioli* BCC1622 Δ gdsG (red), *B. gladioli* BCC1622 Δ gdsG::p022 (blue), *B. gladioli* BCC1622 wild type (black) for $m/z = 506.27 \pm 0.02$, corresponding to $[M+H]^+$ species for both gladiostatin and gladiostatin D, indicating the partial restoration of the original metabolite after in trans expression of sabP and p022.

The inability of *sabP* and *p022* to fully restore production of **49** *in vivo*, in comparative production levels to these observed in the wild type, could be explained by the difference in the species from which these genes derive from as well as substrate preference. The HPLC-MS also provides a snapshot of the substrate tolerance of these two enzymes. P022, which is proposed to catalyse the phosphate rearrangement on 3-methyl phosphorylated butenolide substrates containing C₁₀ and C₁₁ alkyl chains, yielded increased production levels of **49** compared to SabP which is proposed to accept phosphorylated butenolides substituted with shorter alkyl side chains as native substrate.

subsequent dephosphorylation catalysed by either GdsA or SAP to furnish **72** was investigated through UHPLC-ESI-Q-TOF-MS (**Figure 4.3.13**). Incubation of SabP and P022 into the PBS-catalysed reaction revealed a peak of $m/z = 293.08$ which is indicative of the observed peak corresponding to the rearranged phosphorylated intermediate **77** in instances where GdsG is introduced in the reaction (**Figure 4.3.14**). The difference in retention time between the two phosphorylated intermediates is also consistent with previous studies on GdsG. Upon addition of GdsA or SAP an analyte with $m/z = 235.09$ corresponding to the protonated adduct at the same retention time to that of synthetic standard **72**.

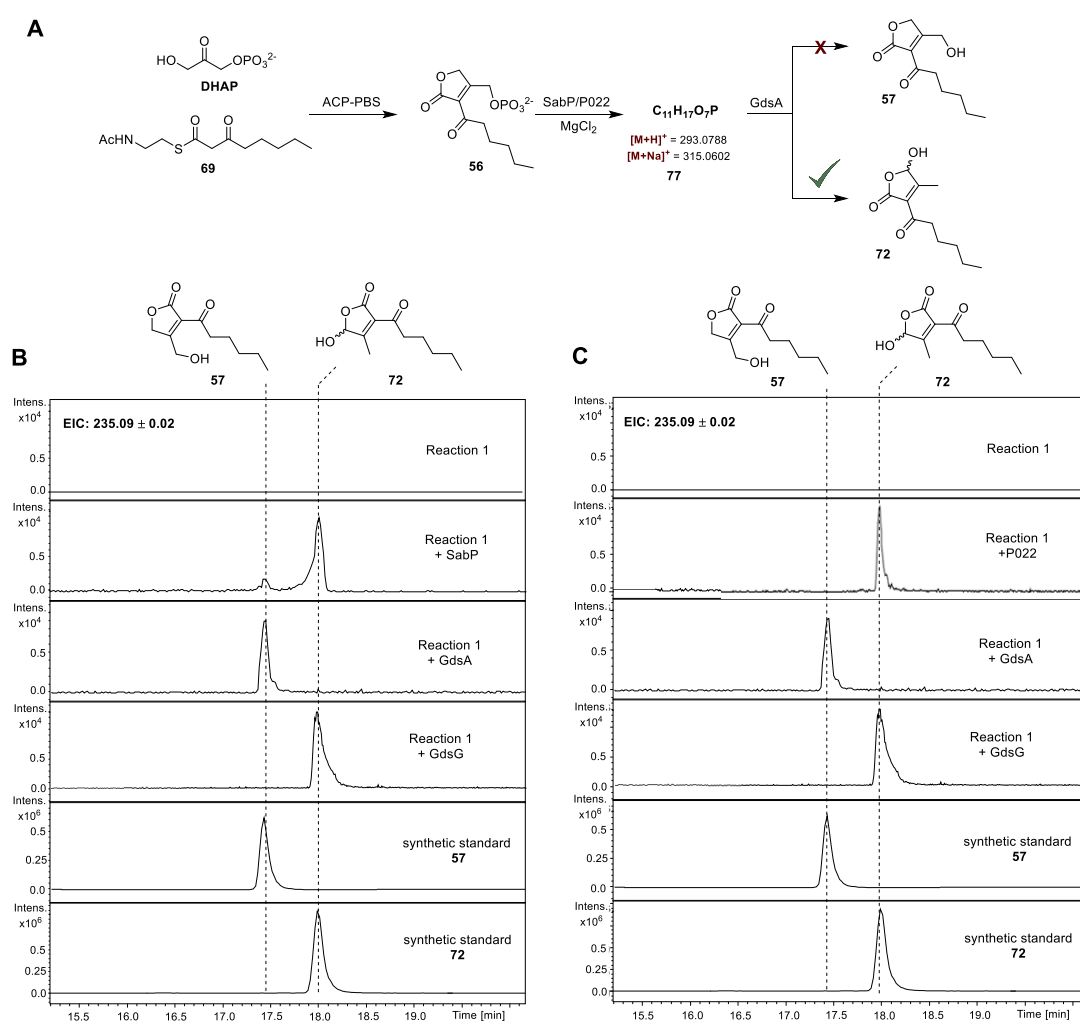


Figure 4.3.14 Enzymatic reaction cascade performed by PBS and GdsG homologues SabP and P022. (A) Schematic representation of condensation reaction of DHAP and the NAC thioester of 3-ketooctanoate by the purified recombinant GdsF ACP-PBS di-domain (denoted as Reaction 1) and

incubation with SabP/P022. **(B)** Extracted-ion chromatograms at $m/z = 315.06$ and comparison to authentic standards **57** and **72**. Heat-inactivated SabP/P022 was used as a control reaction.

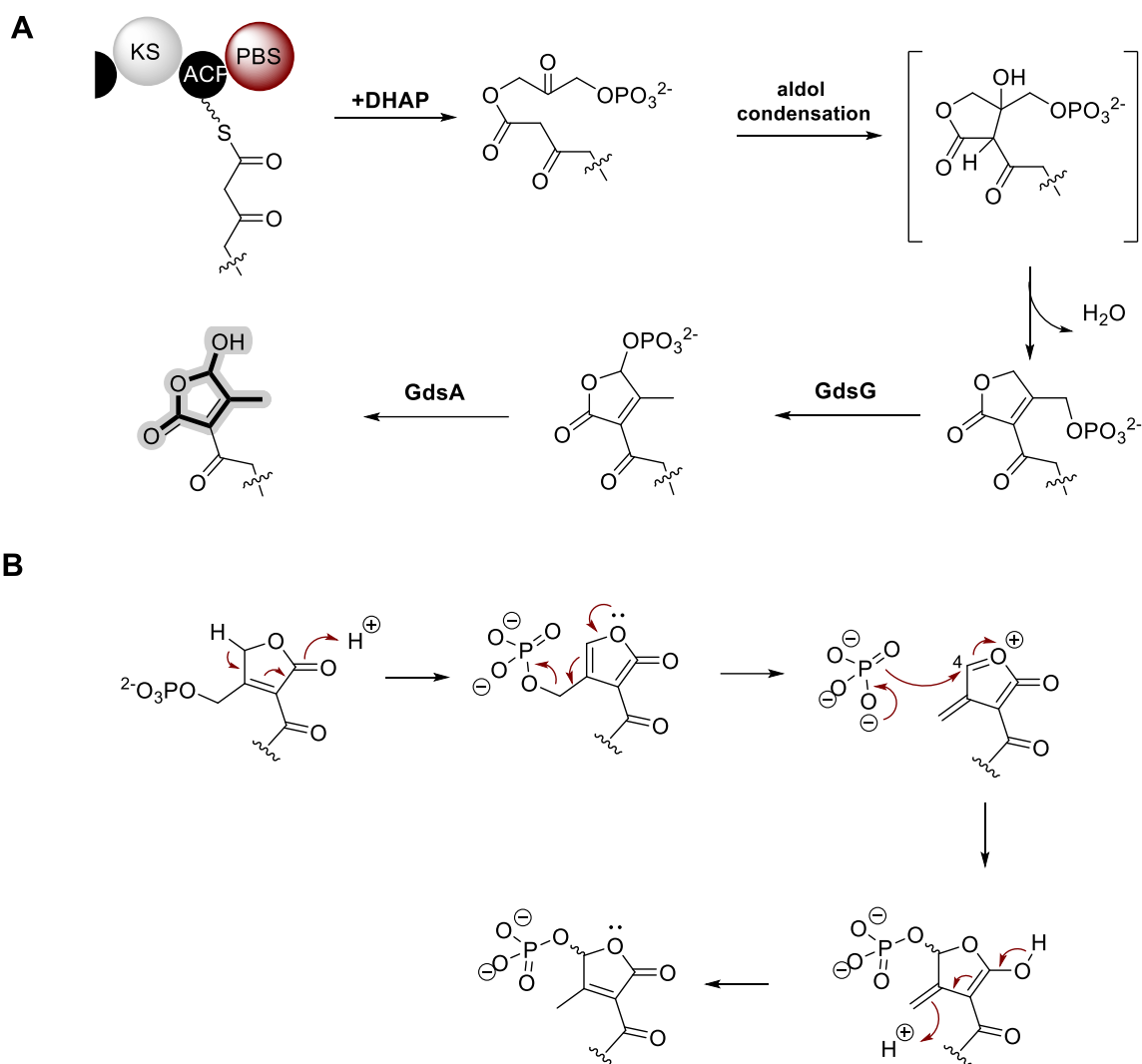
The aforementioned results are consistent with data obtained during the *in vitro* characterisation of GdsG, demonstrating that SabP and P022 are capable of catalysing 1,3 rearrangement of the phosphate group in a similar manner observed for GdsG.

4.4 Conclusions

This study presents insights into the formation of the 2-acyl-4-hydroxy-3-methyl butenolide unit of gladiostatin which is closely related to the butenolide core of the 2-acyl-4-hydroxy-3-methyl butenolide signalling molecules, SABs and SRBs, found to regulate production of antibiotics in *Streptomyces*.^{170,171,220} In addition to genes encoding AfsA-like enzymes, which are responsible for condensation of the final polyketide intermediate tethered to the upstream ACP with DHAP, comparative analysis of *gds*, *srb* and *sab* biosynthetic gene clusters indicated they all contained genes encoding conserved HAD-like phosphatase enzyme. It was therefore postulated that this enzyme could facilitate the required post-assembly modifications from the phosphorylated butenolide **56** to yield the 2-acyl-4-hydroxy-3-methyl butenolide substructure observed in gladiostatin. Along with this assumption, a second putative phosphatase was found to be encoded by the BGC of gladiostatin with a potential of catalysing dephosphorylation of the intermediate **56**. Both phosphatases were subjected to characterisation of their role in the biosynthesis through *in vivo* genetic engineering and *in vitro* biochemical studies. A $\Delta gdsA$ mutant revealed that the enzyme is not essential for the assembly of the secondary metabolite gladiostatin, though the production levels of gladiostatin in the mutant strain was lower indicating that the gene is partially complemented by non-specific phosphatases found in the cell. In-frame deletion of *gdsG* abolished production of gladiostatin showing it has a crucial role in the assembly. Interestingly, a new peak arose with the same *m/z* of that observed for gladiostatin. Structural elucidation of the gladiostatin derivative confirmed the presence of a 2-acyl-3-hydroxymethyl butenolide unit. This suggested that *gdsG* encodes an enzyme which is responsible for the rearrangement of the phosphate group and elimination of this enzyme from the biosynthetic pathway results in accumulation of the desphosphorylated species. The *in vivo* data were supported by *in vitro* biochemical characterisation of the two enzymes. Investigation into the function of these two putative phosphatases, revealed that GdsG does not perform dephosphorylation of generic substrates in contrast to GdsA which was able to catalyse dephosphorylation of *p*NNP. Building upon previous work on the GdsF-PBS chain release mechanism, the biosynthesis of the 2-acyl-4-hydroxy-3-methyl butenolide of gladiostatin has been fully reconstituted *in vitro*.

Gene complementation studies on *B. gladioli* BCC1622 Δ *gdsG* mutant using homologous genes from *sab* and *srb* BGC, partially restored gladiostatin production. The analysis provided a preliminary indication of the substrate tolerance of SabP and P022 demonstrated by the ratio of gladiostatin to the 2-acyl-3-hydroxymethylbutenolide shunt metabolite. P022 which is proposed to perform catalysis using butenolide substrates with shorter alkyl chains presented low production level of the secondary metabolite in comparison to SabP which favour longer alkyl chains.

Using the aforementioned data, a molecular mechanism for the installation of the 2-acyl-4-hydroxy-3-methyl butenolide core in the biosynthesis of gladiostatin could be proposed as outlined in **Scheme 4.4.1**. The mechanism of the phosphato-mutase, GdsG, is proposed to occur through a ping pong mechanism that involves proton transfer that affords a furan intermediate which undergoes dephosphorylation. This is followed by addition of phosphate group at C-4 and subsequent proton transfer to yield the rearranged phosphorylated intermediate (**Scheme 4.4.1**). To our knowledge, this is the first example of a HAD-like phosphatase able to catalyse this type of reaction. Additional structural studies on GdsG and GdsG-homologues will need to be carried out in order to identify key catalytic residues.



Scheme 4.4.1 Proposed pathway for of 2-acyl-4-hydroxy-3-methyl butenolide biosynthesis demonstrated in this study. (A) Condensation of DHAP with the fully-assembled polyketide intermediate is catalyzed by AfsA-like enzyme PBS (previous work). After this condensation step, nonenzymatic intramolecular aldol condensation yields the 3-hydroxymethyl intermediate. GdsG-catalysed phosphate rearrangement in the butenolide unit and subsequent dephosphorylation by GdsA generates the tri-substituted butenolide. **(B)** Proposed mechanism for the 1,3 phosphate rearrangement catalysed by GdsG.

Chapter 5

Comparison of the gladiostatin biosynthetic gene cluster in numerous B. gladioli isolates reveals a novel mechanism for trans-AT polyketide synthase evolution

5.1 Introduction

5.1.1 Engineering of modular PKSs inspired by evolution

Since the discovery of assembly-line PKSs, there has been an enduring interest in reprogramming these megasynthases to generate novel derivatives.²³³ Despite the substantial progress in establishing domain structure-function relationships, genetic manipulation of these systems has proven not to be a trivial task. Reordering domains and modules as well as module swapping results in catalytically compromised assemblies.^{234–236} Electron cryo-microscopy structures of a full-length pikromycin PKS module revealed that interdomain contacts are essential for the function of the module and therefore explaining why exchange of catalytic domains yields poor turn overs for most hybrid assemblies.¹²⁸

Efforts to engineer PKSs in the lab using rational and combinatorial approaches proved to be not efficient, hence tactics towards mimicking the natural evolution processes were pursued. Evolutionary studies have been carried out on modules and individual domains of modular PKSs.^{55,237,238} A phylogenetic tree of KS domains from *cis*-AT PKSs follows the phylogeny of their hosts organisms, with higher sequence identities within synthases and to a lesser degree between different assembly lines within the species (**Figure 5.1.1**).⁵⁵ On the other hand, KS domains from *trans*-AT PKSs do not fall into the same clade with KS domains of the same BGC. Instead they form substrate-specific clades (**Figure 5.1.1**).¹²⁴ This finding offers the potential for prediction of PKS product based on KS specificities in *trans*-AT systems. Therefore it was suggested that *cis*- and *trans*-AT PKSs have evolved through different evolutionary pathways. Due to the high levels of sequence similarity between KS domains within the same PKS and the phylogenetic clustering it is speculated that the multi-modularity of *cis*-AT PKSs arose through duplication of entire lengths of DNA encoding complete modules.^{55,125} On the contrary, *trans*-AT PKSs are believed to be a mosaic of DNA fragments assimilated from horizontal gene transfer events.²³⁹

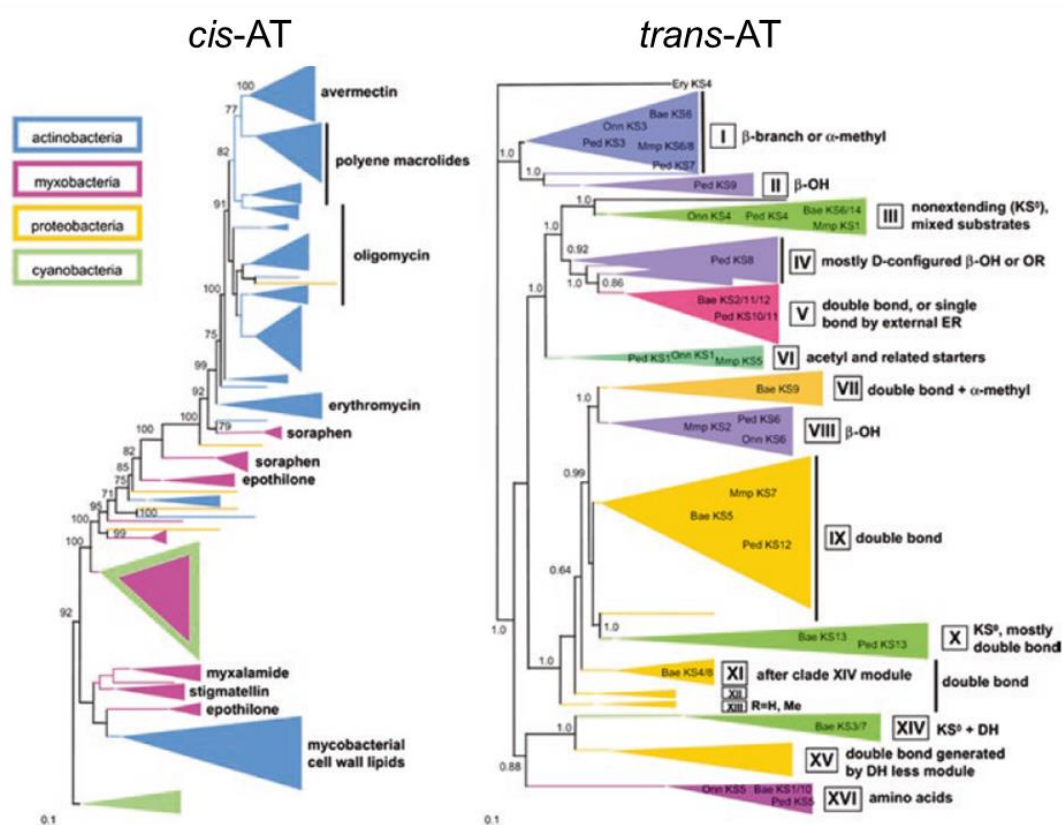


Figure 5.1.1 Phylogenetic analysis of KS domains from *cis*-AT and *trans*-AT PKS systems. Reproduced from ref.¹²⁵

Based on the existing evolutionary model that considers gene duplication as the driving force for *cis*-AT PKS evolution, modules are considered to be the functional unit of duplication.^{235,236} A recent study on the evolutionary relationship of 4 PKS gene sets that direct the biosynthesis of structurally related aminopolyol polyketides suggested KS domains in *cis*-AT PKSs co-evolve with domains in the upstream module, rather than their own module.⁶³ This prompted an alternative definition of the unit of duplication (**Figure 5.1.2**).^{63,64} Swaps of both modules and co-evolved units have been used for successful PKS engineering. In the aureothin and neo-aureothin synthases, splicing between the KS and AT domains was more productive than insertions/deletions between the ACP and KS domains. However, the titers were still not comparable to those for the wild-type system (**Figure 5.1.2**).^{240,241}

Homologous recombination occurs between DNA sequences with high sequence similarity²⁴² and such splice sites have therefore been proposed to be optimal for PKS engineering. This

was first assessed computationally by analysing recombination between PKSs in *Streptomyces* sp., which predicted that most recombination events occurred between pairs of KS domains.²⁴³ The viability of PKS engineering via homologous recombination was later demonstrated by recombining PKS genes from the pikromycin and erythromycin BGCs in *Saccharomyces cerevisiae* (**Figure 5.1.2**).²⁴⁴ Despite the observation of macrolactone production *in vitro* using the resultant chimeric modules, only low levels of production of the corresponding metabolites were observed when the chimeras were introduced into the native pikromycin producer *S. venezuelae*, suggesting they are unstable. More recently, the highly repetitive nature of PKS genes has been exploited to create chimeric PKSs by repair of double strand breaks via homologous recombination in *Streptomyces* sp.²⁴⁵ Using this methodology, which mimics a plausible mechanism for PKS evolution, 17 9 chimeras of the rapamycin and tylosin PKS, respectively, were created. Splicing points were mainly located within KS, AT and ACP domains (**Figure 5.1.2**). The majority of chimeras were functional and produced the expected natural product analogues at titers comparable to those of the wild type strain.

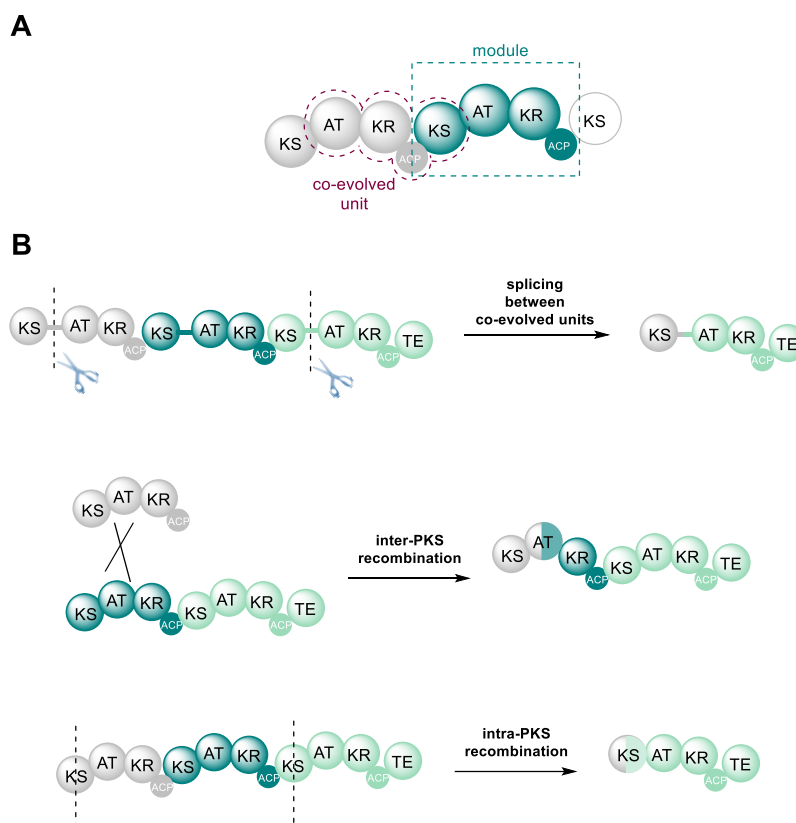


Figure 5.1.2 Evolution-inspired PKS engineering. (A) Module and co-evolved unit boundaries. Boundaries for both models have been successfully used for PKS engineering. (B) Approaches include the use of splice points of co-evolved units or inter- or intra-PKS recombination.

More recently, a repetitive genetic element which appears to participate in accelerating the diversification of BGCs in *Streptomyces* species through gene conversion was observed.²⁴⁶ Genetic repeats of intense nucleotide skews (GRINS) were mainly distributed in AT domain-encoding regions, with relatively frequent occurrence in KS, DH, ER, and KR domains and to a lesser degree in ACP domains and docking regions.²⁴⁶ While the molecular mechanisms associated with the appearance and distribution of GRINS are unknown, their presence in a broad range of bacterial phyla and gene families indicates that they could participate in protein evolution. *In silico* analysis of recombination events observed in the rapamycin PKS indicated that they occurred between GRINS, illustrating their potential utility for PKS engineering.

5.2 Results and Discussion

5.2.1 Distribution of *gds* locus among *B. gladioli*

Following the discovery of gladiostatin, a large-scale genome mining and phylogenetic study was conducted on all available *Burkholderia* genomes to fully understand the distribution of the *gds* locus within the species (work carried out by Dr. Alex Mullins, Cardiff University. *n.b.* some genomes searched are not yet published). The *gds* locus was searched against 4479 genomic assemblies of *Burkholderia*, using the sequence extracted from *B. gladioli* BCC0238 as a representative sequence. The *gds* BGC was found solely in *B. gladioli* isolates. The paired-end Illumina reads of the genome subset were subsequently mapped against the gladiostatin BGC in *B. gladioli* BCC0238 to confirm the presence of the biosynthetic genes. Unrooted core-gene phylogeny of 229 *B. gladioli* genomes based on *gdsF*, revealed that 112 *B. gladioli* genomes containing either the *gds* BGC or variants (**Figure 5.2.1**).

The analysis revealed that the gladiostatin biosynthetic gene cluster is present in approximately 49% of *B. gladioli* genomes. This suggests that gladiostatin plays an important role in the adaption of *B. gladioli* isolates to their environment. The strict distribution of the *gds* locus along a monophyletic clade suggested that intragenic recombination is the driving force for gene cluster distribution rather than horizontal gene transfer.

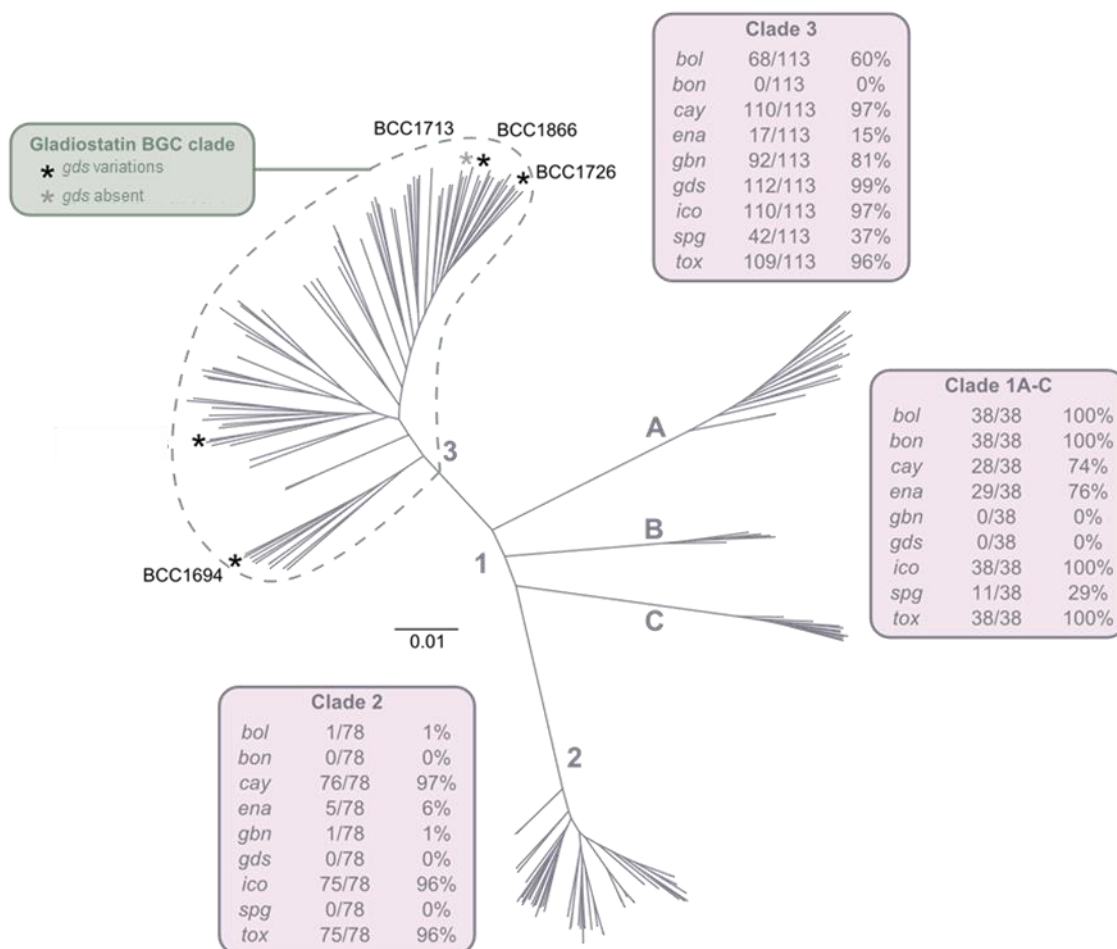


Figure 5.2.1 Unrooted core-gene phylogeny of 229 *Burkholderia gladioli* genomes highlighting the distribution of *gds*-like BGCs. *gds*-like BGCs were restricted to a single clade within the phylogeny and were found in 112 of the 113 genomes in this clade. The maximum-likelihood phylogenetic tree was constructed with 100 bootstraps using a general time reversible model and a gamma model of rate heterogeneity. Bootstrap support is indicated for major clade splits (node circles). The distribution of several other BGCs of known function in *B. gladioli* genomes are also shown. Scale bar represents number of base substitutions per site. The figure was generated by Dr. Alex Mullins.

Comprehensive analysis of the *gds*-like BGCs in *B. gladioli* enabled us to identify differences in the architecture of some of the encoded PKSs. A deletion of approximately 4440 bp was observed in the region corresponding to the GdsF PKS subunit in four strains (BCC1713, BCC1726, BCC1866 and BCC1694) (**Figure 5.2.2**). This deletion corresponds to 1480 aa, suggesting loss of several catalytic domains from the PKS.

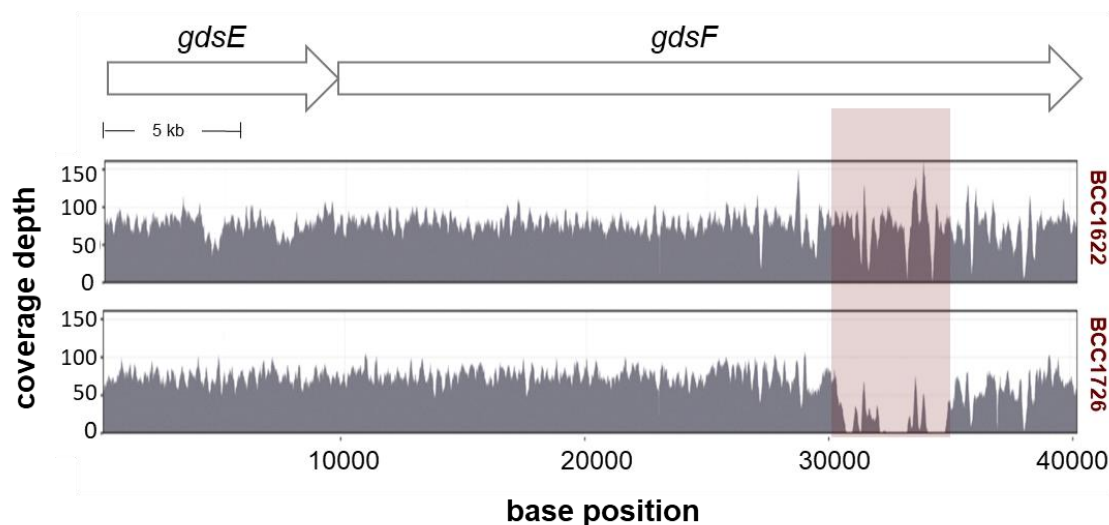


Figure 5.2.2 Read depth graph indicating the 4440 bp deletion region compared to the reference genome of *B. gladioli* BCC1622.

5.2.2 Secondary metabolite analysis reveals the presence of a gladiostatin variant

The deletion in the *gds* BGC in four *B. gladioli* isolates prompted us to explore the potential of these strains to produce derivatives of gladiostatin. EtOAc extracts of *B. gladioli* strains where the *gds* cluster was present, absent or contained a deletion in *gdsF* were subjected to UHPLC-ESI-Q-TOF-MS analysis. These data showed that *B. gladioli* BCC1713 (*gds* cluster absent) does not produce gladiostatin, whereas strains that harboured the 'standard' *gds* cluster in their genome produced high amounts of this metabolite (**Figure 5.2.3**). Interestingly, *B. gladioli* BCC1866, BCC1726 and BCC1694, strains that contain the deletion in the *gdsF* gene, were found to produce a novel metabolite. This gave rise to ions with m/z 478.27 ± 0.02 and 500.22 ± 0.02 , corresponding to $[M+H]^+$ and $[M+Na]^+$, respectively, for a compound with the molecular formula $C_{25}H_{35}NO_8$ (**Figure 5.2.3**). Gladiostatin is not produced by this strain, indicating that this metabolite is the only metabolic product of this variant BGC. The structure of this novel metabolite was elucidated using 1D- and 2D-NMR spectroscopy (**Figure 5.2.3**, **Table 7.10.3**). Analysis of ^{13}C and DEPT-135 NMR spectra revealed the presence of two methyl groups, six

methines, five carbonyl groups, seven quaternary carbons and ten methylenes. HMBC correlations between the exchangeable NH proton and C-2/C-2' and between H-3 and C-1/C-1' confirmed the presence of the 2,6-piperidinedione moiety. The chemical shifts for carbons and protons in this moiety were very similar to those reported for gladiostatin. COSY correlations indicated the link of the 2, 6-piperidinedione moiety to the C-6 methylene. HMBC analysis confirmed that a keto group is juxtaposed between C-6 and C-8. COSY correlations established the connectivity of the C-8 to C-16 region of the molecule and HMBC data confirmed the location of the C-8 methyl group and the C-9/C-10 double bond. The double bond was assigned the *E* configuration, based on a $^3J_{HH}$ value of 15 Hz. The butenolide moiety was assigned based on HMBC correlations between C-16 and C-17/C-18, C-3" protons and C-1"/C-18 and the C-1" proton correlation to C-2"/C-18. These data agreed well with the previously reported NMR data for the corresponding moiety in gladiostatin. Having structurally confirmed this compound as an analogue of **49**, we named it gladiostatin C (**78**).

Given the previous bioactivity profile of gladiostatin (**49**), the biological activity of **78** was also examined against a range of representative ESKAPE pathogens and *C. albicans*, using MIC assays as described in **section 7.11**. Similar to **49**, **78** proved to be inactive against bacterial pathogens as well as *C. albicans* (**Table A2**). Gladiostatin C was also tested against *S. cerevisiae* and showed similar activity to **49** with an MIC of 4 $\mu\text{g mL}^{-1}$ (**Table A2**). Since other glutarimides as well as **49** are active against various cancer cell lines, **78** was screened for its cytotoxicity towards cancer cells lines by Dr. Isolda Romero-Canelón from the School of Pharmacy, University of Birmingham, showing lower potency than gladiostatin.

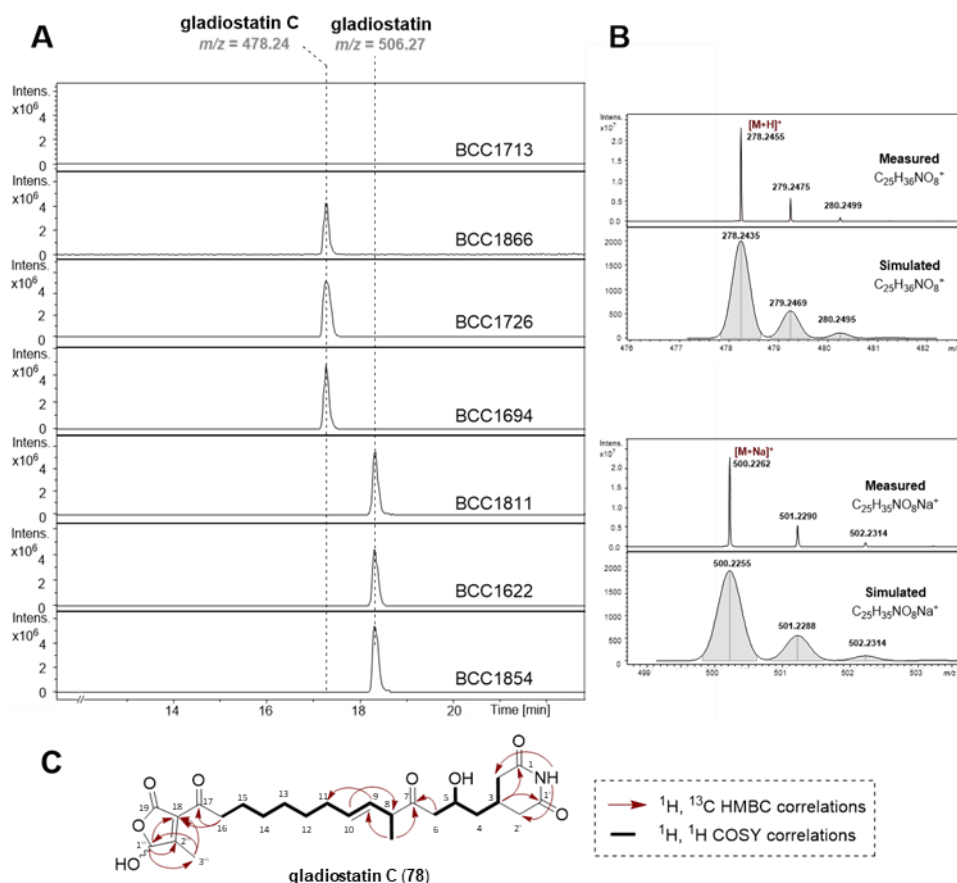
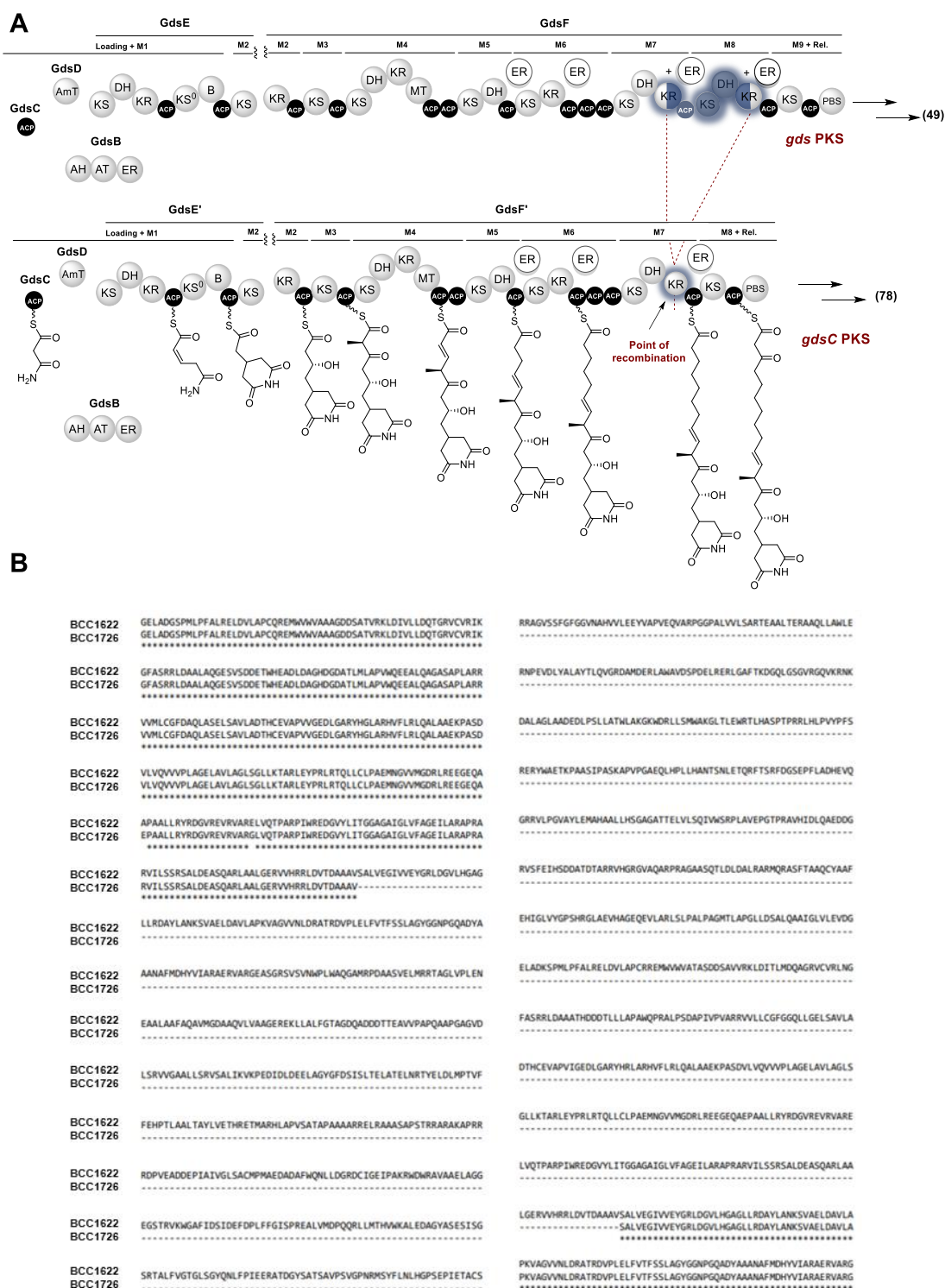


Figure 5.2.3 Production of gladiostatin and the derivative, gladiostatin C (78), by genome sequenced *B. gladioli* strains. (A) Extracted ion chromatograms at $m/z = 478.27 \pm 0.02$ (corresponding to the $[M+H]^+$ for gladiostatin C) and $m/z = 506.27 \pm 0.02$ (corresponding to the $[M+H]^+$ for gladiostatin) from LC-MS analyses of extracts of *B. gladioli* strains where the *gds* cluster was present (in modified and unmodified form) or absent based on read mapping against the *B. gladioli* BCC1622 reference strain containing the *gds* BGC. (B) High-resolution mass spectrometry analysis of **78**. Comparison of measured spectra with simulated spectra for $C_{25}H_{36}NO_8^+$ and $C_{25}H_{35}NO_8Na^+$, corresponding to the $[M+H]^+$ and $[M+Na]^+$ ions, respectively. (C) Key COSY and HMBC correlations observed for **78** (d-chloroform, 1H 700 MHz, ^{13}C 175 MHz).

5.2.3 Domain-level analysis of the recombination event in the *gds* BGC

Structure elucidation of the truncated derivative of **49**, allowed the prediction of possible changes in the module organisation of the PKS that could account for the structural differences. The lack of two methylenes ($-C_2H_4-$) is consistent with a PKS lacking an elongation module. The truncated *gdsF* found in BCC1726 strain was referenced as *gdsF'*. Comparison of the PKSs encoded by *gdsF/gdsF'* indicated a high degree of sequence conservation except module 8 in GdsF which appears to be absent from the PKS encoded by *gdsF'* (**Figure 5.1.4**). The amino acids missing from GdsF' extend from the C-terminal region of the KR domain in module 7 of GdsF to the N-terminal region of the KR domain in module 8 of this PKS subunit. The result is loss of a PKS module and creation of a chimeric KR domain (**Figure 5.1.4**). These data are indicative of a homologous recombination event between the regions of *gdsF* encoding the module 7 and module 8 KR domains.

To more precisely locate the fusion point and the potential impact this may have on the KR domain, a homology model of the GdsF KR_{M7} domain was generated using Phyre2²⁴⁷, as described in the experimental section (**Figure 5.2.5**). The structure of a B-type KR domain from the erythromycin *cis*-AT PKS was used as reference to predict precise domain boundaries of the protein through sequence alignment.⁸⁴ The Ala 12 residues upstream of the conserved (H/L/M/F/Y)xxxW sequence motif was selected as the N-terminal boundary, while the Asp 18 residues upstream of the ACP domain was selected as the C-terminal boundary. The model was based on PksJKR₁, a ketoreductase domain of the bacillaene *trans*-AT PKS that carries out both α - and β -keto reduction reactions during chain assembly.^{182,248}



found in *B. gladioli* BCC1622 and *B. gladioli* BCC1726, respectively. The alignment was generated using Clustal Omega²⁴⁹.

Mapping the sequences onto the model, it can be observed that the splice point is in the middle of the catalytic subdomain of the KR monomer with both “halves” contributing to the NADPH binding pocket (**Figure 5.2.5, 5.2.6**). The dinucleotide binding motif (TGGAGSLG) is in the N-terminal stem region of the GdsF module 7 KR domain (**Figure 5.2.6**). Catalytic residues (Y330; K293; S317) are in the part of the catalytic subdomain that derives from the C-terminus of the GdsF module 8 KR domain (**Figure 5.2.5, 5.2.6**). KR domains in modules 7 and 8 of GdsF are predicted to be B-type. The GdsF' module 7 KR domain also contains a Leu-Arg-Asp triad typical of B-type enzymes (**Figure 5.2.6**).¹⁸² Based on this analysis it was postulated that the splice point lies within the KR domain.

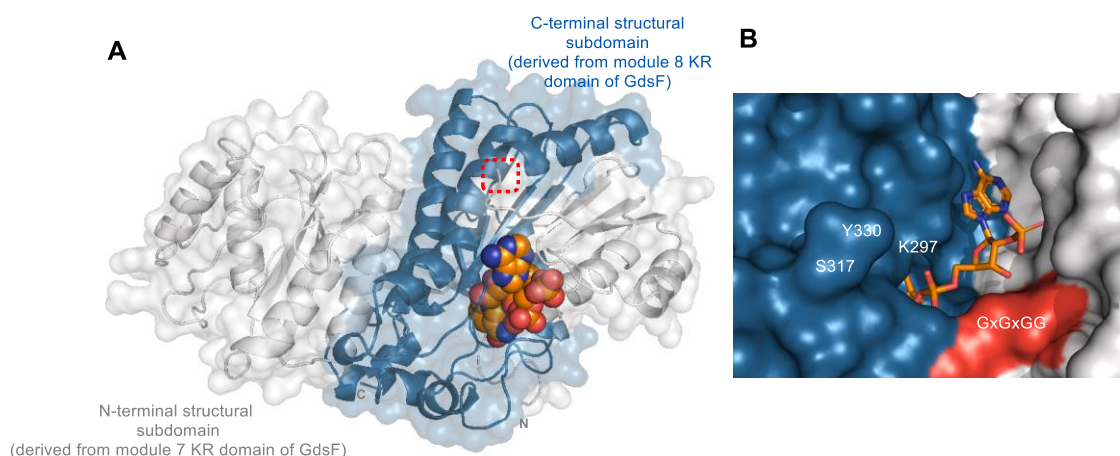


Figure 5.2.5 Homology model of chimeric KR_{M7} responsible for the biosynthesis of gladiostatin C.

(A) Homology model of GdsF' module 7 KR domain showing cartoon and surface representations. The model is colour coded depending on the origin of the peptide sequence, grey for the N-terminal region of the GdsF' module 7 KR domain and blue for the C-terminal region of the GdsF module 8 KR domain. NADPH is represented as spheres. The point of sub-domain fusion is highlighted by a red box. (B) Structure of the GdsF' module 7 KR domain active site showing NADP⁺ and residues that bind it, indicating that both halves contain essential elements for the function of the domain.

[illegible]

137

(5204 bp product), single crossover *B. gladioli* BCC1622 Δ *gdsF*[S7296-D8774] mutant and the *B. gladioli* BCC1622 Δ *gdsF*[S7296-D8774] double crossover mutant (770 bp product).

The metabolic profile of the *B. gladioli* BCC1622 Δ *gdsF*[S7295-D8774] mutant was analysed using UHPLC-ESI-Q-TOF-MS and compared to metabolite profiles of *B. gladioli* BCC1622 wild type (gladiostatin producer) and *B. gladioli* BCC1726 (gladiostatin C producer) (**Figure 5.2.8**). The S7295-D8774 deletion in *gdsF* in BCC1622 resulted in exclusive production of **78**. As expected, the NMR data for this compound matched those for **78** isolated from BCC1726. The relative production levels to those observed for BCC1726 indicated that the chimeric PKS was fully functional, and suggested that the splice point identified results in an optimally functioning PKS, unlike other engineering approaches reported in the literature.²⁵⁵

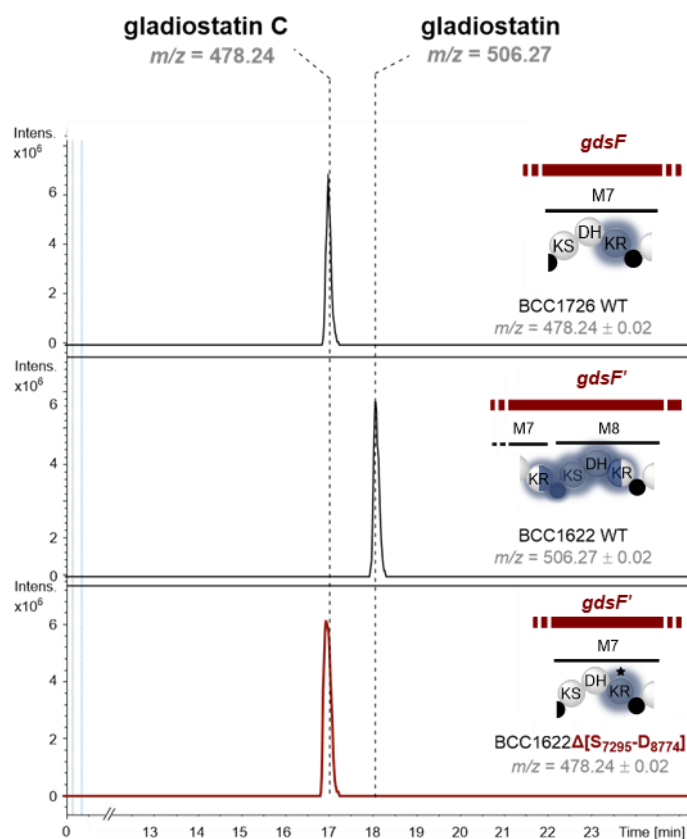


Figure 5.2.8 Engineering gladiostatin C production in the gladiostatin producing strain *B. gladioli* BCC1622. Extracted ion chromatograms at $m/z = 478.27 \pm 0.02$ (corresponding to the $[M+H]^+$ for gladiostatin C) from LC-MS analyses of the *B. gladioli* BCC1622 Δ [S7295-D8774] mutant, wild type *B. gladioli* BCC1622 strain and *B. gladioli* BCC1726. The chimeric KR domain created by the deletion is highlighted with an asterisk.

These results confirm our initial *in silico* observations of a recombination event occurring resulting in a chimeric KR domain in BCC1726. Mimicking natural evolutionary processes, the PKS responsible for gladiostatin assembly was converted into a gladiostatin C producer, without compromising the functionality of the chimeric KR domain.

5.3 Conclusions

In this study, the diversity of metabolites produced by *B. gladioli* isolates containing variant *gds* loci with alterations in PKS genes was explored. A metabolomics analysis revealed that some isolates produce a novel gladiostatin derivative lacking two methylene groups. *In silico* analysis predicted the splice site is in the middle of a KR domain, specifically in the catalytic subdomain. Since it was believed that this constitutes a snapshot of *trans*-AT PKS evolution, a strategy to convert the gladiostatin producer into a gladiostatin C producer by mimicking the natural recombination event was pursued. An in-frame deletion in the region of *gdsF* encoding the C-terminus of the module 7 KR to the N-terminus of the module 8 KR domain was constructed in an effort to reproduce the architecture of the gladiostatin C PKS. The truncated gladiostatin PKS produced exclusively gladiostatin C at wild type levels.

These findings could greatly facilitate future efforts to create productive PKS variants. PKS engineering based on evolution-inspired strategies has gradually gained ground in efforts to expand the structural diversity of polyketide natural products. Previous studies have suggested that recombination in *cis* and *trans*-AT PKS systems likely occurs at KS-AT domain junctions rather than ACP-KS domain junctions.^{63,241} Our study on engineering the gladiostatin *trans*-AT PKS based on natural splice sites further solidifies the hypothesis resulting from accelerated evolution experiments conducted on the rapamycin and tylosin *cis*-AT PKS systems²⁴⁵ that recombination often occurs in the middle of a catalytic domain, and provides new insights into the evolution of *trans*-AT PKS systems, as well as a directions for future research on diversity-oriented biosynthesis of other polyketides. Reliable excisions and insertions could be readily engineered once the exact locations of functional splice sites have been more fully defined.

Chapter 6

Conclusions and future perspectives

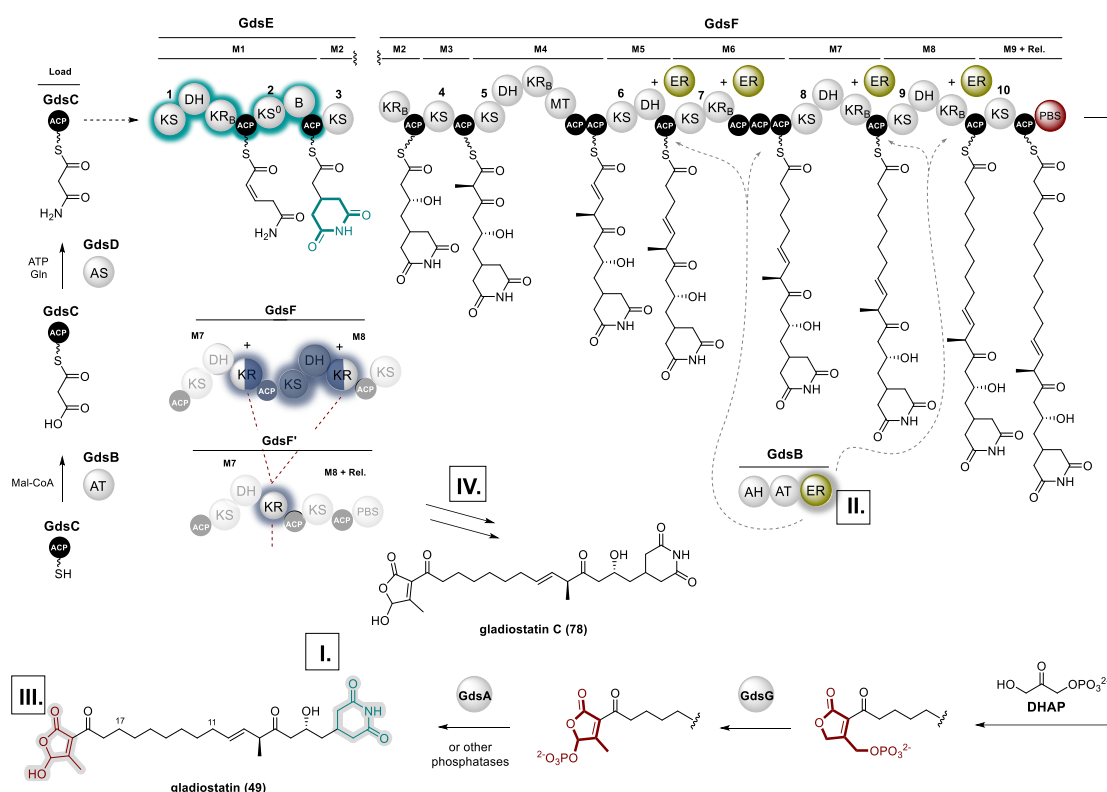


Figure 6.1 Overview of research on the gladiostatin BGC described in this thesis. This includes: (I) the discovery of a novel natural product with unusual motifs; (II) *in vivo* and *in vitro* studies on an unprecedented number of enoyl reductions catalysed by a sole *trans*-acting ER; (III) elucidation of a novel chain off-loading mechanism and (IV) successful engineering of a PKS by mapping the evolution of *trans*-AT PKS on numerous *B. gladioli* isolates.

6.1 Discovery of a polyketide bearing unusual functionalities

(I)

In this thesis, the genomics-driven discovery of a novel polyketide product of cryptic *trans*-AT PKS is described, further demonstrating the ability of *B. gladioli* BCC1622 and BCC0238 to produce a diversity of specialised metabolites.^{139,256,257} The carbon skeleton of gladiostatin (**49**) exhibits biologically active functionalities, as observed for molecules that tend to derive from *Streptomyces* species. As the first glutarimide antibiotic produced from a Gram-negative bacterium, **49** presents differences in structure including a saturated C-11 to C-18 chain and a

2-acyl-4-hydroxyl-3-methyl butenolide. Bioinformatics analysis of the PKS reveal that these alterations derive from different domain architectures compared to known glutarimide PKS organisation. Whilst many changes in the structure can be mapped to PKS domain organisation, some others cannot, including the lack of a C-10 methyl group. Therefore, future efforts should be directed towards this system in order to develop a better understanding of the enzymatic peculiarities of *trans*-AT PKSs, and will facilitate bioengineering approaches to polyketide structural diversification. Gladiostatin (X) was active against the eukaryote *S. cerevisiae* and presented cytotoxicity against a range of cancer cell lines. The results were in good agreement with reported anti-tumour activity of other glutarimide-containing molecules.^{146,151,177} Recently, X-ray crystal structures of lactimidomycin (**34**) and cycloheximide (**25**) bound to the yeast ribose revealed their mode of inhibition which occurred at different elongation stages.¹⁶³ Eukaryote-specific ribosomal inhibitors are increasingly used in research and hold potential for new therapeutics against a wide range of infectious diseases, cancer and genetic disorders.^{258,259} Ongoing efforts to co-crystallise gladiostatin with the yeast ribosome will unravel possible binding sites and modes of action towards inhibition of the eukaryotic ribosome.

6.2 Unravelling an unprecedented number of enoyl reductions performed by a sole *trans*-acting ER (II)

The structure of gladiostatin was found to diverge from the predicted structure based on domain organisation of the *gds* PKS. Like other *trans*-AT PKS systems, the *gds* PKS is proposed to utilise *trans*-acting proteins to facilitate enzymatic steps required for additional structural diversity. The GdsB *trans*-acting ER domain was biochemically characterised and found to perform enoyl reduction *in trans* at multiple points during biosynthesis, resulting the aliphatic region found in the skeleton structure of gladiostatin. This work demonstrates multiple specific interactions between the *trans*-acting ER domain and a set of cognate ACP domains. Whilst *trans*-acting enoyl reduction events are not uncommon in *trans*-AT PKSs, the number of sequential reactions exhibited during the biosynthesis of gladiostatin is unprecedented. The

work described in Chapter 3 sets the stage for the molecular details of this interaction to be studied. The conserved Phe found in all interacting ACPs but not present in non-interacting counterparts could be the starting point of saturation mutagenesis efforts to identify the conserved residues that mediate specific interaction. In addition, these studies can be complemented with NMR, carbene footprinting and crystallography to structurally characterise these interactions. The role of the second standalone oxidoreductase in the cluster remains elusive. Knock-out studies revealed that the gene encoding the GdsH oxidoreductase is essential for gladiostatin production and upon deletion a gladiostatin derivative is formed containing an additional double bond. This suggested that GdsH contributes towards one enoyl reduction event during the assembly of gladiostatin. A parallel study working on the same metabolite produced from a *B. gladioli* beetle symbiont, reported a similar outcome after deleting the homologous gene.²¹⁹ Biochemical characterisation of this enzyme did not provide evidence of interaction of this protein with ACP domains therefore the role of this enzyme requires further investigation. The possibility of this enzyme to perform post-PKS tailoring function is highly unlikely due the less electrophilic nature of the double bond (detailed in Chapter 3).

6.3 An unusual chain off-loading mechanism (III)

The AfsA enzymology, which has been well studied for γ -butyrolactone biosynthesis in *Streptomyces*, has never been found appended to a PKS pathway as a release mechanism. In Chapter 4, the AfsA-like enzymology is characterised for the *gds trans*-AT PKS pathway as a mechanism for catalytic chain release. In addition to the AfsA-enzymology, a novel 1,3-rearrangement of a phosphate group was identified, which generates the final 2-acyl-3-methyl-4-hydroxy butanolide unit of gladiostatin. This reaction is catalysed by a phosphato-mutase, GdsG, which presents homology to HAD-like phosphatase enzymes. The new mechanism for polyketide chain release, in addition to the new biochemical transformation performed by GdsG and homologues could prove to be a valuable addition to the biosynthetic toolbox. Future directions towards obtaining structural information for GdsG or its homologues, SabP and P022, will reveal the catalytic residues responsible for catalytic mechanism. In addition, the loss of the butanolide head over time from the gladiostatin skeleton which appears to be synchronised with

the timing of optimal production of other antibiotics produced by this strain, raises questions with regards to its involvement in the antibiotic production. Therefore studies into the biological role of this functionality could be pursued in the future.

6.4 Evolution-guided PKS engineering (IV)

Species-wide analysis of the *gds* cluster demonstrated that the distribution of the *gds* BGC was found to be solely within the *B. gladioli* species. Interestingly, several *B. gladioli* strains possessed deletions in the *gds* PKS. Analysis of the metabolic profile of these strains identified a gladiostatin derivative lacking two carbons in its structure which was hypothesised to result from one less chain elongation step (plus reductive modification) during gladiostatin biosynthesis. Domain-level inspection confirmed this hypothesis and identified a recombination event within the KR domain region which resulted in a loss of a module. Prompted by this *in silico* analysis, we performed engineering of the gladiostatin PKS based on the observed natural point of recombination. That allowed us to produce a functional chimeric PKS with metabolite production levels comparable to those of the wild-type strain. To our knowledge, this is the first example of engineering of a *trans*-AT PKS. The vast repository of *B. gladioli* strains containing high sequence similarities will allow future exploration of recombination events in PKSs. This logic can also be utilised to study the evolution of non-ribosomal peptide synthetases (NRPS). Even though these type of modular enzymes contain enzymatic domains that belong to different protein families of those found in PKSs, the two assembly lines utilise similar biosynthetic logic to assembly complex peptide products.²⁶⁰ A comprehension of evolutionary procedures will not only add value in leveraging our understanding of these enzymes but also enable us to engineer fully functional chimeric PKSs capable of generating custom-made chemical entities.

Chapter 7

Experimental

7.1 Instruments and Equipment

7.1.1 Semi-preparative HPLC

Isolation of secondary metabolites was performed in a semi-preparative reverse-phase C18 Betasil column (21.2 mm x 150 mm). The elution was conducted using 5% acetonitrile/ 95% water for 5 min followed by linear gradient from 5 to 100% over 45 min under a flow rate of 9 mL/min. The desired fractions were collected and dried using ScanVac CoolSafe freeze dryer (LaboGene).

7.1.2 Mass Spectrometry

For small molecules, UHPLC-ESI-Q-TOF-MS analyses were performed using a Dionex UltiMate 3000 UHPLC connected to a Zorbax Eclipse Plus C18 column (100 × 2.1 mm, 1.8 µm) coupled to a Bruker MaXis IMPACT mass spectrometer. Mobile phases: water (A) and acetonitrile (B), each supplemented with 0.1% formic acid. A gradient of 5% B to 100% B over 30 minutes was employed at a flow rate of 0.2 mL/min. The mass spectrometer was operated in positive ion mode with a scan range of 50-3000 m/z . Source conditions were as follows: end plate offset at -500 V; capillary at -4500 V; nebulizer gas (N₂) at 1.6 bar; dry gas (N₂) at 8 L min⁻¹; dry temperature at 180 °C. Ion transfer conditions were as follows: 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 µs; pre-pulse storage time at 1 µs. Calibration was performed with 1 mM sodium formate through a loop injection of 20 µL at the start of each run.

UHPLC-ESI-Q-TOF-MS analyses of intact proteins were performed on a Dionex UltiMate 3000 RS HPLC integrated to an ACE 3 C4-300 reverse phase column (Advanced Chromatography Technologies; 100 × 2.1 mm, 5 µM, 30 °C), coupled to a Bruker MaXis II mass spectrometer. Proteins were eluted with a linear gradient of 5-100% MeCN containing 0.1% formic acid over 30 min. The mass spectrometer was functioned in positive ion mode with a scan range of 200-3000 m/z . Source conditions were the following: end plate offset at -500 V; capillary at -4500

V; nebulizer gas (N_2) at 1.6 bar; dry gas (N_2) at 9 L/min, dry temperature at 200°C. Ion transfer conditions were: ion funnel RF at 400 Vpp; multiple RF at 200 Vpp; quadrupole low mass at 200 m/z; collision energy at 8.0 eV; collision RF at 2000 Vpp; transfer time at 110.0 μ s and pre-pulse storage time at 10.0 μ s.

7.1.3 NMR spectroscopy

The 1H - and ^{13}C -NMR spectra were recorded using a Bruker Avance AV-300, HD-400 and HD-500 MHz spectrometer. Chemical shifts were expressed in parts per million (ppm) referenced from $CDCl_3$ ($\delta_H = 7.26$ ppm and $\delta_C = 77.2$ ppm). The following abbreviations of multiplicities are used: multiplet (m), singlet (s), doublet (d), triplet (t), quartet (q), doublet of doublets (dd). Coupling constants (J) are rounded to the nearest 0.5 Hertz (Hz). 1H - and ^{13}C assignments were identified based on the basis of COSY, ^{13}C -DEPT, HMBC and HSQC correlations.

7.2 Materials

7.2.1 General

All media were prepared using deionised water and sterilised through autoclaving at 121 °C for 5 min. Broth powders were purchased from Sigma Aldrich. Bacto Casamino Acids, Bacto yeast extract and Bacto agar were purchased from Becton Dickinson Microbiology. All chemical reagents and solvents were purchased from Sigma Aldrich, Thermo Fisher or VWR International.

7.2.2 Culture Media

Luria-Bertani (LB) medium: 25 g of LB broth powder was dissolved in deionised 1L of H_2O and sterilised by autoclaving. For LB agar plates, same procedure was pursued, followed by addition of Bacto Agar (15 g/L) prior to sterilisation.

Basal Salts Medium (BSM): The following components: $K_2HPO_4 \cdot 3H_2O$ (4.3 g/L), NaH_2PO_4 (1.0 g/L), NH_4Cl (2.0 g/L), $MgSO_4 \cdot 7H_2O$ (0.2 g/L), $FeSO_4 \cdot 7H_2O$ (0.01 g/L), $MnSO_4 \cdot H_2O$ (0.003 g/L), $ZnSO_4 \cdot H_2O$ (0.003 g/L), $CoSO_4 \cdot 7H_2O$ (0.001 g/L), nitriloacetic acid (0.1 g/L), Bacto Yeast

Extract (0.5 g/L), Bacto Casaminoacids (0.5 g/L) and glycerol (4.0 g/L) were dissolved in deionised H₂O and the pH was adjusted to 7.2 ± 0.2 . Bacto agar (15 /L) was then added prior to sterilization. When glycerol is replaced by glucose or ribose, 10 ml of a sterile solution of 20% (w/v) ribose/glucose in H₂O was added instead of glycerol after autoclaving the media²⁶¹.

Mueller Hinton (MH) medium: MH broth powder was dissolved in deionised H₂O (22 g/L) and sterilised by autoclaving.

Roswell Park Memorial Institute RPMI 1640 medium: RPMI 1640 buffered with 0.165 mol/L MOPs was prepared by adding powdered RPMI 1640 medium (with glutamine and phenol red, without bicarbonate) in deionised H₂O (10.4 g/L). 34.53 g of MOPS was added and the solution was stirred until dissolved. While stirring, the pH of the solution was adjusted to 7.0 at 25 °C, using sodium hydroxide. The solution was then sterilized using filtration.

M9 Minimal Media: The media was prepared by dissolving the following components in deionised H₂O: Na₂HPO₄ (6.0 g/L), KH₂PO₄ (3.0 g/L), NaCl (0.5 g/L), NH₄Cl (1.0 g/L), sucrose (10.0 g/L), MgSO₄ (0.1 g/L) and CaCl₂ (0.01 g/L). The pH was adjusted to 7.4. Bacto agar (15 g/L) was subsequently added to the mixture, prior to sterilisation.

7.2.3 Antibiotics

All antibiotics were purchased from Sigma Aldrich.

Table 7.2.1 List of antibiotics used for these studies.

Antibiotic	Working Conc.	Solvent	Strain to be applied
Kanamycin	50 µg/mL	H ₂ O	<i>E. coli</i> carrying recombinant pET28a(+)
Trimethoprim	50 µg/mL	DMSO	<i>E. coli</i> carrying recombinant pGPI-Scel
	150 µg/mL		<i>Burkholderia</i> carrying recombinant pGPI-Scel
Polymyxin B	600 U/mL	H ₂ O	triparental mating selection
Tetracycline	50 µg/mL	70% EtOH	<i>E. coli</i> carrying pDAI-Scel

150 µg/mL *Burkholderia* carrying recombinant
pDAI-Scel

7.2.4 Buffers

Tris/Borate/EDTA (TBE) Buffer (10x): Tris-HCl (890 mM), boric acid (890 mM) and EDTA (20 mM) were dissolved in deionised H₂O. The solution was diluted to 1x before use.

Loading Buffer: NaCl (100 mM), Tris base (20 mM) and imidazole (20 mM) were dissolved in deionised H₂O and the pH was adjusted to 8.0 prior to filter sterilisation.

Imidazole Elution Buffers: NaCl (100 mM), Tris base (20 mM) and imidazole concentrations (50-300 mM) were dissolved in deionised H₂O and the pH was adjusted to 8.0 prior to filter sterilisation.

Storage Buffer: NaCl (100 mM) and Tris base (20 mM) were dissolved in deionised H₂O and the pH was adjusted to 7.4, prior to filter sterilisation.

SDS-PAGE Sample Buffer (5x): SDS (10% w/v), 2-mercaptoethanol (10 mM), glycerol (20 % w/v), Tris base (0.2 M) and bromophenol blue (0.05 %, w/v) was dissolved in deionised H₂O and the pH was adjusted to 6.8. The solution was diluted to 1x with protein sample when used.

SDS-PAGE Running Buffer (10x): Tris base (250 mM), glycine (2 M) and SDS (1% w/v) were dissolved in deionised H₂O and the pH was adjusted to 8.0. The solution was diluted to 1x with protein sample when used.

Chemically Competent Cell Preparation Buffer I (CCCI): Potassium acetate (30 mM) and CaCl₂ (80 mM) were dissolved in deionised H₂O and the pH was adjusted to 5.8 prior to sterilisation.

Chemically Competent Cell Preparation Buffer II (CCCI): Potassium acetate (30 mM), CaCl₂ (80 mM) and glycerol (15% w/v) were dissolved in deionised H₂O and the pH was adjusted to 5.8 prior to sterilisation.

7.2.5 Strains and vectors

Table 7.2.2 List of microbial strains used within this project.

Microbial strain	Growth Temperature	Description	Reference
<i>Escherichia coli</i>			
<i>E. coli</i> TOP10	37 °C	cloning and plasmid propagation	NEB
NEB 5-alpha	37 °C	cloning and plasmid propagation	NEB
<i>E. coli</i> BL21 (DE3)	37 °C	Host for recombinant protein expression	NEB
<i>E. coli</i> SY327	37 °C	Donor strain for conjugation between <i>E. coli</i> and <i>Burkholderia</i>	Mahenthiralin gam collection
<i>E. coli</i> HB101/pRK2013	37 °C	Helper strain for conjugation between <i>E. coli</i> and <i>Burkholderia</i>	Mahenthiralin gam collection
<i>Burkholderia gladioli</i>			
BCC1622	30 °C	Wild type, gladiostatin A producing strain	Mahenthiralin gam collection
BCC0238	30 °C	Wild type, gladiostatin A producing strain	Mahenthiralin gam collection
BCC1713	30 °C	Wild type	Mahenthiralin gam collection

BCC1866	30 °C	Wild type	Mahenthiralin gam collection
BCC1726	30 °C	Wild type, gladiostatin B producing strain	Mahenthiralin gam collection
BCC1694	30 °C	Wild type, gladiostatin B producing strain	Mahenthiralin gam collection
BCC1811	30 °C	Wild type, gladiostatin B producing strain	Mahenthiralin gam collection
BCC1854	30 °C	Wild type, gladiostatin A producing strain	Mahenthiralin gam collection
BCC1622 Δ <i>gdsH</i>	30 °C	Biosynthetic studies of gladiostatin	Genetically engineered in this study
BCC1622 Δ <i>gdsG</i>	30 °C	Biosynthetic studies of gladiostatin	Genetically engineered in this study
BCC1622 Δ <i>gdsA</i>	30 °C	Biosynthetic studies of gladiostatin	Genetically engineered in this study
BCC1622:: <i>gdsE</i>	30 °C	Biosynthetic studies of gladiostatin	Genetically engineered in this study
BCC1622 Δ <i>gdsF</i> [S ₇₂₉₆ - D ₈₇₇₄]	30 °C	Biosynthetic studies of gladiostatin	Genetically engineered in this study

Strains used for antimicrobial susceptibility testing

<i>Klebsiella pneumonia</i>		Antimicrobial	DSMZ culture
DSM26371	30 °C	susceptibility testing	collection
<i>Acinetobacter</i>		Antimicrobial	DSMZ culture
<i>baumannii</i> DSM25645	30 °C	susceptibility testing	collection
<i>Pseudomonas</i>		Antimicrobial	DSMZ culture
<i>aeruginosa</i> DSM29239	30 °C	susceptibility testing	collection
<i>Enterobacter cloacae</i>		Antimicrobial	DSMZ culture
DSM16690	30 °C	susceptibility testing	collection
<i>Enterococcus faecium</i>		Antimicrobial	DSMZ culture
DSM25390	30 °C	susceptibility testing	collection
<i>Staphylococcus</i>		Antimicrobial	DSMZ culture
<i>aureus</i> DSM21979	30 °C	susceptibility testing	collection
<i>Saccharomyces</i>		Antimicrobial	ATCC
<i>cerevisiae</i> W303-1a	30 °C	susceptibility testing	collection

Table 7.2.3 List of vectors used within this project.

Name	Description	Source of reference
	Recombinant protein	
pET28a (+)	expression in <i>E. coli</i> (N-terminal His tag)	Novagen
pGPI-SceI	ori _{R6K} <i>Tp^R mob⁺</i> carries <i>I-SceI</i> cut site, for gene deletion in <i>Burkholderia</i>	Mahenthiralingam collection
pDAI-SceI	ori _{pBBRI} <i>Tet^R mob⁺</i> P _{dhfr} FLAG epitope, carries the <i>I-SceI</i>	Mahenthiralingam collection

pMLBAD	<i>pBBR1 ori araC-PBAD Tp^R</i> <i>mob⁺</i> , for gene deletion in <i>Burkholderia</i>	Mahenthiralingam collection
--------	---	--------------------------------

7.3 General DNA manipulation

All kits described in this section and restriction enzymes were purchased from Thermo Fisher Scientific (UK). Q5® Hot Start High-Fidelity 2X Master Mix, OneTaq® Hot Start Quick-Load® 2X Master Mix with GC Buffer, T4 DNA ligase, Q5® Site-Directed Mutagenesis Kit were all purchased from New England Biolabs.

7.3.3 Isolation of genomic DNA/plasmid DNA

B. gladioli BCC1622/BCC0238 genomic DNA was extracted from a 5 mL overnight culture using the GeneJET Genomic DNA Purification Kit, based on manufacturer's recommended protocol for Gram-negative bacteria.

Plasmid DNA was extracted from an overnight culture of *E. coli* (5 mL) using the GeneJET Plasmid Miniprep Kit. The concentration of the purified plasmid DNA was measured using a NanoDrop™ Lite spectrophotometer.

7.3.4 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. PCRs were performed on an Eppendorf Mastercycler® nexus GX2. A typical reaction was set up as described in **Table 7.3.1.**

Table 7.3.1 PCR components per 25 µL total volume.

Component	Volume (µL)
Q5 Hot Start High-Fidelity/ OneTaq Hot Start DNA polymerase	12.5

10 μ M forward primer	1.25
10 μ M reverse primer	1.25
DNA template	variable
DMSO	1-2
Nuclease-free water	to 25

Table 7.3.2 Conditions used in the thermocycler.

Step		Temperature	Time
Initial Denaturation		98 °C	3 min
35 cycles	Denaturation	98 °C	30 sec
	Annealing	58-69 °C	30 sec
	Extension	72 °C	30 sec/kb or 1 min/kb
	Final Extension	72 °C	10 min
Hold		4 °C	∞

30 sec/kb or 1 min/kb was applied at extension step when using Q5 Hot Start High-Fidelity or OneTaq Hot Start DNA polymerase, respectively.

7.4 Agarose gel electrophoresis

Agarose gel electrophoresis was used to isolate DNA products from PCR. 0.7 % agarose was added to 1x TBE buffer (final volume 100 mL) and the mixture was microwave heated until is fully dissolved. The solution was cooled to RT before adding GelRed (0.5 μ L) and left to set on appropriate cast. Once set, the gel was placed into a tank filled with 1x TBE and DNA samples and DNA ladder were loaded (each sampled contained 3 μ L loading dye). The DNA ladders used in this study are: Thermo Scientific GeneRuler 1kb or FastRuler High range DNA ladder. The gel was electrophoresed at 110 V for 45 min. After electrophoresis, the gel fragments can be visualised using UV-transilluminator and the fragments were purified using GeneJET PCR

purification kit. The concentration of the DNA fragment is measured using NanoDrop™ Lite spectrophotometer.

7.5 Construction of recombinant plasmids

7.5.1 Restriction digestion and ligation into pET28a(+)/pMLBAD

PCR amplified fragments (48 µL) with the desired restriction recognition sites, which were included in the designed oligonucleotide primers, were incubated with the appropriate restriction enzymes (1.5 µL each), restriction digestion buffer (5 µL) in a solution of final volume of 50 µL. Similar procedure is applied for the empty pET28a(+) vector. The reaction mixtures were incubated at 37 °C for 1 h and 30 min followed by purification of the digested fragments using GeneJET PCR purification kit. Purified digested fragments and linearized vector plasmid DNA were subsequently ligated using T₄ DNA ligase. The ligation reaction comprises of 50 ng of linearised vector DNA and a 3:1 – 5:1 molar ratio of the digester insert to vector. 1 µL of the T₄ DNA ligase and 4 µL of 5x Rapid DNA Ligation Buffer were added to a final volume solution of 20 µL in dH₂O. Ligation reactions were incubated at room temperature for 3-5 hours and transformed into *E. coli* TOP10 chemical competent cells.

7.5.2 Restriction digestion and ligation into pGPI

The restriction digestion and ligation reaction is performed as described in **section 7.5.1**, however the ligation is incubated at rt overnight. The ligation product is then purified by adding 3 M sodium acetate (2 µL) followed by 95 % EtOH (50 µL). The reaction was left at -20 °C for 5 min. The mixture was then centrifuged at 13000 rpm at 4 °C for 15 min. The supernatant was carefully removed and the pellet was washed with 70% EtOH and left to dry. The purified ligated product was then dissolved in 10 µL of dH₂O before transformation onto *E. coli* SY327 electrocompetent cells.

7.5.3 Preparation of competent cells

For chemical transformation, chemically competent cells of *E. coli* TOP10 or BL21 (DE3) were used. A single colony of the desired *E. coli* strain was inoculated in 5 mL LB and grown at 37 °C overnight in a shaking incubator. 0.5 mL of the overnight culture was subcultured in 50 mL LB in a 250 mL flask at 37°C, 180 rpm shaker until reaching an OD₆₀₀ value of 0.4-0.6. The culture was cooled on an ice bath for 30 min to stop cell growth and cells were harvested by centrifuging for 10 min at 3000 rpm, at 4°C. The cell pellet was resuspended in 30 mL CCC I buffer and kept on ice for 1 h. Cells were harvested and the cell pellet was washed with 30 mL CCC I buffer following resuspension in 1.5 mL CCC II buffer and aliquoted into 50 µL. Aliquots were flash frozen and stored at -80°C.

Electrocompetent *E. coli* SY327 cells were prepared following the same procedure, however 10 % glycerol solution was used instead of CCCI and CCCII.

7.5.4 Chemical transformation

For chemical transformation, chemically competent cells of *E. coli* TOP10 or BL21 (DE3) were used. 50 µL of frozen chemical competent cells were thawed on ice, and 50-100 ng of plasmid DNA was added and mixed gently. The mixture was left on ice for 30 min following heating shocking at 42 °C for 30 sec. The mixture was transferred to an ice box and left for 2 min. 950 µL of sterile LB was added to the mixture and the cell resuspension was incubated for 1 h at 37 °C in a shaking incubator. 100 µL of culture was spread on appropriate selective LB agar plates and incubated at 37 °C overnight.

7.5.5 Electroporation

For transformation using electroporation, electrocompetent cells of *E. coli* SY327 were used. 50 µL of frozen electrocompetent cells were thawed on ice, and 50-100 ng of plasmid DNA was added and mixed gently. The mixture was left on ice for 5 min before being transferred to a pre-cooled 1 mm gap electroporation cuvette (Bio-Rad). The electroporation was performed using a BioRad Gene Pulser® II (200 Ω, 25 µF, 1.8 kV). Subsequently, 950 µL of sterile LB was added

to the mixture and the cell resuspension was incubated for 1 h at 37 °C in a shaking incubator. A series of different volumes of this mixture were then spread onto pre-warmed LB agar okates supplemented with the appropriate antibiotic and the plates were incubated at 37 °C overnight.

7.6 Site-directed mutagenesis

Site-directed mutagenesis was performed using the Q5® Site-Directed Mutagenesis Kit. The pET28a_GdsB-AH-AT-ER tri-domain was used as the template for PCR using Q5® Hot Start High Fidelity Polymerase. The primers were designed using the NEBaseChanger. The resulting linear plasmid (1 µL) was incubated with the kinase, ligase and DpnI (KLD) mix (1 µL) together with the appropriate KLD buffer (5 µL) and H₂O (3 µL). The reaction was left at room temperature for 5 min before transformation using chemically competent cells (*E. coli* TOP10). The generated mutagenesis constructs and the respective primers used for cloning as well as the cell line used for overproduction are summarised in **Table 7.6.2**.

Table 7.6.1 Overview of the primers and mutations for mutant constructs.

Mutant Construct	Primers
pET28a(+)_GdsB (Δ AHAT)	F: GCGGTCCCGCAGATCC
	R: CACCAGGCCGCTGCTGT

7.7 Recombinant protein overproduction and purification

7.7.1 Recombinant protein expression constructs

The constructs generated for recombinant protein overproduction in *E. coli* are shown in **Table 7.7.1**, together with the primers used for cloning. The genomic template used for the generation of the inserts is DNA of *Burkholderia gladioli* BCC1622. Constructs were established by Sanger sequencing (GATC Biotech). The constructs pET28a(+)_SabP, pET28a(+)_P022 and pET24a_ ACP-PBS constructs were generated by Epoch Life Science gene synthesis service.

Table 7.7.1 Recombinant constructs generated using pET28a(+) vector. Size of recombinant protein, insert size and restriction sites are shown.

Protein (kDa / insert size)	Primer
GdsF_M4_ACP_didomain (26.22 kDa/ 786 bp)	F: ATACATATGGTCGATTCCGGCA R: ATAAAGCTTTCAGGTGACGGGA
GdsF_M5_ACP (13.95 kDa/ 382 bp)	F: ATACATATGCCCGCCGAGGAGATTCC R: ATAAAGCTTTTAGACGGCCGCCGCC
GdsF_M6_ACP_tri-domain (42.13 kDa/ 1185 bp)	F: ATACATATGGCGCAGCCGGTCT R: ATAAAGCTTTTACCAGACCTGCGTGA
GdsF_M7_ACP (13.61 kDa/ 381 bp)	F: ATACATATGGTGCCGGCGCCGCA R: ATAAAGCTTTTACGGCGCCGTGGCGGA
GdsF_M8_ACP (14.14 kDa/ 396 bp)	F: ATACATATGCCGGCGGTTCGAGATCAC R: ATAAAGCTTTTACGCAGCCGATGCCTCC
GdsB_tridomain (121.07 kDa/ 3400 bp)	F: ATACATATGCAGCGAAAGACCGTGTTTCTGATGTCT R: ATAGAATTCTCAGACGACGGCCTCGGCCTC
GdsB Δ ATAH (50.73 kDa/ 1425 bp)	F: GCGGTCCCGCAGATCC R: CACCAGGCCGCTGCTGT
GdsH (38.08 kDa/ 1083 bp)	F: ATACATATGACCGCAATGAATCGCCGCAT R: ATAAAGCTTTCAAGCGACCAGCTTCACC ACCT

GdsG	F: ATACATATGAATAGCACCAGCCTGCAAA
(29.50 kDa/ 771 bp)	R: ATAAAGCTTTCACCAGCCGCCT
GdsA	F: ATACATATGCAAAGTCTTTATCTGG
(28.11 kDa/ 759 bp)	R: ATAGAATTCTTACCGATAAGTGAGGAA

7.7.2 Recombinant protein overproduction

To 10 mL sterile LB supplemented with the appropriate antibiotic, a single colony of transformants of *E. coli* BL21 (DE3) was inoculated and the culture was incubated overnight at 37 °C on a shaking incubator. Subsequently, sterile LB (1 L) is inoculated with the overnight culture and incubated at 37 °C on a shaking incubator for approx. 3.5 h to reach an OD₆₀₀ value of 0.6-0.7 followed by induction with 1 mM isopropyl-β-thiogalactopyranoside (IPTG). The culture was left overnight at 15 °C on a shaking incubator to induce protein overproduction.

7.7.3 Recombinant protein purification

The following day the cells were harvested by centrifugation (4000 rpm, 5 °C, 20 min) and the cell pellet was resuspended in loading buffer (10 mL/L of culture). The harvested cells were lysed using a Constant Systems Cell disruptor. The lysate was centrifuged at 17000 rpm, 4 °C for 40 min and the supernatant was filtered with 0.22 µm syringe filter. The filtered supernatant was loaded into 1 mL HiTrap™ HP chelating column (purchased from GE Healthcare). The column was pre-charged with NiSO₄ solution (5 mL, 100 mM) and equilibrated by passing 5 mL of loading buffer. The loaded column was first washed with 20 mL of loading buffer to remove non-specific binding proteins and then washed with a gradient of imidazole solutions [50 mM (5 mL), 100 mM (5 mL), 200 mM (3 mL) and 300 mM (3 mL)]. Fractions with the desired overproduced protein were collected and concentrated using Viva-spin concentrator with various molecular weights, depending on the size of the protein. Upon concentration to 1 mL the buffer was exchanged into storage buffer (2x initial volume of fractions collected) and

concentrated again. 10% glycerol was added to the final volume and the protein aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

7.7.4 SDS-PAGE gel electrophoresis

SDS-PAGE was used to analyse the size of recombinant protein overproduction 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 (protocol described in **Table 7.7.2**), depending on the size of the protein. PageRuler™ Plus Prestained Protein Ladder was used as molecular weight standard. Gels were run using a Mini-PROTEAN® Tetra cell (Bio-Rad, 180 V, 50 min), stained with InstantBlue™ gel staining reagent (Sigma Aldrich) for 30 min and destained overnight in dH₂O supplemented with 70% ethanol.

Table 7.7.2 Solution components and volumes used for the preparation of SDS-PAGE gels. 6% SDS-PAGE gels were used for ≥ 80 kDa proteins, 10% SDS-PAGE gels were used for 20-80 kDa proteins and 15% SDS-PAGE gels were used for ≤ 20 kDa proteins.

Solution Components	Resolving Gel			Stacking Gel	
	6% Volume (mL)	10% Volume (mL)	15% Volume (mL)	Solution Components	Volume (mL)
H ₂ O	2.6	1.9	1.1	H ₂ O	1.4
30% acrylamide	1.0	1.7	2.5	30% acrylamide	0.33
1.5 M Tris (pH 8.8)	1.3	1.3	1.3	1.5 M Tris (pH 6.8)	0.25
10% -SDS	0.05	0.05	0.05	10% -SDS	0.02

10%				10%	
ammonium	0.05	0.05	0.05	ammonium	0.02
persulfate				persulfate	
TEMED	0.004	0.002	0.001	TEMED	0.002

7.8 *In vitro* biochemical assays

7.8.1 *p*-Nitrophenyl (pNPP) colorimetric assay

p-Nitrophenyl was purchased from New England BioLabs. Phosphatase activity was assayed in a reaction mixture of 500 μ L containing 50 mM of pNPP. The reaction was initiated by addition of enzyme (50 μ M) and quenched after 30 min by addition of 1 N NaOH. Phosphatase activity is detected based on alteration in the colour of the reaction.

7.8.2 Enoyl Reduction assay

Substrate loading

For CoA substrate loading assays, 100 μ M ACP was incubated with 2 μ M Sfp (PPtase from *B. subtilis*) 10 mM MgCl₂ and 2 μ L of 20 mM crotonoyl-CoA in Storage Buffer in a total volume of 50 μ L. The reaction was allowed to proceed at rt for 90 min. The reaction was diluted to 20 μ M with the deionised H₂O. The successful loading of the substrate was monitored by UHPLC-ESI-Q-TOF-MS analysis (**section 7.1.2**). All assays were conducted in triplicate.

Monitoring enoyl reduction activity

50 μ M of the loaded ACP with the respective substrate were added to an Eppendorf tube and incubated with 50 μ M of the respective enoyl reductase (GdsB, GdsB Δ ATAH, GdsH), using 500 μ M NADPH as a co-factor. The above reaction was proceeded in a final volume of 20 μ L in

storage buffer. The reaction was allowed to proceed at rt for 240 min. A control reaction was conducted under the same conditions without adding the ER. The enoyl reaction was monitored by injection ion generated from intact ACP species in LC-MS analysis (**section 7.1.2**).

7.8.3 Butenolide formation assay

Storage buffer was used for the enzymatic reactions. 400 μ M of 3-ketooctanoyl NAC thioester (kindly provided by Dr. Shanshan Zhou), 400 μ M of DHAP and 50 μ M of pHis8-ACP-PBS protein in a total volume of 200 μ L were incubated for 90 min at room temperature. 50 μ M of a phosphato-mutase was added subsequently and incubated at room temperature for 50 min. Then, 1 unit of shrimp alkaline phosphatase (New England Biolabs) or GdsA (50 μ M) was added and a further 45-min incubation at 30°C was carried out. On completion of the reaction, 200 μ L of methanol was added to the enzymatic mixture and allowed to stand for 5 min. After centrifugation at 14,200 rpm for 10 min, the supernatant was filtered (0.4 μ m) for UHPLC-ESI-Q-TOF-MS (**section 7.1.2**). All assays were conducted in triplicate.

7.9 Genetic Manipulation of *Burkholderia*

7.9.1 Insertional Mutagenesis in *Burkholderia*

Insertional mutagenesis was performed using the pGPI suicide plasmid.²⁶² The sequences of regions (~ 1 kb) flanking the desired gene were amplified from *B. gladioli* BCC1622 genomic DNA using Q5® Hot Start High-Fidelity DNA Polymerase. Restriction sites were introduced at the 5'- end of the primers to allow the directional cloning of the PCR fragments into the pGPI vector. The resulting construct were transformed into *E. coli* SY327 by electroporation and the transformants were selected on LB agar plates, containing trimethoprim (50 μ g/mL). Plasmids were purified from trimethoprim-resistant colonies using the GeneJET Plasmid MiniPrep Kit and desired constructs were confirmed by Sanger Sequencing (GATC Biotech). The confirmed constructs were subsequently mobilised into *B. gladioli* BCC1622 using triparental mating. *E. coli* SY327 containing the desired plasmid and *E. coli* HB101 (pRK2013) were used as donor and helper strain, respectively. The transconjugants were selected using trimethoprim (150

µg/mL) and polymyxin B (600 U/mL). A single *B. gladioli* insertional mutant was selected, and the correct integration of the mutagenesis plasmids into the genome was confirmed by colony PCR and genome sequencing (GATC Biotech). The generated insertional mutagenesis construct and the respective primers used for cloning are shown in **Table 7.9.1**.

Table 7.9.1 Mutant constructs generated through insertional mutagenesis.

Mutant Construct	Primers (restriction site)	Verification Primers
pGPI- <i>gdsE</i> _Bdomain	F: ATATCTAGACCGAGCC	F: GACAACGTGCTGTCGA
	CGTTCGAGAA	TCGAC
	R: ATAGAATTCGAAGTGC	R:GCCAGGGATGTAACG
	GCGACCT	CACTG

7.9.2 In-frame gene deletions in *Burkholderia*

In-frame deletions of genes from *B. gladioli* BCC1622 were introduced using a double homologous recombination. This was achieved by utilising of the reported pGPI and the pDAI plasmid that encodes a *SceI* endonuclease.¹⁷⁸ Sequences flanking the targeted gene for deletion were amplified from *B. gladioli* BCC1622 genomic DNA and inserted into the pGPI vector using restriction digestion and ligation. The construct was introduced to *E. coli* through electroporation, a process described in **section 7.5.5**. Successful constructs were inserted into *B. gladioli* BCC1622 via tri-parental mating using *E.coli* SY327 (pGPI) and *E. coli* HB101 (pRK2013) as the donor and helper strain, respectively. Transconjugants were selected against trimethoprim (150 µg/mL) and polymyxin (600 U/mL). The correct integration was confirmed through colony PCR. Subsequently, the pDAI plasmid was introduced to the *B. gladioli* single cross-over mutant by triparental mating, using *E. coli* HB101 as a helper strain and *E. coli* SY327 (pDAI) as the donor strain. Transconjugants were selected on LB agar plates containing polymyxin (600 U/mL) and tetracycline (200 µg/mL). Successful mutants were tested for trimethoprim resistance and screened by colony PCR to validate the deletion. The in-frame deletion was confirmed by sequencing. The mutant was grown on minimal medium M9 to remove pDAI-*SceI* which is accompanied with loss of tetracycline resistance. This is an essential step for studying the metabolic profile when mutants have short incubation times since mutants containing tetracycline resistance grow slower. Deletion mutants performed within this project as well as the primers used are shown in **Table 7.9.2**.

Table 7.9.2 Overview of constructs used in in-frame gene deletion within this project.

Mutant construct	Primers (Restriction Enzyme)	Verification Primers
pGPI- <i>gdsA</i>	5F: ATATCTAGAAATAGTCCTTG	F: CAGGATGCGCAGCACCT
	TCCGGC (XbaI)	R: TCGCGCCTGCCTTGAT
	5R: ATACATATGATTTGACCT	
	GGAGCCTC (NdeI)	
	3F: ATACATAGTACCGGCCGTT	
	(NdeI)	
	3R: ATAGGTACCTGCCGTGGACG	
	ATCGT (EcoRI)	
pGPI- <i>gdsG</i>	5F: ATATCTAGAGTCGCCCTG	F: AATCGCCGGGCTTGAAAC
	ATCC (XbaI)	R: AAACGTGCTGCTCAGCG
	5R: ATAAAGCTTATGAAGGAG	
	CCG (HindIII)	
	3F: ATAAAGCTTCTCGCTGCA	
	CAGCTTGTG (HindIII)	
	3R: ATAGAATTCAAGCTGGTG	
	GTGGACGAATC (EcoRI)	
pGPI- <i>gdsF</i> [S ₇₂₉₆ - D ₈₇₇₄]	5F: ATATCTAGAATCGTGCTGC	F: TGGCGCGGGAGCTGGTA
	TCGACCAGACGGGGCGCGT	CAAA
	(XbaI)	R: GTCCCCGGCCGCGCCGA
	5R: ATAAAGCTTCACCGCCGC	ACAGCGCGAGCAACTT
	CGCCGCGTCCGT (HindIII)	
	3F: ATAAAGCTTAGTGCGCTG	
	GTGGAGG (HindIII)	
	R 3: ATAGAATTCCTGCCAGCG	
	TCG (EcoRI)	

pGPI- <i>gdsH</i>	5F: ATATCTAGACAAACGATTCTG	F :ATACATGACCGCAATGA
	CCTCAT (XbaI)	ATCGCCGCAT
	5R: ATAAAGCTTTAGCTCTTGGA	R: ATAACGCTTGCGGTCAA
	TACGCTCA (HindIII)	GATGACTT
	3F:ATATAAAGCTTATCATTGCGG	
	ATCAGG (HindIII)	
	3R: ATAGAATTCATGGCGATCGT	
	CG (EcoRI)	

7.9.3 Gene complementation in *Burkholderia*

For gene complementation in *Burkholderia*, the arabinose-inducible P_{BAD} promoter carrying the expression vector pMLBAD was utilised.¹⁹⁵ The targeted gene was amplified from *B. gladioli* BCC1622 using Q5® Hot Start High-Fidelity DNA Polymerase. Restriction sites were introduced at the 5'- end of the primers to allow the directional cloning of the PCR fragments into the pMLBAD vector. The resulting construct were transformed into *E. coli* TOP10 by chemical transformation and the transformants were selected on LB agar plates, containing trimethoprim (50 µg/mL). Plasmids were purified from trimethoprim-resistant colonies using the GeneJET Plasmid MiniPrep Kit and desired constructs were confirmed by Sanger Sequencing (GATC Biotech). The confirmed constructs were subsequently mobilised into *B. gladioli* BCC1622Δ*gdsG* using triparental mating. *E. coli* SY327 containing the desired plasmid and *E. coli* HB101 (pRK2013) were used as donor and helper strain, respectively. The transconjugants were selected using trimethoprim (150 µg/mL) and the complementation was confirmed by colony PCR and genome sequencing (GATC Biotech). The generated constructs using complementation and their related primers are shown in **Table 7.9.3**.

Table 7.9.3 Constructs created using the pMLBAD vector for gene complementation studies.

Mutant construct	Primers (Restriction Enzyme)
pMLBAD- <i>gdsG</i>	F: ATATCTAGAATGAATAGCACCAGCCTGC (XbaI) R: ATAAGAATTTTACCAGCCGCCT (HindIII)
pMLBAD- <i>gdsH</i>	F: ATATCTAGAATGACCGCAATGAATCGCC (XbaI) R: ATAAAGCTTTCAAGCGACCAGCTTCA (HindIII)
pMLBAD- <i>p022</i>	F: ATATCTAGAATGACGGCACGGCATCT (XbaI) R: ATAAAGCTTTTCAGCCGCCGGT (HindIII)
pMLBAD- <i>sabP</i>	F: ATATCTAGAATGACCGAGCAGC (XbaI) R: ATAAAGCTTTTCACGCCACC (HindIII)

7.10 *Burkholderia* metabolite extraction and analysis

7.10.1 Analysis of metabolite profile of *B. gladioli* strains/knockout mutants and time-course analysis

Burkholderia gladioli strains were handled in a category 2 laboratory equipped with a Bassaire laminar flow hood. To investigate the metabolic profile from *B. gladioli* strains used in this studies as well as investigate changes in gladiostatin biosynthesis because of the generated mutants, the *B. gladioli* strains/mutants were grown overnight in LB (5 mL). The cells were harvested and resuspended in a sterile solution of NaCl (0.9%). The cells were spread on a single plate of BSM containing glucose. Following incubation for one day at 30°C, the cells were removed and the agar was cut into square small pieces and extracted with EtOAc (10 mL). After 2 h, 1 mL the EtOAc extract was added to an Eppendorf tube and centrifuged at 13000 rpm for 15 min. The supernatant was carefully removed and subjected to UHPLC-ESI-Q-TOF-MS analysis (see section 7.1.2).

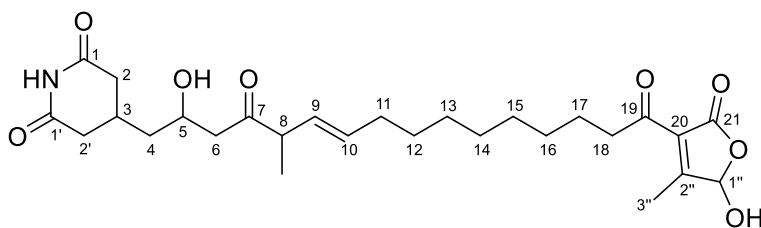
To establish the optimum conditions for gladiostatin production, a time-course analysis was performed by plating *B. gladioli* BCC1622 on Basal Salts Medium (BSM) agar, supplemented

with glucose/ribose/glycerol at 30 °C and extracting using ethyl acetate every 3 hours for 4 days (similar procedure as described above). The extracts were analysed using UHPLC-ESI-Q-TOF-MS.

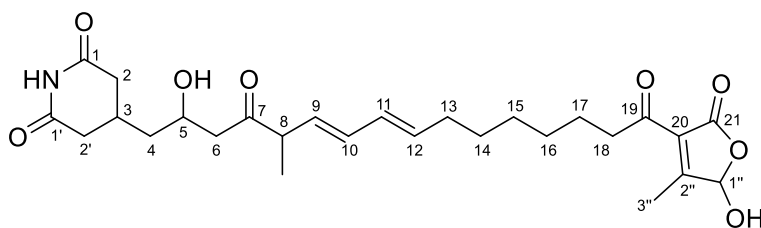
7.10.2 Isolation and structure elucidation of gladiostatin and derivatives and time-course analysis

Ethyl acetate extracts of the agar were fractionated by semi-preparative HPLC on a reverse-phase C18 Betasil column (see **section 7.1.1**), monitoring absorbance at 210 nm. Mobile phases consisted of water (A) and acetonitrile (B). A gradient of 5% B to 100% B over 45 minutes was used at a flow rate of 9 mL/min.

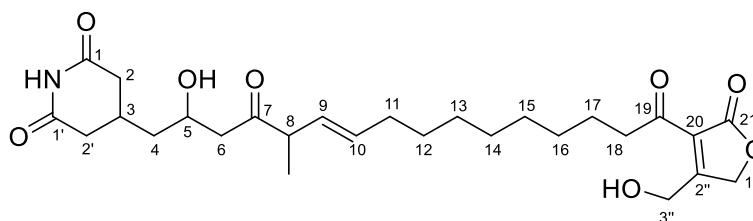
The structure of gladiostatin was elucidated using 1- and 2-D NMR experiments. Purified gladiostatin and gladiostatin derivatives were dissolved in CDCl₃ and ¹H, ¹³C, COSY, HSQC and HSBC spectra were recorded on a Bruker 500 MHz spectrometer equipped with a TCI cryoprobe at 25 °C. The ¹H and ¹³C NMR signals were referenced to the residual protonated solvent at (δ_H 7.26 and δ_C 77.16 for CHCl₃). Assignments of gladiostatin and gladiostatin analogues characterised in this study are shown below in Tables **7.10.1-7.10.4**.

Table 7.10.1 ^1H and ^{13}C NMR data for gladiostatin (49) in CDCl_3 .

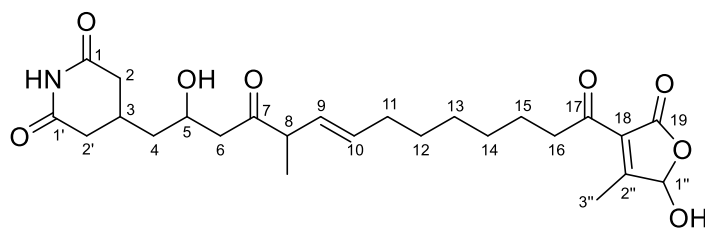
Position	δ_{C} (ppm)	δ_{H} (ppm)	HMBC
1	173.3	-	-
1-NH	-	8.87 (s)	C-1, C-1'
2	37.73	2.27 (m) / 2.73 (m)	C-1, C-3, C-2', C-4
3	27.15	2.44 (m)	C-1, C-2, C-2', C-1', C-4
2'	38.37	2.32 (m) / 2.71 (m)	C-3, C-2, C-1', C-4
1'	173.2	-	-
4	40.90	1.33 (m) / 1.57 (m)	C-2, C-3, C-2', C-5, C-6
5	64.86	4.09 (q, $J = 7.2$ Hz)	C-3, C-4, C-6, C-7
6	47.22	2.57 (m)	C-4, C-5, C-7, C-8
7	213.21	-	-
8	50.96	3.09 (d, $J = 7.4$ Hz)	C-7, C-9, C-10, C-8Me
8-Me	15.93	1.11 (d, $J = 6.8$ Hz)	C-7, C-8, C-9
9	128.33	5.29 (m)	C-8, C-10, C-11, C-8Me
10	134.56	5.56 (dt, $J = 15.4, 6.8$ Hz)	C-8, C-9, C-11
11	32.47	1.98 (t, $J = 7.2$ Hz)	C-9, C-10
12	28.67-29.85	1.23	-
13	28.67-29.85	1.23	-
14	28.67-29.85	1.23	-
15	28.67-29.85	1.23	-
16	28.67-29.85	1.23	-
17	23.24	1.56 (m)	C-18, C-19
18	42.46	2.86 (t, $J = 7.2$ Hz)	C-17, C-19, C-20
19	197.84	-	-
20	126.63	-	-
21	171.82	-	-
1''	97.84	5.91 (s)	C-20, C-2'', C-3''
2''	168.77	-	-
3''	13.69	2.31(s)	C-20, C-2'', C-1''

Table 7.10.2 ^1H and ^{13}C NMR data for gladiostatin B (55) in CDCl_3 .

Position	δ_{C} (ppm)	δ_{H} (ppm)	HMBC
1	173.3	-	-
1-NH	-	8.87 (s)	C-1, C-1'
2	37.73	2.27 (m) / 2.73 (m)	C-1, C-3, C-2', C-4
3	27.15	2.44 (m)	C-1, C-2, C-2', C-1', C-4
2'	38.37	2.32 (m) / 2.71 (m)	C-3, C-2, C-1', C-4
1'	173.2	-	-
4	40.90	1.33 (m) / 1.57 (m)	C-2, C-3, C-2', C-5, C-6
5	64.86	4.09 (q, $J = 7.2$ Hz)	C-3, C-4, C-6, C-7
6	47.22	2.57 (m)	C-4, C-5, C-7, C-8
7	213.21	-	-
8	50.96	3.09 (d, $J = 7.4$ Hz)	C-7, C-9, C-10, C-8Me
8	50.96	3.14 (d, $J = 7.4$ Hz)	C-7, C-9, C-10, C-8Me
8-Me	15.93	1.16 (d, $J = 6.8$ Hz)	C-7, C-8, C-9
9	135.6	5.66 (dt, $J = 14.6, 7.0$ Hz)	C-8, C-10, C-11, C-8Me
10	133.6	6.13 (dd, $J = 15.2, 10.3$ Hz)	C-8, C-9, C-11, C-12
11	129.9	5.98 (dd, $J = 15.2, 10.3$ Hz)	C-9, C-10
12	128.7	5.43 (dd, $J = 15.1, 8.6$ Hz)	C-13, C-14
13	32.4	2.09 (t, $J = 7.2$ Hz)	C-12, C-14, C-15
14	29.7	1.28 (m)	-
15	28.6	1.28 (m)	-
16	28.5	1.28 (m)	-
17	23.24	1.56 (m)	C-18, C-19
19	197.84	-	-
20	126.63	-	-
21	171.82	-	-
1''	97.84	5.91 (s)	C-20, C-2'', C-3''
2''	168.77	-	-
3''	13.69	2.31 (s)	C-20, C-2'', C-1''

Table 7.10.3 ^1H and ^{13}C NMR data for gladiostatin D (59) in acetonitrile- d_3 .

Position	δ_{C} (ppm)	δ_{H} (ppm)	HMBC
1	173.2	-	-
1-NH	-	8.80 (s)	C-1, C-1'
2	37.73	2.27 (m) / 2.73 (m)	C-1, C-3, C-2', C-4
3	27.15	2.44 (m)	C-1, C-2, C-2', C-1', C-4
2'	38.40	2.32 (m) / 2.71 (m)	C-3, C-2, C-1', C-4
1'	173.0	-	-
4	40.70	1.33 (m) / 1.57 (m)	C-2, C-3, C-2', C-5, C-6
5	64.89	4.11 (q, $J = 7.2$ Hz)	C-3, C-4, C-6, C-7
6	47.22	2.57 (m)	C-4, C-5, C-7, C-8
7	213.17	-	-
8	50.96	3.14 (d, $J = 7.4$ Hz)	C-7, C-9, C-10, C-8Me
8-Me	15.93	1.11 (d, $J = 6.8$ Hz)	C-7, C-8, C-9
9	128.33	5.29 (m)	C-8, C-10, C-11, C-8Me
10	134.56	5.56 (dt, $J = 15.4, 6.8$ Hz)	C-8, C-9, C-11
11	32.47	1.98 (t, $J = 7.2$ Hz)	C-9, C-10
12	28.65-29.88	1.25	-
13	28.65-29.88	1.25	-
14	28.65-29.88	1.25	-
15	28.65-29.88	1.25	-
16	28.65-29.88	1.25	-
17	23.24	1.56 (m)	C-18, C-19
18	42.46	2.86 (t, $J = 7.2$ Hz)	C-17, C-19, C-20
19	197.84	-	-
20	125.68	-	-
21	171.72	-	-
1''	71.2	5.15 (s)	C-20, C-21, C-2''
2''	170.23	-	-
3''	60.87	5.02 (s)	C-20, C-1'', C-2''

Table 7.10.4 ^1H and ^{13}C NMR data for gladiostatin C (78) in CDCl_3 .

Position	δ_{C} (ppm)	δ_{H} (ppm)	HMBC
1	173.29	-	-
1-NH	-	8.84 (s)	C-1, C-1'
2	37.71	2.27 (m) / 2.73 (m)	C-1, C-3, C-2', C-4
3	27.15	2.44 (m)	C-1, C-2, C-2', C-1', C-4
2'	38.35	2.32 (m) / 2.71 (m)	C-3, C-2, C-1', C-4
1'	173.2	-	-
4	40.87	1.37 (m) / 1.56 (m)	C-2, C-3, C-2', C-5, C-6
5	64.86	4.09 (q, $J = 7.2$ Hz)	C-3, C-4, C-6, C-7
6	47.22	2.58 (m)	C-4, C-5, C-7, C-8
7	213.21	-	-
8	50.96	3.10 (d, $J = 7.4$ Hz)	C-7, C-9, C-10, C-8Me
8-Me	15.91	1.12 (d, $J = 6.8$ Hz)	C-7, C-8, C-9
9	128.40	5.31 (m)	C-8, C-10, C-11, C-8Me
10	134.35	5.55 (dt, $J = 14.2, 6.8$ Hz)	C-8, C-9, C-11
11	32.34	1.98 (t, $J = 7.2$ Hz)	C-9, C-10
12	29.66	1.23	-
13	28.69	1.23	-
14	28.51	1.23	-
15	23.21	1.57 (dq, , $J = 14.8, 7.1$ Hz)	C-16, C-17
16	42.39	2.87 (t, $J = 7.3$ Hz)	C-15, C-17, C-18
17	197.84	-	-
18	126.61	-	-
19	171.83	-	-
1''	97.86	5.91 (s)	C-18, C-2'', C-3''
2''	168.77	-	-
3''	13.69	2.32(s)	C-18, C-2'', C-1''

7.11 *In vivo* biological activity assays

7.11.1 Broth microdilution method

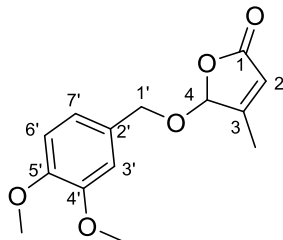
MICs performed in bacterial strains were determined using the broth microdilution method (National Committee for Clinical Laboratory Standards, document M7-A10).²⁶³ Overnight cultures of the test strains were grown in 5 mL Mueller Hinton media at 30 °C. The antibiotics were dissolved in DMSO to a concentration of 5 mg/mL. In a 96-well microtiter plate, 50 µL of 2-fold serial dilutions of the antibiotic (128 µg/mL - 0.0025 µg/mL) in Mueller Hinton media were mixed with 50 µL of microbial suspension that was made up, from the overnight cultures, to McFarland standard No. 0.5 following a 100-fold dilution, performed in Mueller Hinton media. The assays were incubated for 16 hr at 30 °C, and the MIC was defined as the lowest concentration that visibly inhibited bacterial growth. Each assay was performed in triplicate.

7.11.2 Antifungal susceptibility testing against yeast strain

For the antifungal susceptibility testing against yeast strain, a reference method was used following the CSLI approved standard.¹⁷³ *S. cerevisiae* W303-1a was grown overnight in YPD at 30 °C. The culture was resuspended in PBS and standardized using 0.5 McFarland turbidity standard. The working solution was prepared using a 1:1000 dilution in RPMI (1 to 5*10³ cells/ml). 0.1 ml of the final culture dilution was added into each well. Drug dilutions were added to each well before the addition of the culture, using 0.1 µl of serial two-fold dilutions of gladiostatin in RPMI. Sterility control (no cells added) and growth control (DMSO control) were included in the 96-well plate. The plate was incubated without shaking at 35°C for 24h. All MIC determinations were performed in triplicate.

7.12 Chemical synthesis of authentic standard

5-((3,4-dimethoxybenzyl)oxy)-4-methylfuran-2(5H)-one (74)

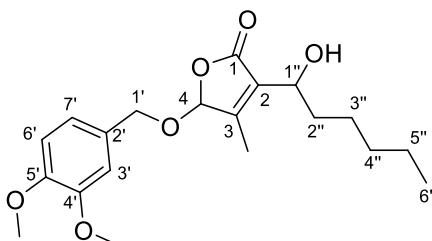


A mixture of 5-hydroxy-4-methylfuran-2(5H)-one (**73**) (1.5 g, 13.14 mmol), dimethoxybenzyl alcohol (2.05 ml, 14.46 mmol, 1.1 equiv.) and catalytic amount of Amberlyst-15 resin (10% w/w, 150 mg) in dry CH_2Cl_2 (30 mL) was refluxed. After 48 h, the crude reaction mixture was filtered using a Whatman filter paper and the residue washed with CH_2Cl_2 , dried over anhydrous Na_2SO_4 , concentrated in *vacuo* and purified using silica chromatography (pet ether : ethyl acetate, 95:5) to afford 5-((3,4-dimethoxybenzyl)oxy)-4-methylfuran-2(5H)-one (**74**) (1.47 g, 43%) as pale yellow oil.

^1H NMR (400 MHz, CDCl_3) δ : 1.93 (s, 3H, 3-Me), 3.84 (s, 3H, 4'-OMe), 3.86 (s, 3H, 5'-OMe), 4.43 (s, 2H, H-1'), 5.82 (s, 1H, H-4) 5.93 (s, 1H, H-2), 6.89 - 6.78 (m, 3H, H-3', H- 6', H-7')

^{13}C NMR (100 MHz, CDCl_3) δ : 171.4 (C-1), 155.3 (C-3), 149.0 (C-4'), 148.6 (C-5'), 130.8 (C-2'), 118.5 (C-2), 121.9 (C-7'), 112.3 (C-3'), 112.1 (C-6'), 103.4 (C-4), 74.3 (C-1'), 55.9 (4'-OMe), 55.8 (5'-OMe), 13.8 (3-Me)

HR-MS: m/z calculated for $\text{C}_{14}\text{H}_{16}\text{NaO}_5$ $[\text{M}+\text{Na}]^+$: 287.0889; found: 287.0892.

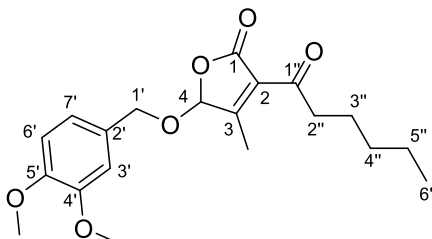
5-((3,4-dimethoxybenzyl)oxy)-3-(1-hydroxyhexyl)-4-methylfuran-2(5H)-one (75)

A solution of *n*-butyllithium (1.65 M in hexane, 2.75 mL, 4.54 mmol, 1.2 equiv.) was added dropwise at 0 °C to a solution of diisopropylamine (629 μ L, 4.54 mmol, 1.2 equiv.) in THF (10 mL). After the system has been stirred for 30 min, a solution of 5-((3,4-dimethoxybenzyl)oxy)-4-methylfuran-2(5H)-one (**74**) (1 g, 3.78 mmol) in THF (12 mL) was added drop-wise at -78 °C and the mixture was stirred at the same temperature for 30 min. Hexanal (0.55 mL, 4.54 mmol, 1.2 equiv.) was added drop-wise at -78 °C and further stirred for 2 h. The solution was quenched with saturated aqueous NH_4Cl (10 mL) and slowly warmed to rt. The layers were separated and the aqueous layer was extracted with EtOAc (3 x 30 mL). The combined organic layers were washed with brine (30 mL), dried over anhydrous Na_2SO_4 and concentrated. The residue was purified using silica chromatography (hexane: EtOAc, 4:1) to afford 5-((3,4-dimethoxybenzyl)oxy)-3-(1-hydroxyhexyl)-4-methylfuran-2(5H)-one (**75**) (438 mg, 32%) as a colorless oil mixture of diastereomers that used directly for the next step.

^1H NMR (300 MHz, CDCl_3) δ : 0.90 (t, J = 7.0 Hz, 3H, H-6''), 1.25-1.36 (m, 6H, H-3'', H-4'', H-5''), 1.47 - 1.68 (m, 2H, H-2'') 1.93 (s, 3H, 3-Me), 3.45 (br. s, 1H, 1''-OH), 3.81 (s, 3H, 4'-OMe), 3.85 (s, 3H, 5'-OMe), 4.43 (s, 2H, H-1'), 4.45-4.50 (m, 1H, H-1''), 5.82 (s, 1H, H-4), 6.89 - 6.78 (m, 3H, H-3', H-6', H-7')

^{13}C NMR (75 MHz, CDCl_3) δ : 171.4 (C-1), 155.3 (C-3), 149.0 (C-4'), 148.6 (C-5'), 130.8 (C-2'), 129.4 (C-2), 121.9 (C-7') 112.3 (C-3'), 112.1 (C-6'), 103.4 (C-4), 74.3 (C-1'), 66.5 (C-1'') 55.9 (4'-OMe), 55.8 (5'-OMe), 35.7 (C-2''), 31.4 (C-5''), 25.3 (C-3''), 22.4 (C-6''), 14.0 (C-6''), 13.8 (3-Me)

HR-MS: m/z calculated for $\text{C}_{20}\text{H}_{28}\text{NaO}_6$ $[\text{M}+\text{Na}]^+$: 387.1778; found: 387.1772.

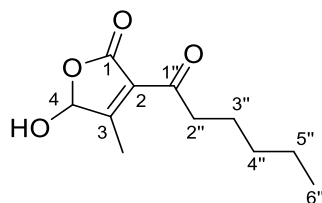
5-((3,4-dimethoxybenzyl)oxy)-3-hexanoyl-4-methylfuran-2(5H)-one (76)

Dess-Martin periodinane (15% wt., 3.42 mL, 1.21 mmol) was added to a solution of 5-((3,4-dimethoxybenzyl)oxy)-3-(1-hydroxyhexyl)-4-methylfuran-2(5H)-one (**75**) (400 mg, 1.1 mmol) in CH_2Cl_2 (10 mL) and stirred at room temperature for 2 h. TLC confirmed complete conversion of starting material to a less polar product. The reaction was quenched by the addition of saturated aqueous solution of $\text{Na}_2\text{S}_2\text{O}_3$ (15 mL) and NaHCO_3 (2 mL) stirred for a further 30 minutes. The layers were separated and the aqueous later was extracted with CH_2Cl_2 (3 x 30 mL). The combined organics were washed with brine (30 mL), dried over MgSO_4 and concentrated. Standard silica gel chromatography was carried out eluting with 20% EtOAc in petroleum ether 40-60 °C to give 5-((3,4-dimethoxybenzyl)oxy)-3-hexanoyl-4-methylfuran-2(5H)-one (**76**) (180 mg, 46%) as a colourless oil.

^1H NMR (400 MHz, CDCl_3) δ : 0.92 (t, $J = 7.0$ Hz, 3H, H-6''), 1.29-1.32 (m, 4H, H-4'', H-5''), 1.59 (q, 2H, H-3''), 1.91 (s, 3H, 3-Me), 2.95 (t, $J = 7.5$ Hz, 2H, H-2''), 3.81 (s, 3H, 4'-OMe), 3.85 (s, 3H, 5'-OMe), 4.43 (s, 2H, H-1'), 5.82 (s, 1H, H-4), 6.89 - 6.78 (m, 3H, H-3', H- 6', H-7')

^{13}C NMR (100 MHz, CDCl_3) δ : 197.1 (C-1''), 172.6 (C-1), 155.3 (C-3), 149.0 (C-4'), 148.6 (C-5'), 130.8 (C-2'), 130.1 (C-2), 121.9 (C-7'), 112.3 (C-3'), 112.1 (C-6'), 103.4 (C-4), 74.3 (C-1'), 55.9 (4'-OMe), 55.7 (5'-OMe), 41.5 (C-2''), 31.2 (C-3''), 22.8 (C-4''), 22.4 (C-5''), 14.0 (C-6''), 13.8 (3-Me)

HR-MS: m/z calculated for $\text{C}_{20}\text{H}_{26}\text{NaO}_6$ $[\text{M}+\text{Na}]^+$: 385.1621; found: 385.1625.

3-hexanoyl-5-hydroxy-4-methylfuran-2(5H)-one (72)

To a stirred solution of 5-((3,4-dimethoxybenzyl)oxy)-3-hexanoyl-4-methylfuran-2(5H)-one (**76**) (200 mg, 0.55 mmol) in dry CH_2Cl_2 (5 mL), phenyliodine (III) bis(trifluoroacetate) (PIFA) (0.47 mg, 1.1 mmol) was added in small portions at room temperature. After 3 h, the reaction was quenched with saturated aqueous $\text{Na}_2\text{S}_2\text{O}_3$ (10 mL) at room temperature, and the mixture was extracted with CHCl_3 (3 x 10 mL). The organic layers were combined and washed with brine (1 x 10 mL), then dried over Na_2SO_4 . Concentration of the solvent in *vacuo* afforded a residue, which was purified by column chromatography (EtOAc/hexane 1:2) to afford 3-hexanoyl-5-hydroxy-4-methylfuran-2(5H)-one (**72**) (66 mg, 57%) as a colourless oil.

^1H NMR (300 MHz, CDCl_3) δ : 0.91 (t, $J = 7.0$ Hz, 3H, H-6''), 1.29-1.32 (m, 4H, H-4'', H-5''), 1.59 (q, 2H, H-3''), 1.98 (s, 3H, 3-Me), 2.95 (t, $J = 7.5$ Hz, 2H, H-2''), 5.85 (s, 1H, H-4)

^{13}C NMR (75 MHz, CDCl_3) δ : 197.0 (C-1''), 172.3 (C-1), 155.3 (C-3), 129.9 (C-2), 98.5 (C-4), 41.5 (C-2''), 31.2 (C-3''), 22.8 (C-4''), 22.4 (C-5''), 13.9 (C-6''), 13.6 (3-Me)

HR-MS: m/z calculated for $\text{C}_{11}\text{H}_{16}\text{NaO}_4$ $[\text{M}+\text{Na}]^+$: 235.0940; found: 235.0945.

7.13 Computational methods

7.13.1 Sequence alignments

Multiple sequence alignment was performed using ClustalX 2.1.²⁴⁹ Protein sequences were obtained from NCBI repository. ESPript²⁵⁴ was used to generate secondary structure annotations.

7.13.2 Phylogenetic analyses

The phylogenetic trees shown in this study were generated using ClustalX 2.1²⁴⁹ using the neighbour joining clustering algorithm. Phylogenetic trees were visualised using FigTree v1.4.3.

7.13.3 Cladogram construction

The ACP cladogram was generated in a similar manner to previous reports.¹⁹⁷ Well-characterised transAT PKSs were extracted from the MiBiG repository: (albicidin, Alb, BGC0001088; bacillaene, Bae, BGC0001089; batumin, Bat, BGC0001099; bongkrelic acid, Bon, BGC0000173; bryostatin, Bry, BGC0000174; calyculin, Cal, BGC0000967; chivosazole, Chi, BGC0001069; coralopyronin, Cor, BGC0001091; cycloheximide, Chx, BGC0000175; difficidin, Dif, BGC0000176; disorazole, Dsz, BGC0001093; elansolid, Ela, BGC0000178; enacyloxin, Ena, BGC0001094; etnangien, Etn, BGC0000179; kirromycin, Kir, BGC0001070; lactimidomycin, Ltm, BGC0000083; leinamycin, Lnm, BGC0001101; macrolactin, Mln, BGC0000181; migrastatin, Mgs, BGC0000177; nosperin, Nsp, BGC0001071; oocydin, Ooc, BGC0001031; onnamide, Onn, BGC0001105; psymberin, Psy, BGC0001110; rhizopodins, Riz, BGC0001111; rhizoxins, Rhi, BGC0001112; sorangicin, Sor, BGC0000184; thailandamide, Tai, BGC0000186, and thailanstatin, Tst, BGC0001114)⁹. ACP cladograms were rooted using representative domains from *cis*-AT assembly lines (amphotericin module 10, Amp10, BGC0000015; concanamycin module 6, Con6, BGC0000040; erythromycin module 4, Ery4, BGC0000055; mycolactone module 4, Myc4, BGC0000103; nystatin module 9, Nys9, BGC0000115; rapamycin module 6, Rap6, BGC0001040; rifamycin module 5, Rif5,

BGC0000136; spinosad module 5, Spn5, BGC0000148; and tylosin module 3, Tyl3, BGC0000166).²⁶⁴ Sequences were aligned using the ClustalX multiple sequence alignment software and RaxML- NG for maximum likelihood phylogenetic inference.²⁶⁵ The cladogram was visualised using FigTree v1.4.4 software.

7.13.4 Homology Modelling

A homology model of the chimeric GdsF' module 7 KR was constructed using intensive modelling mode of Phyre2²⁴⁷. The model was based on PksJ KR1 (PDB: 4J1S), a ketoreductase domain of the bacillaene *trans*-AT PKS. The quality of the model was assessed using Ramachandran plots constructed using MOLProbity Server²⁶⁶. Addition of NADPH cofactor was achieved by alignment of the respective homology model to the, *B. subtilis* module KR of the bacillaene *trans*-AT PKS (PDB: 5KTK).

References

1. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *Journal of Natural Products* (2020).
2. Williams, D. H., Stone, M. J., Hauck, P. R. & Rahman, S. K. Why are secondary metabolites (Natural Products) biosynthesized. *J. Nat. Prod.* **52**, 1189–208 (1989).
3. Demain, A. L. & Fang, A. The natural functions of secondary metabolites. *Adv. Biochem. Eng. Biotechnol.* **69**, 1–39 (2000).
4. Fleming, A. On The Antibacterial Action Of Cultures Of a *Penicillium*, With Special Reference To Their Use In The Isolation Of *B. Influenzae*. *Br. J. Exp. Pathol.* **10**, 26–236 (1929).
5. Waksman, S. A., Bugie, E. & Reilly, H. C. Bacteriostatic and Bactericidal Properties of Antibiotic Substances, with Special Reference to Plant-Pathogenic Bacteria. *Bull. Torrey Bot. Club* **71**, 107–21 (1944).
6. Waksman, S. A., Reilly, H. C., Johnstone, D. B. Isolation of streptomycin-producing strains of *Streptomyces griseus*. *J. Bacteriol.* **52**, 393–397 (1946).
7. Schatz, A. & Waksman, S. A. Effect of Streptomycin and Other Antibiotic Substances upon *Mycobacterium tuberculosis* and Related Organisms. *Proc. Soc. Exp. Biol. Med.* **57**, 244–248 (1944).
8. Duggar, B. M. Aureomycin; a product of the continuing search for new antibiotics. *Ann. N. Y. Acad. Sci.* **51**, 177–81 (1948).
9. Newton, G. G. F. & Abraham, E. P. Cephalosporin C, a new antibiotic containing sulphur and D- α -aminoadipic acid. *Nature* **175**, 548 (1955).
10. Geraci, J. E., Heilman, F. R., Nichols, D. R., Ross, G. T. & Well, W. E. Some laboratory and clinical experiences with a new antibiotic, vancomycin. *Antibiot Annu.*

31, 90–106 (1956).

11. David, B., Wolfender, J. L. & Dias, D. A. The pharmaceutical industry and natural products: historical status and new trends. *Phytochem Rev* **14**, 299–315 (2015).

12. Baltz, R. H. Natural product drug discovery in the genomic era: realities, conjectures, misconceptions, and opportunities. *J. Ind. Microbiol. Biotechnol.* **46**, 281–299 (2019).

13. Lerner, C. G. *et al.* From bacterial genomes to novel antibacterial agents: Discovery, characterization, and antibacterial activity of compounds that bind to HI0065 (YjeE) from *Haemophilus influenzae*. *Chem. Biol. Drug Des.* **69**, 395–404 (2007).

14. Challis, G. L. Genome mining for novel natural product discovery. *J Med Chem.* **51**, 2618–2628 (2008).

15. Gross, H. Genomic mining - A concept for the discovery of new bioactive natural products. *Curr Opin Drug Discov Devel.* **12**, 207–19 (2009).

16. Corre, C. & Challis, G. L. New natural product biosynthetic chemistry discovered by genome mining. *Nat. Prod. Rep.* **26**, 977–86 (2009).

17. Ziemert, N., Alanjary, M. & Weber, T. The evolution of genome mining in microbes-a review. *Nat. Prod. Rep.* **33**, 988–1005 (2016).

18. Amos, G. C. A. *et al.* Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E11121–E11130 (2017).

19. Lee, N. *et al.* Synthetic Biology Tools for Novel Secondary Metabolite Discovery in *Streptomyces*. *J. Microbiol. Biotechnol.* **29**, 667–686 (2019).

20. Laureti, L. *et al.* Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6258–6263 (2011).

21. Pham, V. H. T. & Kim, J. Cultivation of unculturable soil bacteria. *Trends Biotechnol.* **30**, 475–484 (2012).
22. Masschelein, J., Jenner, M. & Challis, G. L. Antibiotics from Gram-negative bacteria: A comprehensive overview and selected biosynthetic highlights. *Nat. Prod. Rep.* **34**, 712–783 (2017).
23. Mahenthiralingam, E., Urban, T. A. & Goldberg, J. B. The multifarious, multireplicon *Burkholderia cepacia* complex. *Nat. Rev. Microbiol.* **3**, 144–156 (2005).
24. Eberl, L. & Vandamme, P. Members of the genus *Burkholderia*: Good and bad guys. *F1000Research* vol. 5 (2016).
25. Sawana, A., Adeolu, M. & Gupta, R. S. Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: Proposal for division of this genus into the emended genus *Burkholderia* containing pathogenic organisms and a new genus *Paraburkholderia* gen. nov. harboring environmental species. *Front. Genet.* (2014).
26. Dobritsa, A. P. & Samadpour, M. Transfer of eleven species of the genus *Burkholderia* to the genus *Paraburkholderia* and proposal of *Caballeronia* gen. nov. to accommodate twelve species of the genera *Burkholderia* and *Paraburkholderia*. *Int. J. Syst. Evol. Microbiol.* **66**, 2836–2846 (2016).
27. Estrada-de los Santos, P. *et al.* Whole genome analyses suggests that *Burkholderia* sensu lato contains two additional novel genera (*Mycetohabitans* gen. nov., and *Trinickia* gen. nov.): Implications for the evolution of diazotrophy and nodulation in the *Burkholderiaceae*. *Genes (Basel)*. **9**, (2018).
28. Walter H. Burkholder. Sour skin, a bacterial rot of Onion bulbs. *Phytopathology* **40**, 115–117 (1950).
29. Kersters, K. *et al.* Recent changes in the classification of the pseudomonads: An overview. *Syst. Appl. Microbiol.* **19**, 465–477 (1996).
30. Palleroni, N. J., Kunisawa, R., Contopoulou, R. & Doudoroff, M. Nucleic acid

- homologies in the genus *Pseudomonas*. *Int. J. Syst. Bacteriol.* **23**, 333–339 (1973).
31. Hotta, H. *et al.* Proposal of *Burkholderia* gen. Anov. 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.* **36**, 1251–1275 (1992).
 32. Parke, J. L. & Gurian-Sherman, D. Diversity of the burkholderia cepacia complex and implications for risk assessment of biological control strains. *Annu. Rev. Phytopathol.* **39**, 225–258 (2001).
 33. Partida-Martinez, L. P. *et al.* *Burkholderia rhizoxinica* sp. nov. and *Burkholderia endofungorum* sp. nov., bacterial endosymbionts of the plant-pathogenic fungus *Rhizopus microsporus*. *Int. J. Syst. Evol. Microbiol.* **57**, 2583–2590 (2007).
 34. LiPuma, J. J. The changing microbial epidemiology in cystic fibrosis. *Clin. Microbiol. Rev.* **23**, 299–323 (2010).
 35. Mahenthiralingam, E., Baldwin, A. & Dowson, C. G. *Burkholderia cepacia* complex bacteria: opportunistic pathogens with important natural biology. *J. Appl. Microbiol.* **104**, 1539–1551 (2008).
 36. Tom, C., Peter, V., W., G. J. R. & J., L. J. Taxonomy and Identification of the *Burkholderia cepacia* Complex. *J. Clin. Microbiol.* **39**, 3427–3436 (2001).
 37. Gilad, J., Harary, I., Dushnitsky, T., Schwartz, D. & Amsalem, Y. *Burkholderia mallei* and *Burkholderia pseudomallei* as bioterrorism agents: National aspects of emergency preparedness. *Isr. Med. Assoc. J.* **9**, 499–503 (2007).
 38. Silva, E. B. & Dow, S. W. Development of *Burkholderia mallei* and *pseudomallei* vaccines. *Front. Cell. Infect. Microbiol.* **3**, (2013).
 39. Jeong, Y. *et al.* Toxoflavin produced by *Burkholderia glumae* causing rice grain rot is responsible for inducing bacterial wilt in many field crops. *Plant Dis.* **87**, 890–895 (2003).

40. Azegami, K. *et al.* *Pseudomonas plantarii* sp. nov., the Causal Agent of Rice Seedling Blight. *Int. J. Syst. Bacteriol.* **37**, 144–152 (1987).
41. Mullins, A. J. *et al.* Genome mining identifies cepacin as a plant-protective metabolite of the biopesticidal bacterium *Burkholderia ambifaria*. *Nat. Microbiol.* **4**, 996–1005 (2019).
42. Wang, X. Q. *et al.* Occidiofungin is an important component responsible for the antifungal activity of *Burkholderia pyrrocinia* strain Lyc2. *J. Appl. Microbiol.* **120**, 607–618 (2016).
43. Mullins, A. J. *et al.* Genomic Assemblies of Members of *Burkholderia* and Related Genera as a Resource for Natural Product Discovery. *Microbiol. Resour. Announc.* **9**, (2020).
44. Wu, Y. & Seyedsayamdost, M. R. The Polyene Natural Product Thailandamide A Inhibits Fatty Acid Biosynthesis in Gram-Positive and Gram-Negative Bacteria. *Biochemistry* **57**, 4247–4251 (2018).
45. Mahenthiralingam, E. *et al.* Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic *burkholderia ambifaria* genomic island. *Chem. Biol.* **18**, 665–77 (2011).
46. Howard, A., O'Donoghue, M., Feeney, A. & Sleator, R. D. *Acinetobacter baumannii* An emerging opportunistic pathogen. *Virulence* **71**, 292–301 (2012).
47. Esmaeel, Q., Pupin, M., Jacques, P. & Leclère, V. Nonribosomal peptides and polyketides of *Burkholderia*: new compounds potentially implicated in biocontrol and pharmaceuticals. *Environ. Sci. Pollut. Res.* **25**, 29794–29807 (2018).
48. Esmaeel, Q. *et al.* *Burkholderia* genome mining for nonribosomal peptide synthetases reveals a great potential for novel siderophores and lipopeptides synthesis. *Microbiologyopen* **5**, 512–26. (2016).
49. Collie, N. & Myers, W. S. The formation of orcinol and other condensation products from dehydracetic acid. *J. Chem. Soc. Trans.* **63**, 122–128 (1893).

50. Birch, A. J., Massy-Westropp, R. A. & Moye, C. J. Studies in relation to biosynthesis. VII. 2-hydroxy-6-methylbenzoic acid in penicillium griseofulvum Dierckx. *Aust. J. Chem.* **8**, 539–544 (1955).
51. Malpartida, F. & Hopwood, D. A. Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. *Nature* **309**, 462–464 (1984).
52. Tuan, J. S. *et al.* Cloning of genes involved in erythromycin biosynthesis from Saccharopolyspora erythraea using a novel actinomycete-Escherichia coli cosmid. *Gene* **90**, 21–29 (1990).
53. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. Modular organization of genes required for complex polyketide biosynthesis. *Science* (80-.). **252**, 675–679 (1991).
54. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. *Nature* **348**, 176–178 (1990).
55. Jenke-Kodama, H., Sandmann, A., Müller, R. & Dittmann, E. Evolutionary implications of bacterial polyketide synthases. *Mol. Biol. Evol.* **22**, 2027–2039 (2005).
56. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: A millennium review. *Nat. Prod. Rep.* **8**, 380–416 (2001).
57. Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **7**, 285–95 (2003).
58. Olano, C. *et al.* Biosynthesis of the Angiogenesis Inhibitor Borrelidin by Streptomyces parvulus Tü4055: Cluster Analysis and Assignment of Functions. *Chem. Biol.* **11**, 87–97 (2004).
59. Gaitatzis, N. *et al.* The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide

- synthase. *J. Biol. Chem.* **277**, 13082–13090 (2002).
60. Bevitt DJ, Cortes J, Haydock SF, L. P. 6-Deoxyerythronolide-B synthase 2 from *Saccharopolyspora erythraea* Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. *Eur. J. Biochem.* **204**, 39–49 (1992).
61. Helfrich, E. J. N. & Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **27**, 996–1047 (2016).
62. Keatinge-Clay, A. T. The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29**, 1050–1073 (2012).
63. Zhang, L. *et al.* Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. *Angew. Chemie - Int. Ed.* **56**, 1740–1745 (2017).
64. Keatinge-Clay, A. T. Polyketide Synthase Modules Redefined. *Angew. Chemie - Int. Ed.* **56**, 4658–4660 (2017).
65. Lambalot, R. H. *et al.* A new enzyme superfamily - The phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923–936 (1996).
66. 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.* **13**, 625–636 (2006).
67. Smith, S. & Tsai, S. C. The type I fatty acid and polyketide synthases: A tale of two megasynthases. *Nat. Prod. Rep.* **4**, 1041–72 (2007).
68. Chen, A. Y., Schnarr, N. A., Kim, C. Y., Cane, D. E. & Khosla, C. Extender unit and acyl carrier protein specificity of ketosynthase domains of the 6-deoxyerythronolide B synthase. *J. Am. Chem. Soc.* **128**, 3067–3074 (2006).
69. 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.* **16**, 2093–2107

(2007).

70. Chandran, S. S., Menzella, H. G., Carney, J. R. & Santi, D. V. Activating Hybrid Modular Interfaces in Synthetic Polyketide Synthases by Cassette Replacement of Ketosynthase Domains. *Chem. Biol.* **13**, 469–474 (2006).
71. Roujeinikova, A. *et al.* Structural Studies of Fatty Acyl-(Acyl Carrier Protein) Thioesters Reveal a Hydrophobic Binding Cavity that Can Expand to Fit Longer Substrates. *J. Mol. Biol.* **365**, 135–145 (2007).
72. Busche, A. *et al.* Characterization of molecular interactions between ACP and halogenase domains in the curacin A polyketide synthase. *ACS Chem. Biol.* **7**, 378–386 (2012).
73. Farmer, R., Thomas, C. M. & Winn, P. J. Structure, function and dynamics in acyl carrier proteins. *PLoS One* **14**, (2019).
74. Serre, L., Verbree, E. C., Dauter, Z., Stuitje, A. R. & Derewenda, Z. S. The Escherichia coli malonyl-CoA:acyl carrier protein transacylase at 1.5-Å resolution: Crystal structure of a fatty acid synthase component. *J. Biol. Chem.* **270**, 12961–12964 (1995).
75. Jensen, K. *et al.* Polyketide proofreading by an acyltransferase-like enzyme. *Chem. Biol.* **19**, 329–39 (2012).
76. Park, H., Kevany, B. M., Dyer, D. H., Thomas, M. G. & Forest, K. T. A polyketide synthase acyltransferase domain structure suggests a recognition mechanism for its hydroxymalonyl-acyl carrier protein substrate. *PLoS One* **9**, (2014).
77. Marsden, A. F. A. *et al.* Stereospecific acyl transfers on the erythromycin-producing polyketide synthase. *Science (80-.)*. **263**, 378–380 (1994).
78. Hopwood, D. A. Complex enzymes in microbial natural product biosynthesis, part B: polyketides, aminocoumarins and carbohydrates. Preface. *Methods Enzymol.* **459**, (2009).

79. Dunn, B. J., Cane, D. E. & Khosla, C. Mechanism and specificity of an acyltransferase domain from a modular polyketide synthase. *Biochemistry* **52**, 1839–1841 (2013).
80. 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* **55**, 4476–4484 (2016).
81. 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.* **281**, 17390–17399 (2006).
82. Kwan, D. H. & Schulz, F. The stereochemistry of complex polyketide biosynthesis by modular polyketide synthases. *Molecules* **16**, 6092–6115 (2011).
83. Kavanagh, K. L., Jörnvall, H., Persson, B. & Oppermann, U. Medium- and short-chain dehydrogenase/reductase gene and protein families: The SDR superfamily: Functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* **65**, 3895–3906 (2008).
84. 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* **14**, 737–748 (2006).
85. Keatinge-Clay, A. T. A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. *Chem. Biol.* **14**, 898–908 (2007).
86. Caffrey, P. Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *ChemBioChem* **4**, 654–657 (2003).
87. Xie, X., Garg, A., Keatinge-Clay, A. T., Khosla, C. & Cane, D. E. Epimerase and Reductase Activities of Polyketide Synthase Ketoreductase Domains Utilize the Same Conserved Tyrosine and Serine Residues. *Biochemistry* **55**, 1179–1186 (2016).

88. Keatinge-Clay, A. Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *J. Mol. Biol.* **384**, 941–53 (2008).
89. Akey, D. L. *et al.* Crystal Structures of Dehydratase Domains from the Curacin Polyketide Biosynthetic Pathway. *Structure* **18**, 94–105 (2010).
90. 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* **52**, 8916–8928 (2013).
91. 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.* **132**, 14694–14696 (2010).
92. Valenzano, C. R. *et al.* Stereospecificity of the dehydratase domain of the erythromycin polyketide synthase. *J. Am. Chem. Soc.* **132**, 14697–14699 (2010).
93. Xie, X. & Cane, D. E. PH-Rate profiles establish that polyketide synthase dehydratase domains utilize a single-base mechanism. *Org. Biomol. Chem.* **16**, 9165–9170 (2018).
94. Persson, B., Zigler, J. S. & Jörnvall, H. A Super-Family of Medium-Chain Dehydrogenases/Reductases (MDR): Sub-Lines including ζ -Crystallin, Alcohol and Polyol Dehydrogenases, Quinone Oxidoreductases, Enoyl Reductases, VAT-1 and other Proteins. *Eur. J. Biochem.* **226**, 15–22 (1994).
95. Zheng, J., Gay, D. C., Demeler, B., White, M. A. & Keatinge-Clay, A. T. Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol.* **8**, 615–621 (2012).
96. Kwan, D. H. *et al.* Prediction and Manipulation of the Stereochemistry of Enoylreduction in Modular Polyketide Synthases. *Chem. Biol.* **15**, 1231–1240 (2008).
97. Rosenthal, R. G. *et al.* The use of ene adducts to study and engineer enoyl-

- thioester reductases. *Nat. Chem. Biol.* **11**, 398–400 (2015).
98. Zhang, L. *et al.* Stereospecificity of Enoylreductase Domains from Modular Polyketide Synthases. *ACS Chem. Biol.* **13**, 871–875 (2018).
99. Kwan, D. H. & Leadlay, P. F. Mutagenesis of a modular polyketide synthase enoylreductase domain reveals insights into catalysis and stereospecificity. *ACS Chem. Biol.* **5**, 829–838 (2010).
100. Skiba, M. A. *et al.* Domain Organization and Active Site Architecture of a Polyketide Synthase C-methyltransferase. *ACS Chem. Biol.* **11**, 3319–3327 (2016).
101. Skiba, M. A. *et al.* Structural Basis of Polyketide Synthase O-Methylation. *ACS Chem. Biol.* **13**, 3221–3228 (2018).
102. Horsman, M. E., Hari, T. P. A. & Boddy, C. N. Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: Logic gate or a victim of fate? *Nat. Prod. Rep.* **33**, 183–202 (2016).
103. Tsai, S. C., Lu, H., Cane, D. E., Khosla, C. & Stroud, R. M. Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry* **41**, 12598–12606 (2002).
104. Kotowska, M. & Pawlik, K. Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Appl. Microbiol. Biotechnol.* **98**, 7735–46 (2014).
105. Harvey, B. M. *et al.* Insights into Polyether Biosynthesis from Analysis of the Nigericin Biosynthetic Gene Cluster in *Streptomyces* sp. DSM4137. *Chem. Biol.* **14**, 703–714 (2007).
106. Liu, R. *et al.* Genomics-driven discovery of the biosynthetic gene cluster of maduramicin and its overproduction in *Actinomadura* sp. J1-007. *J. Ind. Microbiol. Biotechnol.* **47**, 275–285 (2020).
107. Wang, W. G. *et al.* Molecular Basis for the Biosynthesis of an Unusual

- Chain-Fused Polyketide, Gregatin A. *J. Am. Chem. Soc.* **142**, 8464–8472 (2020).
108. Mullowney, M., McClure, R. A., Robey, M. T., Kelleher, N. L. & Thomson, R. J. Natural products from thioester reductase containing biosynthetic pathways. *Nat. Prod. Rep.* **35**, 847–878 (2018).
109. Mo, S. *et al.* Elucidation of the *Streptomyces coelicolor* Pathway to 2-Undecylpyrrole, a Key Intermediate in Undecylprodiginine and Streptorubin B Biosynthesis. *Chem. Biol.* **15**, 137–148 (2008).
110. Hu, D. X., Withall, D. M., Challis, G. L. & Thomson, R. J. Structure, Chemical Synthesis, and Biosynthesis of Prodiginine Natural Products. *Chem. Rev.* **116**, 7818–53 (2016).
111. Masschelein, J. *et al.* A dual transacylation mechanism for polyketide synthase chain release in enacyloxin antibiotic biosynthesis. *Nat. Chem.* **11**, 906–912 (2019).
112. Sun, Y. *et al.* In vitro reconstruction of tetronate RK-682 biosynthesis. *Nat. Chem. Biol.* **6**, 99–101 (2010).
113. Olano, C., Méndez, C. & Salas, J. A. Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Nat. Prod. Rep.* **27**, 571-616. (2010).
114. Musiol, E. M. & Weber, T. Discrete acyltransferases involved in polyketide biosynthesis. *Medchemcomm* **3**, 871–886 (2012).
115. 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.* **100**, 3149–3154 (2003).
116. 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.* **46**, 8195–8197 (2007).
117. 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.* **11**, 33–45 (2004).

118. Musiol, E. M. *et al.* Supramolecular templating in kirromycin biosynthesis: The acyltransferase KirCII loads ethylmalonyl-CoA extender onto a specific ACP of the trans-AT PKS. *Chem. Biol.* **18**, 438–444 (2011).

119. Zhao, C. *et al.* Oxazolomycin biosynthesis in *Streptomyces albus* JA3453 featuring an 'acyltransferase-less' type I polyketide synthase that incorporates two distinct extender units. *J. Biol. Chem.* **285**, 20097–20108 (2010).

120. Menche, D. *et al.* Stereochemical determination and complex biosynthetic assembly of etnangien, a highly potent RNA polymerase inhibitor from the myxobacterium *Sorangium cellulosum*. *J. Am. Chem. Soc.* **130**, 14234–43 (2008).

121. Jenner, M. *et al.* Acyl hydrolases from: Trans -AT polyketide synthases target acetyl units on acyl carrier proteins. *Chem. Commun.* **52**, 5262–5 (2016).

122. 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.* **130**, 11614–11616 (2008).

123. Bukhari, H. S. T., Jakob, R. P. & Maier, T. Evolutionary origins of the multienzyme architecture of giant fungal fatty acid synthase. *Structure* **22**, 1775–1785 (2014).

124. Nguyen, T. A. *et al.* Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* **26**, 225–233 (2008).

125. Hochmuth, T. & Piel, J. Polyketide synthases of bacterial symbionts in sponges - Evolution-based applications in natural products research. *Phytochemistry* **70**, 1841–1849 (2009).

126. Maier, T., Jenni, S. & Ban, N. Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science* **311**, 1258–1262 (2006).

127. Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510**, 512–517 (2014).
128. Whicher, J. R. *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560–4 (2014).
129. Bagde, S. R., Mathews, I. I., Fromme, J. C. & Kim, C. Y. Modular polyketide synthase contains two reaction chambers that operate asynchronously. *Science* (80-.). **374**, 723–729 (2021).
130. P., C. D. *et al.* Mapping the catalytic conformations of an assembly-line polyketide synthase module. *Science* (80-.). **374**, 729–734 (2021).
131. Klaus, M. & Grininger, M. Correction: Engineering strategies for rational polyketide synthase design. *Nat. Prod. Rep.* **38**, 1409 (2021).
132. Kunakom, S. & Eustáquio, A. S. *Burkholderia* as a Source of Natural Products. *J. Nat. Prod.* **82**, 2018–2037 (2019).
133. Blin, K. *et al.* AntiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **49**, W29–W35 (2021).
134. Song, L. *et al.* Discovery and Biosynthesis of Gladiolin: A *Burkholderia gladioli* Antibiotic with Promising Activity against *Mycobacterium tuberculosis*. *J. Am. Chem. Soc.* **139**, 7974–7981 (2017).
135. Jenner, M. *et al.* An unusual *Burkholderia gladioli* double chain-initiating nonribosomal peptide synthetase assembles 'fungal' icosalide antibiotics. *Chem. Sci.* **10**, 5489–5494 (2019).
136. Ruiz, B. *et al.* Production of microbial secondary metabolites: Regulation by the carbon source. *Crit. Rev. Microbiol.* **36**, 146–167 (2010).
137. Aziz, F. A. A. *et al.* Interspecies interactions are an integral determinant of microbial community dynamics. *Front. Microbiol.* **6**, 1148 (2015).
138. Dashti, Y. *et al.* Discovery and Biosynthesis of Bolagladins: Unusual Lipodepsipeptides from *Burkholderia gladioli* Clinical Isolates**. *Angew. Chemie*

132, 21737–21745 (2020).

139. Song, L. *et al.* Discovery and Biosynthesis of Gladiolin: A *Burkholderia gladioli* Antibiotic with Promising Activity against *Mycobacterium tuberculosis*. *J. Am. Chem. Soc.* **139**, 7974–7981 (2017).

140. Yin, M. *et al.* Cycloheximide and actiphenol production in streptomyces sp. YIM56141 governed by single biosynthetic machinery featuring an acyltransferase-less type I polyketide synthase. *Org. Lett.* **16**, 3072–3075 (2014).

141. Kominek, L. A. Cycloheximide production by *Streptomyces griseus*: control mechanisms of cycloheximide biosynthesis. *Antimicrob Agents Chemother.* **7**, 856–860 (1975).

142. Sisler, H. D. & Siegel, M. R. Cycloheximide and Other Glutarimide Antibiotics. in *Mechanism of Action* (1967).

143. Allen, M. S., Becker, A. M. & Rickards, R. W. The glutarimide antibiotic 9-methylstreptimidone: Structure, biogenesis and biological activity. *Aust. J. Chem.* **29**, 673–679 (1976).

144. Lee, B. *et al.* Isolation of new streptimidone derivatives, glutarimide antibiotics from *Streptomyces* sp. W3002 using LC-MS-guided screening. *J. Antibiot. (Tokyo)*. **73**, 184–188 (2020).

145. Frohardt, R. P. *et al.* Chemistry of Streptimidone, A New Antibiotic. *J. Am. Chem. Soc.* **81**, 5500–5506 (1959).

146. Takemoto, Y. *et al.* Migrastatin, a novel 14-membered ring macrolide, inhibits anchorage-independent growth of human small cell lung carcinoma Ms-1 cells. *J. Antibiot. (Tokyo)*. **54**, 1104–1107 (2001).

147. Nakae, K. *et al.* Migrastatin, a new inhibitor of tumor cell migration from *Streptomyces* sp. MK929-43F1 taxonomy, fermentation, isolation and biological activities. *J. Antibiot. (Tokyo)*. **53**, 1130–1136 (2000).

148. Nakamura N, Takahashi Y, Naganawa H, Nakae K, Imoto M, Shiro M,

- Matsumura K, W. H. & T, K. Absolute configuration of migrastatin, a novel 14-membered ring macrolide. *J. Antibiot. (Tokyo)*. **55**, 442–444 (2002).
149. Hochlowski, J. E., Whittern, D. N., Hill, P. & McAlpine, J. B. Dorrigocins: Novel antifungal antibiotics that change the morphology of ras-transformed NIH/3T3 cells to that of normal cells. II. Isolation and elucidation of structures. *J. Antibiot. (Tokyo)*. **47**, 870–874 (1994).
150. Woo, E. J. *et al.* Migrastatin and a new compound, isomigrastatin from *Streptomyces platensis*. *J. Antibiot. (Tokyo)*. **55**, 141–6 (2002).
151. Sugawara, K. *et al.* Lactimidomycin† a new glutarimide group antibiotic production, isolation, structure and biological activity. *J. Antibiot. (Tokyo)*. **45**, 1433–1441 (1992).
152. Takayasu, Y., Tsuchiya, K., Aoyama, T. & Sukenaga, Y. NK30424A and B, novel inhibitors of lipopolysaccharide-induced tumour necrosis factor alpha production, produced by *Streptomyces* sp. NA30424. *J. Antibiot. (Tokyo)*. **54**, 1111–5 (2001).
153. Zhang, D., Yi, W., Ge, H., Zhang, Z. & Wu, B. Bioactive Streptoglutarimides A-J from the Marine-Derived *Streptomyces* sp. ZZ741. *J. Nat. Prod.* **82**, 2800–2808 (2019).
154. Wang, B. *et al.* Biosynthesis of 9-methylstreptimidone involves a new decarboxylative step for polyketide terminal diene formation. *Org. Lett.* **15**, 1278–1281 (2013).
155. Ju, J., Lim, S. K., Jiang, H. & Shen, B. Migrastatin and dorrigocins are shunt metabolites of iso-migrastatin. *J. Am. Chem. Soc.* **127**, 1622–1623 (2005).
156. Lim, S. K. *et al.* Iso-migrastatin, migrastatin, and dorrigocin production in *streptomyces platensis* NRRL 18993 is governed by a single biosynthetic machinery featuring an acyltransferase-less type I polyketide synthase. *J. Biol. Chem.* **284**, 29746–29756 (2009).

157. Seo, J. W. *et al.* Comparative characterization of the lactimidomycin and iso-migrastatin biosynthetic machineries revealing unusual features for acyltransferase-less type I polyketide synthases and providing an opportunity to engineer new analogues. *Biochemistry* **53**, 7854–7865 (2014).
158. Bretschneider, T. *et al.* Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature* **502**, 124–128 (2013).
159. Hong, J. & White, J. D. The chemistry and biology of rhizoxins, novel antitumor macrolides from *Rhizopus chinensis*. *Tetrahedron* **60**, 5653–5681 (2004).
160. Sundaram, S., Heine, D. & Hertweck, C. Polyketide synthase chimeras reveal key role of ketosynthase domain in chain branching. *Nat. Chem. Biol.* **11**, 949–51 (2015).
161. Kato, J. Y., Funa, N., Watanabe, H., Ohnishi, Y. & Horinouchi, S. Biosynthesis of γ -butyrolactone autoregulators that switch on secondary metabolism and morphological development in *Streptomyces*. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2378–83 (2007).
162. Schneider-Poetsch, T. *et al.* Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat. Chem. Biol.* **6**, 209–217 (2010).
163. Garreau De Loubresse, N. *et al.* Structural basis for the inhibition of the eukaryotic ribosome. *Nature* **513**, 517–522 (2014).
164. Gaul, C. *et al.* The migrastatin family: Discovery of potent cell migration inhibitors by chemical synthesis. *J. Am. Chem. Soc.* **126**, 11326–11337 (2004).
165. Carocci, M. & Yang, P. L. Lactimidomycin is a broad-spectrum inhibitor of dengue and other RNA viruses. *Antiviral Res.* **5**, 62 (2016).
166. Saito, N., Suzuki, F., Sasaki, K. & Ishida, N. Antiviral and interferon inducing activity of a new glutarimide antibiotic, 9 methylstreptimidone. *Antimicrob. Agents Chemother.* **10**, 14–19 (1976).
167. Suzuki, F., Saito, N. & Ishida, N. Effect of an interferon inducer, 9-

methylestreptimidone, on influenza virus infection in mice. *Ann. N. Y. Acad. Sci.* **284**, 667–675 (1977).

168. Garreau De Loubresse, N. *et al.* Structural basis for the inhibition of the eukaryotic ribosome. *Nature* **513**, 517–522 (2014).

169. Yin, M. *et al.* Cycloheximide and actiphenol production in streptomyces sp. YIM56141 governed by single biosynthetic machinery featuring an acyltransferase-less type I polyketide synthase. *Org. Lett.* **16**, 3072–3075 (2014).

170. Arakawa, K., Tsuda, N., Taniguchi, A. & Kinashi, H. The Butenolide Signaling Molecules SRB1 and SRB2 Induce Lankacidin and Lankamycin Production in *Streptomyces rochei*. *ChemBioChem* **13**, 1447–57 (2012).

171. Wang, W. *et al.* Identification of a butenolide signaling system that regulates nikkomycin biosynthesis in *Streptomyces*. *J. Biol. Chem.* **293**, 20029–20040. (2018).

172. Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327 (2018).

173. Clinical and Laboratory Standards Institute (CLSI). Reference method for broth dilution. *M27-A3 Ref. Method Broth Dilution Antifung. Susceptibility Test. Yeasts; Approv. Stand. - Third Ed.* (2008).

174. Goldstein, A. L. & McCusker, J. H. Development of *Saccharomyces cerevisiae* as a model pathogen: A system for the genetic identification of gene products required for survival in the mammalian host environment. *Genetics* **159**, 499–513 (2001).

175. Vignjevic, D. & Montagnac, G. Reorganisation of the dendritic actin network during cancer cell migration and invasion. *Semin. Cancer Biol.* **18**, 12–22 (2008).

176. Yamaguchi, H. & Condeelis, J. Regulation of the actin cytoskeleton in cancer cell migration and invasion. *Biochim. Biophys. Acta - Mol. Cell Res.* **1773**, 642–652 (2007).

177. Micoine, K. *et al.* Total syntheses and biological reassessment of lactimidomycin, isomigrastatin and congener glutarimide antibiotics. *Chem. - A Eur. J.* **19**, 7370–7383 (2013).
178. Flannagan, R. S., Linn, T. & Valvano, M. A. A system for the construction of targeted unmarked gene deletions in the genus *Burkholderia*. *Environ. Microbiol.* **10**, 1652–60 (2008).
179. Tang, Y., Kim, C. Y., Mathews, I. I., Cane, D. E. & Khosla, C. The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11124–11129 (2006).
180. Crosby, J. & Crump, M. P. The structural role of the carrier protein - Active controller or passive carrier. *Nat. Prod. Rep.* **29**, 1111–1137 (2012).
181. Reid, R. *et al.* A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry* **42**, 72–79 (2003).
182. Piasecki, S. K., Zheng, J., Axelrod, A. J., Detelich, M. E. & Keatinge-Clay, A. T. Structural and functional studies of a trans-acyltransferase polyketide assembly line enzyme that catalyzes stereoselective α - and β -ketoreduction. *Proteins Struct. Funct. Bioinforma.* **82**, 2067–2077 (2014).
183. Helfrich, E. J. N. *et al.* Automated structure prediction of trans-acyltransferase polyketide synthase products. *Nat. Chem. Biol.* **15**, 813–821 (2019).
184. Uytterhoeven, B., Lathouwers, T., Voet, M., Michiels, C. W. & Lavigne, R. A protein interaction map of the kalimantacin biosynthesis assembly line. *Front. Microbiol.* **7**, 1726 (2016).
185. Piel, J. *et al.* Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16222–7 (2004).
186. El-Sayed, A. K. *et al.* Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem. Biol.* **10**, 419–30

(2003).

187. Mattheus, W. *et al.* Isolation and Purification of a New Kalimantacin/Batumin-Related Polyketide Antibiotic and Elucidation of Its Biosynthesis Gene Cluster. *Chem. Biol.* **17**, 149–59 (2010).

188. Geoghegan, K. F. *et al.* Spontaneous α -N-6-phosphogluconoylation of a 'His tag' in *Escherichia coli*: The cause of extra mass of 258 or 178 Da in fusion proteins. *Anal. Biochem.* **267**, 169–84 (1999).

189. Owen, J. G., Robins, K. J., Parachin, N. S. & Ackerley, D. F. A functional screen for recovery of 4'-phosphopantetheinyl transferase and associated natural product biosynthesis genes from metagenome libraries. *Environ. Microbiol.* **14**, 1198–209 (2012).

190. Izoré, T. *et al.* Structures of a non-ribosomal peptide synthetase condensation domain suggest the basis of substrate selectivity. *Nat. Commun.* **12**, 2511 (2021).

191. Worthington, A. S., Rivera, H., Torpey, J. W., Alexander, M. D. & Burkart, M. D. Mechanism-based protein cross-linking probes to investigate carrier protein-mediated biosynthesis. *ACS Chem. Biol.* **1**, 687–691 (2006).

192. Quadri, L. E. N. *et al.* Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585–95 (1998).

193. Kapur, S. *et al.* Mechanism based protein crosslinking of domains from the 6-deoxyerythronolide B synthase. *Bioorganic Med. Chem. Lett.* **18**, 3034–8 (2008).

194. Dorrestein, P. C. *et al.* Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry. *Biochemistry* **45**, 12756–66 (2006).

195. Lefebvre, M. D. & Valvano, M. A. Construction and evaluation of plasmid vectors optimized for constitutive and regulated gene expression in *Burkholderia*

- cepacia* complex isolates. *Appl. Environ. Microbiol.* **68**, 5956–64 (2002).
196. Vander Wood, D. A. & Keatinge-Clay, A. T. The modules of *trans* -acyltransferase assembly lines redefined with a central acyl carrier protein. *Proteins Struct. Funct. Bioinforma.* **86**, 664–675 (2018).
197. Vander Wood, D. A. & Keatinge-Clay, A. T. The modules of *trans*-acyltransferase assembly lines redefined with a central acyl carrier protein. *Proteins Struct. Funct. Bioinforma.* **86**, 664–675 (2018).
198. Ma, M. *et al.* Post-polyketide synthase steps in iso-migrastatin biosynthesis, featuring tailoring enzymes with broad substrate specificity. *J. Am. Chem. Soc.* **135**, 2489–2492 (2013).
199. Willey, J. M. & Gaskell, A. A. Morphogenetic signaling molecules of the streptomycetes. *Chem. Rev.* **111**, 174–187 (2011).
200. Li, Y. & Tan, H. Biosynthesis and molecular regulation of secondary metabolites in microorganisms. *Sci. China Life Sci.* **60**, 935–938 (2017).
201. Khokhlov, A. S. *et al.* The A-factor, responsible for streptomycin biosynthesis by mutant strains of *Actinomyces streptomycini*. *Dokl. Akad. Nauk SSSR* **177**, 232–235 (1967).
202. Hara, O. & Beppu, T. Mutants blocked in streptomycin production in *Streptomyces griseus* - the role of A-factor. *J. Antibiot. (Tokyo)*. **35**, 349–358 (1982).
203. Sakuda, S. & Yamada, Y. Stereochemistry of butyrolactone autoregulators from streptomyces. *Tetrahedron Lett.* **32**, 1817–1820 (1991).
204. Ruengjitchatchawalya, M., Nihira, T. & Yamada, Y. Purification and characterization of the IM-2-binding protein from *Streptomyces* sp. strain FRI-5. *J. Bacteriol.* **177**, 551–557 (1995).
205. Takano, E., Chakraborty, R., Nihira, T., Yamada, Y. & Bibb, M. J. A complex role for the gamma-butyrolactone SCB1 in regulating antibiotic production in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **41**, 1015–1028 (2001).

206. Sidda, J. D. *et al.* Overproduction and identification of butyrolactones SCB1-8 in the antibiotic production superhost: *Streptomyces* M1152. *Org. Biomol. Chem.* **14**, 6390–6393 (2016).
207. Lee, Y. J., Kitani, S. & Nihira, T. Null mutation analysis of an afsA-family gene, barX, that is involved in biosynthesis of the γ -butyrolactone autoregulator in *Streptomyces virginiae*. *Microbiology* **156**, 206–210 (2010).
208. Hsiao, N. H. *et al.* Analysis of Two Additional Signaling Molecules in *Streptomyces coelicolor* and the Development of a Butyrolactone-Specific Reporter System. *Chem. Biol.* **16**, 951–960 (2009).
209. Van Der Heul, H. U., Bilyk, B. L., McDowall, K. J., Seipke, R. F. & Van Wezel, G. P. Regulation of antibiotic production in Actinobacteria: New perspectives from the post-genomic era. *Nat. Prod. Rep.* **35**, 575–604 (2018).
210. Corre, C., Song, L., O'Rourke, S., Chater, K. F. & Challis, G. L. 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by *Streptomyces coelicolor* genome mining. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 17510–17515 (2008).
211. Arakawa, K., Tsuda, N., Taniguchi, A. & Kinashi, H. The Butenolide Signaling Molecules SRB1 and SRB2 Induce Lankacidin and Lankamycin Production in *Streptomyces rochei*. *ChemBioChem* **13**, 1447–57 (2012).
212. Zhou, S. *et al.* Molecular basis for control of antibiotic production by a bacterial hormone. *Nature* **590**, 463–467 (2021).
213. Hsiao, N. H. *et al.* ScbA from *Streptomyces coelicolor* A3(2) has homology to fatty acid synthases and is able to synthesize γ -butyrolactones. *Microbiology* **153**, 1394–1404 (2007).
214. Zhou, S., Malet, N. R., Song, L., Corre, C. & Challis, G. L. MmfL catalyses formation of a phosphorylated butenolide intermediate in methylenomycin furan biosynthesis. *Chem. Commun.* **56**, 14443–14446 (2020).

215. Pandya, C., Dunaway-Mariano, D., Xia, Y. & Allen, K. N. Structure-guided approach for detecting large domain inserts in protein sequences as illustrated using the haloacid dehalogenase superfamily. *Proteins Struct. Funct. Bioinforma.* **82**, 1896–906 (2014).
216. Corre, C., Haynes, S. W., Malet, N., Song, L. & Challis, G. L. A butenolide intermediate in methylenomycin furan biosynthesis is implied by incorporation of stereospecifically ¹³C-labelled glycerols. *Chem. Commun.* **46**, 4079–4081 (2010).
217. Little, R. F. & Hertweck, C. Chain release mechanisms in polyketide and non-ribosomal peptide biosynthesis. *Nat. Prod. Rep.* **11**, 906–912 (2021).
218. Nakou, I. T. *et al.* Genomics-Driven Discovery of a Novel Glutarimide Antibiotic from *Burkholderia gladioli* Reveals an Unusual Polyketide Synthase Chain Release Mechanism. *Angew. Chemie - Int. Ed.* **59**, 23145–23153 (2020).
219. Niehs, S. P. *et al.* Insect-Associated Bacteria Assemble the Antifungal Butenolide Gladiofungin by Non-Canonical Polyketide Chain Termination. *Angew. Chemie* **59**, 23122–23126 (2020).
220. Nakou, I. T. *et al.* Genomics-Driven Discovery of a Novel Glutarimide Antibiotic from *Burkholderia gladioli* Reveals an Unusual Polyketide Synthase Chain Release Mechanism. *Angew. Chemie - Int. Ed.* **59**, 23145–23153 (2020).
221. Koonin, E. V. & Tatusove, R. L. Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity. Application of an iterative approach to database search. *J. Mol. Biol.* **244**, 125–132 (1994).
222. Aravind, L., Galperin, M. Y. & Koonin, E. V. The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. *Trends Biochem. Sci.* **23**, 127–129 (1998).
223. Morais, M. C. *et al.* The crystal structure of *Bacillus cereus* phosphonoacetaldehyde hydrolase: Insight into catalysis of phosphorus bond cleavage and catalytic diversification within the had enzyme superfamily.

Biochemistry **39**, 10385–10396 (2000).

224. Baker, A. S. *et al.* Insights into the mechanism of catalysis by the P-C bond-cleaving enzyme phosphonoacetaldehyde hydrolase derived from gene sequence analysis and mutagenesis. *Biochemistry* **37**, 9305–9315 (1998).

225. Qian, N., Stanley, G. A., Hahn-Hagerdal, B. & Radstrom, P. Purification and characterization of two phosphoglucomutases from *Lactococcus lactis* subsp. *lactis* and their regulation in maltose- and glucose-utilizing cells. *J. Bacteriol.* **176**, 5304–5311 (1994).

226. Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D. & Aravind, L. Evolutionary Genomics of the HAD Superfamily: Understanding the Structural Adaptations and Catalytic Diversity in a Superfamily of Phosphoesterases and Allied Enzymes. *J. Mol. Biol.* **361**, 1003–34 (2006).

227. Kuznetsova, E. *et al.* Genome-wide analysis of substrate specificities of the *Escherichia coli* haloacid dehalogenase-like phosphatase family. *J. Biol. Chem.* **281**, 36149–36161 (2006).

228. Rigden, D. J. The histidine phosphatase superfamily: Structure and function. *Biochem. J.* **409**, 333–48 (2008).

229. Kretschmer, M., Langer, C. & Prinz, W. Mutation of Monofunctional 6-Phosphofructo-2-kinase in Yeast to Bifunctional 6-Phosphofructo-2-kinase/Fructose 2,6-Bisphosphatase. *Biochemistry* **32**, 11143–11148 (1993).

230. Hamda, K. *et al.* Crystal structure of the protein histidine phosphatase SixA in the multistep His-Asp phosphorelay. *Genes to Cells* **10**, 1–11 (2005).

231. Fahs, S., Lujan, P. & Köhn, M. Approaches to study phosphatases. *ACS Chem. Biol.* **11**, 2944–2961 (2016).

232. Ridder, I. S. & Dijkstra, B. W. Identification of the Mg²⁺-binding site in the P-type ATPase and phosphatase members of the HAD (haloacid dehalogenase) superfamily by structural similarity to the response regulator protein CheY. *Biochem.*

J. **339**, 223–226 (1999).

233. Nivina, A., Yuet, K. P., Hsu, J. & Khosla, C. Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem. Rev.* **119**, 12524–12547 (2019).

234. Weissman, K. J. & Leadlay, P. F. Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Microbiol.* **3**, 925–36 (2005).

235. Menzella, H. G. *et al.* Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171–1176 (2005).

236. Menzella, H. G., Carney, J. R. & Santi, D. V. Rational Design and Assembly of Synthetic Trimodular Polyketide Synthases. *Chem. Biol.* **14**, 143–51 (2007).

237. Poptsova, M. Testing phylogenetic methods to identify horizontal gene transfer. *Methods Mol. Biol.* **532**, 227–240 (2009).

238. Ridley, C. P., Ho, Y. L. & Khosla, C. Evolution of polyketide synthases in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 4595–600 (2008).

239. Ginolhac, A. *et al.* Type I polyketide synthases may have evolved through horizontal gene transfer. *J. Mol. Evol.* **60**, 716–25 (2005).

240. Sugimoto, Y., Ding, L., Ishida, K. & Hertweck, C. Rational design of modular polyketide synthases: Morphing the aureothin pathway into a luteoreticuliculin assembly line. *Angew. Chemie - Int. Ed.* **53**, 1560–1564 (2014).

241. Peng, H., Ishida, K., Sugimoto, Y., Jenke-Kodama, H. & Hertweck, C. Emulating evolutionary processes to morph aureothin-type modular polyketide synthases and associated oxygenases. *Nat. Commun.* **10**, 3918 (2019).

242. Shen, P. & Huang, H. V. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* **112**, 441–457 (1986).

243. Starcevic, A. *et al.* A novel docking domain interface model predicting recombination between homoeologous modular biosynthetic gene clusters. *J. Ind. Microbiol. Biotechnol.* **38**, 1295–1304 (2011).

244. Chemler, J. A. *et al.* Evolution of Efficient Modular Polyketide Synthases by Homologous Recombination. *J. Am. Chem. Soc.* **137**, 10603–10609 (2015).
245. Wlodek, A. *et al.* Diversity oriented biosynthesis via accelerated evolution of modular gene clusters. *Nat. Commun.* **8**, 1206 (2017).
246. Nivina, A., Paredes, S. H., Fraser, H. B. & Khosla, C. GRINS: Genetic elements that recode assembly-line polyketide synthases and accelerate their diversification. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2100751118 (2021).
247. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–58 (2015).
248. Calderone, C. T., Bumpus, S. B., Kelleher, N. L., Walsh, C. T. & Magarvey, N. A. A ketoreductase domain in the PksJ protein of the bacillaene assembly line carries out both α - and β -ketone reduction during chain growth. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 12809–12814 (2008).
249. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–8 (2007).
250. Zheng, J., Fage, C. D., Demeler, B., Hoffman, D. W. & Keatinge-Clay, A. T. The missing linker: A dimerization motif located within polyketide synthase modules. *ACS Chem. Biol.* **8**, 1263–1270 (2013).
251. 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* **18**, 913–922 (2010).
252. Bonnett, S. A. *et al.* Structural and stereochemical analysis of a modular polyketide synthase ketoreductase domain required for the generation of a cis-alkene. *Chem. Biol.* **20**, 772–783 (2013).
253. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting,

position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).

254. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323 (2003).

255. Weissman, K. J. Genetic engineering of modular PKSs: From combinatorial biosynthesis to synthetic biology. *Nat. Prod. Rep.* **33**, 203–230 (2016).

256. Jenner, M. *et al.* An unusual *Burkholderia gladioli* double chain-initiating nonribosomal peptide synthetase assembles 'fungal' icosalide antibiotics. *Chem. Sci.* **10**, 5489–5494 (2019).

257. Dashti, Y. *et al.* Discovery and Biosynthesis of Bolagladins: Unusual Lipodepsipeptides from *Burkholderia gladioli* Clinical Isolates**. *Angew. Chemie* **132**, 21737–21745 (2020).

258. Lu, W., Roongsawang, N. & Mahmud, T. Biosynthetic Studies and Genetic Engineering of Pactamycin Analogs with Improved Selectivity toward Malarial Parasites. *Chem. Biol.* **18**, 425–431 (2011).

259. Bidou, L., Allamand, V., Rousset, J. P. & Namy, O. Sense from nonsense: Therapies for premature stop codon diseases. *Trends Mol. Med.* **18**, 679–88 (2012).

260. Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9259–9264 (2014).

261. O'Sullivan, L. A. *et al.* Identifying the genetic basis of ecologically and biotechnologically useful functions of the bacterium *Burkholderia vietnamiensis*. *Environ. Microbiol.* **9**, 1017–1034 (2007).

262. Flannagan, R. S., Aubert, D., Kooi, C., Sokol, P. A. & Valvano, M. A. *Burkholderia cenocepacia* requires a periplasmic HtrA protease for growth under

thermal and osmotic stress and for survival in vivo. *Infect. Immun.* **75**, 1679–89 (2007).

263. National Committee for Clinical Laboratory Standard. NCCLS M7-A6, Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically. in *Wayne, PA: Clinical and Laboratory Standards Institute* 60 (2003).

264. Kautsar, S. A. *et al.* MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* **48**, 454–458 (2020).

265. Kozlov, A. M. *et al.* RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).

266. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).

Appendix

Protein sequences

N-His₆-GdsH oxidoreductase

```

    10      20      30      40      50      60
MTAMNRRIVL ASRPRQAEV GNFRLEAVPL VQPGEDQVLV RQHYLSLDPY MRDRMSVSKS

    70      80      90     100     110     120
YAAPQPLDEV MIGGTVGEVL ASRNARFKPG DfvAGMGGWQ EYALLDGAEL GGWRHVDPAR

    130     140     150     160     170     180
APLSAYLGAI GMPGVTAWYG LARIIPQAG ETVVISSAAG AVGGAAGQLA RARGARVIGI

    190     200     210     220     230     240
AGGPRKCAYV VEELGFDACL DHRQYGDLKS MSAALNRACP NGIDGYFENV GGMLLDVAVLM

    250     260     270     280     290     300
RCNPFARVAL CGMIAGYNGQ AIPLTMPQLL LINRIRLEGF IIADQGDAWF DALEELGALV

    310     320     330     340
GAGKLRYHES VADGLARAPE AFITLLGGGN HGKQVVKLVA

```

N-His₆-GdsG phosphato-mutase

```

    10      20      30      40      50      60
MGSSHHHHHH SSGLVPRGSH MNSTSLQTIR LIAVDLDGPL LIDTFSPIMH KLCSEYYRID

    70      80      90     100     110     120
YTRDLERNTF SRSRAEVVDY LRRQIGEQMS ETERKQSDEE SIASYFRYRD EYMRDHPHGM

    130     140     150     160     170     180
KPEVPAFLDL LTSLGVTVIC YGGLDEDYMR RGLGEQAERF ATYICTNEFR PGVREIVRDF

    190     200     210     220     230     240
YKLAPHQALF IDDVNFVAEH AKMLGTPFIG VPSNEAWSWQ KHDMKETGVR RIVDSVSQID

    250
LALLQEIDAA AVHGGGW

```

N-His₆-GdsA histidine phosphatase

```

      10      20      30      40      50      60
MGSSHHHHHH SSGLVPRGSH MQSLYLARHG QASFGSDDYD RLTDVGTRQC TLLGERLNSL

      70      80      90     100     110     120
QPQRPVLVGG SHRRHLQSAL ACAEAWGPVL ETEYGVDAGF DEYDHTTEILL QATPQFQDLR

     130     140     150     160     170     180
ALSAHVAQQD DPPRAFQRMF ADAAARWVSG KFDGEYTQSW SAFADRCWAT LRGLPARHGA

     190     200     210     220     230     240
TQPVMVFTSG GVIGAICQRL LGIPDERIFD LTWSLQNAAGL TRLMVDDDDGH FKLGYLNCIA

     250
HLDTS DGRFL TYR

```

N-His₆-GdsF-M4-ACP di-domain

```

      10      20      30      40      50      60
MGSSHHHHHH SSGLVPRGSH MVDSAAAPAA LSRGAASDYL RELVAEVIRI PAEEIDPAEP

      70      80      90     100     110     120
LESYGIDSVL VVQLTARLRS VLPNIGGTLL FEVATIDALS ERLSTHAEP LGALLSGAQA

     130     140     150     160     170     180
AAPVAAEASG ATGTISEAVV PVVAVAAEPV VTVSAAAAAA SADTGGDAAL RAGALDYLRG

     190     200     210     220     230     240
LVAEVIRIPA TEIDPAEPL ESYGIDSVLVV QL TARLRSVL PNIGGTLLFE VTSVAGLAER

     250     260
LLVTHAPALR ALVAGAGAAP VT

```

N-His₆-GdsF-M5-ACP

```

      10      20      30      40      50      60
MGSSHHHHHH SSGLVPRGSH MPAAEIPLST DTQLHERTER YLKSLLANAT RMAAEQITPD

      70      80      90     100     110     120
VPFEEYGLDS IMVIKLTSEL EKHF GALS KT LFF EY GSLAE LAGYFVESHA GQLDAVLGSS

RR TAAAV

```

N-His₆-GdsF-M6-ACP tri-domain

```

10      20      30      40      50      60
MGSSHHHHH SSGLVPRGSH MAQPVSAEIA DSELLDRLLA RLREQVASLI KVEPRLVEFD

      70      80      90      100     110     120
AELSDYGYDS VSLTDLADRL NREFGLDLTP ALFFEFPTLQ RMAEHLLATQ RTPLAARLAP

      130     140     150     160     170     180
AVAIPAALAA PAAPTMPAAP AAPAVVAAPV AQSPARSPAQ SGAAPATRDA AISATELRDR

      190     200     210     220     230     240
VGAYLREQAA ALIKVEADVI DLDCELSEYG YDSISLTDLA DRVNRELGLE LTPALFFEYP

      250     260     270     280     290     300
SLQGLAGHLV DAHAALLTRR FAPATVAAPI HETPAPAIEV EAAVAAREPD LAVAASAVPV

      310     320     330     340     350     360
GTPGAVGAEE LRDRLCGALR ERVAGRLHVV PEQIELGTAL SEYGYDSVQL TELMNELNAA

      370     380     390
FGFALTPALF FEFPTLQALA EYLLRTEQER LTQVW

```

N-His₆-GdsF-M7-ACP

```

      10      20      30      40      50      60
MGSSHHHHH SSGLVPRGSH MVPAPQAAPG AGVDLSRVVG AALLSRVSAL IKVKPEDIDL

      70      80      90      100     110     120
DEELAGYGFD SISLTELATE LNRTYELDLM PTVFFEHTPL AALTAYLVET HRETMARHLA

PVSATAP

```

N-His₆-GdsF-M8-ACP

```

      10      20      30      40      50      60
MGSSHHHHH SSGLVPRGSH MPAVEITAAR SPAAGGADLS RMVSAALLDQ ASVLIKVKPE

      70      80      90      100     110     120
DIDLDEELAG YGFDSISLTE LATELNRTYE LDLMPTVFFE HPTLAALTAY LVDARETMA

      130
RHFAPATEAS AA

```

N-His₆-GdsB-AH-AT-ER

10	20	30	40	50	60
MGSSHHHHHH	SSGLVPRGSH	MQRKTVFLMS	GQGSQYYHMG	IDLHAAEPVF	RRSMEALDAV
70	80	90	100	110	120
VGEEAGFSL	AVLYGGNVKR	TTPFDRTLHT	HPALFMLQYS	LAQTLLARGI	EPDYLLGTSL
130	140	150	160	170	180
GEYVAAALAG	VGSPEQVLRI	LLHKAQLAEQ	HCGPGAMLAV	LDDPARYERD	PVLHRNSALA
190	200	210	220	230	240
GVNHDQHFLV	SGEREALKRV	HAHLQAQQVV	CLMLPISHGF	HSSAMDPCLP	ALLDTAPQTL
250	260	270	280	290	300
AAPRIPLISC	LLGDRLTAAP	DKDYFWQIGR	RPILFREAVA	ALLSAEPGGV	EFIDVGPAT
310	320	330	340	350	360
LSGLLKRLLP	AGHPHSIRAL	LSPFAGELDG	LHALAPRSDA	RATTVGDPKN	MYAFVFPQG
370	380	390	400	410	420
AQKKGMGAAL	FEKYPERVAS	ADRILGYSIA	SLCLEDPHGQ	LDQTLYTQPA	LYVVNALSYL
430	440	450	460	470	480
DASEQEAVPP	AFVAGHSLGE	YAALFAAGVV	DFETGLRLVQ	KRATLMSQAT	AGAMSAVLGL
490	500	510	520	530	540
DAVTLRVLLD	ANGGQEIDIA	NLNTPSQIVI	SGPRDAVLAI	KPVLEQAPGC	RGVIQLAVSG
550	560	570	580	590	600
AFHSRLMLDA	QREFERFLAG	FELAAPRIPV	ISNVTARPYG	PGEAKRLLAT	QITHSVNWVD
610	620	630	640	650	660
SVRYMWGQGV	ETFRELGHGR	VLTNLVTKIR	SEATPLVEAA	PTAHAEPASR	AVQAPARSAS
670	680	690	700	710	720
PASAPAPAAG	GRAAVPQIHA	TSLGSAAFRA	DYKLKYAYVG	GAMVHGIAV	AMVVRMARAG
730	740	750	760	770	780
MLAFFGTGGL	RPADVERAIL	EIRAALGPDA	TFGMNLLNGS	REEAMVALLL	EHGVRNVEAS
790	800	810	820	830	840
AYMQVPSPLV	LYRLRGLQAG	SDGRPVARNR	IMAKLSRPEV	AAGFLKPAPA	AIVDKLLQSG
850	860	870	880	890	900
AITAQQAAWA	RELPMADDIC	VEADSGGHTD	SGVAAALLPV	IIAQRDEAMA	AHGYAKPIRI
910	920	930	940	950	960
GSAGGIGTPE	AAASAFVLGA	DFILTGSINQ	CTVEAGTSDL	VKDMLQDVEV	QDQDYAPAGD
970	980	990	1000	1010	1020
MFEIGARVQV	LKRGVFFPAR	ANKLYDVYRS	VDSLEQIDAK	TARMIQDKYF	QRSFEAVYEE
1030	1040	1050	1060	1070	1080
CRKHYPAEII	AAADQAPKKK	MALVFKWYFA	YSNRLALTGE	SGGKVDFQIH	CGPALGGFNR
1090	1100	1110	1120	1130	
LVKGTPLESW	RNRHVDEIGV	RLMESTAALL	NTRFTQLLRQ	AATAPAPEAE	AVV

N-His₆-GdsB-excised ER

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MGSSHHHHHH	SSGLVAVPQI	HATSLGSAAF	RADYKLYAY	VGGAMVHGIA	SVAMVVRMAR
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
AGMLAFFGTG	GLRPADVERA	ILEIRAALGP	DATFGMNLN	GSREEAMVAL	LLEHGVNRVE
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
ASAYMQVSPS	LVLYRLRGLQ	AGSDGRPVAR	NRIMAKLSRP	EVAAGFLKPA	PAAIVDKLLQ
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
SGAITAQQA	WARELPMADD	ICVEADSGGH	TDSGVAAALL	PVIIAQRDEA	MAAHGYAKPI
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>
RIGSAGGIGT	PEAAASAFVL	GADFILTGSI	NQCTVEAGTS	DLVKDMLQDV	EVQDQDYAPA
<u>310</u>	<u>320</u>	<u>330</u>	<u>340</u>	<u>350</u>	<u>360</u>
GDMFEIGARV	QVLKRGVFFP	ARANKLYDVY	RSVDSLEQID	AKTARMIQDK	YFQRSFEAVY
<u>370</u>	<u>380</u>	<u>390</u>	<u>400</u>	<u>410</u>	<u>420</u>
EECRKHYPAE	IIAAADQAPK	KKMALVFKWY	FAYSNRLALT	GESGGKVDFQ	IHCGPALGGF
<u>430</u>	<u>440</u>	<u>450</u>	<u>460</u>	<u>470</u>	
NRLVKGTPLE	SWNRHVDEI	GVRLMESTAA	LLNTRFTQLL	RQAATAPAPE	AEAVV

N-His₆-P022 phosphato-mutase

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MGSSHHHHHH	SSGLVPRGSH	MTARHLRLVA	VNIDGVLLND	TFSPVIHRFV	VGRGGRYTAE
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
LERALFSQPQ	RTAAAAAAG	CGWQASPEEL	LAAYFAEREA	YLRDHPVRVL	DGAQRLLARL
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
RDLGVTVVY	GGLDRSHFDR	HLGGLAGYFD	APGYVCTDGF	RPGVKEITRD	IVGVRFDQAL
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
FIDVARVAE	TARSLGAIFI	GHPSPFAGHF	QAEAMRRAGV	RHVVRSLDAI	DERLLHALDA
<u>250</u>	<u>260</u>				
EAAAVHRAGP	GAGGTGAGFD	HPVTGG			

N-His₆-SabP phosphato-mutase

```

      10      20      30      40      50      60
MGSSHHHHHH SSGLVPRGSH MTEQLEKLRM VALNIDGVLL NDTFSPVIHH FVTSRGGTYS

      70      80      90     100     110     120
ADLERRIFSQ PQAVAGKILA EVCGMRTSPM ETIAVYFEDR ARYLADHPVR VLDGAAELLR

     130     140     150     160     170     180
RIRALGLRTV CYGGLDASHF ERYLGHLTDL FDEPRYVCTN DIRPGLYEIA VDHAGLACDE

     190     200     210     220     230     240
VLFVDDVARV AEAARELGAP FIGHPSTFEH GYQRQLMAET GARHVVGSLD EIDDALLRTI

     250
DAEARAGTLW AGAPGK

```

Protein ESI-MS data

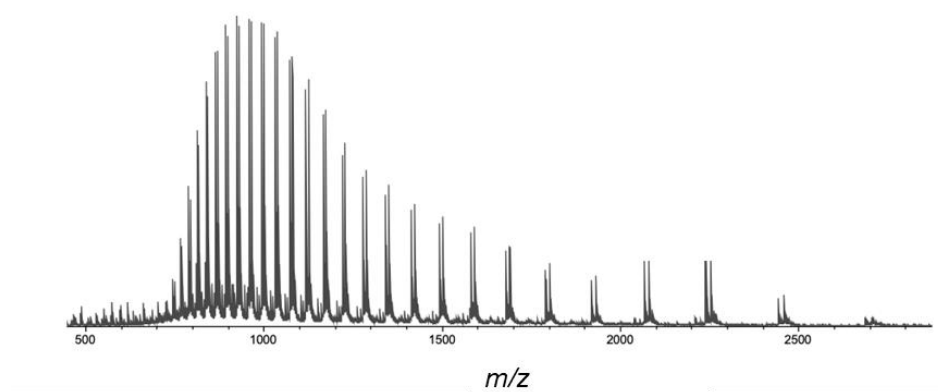


Figure A1 High resolution mass spectrum of N-His₆-GdsF-M4-ACP di-domain.

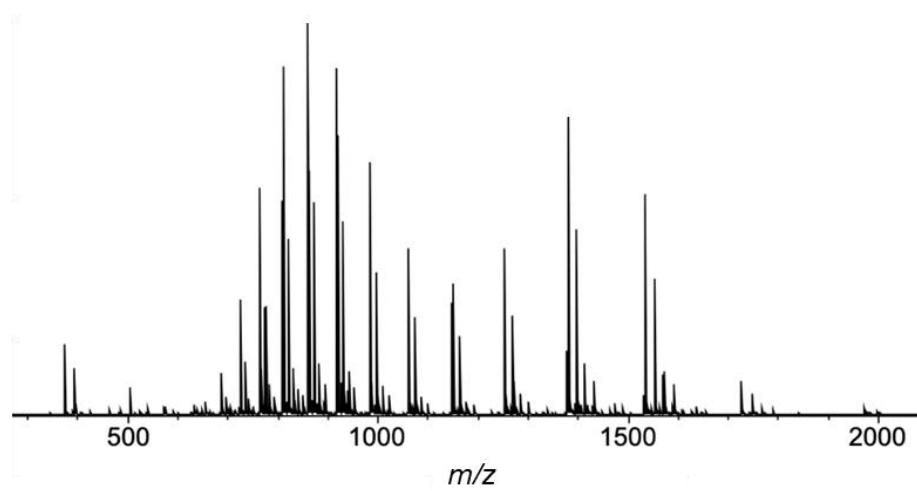


Figure A2 High resolution mass spectrum of N-His₆-GdsF-M5-ACP.

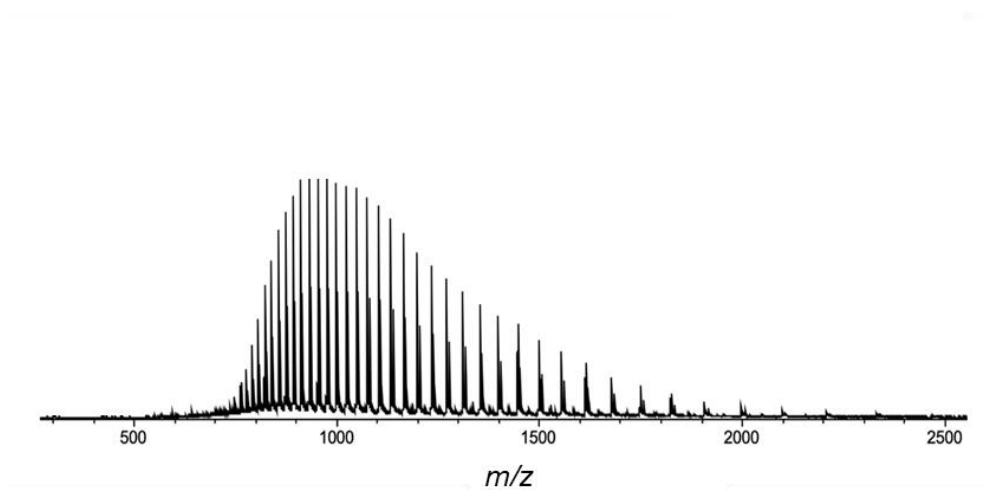


Figure A3 High resolution mass spectrum of N-His₆-GdsF-M6-ACP tri-domain.

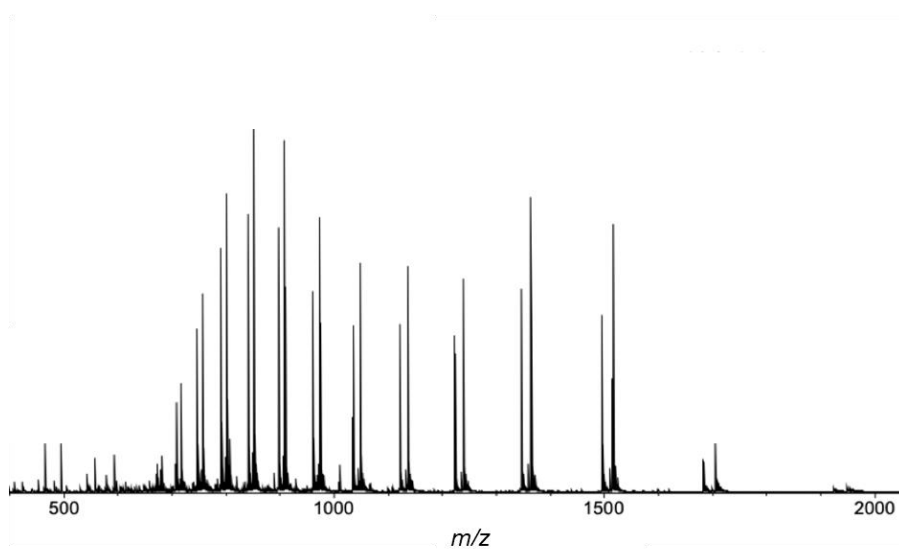


Figure A4 High resolution mass spectrum of N-His₆-GdsF-M7-ACP.

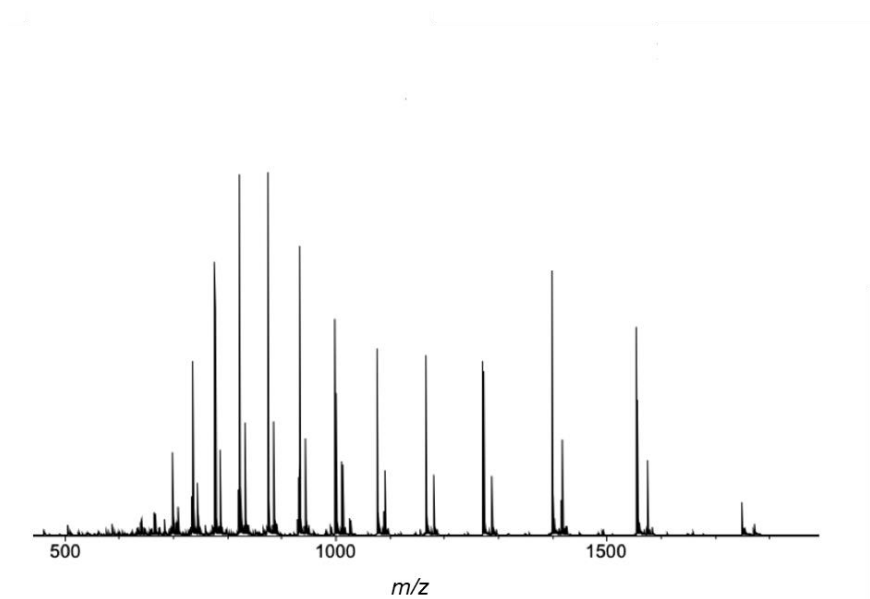


Figure A5 High resolution mass spectrum of N-His₆-GdsF-M8-ACP.

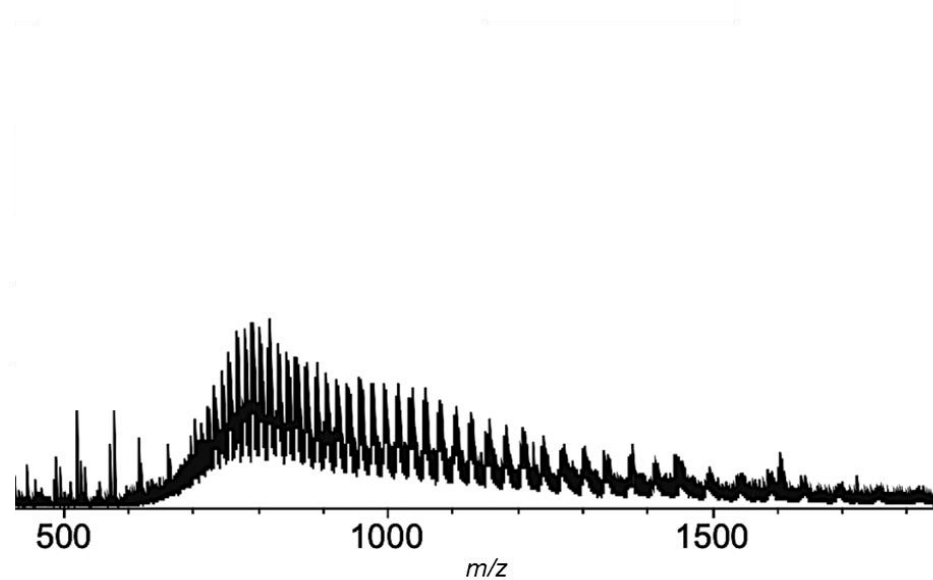


Figure A6 High resolution mass spectrum of N-His₆-GdsB-AH-AT-ER.

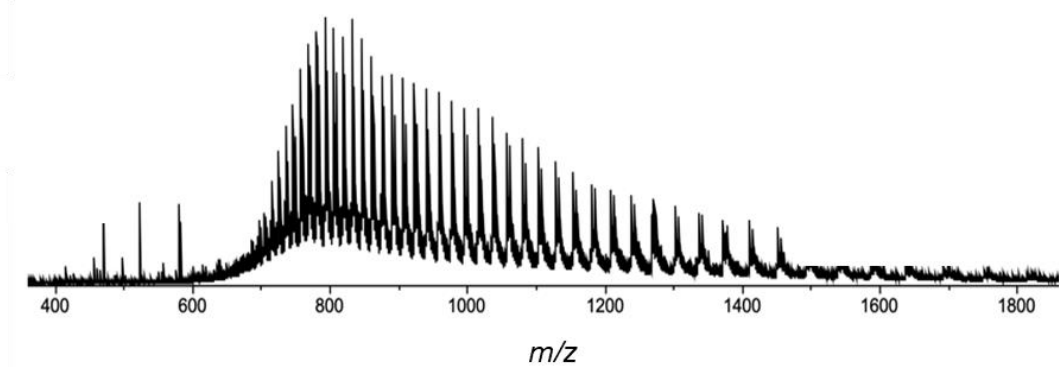


Figure A7 High resolution mass spectrum of N-His₆-GdsB excised ER.

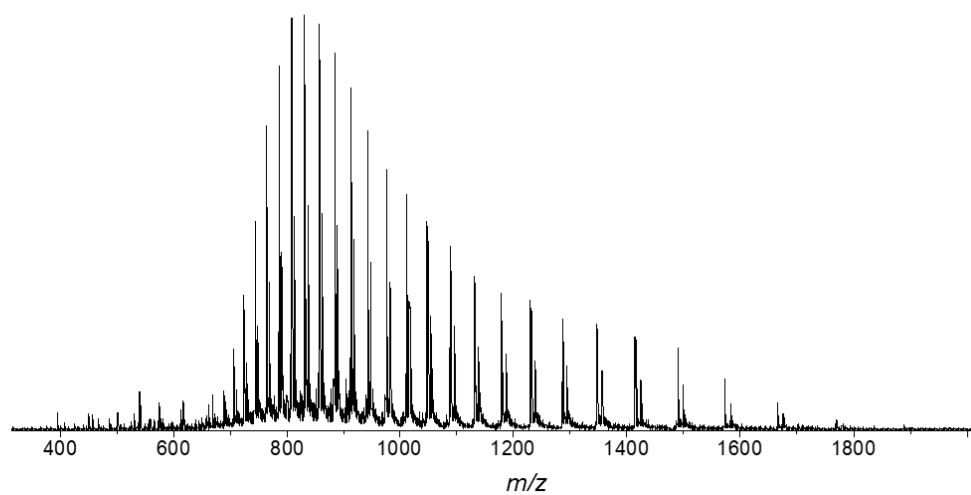


Figure A8 High resolution mass spectrum of N-His₆-SabP.

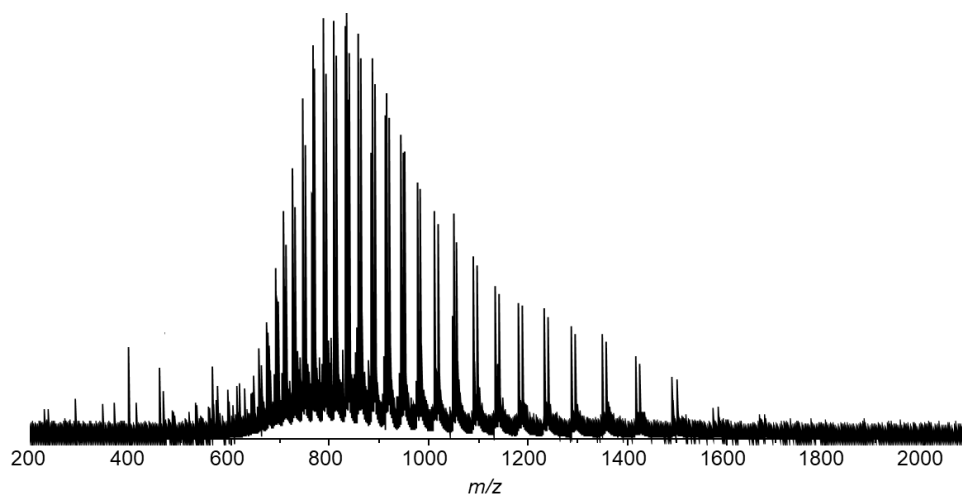


Figure A9 High resolution mass spectrum of N-His₆-GdsG.

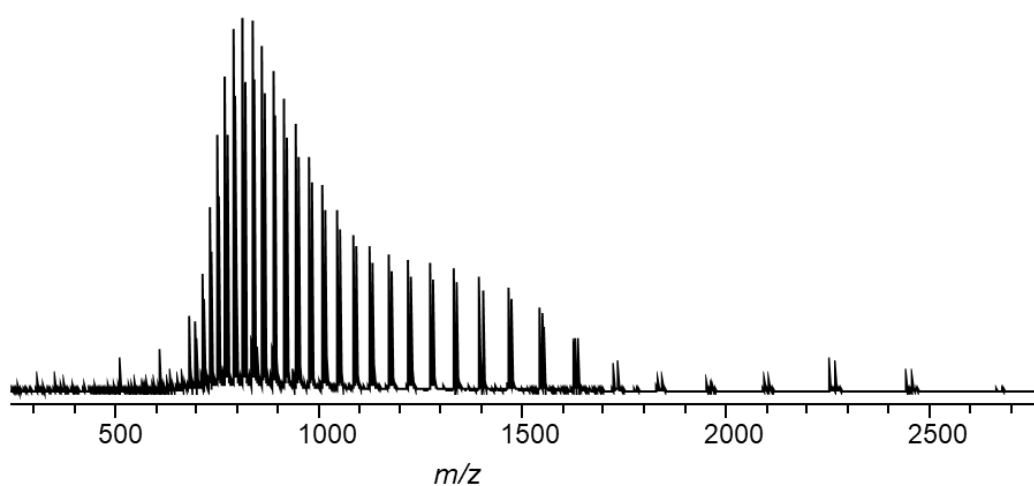


Figure A10 High resolution mass spectrum of N-His₆-GdsH.

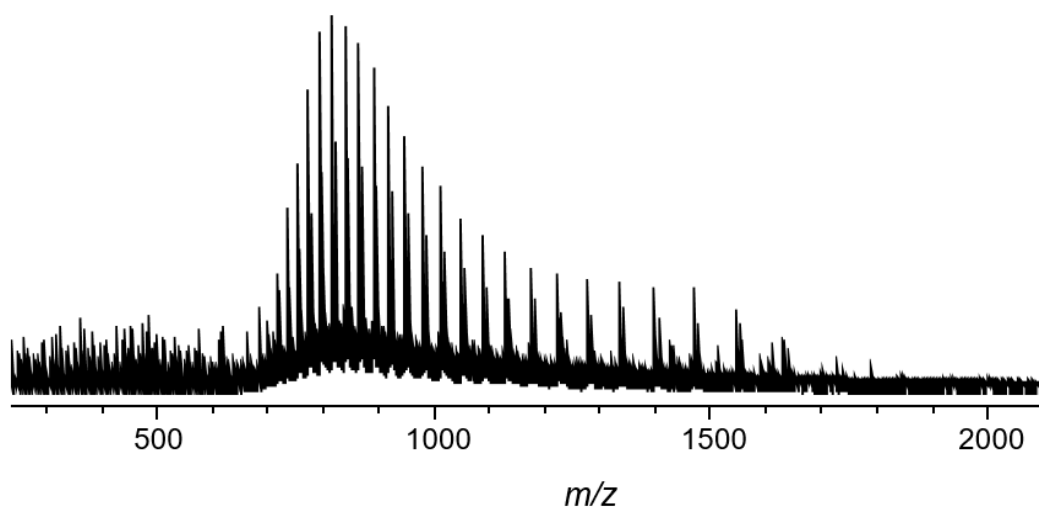


Figure A11 High resolution mass spectrum of N-His₆-P022.

Ramachandran plot for chimeric of GdsF module 7 KR

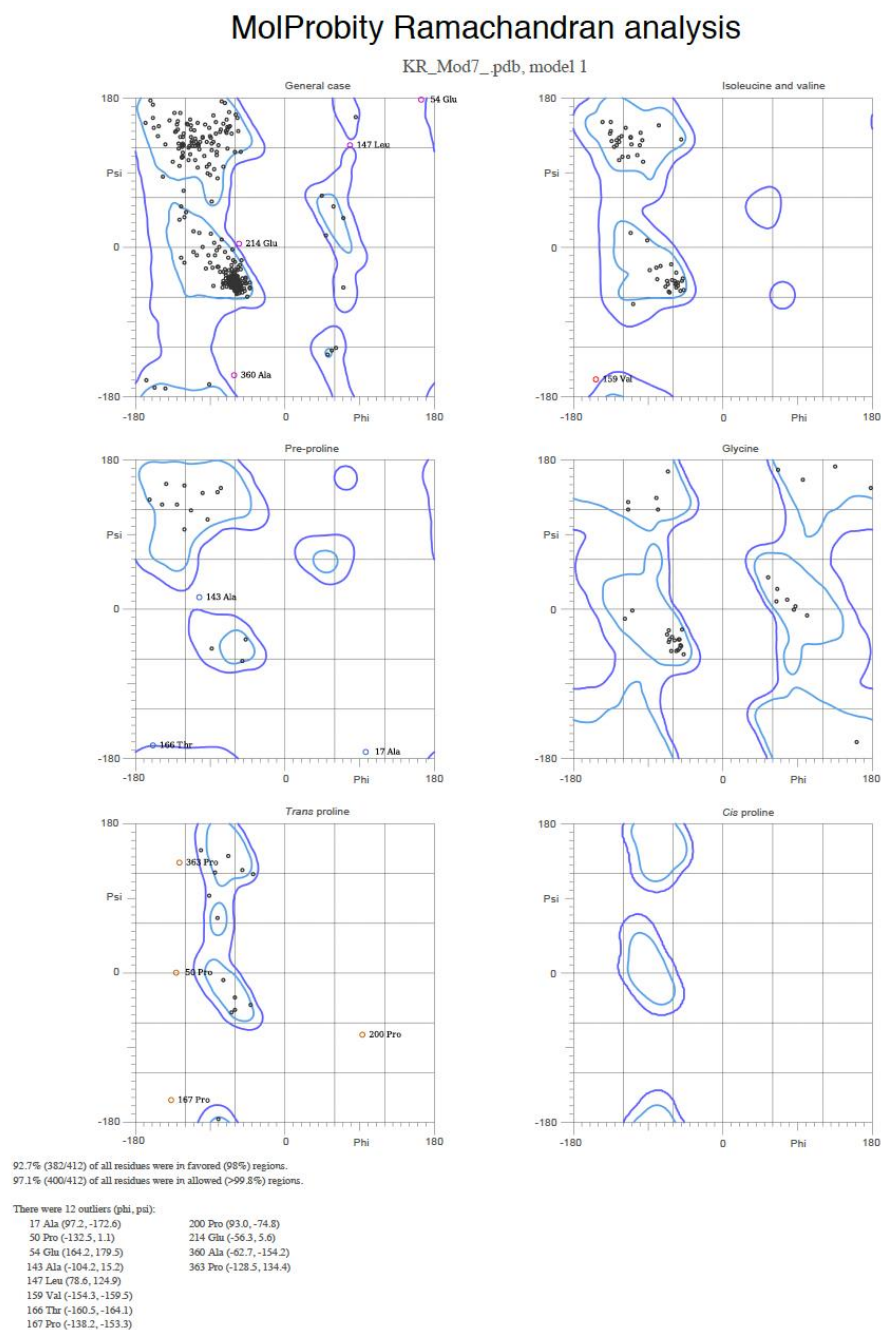
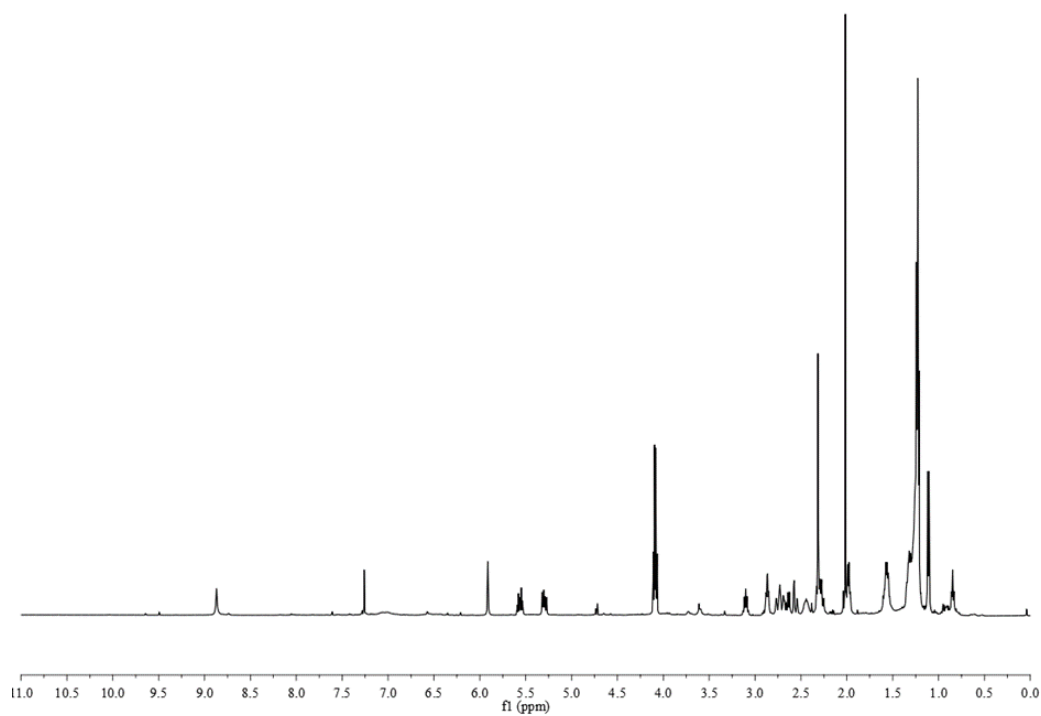
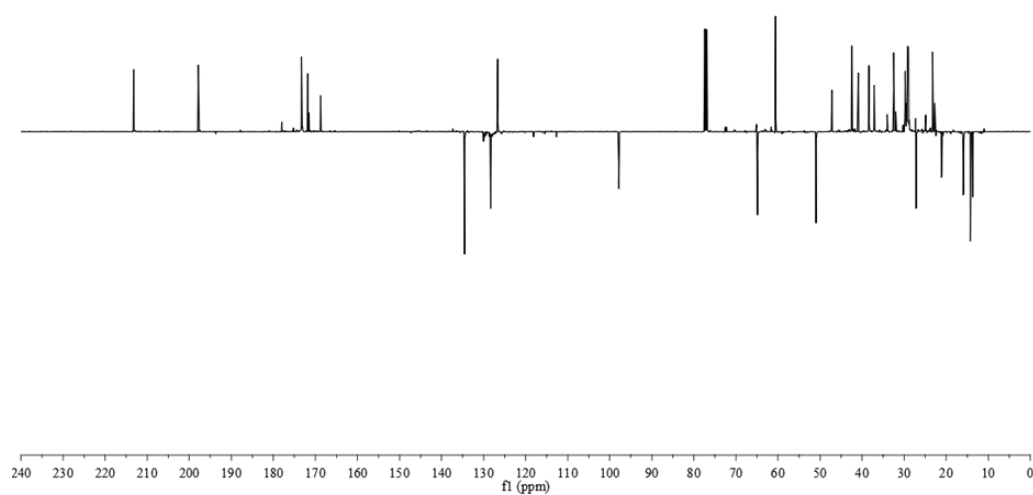


Figure A12 Ramachandran plot created using Molprobity software to assess the quality of the homology model generated for the chimeric GdsF module 7 KR.

NMR spectra

Figure A13 ^1H NMR spectrum of gladiostatin in CDCl_3 .Figure A14 ^{13}C NMR spectrum of gladiostatin in CDCl_3 .

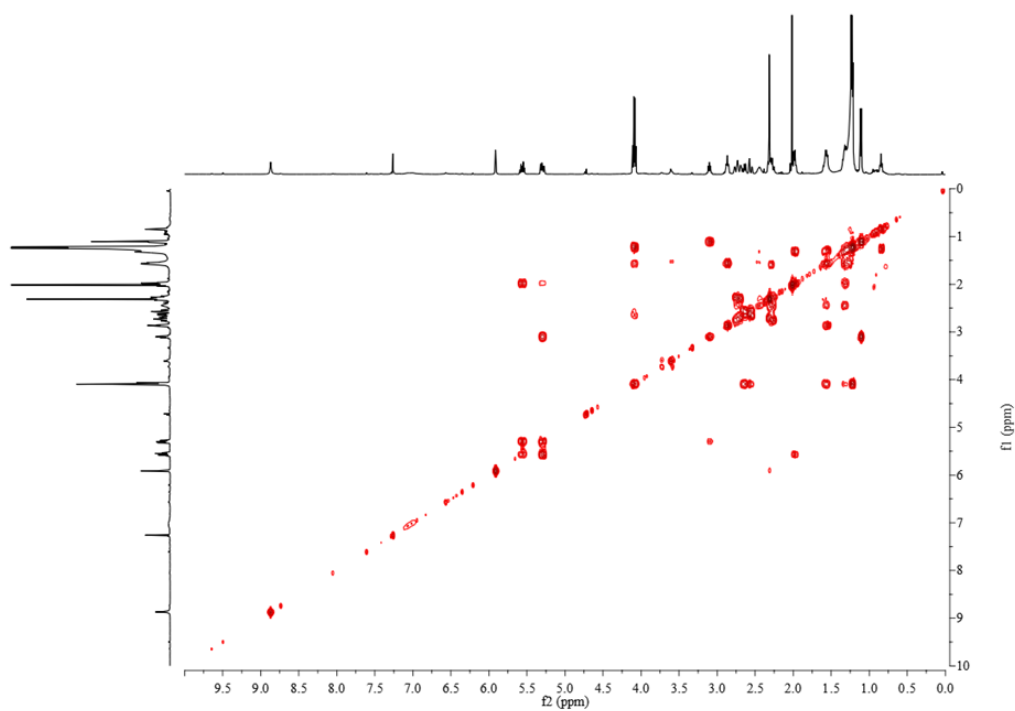


Figure A15 COSY NMR spectrum of gladiostatin in CDCl₃.

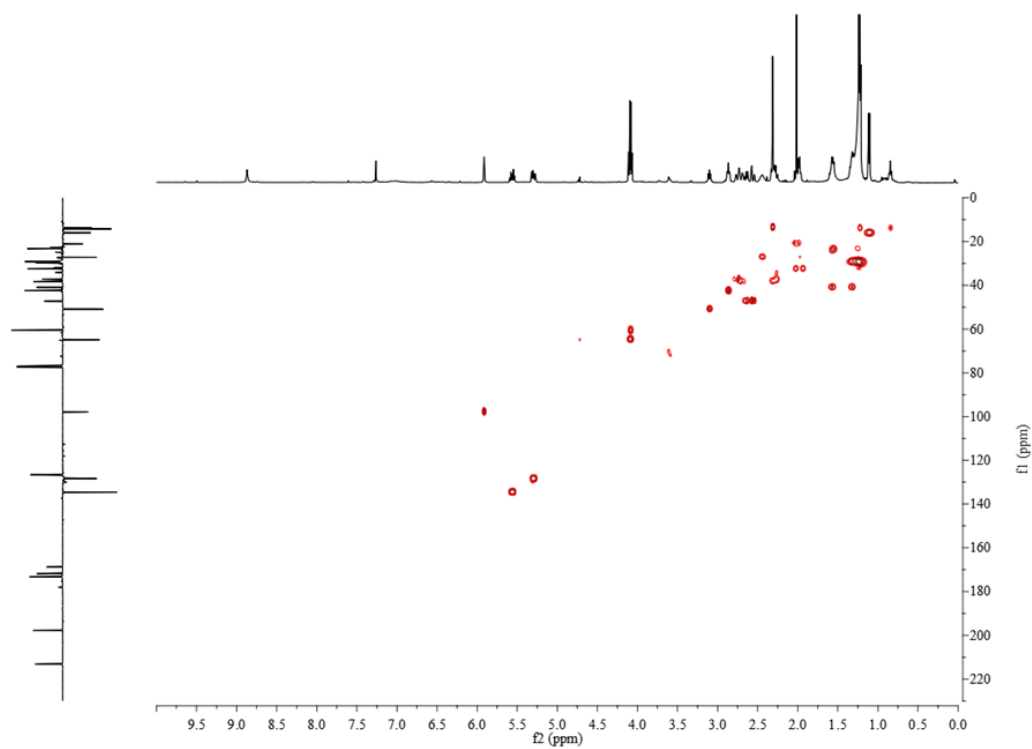


Figure A16 HSQC NMR spectrum of gladiostatin in CDCl₃.

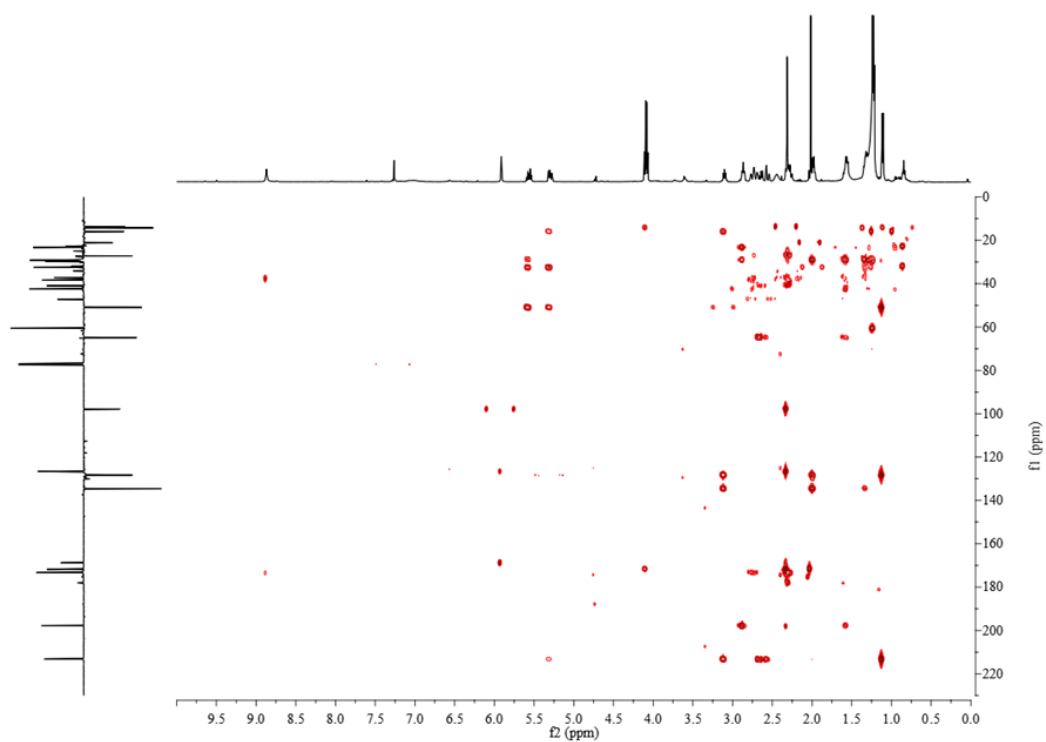


Figure A17 HMBC NMR spectrum of gladiostatin in CDCl₃.

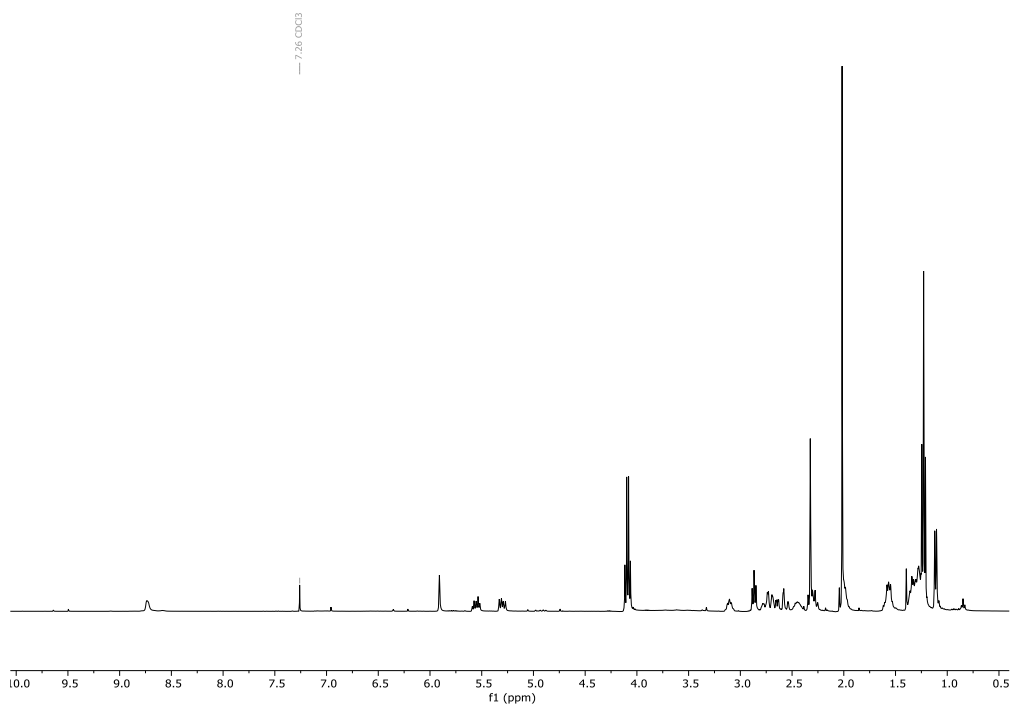


Figure A18 ¹H NMR spectrum of gladiostatin B in CDCl₃.

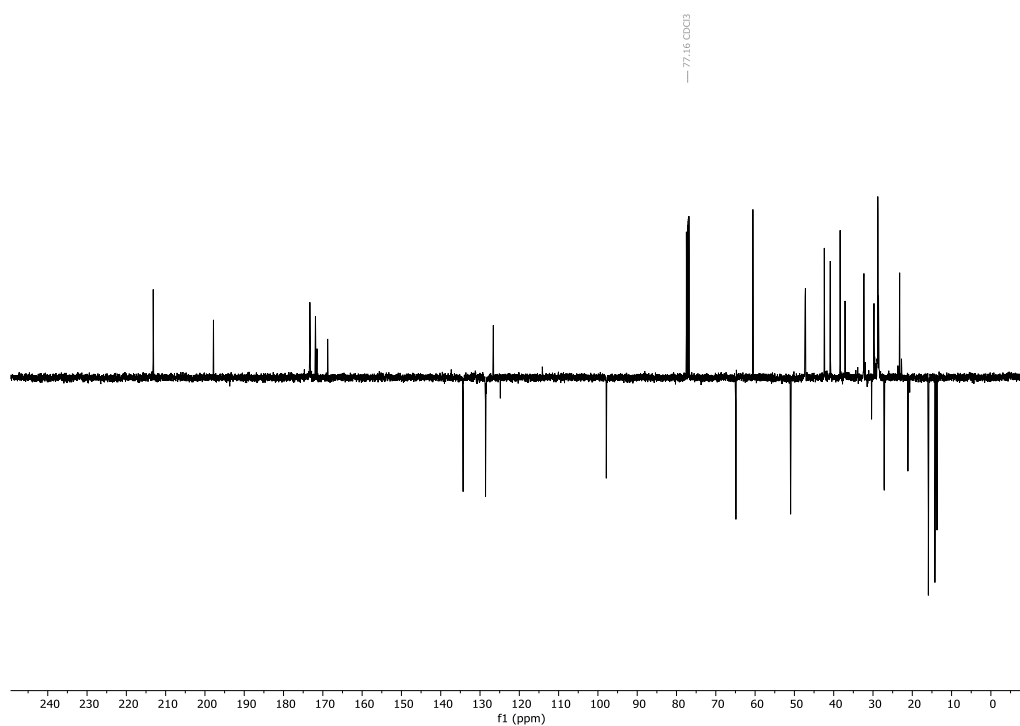


Figure A19 ^{13}C NMR spectrum of gladiostatin B in CDCl_3 .

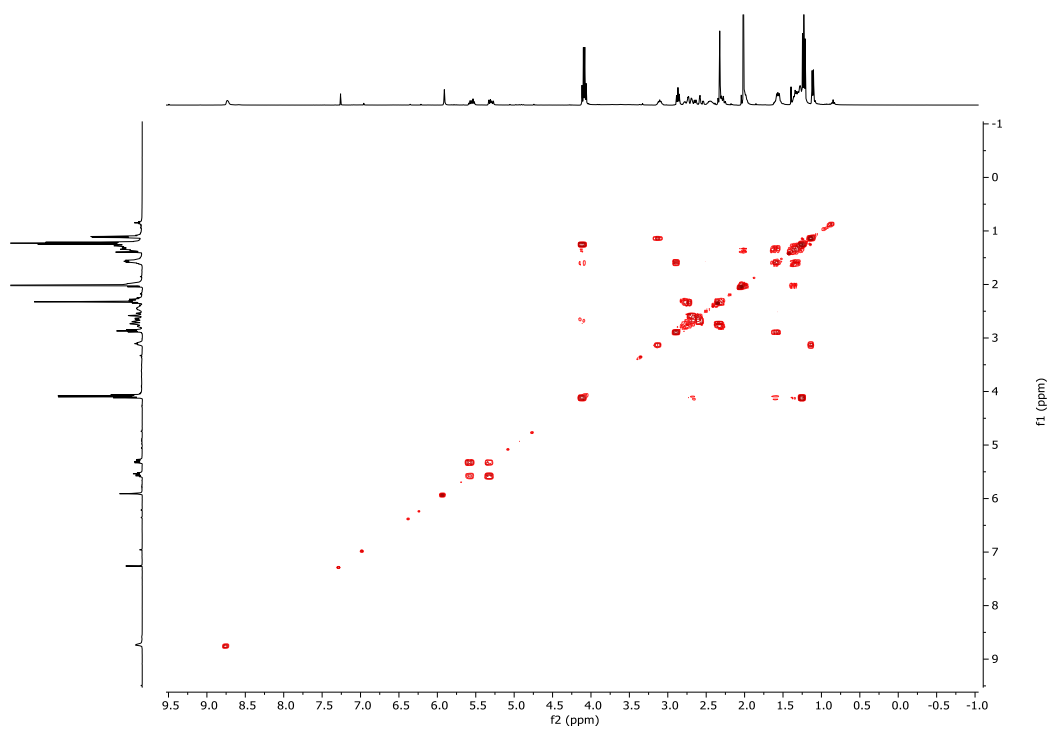


Figure A20 COSY spectrum of gladiostatin B in CDCl_3 .

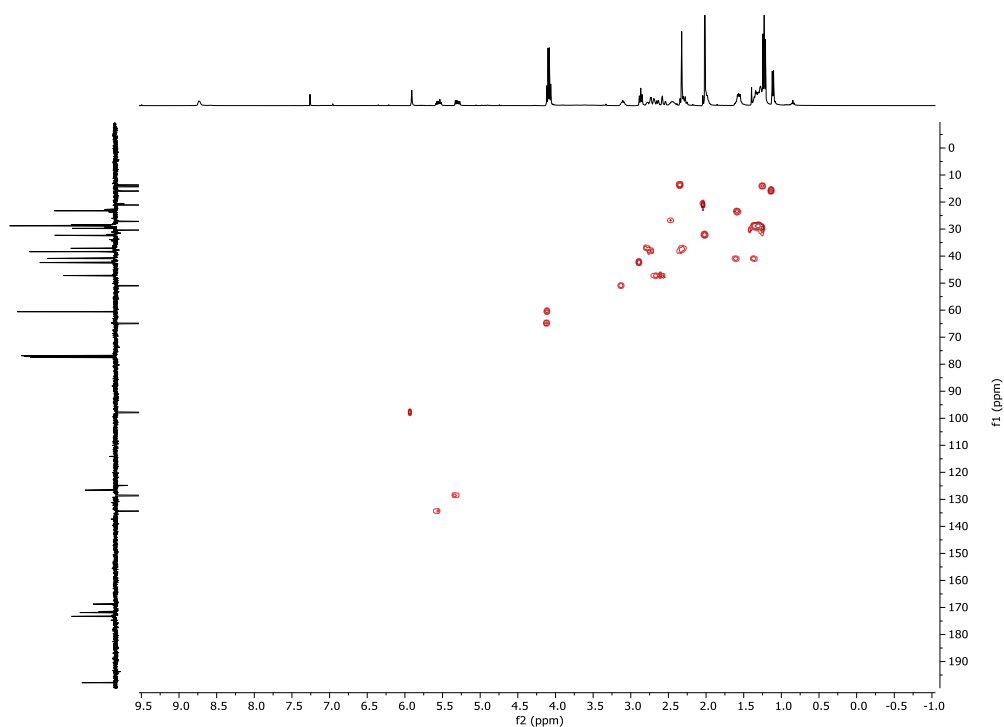


Figure A21 HSQC spectrum of gladiostatin B in CDCl_3 .

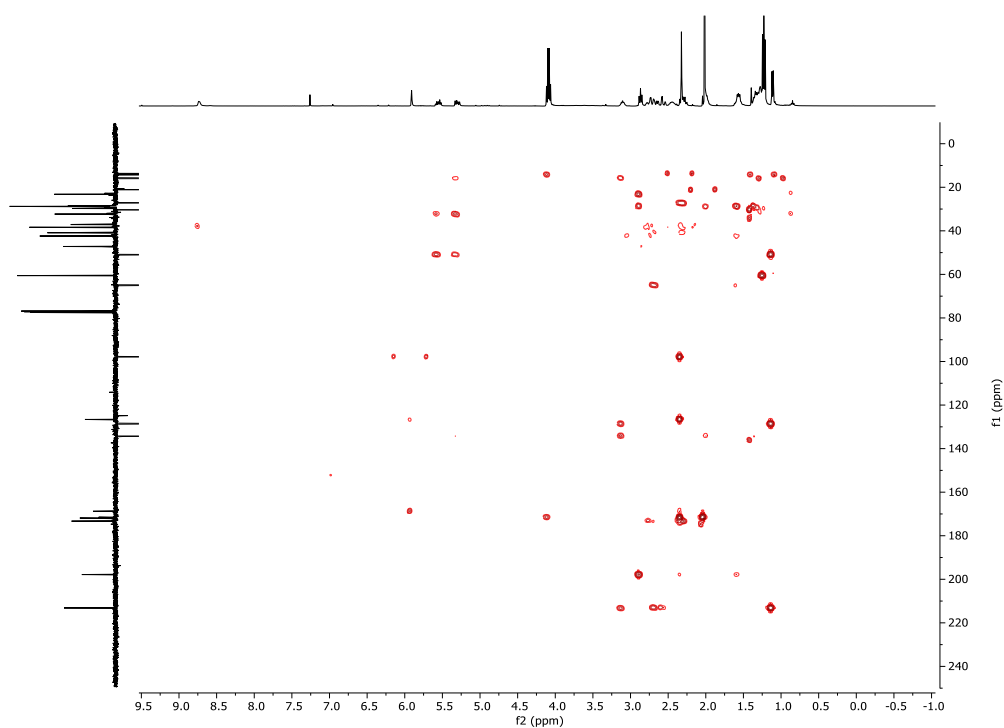


Figure A22 HMBC spectrum of gladiostatin B in CDCl_3 .

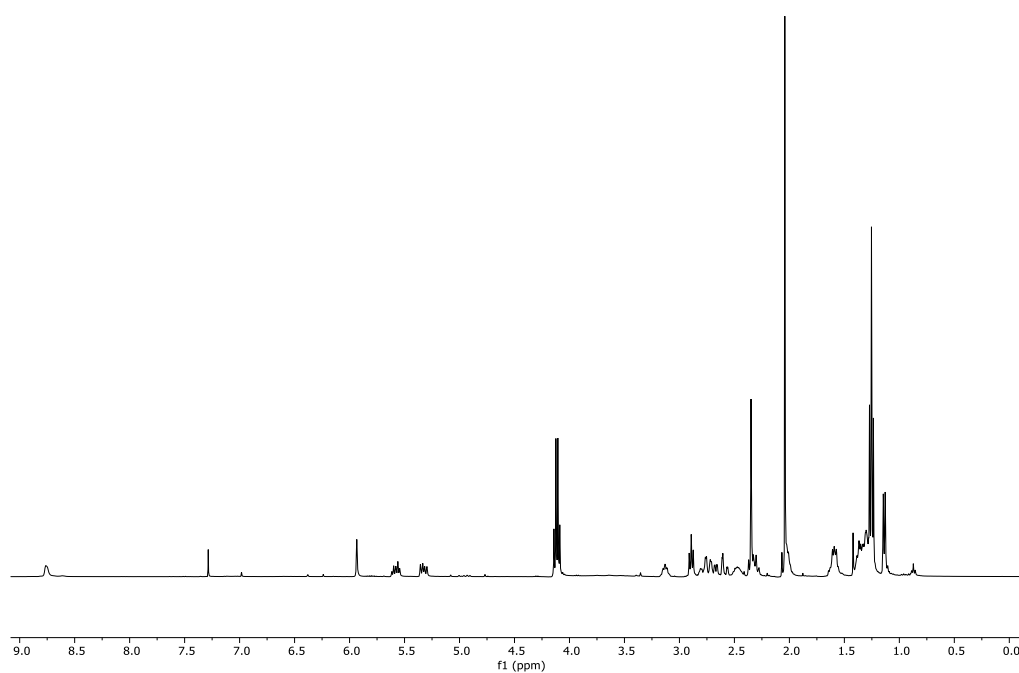


Figure A23 ^1H NMR spectrum of gladiostatin C in CDCl_3 .

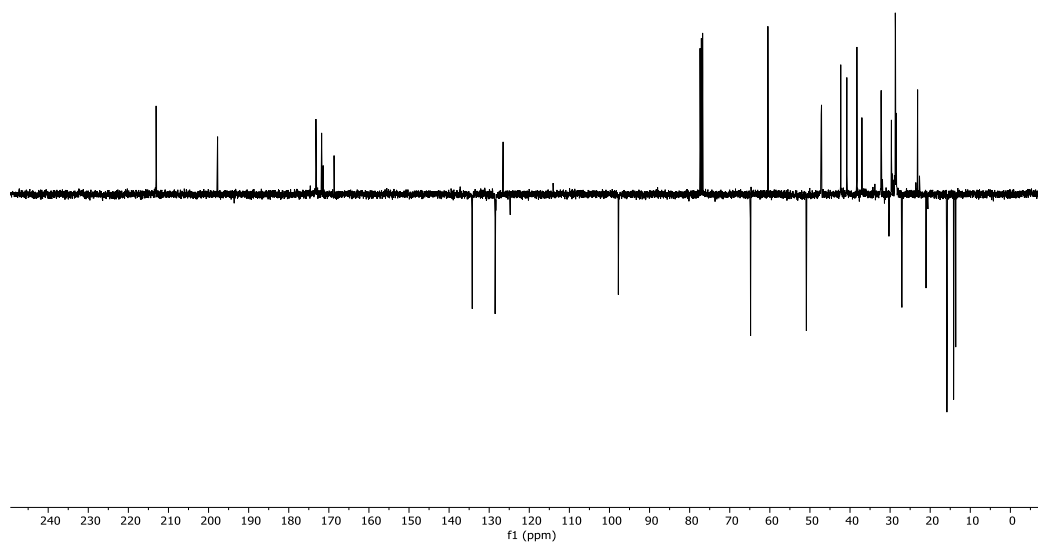


Figure A24 ^{13}C NMR spectrum of gladiostatin C in CDCl_3 .

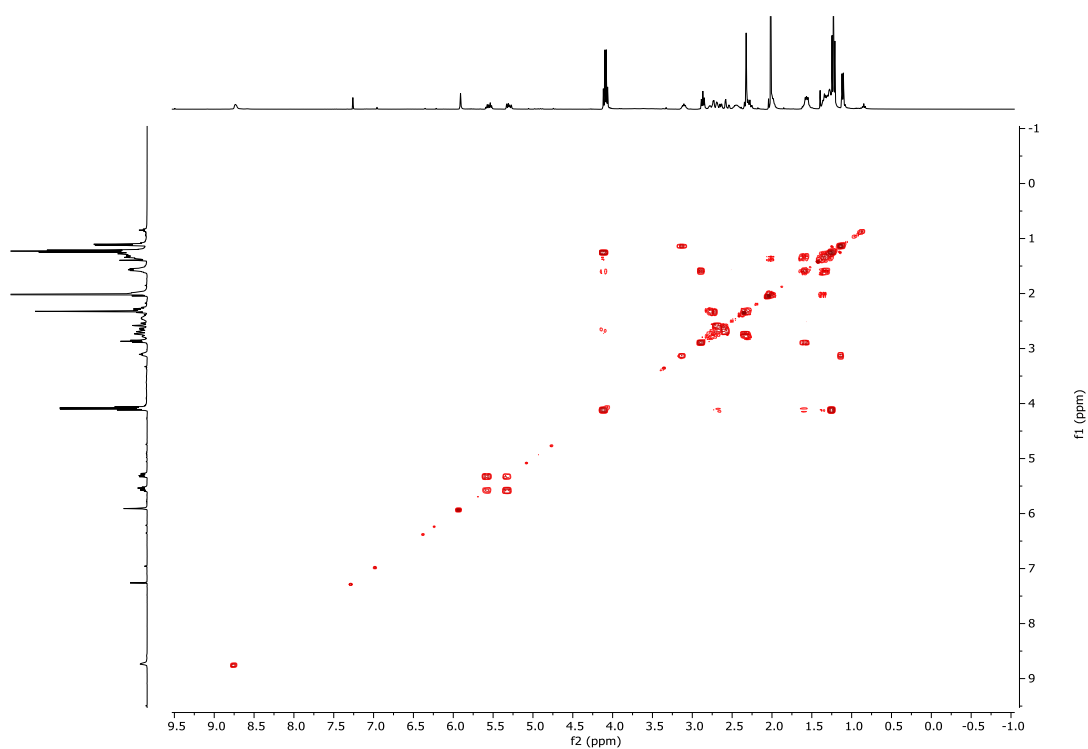


Figure A25 COSY spectrum of gladiostatin C in CDCl₃.

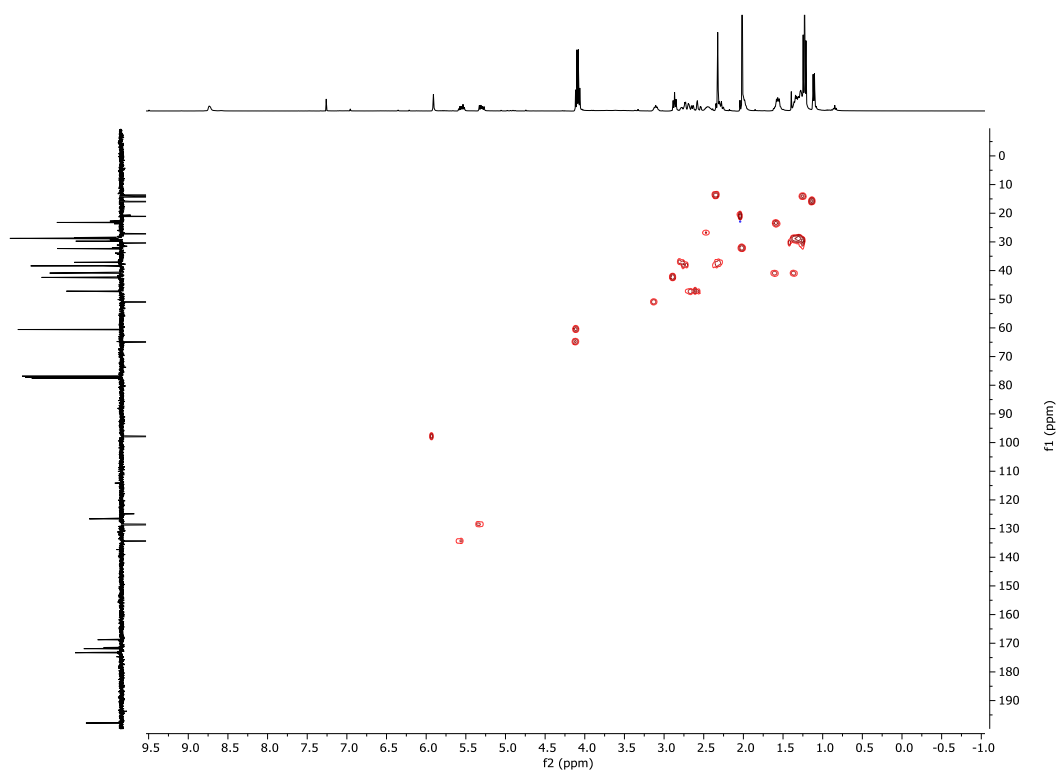


Figure A26 HSQC spectrum of gladiostatin C in CDCl₃.

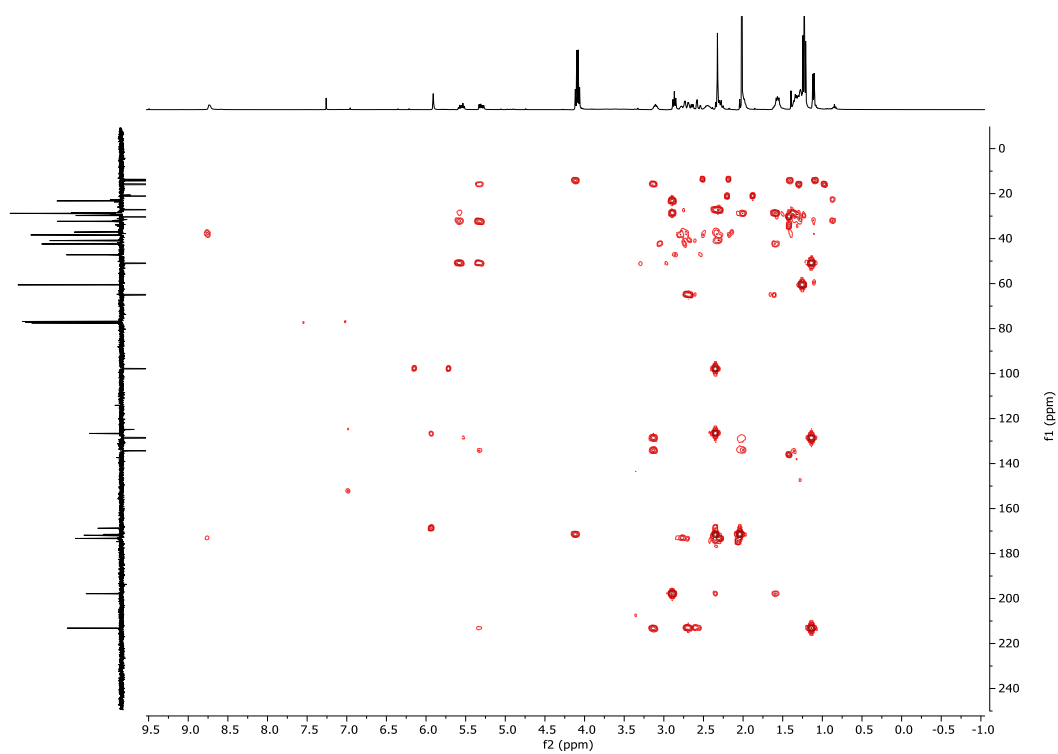


Figure A27 HMBC spectrum of gladiostatin C in CDCl_3 .

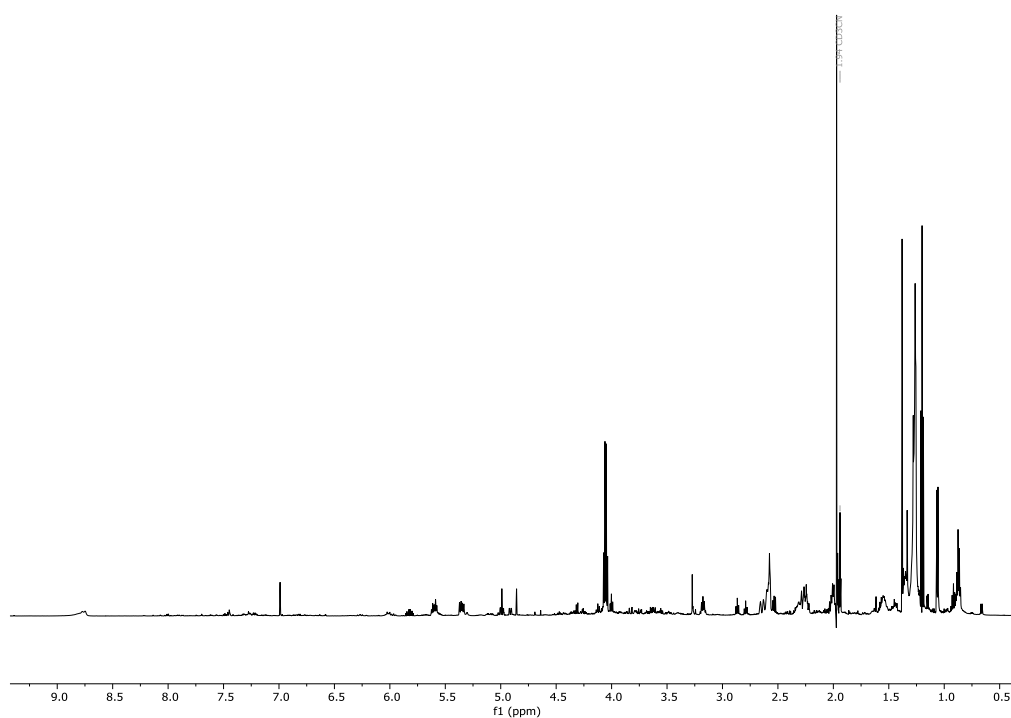


Figure A28 ^1H NMR spectrum of gladiostatin D in acetonitrile-d_3 .

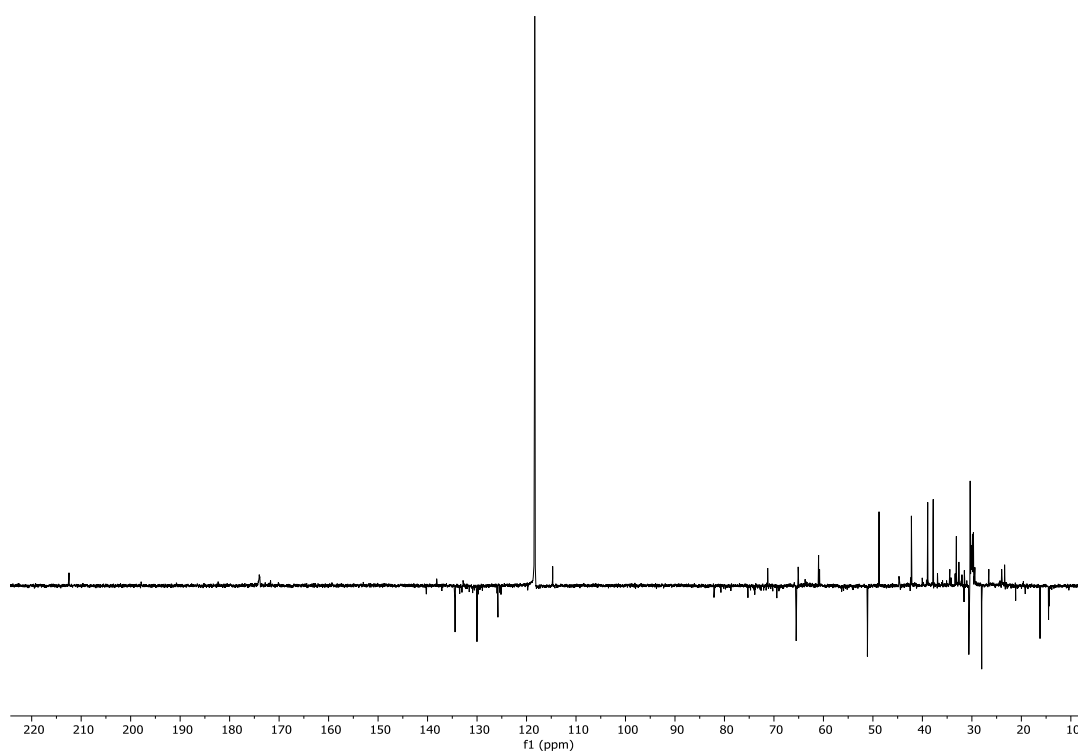


Figure A29 ^{13}C NMR spectrum of gladiostatin D in acetonitrile- d_3 .

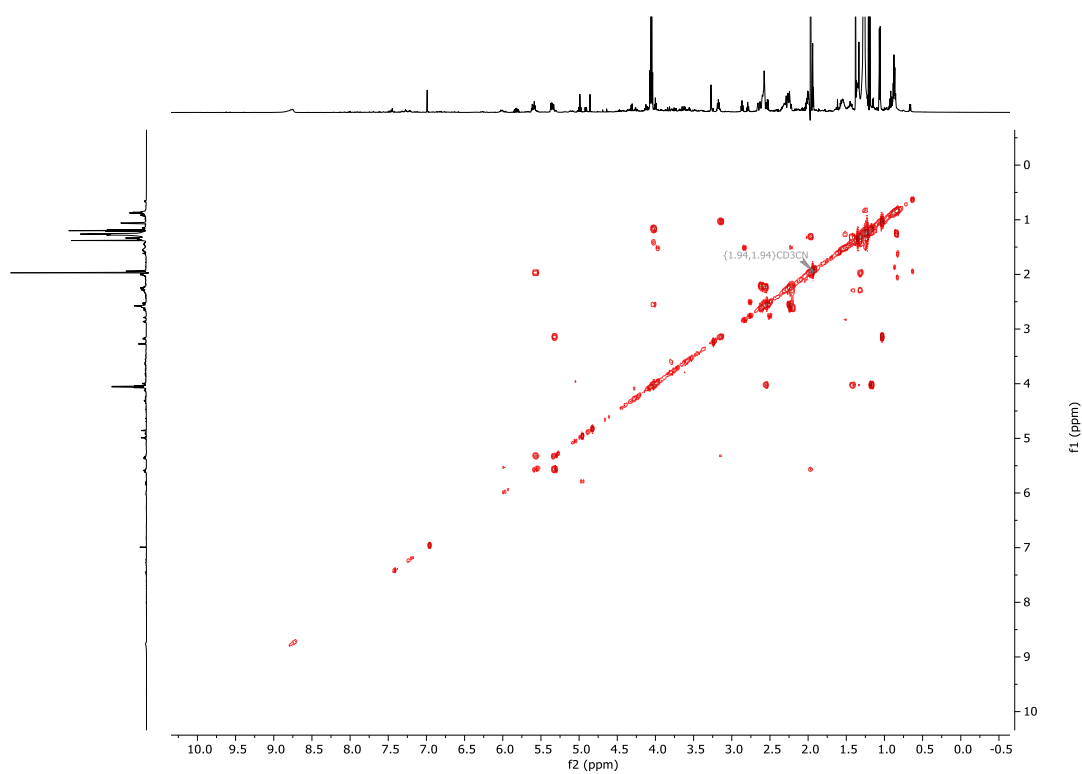


Figure A30 COSY spectrum of gladiostatin D in acetonitrile- d_3 .

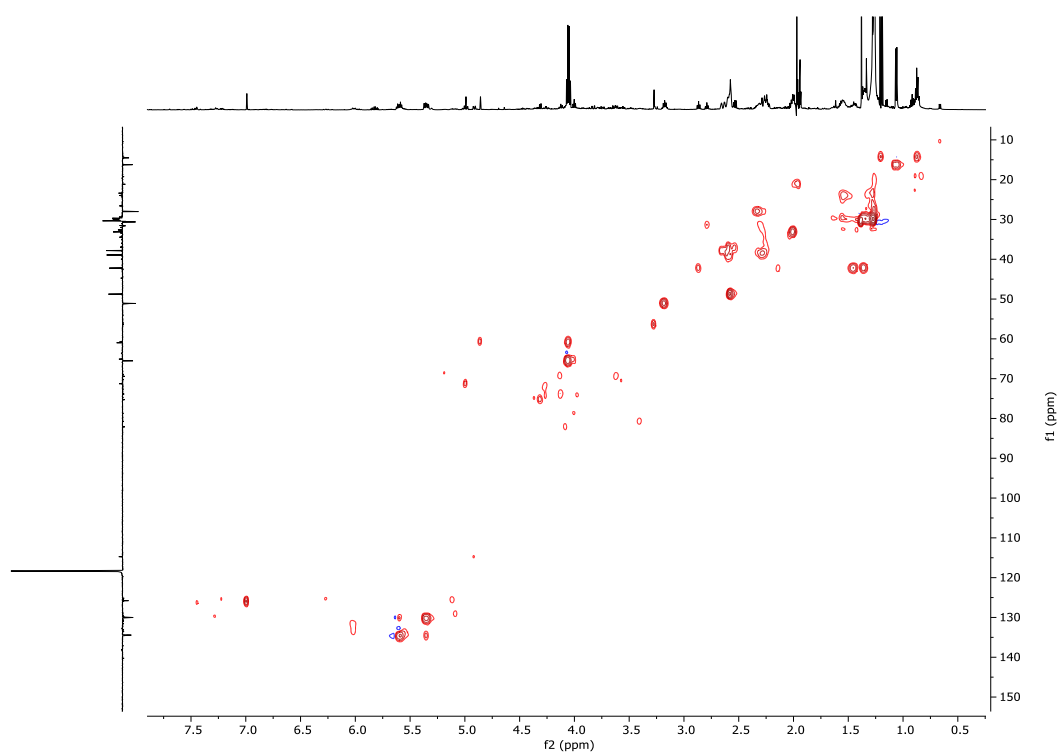


Figure A31 HSQC spectrum of gladiostatin D in acetonitrile-d₃.

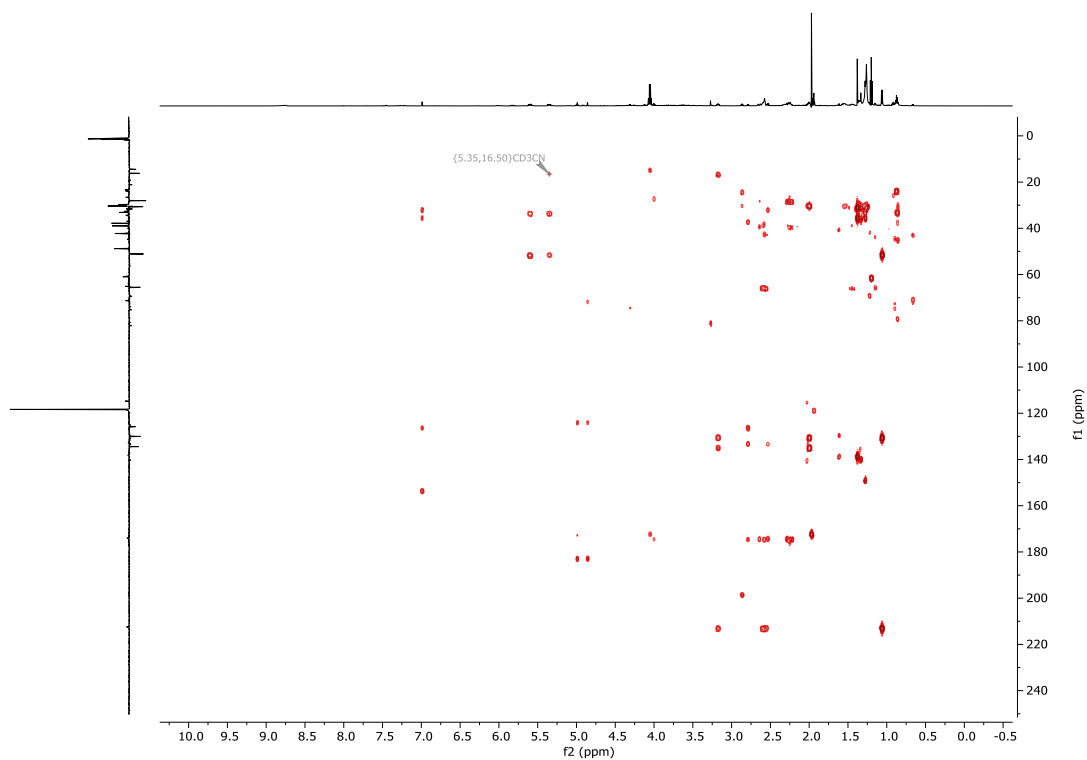


Figure A32 HMBC spectrum of gladiostatin C in acetonitrile-d₃.

Miscellaneous

Table A1 Putative functions of encoding genes within the gladiostatin biosynthetic gene cluster.

Gene/Protein	Length bp/aa	Similar Proteins	% Identity
<i>gdsA/GdsA</i>	699/233	Phosphoglycerate mutase (<i>Collimonas arenae</i>)	50
		Histidine phosphatase protein family (<i>Myxococcus xanthus</i>)	44
		Histidine phosphatase family protein (<i>Sinobacteraceae bacterium</i>)	42
<i>gdsB/GdsB</i>	3339/1113	ACP S-malonyltransferase (<i>Bacillus cereus</i>)	50
		ACP S-malonyltransferase (<i>Bacillus subtilis</i>)	48
		Malonyl CoA-ACP transacylase (<i>Paenibacillus riograndensis</i>)	47
<i>gdsC/GdsC</i>	240/80	Acyl carrier protein (<i>Streptomyces eurocidicus</i>)	55
		Acyl carrier protein (<i>Streptomyces</i> sp. W007)	52
		Acyl carrier protein (<i>Pseudomonas syringae</i>)	42
<i>gdsD/GdsD</i>	1995/665	Asparagine synthase (glutamine-hydrolyzing) (<i>Robbsia andropogonis</i>)	79
		Asparagine synthase (glutamine-hydrolyzing) (<i>Streptomyces griseus</i>)	61
		Asparagine synthase (glutamine-hydrolyzing) (<i>Stappia indica</i>)	58
<i>gdsE/GdsE</i>	9744/3248	SDR family NAD(P)-dependent oxidoreductase (<i>Robbsia andropogonis</i>)	67
		Polyketide synthase (<i>Streptomyces olivoreticuli</i>)	49
		ChxE (<i>Streptomyces</i> sp. YIM 56141)	47
<i>gdsF/GdsF</i>	30543/10181	Polyketide synthase PksN (<i>Methylobacillus anaerophila</i>)	49
		Non-ribosomal peptide synthase (<i>Bacillus swezeyi</i>)	41
		Polyketide synthase PksN (<i>Paenibacillus halotolerans</i>)	40
<i>gdsG/GdsG</i>	711/237	HAD family phosphatase (<i>Streptomyces</i> sp. BK329)	44
		Phosphoglycolate phosphatase-like HAD superfamily hydrolase (<i>Streptomyces rochei</i>)	44
		HAD family hydrolase (<i>Streptococcus thermophilus</i>)	43
<i>gdsH/GdsH</i>	1020/340	NADP-dependent oxidoreductase (<i>Paucibacter toxinivorans</i>)	65
		NADP-dependent oxidoreductase (<i>Pelomonas saccharophila</i>)	65
		NADP-dependent oxidoreductase (<i>Xylophilus ampelinus</i>)	60
<i>gdsI/GdsI</i>	906/302	LysR family transcriptional regulator (<i>Caballeronia calidae</i>)	80
		LysR family transcriptional regulator (<i>Caballeronia sordidicola</i>)	74
		LysR family transcriptional regulator (<i>Paraburkholderia caribensis</i>)	72

Table A2 MIC values determined for gladiostatin C against a range of Gram-positive, Gram-negative bacteria, fungi and yeast.

Strain	gladiostatin C
	MIC [$\mu\text{g mL}^{-1}$]
Gram-negative Bacteria	
<i>Klebsiella pneumonia</i> DSM 26371	>64
<i>Acinetobacter baumannii</i> DSM25645	>64
<i>Pseudomonas aeruginosa</i> DSM29239	>64
<i>Enterobacter cloacae</i> DSM 16690	>64
Gram-positive Bacteria	
<i>Enterococcus faecium</i> DSM25390	>64
<i>Staphylococcus aureus</i> DSM21979	>64
Fungi	
<i>Candida albicans</i> SC 5314	>64
Yeast	
<i>S. cerevisiae</i> W303-1a	4