Evolution of Diversity in Actinobacterial Assembly-Line Biosynthesis

Thomas James Booth

John Innes Centre

Department of Molecular Microbiology

Norwich Research Park, Colney Lane, Norwich NR4 7UH, UK

Thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

September 2018

Evolution of Diversity in Actinobacterial Assembly-Line Biosynthesis

Thomas James Booth, John Innes Centre, September 2018

Abstract

Actinobacteria are prolific producers of natural products of significant clinical and industrial importance. Among the most common natural product biosynthetic pathways are the type I modular polyketide synthases (PKSs) and the non-ribosomal peptide synthetases (NRPSs); large assembly-line megasynthases. This thesis investigates the biosynthesis and evolution of five families of assembly-line natural products.

Firstly, I report the desotamide (*dsa*) and wollamide (*wol*) biosynthetic gene clusters (BGCs) from *Streptomyes* sp. MST-70754 and *Streptomyes* sp. MST-110588, respectively. The *wol* BGC was found to encode a unique bifurcated NRPS assembly-line capable of producing both the desotamides and the wollamides. Genomic analysis supports the emergence of the *wol* BGC through an ancestral intergenic gene duplication followed by an intragenomic recombination event.

Secondly, I report the BGC of the β-amino acid containing polyketide macrolactam (βPM) heronamide C and characterise its spontaneous thermal [6π + 4π] and photochemical [6π + 6π] intramolecular cycloadditions. Furthermore, sub-cluster genome mining using conserved genes involved in the incorporation of the β-amino acid identified 40 orphan BGCs putatively belonging to the βPM family. Phylogenetic analysis of these BGCs allowed strains to be prioritised for metabolic analysis, leading to the identification of candidate βPMs from *Nocardiopsis gilva* DSM44841.

Finally, I report the BGCs of the polyketide macrolide pladienolide and the polyketide spirotetronate spirohexenolide and activated their expression in strains of *Streptomyces platensis* through the overexpression of LAL-family and SARP-family transcriptional activators.

Acknowledgments

Foremost, I want to thank Professor Barrie Wilkinson for his supervision and mentorship. He has allowed me to pursue my interests and has taken great interest in my future and for this I am eternally grateful. Working together over the last four years has been a pleasure. Furthermore, I wish to extend my thanks to Dr. Andrew Truman and Juan Pablo Gomez-Escribano for their input on this thesis as members of my supervisory committee; Professor Jason Micklefield, Dr. Lu Shin Wong and Dr. Sarah Shepherd for hosting me at the University of Manchester and; Dr. Ernest Lacey who provided many of the strains discussed in this thesis.

I am grateful to members of the Wilkinson lab past and present. In particular: Dr. Silke Alt for helping me to learn molecular biological techniques when I first arrived in the lab, Dr. Kenan Bozhueyuek for many inspiring discussions; Dr. David Widdick for his infinite patience in answering my questions and; my fellow PhD students Dr. Thomas Scott, Mr. Tom Eyles, Mr. Sebastian Samwald and Ms. Eleni Vikeli for their comradery. I consider myself extremely fortunate to have been given the opportunity to work in the Department of Microbiology and would like to thank all its members (past and current) for creating such a positive working environment and acknowledge the hard work of the support staff in keeping the department running smoothly. In addition, I want to acknowledge the fantastic community of PhD students at the John Innes Centre and across the Norwich Research Park and would like special thanks to Mr. Sylvain Capdevielle and Ms. Nicola Capstaff for their unwavering support and, Mr. Billy Aldridge and Ms. Jessica Garnet for helping me stay sane during the writing process.

Finally, I wish to thank my friends and family, in particular: my sister Lucy who is an endless source of inspiration and, my mum and dad to whom I owe everything.

Authors Declaration

The research described in this thesis was conducted at the John Innes Centre between October 2014 and September 2018. All data described here are original and were obtained by the author, except where specific acknowledgement has been made. No part of this thesis has previously been submitted for a degree at this or any other academic institution.

Yeast mediated assembly of pBO1 was performed by Dr. Kenan Bozhueyuek
The genome sequence for *Streptomyces* sp. CMB-0406 was obtained by Dr. Silke Alt.
Assembly of plasmin pGP9-*spxR2* was performed by Sebastian Samwald.

Chapter 3 of the thesis has been the subject of the following publication:

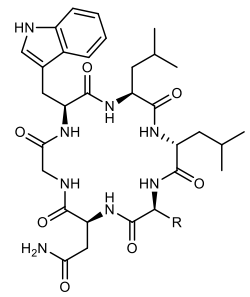
> Booth, T. J., Alt, S., Capon, R. J. and Wilkinson, B. Synchronous Intramolecular Cycloadditions of the Polyene Macrolactam Polyketide Heronamide C. *Chem. Commun. (Camb).* **2016**, *52* (38), 6383–6386.

Abbreviations

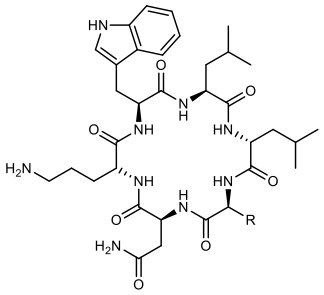
ACP	Acyl carrier protein
A-domain	Adenylation domain
AMP	Adenosine monophosphate
Arg	Argenine
Asn	Asparagine
AT	Acyltransferase
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
BGC	Biosynthetic gene cluster
C-domain	Condensation domain
CoA	Coenzyme A
DAase	Diels-Alderase
DH-domain	Dehyratase domain
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
E-domain	Epimerisation domain
EMS/MMS	Ethyl/methyl methanesulfonate
ER	Enoylreductase
Gly	Glycine
HPLC	High performance liquid chromatography
Ile	Isoleucine
KR	Ketoreductase
KS	Ketosynthase
Kyn	Kynurenine
LAL	large ATP-binding regulators of the LuxR family
LCMS	Liquid chromatography mass spectrometry
Leu	Leucine
mCPBA	Meta-chloroperoxybenzoic acid
MNNG	methylnitronitrosoguanidin
MRSE	Methicillin resistant *Staphylococcus epidermis*
NRPS	Non-ribosomal peptide synthetase
ORF	Open reading frame
Orn	Ornithine
OSMAC	One strain many compounds

PAC	P1 artificial chromosome
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
Phe	Phenylalanine
PKS	Polyketide synthase
SARP	Streptomyces antibiotic regulatory protein
Ser	Serine
T1PKS	Type I polyketide synthase
T1TE	Type I thioesterase
T2PKS	Type II polyketide synthase
T2TE	Type II thioesterase
TAR	Transformation associated recombination
T-domain	Thiolation domain
TE	Thioesterase
TES	Triethylsilane
Trp	Tryptophan
Val	Valine
WGS	Whole genome shotgun
βPM	β-amino acid containing polyketide macrolactam

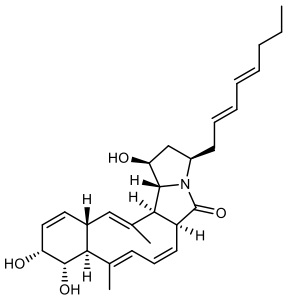
Compound Key



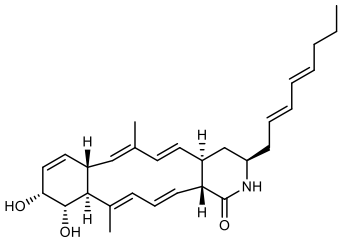
1: Desotamide A, R = L-*allo*-Ile
2: Desotamide B, R = L-val



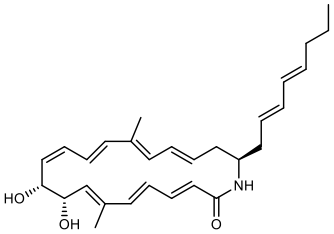
3: Wollamide A, R = L-*allo*-Ile
4: Wollamide B, R = L-val



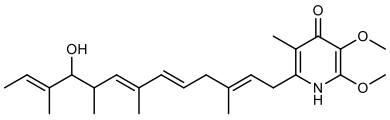
5: Heronamide A



6: Heronamide B

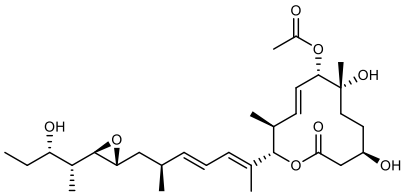


7: Heronamide C

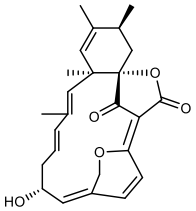


8: Piericidin A

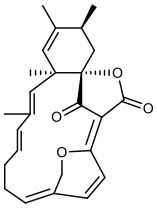
9 and 10: $C_{29}H_{40}N_2O_4$
11: $C_{23}H_{29}NO_2$
12: $C_{23}H_{27}NO_2$



13: Pladienolide B

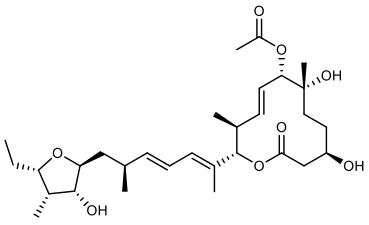


14: Spirohexenolide A



15: Spirohexenolide B

16: M = 673.3562
17: M = 600.3521
18: M = 586.3305
19: M = 362.1408



20: Isopladienolide B

Table of Contents

Index of Figures

Index of Tables

Chapter One: Introduction

Contents

1.1 A Historical Perspective on Natural Product Discovery

The discovery of actinomycin by Selman Waksman in 1940 was a turning point in the history of microbial natural products[1]. Although preceded by the discovery of other antibiotic substances, most notably penicillin by Fleming in 1929[2] (by Chain et al., 1940[3]) and tyrothricin by René Dubos in 1939[4,5], the discovery of actinomycin marked a step-change in the rate of natural product discovery. Unlike penicillin and tyrothricin, actinomycin would not find clinical use as an antibiotic, however it would eventually become the first bacterial agent used to treat cancer[6]. Perhaps more importantly, the discovery of actinomycin marked a profound increase in the rate at which new antibiotics were discovered. Inspired by Dubos' work with the soil bacterium *Bacillus brevis*[4,5], Waksman pursued a strategy of specifically screening microbes from the soil, which consequently led to the discovery of prolific producers of bioactive natural products, namely species of actinobacteria, firmicutes and ascomycete fungi [1,7].

The next twenty years marked what is commonly referred to as the 'Golden Age of Antibiotic Discovery.' During this time, many of the major groups of natural products were discovered or significantly expanded, most prevalently: aminoglycosides (e.g. streptomycin, neomycin and kanamycin[8–10]), polyketides (e.g. tetracyclins and erythromycin[11–13]) and non-ribosomal peptides (e.g. bacitracin and vancomycin[14,15]), including β-lactams (e.g. cephalosporins[16,17]) (Figure 1. 1). However, by the end of the 1960s, interest in antimicrobial discovery would begin to decline. Writing in 1965, Waksman discusses the 'duplication problem' and comments on the high frequency of neomycin rediscovery[18]. He cites an incidence in the

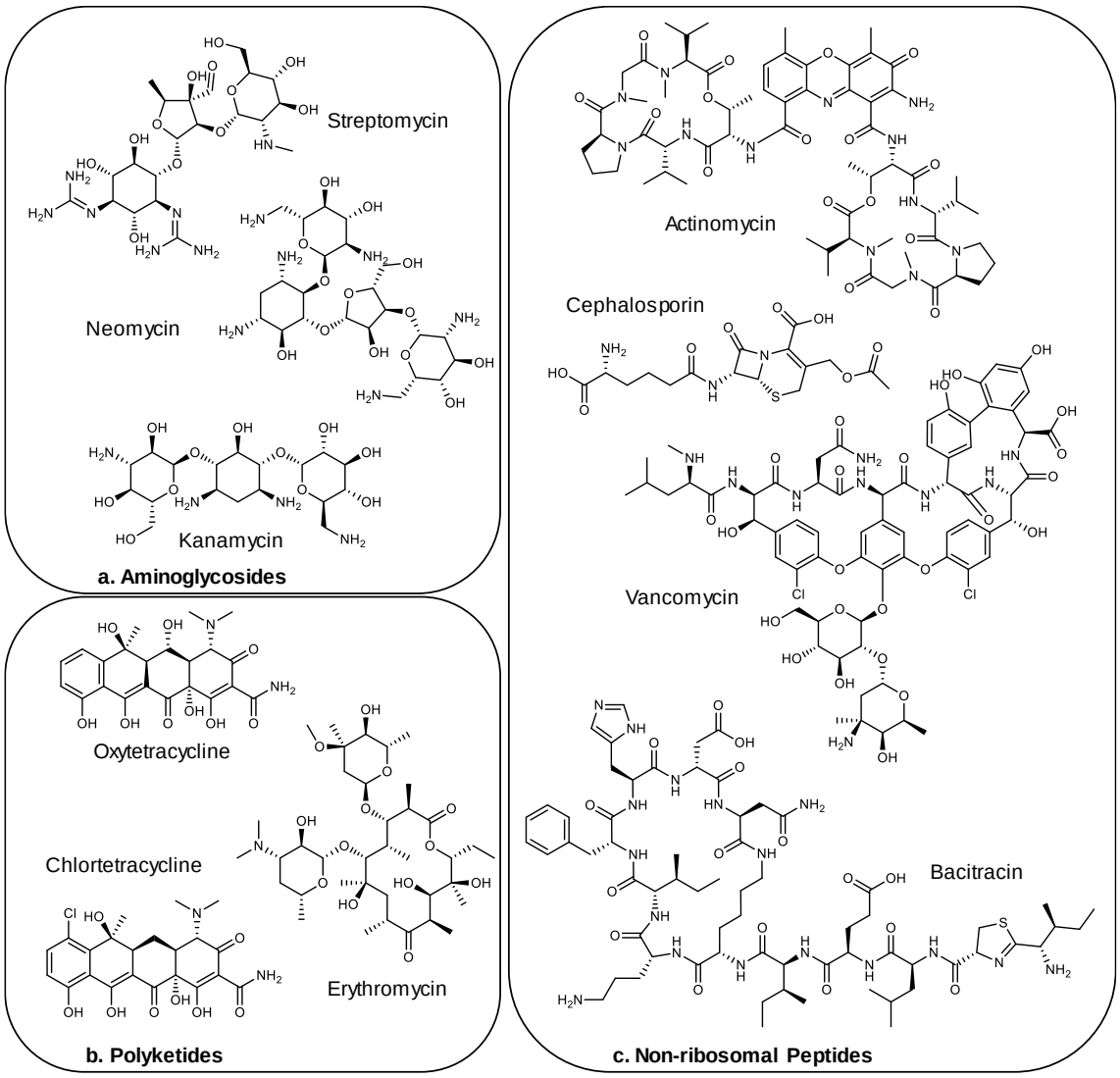


Figure 1. 1: Golden Age antibiotics. Representative compounds of the three main families of antibiotics discovered during the 'golden age' of antibiotic discovery, 1940-1960.

Soviet Union where three 'new' antibiotic substances were permitted for medicinal use over a six-year period, all were in fact neomycin – discovered a decade earlier. However, rediscovery was only a single factor. The introduction of stricter standards for the approval of new clinical drugs, increased time and cost associated with the development of lead compounds for the clinic, and the difficulty of new drugs to compete within already established markets are the most often cited causes for this decline[19].

As a result of these diminishing returns, research focus shifted from the discovery of new antibiotic compounds to other avenues, notably mode of action, structural elucidation and biosynthesis[19]. By the mid-1980s, advances in molecular biology, such as *in vitro* recombination and later PCR and Sanger sequencing[20,21], allowed the first natural product biosynthetic genes to be cloned. Among the first of these genes were those responsible for the biosynthesis of actinorhodin. This system was a useful model to probe biosynthesis due to its characteristic blue colour (from which the model organism *Streptomyces coelicolor* derives its name)[22]. The identification of the genes involved in actinorhodin biosynthesis provided the first opportunity for the genetic engineering of novel natural products, in the form of hybrid antibiotics. By transforming medermycin and granaticin producing *Streptomyces* spp. with plasmids containing different transcriptional units of actinorhodin, Hopwood et al., were able to isolate strains producing the hybrid compounds merderrhodin A and dihydrogranaticin[23]. In addition, these compounds, like their progenitors, demonstrated biological activity, validating the engineering of biosynthetic gene clusters (BGCs) as a strategy for developing new antimicrobials. Following the identification of the actinorhodin BGC, the genes responsible for the production of many more natural products, including tyrocidine[24,25] and gramicidin[26] (the constituents of tyrothricin), and erythromycin[27,28], discussed below, would follow.

In 2002, the publication of the *Streptomyces coelicolor* A3(2) genome[29] led to another paradigm shift in the field of microbial natural products. The genome sequence allowed BGCs to be assigned to previously orphan metabolites, such as geosmin[29,30] and desferrioxamine E and G1[29,31]. Perhaps more importantly, the genome was shown to contain BGCs for compounds that had yet to be observed from *S. coelicolor* A3(2) cultures. The sequencing of other Actinobacteria revealed that this 'cryptic metabolism' was commonplace[32–35], indicating that a wealth of chemical diversity existed that was not observed under standard laboratory conditions. The accumulation of new genome sequences coupled with the abundance of undescribed biosynthetic pathways popularised the strategy of genome mining; isolating new natural products on the basis of genetic information. Although genome mining efforts, particularly of polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs), had been possible prior to genome sequencing (e.g. the discovery of coelichelin from *S. coelicolor*[36]), access to genomic data massively expanded the scope of genome mining projects[37]. This was compounded by the development of 2nd generation sequencing technologies (pyrosequencing, ion torrent and bridge

amplification[38–41]) leading to an exponential decrease in the cost of whole genome sequencing (Figure 1. *2*)[42,43]. This meant that genome sequencing was no longer limited to well-resourced laboratories or conglomerates. As a result, the number of publicly available microbial genomes has increased exponentially (at the time of writing there are over 160,000 prokaryotic genome sequences in GenBank alone[44]). Estimates of the total diversity of natural product biosynthesis pathways in Nature vary considerably[45,46]. However, novel culture independent approaches to genome mining, such as single cell genomics or the cloning of metagenomic BGCs increase the available diversity considerably[47–49].

This abundance of data means we are more capable than ever to ask fundamental questions about natural product biosynthesis and their evolution[50]. Furthermore, genome sequences provide broader insight into the biology of producing organisms and their relationships. Recent genome-based phylogeny of Actinobacteria and studies into the rates of horizontal gene transfer are good examples of this[51,52].

1.2 Assembly-line Biosynthesis

The most abundant natural product biosynthetic pathways in Actinobacteria (with the possible exception of glycosides[46]) are the modular assembly-lines: non-ribosomal peptide synthetases (NRPSs) and type I polyketide synthases (T1PKSs). The assembly-lines consist of complexes of large proteins with repetitive peptide sequences, each repeat encoding several individual enzymatic domains. These repeats are known as modules, and each module is responsible for the incorporation (and often modification) of a single monomer into the polymer chain. Due to their modular nature, biosynthetic assembly-lines are attractive targets for synthetic biology.

1.2.1 Non-Ribosomal Peptide Synthetases

The study of NRPSs began in the 1960s, with the observation that certain polypeptides are produced in particle-free cell extracts, even in the presence of ribosome inhibitors[53,54]. The majority of the early biochemical characterisation of NRPSs was focused on the biosynthesis of gramicidin and tyrocidine, the components of tyrothricin discussed above. Prior to the publication of the first NRPS gene sequences, it was already apparent that non-ribosomal peptides were synthesised in an ATP-dependent stepwise fashion from modular proteins, through the condensation of covalently bound amino acids[55–57]. The first NRPS nucleotide sequences emerged in the late 1980s. The first was the sequence of the tyrocidine synthetase gene *tycA*[24], followed closely by the sequence of the gramicidin synthetase[58]. These sequences, and others[59,60], showed that NRPSs are encoded by repetitive sequences, with the number of repeats correlating with the

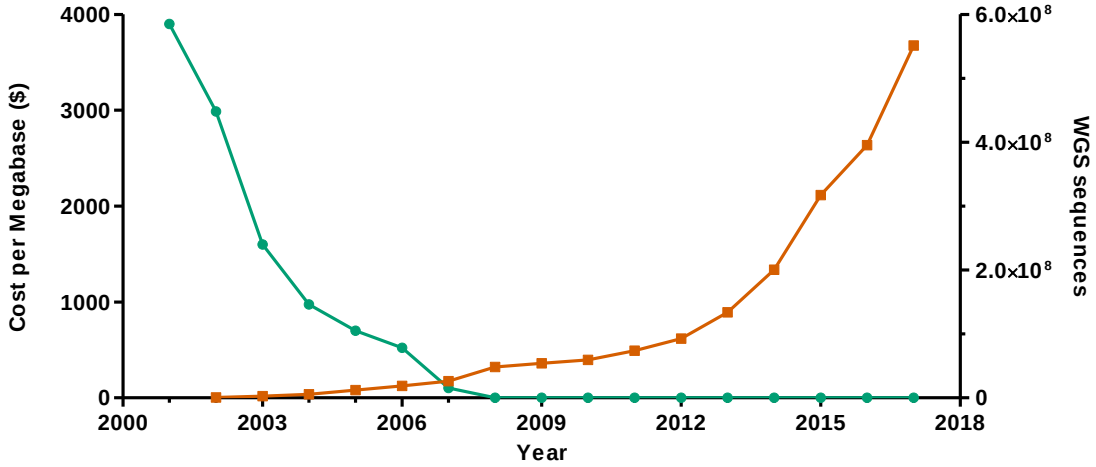


Figure 1. 2: Change in the cost and frequency of sequencing over time. Cost per megabase according to reference Wetterstrand[42] vs. number of whole genome shotgun (WGS) sequences in GenBank. In both cases figure is taken from the latest data point in a given year.

number of amino acids in the peptide product, confirming the notion of modularity that was developed into what we recognise today as the multiple carrier model[61].

Each module is responsible for the incorporation of a single amino acid and consists of, at minimum, three core domains: an adenylation (A)-domain, a thiolation (T)- domain or peptidyl-carrier protein (PCP), and a condensation (C)-domain (Figure 1. 3). The A-domains are responsible for the selection of a specific amino acid and catalyse the ATP-dependent activation of the carboxyl carbon[62,63]. Unlike in ribosomal peptide synthesis, this process is not restricted to proteinogenic amino acid substrates, meaning NRPSs have access to much broader structural diversity – even before taking into account any downstream modifications. Estimates put the number of accessible monomers at over 500, but theoretically this number could be much greater[64]. For example, gramicidin and tyrocidine both contain ornithine residues and a particularly remarkable example is the hexadepsipeptide kutzneride, which contains six highly modified non-proteinogenic residues (Figure 1. 4)[65]. Additionally, the substrates of adenylation domains are not limited to α-amino acids and may include β-amino or α-hydroxy acids[65,66]. The adenylation domain also catalyses the transfer of the activated amino acid to the thiol group of a prosthetic phosphopantetheine tethered to the T-domain. Upon the formation of the aminoacyl thioester, conformational change in the adenylation domain will reposition the T-domain so that the aminoacyl thioester is positioned in the active site of the C-domain. If both upstream and downstream substrates are present in the C-domain active site, it can catalyse peptide bond formation. This cycle will continue in a stepwise manner through each module until the final polypeptide is synthesised. Following the final catalytic cycle, the peptide is released. This can be catalysed through the action of a thioesterase, terminal reductase or, in fungal NRPS systems, terminal condensation (C_T) domains[67–69]. These domains however are not observed in every NRPS system and so are not considered essential.

Additional domains may further modify the nascent peptide. By far the most common of these tailoring domains is the epimerisation (E)– domain. Following the incorporation of an L-amino acid to the nascent peptide, E-domains convert the amino acid between the L- and D-configurations[70]. Condensation domains specific for the D-amino acid in the growing chain ($^{D}C_{L}$-domains) then catalyse the elongation step resulting in a stereospecific product[70,71]. In addition to the E-domains, there are several other functions that can be incorporated into a module. Common examples include N-methyltransferases as observed in the biosynthesis of cyclosporins[72] or the formation of oxazole and thiazole rings observed in the biosynthesis of pyochelin and vibriobactin, among others[73,74].

Considerable effort has been expended on elucidating the structure of NRPS domains and modules. The structures of all core domains have been determined in a number of structural states,

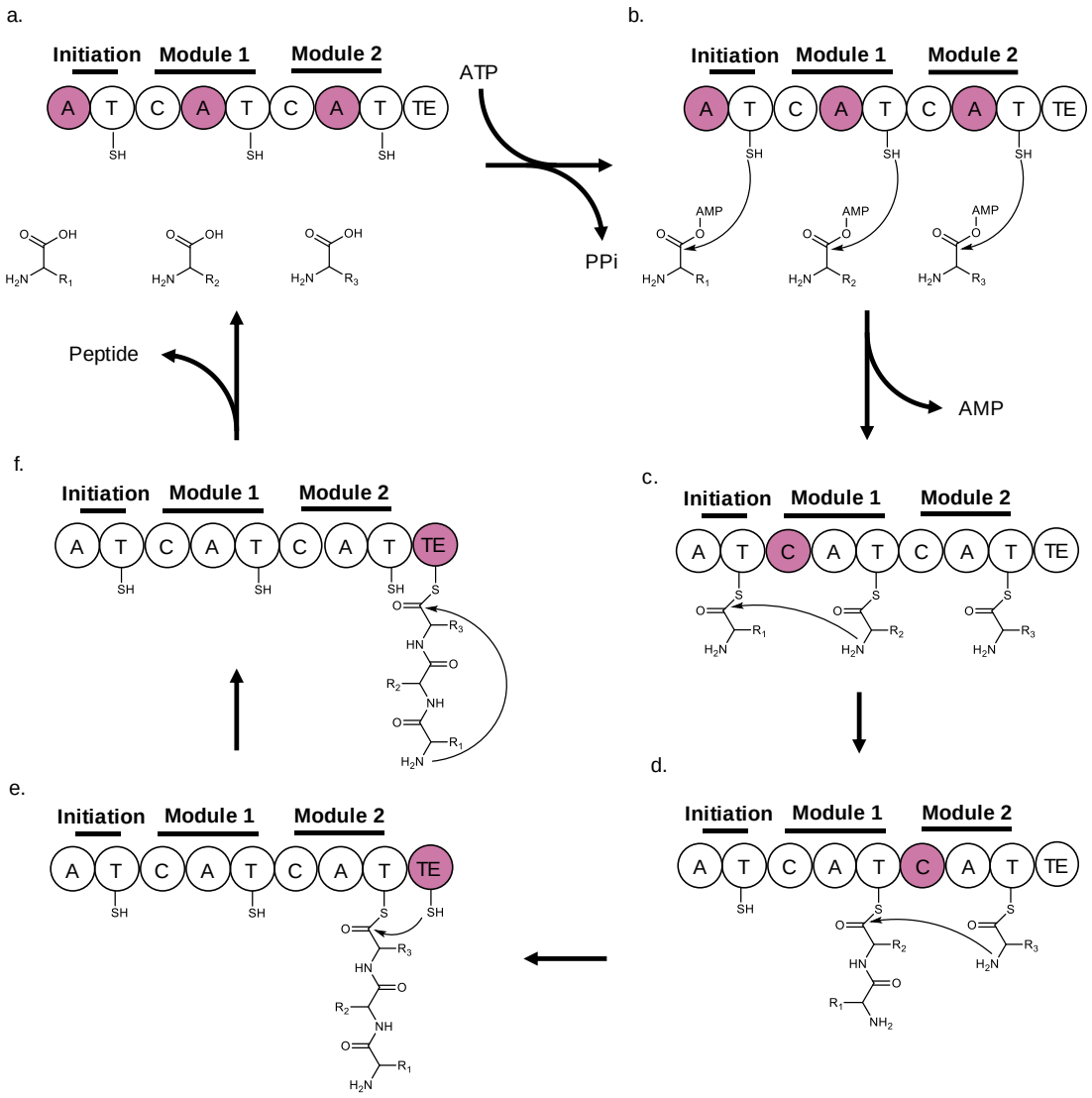


Figure 1. 3: Schematic of the archetypical NRPS cycle. Domains responsible for a catalytic step are highlighted. a.) The adenylation (A)-domains select specific amino acids; b.) the selected amino acids are adenylated and transferred to the thiol group of the thiolation (T)-domain; c. - d.) condensation (C)-domains catalyse the stepwise elongation of the polypeptide chain; e.) transthioesterification; and f.) chain termination and cyclisation by the thioesterase (TE)-domain.

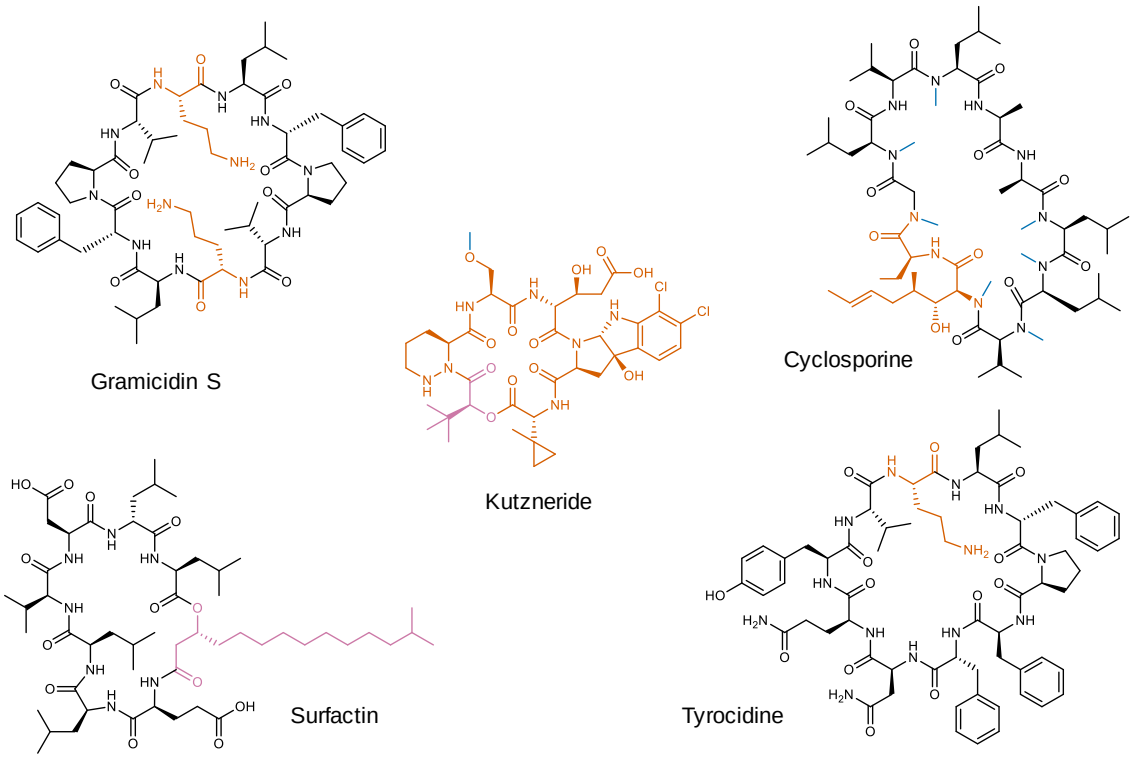


Figure 1. 4: Non-ribosomal peptide diversity. Examples of non-ribosomal peptides discussed in this section. Non-proteinogenic amino acids are coloured red, non-amino acids are coloured purple and methylations installed by modular methyltransferases are coloured blue.

revealing the dynamics of NRPS catalytic cycles. The first domain to be structurally characterised was the A-domain of GrsA (PheA)[75]. The elucidation of this structure led to the identification of the substrate binding pocket and ultimately the description of the substrate specificity code (colloquially referred to as the Stachelhaus code). This comprises ten residues that can be used to predict the substrate specificity of an A-domain with reasonable accuracy[76,77]. Comparison of the PheA structure with that of the previously published standalone adenylation enzyme firefly luciferase revealed the rotation of the C-terminal domain by 94°[75]. Such a rotation was later observed in A-T domain structures confirming the hypothesis that A-domain rotation guides the T-domain into the A-domain active site to enable thioesterformation[78,79]. More recently, full module structure have been elucidated, the first of which was the terminal module of the surfactin synthase SfcA-C[62]. This structure revealed the large interface between the A-domain and the C-domain and that conformation change in the A-domain was sufficient to shuttle the T-domain the three conformations required for biosynthesis. Additionally, structures of other complete modules of EntF and AB3403 confirmed that only three conformations are required[80].

1.2.2 Type I Polyketide Synthases

As mentioned above, the first PKS genes, responsible for the biosynthesis of actinorhodin, were identified by Malpartida and Hopwood in 1984[22]. Actinorhodin is a phenolic aromatic polyketide, synthesised in a recursive fashion by a type II polyketide synthase (T2PKS), so-called due to its observed similarity with type II fatty acid synthases. T2PKSs are comprised of several single domain proteins which incorporate acetate, onto an acyl carrier protein (ACP), which is transferred to the active site of a ketosynthase (KS) and then undergoes iterative elongation by Claisen condensation with malonyl-coenzyme A (CoA) derrivatives[81]. The resulting polyketide is then cyclised to produce polycyclic compounds. It wouldn't be until the early 1990s that it became apparent that polyketides were not necessarily produced by single monofunctional proteins, but also by modular megasynthases reminiscent of the NRPSs[27,28]. Seminal work by the Leadlay and Katz groups on the biosynthesis erythromycin laid the foundation for modern understanding of the type I polyketide synthase (T1PKS)[27,28]. By sequencing outwards from the identified resistance gene, Cortes et al. were able to identify a gene encoding two modules of the erythromycin PKS[28]. It was unclear however, whether this was sufficient for the biosynthesis of the erythromycin aglycone or whether additional genes were required. The following year, Donadio et al. described three open reading frames (ORFs) required for erythromycin aglycone biosynthesis[82]. These ORFs encoded six repeating units (modules) in addition to a loading and termination module acting in a 'colinear' fashion (i.e. the order of the genes reflects the order of the biosynthesis). This represents the first depiction of a T1PKS and the model was further clarified in the following years[83,84]. The discovery of the erythromycin

synthase encouraged the discovery of many more T1PKS BCGs, for example the industrially important rapamycin and avermectin synthases[85,86].

Akin to the NRPSs, the typical T1PKS module consist of a minimum of three, core domains: the ketosynthase (KS)-domain, the acyltransferase (AT)-domain and the acyl-carrier protein (ACP). The AT-domain is responsible for the selection of acyl-coenzyme A (CoA) substrates. The substrates of AT-domains vary, ATs in the loading module typically accommodate acetyl-CoA, while the ATs in subsequent modules typically activate malony- or methyl malonyl-CoA, however a diverse range of starter and extender units have been identified (see Zwittermycin, Figure 1. 5)[87]. Subsequent, acyl-transfer to the phosphopantetheine of the ACP primes the PKS. The upstream KS-domain is then transacylated with the growing polyketide chain, enabling malonyl-ACP is condensed with the upstream acyl-KS substrate through Claisen condensation and the liberation of CO_2 (Figure 1. 6). In this fashion, each module of a PKS can incorporate a C2 acetyl unit (Figure 1. 6). In most cases, the newly formed β-keto group will undergo some level of reduction. In type I fatty acid biosynthesis this is always to an olefin, however in T1PK biosynthesis reduction is variable allowing greater structural diversity of the polyketide chain. Ketoreductase (KR) domains catalyse the reduction of the β-keto group to a hydroxyl. Given that the hydroxyl product is chiral, KR-domains play an important role in the introduction of stereocentres into polyketides. The ketoreduction mechanism proceeds *via* an enol intermediate that allows the stereochemistry to be controlled at both the C2 and C3 positions (Figure 1. 7). KR-domains are categorised according to the stereochemistry of their products, A or B type for R- or S-hydroxyl products and 1 or 2 for S- or R- configuration of the R-group, hence four categories: A1 (*2R, 3S*), B1, (*2S, 3S*), A2 (*2R, 3R*) and B2 (*2S, 3R*)[88–91]. Dehydratase (DH) domains dehydrate the β-hydroxyl group to produce a *cis*- or *trans*- double bond. In the majority of cases, the stereochemistry of the hydroxyl group is predictive of the configuration of the double bond, with *R*-hydroxyls being reduced to *trans*-double bonds and *S*-hydroxyls reduced to *cis*-double bonds[90,91]. The resulting olefin may then be reduced once more by the action of an enoylreductase (ER) domain to produce a single bond. Again, this presents another opportunity to control the chirality of the C2 R-group if present[92,93]. During type I fatty acid biosynthesis, the polyketide is always reduced to a fully saturated alkyl chain, however in the case of T1PK biosynthesis, the presence of these domains is variable and as such so is the product. Although less common, tailoring domains such as C-methyl transferases have been identified[94]. Following polyketide biosynthesis, chain termination is catalysed by the action of a thioesterase (Figure 1. 6). There are also *trans*-AT PKSs that encode their AT domains separately from the modular assembly-line. This family of PKSs can incorperate a number of interesting features including the formation of carbon branches at the β-position and often deviate from the coliearity rule[95].

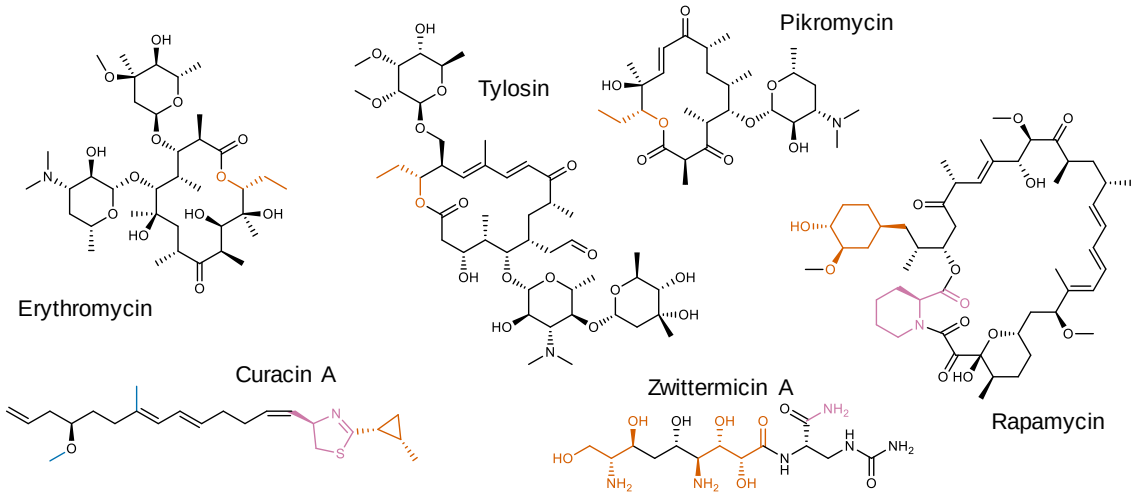


Figure 1. 5: Diversity of type I polyketide natural products. Starter units and unusual extender units are highlighted in red, amino acid derived moieties in purple and modifications of modular methyltransferases in blue.

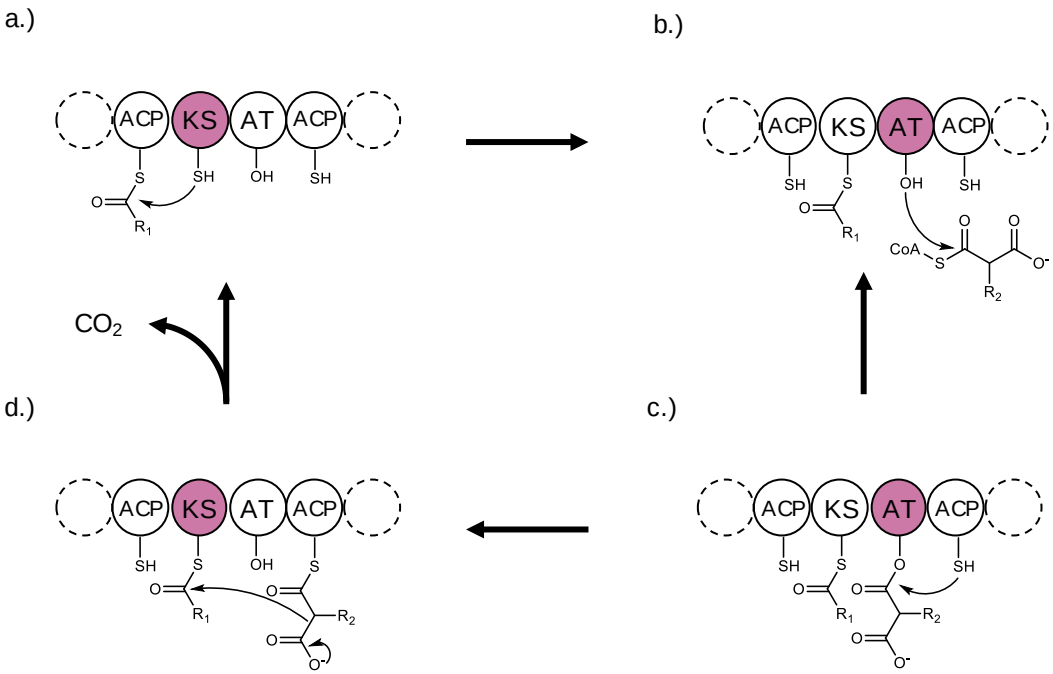


Figure 1. 6: Catalytic cycle of a type I polyketide synthase (T1PKS). Domains responsible for catalytic activity highlighted in purple. a.) Acyl carrier protein (ACP) bound growing chain is transacylated to the ketosynthase (KS)-domain; b.- c.) acyltransferase (AT)-domain selects malonyl-coenzyme A (CoA) substrate, which is transferred to the downstream ACP; d.) KS-domain catalyses the decarboxylative condensation of the two covalently bound substrates.

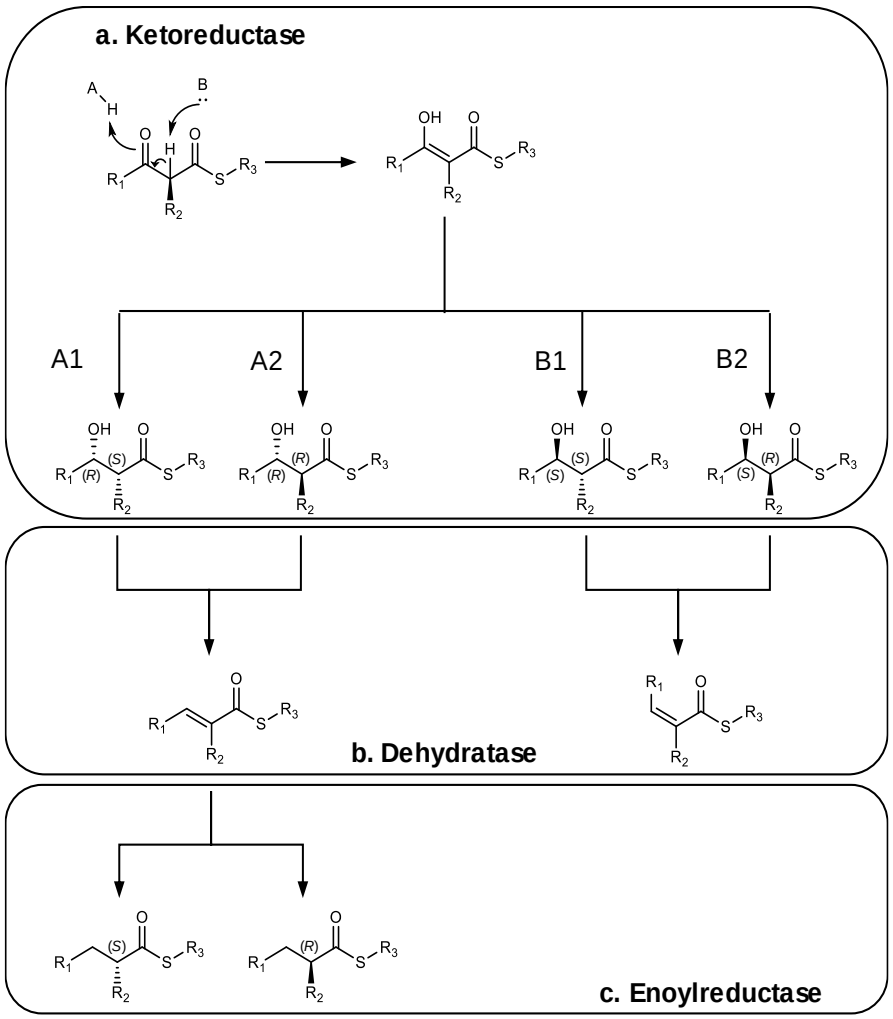


Figure 1. 7: Sequential reduction of β-keto group by ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) domains. a. Reduction of the β-keto group by the KR-domain can result in four different stereo-configurations; b. DH domains dehydrate the β-carbon to produce cis- or trans- double bonds depending on the stereochemistry of the β-hydroxyl group; c.) ER domains are capable of reducing trans- double bonds to single bonds.

Structures of all the main domains of the PKS have been solved[96–101]. Perhaps the most significant structure however is the cryo-electron microscopy model of the pikromycin biosynthetic module PikAIII[102]. This model showed that, unlike the NRPSs, which are monomeric, T1PKSs are homodimeric. The KS, AT, KR and ACP domains forming a single chamber and each KS domain has two openings for both the upstream and downstream ACPs. It is hypothesised that this architecture allows for the ACP bound substrate to access each domain while excluding other ACPs in the assembly-line.

1.2.3 Assembly-Line Engineering

Following the cloning of the actinorhodin BGC, Malpartida and Hopwood[22] state that: 'the application of molecular cloning to antibiotic-producing microorganisms should lead to enhanced antibiotic productivity and to the biosynthesis of novel antibiotics.' The following year, Hopwood et al. published the first hybrid antibiotics - merderrhodin A and dihydrogranaticin[23]. The advances in the understanding of assembly line biosynthesis described above have inspired efforts to engineer assembly lines to enhance production or generate novel metabolites. Attempts to engineer assembly-lines vary in scope, from the substitution of amino acids to *de novo* assembly.

The smallest modifications are the mutation of individual residues. Mutation of A-domain substrate binding pockets have been used to alter the substrate specificity of NRPSs[103,104]. For example, mutating tryptophan to serine in the binding pocket of GrsA was sufficient to change the substrate specificity from L-phenylalanine to L-tyrosine[104]. Furthermore, this modified module could also accept O-propargyl tyrosine when fed into the fermentation media. Specific mutation of the substrate binding pocket has also been used to fix the activity of a promiscuous A-domain to a preferred substrate[103,105]. Analogously in PKS systems, point mutations in the AT-domains are sufficient to alter substrate specificity. The combination of point mutation and mutasynthesis can result in the generation of novel compounds. Large mutant libraries can be screened for the incorporation of the monomer of interest[106,107]. Point mutation has also been used as a strategy to alter the function of KR-domains. In the erythromycin ketoreductase domain EryK1, for example it was shown that only two point mutations were required to reverse the stereoselectivity of the domain[108].

It has also been shown that swapping individual sub-domains, domains and even whole modules can be used to reprogramme assembly-lines. Perhaps the most straightforward approach is the swaping of assembly-line subunits. Closely related PKSs and NRPSs sometimes share the same genetic architecture while incorporating different monomers. Knocking out a native gene and complimenting it with a homologue from a related yet distinct BGC can lead to the generation of a hybrid assembly-line. This was first achieved by complimenting the genes for modules 5 and 6 of pikromycin biosynthesis with homologous genes from erythromycin and oleandomycin

BGCs[109]. The first example of this type of reprograming in NRPSs was of the daptomycin BGC. The daptomycin core peptide is encoded by three genes: *dptA*, *dptBC* and *dptD*. By knocking out *dptBC* or *dptD* and replacing it with genes from related BGCs with similar architecture (encoding the A54145 and calcium dependent antibiotic assembly-lines) Baltz et al. were able to generate novel lipopeptides[110]. It is also possible to delete, swap and insert intergenic domains and modules. Many attempts to engineer the subunits responsible for the biosynthesis of erythromycin (deoxyerythronolide B synthase, DEBS) have been reported. For example, the disruption of KR and ER domains has enabled the production of an array of erythromycin analogues[111], the entire loading module of DEBS1 was swapped with that from the avermectin PKS[112], and the size of the assembly-line has been extended by the addition of a module from the rapamycin PKS[113]. Combinations of different approaches can enable potentially large libraries of analogues to be created[114]. These strategies however have varying levels of success. The complicated nature of such modifications is exemplified by attempts to reprogramme the NRPS for pyroverdine, a siderophore produced by *Pseudomonas* spp.. The wild-type pyroverdine NRPS incorporates L-threonine in the final module. Attempts to swap the final A-domain of the NRPS showed production comparable to wild-type only when the substrate of the inserted A-domain was also L-threonine. Additionally, mass spectral analysis revealed that wild-type pyroverdines were still produced in strains where the inserted domain was not specific for L-threonine, leading to the conclusion that selectivity at the acceptor site of the C-domain was preventing the incorporation of other amino acids[115]. Based on this hypothesis several C-A domain swaps were engineered which allowed the the incorporation of L-lysine or L-serine in the place of L-threonine. However most of the C-A substitutions failed to produce a mature peptide, presumably due to disruption of the interface between domains. Sub-domain swaps have been suggested as a potential method to overcome the problem of domain interference[116,117]. This is discussed in detail in Chapter 2.

Perhaps the holy grail of synthetic biology is the *de novo* assembly of biosynthetic pathways. Although this goal has yet to be achieved in full, considerable advances have been made in the last few years, particularly with NRPSs. Bozhüyük et al. recently published their method for the *de novo* assembly of NRPSs[118]. By taking into account condensation domain specificities it was possible to build libraries of novel NRPSs from different exchange units (XUs) consisting of A-T-C domains. A recent modification to this strategy allows the specificity of the C-domain to be overcome, significantly improving the utility of the technique. This was achieved by splitting the C-domain between each XU at the flexible linker between the donor and acceptor site of the condensation domain[118].

The examples above all describe attempts to rationally design assembly-lines. Some techniques however seek to take advantage of evolutionary processes through artificial selection to generate diversity. Directed evolution can improve chimeric assembly-lines significantly. Through rounds of PCR-mediated mutagenesis and artificial selection, Fischbach et al. were able

to improve the catalytic activity of a chimeric NRPS by 10-fold[119]. Remarkably, between only two and nine amino acid substitutions were required to restore activity, requiring only three generations to restore activity to wild-type levels. These mutations however did not appear to follow a uniform pattern. Directed evolution has also been used to completely alter the substrate specify of adenylation domains. Through successive rounds of saturation mutation of residues responsible for substrate specificity Villiers and Hollfelder were able to increase the specificity of the TycA adenylation domain for L-alanine by 170-fold while all but eliminating activity for its original substrate L-phenylalanine[120]. Rather than dedicated screening for a compound of interest, it is also possible to apply a much broader approach and select for diverse, novel products. Recombination is a useful tool for the rapid generation of diversity. Homologous recombination in yeast has been utilised to produce large libraries of hybrid PKSs[121]. More recently, so-called 'accelerated evolution' was discovered fortuitously while attempting to rationally design rapamycin analogues through allelic exchange. Instead of the expected products, Wlodek et al generated analogues of rapamycin with varying ring size[122]. They hypothesised that the incorporation of the plasmid pKC1139, used to introduce gene disruptions, caused instability in the genomic DNA leading to increased rates of recombination in the KS, AT and ACP-linker regions[122]. This study is a fascinating example of not only how understanding of evolution processes impacts synthetic biology, but how synthetic biology can expand our understanding of evolution.

1.2.4 Evolution of Chemical Diversity in Assembly-Line Biosynthesis

Since their discovery, the relationship between PKSs NRPSs and fatty acid synthases (FASs) have been speculated[55,82]. In the case of T1PKSs, this relationship was first formalised by Donadio and Katz[83]. By creating a phylogeny of ACP domain sequences, they were able to show that T1PKS shared a common ancestor with vertebrate fatty acid synthases. Later studies of KS-domains validated this relationship[123,124]. Elucidating the evolutionary origins of the NRPSs is a little more complex, given that only the less conserved ACP and TE domains are shared between the PKS, FAS and NRPS.

The current model of assembly-line evolution is based on two main observations. Firstly, phylogenetic studies of biosynthetic domains show that, in the majority of cases, domains are more closely related to homologues within the same BGC than homologues from other BGCs[50,125]. This is most commonly interpreted as the result of multiple duplications of biosynthetic modules during the evolution of the BGC, however, there is also evidence that concerted evolution (the homogenisation of paralogous sequences through homologous recombination) has contributed to this phenomenon[50,126,127]. The second common observation is that coding regions responsible for substrate specificity are particularly amenable to recombination[117,128–130]. Perhaps the most

detailed examination of recombination in the evolution of NRPSs has been carried out on the microcystin producing *myc* BCGs[128–131]. Microcystins are a large group of cyanobacterial heptapeptides with a well-defined cyclic structure[132]. Despite the overall conservation of the peptide structure, the second and fourth residues of the peptide, encoded by *mycB* and *mycC*, are highly variable. The adenylation domains of *mycB* and *mycC* have been identified as hotspots for recombination[131] and there is evidence that multiple, genus-level recombination events within the adenylation domains of *mcyB* and *mcyC* have shaped the substrate variability of microcystin synthetases[129].

1.3 Thesis Outline

In this thesis, I will discuss my investigations into the biosynthesis of several assembly-line natural products. The research was predicated on the following questions:

a.) how is the structural diversity of natural products generated on a biochemical level;

b.) how has evolution generated this diversity; and

c.) can we use knowledge of these evolution processes to aid in the historical challenges of the discovery and engineering of natural product biosynthetic pathways?

Chapter 2 examines the evolution of an unusual bifurcated NRPS. Chapters 3 and 4, explore the structural diversity of β-amino acid containing polyketide macrolactams, first describing the intramolecular cycloadditions of heronamide C, and then continuing to a wider analysis of the entire family of compounds. Chapter 5 describes the activation of two polyketide producing BGCs through the overexpression of positive regulators. Finally, Chapter 6 will discuss the implications of these results on genome mining and the evolution and engineering of biosynthetic assembly-lines.

1.4 References

(1) Waksman, S. A.; Woodruff, H. B. The Soil as a Source of Microorganisms Antagonistic to Disease-Producing Bacteria. *J. Bacteriol.* **1940**, *40* (4), 581–600.

(2) Fleming, A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to Their Use in the Isolation of B. Influenzæ. *Br. J. Exp. Pathol.* **1929**, *10* (3), 226-236.

(3) Chain, E.; Florey, H. W.; Gardner, A. D.; Heatley, N. G.; Jennings, M. A.; Orr-Ewing, J.; Sanders, A. G. Penicillin As A Chemotherapeutic Agent. *Lancet* **1940**, *236* (6104), 226–228.

(4) Dubos, R. J. Studies On A Bactericidal Agent Extracted From A Soil Bacillus : I. Preparation Of The Agent. Its Activity In Vitro. *J. Exp. Med.* **1939**, *70* (1), 1–10.

(5) Dubos, R. J. Studies On A Bactericidal Agent Extracted From A Soil Bacillus : II.

Protective Effect Of The Bactericidal Agent Against Experimental Pneumococcus Infections In Mice. *J. Exp. Med.* **1939**, *70* (1), 11–17.

(6) Hollstein, U. Actinomycin. Chemistry and Mechanism of Action. *Chem. Rev.* **1974**, *74* (6), 625–652.

(7) Waksman, S. A.; Woodruff, H. B. Selective Antibiotic Action of Various Substances of Microbial Origin. *J. Bacteriol.* **1942**, *44* (3), 373–384.

(8) Schatz, A.; Bugle, E.; Waksman, S. A. Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria. *Exp. Biol. Med.* **1944**, *55* (1), 66–69.

(9) Waksman, S. A.; Lechevalier, H. A. Neomycin, a New Antibiotic Active against Streptomycin-Resistant Bacteria, Including Tuberculosis Organisms. *Science.* **1949**, *109* (2830), 305–307.

(10) Umezawa, H.; Ueda, M.; Maeda, K.; Yagishita, K.; Kondo, S.; Okami, Y.; Utahara, R.; Osato, Y.; Nitta, K.; Takeuchi, T. Production and Isolation of a New Antibiotic: Kanamycin. *J. Antibiot. (Tokyo).* **1957**, *10* (5), 181–188.

(11) Duggar, B. M. Aureomycin: A Product Of The Continuing Search For New Antibiotics. *Ann. N. Y. Acad. Sci.* **1948**, *51* (2), 177–181.

(12) Finlay, A. C.; Hobby, G. L. Terramycin, a New Antibiotic. *Science* **1950**, *111* (2874), 85.

(13) Mcguire, J. M.; Bunch, R. L.; Anderson, R. C.; Boaz, H. E.; Flynn, E. H.; Powell, H. M.; Smith, J. W. Ilotycin, a New Antibiotic. *Antibiot. Chemother. (Northfield, Ill.)* **1952**, *2* (6), 281–283.

(14) Johnson, B. A.; Anker, H.; Meleney, F. L. Bacitracin: A New Antibiotic Produced By A Member Of The B. Subtilis Group. *Science.* **1945**, *102* (2650), 376–377.

(15) Mccormick, M. H.; Mcguire, J. M.; Pittenger, G. E.; Pittenger, R. C.; Stark, W. M. Vancomycin, a New Antibiotic. I. Chemical and Biologic Properties. *Antibiot. Annu.* **1955-1956,** *3*, 606–611.

(16) Burton, H. S.; Abraham, E. P. Isolation of Antibiotics from a Species of Cephalosporium; Cephalosporins P1, P2, P3, P4, and P5. *Biochem. J.* **1951**, *50* (2), 168–174.

(17) Abraham, E. P.; Newton, G. G. F.; Crawford, K.; Burton, H. S.; Hale, C. W. Cephalosporin N: A New Type of Penicillin. *Nature* **1953**, *171* (4347), 343.

(18) Waksman, S. A. Antibiotics: The Duplication Problem. *Science* **1965**, *147* (3664), 1396–1397.

(19) Béahdy, J. Recent Developments of Antibiotic Research and Classification of Antibiotics According to Chemical Structure. *Adv. Appl. Microbiol.* **1974**, *18*, 309–406.

(20) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74* (12), 5463–5467.

(21) Saiki, R. K.; Gelfand, D. H.; Stoffel, S.; Scharf, S. J.; Higuchi, R.; Horn, G. T.; Mullis, K.

B.; Erlich, H. A. Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science* **1988**, *239* (4839), 487–491.

(22) Malpartida, F.; Hopwood, D. A. Molecular Cloning of the Whole Biosynthetic Pathway of a Streptomyces Antibiotic and Its Expression in a Heterologous Host. *Nature* **1984** *309* (5967), 462–464.

(23) Hopwood, D. A.; Malpartida, F.; Kieser, H. M.; Ikeda, H.; Duncan, J.; Fujii, I.; Rudd, B. A. M.; Floss, H. G.; Ōmura, S. Production of 'Hybrid' Antibiotics by Genetic Engineering. *Nature* **1985**, *314* (6012), 642–644.

(24) Weckermann, R.; Fürbass, R.; Marahiel, M. A. Complete Nucleotide Sequence of the TycA Gene Coding the Tyrocidine Synthetase 1 from Bacillus brevis. *Nucleic Acids Res.* **1988**, *16* (24), 11841.

(25) Mootz, H. D.; Marahiel, M. A. The Tyrocidine Biosynthesis Operon of Bacillus Brevis: Complete Nucleotide Sequence and Biochemical Characterization of Functional Internal Adenylation Domains. *J. Bacteriol.* **1997**, *179* (21), 6843–6850.

(26) Krause, M.; Marahiel, M. A. Organization of the Biosynthesis Genes for the Peptide Antibiotic Gramicidin S. *J. Bacteriol.* **1988**, *170* (10), 4669–4674.

(27) Tuan, J. S.; Weber, J. M.; Staver, M. J.; Leung, J. O.; Donadio, S.; Katz, L. Cloning of Genes Involved in Erythromycin Biosynthesis from Saccharopolyspora erythraea Using a Novel Actinomycete-Escherichia Coli Cosmid. *Gene* **1990**, *90* (1), 21–29.

(28) 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* **1990**, *348* (6297), 176–178.

(29) Bentley, S. D.; Chater, K. F.; Cerdeño-Tárraga, A.-M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; et al. Complete Genome Sequence of the Model Actinomycete Streptomyces coelicolor A3(2). *Nature* **2002**, *417* (6885), 141–147.

(30) Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. PCR-Targeted Streptomyces Gene Replacement Identifies a Protein Domain Needed for Biosynthesis of the Sesquiterpene Soil Odor Geosmin. *Proc. Natl. Acad. Sci.* **2003**, *100* (4), 1541–1546.

(31) Barona-Gómez, F.; Wong, U.; Giannakopulos, A. E.; Derrick, P. J.; Challis, G. L. Identification of a Cluster of Genes That Directs Desferrioxamine Biosynthesis in *Streptomyces coelicolor* M145. *J. Am. Chem. Soc.* **2004**, *126* (50), 16282–16283.

(32) Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Ōmura, S. Complete Genome Sequence and Comparative Analysis of the Industrial Microorganism Streptomyces Avermitilis. *Nat. Biotechnol.* **2003**, *21* (5), 526–531.

(33) Ohnishi, Y.; Ishikawa, J.; Hara, H.; Suzuki, H.; Ikenoya, M.; Ikeda, H.; Yamashita, A.;

Hattori, M.; Horinouchi, S. Genome Sequence of the Streptomycin-Producing Microorganism Streptomyces griseus IFO 13350. *J. Bacteriol.* **2008**, *190* (11), 4050–4060.

(34) Oliynyk, M.; Stark, C. B. W.; Bhatt, A.; Jones, M. A.; Hughes-Thomas, Z. A.; Wilkinson, C.; Oliynyk, Z.; Demydchuk, Y.; Staunton, J.; Leadlay, P. F. Analysis of the Biosynthetic Gene Cluster for the Polyether Antibiotic Monensin in Streptomyces Cinnamonensis and Evidence for the Role of MonB and MonC Genes in Oxidative Cyclization. *Mol. Microbiol.* **2003**, *49* (5), 1179–1190.

(35) Udwary, D. W.; Zeigler, L.; Asolkar, R. N.; Singan, V.; Lapidus, A.; Fenical, W.; Jensen, P. R.; Moore, B. S. Genome Sequencing Reveals Complex Secondary Metabolome in the Marine Actinomycete Salinispora Tropica. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (25), 10376–10381.

(36) Challis, G. L.; Ravel, J. Coelichelin, a New Peptide Siderophore Encoded by the Streptomyces Coelicolor Genome: Structure Prediction from the Sequence of Its Non-Ribosomal Peptide Synthetase. *FEMS Microbiol. Lett.* **2000**, *187* (2), 111–114.

(37) Wilkinson, B.; Micklefield, J. Mining and Engineering Natural-Product Biosynthetic Pathways. *Nat. Chem. Biol.* **2007**, *3* (7), 379–386.

(38) Margulies, M.; Egholm, M.; Altman, W. E.; Attiya, S.; Bader, J. S.; Bemben, L. A.; Berka, J.; Braverman, M. S.; Chen, Y.-J.; Chen, Z.; et al. Genome Sequencing in Microfabricated High-Density Picolitre Reactors. *Nature* **2005**, *437* (7057), 376–380.

(39) Fedurco, M.; Romieu, A.; Williams, S.; Lawrence, I.; Turcatti, G. BTA, a Novel Reagent for DNA Attachment on Glass and Efficient Generation of Solid-Phase Amplified DNA Colonies. *Nucleic Acids Res.* **2006**, *34* (3), e22.

(40) Bentley, D. R.; Balasubramanian, S.; Swerdlow, H. P.; Smith, G. P.; Milton, J.; Brown, C. G.; Hall, K. P.; Evers, D. J.; Barnes, C. L.; Bignell, H. R.; et al. Accurate Whole Human Genome Sequencing Using Reversible Terminator Chemistry. *Nature* **2008**, *456* (7218), 53–59.

(41) Rothberg, J. M.; Hinz, W.; Rearick, T. M.; Schultz, J.; Mileski, W.; Davey, M.; Leamon, J. H.; Johnson, K.; Milgrew, M. J.; Edwards, M.; et al. An Integrated Semiconductor Device Enabling Non-Optical Genome Sequencing. *Nature* **2011**, *475* (7356), 348–352.

(42) Wetterstrand K A. DNA Sequencing Costs: Data - National Human Genome Research Institute (NHGRI) https://www.genome.gov/27541954/dna-sequencing-costs-data/ (accessed Sep 12, 2018).

(43) Heather, J. M.; Chain, B. The Sequence of Sequencers: The History of Sequencing DNA. *Genomics* **2016**, *107* (1), 1–8.

(44) NCBI. Genbank Prokaryotic Genome Database https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/ (accessed Sep 22, 2018).

(45) Doroghazi, J. R.; Albright, J. C.; Goering, A. W.; Ju, K.-S.; Haines, R. R.; Tchalukov, K. A.; Labeda, D. P.; Kelleher, N. L.; Metcalf, W. W. A Roadmap for Natural Product Discovery Based on Large-Scale Genomics and Metabolomics. *Nat. Chem. Biol.* **2014**, *10* (11), 963–968.

(46) Cimermancic, P.; Medema, M. H.; Claesen, J.; Kurita, K.; Wieland Brown, L. C.; Mavrommatis, K.; Pati, A.; Godfrey, P. A.; Koehrsen, M.; Clardy, J.; et al. Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. *Cell* **2014**, *158* (2), 412–421.

(47) Freeman, M. F.; Gurgui, C.; Helf, M. J.; Morinaka, B. I.; Uria, A. R.; Oldham, N. J.; Sahl, H.-G.; Matsunaga, S.; Piel, J. Metagenome Mining Reveals Polytheonamides as Posttranslationally Modified Ribosomal Peptides. *Science* **2012**, *338* (6105), 387–390.

(48) Wilson, M. C.; Mori, T.; Rückert, C.; Uria, A. R.; Helf, M. J.; Takada, K.; Gernert, C.; Steffens, U. A. E.; Heycke, N.; Schmitt, S.; et al. An Environmental Bacterial Taxon with a Large and Distinct Metabolic Repertoire. *Nature* **2014**, *506* (7486), 58–62.

(49) Owen, J. G.; Charlop-Powers, Z.; Smith, A. G.; Ternei, M. A.; Calle, P. Y.; Reddy, B. V. B.; Montiel, D.; Brady, S. F. Multiplexed Metagenome Mining Using Short DNA Sequence Tags Facilitates Targeted Discovery of Epoxyketone Proteasome Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (14), 4221–4226.

(50) Medema, M. H.; Cimermancic, P.; Sali, A.; Takano, E.; Fischbach, M. A. A Systematic Computational Analysis of Biosynthetic Gene Cluster Evolution: Lessons for Engineering Biosynthesis. *PLoS Comput. Biol.* **2014**, *10* (12), e1004016.

(51) McDonald, B. R.; Currie, C. R. Lateral Gene Transfer Dynamics in the Ancient Bacterial Genus *Streptomyces*. *MBio* **2017**, *8* (3), e00644-17.

(52) Nouioui, I.; Carro, L.; García-López, M.; Meier-Kolthoff, J. P.; Woyke, T.; Kyrpides, N. C.; Pukall, R.; Klenk, H.-P.; Goodfellow, M.; Göker, M. Genome-Based Taxonomic Classification of the Phylum Actinobacteria. *Front. Microbiol.* **2018**, *9*, 2007.

(53) Mach, B.; Reich, E.; Tatum, E. L. Separation Of The Biosynthesis Of The Antibiotic Polypeptide Tyrocidine From Protein Biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **1963**, *50* (1), 175–181.

(54) Berg, T. L.; Froholm, L. O.; Laland, S. G. The Biosynthesis Of Gramicidin S In A Cell-Free System. *Biochem. J.* **1965**, *96*, 43–52.

(55) Lipmann, F. Attempts to Map a Process Evolution of Peptide Biosynthesis. *Science* **1971**, *173* (4000), 875–884.

(56) Laland, S. G.; Zimmer, T. L. The Protein Thiotemplate Mechanism of Synthesis for the Peptide Antibiotics Produced by Bacillus brevis. *Essays Biochem.* **1973**, *9*, 31–57.

(57) Kurahashi, K. Biosynthesis of Small Peptides. *Annu. Rev. Biochem.* **1974**, *43* (1), 445–459.

(58) Krätzschmar, J.; Krause, M.; Marahiel, M. A. Gramicidin S Biosynthesis Operon Containing the Structural Genes GrsA and GrsB Has an Open Reading Frame Encoding a Protein Homologous to Fatty Acid Thioesterases. *J. Bacteriol.* **1989**, *171* (10), 5422–5429.

(59) Rusnak, F.; Sakaitani, M.; Drueckhammer, D.; Reichert, J.; Walsh, C. T. Biosynthesis of the Escherichia Coli Siderophore Enterobactin: Sequence of the EntF Gene, Expression and Purification of EntF, and Analysis of Covalent Phosphopantetheine. *Biochemistry* **1991**, *30* (11), 2916–2927.

(60) Nakano, M. M.; Magnuson, R.; Myers, A.; Curry, J.; Grossman, A. D.; Zuber, P. SrfA Is an Operon Required for Surfactin Production, Competence Development, and Efficient Sporulation in Bacillus Subtilis. *J. Bacteriol.* **1991**, *173* (5), 1770–1778.

(61) Stein, T.; Vater, J.; Kruft, V.; Otto, A.; Wittmann-Liebold, B.; Franke, P.; Panico, M.; McDowell, R.; Morris, H. R. The Multiple Carrier Model of Nonribosomal Peptide Biosynthesis at Modular Multienzymatic Templates. *J. Biol. Chem.* **1996**, *271* (26), 15428–15435.

(62) Tanovic, A.; Samel, S. A.; Essen, L.-O.; Marahiel, M. A. Crystal Structure of the Termination Module of a Nonribosomal Peptide Synthetase. *Science.* **2008**, *321* (5889), 659–663.

(63) Reimer, J. M.; Haque, A. S.; Tarry, M. J.; Schmeing, T. M. Piecing Together Nonribosomal Peptide Synthesis. *Curr. Opin. Struct. Biol.* **2018**, *49*, 104–113.

(64) Caboche, S.; Leclère, V.; Pupin, M.; Kucherov, G.; Jacques, P. Diversity of Monomers in Nonribosomal Peptides: Towards the Prediction of Origin and Biological Activity. *J. Bacteriol.* **2010**, *192* (19), 5143–5150.

(65) Fujimori, D. G.; Hrvatin, S.; Neumann, C. S.; Strieker, M.; Marahiel, M. A.; Walsh, C. T. Cloning and Characterization of the Biosynthetic Gene Cluster for Kutznerides. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (42), 16498–16503.

(66) Kudo, F.; Miyanaga, A.; Eguchi, T. Biosynthesis of Natural Products Containing β-Amino Acids. *Nat. Prod. Rep.* **2014**, *31* (8), 1056–1073.

(67) Keating, T. A.; Ehmann, D. E.; Kohli, R. M.; Marshall, C. G.; Trauger, J. W.; Walsh, C. T. *Chain Termination Steps in Nonribosomal Peptide Synthetase Assembly Lines: Directed Acyl-S-Enzyme Breakdown in Antibiotic and Siderophore Biosynthesis*; 2001; Vol. 2.

(68) Frueh, D. P.; Arthanari, H.; Koglin, A.; Vosburg, D. A.; Bennett, A. E.; Walsh, C. T.; Wagner, G. Dynamic Thiolation–thioesterase Structure of a Non-Ribosomal Peptide Synthetase. *Nature* **2008**, *454* (7206), 903–906.

(69) Gao, X.; Haynes, S. W.; Ames, B. D.; Wang, P.; Vien, L. P.; Walsh, C. T.; Tang, Y. Cyclization of Fungal Nonribosomal Peptides by a Terminal Condensation-like Domain. *Nat. Chem. Biol.* **2012**, *8* (10), 823–830.

(70) Clugston, S. L.; Sieber, S. A.; Marahiel, M. A. and; Walsh, C. T. Chirality of Peptide Bond-Forming Condensation Domains in Nonribosomal Peptide Synthetases: The C5 Domain of Tyrocidine Synthetase Is a DCL Catalyst. *Biochemistry* **2003** *42* (41) 12095-12104.

(71) Rausch, C.; Hoof, I.; Weber, T.; Wohlleben, W.; Huson, D. H. Phylogenetic Analysis of Condensation Domains in NRPS Sheds Light on Their Functional Evolution. *BMC Evol. Biol.* **2007**, *7* (1), 78.

(72) Velkov, T.; Horne, J.; Scanlon, M. J.; Capuano, B.; Yuriev, E.; Lawen, A. Characterization of the N-Methyltransferase Activities of the Multifunctional Polypeptide Cyclosporin Synthetase. *Chem. Biol.* **2011**, *18* (4), 464–475.

(73) Quadri, L. E.; Keating, T. A.; Patel, H. M.; Walsh, C. T. Assembly of the Pseudomonas Aeruginosa Nonribosomal Peptide Siderophore Pyochelin: In Vitro Reconstitution of Aryl-4, 2-Bisthiazoline Synthetase Activity from PchD, PchE, and PchF. *Biochemistry* **1999**, *38* (45), 14941–14954.

(74) Keating, T. A.; Marshall, C. G.; Walsh, C. T. Reconstitution and Characterization of the Vibrio Cholerae Vibriobactin Synthetase from VibB, VibE, VibF, and VibH. *Biochemistry* **2000**, *39* (50), 15522–15530.

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

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

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

(78) Mitchell, C. A.; Shi, C.; Aldrich, C. C.; Gulick, A. M. Structure of PA1221, a Nonribosomal Peptide Synthetase Containing Adenylation and Peptidyl Carrier Protein Domains. *Biochemistry* **2012**, *51* (15), 3252–3263.

(79) Sundlov, J. A.; Shi, C.; Wilson, D. J.; Aldrich, C. C.; Gulick, A. M. Structural and Functional Investigation of the Intermolecular Interaction between NRPS Adenylation and Carrier Protein Domains. *Chem. Biol.* **2012**, *19* (2), 188–198.

(80) Drake, E. J.; Miller, B. R.; Shi, C.; Tarrasch, J. T.; Sundlov, J. A.; Leigh Allen, C.; Skiniotis, G.; Aldrich, C. C.; Gulick, A. M. Structures of Two Distinct Conformations of Holo-Non-Ribosomal Peptide Synthetases. *Nature* **2016**, *529* (7585), 235–238.

(81) Das, A.; Khosla, C. Biosynthesis of Aromatic Polyketides in Bacteria. *Acc. Chem. Res.* **2009**, *42* (5), 631–639.

(82) Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. Modular Organization of Genes Required for Complex Polyketide Biosynthesis. *Science* **1991**, *252* (5006), 675–679.

(83) Donadio, S.; Katz, L. Organization of the Enzymatic Domains in the Multifunctional Polyketide Synthase Involved in Erythromycin Formation in Saccharopolyspora Erythraea. *Gene* **1992**, *111* (1), 51–60.

(84) Bevitt, D. J.; Cortes, J.; Haydock, S. F.; Leadlay, P. F. 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.* **1992**, *204* (1), 39–49.

(85) Macneil, D. J.; Occi, J. L.; Gewain, K. M.; Macneil, T. Correlation of the Avermectin Polyketide Synthase Genes to the Avermectin Structure: Implications for Designing Novel Avermectins. *Ann. N. Y. Acad. Sci.* **1994**, *721*, 123–132.

(86) Schwecke, T.; Aparicio, J. F.; Molnár, I.; König, A.; Khaw, L. E.; Haydock, S. F.; Oliynyk, M.; Caffrey, P.; Cortés, J.; Lester, J. B. The Biosynthetic Gene Cluster for the Polyketide Immunosuppressant Rapamycin. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92* (17), 7839–7843.

(87) Ray, L.; Moore, B. S. Recent Advances in the Biosynthesis of Unusual Polyketide Synthase Substrates. *Nat. Prod. Rep.* **2016**, *33* (2), 150–161.

(88) Caffrey, P. Conserved Amino Acid Residues Correlating with Ketoreductase Stereospecificity in Modular Polyketide Synthases. *Chembiochem* **2003**, *4* (7), 654–657.

(89) Reid, R.; Piagentini, M.; Rodriguez, E.; Ashley, G.; Viswanathan, N.; Carney, J.; Santi, D. V.; Hutchinson, C. R. and; McDaniel, R. A. Model of Structure and Catalysis for Ketoreductase Domains in Modular Polyketide Synthases. *Biochemistry* **2003** *42* (1) 72-79.

(90) Caffrey, P. The Stereochemistry of Ketoreduction. *Chem. Biol.* **2005**, *12* (10), 1060–1062.

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

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

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

(94) Skiba, M. A.; Sikkema, A. P.; Fiers, W. D.; Gerwick, W. H.; Sherman, D. H.; Aldrich, C. C.; Smith, J. L. Domain Organization and Active Site Architecture of a Polyketide Synthase C-Methyltransferase. *ACS Chem. Biol.* **2016**, *11* (12), 3319–3327.

(95) Piel, J. Biosynthesis of Polyketides by Trans-AT Polyketide Synthases. *Nat. Prod. Rep.*

2010, *27* (7), 996-1047.

(96) 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-A Resolution. Crystal Structure of a Fatty Acid Synthase Component. *J. Biol. Chem.* **1995**, *270* (22), 12961–12964.

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

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

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

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

(101) Ames, B. D.; Nguyen, C.; Bruegger, J.; Smith, P.; Xu, W.; Ma, S.; Wong, E.; Wong, S.; Xie, X.; Li, J. W.-H.; et al. Crystal Structure and Biochemical Studies of the Trans-Acting Polyketide Enoyl Reductase LovC from Lovastatin Biosynthesis. *Proc. Natl. Acad. Sci.* **2012**, *109* (28), 11144–11149.

(102) Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R. H.; Håkansson, K.; Sherman, D. H.; Smith, J. L.; et al. Structure of a Modular Polyketide Synthase. *Nature* **2014**, *510* (7506), 512–517.

(103) Han, J. W.; Kim, E. Y.; Lee, J. M.; Kim, Y. S.; Bang, E.; Kim, B. S. Site-Directed Modification of the Adenylation Domain of the Fusaricidin Nonribosomal Peptide Synthetase for Enhanced Production of Fusaricidin Analogs. *Biotechnol. Lett.* **2012**, *34* (7), 1327–1334.

(104) Kries, H.; Wachtel, R.; Pabst, A.; Wanner, B.; Niquille, D.; Hilvert, D. Reprogramming Nonribosomal Peptide Synthetases for Clickable Amino Acids. *Angew. Chem. Int. Ed. Engl.* **2014**, *53* (38), 10105–10108.

(105) Bian, X.; Plaza, A.; Yan, F.; Zhang, Y.; Müller, R. Rational and Efficient Site-Directed Mutagenesis of Adenylation Domain Alters Relative Yields of Luminmide Derivatives in Vivo. *Biotechnol. Bioeng.* **2015**, *112* (7), 1343–1353.

(106) Sundermann, U.; Bravo-Rodriguez, K.; Klopries, S.; Kushnir, S.; Gomez, H.; Sanchez-Garcia, E.; Schulz, F. Enzyme-Directed Mutasynthesis: A Combined Experimental and Theoretical Approach to Substrate Recognition of a Polyketide Synthase. *ACS Chem. Biol.* **2013**, *8* (2), 443–450.

(107) Kushnir, S.; Sundermann, U.; Yahiaoui, S.; Brockmeyer, A.; Janning, P.; Schulz, F. Minimally Invasive Mutagenesis Gives Rise to a Biosynthetic Polyketide Library. *Angew. Chemie Int. Ed.* **2012**, *51* (42), 10664–10669.

(108) Bailey, C. B.; Pasman, M. E.; Keatinge-Clay, A. T. Substrate Structure–activity Relationships Guide Rational Engineering of Modular Polyketide Synthase Ketoreductases. *Chem. Commun.* **2016**, *52* (4), 792–795.

(109) Tang, L.; Fu, H.; McDaniel, R. Formation of Functional Heterologous Complexes Using Subunits from the Picromycin, Erythromycin and Oleandomycin Polyketide Synthases. *Chem. Biol.* **2000**, *7* (2), 77–84.

(110) Baltz, R. H.; Brian, P.; Miao, V.; Wrigley, S. K. Combinatorial Biosynthesis of Lipopeptide Antibiotics in Streptomyces Roseosporus. *J. Ind. Microbiol. Biotechnol.* **2006**, *33* (2), 66–74.

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

(112) Marsden, A. F.; Wilkinson, B.; Cortés, J.; Dunster, N. J.; Staunton, J.; Leadlay, P. F. Engineering Broader Specificity into an Antibiotic-Producing Polyketide Synthase. *Science* **1998**, *279* (5348), 199–202.

(113) Rowe, C. J.; Böhm, I. U.; Thomas, I. P.; Wilkinson, B.; Rudd, B. A.; Foster, G.; Blackaby, A. P.; Sidebottom, P. J.; Roddis, Y.; Buss, A. D.; et al. Engineering a Polyketide with a Longer Chain by Insertion of an Extra Module into the Erythromycin-Producing Polyketide Synthase. *Chem. Biol.* **2001**, *8* (5), 475–485.

(114) McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Ashley, G. Multiple Genetic Modifications of the Erythromycin Polyketide Synthase to Produce a Library of Novel "Unnatural" Natural Products. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (5), 1846–1851.

(115) Calcott, M. J.; Owen, J. G.; Lamont, I. L.; Ackerley, D. F. Biosynthesis of Novel Pyoverdines by Domain Substitution in a Nonribosomal Peptide Synthetase of Pseudomonas Aeruginosa. *Appl. Environ. Microbiol.* **2014**, *80* (18), 5723–5731.

(116) Kries, H.; Niquille, D. L.; Hilvert, D. A Subdomain Swap Strategy for Reengineering Nonribosomal Peptides. *Chem. Biol.* **2015**, *22* (5), 640–648.

(117) Crüsemann, M.; Kohlhaas, C.; Piel, J. Evolution-Guided Engineering of Nonribosomal Peptide Synthetase Adenylation Domains. *Chem. Sci.* **2013**, *4* (3), 1041–1045.

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

(119) Fischbach, M. A.; Lai, J. R.; Roche, E. D.; Walsh, C. T.; Liu, D. R. Directed Evolution

Can Rapidly Improve the Activity of Chimeric Assembly-Line Enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (29), 11951–11956.

(120) Villiers, B.; Hollfelder, F. Directed Evolution of a Gatekeeper Domain in Nonribosomal Peptide Synthesis. *Chem. Biol.* **2011**, *18* (10), 1290–1299.

(121) Chemler, J. A.; Tripathi, A.; Hansen, D. A.; O'Neil-Johnson, M.; Williams, R. B.; Starks, C.; Park, S. R.; Sherman, D. H. Evolution of Efficient Modular Polyketide Synthases by Homologous Recombination. *J. Am. Chem. Soc.* **2015**, *137* (33), 10603–10609.

(122) Wlodek, A.; Kendrew, S. G.; Coates, N. J.; Hold, A.; Pogwizd, J.; Rudder, S.; Sheehan, L. S.; Higginbotham, S. J.; Stanley-Smith, A. E.; Warneck, T.; et al. Diversity Oriented Biosynthesis via Accelerated Evolution of Modular Gene Clusters. *Nat. Commun.* **2017**, *8* (1), 1206.

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

(124) Jenke-Kodama, H.; Dittmann, E. Evolution of Metabolic Diversity: Insights from Microbial Polyketide Synthases. *Phytochemistry* **2009**, *70* (15–16), 1858–1866.

(125) Jenke-Kodama, H.; Börner, T.; Dittmann, E. Natural Biocombinatorics in the Polyketide Synthase Genes of the Actinobacterium Streptomyces Avermitilis. *PLoS Comput. Biol.* **2006**, *2* (10), e132.

(126) Liao, D. Concerted Evolution: Molecular Mechanism and Biological Implications. *Am. J. Hum. Genet.* **1999**, *64* (1), 24–30.

(127) Santoyo, G.; Romero, D. Gene Conversion and Concerted Evolution in Bacterial Genomes. *FEMS Microbiol. Rev.* **2005**, *29* (2), 169–183.

(128) Fewer, D. P.; Rouhiainen, L.; Jokela, J.; Wahlsten, M.; Laakso, K.; Wang, H.; Sivonen, K. Recurrent Adenylation Domain Replacement in the Microcystin Synthetase Gene Cluster. *BMC Evol. Biol.* **2007**, *7* (1), 183.

(129) Tooming-Klunderud, A.; Fewer, D. P.; Rohrlack, T.; Jokela, J.; Rouhiainen, L.; Sivonen, K.; Kristensen, T.; Jakobsen, K. S. Evidence for Positive Selection Acting on Microcystin Synthetase Adenylation Domains in Three Cyanobacterial Genera. *BMC Evol. Biol.* **2008**, *8* (1), 256.

(130) Shishido, T.; Kaasalainen, U.; Fewer, D. P.; Rouhiainen, L.; Jokela, J.; Wahlsten, M.; Fiore, M.; Yunes, J.; Rikkinen, J.; Sivonen, K. Convergent Evolution of [D-Leucine1] Microcystin-LR in Taxonomically Disparate Cyanobacteria. *BMC Evol. Biol.* **2013**, *13* (1), 86.

(131) Tanabe, Y.; Kaya, K.; Watanabe, M. M. Evidence for Recombination in the Microcystin Synthetase (Mcy) Genes OfToxic Cyanobacteria Microcystis Spp. *J. Mol. Evol.* **2004**, *58* (6), 633–641.

(132) Dittmann, E.; Fewer, D. P.; Neilan, B. A. Cyanobacterial Toxins: Biosynthetic Routes and

Evolutionary Roots. *FEMS Microbiol. Rev.* **2013**, *37* (1), 23–43.

Chapter Two: Bifurcated Biosynthesis and Evolution of Non-Ribosomal Hexapeptide Antibiotics

Contents

2. 1. Introduction

The previous Chapter discusses the concept of assembly-line evolution in terms of duplication and refunctionalisation[1,2]. This hypothesis is built upon indirect observation, i.e. sequence data, as intermediate BGCs have not been identified. Due to their large effective population sizes and short generation times, bacterial genomes evolve rapidly and consequently mutations that provide a selective advantage will quickly reach fixation in a given population, while mutations that reduce fitness will be lost[3–5]. The upshot of this for the evolutionary biologist is that transitional BGCs (missing links) are fleeting and therefore extremely rare. It has long been hypothesised that the promiscuous activities of enzymes plays an important role in their evolution[6,7]. Therefore, we hypothesised that strains capable of producing multiple related yet distinct compounds can provide insight into evolutionary processes. One such strain is *Streptomyces sp.* MST-110588, which produces two groups of hexapeptides, the desotamides and the wollamides[8]. In this chapter, I describe the identification of the BGC responsible for the production of these distinct peptides, the unique mode of their biosynthesis, and the implications of these results on the evolution of biosynthetic assembly-lines.

2. 1. 1 The Desotamides and the Wollamides

The desotamides (A-F) (Figure 2. 1) are a group of cyclic hexapeptides produced by *Streptomyces scopuliridis* SCSIO ZJ46[9], *Streptomyces sp.* NRRL 21611[10] and *Streptomyces sp.* MST-110588[8]. Desotamide A and B (**1** and **2**) have demonstrated antibiotic activity against Gram-positive bacteria including methicillin resistant *Staphylococcus epidermidis* (MRSE)[9]. The desotamide BGC (*dsa*) encodes twenty-five genes including a six module NRPS (*dsaIHG*) and genes for the biosynthesis of the non-proteinogenic amino acid L-*allo*-isoleucine (*dsaDE*).

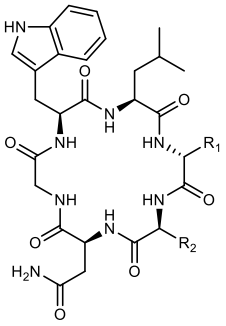
Of the producing strains, MST-110588 is remarkable in its ability to produce, in addition to the desotamides, the structurally related peptides wollamide A (**3**) and wollamide B (**4**) (Figure 2. 1). The wollamides differ from the desotamides by the substitution of glycine for D-ornithine in the sixth position of the hexapeptide[8]. In addition, they have shown to be promising leads for drug development due to their antitubercular activity, lack of cytotoxicity and high stability in blood plasma[8,11]. Several analogues with improved activities and stabilities have been generated synthetically (Table 2. 1)[11,12].

Although the desotamide (*dsa*) biosynthetic gene cluster (BGC) has been previously identified[13], wollamide biosynthesis has yet to be described. There are three possible hypotheses that explain the production of both the wollamides and desotamides by MST-110588: i.) promiscuous activity of the module six adenylation domain; ii.) duplication and refunctionalisaton of the final assembly-line protein and; iii.) duplication and refunctionalisation of the entire cluster (Figure 2. 2). In this Chapter, I present the BGCs of MST-110588 and a novel desotamide producer *Streptomyces sp.* MST-70754 and provide genetic and biochemical evidence for the bifurcated biosynthesis of both the wollamides and desotamides in MST-110588. Further *in silico* analyses support the hypothesis that this BGC has arisen through the process of gene duplication and refunctionalisation through intragenomic recombination. To our knowledge this is the first example of such an NRPS BGC in the literature.

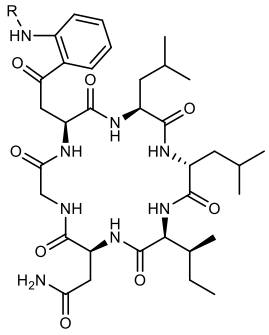
2. 2. Results and Discussion

2. 2. 1. Production of the Wollamides and Desotamides

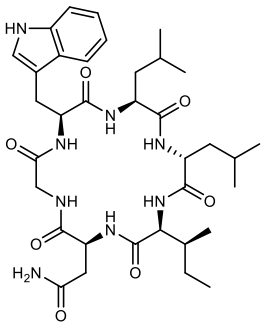
Streptomyces sp. MST-110588 was grown on SFM agar for 7 days. 1 cm agar plugs from each plate were extracted with 1 ml ethyl acetate. The organic phase was removed and evaporated and the resulting residue solved in 200 µl methanol. LCMS analysis of culture extracts confirmed the production of desotamide A (**1**), desotamide B (**2**), wollamide A (**3**) and wollamide B (**4**) (Figure 2. 3.a, Table 2. 2). In addition, we identified a novel desotamide producing strain *Streptomyces* sp. MST-70754 (Ernest Lacey, personal communication). Metabolic extracts were



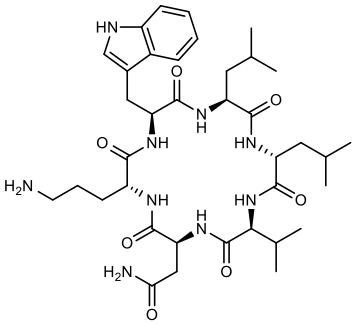
Desotamide A (**1**): R_1 = D-Leu, R_2 = L-*allo*-Ile
Desotamide B (**2**): R_1 = D-Leu, R_2 = L-Val
Desotamide E: R_1 = D-Val, R_2 = L-*allo*-Ile
Desotamide F: R_1 = L-*allo*-Ile, R_2 = L-*allo*-Ile



Desotamide C: R_1 = CHO
Desotamide D: R_1 = H



Desotamide G



Wollamide A (**3**): R_1 = L-*allo*-Ile
Wollamide B (**4**): R_2 = L-Val

Figure 2. 8: Structures of the desotamides (A-G) and wollamides (A and B). Desotamide A-F were isolated from Streptomyces spp. metabolic extracts. Desotamide G was produced when the dsa BGC was expressed heterologously in Streptomyces coelicolour A3(2). Wollamide A and B (**3** and **4**) were isolated alongside desotamide A and B (**1** and **2**) from Streptomyces sp. MST-110588.

	Name	Position					
		I	II	III	IV	V	VI
Naturally Occurring	Desotamide A (**1**)	L-Trp	L-Ile	D-Leu	L-allo-Ile	L-Asn	Gly
	Desotamide B (**2**)	L-Trp	L-Ile	D-Leu	L-Val	L-Asn	Gly
	Desotamide C	N-Formyl-L-Kyn	L-Ile	D-Leu	L-allo-Ile	L-Asn	Gly
	Desotamide D	L-Kyn	L-Ile	D-Leu	L-allo-Ile	L-Asn	Gly
	Desotamide E	L-Trp	L-Ile	D-Val	L-allo-Ile	L-Asn	Gly
	Desotamide F	L-Trp	L-Ile	D-allo-Ile	L-allo-Ile	L-Asn	Gly
	Desotamide G	L-Trp	L-Ile	D-Leu	L-allo-Ile	L-Asp	Gly
	Wollamide A (**3**)	L-Trp	L-Ile	D-Leu	L-allo-Ile	L-Asn	D-Orn
	Wollamide B (**4**)	L-Trp	L-Ile	D-Leu	L-Val	L-Asn	D-Orn
Synthetic	Wollamide 7c	L-Trp	L-Leu	D-Leu	L-Ile	L-Asn	D-Orn
	Wollamide 9c	L-Trp	L-Leu	D-Leu	L-Val	D-Orn	D-Orn
	Wollamide 10c	L-Trp	L-Leu	D-Leu	L-Ile	L-Ser	D-Orn
	Wollamide 13c	L-Trp	L-Leu	D-Leu	L-Ile	L-Asn	D-Arg
	Wollamide 16c	L-Trp	L-Leu	D-Phe(4-Cl)	L-Val	L-Asn	D-Orn

Table 2. 1: Peptide structure of the desotamides, wollamides and synthetic derivatives. Position is relative to the order of incorporation by the desotamide (*dsa*) non-ribosomal peptide synthetase (NRPS).

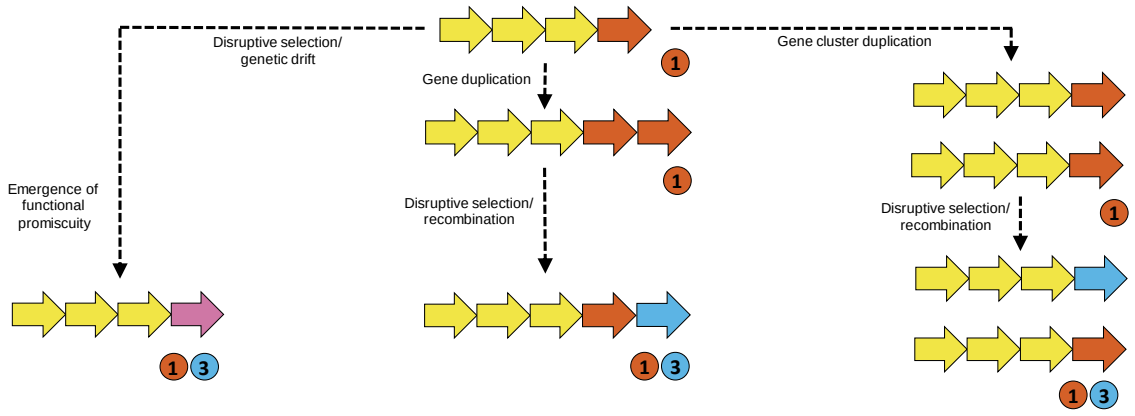


Figure 2. 9: Evolutionary hypotheses for the biosynthesis of desotamide A (1) and wollamide A (3) by *Streptomyces* sp. MST-110588. Cartoon showing from left to right: i) promiscuous activity of the module six adenylation domain; ii) duplication and refunctionalisaton of the final gene and; iii) duplication and refunctionalisation of the entire cluster.

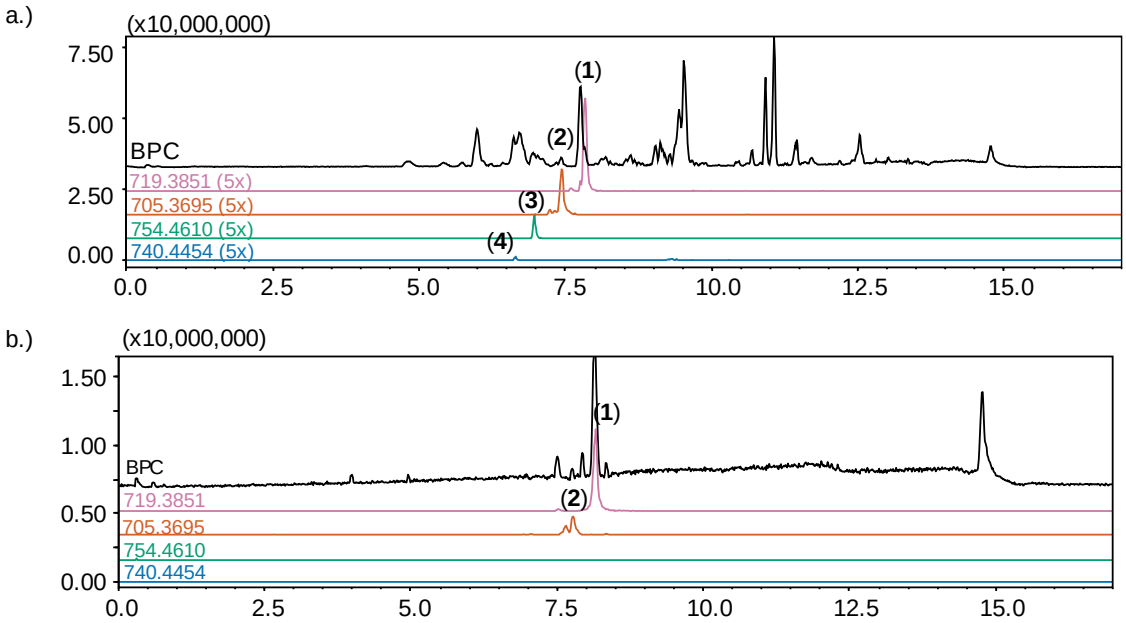


Figure 2. 10: Production of desotamide A and B (1 and 2) and wollamide A and B (3 and 4) by Streptomyces sp. MST-110588 and Streptomyces sp. MST-70754. LCMS spectra of crude extracts of a.) *Streptomyces*. sp MST-110588 and b.) *Streptomyces*. sp. 70754. Base peak chromatograms and extracted ion chromatograms for desotamide A (**1**)([M+Na]$^+$ = 719.3851), desotamide B (**2**)([M+Na]$^+$ = 705.3695) and wollamide A (**3**)([M+H]$^+$ = 754.4610), wollamide B (**4**)([M+H]$^+$ = 740.3695) are shown for both strains. Extracted ion chromatograms from *Streptomyces*. sp MST-110588 extracts are amplified (5x) for clarity.

Compound	Formula	Theoretical m/z	Observed m/z	Mass Difference (Da)	Mass Error (ppm)
1	$C_{35}H_{53}N_8O_7^+$	697.4032	697.4084	-0.0052	7.4949
	$C_{35}H_{52}N_8O_7Na^+$	719.3851	719.3860	-0.0009	1.2511
2	$C_{34}H_{51}N_8O_7^+$	683.3875	683.3886	-0.0011	1.6096
	$C_{34}H_{50}N_8O_7Na^+$	705.3695	705.3726	-0.0031	4.3949
3	$C_{38}H_{60}N_8O_7^+$	754.4610	754.4636	-0.0026	3.4462
	$C_{38}H_{59}N_8O_7Na^+$	776.4430	776.4413	0.0017	-2.1895
4	$C_{37}H_{58}N_8O_7^+$	740.4454	740.4460	-0.0006	0.8103
	$C_{37}H_{57}N_8O_7Na^+$	762.4273	762.4256	0.0017	-2.2297

Table 2. 2: Observed masses for the desotamides and the wollamides. Comparison of the theoretical and observed *m/z* values for hydrogen and sodium adducts of desotamide A (**1**), desotamide B (**2**), wollamide A (**3**) and wollamide B (**4**). Mass differences and mass errors also provided.

generated as above and masses for **1** ($[M+Na]^+$ = 719.3832) and **2** ($[M+Na]^+$ = 705.3698) were detected. Masses corresponding to **3** or **4** could not be detected (Figure 2. 3.b).

2. 2. 2 Identification of the Wollamide BGC

As stated in section 2. 1. 1, we hypothesised three mechanisms that might account for combined biosynthesis of the desotamides and wollamides: i) the presence of a final NRPS module showing promiscuous selectivity able to accept both glycine and L-ornithine as substrates; ii) duplication and diversification of the final NRPS module, creating a bifurcated biosynthetic pathway, or; iii) duplication and diversification of the entire BGC (Figure 2. 2). To test these hypotheses, high quality genomic assemblies of MST-110588 and MST-70754 were obtained using the Pacific Biosciences RSII[14] and Illumina sequencing platforms respectively. Sequencing of MST-110588 yielded a single 7.91 Mb contig. Sequencing of MST-70754 yielded a large contig of 9.01 Mb and four smaller contigs of 86.32, 12.07, 5.38 and 4.41 Kb. All contigs were submitted to antiSMASH v4.02[15] for the identification of specialised metabolic BGCs. There was little similarity between the BGCs present in each strain, however both strains contained a BGC with clear similarity to the previously published (*dsa*) BGC[13] (Figure 2. 4, Table 2. 4 and Table 2. 3). Due to its ability to biosynthesise the wollamides, we designated the BGC from *S.* sp. MST-110588 as *wol*.

Upon comparison, the *wol* and the two *dsa* BGCs have similar topologies (Figure 2. 4, Table 2. 4), except for an apparent duplication of two genes, the NRPS-encoding *dsaG* homologues (*wolG1* and *wolG2*) and the MbtH-like protein-encoding *dsaF* homologues (*wolF1* and *wolF2*). The *wol* BGC also contains six additional genes *wolRSTUVW* predicted to be involved in the biosynthesis of L-ornithine (Table 2. 4). *In silico* analyses of adenylation domain specificities predicted the substrates for the final adenylation (A) -domains of WolG1 and WolG2 (henceforth WolG1A2 and WolG2A2) to be glycine and L-ornithine respectively[15–18]. This supports a previously undescribed bifurcated model of NRP biosynthesis in which the first four rounds of peptide elongation proceed *via* the colinear activity of WolI and WoH, but the final two elongation steps are catalysed independently by WolG1 or WolG2, yielding desotamide or wollamide products respectively (Figure 2. 5).

2. 2. 3 Duplication and Recombination of an Ancestral NRPS

The unusual architecture of the *wol* BGC led us to consider potential mechanisms for its evolution. The high similarity of *wolG1* to *wolG2* (80.9%) and *wolG1*/*wolG2* to *dsaG* (74.3%/71.2% respectively) is indicative of an ancestral gene duplication event. Despite high similarity between *wolG1* and *wolG2*, there is a notable drop in nucleotide identity within the

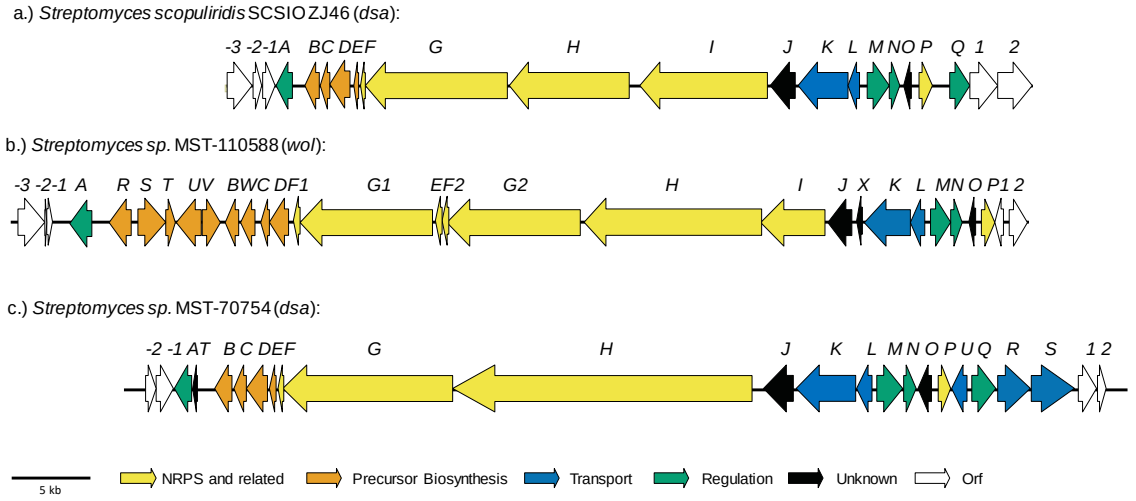


Figure 2. 11: Comparison of wollamide (*wol*)and desotamide (*dsa*) biosynthetic gene clusters (BGCs). a.) Previously published *Streptomyces scopuliridis* SCSIOZJ46 *dsa* loci; b.) *Streptomyces* sp. MST-11058 *wol* loci containing duplicated G (NRPS encoding) and F (MbtH-like protein encoding) genes and additional genes for amino acid biosynthesis and; c.) *Streptomyces* sp. MST-70754 loci were the first four modules are encoded by a single gene (*dsaH*).

Gene	Size (bp)	Size (aa)	Putative Function	Protein Homologue	% Identity	% Similarity
orf(-3)	1458	485	-	M6 family metalloprotease domain-containing protein [Streptomyces sp. FxanaC1]	65	76
orf(-2)	165	54	-	hypothetical protein BG653_04062 [Streptomyces platensis]	80	87
orf(-1)	318	105	-	DUF4190 domain-containing protein [Streptomyces varsoviensis]	70	77
wolA	771	256	regulatory element	DsaA [Streptomyces scopuliridis]	76	84
wolR	1344	447	amino acid biosynthesis	phospho-2-dehydro-3-deoxyheptonate aldolase [Streptomyces sp. KCB13F003]	85	88
wolS	1647	548	amino acid biosynthesis	anthranilate synthase component I family protein [Streptomyces sp. NRRL S-1813]	69	76
wolT	576	191	amino acid biosynthesis	glutamine amidotransferase [Streptomyces decoyicus]	85	89
wolU	1422	473	amino acid biosynthesis	PLP-dependent aminotransferase family protein [Streptomyces decoyicus]	82	87
wolV	1059	352	amino acid biosynthesis	anthranilate phosphoribosyltransferase [Streptomyces decoyicus]	83	89
wolB	792	263	amino acid biosynthesis	indole-3-glycerol phosphate synthase TrpC [Streptomyces sp. NRRL S-1813]	76	82
wolW	855	284	amino acid biosynthesis	amidinotransferase [Streptomyces sp. KCB13F003]	91	94
wolC	504	167	unknown	DsaC [Streptomyces scopuliridis]	81	87
wolD	1170	389	isoleucine biosyntheis	DsaD [Streptomyces scopuliridis]	82	87
wolF1	252	83	MbtH protein	MbtH protein [Streptomyces sp. KCB13F003]	61	68
wolG1	7905	2634	NRPS	DsaG [Streptomyces scopuliridis]	72	79
wolE	381	126	isoleucine biosyntheis	DsaE [Streptomyces scopuliridis]	81	92
wolF2	210	69	MbtH protein	MbtH protein [Streptomyces sp. KCB13F003]	87	97
wolG2	7926	2641	NRPS	DsaG [Streptomyces scopuliridis]	66	75
wolH	10587	3528	NRPS	NRPS [Streptomyces sp. KCB13F003]	67	76
wolI	3732	1243	NRPS	DsaI [Streptomyces scopuliridis]	71	79
wolJ	1380	459	unknown/transpeptidase	DsaJ [Streptomyces scopuliridis]	68	78
wolX	420	139	-	hypothetical protein [Streptomyces sp. KCB13F003]	59	71
wolK	2772	923	transport	ABC transporter permease [Streptomyces sp. KCB13F003]	71	81
wolL	819	272	transport	ABC transporter ATP-binding protein [Streptomyces yunnanensis]	79	89
wolM	1212	403	regulatory element	DsaM [Streptomyces scopuliridis]	81	88
wolN	669	222	regulatory element	LuxR family transcriptional regulator [Streptomyces sp. KCB13F003]	95	97
wolO	414	137	unknown	DsaO [Streptomyces scopuliridis]	54	77
wolP	789	262	thioesterase	thioesterase [Streptomyces sp. 769]	69	79
orf(+1)	558	185	transport	ABC transporter [Streptomyces sp. KCB13F003]	63	73
orf(+2)	1035	344	-	PREDICTED: uncharacterized protein LOC108617933 isoform X3 [Drosophila arizonae]	43	65

Table 2. 3: Description of proteins encoded by the *Streptomyces* sp. MST-110588 wollamide (*wol*) biosynthetic gene cluster (BGC). Proposed function for each protein and closest homologues are shown.

Gene	Size (bp)	Size (aa)	Putative Function	Protein Homologue	Homologue Coverage	Homologue Identity
orf(+3)	1,401	467	-	hypothetical protein [Streptomyces scopuliridis]	100	89
orf(+2)	483	161	-	transcriptional regulator [Streptomyces scopuliridis]	100	100
orf(+1)	702	234	-	hydrolase [Streptomyces scopuliridis]	100	98
dsaA	918	306	regulatory element	DsaA [Streptomyces scopuliridis]	100	99
dsaT	186	62	unknown	-	-	-
dsaB	813	271	unknown	DsaB [Streptomyces scopuliridis]	100	98
dsaC	504	168	unknown	DsaC [Streptomyces scopuliridis]	100	99
dsaD	1,134	378	L-*allo*-isoleucine biosynthesis	DsaD [Streptomyces scopuliridis]	100	98
dsaE	375	125	L-*allo*-isoleucine biosynthesis	DsaE [Streptomyces scopuliridis]	100	98
dsaF	210	70	MbtH protein	DsaF [Streptomyces scopuliridis]	100	100
dsaG	7,944	2648	NRPS	DsaG [Streptomyces scopuliridis]	100	97
dsaH	14,373	4791	NRPS	NRPS [Streptomyces sp. KCB13F003]	100	65
dsaJ	1,410	470	unknown	DsaJ [Streptomyces scopuliridis]	100	99
dsaK	2,760	920	transport	DsaK [Streptomyces scopuliridis]	100	99
dsaL	819	273	transport	DsaL [Streptomyces scopuliridis]	100	99
dsaM	1,212	404	regulatory element	DsaM [Streptomyces scopuliridis]	100	98
dsaN	669	223	regulatory element	DsaN [Streptomyces scopuliridis]	100	99
dsaO	450	150	unknown	DsaO [Streptomyces scopuliridis]	100	98
dsaP	672	224	thioesterase	DsaP [Streptomyces scopuliridis]	100	97
dsaU	618	206	transport	ABC transporter [Streptomyces sp. KCB13F003]	89	62
dsaQ	1,140	380	regulatory element	DsaQ [Streptomyces scopuliridis]	100	98
dsaR	1,557	519	transport	ATP-binding protein [Streptomyces scopuliridis]	100	99
dsaS	1,977	659	transport	inner-membrane translocator [Streptomyces scopuliridis]	99	99
orf(-1)	981	327	-	ribokinase [Streptomyces scopuliridis]	100	89
orf(-2)	387	129	-	D-ribose pyranase [Streptomyces scopuliridis]	100	94

Table 2. 4: Description of proteins encoded by the *Streptomyces* sp. MST-70754 desotamide (*dsa*) biosynthetic gene cluster (BGC). Proposed function for each protein and closest homologues are shown.

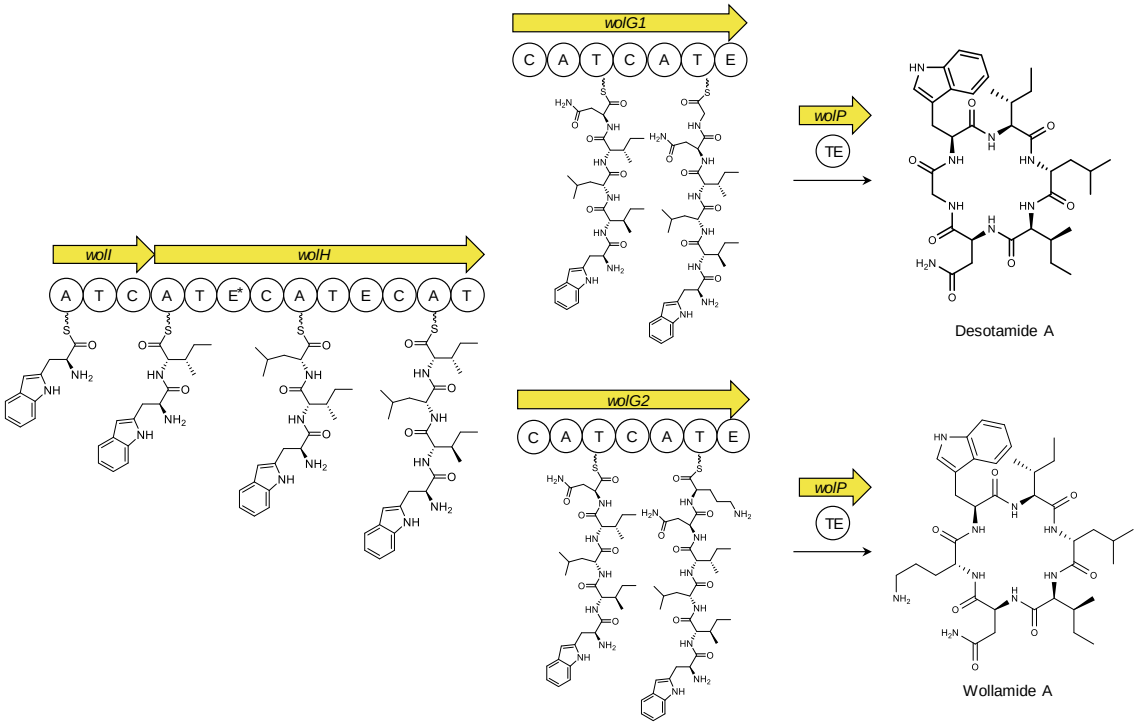


Figure 2. 12: Bifurcated biosynthesis of the wollamides and desotamides by Streptomyces MST-110588. Non-ribosomal peptide synthesis proceeds in a colinear fashion for the first four modules. The amino acids incorporated by the final two rounds of condensation are dependent upon the utilisation of WolG1 (yielding desotamide) or WolG2 (yielding wollamide).

region coding for the final adenylation domains which is manifest in the gene products (54.6% nucleotide and 29.2% protein sequence similarity) (Figure 2. 6). This region corresponded to a large increase in the number of non-synonymous mutations, and therefore the d_s/d_N ratio. Given the overall similarity of these genes we deemed it unlikely that such high sequence variation could emerge through speciation (point mutation) alone. Similar patterns have been observed in other NRPS clusters, for example Crüsemann et al.[19] hypothesised that recombinations within the adenylation domain coding regions of the hormaomycin BGC had led to a switch in substrate specificity, and this was later ascribed to substitution of the flavodoxin-like sub-domain by Kries et al.,[20] which comprise the key substrate selectivity motifs. *In silico* analyses of the genes encoding microcystin NRPSs also revealed evidence for recombination of regions of around 1 kb, including the flavodoxin-like subdomain[2,21].

To assess the role of intragenomic recombination we generated a nucleotide sequence alignment of all NRPS-associated adenylation domains present in the *S.* sp. MST-110588 genome for analysis with the Recombination Detection Programme 4 (RDP4)[22]. RDP4 combines the output of seven algorithms (RDP[23], GENECONV[24], BOOTSCAN[25], MAXCHI[26], CHIMERA[27], SISCAN[28] and 3SEQ[29]) to rank the likelihood of recombination events and identifies sequence triplets consisting of a recombinant, its major parent (contributing the majority of the recombinant DNA) and/or its minor parent (contributing the smaller, internal portion of the recombinant DNA). RDP4 predicted 29 potential recombination events (Table 2. 5), 11 of which were supported by two or more algorithms, allowing recombination breakpoints to be predicted (Figure 2. 8, Table 2. 6). Only two recombination events were supported unanimously, among which *wolG1A2* was identified as a recombinant sequence with *wolG2A2* as the major parent and an adenylation domain encoding sequence from elsewhere in the genome, *orf6595A* (see below), as the minor parent. A model of the terminal module of WolG1 was generated in Phyre2[30] based upon the existing structure for terminal module of surfactin synthetase (SrfA-C, PDB: 2VSQ)[31] (Figure 2. 7). These structures confirmed that the proposed sub-domain swap could occur without perturbing the interactions with the upstream C-domain or downstream domains. The recombinant region lies between well characterised conserved motifs: A2 (N-terminal) and A6 (C-terminal), thus comprising the flavodoxin subdomain and a large portion of the N-terminal subdomain (Figure 2. 8). Crucially, this would allow for the substitution of the amino acid binding pocket and catalytic P-loop while maintaining the condensation-adenylation domain interface. All of the five recombination events predicted to swap the flavodoxin subdomain are predicted to include the P-loop (Figure 2. 8, Table 2. 6) suggesting there may be an advantage to maintaining the structural relationships between the P-loop and substrate binding pocket, a relationship that has not been considered in previous A-domain engineering experiments[19,20]. This result also

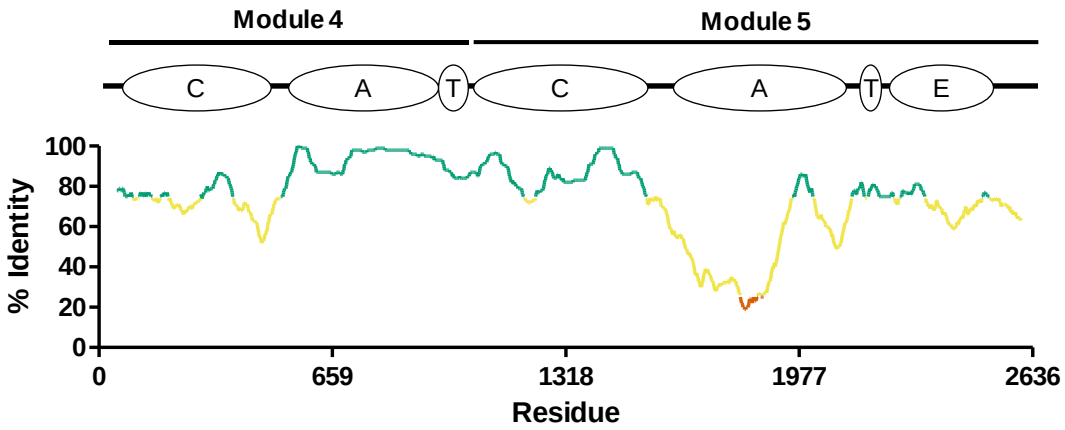


Figure 2. 13: Pairwise identity between WolG1 and WolG2. Generated from protein sequences with a sliding window of 50 amino acids. Plot is coloured by identity at 75% (green), 25-75% (yellow) and <25% (red) intervals. Modular condensation (C)-, adenylation (A)- and epimerisation (E)-domains shown above.

Events	Domains			Detection Methods						
	Recombinant	Major parent	Minor parent	R	G	B	M	C	S	T
1	orf1237A	Unknown	orf1236A	+	+	+	+	+	+	+
1	wolG1A2	wolG2A2	orf6595A	+	+	+	+	+	+	+
1	orf6593A2	orf1236A	Unknown	+	-	+	-	+	+	-
4	wolHA1	wolHA3	orf1227A	-	-	-	+	+	+	-
1	orf3282A	orf6595A	Unknown	-	-	+	+	-	+	-
1	orf1237A	orf6652A	Unknown	+	+	-	+	-	-	-
1	orf1224A	orf6593A1	orf289A	-	-	-	+	+	-	-
1	orf1236A	Unknown	orf6594	-	-	-	+	-	+	-
1	orf252A	orf253A	orf6595A	-	-	-	+	+	-	-
1	orf363A	wolG2A2	orf6625	-	-	-	+	+	-	-
1	orf6593A1	orf6652A	Unknown	+	-	-	-	-	+	-
1	orf1236A	wolHA3	orf1237A	-	-	-	+	-	-	-
1	wolG1A2	orf288A	orf248A	-	-	-	+	-	-	-
1	orf288A	orf6172A	wolG1A2	-	-	-	+	-	-	-
1	orf288A	orf6172A	Unknown	-	-	-	+	-	-	-
1	orf365A	orf6652A	Unknown	-	+	-	-	-	-	-
1	wolG1A1	wolG2A1	orf3282	-	+	-	-	-	-	-
1	orf1224A	orf365A	wolHA3	-	+	-	-	-	-	-
1	orf6595A	orf6594A	Unknown	-	-	-	+	-	-	-
1	orf3282A	Unknown	orf6595A	-	+	-	-	-	-	-
1	orf1227A	orf6172A	orf288A	-	-	-	+	-	-	-
1	orf1149A	orf321A	Unknown	-	-	-	+	-	-	-
1	orf1158A	orf287A	orf6172	-	-	-	+	-	-	-
2	wolHA3	orf6593A1	orf6625	-	-	-	+	-	-	-
1	orf6225A	Unknown	orf1158A	-	-	-	+	-	-	-
1	orf1158A	orf248A	Unknown	-	-	-	+	-	-	-
1	orf3281A	Unknown	wolHA2	-	-	-	+	-	-	-
1	orf253A	orf6595A	Unknown	+	-	-	-	-	-	-
1	orf365A	orf1149A	Unknown	+	-	-	-	-	-	-

Table 2. 5: Detected Recombination Events. Showing predicted recombinant domains and relative major and minor parents. The detection of a given recombination event by a given algorithm is displayed (+ for detected, - for not detected), where ‘R’ is RDP, ‘G’ is GENECONV, ‘B’ is BOOTSCAN, ‘M’ is MAXCHI, ‘C’ is CHIMERA, ‘S’ is SISCAN and ‘T’ is 3SEQ. The event proposed by this paper is highlighted in purple and is supported by all seven algorithms. Number of events >1 indicates that there is evidence for such an event in related sequences indicating that the event may have occurred in the ancestral sequence.

Recombinant Domains			Recombinant Region	Detection Methods						
Recombinant	Major parent	Minor parent		R	G	B	M	C	S	T
wolG1A2	***wolG2A2***	***orf6595A***	**Y116-G343**	**+**	**+**	**+**	**+**	**+**	**+**	**+**
orf1237A	Unknown	*orf1236A*	I352-I448	+	+	+	+	+	+	+
orf6593A2	*orf1236A*	Unknown	M139-N354	+	-	+	-	+	+	-
orf1237A	*orf6652A*	Unknown	K223-M241	+	+	-	+	-	-	-
orf3282A	*orf6595A*	Unknown	I357-H452	-	-	+	+	-	+	-
wolHA1	*wolHA3*	*orf1227A*	N219-K316	-	-	-	+	+	+	-
wolHA2	*wolHA3*	*orf1227A*	N219-K316	-	-	-	+	+	+	-
wolG1A1	*wolHA3*	*orf1227A*	N219-K316	-	-	-	+	+	+	-
wolG2A1	*wolHA3*	*orf1227A*	N219-K316	-	-	-	+	+	+	-
orf6593A1	*orf6652A*	Unknown	K314-R389	+	-	-	-	-	+	-
orf252A	*orf253A*	*orf6595A*	T173-L202	-	-	-	+	+	-	-
orf363A	*wolG2A2*	*orf6625*	L91-L450	-	-	-	+	+	-	-
orf1224A	*orf6593A1*	*orf289A*	K84-A332	-	-	-	+	+	-	-
orf1236A	Unknown	*orf6594*	M95-I377	-	-	-	+	-	+	-

Table 2. 6: Recombination breakpoint analysis. Predicted recombinants from RDP4 for which breakpoint analysis could be applied. Recombination region refers to the approximate homologous positions in the L-phenylalanine activating domain of gramicidin synthase (1AMU). Events are sorted by support from different detection methods. The rows shaded in grey indicate a shared recombination event in the ancestor of these genes.

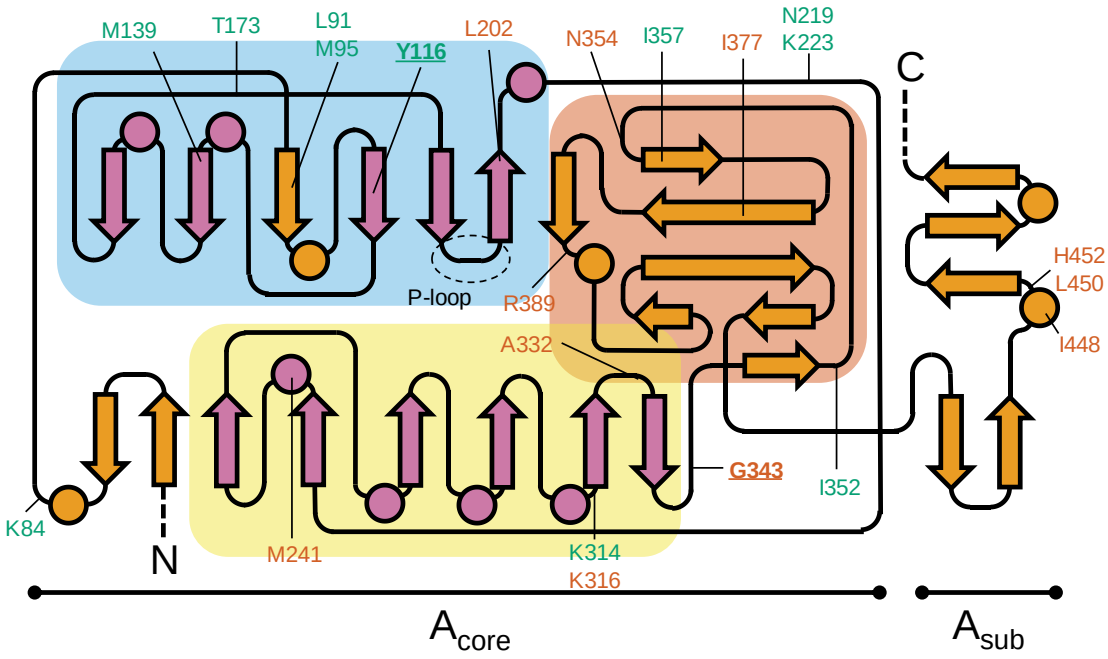


Figure 2. 14: Adenylation domain topology with predicted breakpoints. Topology based on the L-phenylalanine activating domain of gramicidin synthase (1AMU). Circles represent α-helices and arrows represent β-strands. Regions corresponding to the large N-terminal domain (A_{core}) and small C-terminal domain (A_{sub}) are highlighted and subdomains shaded (subdomain 1 = blue; subdomain 2 (flavodoxin-like subdomain) = yellow; subdomain 3 = red). The residues indicated correspond to approximate sites of recombination from Table 2.6 (green = N-terminal; red = C-terminal). Recombinant region of *wolG1* coloured in purple and associated breakpoints are underlined.

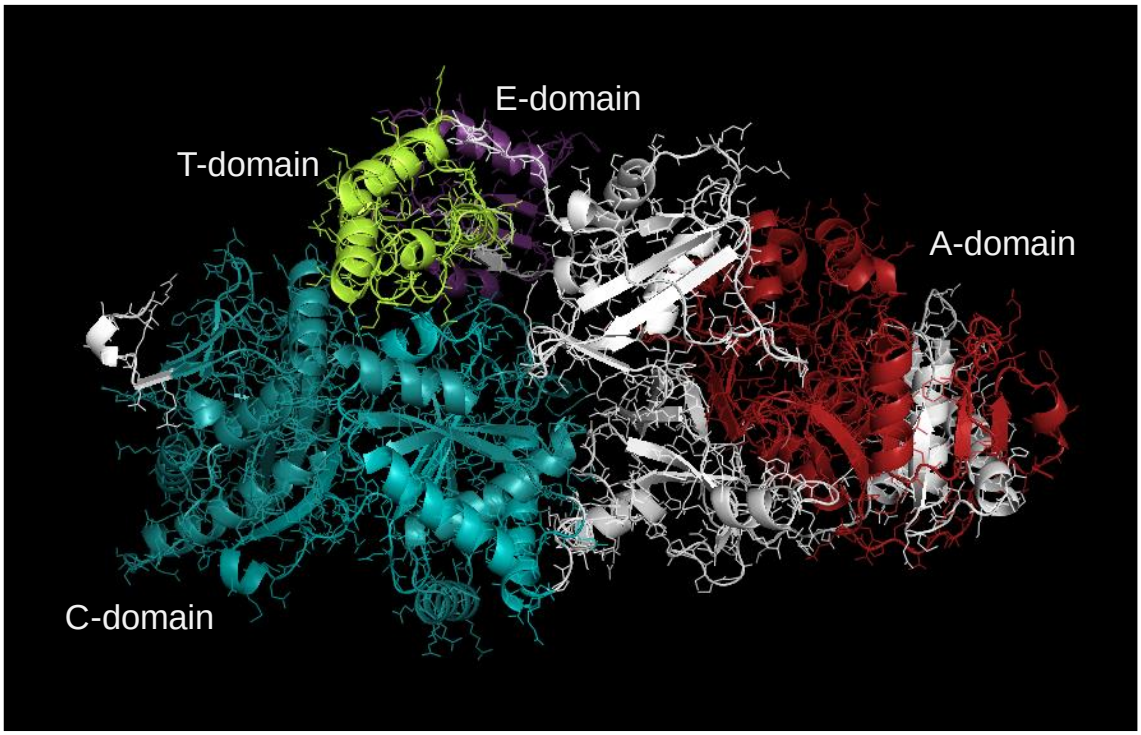


Figure 2. 15: Structural model of the final module of WolG1. Model built in Phyre2 based on the structure of surfactin synthetase (SrfA-C, PDB: 2VSQ). Domains are highlighted as follows: condensation (C)-domain in blue, adenylation (A)-domain in white, thiolation (T)-domain in yellow and epimerase (E)-domain in purple. Predicted recombinant region coloured in red.

mirrors previous work on the evolution of polyketide synthases (PKSs) which suggested that intragenomic recombination was sufficient to explain the diversity of PKSs in the *Streptomyces avermilitis* genome[32]. Considering the low rates of horizontal gene transfer between *Streptomyces* spp.[33], it seems likely that intragenomic recombination plays a major role in the diversification of modular biosynthetic assemblies in these organisms.

2. 2. 4 *In vitro* Analysis of *wol* Module Six Adenylation Domains

The minor parent, *orf6595A*, is a domain originating from *orf6595*, a gene within an NRPS BGC. *orf6595* encodes a single NRPS module that is predicted to select glycine as a substrate for activation (Figure 2. 9, Table 2. 7). To examine the substrate specificity of the A-domains of interest, pET28a(+) hexa-histidine tagged A-domain WolG1A2, WolG2A2 and ORF6595A constructs were expressed in *E. coli* NiCo21(DE3) *pLysS* (using A-domain boundaries selected as described in Crüsemann et al[19]) and purified using Ni-Affinity chromatography. Initially, the resulting protein was insoluble, however co-expression with the MbtH-like protein WolF2 (expressed from pCDFDuet-1) enabled the purification of soluble protein (Figure 2. 10). The ability of these A-domains to activate each of the twenty proteinogenic amino acids and L-ornithine was then measured using the hydroxylamine trapping assay **(**Figure 2. 11)[34]. WolG2A2 adenylates L-ornithine, in line with our hypothesis; however, it was also capable of activating other substrates albeit with lower efficiency. Most noticeably, WolG2A2 accepted L-aspartate (58 % activity relative to L-ornithine) and L-asparagine (44 % activity relative to L-ornithine) as substrates, but wollamide analogues in which aspartate or asparagine were substituted for ornithine were not identified in culture extracts of *S.* sp. MST-110588 using targeted LCMS analysis. In contrast, both WolG1A2 and Orf6595A activate glycine in a highly specific manner, with little or no activity for the other amino acids tested. These data demonstrate the possibility that an historic recombination with *orf6595* could alter the substrate specificity of the module six adenylation domain from L-ornithine to glycine. Such an event would also explain the presence of the, apparently redundant, epimerase domain in the terminal module of *DsaG*.

Based on these data, we predict the existence of an ancestral BGC that exclusively produced wollamides (or related hexapeptides). Duplication of the gene encoding the final two modules of this NRPS followed by intra-genomic recombination with a homologous sequence encoding the glycine-specific adenylation domain of *orf6595* (and subsequent point mutations), resulted in the bifurcated biosynthesis pathway observed in the genome of *S. sp*. MST-110588. In a divergent lineage, the ancestral gene encoding the L-ornithine-specific adenylation domain along with associated genes encoding amino acid biosynthesis are lost through gene deletion, ultimately resulting in the contemporary *dsa* BGCs (producing desotamides only)(Figure 2. 12). This suggests an important role for gene duplication in the evolution of biosynthetic assembly-lines. The likely role of intragenic recombination as a means to increase the number of

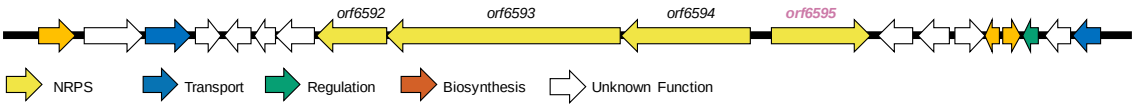


Figure 2. 16: Biosynthetic gene cluster (BGC) associated with *orf6595.* This gene (highlighted in purple) is the predicted minor parent conferring glycine substrate specificity to wolG1. See Table 2. 7 for a description of predicted gene product homologues.

Gene	Protein Homologue	%Identities	%Similarities
orf6585	MULTISPECIES: ATP-grasp domain-containing protein [Streptomyces]	83	88
orf6586	phosphoenolpyruvate synthase [Streptomyces auratus]	83	89
orf6587	MFS transporter [Streptomyces sp. WMMB 714]	80	86
orf6588	hypothetical protein [Streptomyces sp. NRRL F-5755]	63	70
orf6589	MULTISPECIES: taurine dioxygenase [Streptomyces]	83	91
orf6590	hypothetical protein DF18_35265 [Streptomyces rimosus]	76	82
orf6591	hypothetical protein [Streptomyces aurantiacus]	72	79
orf6592	non-ribosomal peptide synthetase [Streptomyces sp. OK885]	66	74
orf6593	MULTISPECIES: non-ribosomal peptide synthetase [Streptomyces]	73	78
orf6594	MULTISPECIES: non-ribosomal peptide synthetase [Streptomyces]	72	80
orf6595	non-ribosomal peptide synthetase [Streptomyces rimosus]	73	80
orf6596	MULTISPECIES: chitinase [Streptomyces]	83	89
orf6597	ATP-dependent DNA ligase [Streptomyces sp. DvalAA-14]	88	91
orf6598	ATP-dependent DNA ligase [Streptomyces albus]	91	94
orf6599	HAD family phosphatase [Streptomyces monomycini]	71	78
orf6600	GNAT family N-acetyltransferase [Streptomyces xanthophaeus]	63	71
orf6601	Lrp/AsnC family transcriptional regulator [Streptomyces sp. NRRL F-5755]	95	95
orf6602	hypothetical protein [Streptomyces sp. SceaMP-e96]	81	87
orf6603	ABC transporter ATP-binding protein [Streptomyces sp. CB02923]	80	87

Table 2. 7: Homologues of proteins encoded by genes proximal to *orf6595*. These proteins relate to the biosynthetic gene cluster (BGC) in Figure 2.9.

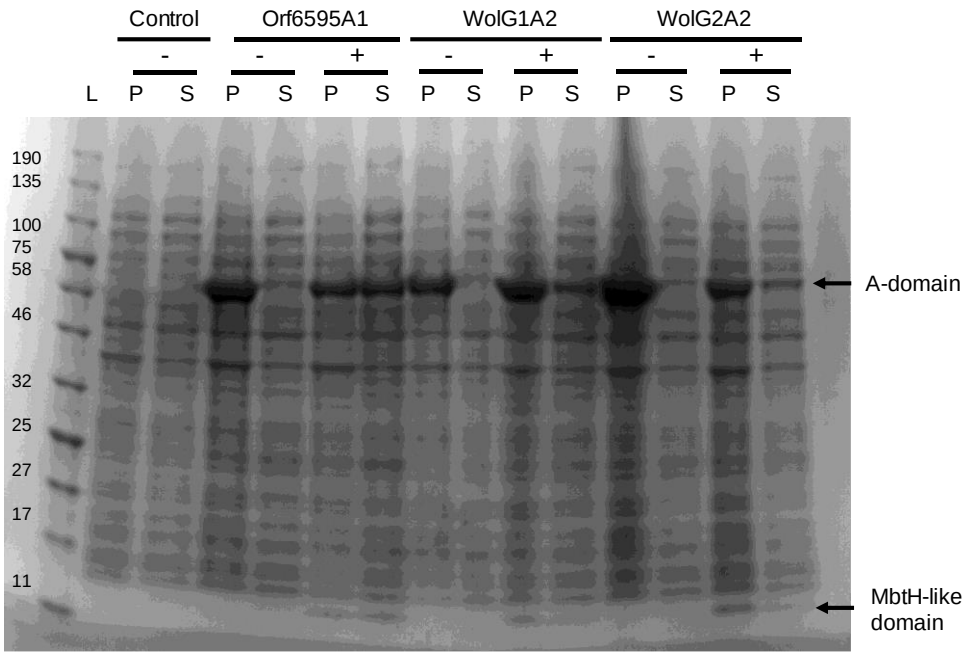


Figure 2. 17: MbtH-domain dependent solubility of heterologously expressed adenylation (A)-domains. SDS-PAGE of *E. coli* NiCo21(DE3) *plysS* lysates to assess adenylation domain solubility. Strains only expressing the pET28a(+) derivative (-) only have an adenylation domain corresponding band (~60 kDa) in the pellet (P), whereas strains co-expressing the mbtH-like domain wolF2 (+) have visible levels of soluble protein (S).

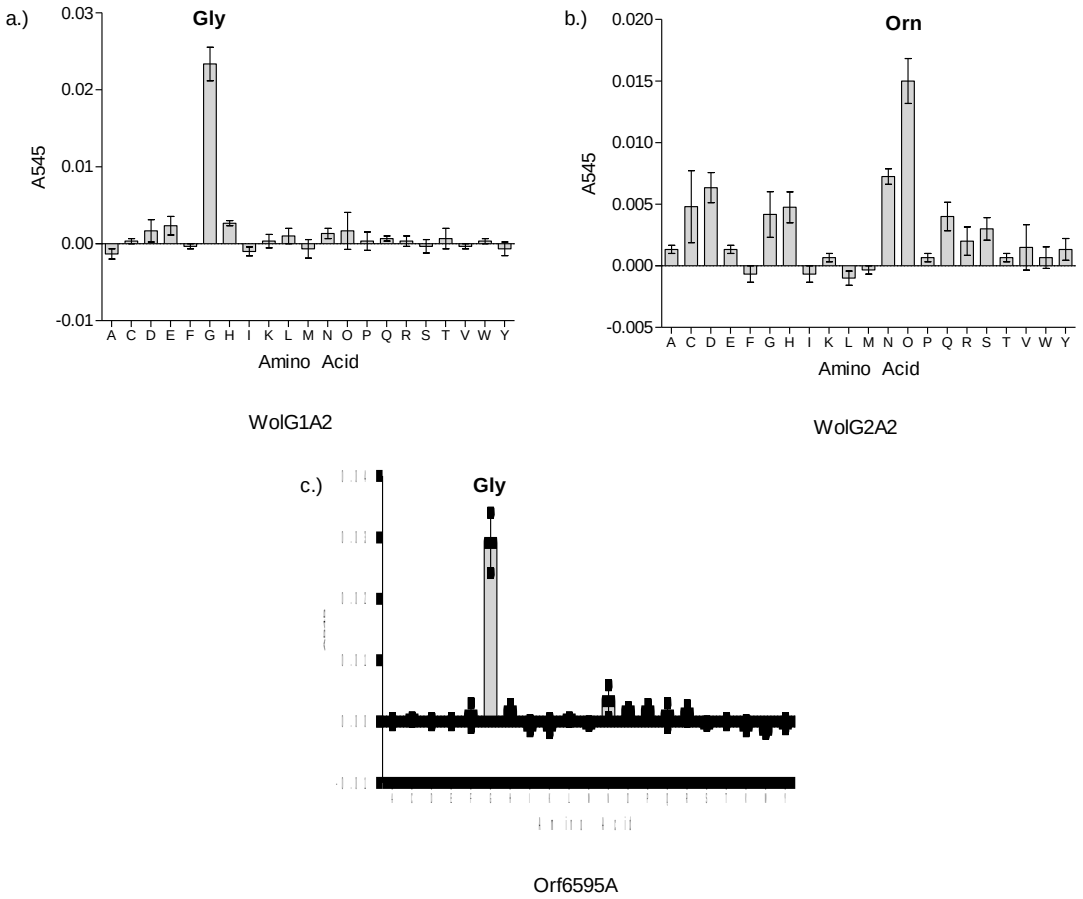


Figure 2. 18: Substrate specificities adenylation (A) domains as determined by hydroxylamine trapping. Absorption at 545 nm, amino acids are represented by their single letter codes. a.) WolG1A2, b.) WolG2A2 and c.) Orf6595A.

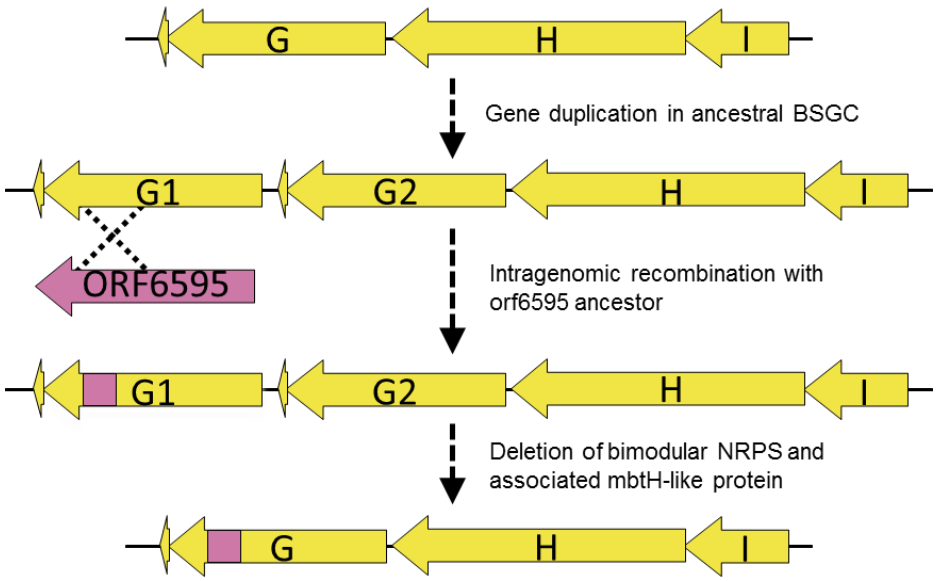


Figure 2. 19: Hypothesis for the evolution of wollamide (*wol)* and desotamide *(dsa)* BGCs**.** A gene duplication event in an ancestral wollamide (or wollamide-like) cluster results in a redundant copy of the bimodular NRPS encoding modules 5 and 6. Subsequently, an intragenomic recombination event between the adenylation domains of module six and *orf6595* results in a bifurcated NRPS of intermediate chemotype (diverging into the lineage of the *wol* BSGC). Loss of the original NRPS gene and genes involved in amino acid biosynthesis results in the contemporary *dsa* BSGC.

biosynthetic modules for PKSs and NRPSs is well documented[2,19,35]. Whereas intragenic duplication increases the length of the assembly-line, whole gene duplication results in the generation of bifurcated systems as evidenced by the *wol* BGC (Figure 2. 13). Gene duplication reduces selective pressure on one copy of the gene due to functional redundancy, and so, in this fashion, a BGC can evolve new functionality without disrupting the original function. This means that refunctionalisation is possible without strong selective pressure and may follow a more gradualistic route.

2. 2. 5. *In vivo* Reconstitution of Ancestral Metabolism

Given the evolutionary relationship of the wollamides and desotamides, we rationalised that it should be possible to resurrect production of the wollamides in a desotamide producing strain through the heterologous expression of *wolG2*. To express *wolG2* in *S. sp* MST-70754, we generated the plasmid pBO1 by yeast mediated assembly of pGP9 and the yeast selective marker (kanMX4) and origin of replication (2μ) of pFF62 A (Figure 2. 14)[36,37]. pBO1 is an integrative *Streptomyces* expression vector capable of propagating in both *E. coli* and *S. cerevisae*. This allows for large genes (such as NRPSs) to be assembled efficiently in yeast *via* homologous recombination and subsequently transferred from *E. coli* to *Streptomyces spp.* for expression. The *wolG2* sequence was cloned into the NdeI site of pBO1 by co-transformation with the NdeI linearised vector into *S. cerevisae* CEN.PK 2-1C. The resulting pBO1-*wolG2* plasmid was transformed into *E. coli* ET12567 pUZ8002 *via* electroporation and conjugated into *S. sp* MST-70754[38]. Methanolic culture extracts of wild-type and mutant strains were analysed with LCMS and the presence of **1, 2** and **3** was confirmed by retention time, isotopic masses and MS/MS fragmentation (Figure 2. 15). These results confirmed the production of the wollamides and desotamides in the *wolG2* overexpression strains. Importantly, this indicates that the pathways have not yet diverged far enough to interfere with the protein-protein interactions between *wolG2* and *dsaG* required for modular interaction. This is perhaps unsurprising given the high similarity of the proteins in both pathways overall and the presumably strong selection to maintain protein-protein intereactions between assembly-line sub units. As in MST-110588, **3** production in the mutant strain low relative to the production of the desotamides. This may be due to competition between WolG2 and DsaG/WolG1 for interaction with the upstream NRPS. In order to test this hypothesis, and to reconstruct the hypothetical ancestral BGC, we plan to knock out *dsaG* from MST-70754 and replace it with *wolG2.*

BlastP results (Table 2. 3) showed that *wolH* from MST-110588 and *dsaH* from MST-70754 shared homology with *ulmA* a NRPS encoding gene from the ulleungmycin assembly-line (Table 2. 3). The ulleungmycins[39] (and related longicatenamycins)[40] are hexapeptides containing the unusual non-proteinogenic amino acids unusual amino acids 5-chloro-L-tryptophan, D-

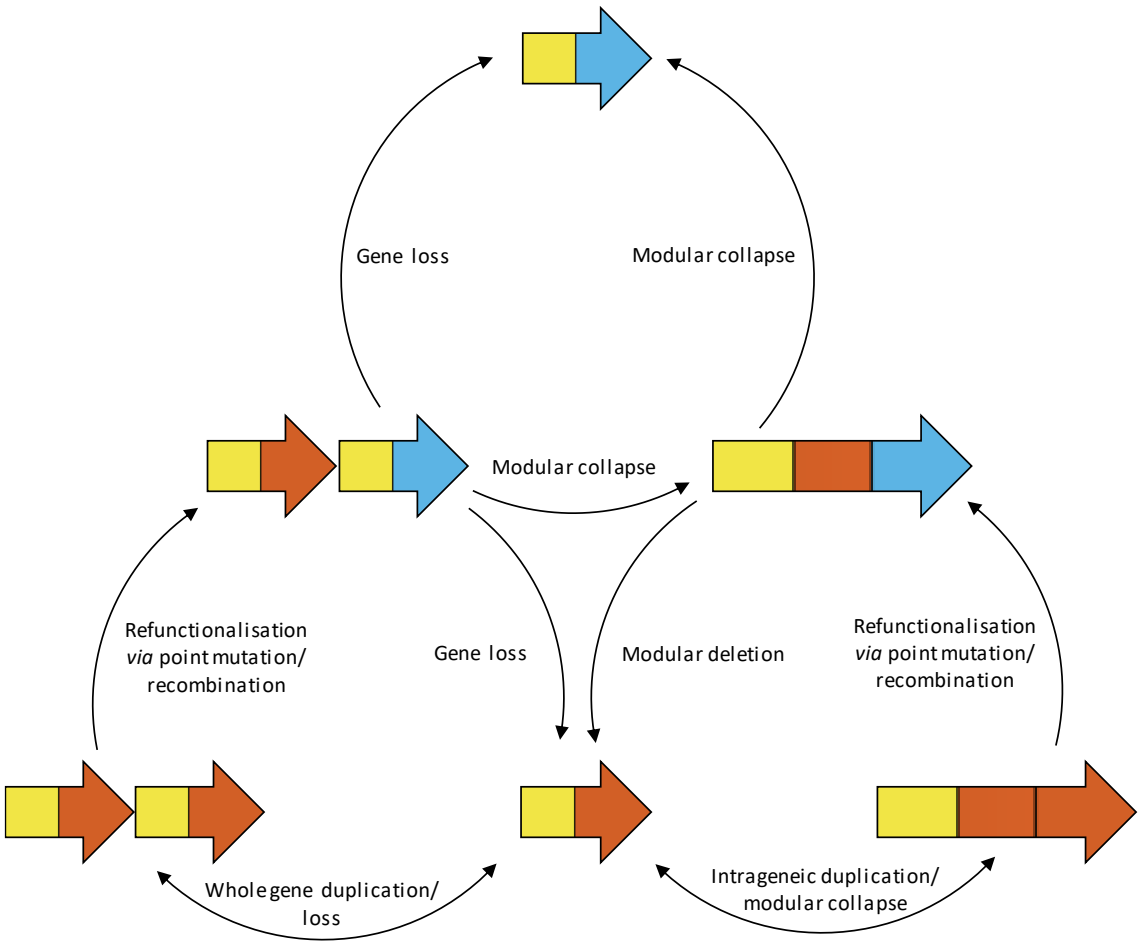


Figure 2. 20: Modes of assembly-line evolution. The different potential mechanisms involved in the evolution of assembly-line biosynthesis. Genes are represented by the large arrows with each segment representing a different module, coloured by function. The small arrows linking genes represent evolutionary processes. This model incorporates whole gene duplication and refunctionalisation (this study), intragenic duplication and refunctionalisation and modular collapse.

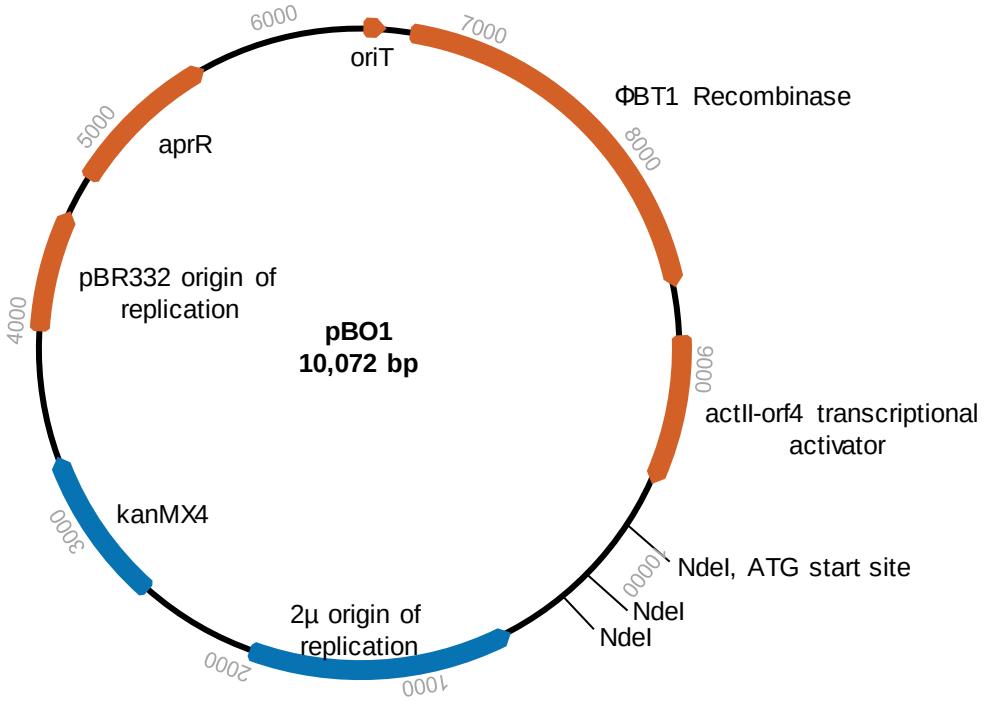


Figure 2. 21: Plasmid map of pBO1. Elements derived from pGP9 are shown in red, elements derived from pFF62A are shown in blue.

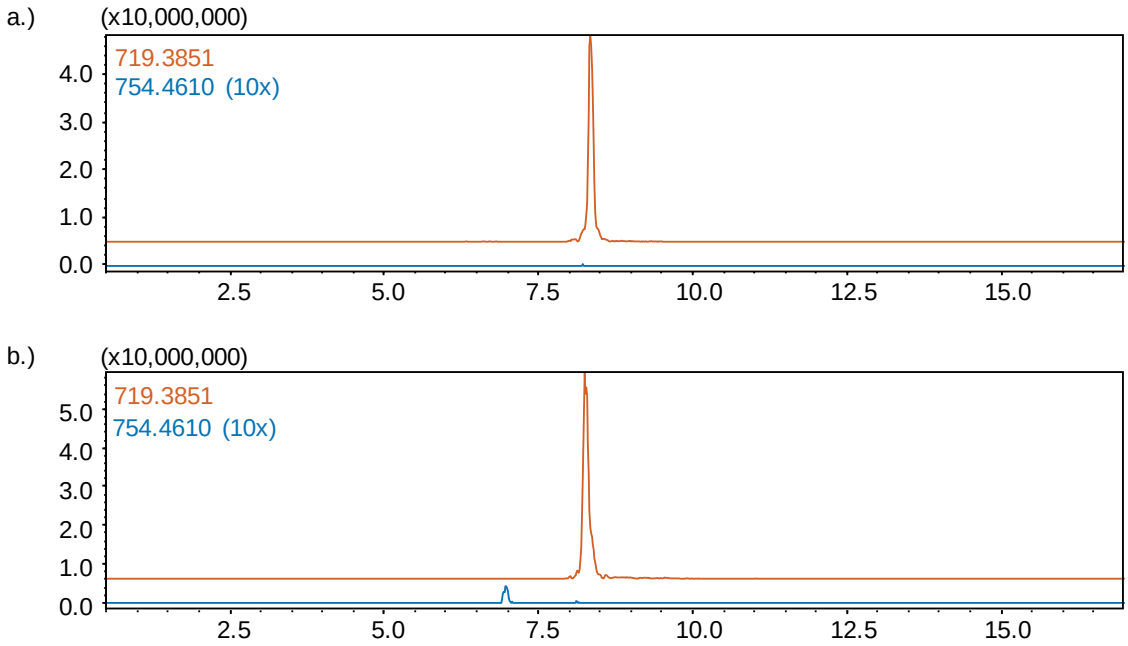


Figure 2. 22: Production of wollamide by *Streptomyces* sp. MST-70754 pGP9-wolG2. Showing the extracted ion chromatogram for desotamide A (**1;** $[M+Na]^+$ = 719.3851) and wollamide A (**1;** $[M+H]^+$ = 754.4610) of a.) *Streptomyces* sp. MST-70754 pGP9 and b.) *Streptomyces* sp. MST-70754 pGP9-wolG2 culture extracts

homoleucine and L-threo-β-OH-aspartate. They also contain structural similarities to the wollamides and desotamides, such as the incorporation of glycine D-ornithine. Comparison of these hexapeptide assembly-lines may yield further information about the formation of biosynthetic assembly-lines. Furthermore, subunit swaps between these two pathways may yield novel products.

2. 3. Conclusion

Here, we have described for the first time, a bifurcated NRPS capable of producing structurally distinct hexapeptides. The *wol* BGC is of unique significance to the evolution of assembly-line biosynthesis and points to a broader role for gene duplication than previously described. We argue that, in addition to its role in modular extension, whole gene duplication might drive cycles of BGC expansion and contraction, providing a new mechanism for assembly-line refunctionalisation (see Chapter Six). This study also highlights the importance of intragenomic recombination as a mechanism for diversification of NRPSs, which is particularly relevant in *Streptomyces,* where horizontal gene transfer has been shown to be relatively rare[33]. The *wol* BGC provides unique insights into how nature adapts the substrate-specificity of an NRPS and differences between *wolG1* and *wolG2* will help identify regions of interest when designing synthetic NRPSs. Finally, we used this knowledge of evolutionary relationships to reconstruct the ancestral phenotype *in vivo*, successfully rescuing wollamide A (**3**) production in an exclusively desotamide producing strain. Experiments are underway to produce the ancestral wollamide-exclusive phenotype either through the knocking out of *dsaG* or reconstitution of *dsaI, dsaH and wolG2* in a heterologous host.

2. 4. Bibliography

(1) Medema, M. H.; Cimermancic, P.; Sali, A.; Takano, E.; Fischbach, M. A. A Systematic Computational Analysis of Biosynthetic Gene Cluster Evolution: Lessons for Engineering Biosynthesis. *PLoS Comput. Biol.* **2014**, *10* (12), e1004016.

(2) Tooming-Klunderud, A.; Fewer, D. P.; Rohrlack, T.; Jokela, J.; Rouhiainen, L.; Sivonen, K.; Kristensen, T.; Jakobsen, K. S. Evidence for Positive Selection Acting on Microcystin Synthetase Adenylation Domains in Three Cyanobacterial Genera. *BMC Evol. Biol.* **2008**, *8* (1), 256.

(3) Otto, S. P.; Whitlock, M. C. The Probability of Fixation in Populations of Changing Size. *Genetics* **1997**, *146* (2) 723-733.

(4) Lawrence, J. G.; Hendrix, R. W.; Casjens, S. Where Are the Pseudogenes in Bacterial Genomes? *Trends Microbiol.* **2001**, *9* (11), 535–540.

(5) Lynch, M. Streamlining and Simplification of Microbial Genome Architecture. *Annu. Rev.*

Microbiol. **2006**, *60* (1), 327–349.

(6) Aharoni, A.; Gaidukov, L.; Khersonsky, O.; Gould, S. M.; Roodveldt, C.; Tawfik, D. S. The "evolvability" of Promiscuous Protein Functions. *Nat. Genet.* **2005**, *37* (1), 73–76.

(7) Brown, S. D.; Babbitt, P. C. New Insights about Enzyme Evolution from Large Scale Studies of Sequence and Structure Relationships. *J. Biol. Chem.* **2014**, *289* (44), 30221–30228.

(8) Khalil, Z. G.; Salim, A. A.; Lacey, E.; Blumenthal, A.; Capon, R. J. Wollamides: Antimycobacterial Cyclic Hexapeptides from an Australian Soil *Streptomyces*. *Org. Lett.* **2014**, *16* (19), 5120–5123.

(9) Song, Y.; Li, Q.; Liu, X.; Chen, Y.; Zhang, Y.; Sun, A.; Zhang, W.; Zhang, J.; Ju, J. Cyclic Hexapeptides from the Deep South China Sea-Derived *Streptomyces Scopuliridis* SCSIO ZJ46 Active Against Pathogenic Gram-Positive Bacteria. *J. Nat. Prod.* **2014**, *77* (8), 1937–1941.

(10) Miao, S.; Anstee, M. R.; LaMarco, K.; Matthew, J.; Huang, L. H. T. and; Brasseur, M. M. Inhibition of Bacterial RNA Polymerases. Peptide Metabolites from the Cultures of *Streptomyces* Sp. *J. Nat. Prod.*, **1997**, *60* (8), 858–861.

(11) Asfaw, H.; Laqua, K.; Walkowska, A. M.; Cunningham, F.; Martinez-Martinez, M. S.; Cuevas-Zurita, J. C.; Ballell-Pages, L.; Imming, P. Design, Synthesis and Structure-Activity Relationship Study of Wollamide B; a New Potential Anti TB Agent. *PLoS One* **2017**, *12* (4), e0176088.

(12) Tsutsumi, L. S.; Elmore, J. M.; Dang, U. T.; Wallace, M. J.; Marreddy, R.; Lee, R. B.; Tan, G. T.; Hurdle, J. G.; Lee, R. E.; Sun, D. Solid-Phase Synthesis and Antibacterial Activity of Cyclohexapeptide Wollamide B Analogs. *ACS Comb. Sci.* **2018**, *20* (3), 172–185.

(13) Li, Q.; Song, Y.; Qin, X.; Zhang, X.; Sun, A.; Ju, J. Identification of the Biosynthetic Gene Cluster for the Anti-Infective Desotamides and Production of a New Analogue in a Heterologous Host. *J. Nat. Prod.* **2015**, *78* (4), 944–948.

(14) Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **2009**, *323* (5910), 133–138.

(15) Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran, H. G.; de los Santos, E. L. C.; Kim, H. U.; Nave, M.; et al. AntiSMASH 4.0—improvements in Chemistry Prediction and Gene Cluster Boundary Identification. *Nucleic Acids Res.* **2017**, *45* (W1), W36–W41.

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

(17) Minowa, Y.; Araki, M.; Kanehisa, M. Comprehensive Analysis of Distinctive Polyketide and Nonribosomal Peptide Structural Motifs Encoded in Microbial Genomes. *J. Mol. Biol.* **2007**, *368* (5), 1500–1517.

(18) Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. NRPSpredictor2--a Web Server for Predicting NRPS Adenylation Domain Specificity. *Nucleic Acids Res.* **2011**, *39* (suppl. 2), W362-7.

(19) Crüsemann, M.; Kohlhaas, C.; Piel, J. Evolution-Guided Engineering of Nonribosomal Peptide Synthetase Adenylation Domains. *Chem. Sci.* **2013**, *4* (3), 1041–1045.

(20) Kries, H.; Niquille, D. L.; Hilvert, D. A Subdomain Swap Strategy for Reengineering Nonribosomal Peptides. *Chem. Biol.* **2015**, *22* (5), 640–648.

(21) Shishido, T.; Kaasalainen, U.; Fewer, D. P.; Rouhiainen, L.; Jokela, J.; Wahlsten, M.; Fiore, M.; Yunes, J.; Rikkinen, J.; Sivonen, K. Convergent Evolution of [D-Leucine1] Microcystin-LR in Taxonomically Disparate Cyanobacteria. *BMC Evol. Biol.* **2013**, *13* (1), 86.

(22) Martin, D. P.; Murrell, B.; Golden, M.; Khoosal, A.; Muhire, B. RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes. *Virus Evol.* **2015**, *1* (1), vev003.

(23) Martin, D.; Rybicki, E. RDP: Detection of Recombination amongst Aligned Sequences. *Bioinformatics* **2000**, *16* (6), 562–563.

(24) Padidam, M.; Sawyer, S.; Fauquet, C. M. Possible Emergence of New Geminiviruses by Frequent Recombination. *Virology* **1999**, *265* (2), 218–225.

(25) SALMINEN, M. O.; CARR, J. K.; BURKE, D. S.; McCUTCHAN, F. E. Identification of Breakpoints in Intergenotypic Recombinants of HIV Type 1 by Bootscanning. *AIDS Res. Hum. Retroviruses* **1995**, *11* (11), 1423–1425.

(26) Smith, J. M. Analyzing the Mosaic Structure of Genes. *J. Mol. Evol.* **1992**, *34* (2), 126–129.

(27) Posada, D.; Crandall, K. A. Evaluation of Methods for Detecting Recombination from DNA Sequences: Computer Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (24), 13757–13762.

(28) Gibbs, M. J.; Armstrong, J. S.; Gibbs, A. J. Sister-Scanning: A Monte Carlo Procedure for Assessing Signals in Recombinant Sequences. *Bioinformatics* **2000**, *16* (7), 573–582.

(29) Boni, M. F.; Posada, D.; Feldman, M. W. An Exact Nonparametric Method for Inferring Mosaic Structure in Sequence Triplets. *Genetics* **2007**, *176* (2), 1035–1047.

(30) 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.* **2015**, *10* (6), 845–858.

(31) Tanovic, A.; Samel, S. A.; Essen, L.-O.; Marahiel, M. A. Crystal Structure of the Termination Module of a Nonribosomal Peptide Synthetase. *Science (80-.).* **2008**, *321*

(5889), 659–663.

(32) Jenke-Kodama, H.; Börner, T.; Dittmann, E. Natural Biocombinatorics in the Polyketide Synthase Genes of the Actinobacterium Streptomyces Avermitilis. *PLoS Comput. Biol.* **2006**, *2* (10), e132.

(33) McDonald, B. R.; Currie, C. R. Lateral Gene Transfer Dynamics in the Ancient Bacterial Genus *Streptomyces*. *MBio* **2017**, *8* (3), e00644-17.

(34) Kadi, N.; Challis, G. L. Chapter 17 Siderophore Biosynthesis. In *Methods in enzymology*; 2009; Vol. 458, pp 431–457.

(35) Fewer, D. P.; Rouhiainen, L.; Jokela, J.; Wahlsten, M.; Laakso, K.; Wang, H.; Sivonen, K. Recurrent Adenylation Domain Replacement in the Microcystin Synthetase Gene Cluster. *BMC Evol. Biol.* **2007**, *7* (1), 183.

(36) Oldenburg, K.; Vo, K. T.; Michaelis, S.; Paddon, C. Recombination-Mediated PCR-Directed Plasmid Construction in Vivo in Yeast. *Nucleic Acids Res.* **1997**, *25* (2), 451–452.

(37) Gietz, R. D.; Schiestl, R. H. Quick and Easy Yeast Transformation Using the LiAc/SS Carrier DNA/PEG Method. *Nat. Protoc.* **2007**, *2* (1), 35–37.

(38) Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A.; John Innes Foundation. *Practical Streptomyces Genetics*; John Innes Foundation, 2000.

(39) Son, S.; Hong, Y.-S.; Jang, M.; Heo, K. T.; Lee, B.; Jang, J.-P.; Kim, J.-W.; Ryoo, I.-J.; Kim, W.-G.; Ko, S.-K.; et al. Genomics-Driven Discovery of Chlorinated Cyclic Hexapeptides Ulleungmycins A and B from a Streptomyces Species. *J. Nat. Prod.* **2017**, *80* (11), 3025–3031.

(40) Shiba, T.; Mukunoki, Y. The Total Structure of the Antibiotic Longicatenamycin. *J. Antibiot. (Tokyo).* **1975**, *28* (8), 561–566.

Chapter Three: Synchronous Cycloadditions of the Macrolactam Polyketide Heronamide C

Note: This chapter is adapted from Booth et al., 2016[1]

Contents

3. 1. Introduction

3. 1. 1 Pericyclic Reactions in Nature

Pericyclic reactions are an important tool in synthetic chemistry and are vital in Nature. Pericyclic reactions involve the concerted rearrangement of bonds through an uncharged, cyclic transition state and include electrocyclisations, cycloadditions and sigmatropic rearrangements (Figure 3. 1). A diverse range of pericyclic reactions have been characterised and involve varying π-systems, which may be hetero- or homo-atomic[2–4]. Pericyclic mechanisms occur through either thermal or photochemical mechanisms and generate specific stereochemistries depending on which route is taken. In Nature, the majority of pericyclic reactions are believed to occur spontaneously. The biosynthesis of vitamin D is perhaps the most well know example. In a twostep process, exposure of 7-dehydrocholesterol to UV light catalyses an electrocyclic ring opening to produce previtamin D, which then isomerises through a thermal [1,7] sigmatropic hydride shift to produce vitamin D (Figure 3. 2)[5,6]. In this manner, vitamin D biosynthesis highlights the importance of both thermal and photochemical pericyclic reactions. Although much

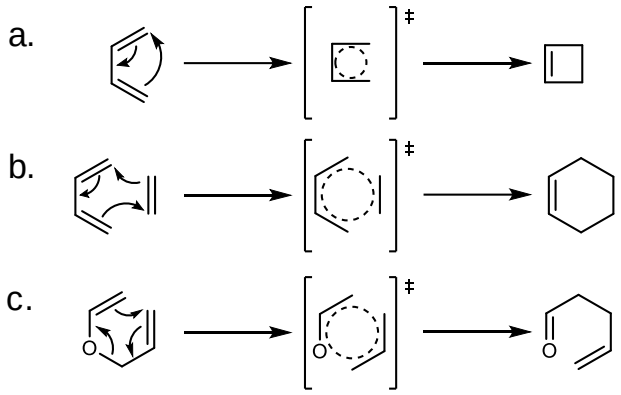


Figure 3. 23: Examples of pericyclic reactions and their transition states. a.) The electocyclisation of 1-3, butadiene to cyclobutene; b) Diels-Alder [4π + 2π] cycloaddition of 1,3 butadiene and ethylene to produce cyclohexene; and c.) the Claisen [3, 3] sigmatropic rearrangement of ethenyloxy-propene to 4-pentenal.

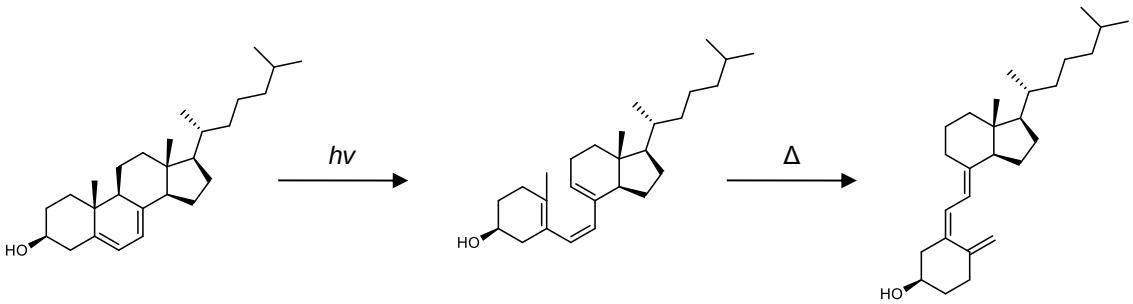


Figure 3. 24: Spontaneous conversion of provitamin D (7-dehydrocholesterol) to vitamin D. The exposure of provitamin D to UVB causes an electrocyclic ring opening yielding previtamin D. Previtamin D is subsequently converted to vitamin D via [1,7] hydride shift.

rarer, examples of enzymatically catalysed pericyclic reactions have been reported. The first example to be described was chorismate mutase, which catalyses the Claisen rearrangement of chorismate to prephenic acid during the biosynthesis of aromatic amino acids, enhancing the rate of this conversion by over one million fold[7,8]. Chorismate mutases vary substantially in structure, but all function through similar mechanisms of conformational control (Figure 3. 3)[9,10].

With regards to specialised metabolism, much effort has been focused on the [4π+2π] Diels-Alder cycloaddition. The biosynthesis lovastatin, spinosyn, and the spirotetramate and spirotetronate families (see Chapter 5) all involve enzymatically catalysed Diels-Alder reactions[11–16]. Although there has been much debate over whether these enzymes are true Diels-Alderases (i.e. whether or not the reaction is catalysed through a concerted rather than stepwise fashion – or lower the activation energy of reaction transition states)[17–19], structural and computational studies of SpnF and AbyU (of the spinosyn and abyssomycin pathways respectively) provide compelling evidence that the reaction proceeds through a concerted mechanism in these examples. Despite different evolutionary origins and structures, AbyU and SpnF both function by encapsulating their entire substrate and reducing enthalpy through the stabilisation of the reactive geometry and transition state through hydrogen-bonding and hydrophobic interations[13,16,20–23]. Other enzymatically catalysed putative pericyclic reactions include the aza [4π + 2π] (hetero-Diels-Alder) cycloaddition in the formation of the pyradine ring during thiopeptide biosynthesis[24,25] and [3,3]-retro-Claisen rearrangements in the formation of didehydropyan during leporin biosynthesis[26]. It is worth noting, that in some cases enzymatic catalysis is not strictly required as the reaction can occur spontaneously, as is the case with spinosyn. Some natural products are hypothesised to rely solely on spontaneous reactions, examples include include [4π + 2π] decalin ring formation in anthracimycin[27] and the spontaneous Claisen rearrangement of prenyl-tyrosine residues in cyanobactin[28].

In this chapter, I discuss the [6π + 4π] and [6π + 6π] intramolecular cycloadditions involved in the biosynthesis of the polyene macrolactams heronamide A (**5**) and B (**6**) (Figure 3. 4) and our efforts to determine the nature of this transformation.

3. 1. 2. The Heronamides

The heronamides A-C (**5**-**7**) are 20-membered polyene macrolactams produced by *Streptomyces sp.* CMB-0406 isolated from marine sediment collected off the coast of Heron Island, Australia[29]. Although **5** and **6** have no reported biological activity, **7** exhibits inhibitory activity against fission yeast. As with other macrocyclic polyketides, the heronamides have been shown to function through interaction with the membrane. Using surface plasmon resonance, Sugiyama *et al.*[30] demonstrated that **7** and its biosynthetic precursor (8-deoxy heronamide C)

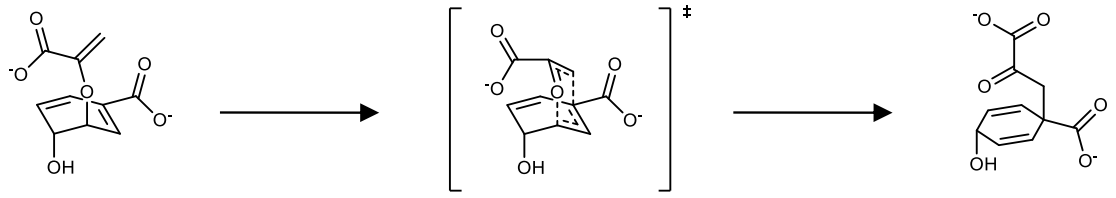


Figure 3. 25: Claisen rearrangement of chorismate to prephenic acid. This reaction is catalysed by chorismite mutase by conformational control; the transition state is shown.

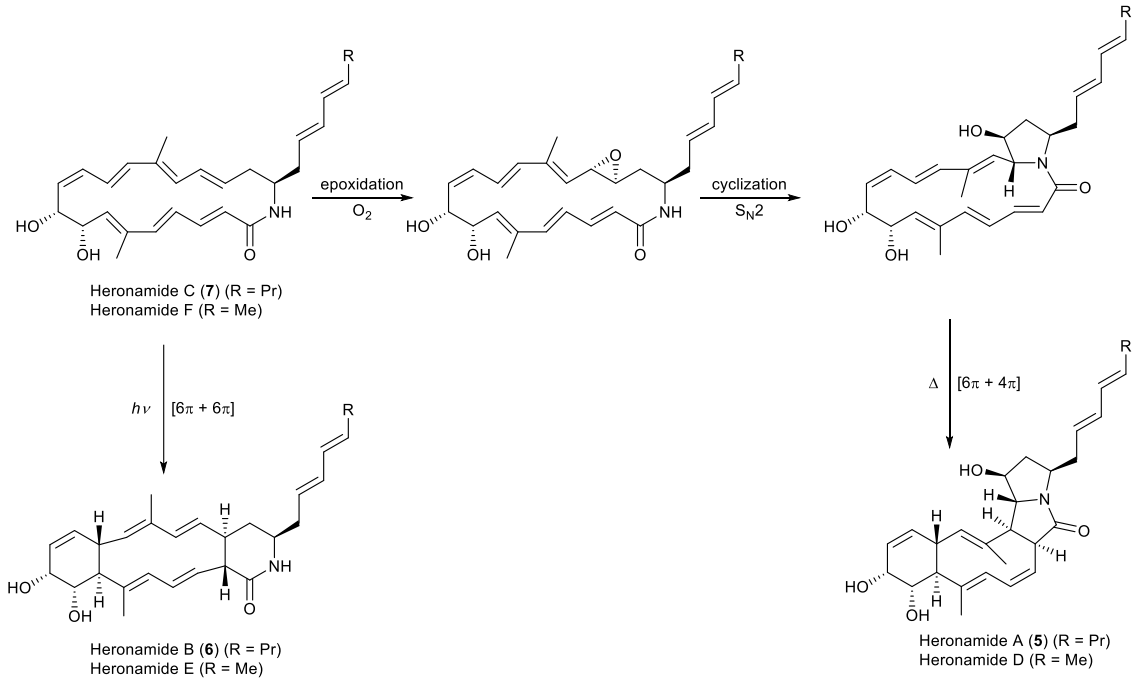


Figure 3. 26: Proposed biosynthetic pathways for the heronamides A/B (5/6) and D/E from heronamides C (7) and F. The mechanism was originally hypothesised by Raju et al. and is supported experimentally in this chapter.

irreversibly bind the lipids 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine and sphingomyelin, which are crucial components of the cell membrane. This results in the abnormal accumulation of cell wall material at both the cell tips and septa. This activity is dependent upon the activity of the β-glucan synthase (BgsI) and its activator, the small GTPase RhoI responsible for the formation of (1-3) β-D-glycosidic bonds during cell wall biogenesis[31,32]. In addition to their anti-fungal activity Raju et al.[29] reported that **7** elicited an unusual response from mammalian cells. HeLa cells grown on **7** (50 μM) formed large, vesicle-like structures that dissipated when it was removed from the media. A similar morphological phenomenon can be induced by the related macrolactam vicenistatin[33].

The heronamides are part of a family of related macrolactams that incorporate β-amino fatty acids as starter units. The family includes the heronamides A-F[29,34], BE-14106[35], ML-449[36], and the more distantly related cremimycin[37] (see Chapter Four for a complete description of β-amino acid containing macrolactams). The biosynthesis of two closely related polyketides, BE-14106, ML-449 and cremimycin have previously been described[36–38]. Polyketide synthesis for this family occurs in two stages. In the biosynthesis of ML-449, the closest structural relative of **7**, an initial three or four rounds of polyketide synthesis leads to the production of an unsaturated tetraketide, which is converted to a β-amino fatty acid through though Michael addition of glycine and subsequent hydrolytic cleavage of glyoxylate[39]. The β-amino fatty acid is then incorporated as the starter unit for the second round of polyketide synthesis to produce the final macrolactam scaffold[36,38]. Although the polyketide biosynthesis of ML-449 and BE-14106 has been characterised, cyclic congeners have never been reported.

Based on the structures of **5**-**7**, it was hypothesised that **7** may undergo unique, transannular intramolecular cycloadditions during the biosynthesis of **5** and **6** (Figure 3. 4). The biosynthesis of **1** is likely to involve an C15-C16 epoxidation of **3** and subsequent intramolecular S_N2 attack by the proximal amide nitrogen to yield an intermediate pyrrolidinol. It is hypothesised that this intermediate can then undergo [6π+4π] cycloaddition to yield **5**. Alternatively, **7** was hypothesised to undergo a [6π+6π] cycloaddition to produce **6**. Given the unique nature of these reactions and their lack of characterisation we sought to examine the production of the heronamides in detail and understand whether or not there was a biological basis for these reactions.

3. 2. Results and Discussion

3. 2. 1. Heronamide C Biosynthetic Gene Cluster (*hrn* BGC)

The genome sequence of *Streptomyces sp.* CMB-0406 was obtained *via* the Pacific Biosciences (PacBio) RSII sequencing platform[40]. The sequenced genome was submitted to

antiSMASH v3.0[41], which identified 25 putative BGCs. The putative heronamide BGC (*hrn*) was identified due to its similaritly to the BE-14106 producing *bec* BGC[38]. The *hrn* BGC spans 80166 base pairs and consists of 21 predicted genes (Figure 3. 5, Table 3. 1). Interestingly, the *hrn* BGC is more closely related to the *mla* and *bec* BGC than the heronamide D producing cluster *her* (published during the course of this study)[42]. In fact, the architecture of the *hrn* BGC was identical to that of *mla*.

ML-449 was reported to differ from **7** due to the presence of a C6-C7 *cis*-double bond. Given the similarity of the two BGCs, we were interested to explore the mechanism behind the altered stereochemistry and hypothesised this may prevent cyclisation. An alignment of predicted protein sequences for the ketoreductase and dehydratase domains of the PKS of **7**, ML-449 and other polyketides was generated. The alignment revealed that all *her*, *bec* and *hrn* KR domains were canonical B-type KR domains[43,44] containing the conserved LDD motif and P144 and N148 residues with the exception of the KR domain in module 4 of the macrolactam PKS responsible for the C10-C11 *cis*-double bond where the LDD sequence was replaced by VDN and the substitution of alanine at position 144; this was assigned as an A-type KR domain. The KR domain of module 6 in the *mla* BGC was 98.9% identical to its counterpart in the *hrn* BGC strongly indicating that the assignment of a C6-C7 cis-double bond most probably erroneous. This is supported by the reported NMR spectrum of BE-14106[35] and implies that BE-14106, ML-449 and the heronamides all share a common macrolactam ring structure. These means that ML-449 and **7** are in fact the same compound. In addition, this analysis supports the findings of Zhu et al.[42] who reported the absence of a 'shift module,'[45,46] in the heronamide F BSGC.

Due to advances in the understanding polyketide macrolactam biosynthesis since the publication of the *mla* BGC, we can provide a more accurate biosynthetic hypothesis for **7** then previously described (Figure 3. 6). Firstly, an 11 carbon tetraketide is synthesised by the KS_Q initiated HrnA1, HrnA2 and HrnC polyketide synthases. A β-amino group is introduced to through Michael addition of glycine by HrnU and subsequent cleavage of glyoxylate by HrnI leads to chain release[39]. The free β-amino fatty acid is then adenylated and transferred to the free ACP HrnS by the action of HrnJ. The amino group is then aminoacylated by the adenylation domain HrnL to protect from premature cyclisation of the macrolactam[39,47] prior to acyl transfer by HrnK to a second PKS complex (HrnBDEFG) for the remaining eight rounds of polyketide synthesis. Finally, the incorporated amino acid is deacylated by HrnP, allowing for macrolactamisation and release from the synthase machinery (see Chapter Four for further information on this mechanism).

Comparison of the *hrn* BGC to related polyene macrolactam BGCs did not highlight any unique genes that might be specific to the biosynthesis of **5** and **6** from **7**. Notably, no gene is present in any of the clusters whose product can be attributed to epoxidation of the C16–C17 double bond[36,38,42]. This was surprising as although polyenes are known to autoxidize in air[48] it

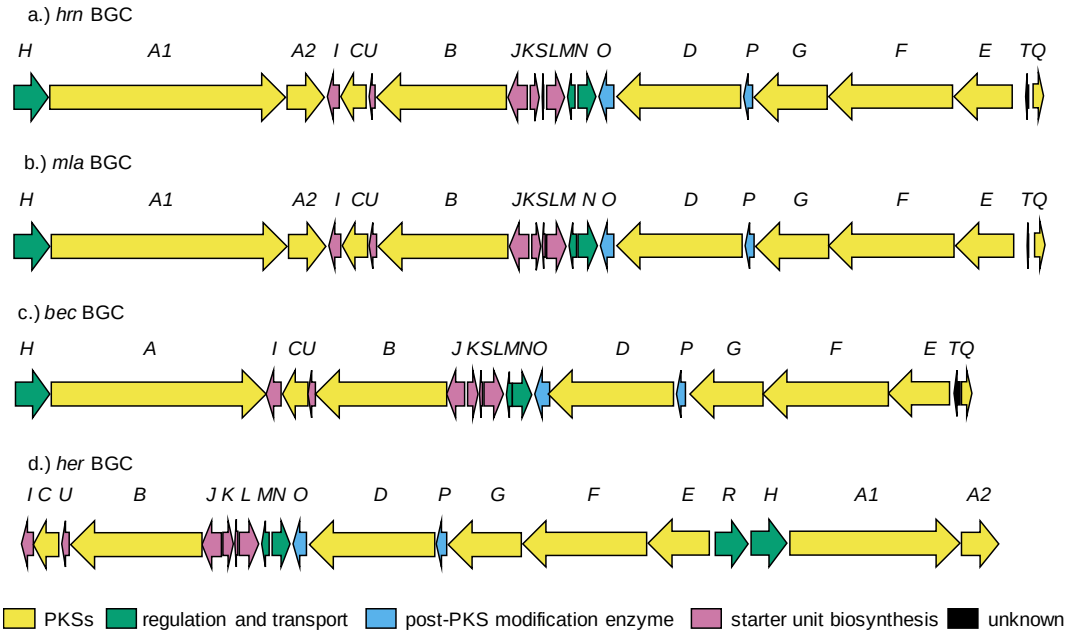


Figure 3. 27: Comparison of the heronamide family of BGCs. a.) hrn (heronamide C (**7**)) and b.) mla (ML-449) loci show identical overall topologies. c.) bec (BE-14106) lacks a portion of the A1 gene and has no A2 gene, reflected in the loss of a whole PKS module. d.) her (heronamide F) contains an additional regulatory gene, herR, and has a different overall topology to the other clusters in the family.

Protein	size (aa)	Proposed function	% Similarities/ % Identities Homolog from		
			mla BGC	*bec* BGC	*her* BGC
HrnH	950	LuxR-type transcriptional regulator	MlaH 98/95	BecH 94/88	HerH 74/64
HrnA1	6334	Polyketide synthase type I	MlaA1 96/94	BecA 90/86	HerA1 74/64
HrnA2	1043	Polyketide synthase type I	MlaA2 98/97	BecA 84/77	HerA2 83/76
HrnI	363	Glycine oxidase/FAD-dependent oxidoreductase	MlaI 95/93	BecI 88/79	HerI 83/73
HrnC	695	Polyketide synthase type I	MlaC 97/95	BecC 90/84	HerC 84/75
HrnU	187	putative NRPS accessory protein	MlaU 100/99	BecU 95/89	HerU 88/81
HrnB	3526	Polyketide synthase type I	MlaB 96/95	BecB 92/87	HerB 85/78
HrnJ	532	putative AMP-dependent acyl-CoA synthetase/ ligase	MlaJ 97/97	BecU 92/87	HerJ 87/79
HrnK	313	Acyltransferase	MlaK 97/96	BecK 90/84	HerK 85/77
HrnS	78	Peptidyl carrier protein	MlaS 98/95	BecS 94/92	HerS 91/82
HrnL	504	NRPS adenylation domain	MlaL 96/94	BecL 89/83	HerL 89/81
HrnM	199	TetR-type transcriptional regulator	MlaM 98/96	BecM 89/86	HerM 87/81
HrnN	523	MFS-type efflux pump	MlaN 97/95	BecN 93/87	HerN 85/77
HrnO	411	P450 monooxygenase	MlaO 98/96	BecO 95/91	HerO 87/76
HrnD	3368	Polyketide synthase type I	MlaD 96/95	BecD 92/88	HerD 83/77
HrnP	311	Putative L-amino acid amidase/ proline iminopeptidase	MlaG 98/98	BecP 91/83	HerP 90/83
HrnG	2001	Polyketide synthase type I	MlaG 95/94	BecG 89/85	HerG 80/72
HrnF	3382	Polyketide synthase type I	MlaF 97/94	BecF 93/88	HerF 84/76
HrnE	1637	Polyketide synthase type I	MlaE 95/93	BecE 91/87	HerE 81/72
HrnT	88	SimX-like protein	MlaT 92/90	BecT 80/75	---
HrnQ	256	Thioesterase type II	MlaQ 98/96	BecQ 93/86	---

Table 3. 8: Description of proteins encoded in genes identified in the heronamide biosynthetic gene cluster. Proposed function and homologs from heronamide family BGCs also shown.

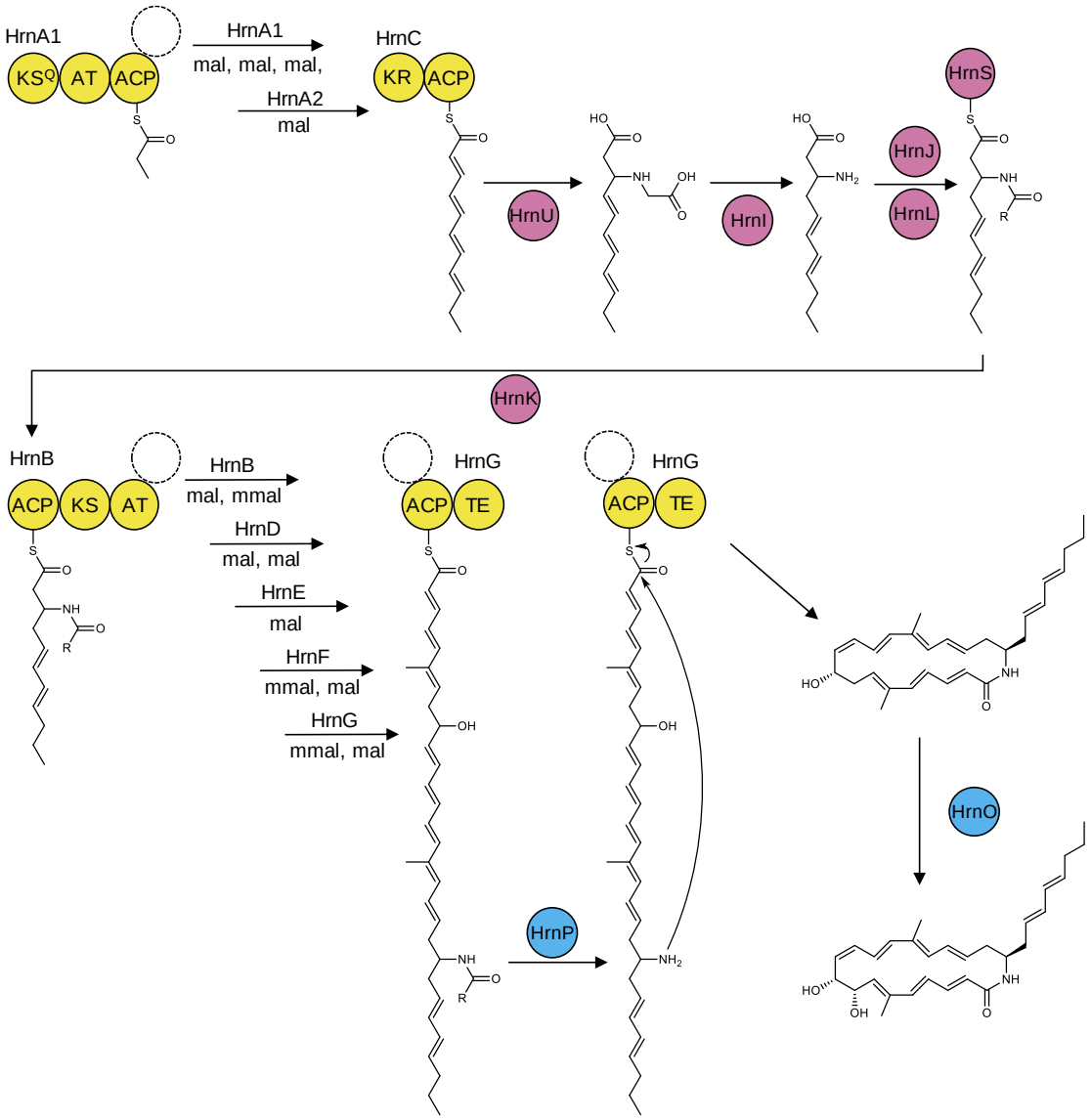


Figure 3. 28: Heronamide C Biosynthesis. Four rounds of polyketide synthesis followed by the Michael addition of glycine followed by release of the polyketide chain, hydrolytic cleavage and amino-acylation. It is important to note that an αβ-βγ double bond shift and reduction occur at some stage during this process, however the precise timing has yet to be elucidated. An additional nine rounds of polyketide synthesis, diacylation and macrocyclization lead to the release of 8-deoxy-**7**. For simplicity, PKS assembly lines are not shown in full however, malonyl-(mal) and methylmalonyl-(mmal) CoA substrates for each PKS are shown.

seemed unlikely that this might occur with both the regio- and stereo- selectivity required to produce **1**.

3. 2. 2. Heronamide Production by *Streptomyces sp.* CMB-0406

We hypothesised that the the regio- and stereoselective oxidation of the pyrrolidinol precursor was the product of biocatalysis. We performed a time course study following the production of **1**–**3** and dry cell mass over 10 days. Cultures of *Streptomyces sp.* CMB-0406 were grown in SV2 medium and the production of **5**, **6** and **7** were calculated by comparison of absorbance to a standard curve. Biosynthesis of **7** correlates with growth however, the production of **5** and **6** did not. **5** was produced steadily, whereas **6** was present in constant but low levels (Figure 3. 7).

To assess the possibility of the reactions being catalysed by other endogenous enzymes or physiological conditions, the strain was cultured in the presence of cyctochrome P450 inhibitors (menadione, clotrimazole and allylsulfide). In addition, we hypothesised that interspecies interactions may activate production and co-cultured *Streptomyces sp.* CMB-0406 with various Gram-positive and Gram-negative bacteria. In all these experiments, no change in the production **5** or **6** was observed relative to the production of **7**. These data along with the genomic sequencing results cast doubt on a biocatalytic route for the production of **5** or **6**.

3. 2. 3. Spontaneous Cycloadditions of Heronamide C

To test the spontaneous reactivity of **7** a reasonable quantity of pure compound was required. *Streptomyces* sp. CMB-0406 was cultured in SV2 medium (6 x 400 mL) for 10 days at 30°C. The resulting broth (2.4 L) was extracted with an equal volume of ethylacetate. The organic phase was separated and ethylacetate was evaporated yielding the crude extract (1.8 g). The crude extract was triturated with an equal volume of water and the insoluble fraction was collected. The insoluble fraction (67 mg) was fractionated by preparative HPLC yielding **5** (1.2 mg), **6** (0.3 mg) and **7** (4.8 mg).

3. 2. 3. 1. Thermally Induced Production of Heronamide A

Previous studies on the heronamides had reported spontaneous conversion of **7** to **5** and **6** in the presence of DMSO[42], however the reaction was poorly characterised. Given that the first step in heronamide A synthesis is oxidation by the introduction of the C15-C16 epoxide, aliquots of **7** (50µM) were incubated for 1 hour with varying concentrations (5 – 500 µM) of the oxidising agent meta-chloroperoxybenzoic acid (mCPBA) - a reagent commonly used for the epoxidation of alkenes[49]. **3** was very sensitive to oxidation under these conditions and multiple products were

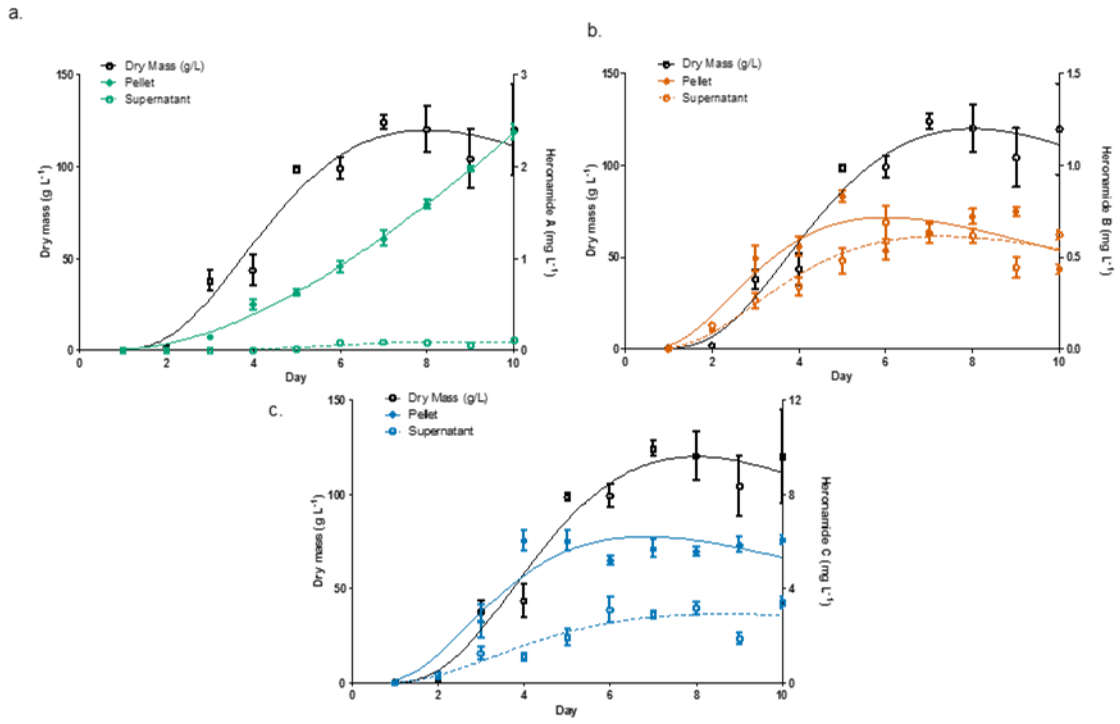


Figure 3. 29: Production of the heronamides by Streptomyces sp. CMB-0406. Dry cell mass and concentration of a.) heronamide A (**5**); b.) heronamide B (**6**) and c.) heronamide C (**7**) in the cell pellet and supernatant.

observed including **5** ($[M+H]^+$ = 466.29]) and other apparently oxidised products of **7** (e.g. $[M+H]^+$ = 482.28]).

Next, we sought to test milder oxidising conditions. Aliquots of **7** (225 mM) were dissolved in methanol with 0 – 50 % DMSO and exposed to air. Experiments were carried out at a range of temperatures (4°C, 30°C and 60°C). Following 7 days incubation samples were analysed using HPLC and LCMS. When exposed to air under ambient conditions production of **5** was observed at significantly higher levels, the most efficient conversion of **7** to **5** occurring in methanol containing 10% DMSO (Figure 3. 8). This may reflect the mild oxidizing conditions, or the accelerated rate of cycloaddition often observed in polar solvents[50,51]. Despite producing greater yields of **5**, incubation at higher temperatures lead to the production of numerous other products accounting for around 30% of the molar yield at 60°C. Although side products were formed, an excellent molar conversion ratio of **7** to **5** was observed, implying an autoxidation of **7** with remarkable facial and regio-selectivity at the C16–C17 double bond. Computational studies support this proposal and suggest a thermal reaction occurring through an ambimodal transition state that leads to a [4π+2π] and [6π+4π] adducts[52]. Interconversion of these adducts was predicted to occur via a facile Cope rearrangement leading to accumulation of the thermodynamically more stable [6π+4π] adduct (Figure 3. 9). Furthermore, the lack of any apparent pathway intermediates suggests the autoxidation step may be rate limiting. This would be consistent with *in silico* experiments that estimated the half-life of the pyrrolidinol precursor as approximately three minutes[52].

Since the publication of this result in 2016[1], other cyclic congeners of the heronamide-like macrocycle have been described. Ding et al.[53], isolated six new heronamides (G-L) from a 140 L culture of *Streptomyces niveus* YIM 32862 showing a range of oxidations and oxidative cyclisations, likely related to the compounds observed in the above experiments. Other heronamide A-like pyrrolidinol containing macrolactams have been identified with larger, 26-membered macrocycles: niizalactam A, dracolactam A and mirilactam C[54–56]. Interestingly, production of these compounds requires coculture of the producing organisms with mycolic acid containing bacterium (MACB). In each case however, feeding the precursor alone to MACB was not sufficient to induce oxidation and cyclisation suggesting that this process is not achieved through bioconversion.

Our experiments confirmed previous reports that the conversion of **7** to **5** may be spontaneous[57] and were replicated around the same time by Kanoh et al.[58] who reported production of **5** in a solution of dimethylformamide (DMF) and DMSO (1:1) when incubated at 50 °C for 7 days. **6** was not produced in any of the above experiments. As light was excluded during these experiments, we hypothesised that the conversion of **7** to **6** proceeds through an alternate route.

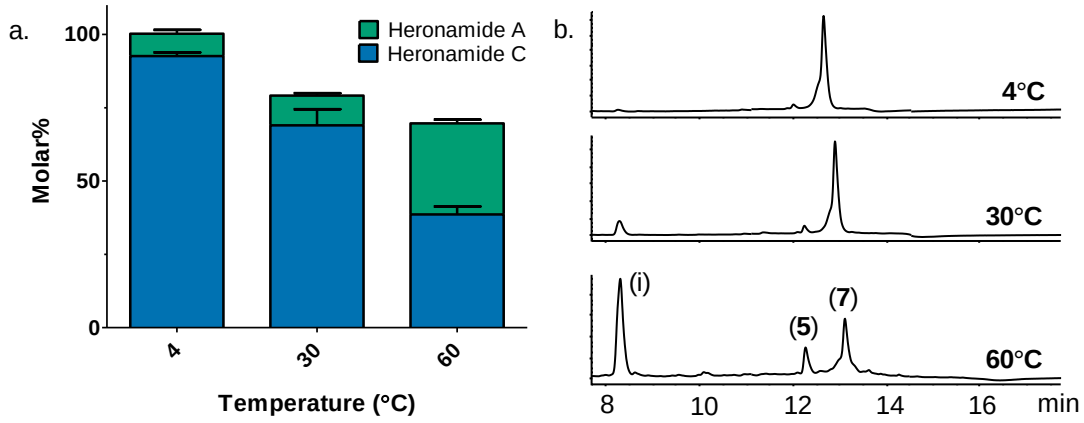


Figure 3. 30: Temperature sensitive conversion of heronamide C (7) to heronamide A (5) after seven days of incubation at 4, 30 and 60°C. a.) The proportion of **7** and **5** as a molar percentage of the starting material; and b.) a section of the HPLC trace at 330 nm showing **5**, **7** and the main decomposition product (i).

.

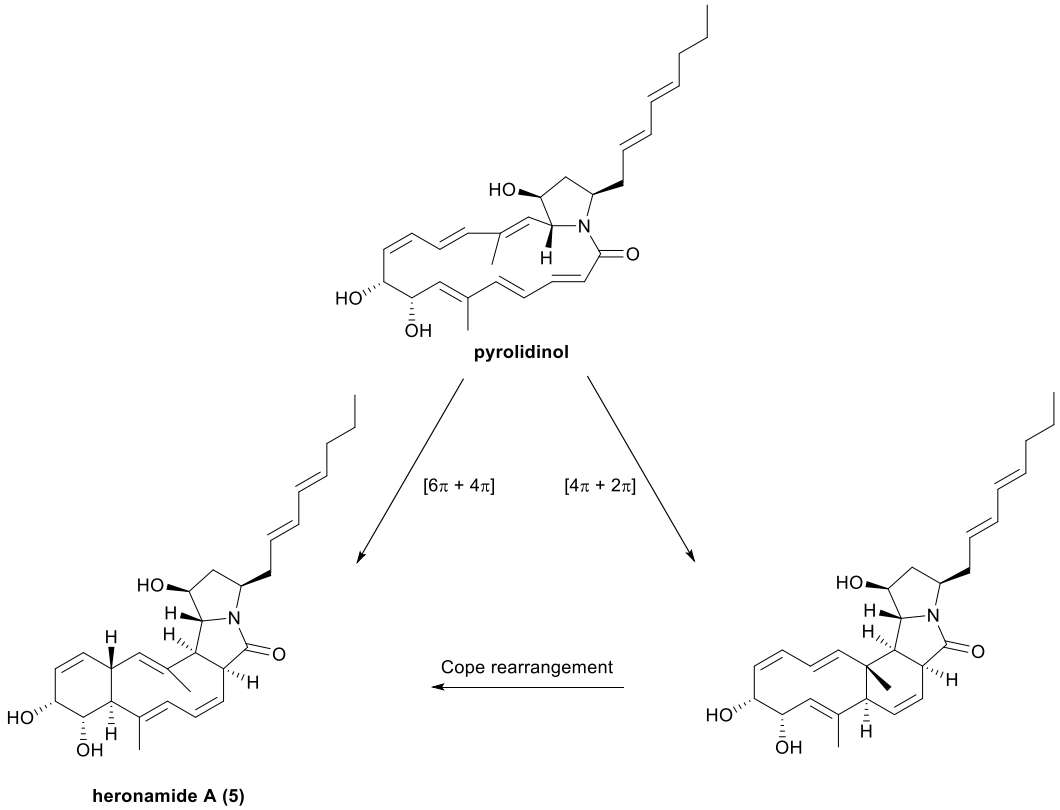


Figure 3. 31: Formation of heronamide A (5) from the hypothetical pyrolinidol via [6π + 4π] cycloaddition and cope rearrangement of the hypothetical [4π + 2π] adduct. Adapted from Yu et al., 2015.

3. 2. 3. 2. Photochemical Production of Heronamide B

Spontaneous cycloadditions follow either thermal or photochemical routes. According to Woodward-Hoffman rules, a [6π + 6π] cycloaddition is thermally forbidden as suprafacial ring closure can only occur through the photochemical route. This is reflected by the 2R, 7R 12S, 17R, stereochemistry of **6**. To test this hypothesis, purified **7** was dissolved in methanol (225 mM) and aliquoted into borosilicate vials to allow naturally occurring UV to pass through the samples. Vials were then incubated at RT in direct sunlight, or darkness for 1 h. The samples were analysed by HPLC-LCMS. The samples placed in direct sunlight showed conversion to **6** (approx. 60% molar conversion), whereas the samples excluded from light showed no conversion (Figure 3. 10).

To characterize this reaction further samples of **3** dissolved in methanol (225 mM) were exposed to several wavelengths of light (330, 375 and 405 nm) using a narrowband (ca. 5 nm) LED UV for several time intervals (10 to 160 s) (Figure 3. 11). Quantities of both **6** and **7** were calculated by comparison to a standard curve. The photochemical consumption of **3** follows first-order kinetics with significantly faster rates towards the ultraviolet range yielding **6** as the major product at all monitored wavelengths. The photon flux at each wavelength was calculated using ferrioxalate chemical actinometry to allow for normalization across wavelengths. From these data, the half-life, rate constant and quantum yields were calculated (Table 3. 2). $\Phi_{apparent}$ was the greatest at 330 nm, presumably because photons absorbed by **7** at this wavelength are more likely to be absorbed by the conjugated polyene resulting in isomerization. Prolonged exposure to shorter wavelengths led to the production of additional side products of identical *m/z* to **6** and **7** (Figure 3. 11.d).

These results demonstrate the wavelength-specific photochemical conversion of **7** to **6**. A similar [6π + 6π] cycloaddition has also been reported to transform the related macrolactam ciromicin A to ciromicin B in a similarly wavelength-specific manner[59]. Shortly after the publication of this work[1], photocatalysis of heronamide C was used in the total synthesis of heronamide B[58]. This allowed them to confirm the previously ambiguous stereochemistry of the compounds[30,60].

3. 3. Conclusion

We have shown that the conversion of heronamide C (**7**) to heronamide A (**5**) and heronamide B (**6**) occur *via* independent, spontaneous thermal [6π + 4π] or photochemical [6π + 6π] cycloadditions. The stereo- and regiospecificity of the olefin oxidation required during the biosynthesis of **5** is particularly noteworthy and may provide important insights for synthetic chemistry. Given the large number of polyene macrolactam natural products related to **7** which have been reported, we envisage that many additional compounds arising from similar reactions

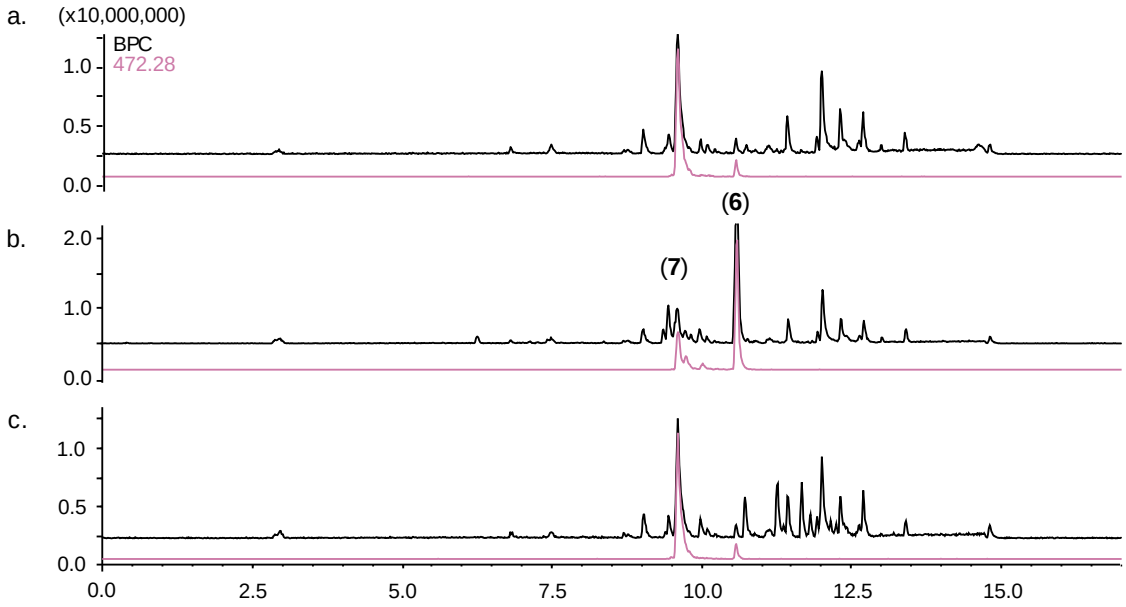


Figure 3. 32: Photochemical conversion of heronamide C (7) to heronamide B (6). LCMS spectra of a.) crude extracts of S. sp. CMB-0406 and b.) crude extracts incubated for 1 hour in: b.) direct sunlight and c.) darkness. The base peak chromatogram (BPC)(black) and the extracted ion chromatogram for $[M+Na]^+$=472.28 are shown for each condition.

Wavelength, λ (nm)	photons delivered to cuvette, q (mol s^{-1})	power delivered to cuvette, P (mW)	UV-Vis absorption 225 μM heronamide C, A	$1\text{-}10^{-A}$, χ	Time, t (s)	q χ t, $N_{photons}$ (mol)	Heronamide B, N_{mole} (mol)	$\Phi_{apparent}$
330	7.89E-09	2.86	0.95	0.89	10	7.00E-08	4.48E-08	0.6393
					20	1.40E-07	7.90E-08	0.5638
375	7.82E-08	24.93	0.34	0.54	10	4.25E-07	6.30E-09	0.0148
					20	8.50E-07	1.87E-08	0.0220
405	7.11E-07	210.02	0.31	0.51	10	3.63E-06	3.60E-09	0.0010
					20	7.26E-06	7.43E-09	0.0010

Table 3. 9: Photochemistry of heronamide C (7). Measurements were taken at three wavelengths. The rate constant (k) for the consumption of **7**, associated half-lives in addition to the actinometry results and the apparent quantum yields ($\Phi_{apparent}$) are presented.

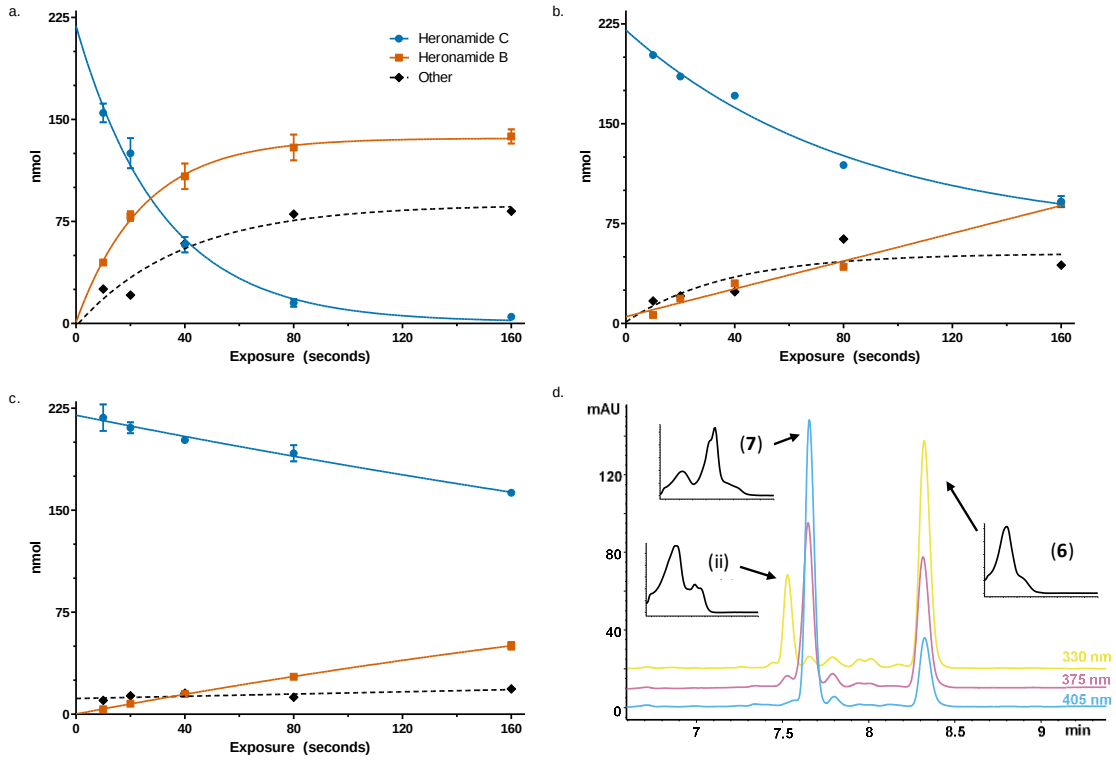


Figure 3. 33: Photochemical conversion of heronamide C to heronamide B and a putative isomer. The concentration of heronamide C (**7)** (blue) and heronamide B (**6**) (red) are plotted over time at: a.) 330nm, b.) 375 nm and, c.) 405 nm. The concentration of side products (grey) are inferred from the quantity of starting material and product. d.) HPLC chromatogram following exposure at the above wavelengths. Labeled are heronamide C (**7)** and heronamide B (**6)** and and unknown product (ii).

remain to be discovered. Indeed, deliberately exposing natural or synthetic polyene macrocycles to a range of oxidative and photochemical conditions may represent a route for expanding molecular diversity.

3. 4. Bibliography

(1) Booth, T. J.; Alt, S.; Capon, R. J.; Wilkinson, B. Synchronous Intramolecular Cycloadditions of the Polyene Macrolactam Polyketide Heronamide C. *Chem. Commun. (Camb).* **2016**, *52* (38), 6383–6386.

(2) Beaudry, C. M.; Malerich, J. P.; Trauner, D. Biosynthetic and Biomimetic Electrocyclizations. **2005**, 4757–4778.

(3) Bian, M.; Li, L.; Ding, H. Recent Advances on the Application of Electrocyclic Reactions in Complex Natural Product Synthesis. *Synthesis (Stuttg).* **2017**, *28* (19), 4383–4413.

(4) Walsh, C. T.; Tang, Y. Recent Advances in Enzymatic Complexity Generation: Cyclization Reactions. *Biochemistry* **2018**, *57* (22), 3087–3104.

(5) Holick, M. F.; Richtand, N. M.; McNeill, S. C.; Holick, S. A.; Frommer, J. E.; Henley, J. W.; Potts, J. T. Isolation and Identification of Previtamin D3 from the Skin of Rats Exposed to Ultraviolet Irradiation. *Biochemistry* **1979**, *18* (6), 1003–1008.

(6) Holick, M. F.; MacLaughlin, J. A.; Clark, M. B.; Holick, S. A.; Potts, J. T.; Anderson, R. R.; Blank, I. H.; Parrish, J. A.; Elias, P. Photosynthesis of Previtamin D3 in Human Skin and the Physiologic Consequences. *Science* **1980**, *210* (4466), 203–205.

(7) Goerisch, H. On the Mechanism of the Chorismate Mutase Reaction. *Biochemistry* **1978**, *17* (18), 3700–3705.

(8) Khanjin, N. A.; Snyder, J. P.; Menger, F. M. Mechanism of Chorismate Mutase: Contribution of Conformational Restriction to Catalysis in the Claisen Rearrangement. *J. Am. Chem. Soc.* **1999**, *121* (50), 11831–11846.

(9) Lee, A. Y.; Stewart, J. D.; Clardy, J.; Ganem, B. New Insight into the Catalytic Mechanism of Chorismate Mutases from Structural Studies. *Chem. Biol.* **1995**, *2* (4), 195–203.

(10) Lee, A. Y.; Karplus, P. A.; Ganem, B.; Clardy, J. Atomic Structure of the Buried Catalytic Pocket of Escherichia Coli Chorismate Mutase. *J. Am. Chem. Soc.* **1995**, *117* (12), 3627–3628.

(11) Auclair, K.; Sutherland, A.; Kennedy, J.; Witter, D. J.; Van den Heever, J. P.; Hutchinson, C. R.; Vederas, J. C. Lovastatin Nonaketide Synthase Catalyzes an Intramolecular Diels−Alder Reaction of a Substrate Analogue. *J. Am. Chem. Soc.* **2000**, *122* (46), 11519–11520.

(12) Kim, H. J.; Ruszczycky, M. W.; Choi, S.; Liu, Y.; Liu, H. Enzyme-Catalysed [4+2] Cycloaddition Is a Key Step in the Biosynthesis of Spinosyn A. *Nature* **2011**, *473* (7345), 109–112.

(13) Fage, C. D.; Isiorho, E. A.; Liu, Y.; Wagner, D. T.; Liu, H.; Keatinge-Clay, A. T. The Structure of SpnF, a Standalone Enzyme That Catalyzes [4 + 2] Cycloaddition. *Nat. Chem. Biol.* **2015**, *11* (4), 256–258.

(14) Hashimoto, T.; Hashimoto, J.; Teruya, K.; Hirano, T.; Shin-ya, K.; Ikeda, H.; Liu, H.; Nishiyama, M.; Kuzuyama, T. Biosynthesis of Versipelostatin: Identification of an Enzyme-Catalyzed [4+2]-Cycloaddition Required for Macrocyclization of Spirotetronate-Containing Polyketides. *J. Am. Chem. Soc.* **2015**, *137* (2), 572–575.

(15) Tian, Z.; Sun, P.; Yan, Y.; Wu, Z.; Zheng, Q.; Zhou, S.; Zhang, H.; Yu, F.; Jia, X.; Chen,

D.; et al. An Enzymatic [4+2] Cyclization Cascade Creates the Pentacyclic Core of Pyrroindomycins. *Nat. Chem. Biol.* **2015**, *11* (4), 259–265.

(16) Byrne, M. J.; Lees, N. R.; Han, L.-C.; van der Kamp, M. W.; Mulholland, A. J.; Stach, J. E. M.; Willis, C. L.; Race, P. R. The Catalytic Mechanism of a Natural Diels–Alderase Revealed in Molecular Detail. *J. Am. Chem. Soc.* **2016**, *138* (19), 6095–6098.

(17) Guimarães, C. R. W.; Udier-Blagović, M.; Jorgensen, W. L. Macrophomate Synthase: QM/MM Simulations Address the Diels−Alder versus Michael−Aldol Reaction Mechanism. *J. Am. Chem. Soc.* **2005**, *127* (10), 3577–3588.

(18) Kelly, W. L. Intramolecular Cyclizations of Polyketide Biosynthesis: Mining for a "Diels–Alderase"? *Org. Biomol. Chem.* **2008**, *6* (24), 4483.

(19) Townsend, C. A. A "Diels-Alderase" at Last. *Chembiochem* **2011**, *12* (15), 2267–2269.

(20) Hess, B. A.; Smentek, L. Concerted, Highly Asynchronous, Enzyme-Catalyzed [4 + 2] Cycloaddition in the Biosynthesis of Spinosyn A; Computational Evidence. *Org. Biomol. Chem.* **2012**, *10* (37), 7503.

(21) Zheng, Y.; Thiel, W. Computational Insights into an Enzyme-Catalyzed [4+2] Cycloaddition. *J. Org. Chem.* **2017**, *82* (24), 13563–13571.

(22) Jeon, B.-S.; Ruszczycky, M. W.; Russell, W. K.; Lin, G.-M.; Kim, N.; Choi, S.-H.; Wang, S.-A.; Liu, Y.-N.; Patrick, J. W.; Russell, D. H.; et al. Investigation of the Mechanism of the SpnF-Catalyzed [4+2]-Cycloaddition Reaction in the Biosynthesis of Spinosyn A. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (39), 10408–10413.

(23) Yang, Z.; Yang, S.; Yu, P.; Li, Y.; Doubleday, C.; Park, J.; Patel, A.; Jeon, B.-S.; Russell, W. K.; Liu, H.-W.; et al. Influence of Water and Enzyme SpnF on the Dynamics and Energetics of the Ambimodal [6+4]/[4+2] Cycloaddition. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (5), E848–E855.

(24) Bowers, A. A.; Walsh, C. T.; Acker, M. G. Genetic Interception and Structural Characterization of Thiopeptide Cyclization Precursors from *Bacillus Cereus*. *J. Am. Chem. Soc.* **2010**, *132* (35), 12182–12184.

(25) Wever, W. J.; Bogart, J. W.; Baccile, J. A.; Chan, A. N.; Schroeder, F. C.; Bowers, A. A. Chemoenzymatic Synthesis of Thiazolyl Peptide Natural Products Featuring an Enzyme-Catalyzed Formal [4 + 2] Cycloaddition. *J. Am. Chem. Soc.* **2015**, *137* (10), 3494–3497.

(26) Ohashi, M.; Liu, F.; Hai, Y.; Chen, M.; Tang, M.; Yang, Z.; Sato, M.; Watanabe, K.; Houk, K. N.; Tang, Y. SAM-Dependent Enzyme-Catalysed Pericyclic Reactions in Natural Product Biosynthesis. *Nature* **2017**, *549* (7673), 502–506.

(27) Alt, S.; Wilkinson, B. Biosynthesis of the Novel Macrolide Antibiotic Anthracimycin. *ACS Chem. Biol.* **2015**, *10* (11), 2468–2479.

(28) McIntosh, J. A.; Donia, M. S.; Nair, S. K.; Schmidt, E. W. Enzymatic Basis of Ribosomal Peptide Prenylation in Cyanobacteria. *J. Am. Chem. Soc.* **2011**, *133* (34), 13698–13705.

(29) Raju, R.; Piggott, A. M.; Conte, M. M.; Capon, R. J. Heronamides A–C, New Polyketide Macrolactams from an Australian Marine-Derived Streptomyces Sp. A Biosynthetic Case for Synchronized Tandem Electrocyclization. *Org. Biomol. Chem.* **2010**, *8* (20), 4682.

(30) Sugiyama, R.; Nishimura, S.; Matsumori, N.; Tsunematsu, Y.; Hattori, A.; Kakeya, H. Structure and Biological Activity of 8-Deoxyheronamide C from a Marine-Derived *Streptomyces* sp.: Heronamides Target Saturated Hydrocarbon Chains in Lipid Membranes. *J. Am. Chem. Soc.* **2014**, *136* (14), 5209–5212.

(31) Arellano, M.; Durán, A.; Pérez, P. Rho 1 GTPase Activates the (1-3)Beta-D-Glucan Synthase and Is Involved in Schizosaccharomyces Pombe Morphogenesis. *EMBO J.* **1996**, *15* (17), 4584–4591.

(32) Nakano, K.; Arai, R.; Mabuchi, I. The Small GTP-Binding Protein Rho1 Is a

Multifunctional Protein That Regulates Actin Localization, Cell Polarity, and Septum Formation in the Fission Yeast Schizosaccharomyces pombe. *Genes Cells* **1997**, *2* (11), 679–694.

(33) Nishiyama, Y.; Ohmichi, T.; Kazami, S.; Iwasaki, H.; Mano, K.; Nagumo, Y.; Kudo, F.; Ichikawa, S.; Iwabuchi, Y.; Kanoh, N.; et al. Vicenistatin Induces Early Endosome-Derived Vacuole Formation in Mammalian Cells. *Biosci. Biotechnol. Biochem.* **2016**, *80* (5), 902–910.

(34) Zhang, W.; Li, S.; Zhu, Y.; Chen, Y.; Chen, Y.; Zhang, H.; Zhang, G.; Tian, X.; Pan, Y.; Zhang, S.; et al. Heronamides D-F, Polyketide Macrolactams from the Deep-Sea-Derived Streptomyces Sp. SCSIO 03032. *J. Nat. Prod.* **2014**, *77* (2), 388–391.

(35) Kojiri, K.; Nakajima, S.; Suzuki, H.; Kondo, H.; Suda, H. A New Macrocyclic Lactam Antibiotic, BE-14106. I. Taxonomy, Isolation, Biological Activity and Structural Elucidation. *J. Antibiot. (Tokyo).* **1992**, *45* (6), 868–874.

(36) Jørgensen, H.; Degnes, K. F.; Dikiy, A.; Fjærvik, E.; Klinkenberg, G.; Zotchev, S. B. Insights into the Evolution of Macrolactam Biosynthesis through Cloning and Comparative Analysis of the Biosynthetic Gene Cluster for a Novel Macrocyclic Lactam, ML-449. *Appl. Environ. Microbiol.* **2010**, *76* (1), 283–293.

(37) Amagai, K.; Takaku, R.; Kudo, F.; Eguchi, T. A Unique Amino Transfer Mechanism for Constructing the β-Amino Fatty Acid Starter Unit in the Biosynthesis of the Macrolactam Antibiotic Cremimycin. *ChemBioChem* **2013**, *14* (15), 1998–2006.

(38) Jørgensen, H.; Degnes, K. F.; Sletta, H.; Fjærvik, E.; Dikiy, A.; Herfindal, L.; Bruheim, P.; Klinkenberg, G.; Bredholt, H.; Nygård, G.; et al. Biosynthesis of Macrolactam BE-14106 Involves Two Distinct PKS Systems and Amino Acid Processing Enzymes for Generation of the Aminoacyl Starter Unit. **2009**, *16* (10), 1109–1121.

(39) Amagai, K.; Takaku, R.; Kudo, F.; Eguchi, T. A Unique Amino Transfer Mechanism for Constructing the β-Amino Fatty Acid Starter Unit in the Biosynthesis of the Macrolactam Antibiotic Cremimycin. *ChemBioChem* **2013**, *14* (15), 1998–2006.

(40) Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **2009**, *323* (5910), 133–138.

(41) Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Muller, R.; Wohlleben, W.; et al. AntiSMASH 3.0: A Comprehensive Resource for the Genome Mining of Biosynthetic Gene Clusters. *Nucleic Acids Res.* **2015**, *43* (W1), W237–W243.

(42) Zhu, Y.; Zhang, W.; Chen, Y.; Yuan, C.; Zhang, H.; Zhang, G.; Ma, L.; Zhang, Q.; Tian, X.; Zhang, S.; et al. Characterization of Heronamide Biosynthesis Reveals a Tailoring Hydroxylase and Indicates Migrated Double Bonds. *ChemBioChem* **2015**, *16* (14), 2086–2093.

(43) Caffrey, P. The Stereochemistry of Ketoreduction. *Chem. Biol.* **2005**, *12* (10), 1060–1062.

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

(45) Kusebauch, B.; Busch, B.; Scherlach, K.; Roth, M.; Hertweck, C. Functionally Distinct Modules Operate Two Consecutive α,B→β,γ Double-Bond Shifts in the Rhizoxin Polyketide Assembly Line. *Angew. Chemie Int. Ed.* **2010**, *49* (8), 1460–1464.

(46) Lohr, F.; Jenniches, I.; Frizler, M.; Meehan, M. J.; Sylvester, M.; Schmitz, A.; Gütschow, M.; Dorrestein, P. C.; König, G. M.; Schäberle, T. F. α,β → β,γ Double Bond Migration in Corallopyronin A Biosynthesis. *Chem. Sci.* **2013**, *4* (11), 4175.

(47) Shinohara, Y.; Kudo, F.; Eguchi, T. A Natural Protecting Group Strategy to Carry an Amino Acid Starter Unit in the Biosynthesis of Macrolactam Polyketide Antibiotics. *J. Am. Chem. Soc.* **2011**, *133* (45), 18134–18137.

(48) Rickards, R. W.; Smith, R. M.; Golding, B. T. Macrolide Antibiotic Studies. XV. *J. Antibiot. (Tokyo).* **1970**, *23* (12), 603–612.

(49) Kim, C.; Traylor, T. G. and; Perrin, C. L. MCPBA Epoxidation of Alkenes: Reinvestigation of Correlation between Rate and Ionization Potential. *J. Am. Chem. Soc.* **1998**, *120* (37), 9513–9516.

(50) Breslow, R.; Maitra, U.; Rideout, D. Selective Diels-Alder Reactions in Aqueous Solutions and Suspensions. *Tetrahedron Lett.* **1983**, *24* (18), 1901–1904.

(51) Breslow, R.; Guo, T. Diels-Alder Reactions in Nonaqueous Polar Solvents. Kinetic Effects of Chaotropic and Antichaotropic Agents and of .Beta.-Cyclodextrin. *J. Am. Chem. Soc.* **1988**, *110* (17), 5613–5617.

(52) Yu, P.; Patel, A.; Houk, K. N. Transannular [6 + 4] and Ambimodal Cycloaddition in the Biosynthesis of Heronamide A. *J. Am. Chem. Soc.* **2015**, *137* (42), 13518–13523.

(53) Ding, N.; Han, L.; Jiang, Y.; Li, G.; Zheng, Z.; Cao, B.; Guan, P.; Mu, Y.; Lin, B.; Huang, X. Heronamides G–L, Polyene Macrolactams from *Streptomyces niveus*. *RSC Adv.* **2018**, *8* (31), 17121–17131.

(54) Hoshino, S.; Okada, M.; Wakimoto, T.; Zhang, H.; Hayashi, F.; Onaka, H.; Abe, I. Niizalactams A–C, Multicyclic Macrolactams Isolated from Combined Culture of *Streptomyces* with Mycolic Acid-Containing Bacterium. *J. Nat. Prod.* **2015**, *78* (12), 3011–3017.

(55) Hoshino, S.; Okada, M.; Awakawa, T.; Asamizu, S.; Onaka, H.; Abe, I. Mycolic Acid Containing Bacterium Stimulates Tandem Cyclization of Polyene Macrolactam in a Lake Sediment Derived Rare Actinomycete. *Org. Lett.* **2017**, *19* (18), 4992–4995.

(56) Hoshino, S.; Ozeki, M.; Wong, C. P.; Zhang, H.; Hayashi, F.; Awakawa, T.; Morita, H.; Onaka, H.; Abe, I. Mirilactams C–E, Novel Polycyclic Macrolactams Isolated from Combined-Culture of Actinosynnema mirum; NBRC 14064 and Mycolic Acid-Containing Bacterium. *Chem. Pharm. Bull.* **2018**, *66* (6), 660–667.

(57) Zhu, Y.; Zhang, W.; Chen, Y.; Yuan, C.; Zhang, H.; Zhang, G.; Ma, L.; Zhang, Q.; Tian, X.; Zhang, S.; et al. Characterization of Heronamide Biosynthesis Reveals a Tailoring Hydroxylase and Indicates Migrated Double Bonds. *Chembiochem* **2015**, *16* (14), 2086–2093.

(58) Kanoh, N.; Itoh, S.; Fujita, K.; Sakanishi, K.; Sugiyama, R.; Terajima, Y.; Iwabuchi, Y.; Nishimura, S.; Kakeya, H. Asymmetric Total Synthesis of Heronamides A-C: Stereochemical Confirmation and Impact of Long-Range Stereochemical Communication on the Biological Activity. *Chemistry* **2016**, *22* (25), 8586–8595.

(59) Derewacz, D. K.; Covington, B. C.; McLean, J. A.; Bachmann, B. O. Mapping Microbial Response Metabolomes for Induced Natural Product Discovery. *ACS Chem. Biol.* **2015**, *10* (9), 1998–2006.

(60) Sakanishi, K.; Itoh, S.; Sugiyama, R.; Nishimura, S.; Kakeya, H.; Iwabuchi, Y.; Kanoh, N. Total Synthesis of the Proposed Structure of Heronamide C. *Eur. J. Org. Chem.* **2014**, *2014* (7), 1376–1380.

(61) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23* (21), 2947–2948.

Chapter Four: A Roadmap for Macrolactam Discovery

Contents

4. 1. Introduction

4. 1. 1 β-amino acid containing polyketide macrolactams

The heronamides discussed in the previous chapter are members of a larger family of β-amino acid containing polyketide macrolactams (βPMs). The first βPM to be discovered was Stubomycin, described in 1981 by researchers at the Kitasato Institute in Tokyo, Japan[1]. A year later, the structure was solved under the now more commonly used name hitachimycin[2]. Based on the structure, it was proposed that hitachimycin had an unusual biosynthesis involving the incorporation of a phenylalanine. Meanwhile studies on its bioactivity revealed cytotoxic activity against mammalian and fungal cells – including antitumor activity [2–4]. In the proceeding three and a half decades, the βPM family has expanded to include over twenty members (Table 4. 1), which exhibit a range of biological activities. βPMs are distributed throughout the actinobacteria and have been observed in the genera *Actinosynnema*, *Micromonospora*, *Nocardiopsis*, *Salinospora* and *Streptomyces* (Table 4. 1). In addition to their biological activities, there has been much interest in their biosynthesis, both the incorporation of β-amino acid starter unit into the polyketide chain and generation of various cyclised congeners typified by the [6π + 4π] and [6π + 6π] cycloadditions described in Chapter 3. As such, βPMs present a natural reservoir of compounds that can provide valuable insight into polyketide biosynthesis and polyene chemistry. This chapter explores our efforts to expand and consolidate existing knowledge on βPMs through genome mining, and genomic and evolutionary analysis.

4. 1. 1. 1 Biosynthesis

4. 1. 1. 1. 1 Incorporation of the β-Amino Acid Starter Unit and Polyketide Synthesis

βPMs are characterised by their unique biosynthesis whereby a β-amino acid is incorporated as a starter unit of the polyketide chain. Miyagana et al.[5] describe five types of βPMs categorised by the β-amino acid utilised in their biosynthesis, they are: dicarboxylate-type (derived from L-aspartate and L-methyl-*threo*-3-aspartate), short-chain fatty acid-type (3-aminobutyrate), aromatic-type (β-phenylalanine) and the saturated and unsaturated medium-chain fatty acid-types (Figure 4. 1).

The first biosynthetic step is the generation of the starter unit, which typically involves the isomerisation of an α-amino acid precursor through a biosynthetic gene cluster (BGC) specific mutase. In hitachimycin biosynthesis, *hitA* encodes a L-phenylalanine-2,3-aminomutase capable of isomerising readily available L-phenylalanine into L-β-phenylalanine[6]. Both L-methyl-*threo*-3-aspartate and 3-amino butyrate are derived from L-glutamate, either through isomeration of the carboxyl-group (to produce methyl-aspartate)[7] or the amino group (to produce β-glutamate)[6,8].

Compound	Producer	Amino-sugar	β-amino acid	Biological Activity
Aureoverticillactam	*Streptomyces aureoverticillatus* NPS001583	Y	β-amino fatty acid	Cytotoxic
BE-14106	*Streptomyces* sp. A14106	N	β-amino fatty acid	Cytotoxic
Ciromicin A	*Nocardiopsis* sp. FU40	Y	3-methylaspartate	Cytotoxic
Cremimycin	*Streptomyces* sp. MJ635-86F5	N	β-amino fatty acid	Cytotoxic
Cyclamenol	*Streptomyces* sp. MHW 846	N	3-methylaspartate	Antiinflamatory
Fluvirucin B	*Actinomedura* spp.	Y	L-aspartate	Inhibits influenza infection
Heronamide C	*Streptomyces* sp. CMB-0406	N	β-amino fatty acid	Anti-fungal
Heronamide F	*Streptomyces* sp. SCSIO 0303	N	β-amino fatty acid	No reported activity
Hitachimycin	*Streptomyces* sp. KG-224	N	β-phenylalanine	Cytotoxic/antitumour
Incednine	*Streptomyces* sp. ML649-90F3	Y	3-aminobutyrate	Sensitises tumour cells
Lobosamide A	*Micromonospora* sp. RL09-050-HVF-A	N	3-aminobutyrate	Antitrypansomal
Nivelactam	*Streptomyces niveus* YIM 32860	N	β-amino fatty acid	Cytotoxic
Macrotermycin B	*Amycolatopsis* sp. M39	Y	3-methylaspartate	Anti-Gram-positive/antifungal
Micromonolactam	*Micromonospora* spp. CMS I1-30 and CMS I2-32.	N	3-aminobutyrate	No reported activity
Micromonosporin	*Micromonospora* TT1-11	N	3-aminobutyrate	No reported activity
Mirilactam A	*Actinosynnema mirum* ATCC 29888	N	3-aminobutyrate	No reported activity
Piceamycin	*Streptomyces* sp. AcH 50	N	3-methylaspartate	Anti-Gram-positve/antitumour
Salinilactam A	*Salinispora tropica* CNB 440	N	3-aminobutyrate	No reported activity
Sceliphrolactam	*Streptomyces* sp. C113 and C122	N	3-methylaspartate	Antifungal
Silvalactam	*Streptomyces* sp. Tü 6392.	Y	3-aminobutyrate	Anti-Gram-positive
Sipanmycin	*Streptomyces* sp. CS149	Y	3-aminobutyrate	Cytotoxic
Vicenistatin	*Streptomyces halstedii* HC34	Y	3-methylaspartate	Cytotoxic
Viridenomycin	*Streptomyces viridochromogenes* T-24146.	N	β-phenylalanine	Antiprotozoal/anti-gram-positive

Table 4. 10: Known β-amino acid containing polyketide macrolactams (βPMs), their structural features and bioactivities.

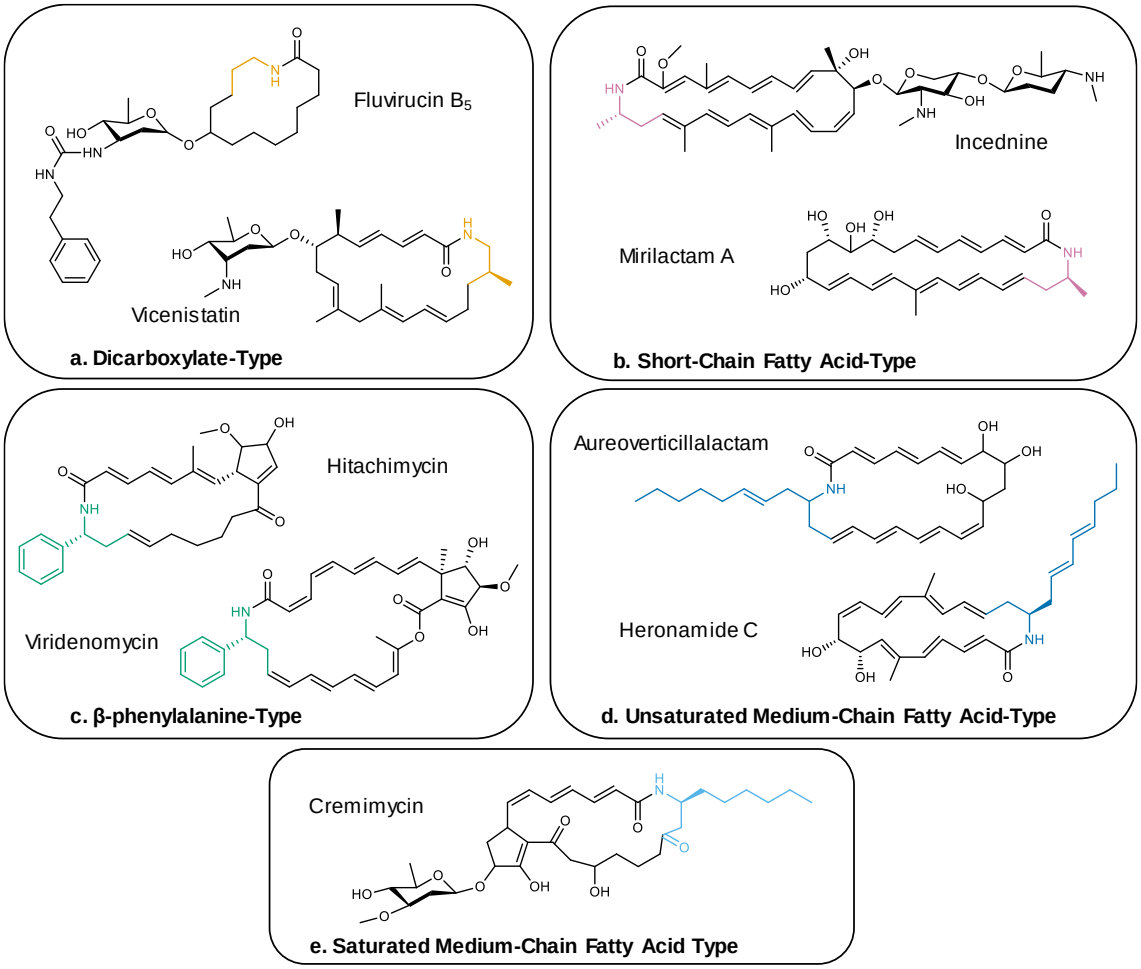


Figure 4. 34: Structural diversity of polyketide macrolactams. The five groups described by Miyagana et al[5] are shown. The region of the compound comprised of the amino acid is highlighted.

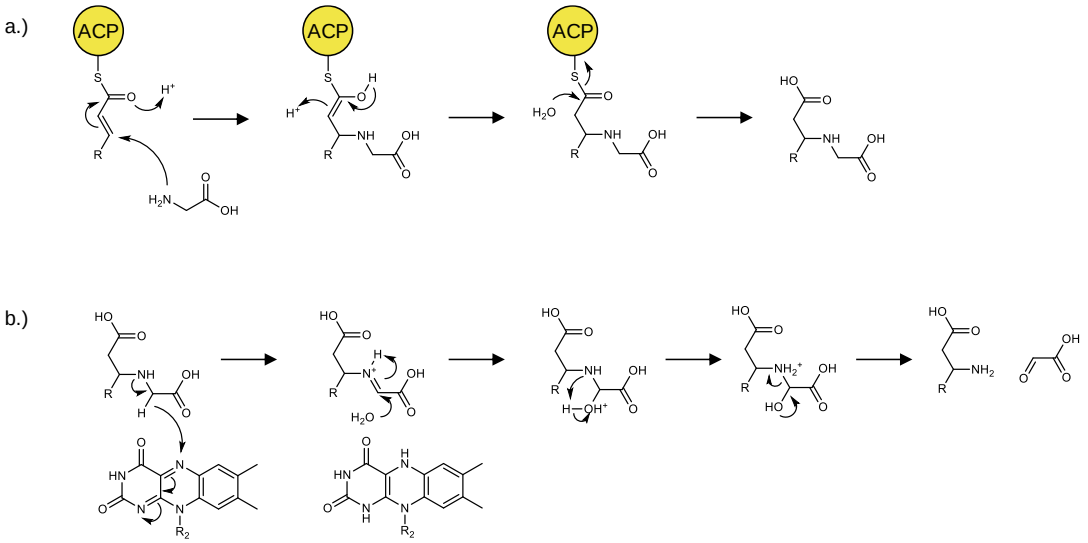


Figure 4. 35: Proposed mechanism for the generation of β-amino fatty acids in the biosynthesis if β-amino acid containing macrolactams (βPMs). Showing the mechanism of a.) CmiS1-like thioesterases and b.) CmiS2-like FAD-dependent oxidase as suggested by Amagai et al[11].

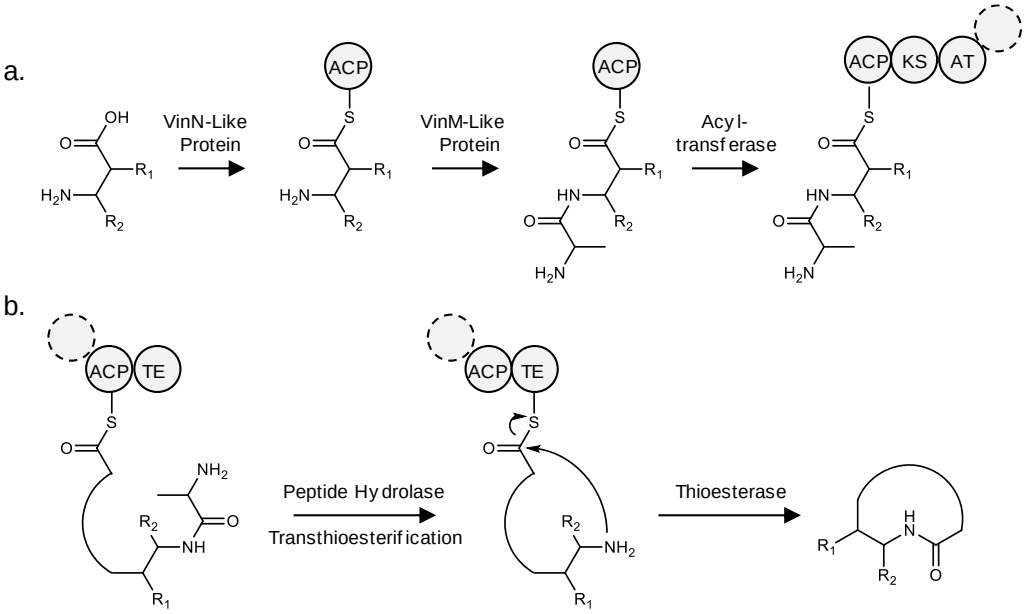


Figure 4. 36: General scheme for polyketide macrolactam biosynthesis. a.) Firstly, the VinN-like adenylation protein will transfer the selected amino acid to a standalone acyl-carrier protein (ACP). Peptidyl-ACP is then amino-acylated by a second adenylation domain protein (VinM-like). A standalone acyl transferase transfers the dipeptide to the polyketide synthase. b.) Following polyketide biosynthesis, alanine is liberated by hydrolysis allowing for macrocyclization.

Both methyl-aspartate and β-glutamate will be decarboxylated prior to incorporation, however this may happen later in the biosynthesis[6]. In the biosynthesis of medium chain fatty acid starter units, a linear fatty acid is generated by a set of type-I polyketide synthases (PKSs) encoded within the BGC[9,10]. Although saturation of the fatty acid is variable, the final PKS module will exclusively process the β-keto group to yield an α-β double bond. Michael addition of glycine and subsequent hydrolysis by a conserved thioesterase (CmiS1-like protein) and FAD-dependent oxidase (CmiS2-like protein) generates the β-amino group (Figure 4. 2)[11,12].

The mechanism for utilising the β-amino acid as the starter unit is highly conserved throughout known biosynthetic pathways. Five proteins are required for the incorporation of the β-amino acid into the final macrolactam (Figure 4. 3). The first step is the loading of the β-amino acid to a free ACP by an adenylation domain protein. As with α-amino acid adenylation domains, the specificity of these enzymes is dictated by the ten residues that constitute the substrate binding pocket[13]. Two of these residues are completely conserved, Asp230 for recognition of the amino group, giving high selectivity for β-amino acids and Lys510 that is hypothesised to interact with the α-carboxyl group of the amino acid and the ribose moiety of adenylate. As with α-amino acid adenylating proteins, binding pocket residues can provide insight into substrate specificity. For example, in dicarboxylate adenylation enzymes such as VinN or CirA2, K330 and R331 form salt-bridges with the C1-carboxyl group[13,14]. Where other substrates are accepted, positions 330 and 331 are mutated to typically hydrophobic residues[5].

The next step in the pathway is aminoacylation of the ACP-bound substrate with L-alanine by a second adenylation enzyme. This has been described as a 'protecting group' strategy, prevent premature cyclisation of the nascent chain[11,15]. The resulting dipeptidyl group is then transferred to the ACP of the PKS's loading module by a trans-acting acyltransferase[16]. Finally, following polyketide synthesis, the polyketide chain is deacylated by hydrolytic cleavage of the L-alanine protecting group amino acid, allowing macrocyclisation and release of the final product from the PKS[17].

4. 1. 1. 1. 2 Aminoglycosylation

βPMs, as with most polyketide natural products, are subject to post-PKS tailoring reactions that increase molecular diversity. The most common tailoring reaction in BPMs is hydroxylation by cytochrome P450s (e.g. Zhu et al., 2015[18]). Several BPMs are also glycosylated with a variety of sugars (Figure 4. 4). Glycosyl donors are nucleotide diphosphate (NDP) amino sugars derived from uridine diphosphate (UDP) and thymidine diphosphate (TDP) activated monosaccharides[6,14,19,20].

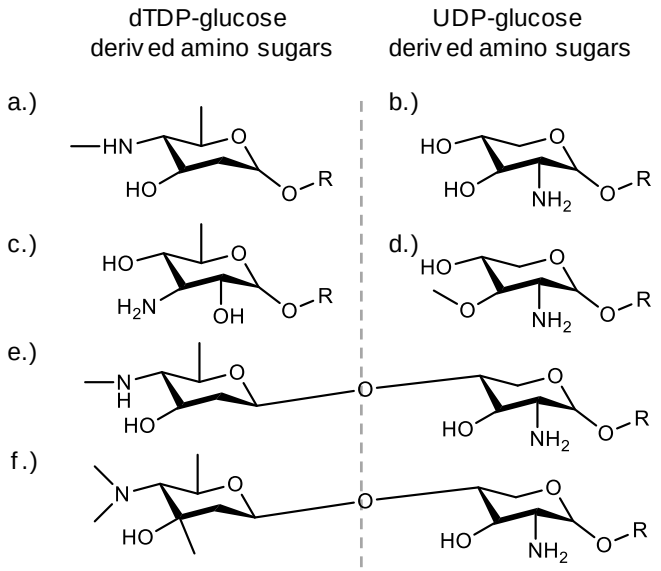
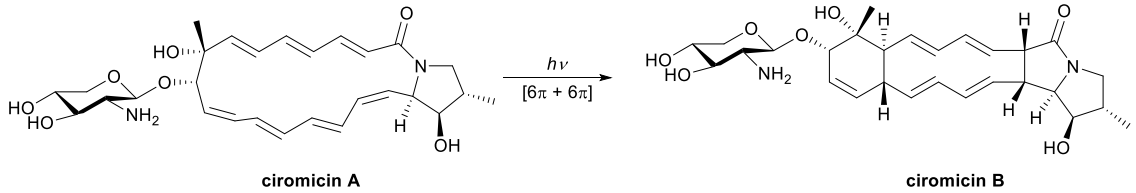


Figure 4. 37: dTDP and UDP derived amino sugars in the biosynthesis of βPMs. Sugars are involved in the biosynthesis of the following compounds: a.) ciromicin and macrotermycin, b.) sannastatin and vicenistatin, c.) fluvirumicin, d.) silvalactam e.) incednine and f.) sipanmycin.

Ciromicin, silvalactam, incednine and sipanmycin all contain sugars derived from UDP-N-acetylglucosamine (UDP-GlcNAc)[6,14,20,21]. The ciromicin BGC encodes three enzymes: CirS1, a GlcNAc de-N-acetylase; CirS2; a UDP-glucose dehydratase and; CirS3 currently annotated as a NAD-dependent epimerase/dehydrogenase[14], but is more likely to function as a GlcA decarboxylase (a common misannotation by predictive software). Additional N- and O-methylation steps may occur to further modify the sugars, as seen in incednine and silvalactam biosynthesis[6,22]. Vicenistatin, incednine, sipanmycin and the fluvirucins derive sugars from glucose-1-phosphate [6,21,23–25]. In the case of vicenistatin the sugar vicenisamine is synthesised by six enzymes VinABDEFG. VinA converts glucose-1-phosphate to D-glucose. VinBDE then catalyse a series of dehydrations and reductions yielding 4-keto-2,6-dideoxy-D-glucose. VinF then catalyses transamination at the 4-keto position followed by N-methylation by VinG[26]. In the fluvirucins, the amino group is often acetylated or, in the case of fluvirucin B_5, is replaced by 2-phenyethylureide[23]. The incednine BGC contains additional reductases resulting in complete reduction of C2[6].

The biosynthesis of incednine and sipanmycin is particularly interesting as it requires the synthesis of both UDP and TDP activated sugars[6,21]. Each monosaccharide is derived from independent nucleotide diphosphates, in the case of incednine 5-deoxy-N-methyl-D-glucosamine and 4-methylamino-2,3,4,6-tetradeoxy-α-D-glucose, derived from UDP and TDP sugars respectively (Figure 4. 4.e-f). As presence or absence of specific amino sugars has been shown to be important for biological activity of βPMs (see below), the production of non-specific glycons would be undesirable. Since nucleotide diphosphates are often required for recognition by biosynthetic enzymes it is possible utilising two distinct NDPs allows specific modifications to be made to each sugar during incednine and sipanmycin biosynthesis. Indeed the correct glycosylation of sipanmycin is dependent on the coordination of its glycosyl transferases[21].

The importance of the aminosugar for activity suggests βPMs may provide interesting targets for glycodiversification – the synthesis of new analogues through glycosylation with different carbohydrates[27]. Due to its ability to accept a variety of substrates, the vicenistatin glycosyltransferase VinC has been identified as a potentially useful tool for glycodiversification[28]. Providing the size and position of polar groups are roughly similar to that of vicenilactam (the biosynthetic aglycon precursor to vicenistatin) VinC can accept a range of glycosyl acceptors. It can also accept both D- and L- sugars to produce a variety of α- and β- vicenilactam glycosides. By combining ten glycosyl donors and five aglycones, Minami and Eguchi generated a library of 22 novel glycosylated polyketides[28].



ciromicin A ciromicin B

Figure 4. 38: Photocatalysed [6π + 6π] intramolecular cycloaddition of ciromicin A.

4. 1. 1. 1. 3 Intramolecular cycloadditions

Due to the nature of the polyene ring, several βPMs are able to undergo intramolecular cycloadditions (see Chapter 3). [6π + 6π] and [6π + 4π] cycloadditions of the polyene macrocycles were first hypothesised upon the discovery of the heronamides (A - C) (**5** - **7**)[29]. Heronamide C (**7**) is the precursor to both heronamide A (**5**) and heronamide B (**6**). A [6π + 6π] intramolecular cycloaddition leads to the production of **6**. **5** is produced by an epoxidation of the C16-C17 olefin and subsequent nucleophilic attack by the proximal nitrogen followed by [6π + 4π] cycloaddition to produce **5**. As demonstrated in the previous chapter, these reactions occur spontaneously through photochemical and thermally dependent routes (Figure 3. 4)[30].

As a consequence of their polyene structures, several βPMs are sensitive to light. The first characterised example involved the conversion of ciromicin A to ciromicin B *via* a [6π + 6π] cycloaddition (Figure 4. 5)[14]. The reaction was photocatalysed in a UV-vis dependent manner, with maximum conversion occurring at 300 nm. The [6π + 6π] cycloaddition of **7** has also been shown to proceed rapidly upon exposure to light[30]. Analysis of this reaction revealed that in addition to greater rates of reaction, the quantum yield was also much higher towards the lower wavelengths of the UV spectrum (Figure 3.11 and Table 3.2). This is likely to be due to the increased likelihood of a photon being absorbed by the conjugated trienes.

Heronamide F was shown to convert to heronamide E (an analogue of **5**) *in vitro* in the presence of DMSO[31]. This result suggested that the formation of the pyrrolidinol and subsequent [6π + 4π] cycloaddition could occur spontaneously in the presence of an oxidising agent. However, the reaction was poorly characterised and apparently occurred at a rate much slower than had been observed *in vivo* (taking 14 days to produce only trace products). Sequencing of the heronamide C (*hrn*), heronamide E (*her*) and ciromicin (*cir*) BGCs revealed that they all lack the enzymatic capacity to carry out such an oxidation (each cluster contains only a single P450 monooxygenase that has been demonstrated to act as a hydroxylase)[14,30,31]. We demonstrated in the previous chapter that production of **5** was unlikely to result from biological activity and that **5** was produced in a thermally-dependent manner. This has been proposed to occur through an ambimodal transition state that leads to a [4π+2π] and [6π+4π] adducts[32]. Interconversion of these leads to accumulation of the thermodynamically more stable [6π+4π] adduct, however this is apparently not always the case however, as with dracolactams and mirilactams, the [4π + 2π] product is apparently stabilised[33,34]

A recent report by Ding et al.[35] revealed the extent of heronamide spontaneous chemistry by identifiying six additional cyclic congeners. This study raises questions about the unexplored diversity of polyene βPMs. As mentioned above, several other βPMs are reportedly sensitive to light and temperature. The exposure of the pentaene-tetraene rings of incednine and sceliphrolactam are of particular interest as is possible that such ring systems may be capable of novel pericyclic rearrangements and could provide new insights into polyene chemistry.

4. 1. 1. 2 Bioactivity and Structural-Functional relationships

βPMs exhibit a wide variety of biological activities (Table 4. 1). For example, cyclamenol A (currently under patent by Bayer AG) exhibits anti-inflammatory activity by inhibiting the adhesion of leukocytes to endothelial cells during immune response[36]. Lobosamide A was discovered due to its ability to inhibit the growth of the protozoa *Trypanosoma brucei*, the parasite responsible for human African trypanosomiasis (sleeping sickness)[37]. Fluviricin inhibits infection of influenza virus type A[23]. Additionally, several βPMs have been shown to inhibit the growth of both Gram-positive bacteria and Gram-negative bacteria[22,38,39]. Most commonly however, βPMs exhibit cytotoxic or antifungal activities.

4. 1. 1. 2. 1 Cytotoxicity

Roughly half of known βPMs exhibit some degree of cytotoxic activity against mammalian cells (Table 4. 1). It is likely that this activity occurs in a similar fashion to that of other known polyene antibiotics through perturbation of the cell membrane[40–42]. Some βPMs however are known to target specific pathways. Incednine, for example, induces apoptosis in cell lines overexpressing the anti-apoptotic oncoproteins Bcl-2 and Bcl-xL[6].

Particular structural motifs have been shown to play an important role in determining the cytotoxicity of βPMs, such as the presence of an amino-sugar. Despite vicenistatin's potent cytotoxic activity, its homologue vicenistatin M - where D-vicenisamine is substituted for D-mycarose[43] – exhibited no activity against a range of human cell lines. Similarly, sipanmycin also requires the correct installation of both amino sugars to exhibit antiproliferative activities[21]. Indeed, the presence of an amino sugar is a strong predictor of cytotoxicity. The only BPMs that exhibit cytotoxicity that lack an amino sugar are those which incorporate hydrophobic β-amino acids[10,44,45]. βPMs lacking both moieties are usually void of activity.

The configuration of the macrocycle also appears to have an effect on activity. Although **5** and **6** lack the anti-fungal properties of **7** (see below), it has been demonstrated that products of intramolecular cycloaddition can retain activities (see ciromicin A and B[14]) or gain novel activities. Nivelactam is a [6π + 4π] product of BE-14106 with a structure distinct from that of **5**. Nivelactam shows increased cyctotoxicity against certain human tumour cell lines compared to its precursor[46]. Methylation of the macrocycle is also important. The total synthesis of vicenistatin derivatives revealed that loss of the C20- and C23-methyl groups resulted in a cumulative decrease in activity[47].

4. 1. 1. 2. 2 Antifungal activity

Many βPMs inhibit the growth of a variety of fungi including *Candida albicans*, *Saccharomyces cerevisae*, the plant pathogens *Magnaporthe grisea* (*Pyricularia oryzae*) and *Sclerotinia sclerotiorum*, and fission yeast *Schizosaccharomyces pombe*[4,29,48–51]. Recently, studies on the heronamides and related 20-membered BPMs have provided insight into their mode of action and structure-activity relationships.

As with other βPMs, the heronamides have been shown to function by irreversibly binding to membrane lipids (specifically 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine and sphingomyelin)[50]. Comparison of **7** with BE-14106, revealed that despite sharing a mode of action, BE-14106 exhibited 4-fold lower inhibitory activity. As the two compounds share a stereo-chemically identical macro-lactam ring, the observed change in activity was attributed to the reduced length of the hydrocarbon tail. The configuration of the macrocycle itself also appears to be a requisite for activity. Despite the potent activity of monocyclic **7**, tetracyclic **5** and tricyclic **6** exhibit no anti-fungal activity[29,50]. Hydroxylation is another important factor. The MIC for 8-deoxy-**7** is 20-fold greater than that of the hydroxylated product. In addition, synthetically produced diacetylated **7** lacked both activity and membrane affinity[50].

4. 1. 1. 2. 3 Inducers of vacuolation

In addition to their anti-fungal activity, Raju et al.[52] reported that **7** elicited an unusual response from mammalian cells. HeLa cells grown on 50μM **7** formed large, vesicle-like structures that dissipated when it was removed from the media. A similar morphological phenomena can be induced by vicenistatin[53]. Large vacuoles formed in rat fibroblasts (3Y) treated with 300nM vicenistatin, reportedly occupying around half of the cells volume. Unlike **7** induced vacuolation, these changes were apparently irreversible. The vacuoles were formed from continuous, uncontrolled homolytic fusion of smaller, early-endosomally derived vesicles. Nishiyama et al.[53] demonstrated that vacuolation is induced by the activation of the Rab5-PAS pathway and a consequent decrease in cellular phosphatidylinositol (3,5)-bisphosphate and is phenotypically consistant with other preturbations to this pathway such as decreased activity of the phosphokinase PIKfyve. This mode of action draws parallels with the *RhoI-bgsI* dependent anti-fungal activity of **7** (see above). It is possible that the accumulation of cell wall material in the septa of fission yeast caused by treatment with **7** and vacuolation of mammalian cells are the result of similar modes of action.

βPMs may provide useful probes for studying the process of vacuolation in mammalian systems. It seems unlikely that the ability to induce vacuolation is be limited to heronamide C and vicenistatin. As the process occurs at sub-cytotoxic levels it is possible that the phenomina has gone unreported during the screening other βPMs.

4. 1.2 Genome-mining novel macrolactams

Given the vast amount of genomic data available, genome mining has become a powerful tool in the natural product discovery toolbox (see Chapter 1). Previous attempts at genome mining for βPMs have been successful. The mirilactams (A and B) were discovered by looking for sequences with homology to the lobosamide PKS sequence[37]. This technique is helpful in identifying closely related compounds (the mirilactams and lobosamides differ by the presence of a methyl group). However due to the high similarity of PKS encoding genes, it is not very successful at identifying more distantly related BGCs. With this in mind, we developed a strategy to survey the diversity of βPMs in the public databases by searching for functional sub-clusters as opposed to whole BGCs. As the process of β-amino acid incorporation is highly conserved among all known examples, it follows that searching for this sub-cluster would allow us to identify βPMs in the public databases without the bias introduced by using assembly line encoding genes.

4. 2. Results and Discussion

4. 2. 1 Targeted Genome Mining

The amino acid sequences for the conserved βPM biosynthetic genes from the vicenistatin BGC (VinJKLMN[19]) and the heronamide C BGC (HrnIJKSLU[30]) were extracted and run in independent MultiGeneBLAST[54] architecture searches against genbank actinobacterial and microbial sequence databases. Results containing two or more homologous genes were considered BGCs of interest. The combined analyses yielded a total of 76 BGCs (Table 4. 2). The BGCs identified were widespread among the actinobacteria, identified in 16 genera of the families Streptomycetales, Streptosporangiales, Micromonoporales, Pseudonocardiales and Corynebacteriales (based upon Nouioui et al.'s classification[55]). This distribution is significant as it indicates a strict phylogenetic delineation between Actinobacterial clades and suggests an origin of these clusters during the diversification of these genera from the rest of the Actinobacteria (Figure 4. 6).

4. 2. 1. 1 PKS-associated BGCs

A total of 54 BGCs were identified sharing at least two of the conserved genes and a PKS encoding gene (Table 4. 2). 46 of the extracted sequences represented entire putative BGCs and 8, extracted from smaller contigs, appeared to contain partial BGCs. Nonetheless, this highlights the advantage of this approach to identify putative BGCs from not only full genome assemblies, but also smaller contigs. 12 BGCs, including *vin* and *hrn* were previously identified and 2 BGCs had architectures identical to known BGCs: *Streptomyces niveus* NCIMB 11891 (BE-14106)[9] and

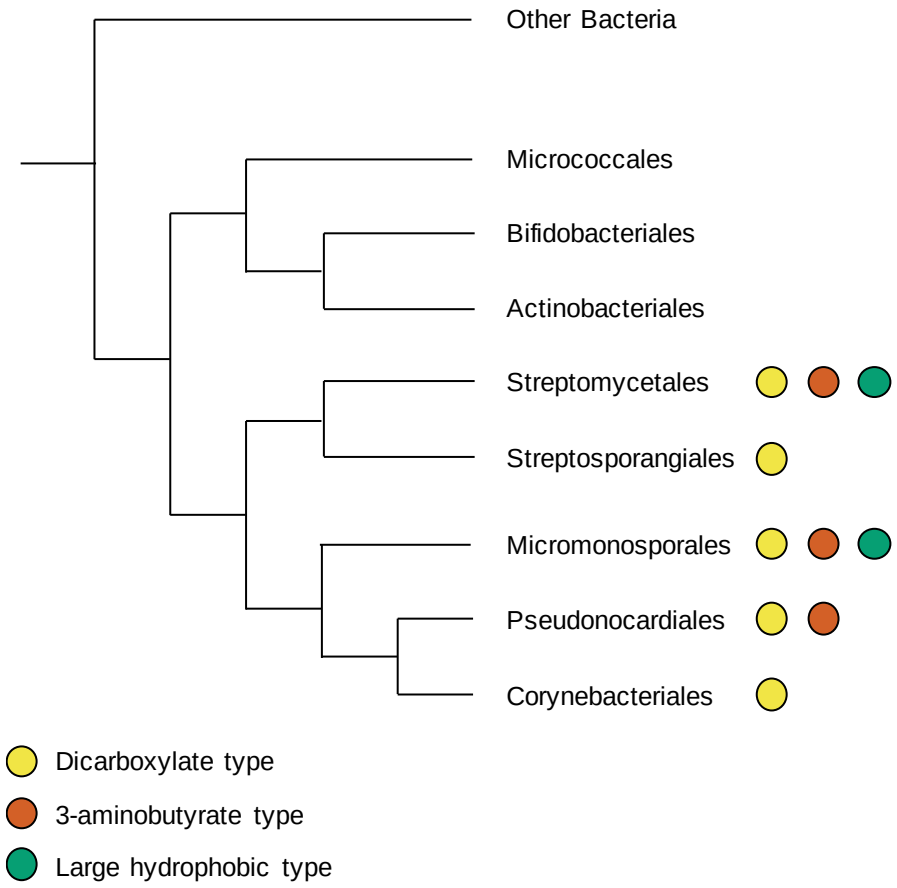


Figure 4. 39: Predicted distribution of β-amino acid containing macrolactams in actinobacteria. The presence of particular β-amino acid containing macrolactam (βPM) biosynthetic gene clusters (BGCs) is signified by the coloured circle: dicarboxylate type (L-aspartate and L-methyl-threo-3-aspartate) in yellow, 3-amino butyrate type in red and large hydrophobic type (L-β-phenylalanine and β-amino fatty acids) in green.

	Genus	Species	Strain	Full/Partial Sequence	Product	Starter Unit	Ring Size	Glycosyl-transferase
Known Clusters	*Actinomadura*	*fulva*	ATC 53714	Full	Fluvirucin	l-aspartate	14	Y
	Micromonospora	sp.	RL09-05-HVF-A	Full	Lobosamide	3-aminobutyrate	26	N
	Salinispora	*tropica*	CNB-440	Full	Salinilactam	3-aminobutyrate	26	N
	Streptomyces	*halstedii*	HC34	Full	Vicenistatin	3-methylaspartate	20	Y
	Streptomyces	*scabrisporus*	KM-4927(T)	Full	Hitachimycin	β-phenylalanine	19(22)	N
	Streptomyces	sp.	CMB-0406	Full	Heronamide C	β-amino fatty acid	20	N
	Streptomyces	sp.	DSM 21069	Full	BE-14106	β-amino fatty acid	20	N
	Streptomyces	sp.	MJ635-86F5	Full	Cremimycin	β-amino fatty acid	19 (22)	Y
	Streptomyces	sp.	ML694-90F3	Full	Incednine	3-aminobutyrate	24	Y
	Streptomyces	sp.	MP39-85	Full	ML-449	β-amino fatty acid	20	N
	Streptomyces	sp.	SCSIO 03032	Full	Heronamide F	β-amino fatty acid	20	N
	Streptomyces	sp.	SD85	Full	Sceliphrolactam	3-methylaspartate	26	N
Predicted BPMs	*Actinokineospora*	*enzanensis*	DSM 44649	Full	Unknown	Unknown	Unknown	Y
	Actinoplanes	sp.	SE50/110	Full	Unknown	Unknown	Unknown	N
	Actinosynnema	*mirum*	DSM 43827	Full	Unknown	3-aminobutyrate	26	N
	Actinosynnema	*pretiosum*	X47	Full	Unknown	3-aminobutyrate	26	N
	Amycolatopsis	*orientalis*	DSM 46075	Full	Unknown	Unknown	Unknown	Y
	Amycolatopsis	sp.	AA4	Full	Unknown	Unknown	Unknown	Y
	Hamadaea	*tsunoensis*	DSM 44101	Full	Unknown	Unknown	Unknown	Y
	Kitasatospora	sp.	Root187	Full	Unknown	Unknown	Unknown	N
	Kutzneria	*albida*	DSM 43870	Full	Unknown	Unknown	Unknown	Y
	Kutzneria	sp.	774	Full	Unknown	Unknown	Unknown	N
	Micromonospora	*auratinigra*	DSM 44815	Full	Unknown	3-aminobutyrate	26	N
	Micromonospora	*purpureochromogenes*	DSM 43821	Full	Unknown	Hydrophobic	Unknown	N
	Micromonospora	*rifamycinica*	DSM 44983	Full	Unknown	3-aminobutyrate	26	N
	Micromonospora	sp.	ATCC 39149	Full	Unknown	Unknown	Unknown	Y
	Nocardia	*tenerifensis*	NBRC 101015	Partial	Unknown	Unknown	Unknown	N
	Nocardiopisis	*gilva*	YIM 90087	Partial	Unknown	Unknown	Unknown	Unknown
	Saccharothrix	sp.	ST-888	Partial	Unknown	Unknown	Unknown	Unknown
	Saccharothrix	*syringae*	NRRL B-16468	Full	Unknown	3-aminobutyrate	Unknown	N
	Streptomyces	*albireticuli*	MDJK11	Full	Unknown	Unknown	Unknown	Y
	Streptomyces	*hygroscopicus*	5008	Full	Unknown	3-methylaspartate	26	Y
	Streptomyces	*hygroscopicus*	ATCC 53653	Full	Unknown	β-amino fatty acid	20	N
	Streptomyces	*hygroscopicus*	KCTC 1717	Full	Unknown	3-methylaspartate	26	Y
	Streptomyces	*hygroscopicus*	TL01	Full	Unknown	3-methylaspartate	26	Y
	Streptomyces	*mobaraensis*	NBRC 13819/DSM 40847	Partial	Unknown	Unknown	Unknown	Unknown
	Streptomyces	*niveus*	NCIMB 11891	Full	BE-14106	β-amino fatty acid	20	N
	Streptomyces	*pratensis*	ATCC 33331	Full	Unknown	3-methylaspartate	26	N
	Streptomyces	*reticuli*	Tu 45	Full	Unknown	3-methylaspartate	26	Y
	Streptomyces	*roseochromogenes*	DS12.976	Full	Unknown	3-methylaspartate	Unknown	N
	Streptomyces	*roseosporus*	NRRL 11379	Full	Unknown	Unknown	Unknown	Y
	Streptomyces	*roseosporus*	NRRL 15998	Full	Unknown	3-aminobutyrate	Unknown	Y
	Streptomyces	*scabrisporus*	DSM 41855	Partial	Hitachimycin	β-phenylalanine	Unknown	Y
	Streptomyces	sp.	2323.1	Full	Unknown	Unknown	Unknown	Y
	Streptomyces	sp.	AA4	Full	Unknown	Unknown	Unknown	Y
	Streptomyces	sp.	DvalAA-83	Full	Unknown	3-methylaspartate	26	Y
	Streptomyces	sp.	LaPpAH-95	Partial	Unknown	Unknown	Unknown	Unknown
	Streptomyces	sp.	PAMC26508	Full	Unknown	3-methylaspartate	26	Y
	Streptomyces	sp.	PBH 53	Full	Unknown	3-methylaspartate	26	N
	Streptomyces	sp.	SirexAA-E	Full	Unknown	Unknown	26	Y
	Streptomyces	sp.	TP-A0875	Partial	Unknown	Unknown	Unknown	Unknown
	Streptomyces	sp.	Tu 6075	Full	Unknown	3-aminobutyrate	Unknown	Y
	Streptomyces	sp.	Tu 6176	Full	Unknown	Unknown	Unknown	Y
	Streptomyces	*tsukubensis*	NRRL 18488	Partial	Unknown	β-amino fatty acid	Unknown	Unknown
Other Clusters	*Actinobacteria*	*bacterium*	OV320	Full	Unknown	N/A	N/A	N/A
	Actinokineospora	*enzanensis*	DSM 44649	Partial/Full	Unknown	N/A	N/A	N/A
	Amycolatopsis	*vancoresmycina*	NRRL B-24208	Partial/Full	Unknown	N/A	N/A	N/A
	Kutzneria	*albida*	DSM 43870	Full	Unknown	N/A	N/A	N/A
	Kutzneria	sp.	744	Full	Unknown	N/A	N/A	N/A
	Lechevalieria	*aerocolonigenes*	NRRL B-3298	Full	Unknown	N/A	N/A	N/A
	Micromonospora	*aurantiaca*	ATCC 27029	Full	Unknown	N/A	N/A	N/A
	Micromonospora	*echinofusca*	DSM 43913	Full	Unknown	N/A	N/A	N/A
	Saccharothrix	sp.	NRRL B-16348	Full	Unknown	N/A	N/A	N/A
	Salinispora	*arenicola*	CNS-205	Full	Unknown	N/A	N/A	N/A
	Salinispora	*tropica*	CNB-440	Full	Unknown	N/A	N/A	N/A
	Streptomyces	*albulus*	ZPM	Full	Unknown	N/A	N/A	N/A
	Streptomyces	*bingchenggensis*	BCW-1	Full	Unknown	N/A	N/A	N/A
	Streptomyces	*collinus*	Tu 365	Full	Unknown	N/A	N/A	N/A
	Streptomyces	*cyaneogriseus*	NMWT 1	Full	Unknown	N/A	N/A	N/A
	Streptomyces	*griseoplanus*	NRRL B-3064	Partial	Unknown	N/A	N/A	N/A
	Streptomyces	*hygroscopicus*	ATCC 53653	Full	Unknown	N/A	N/A	N/A
	Streptomyces	*lincolnensis*	NRRL-2936	Full	Unknown	N/A	N/A	N/A
	Streptomyces	sp.	150FB	Full	Unknown	N/A	N/A	N/A
	Streptomyces	sp.	CdTB01	Full	Unknown	N/A	N/A	N/A
	Streptomyces	sp.	P3	Full	Unknown	N/A	N/A	N/A

Table 4. 11: β-amino acid containing macrolactams biosynthetic gene clusters predicted by sub-cluster mining. Contains information on the strain and observed structural features. Predicted starter unit (based on the phylogeny of VinN-like proteins) and macrocycle size are highlighted in purple. Strains containing non-polyketide synthase biosynthetic gene clusters (BGCs) also shown.

Streptomyces scabrisporus DSM 41855 (hitachimycin)[56]. The remaining 40 BGCs could not be confidently assigned to any known βPM.

4. 2. 1. 2 Non-PKS-associated BGCs

In addition to identifying a number of type I PKS BGCs, 21 BGCs were identified that lacked PKS encoding genes, but had genes sharing homology with VinN and VinM (Table 4. 2). The contents of each BGC are variable but some features appeared to be well conserved among BGCs from several species, in addition to adenylation domains, typically contain acyl-carrier protein, thioesterase and aminotransferase encoding genes. These BGCs were not consistently annotated by antiSMASH, and have no apparent homology with other known BGCs, with the exception of the βPMs.

4. 2. 2. Evolutionary Analyses

4. 2. 2. 1. VinN-like like Proteins

The major distinguishing feature of βPMs, is the utilisation of different starter units. The first step of this process is the selection of an amino acid by a standalone adenylating enzyme. In vicenestatin biosynthesis, this step is carried out by VinN[19], homologues of which can be found in all known βPM BGCs. we hypothesised that a phylogenetic study of our expanded dataset could improve understanding of the diversity of VinN-like proteins and allow us to predict the substrates of some of these novel BGCs. Annotated AMP-binding domains were extracted from all sequences. The resulting set of sequences included both VinN-like proteins, VinM-like proteins (see section 4. 2. 2. 2) and adenylation domains from non-ribosomal peptide synthetases (NRPSs) proximal to BGCs in our dataset. VinN-like proteins were then determined by filtering sequences for standalone adenylation domain proteins and by homology to VinN. The sequences of VinN-like proteins and A-domains of Strop_2821 (57 sequences) were aligned in ClustalW[57] and a maximum likelihood phylogenetic tree was generated using RaxML version 8[58]. VinN-like proteins formed a single highly supported clade (Figure 4. 7). The tree formed two highly supported clades, the first (Clade I) containing proteins required for the adenylation of large hydrophobic β-amino acids and 3-aminobutyrate.and the second (Clade II) containing L-methyl-*threo*-3-aspartate /L-aspartate adenylating proteins.

Clade I contained two well supported subclades. One contained the unsaturated β-amino acid adenylating proteins BecJ, HerJ, HrnJ and MlaJ, along with proteins from *Streptomyces niveus* CIMB 11891 and *Streptomyces tsukubensis* NRRL18488. The BGC from *Streptomyces niveus* CIMB 11891 shared an architecture with the *bec* BGC. Given this homology and it's position in the phylogeny, it is likely that this BGC produces the BE-14101 also. The BGC from *Streptomyces tsukubensis* NRRL18488 is not as similar, and may produce a novel unsaturated

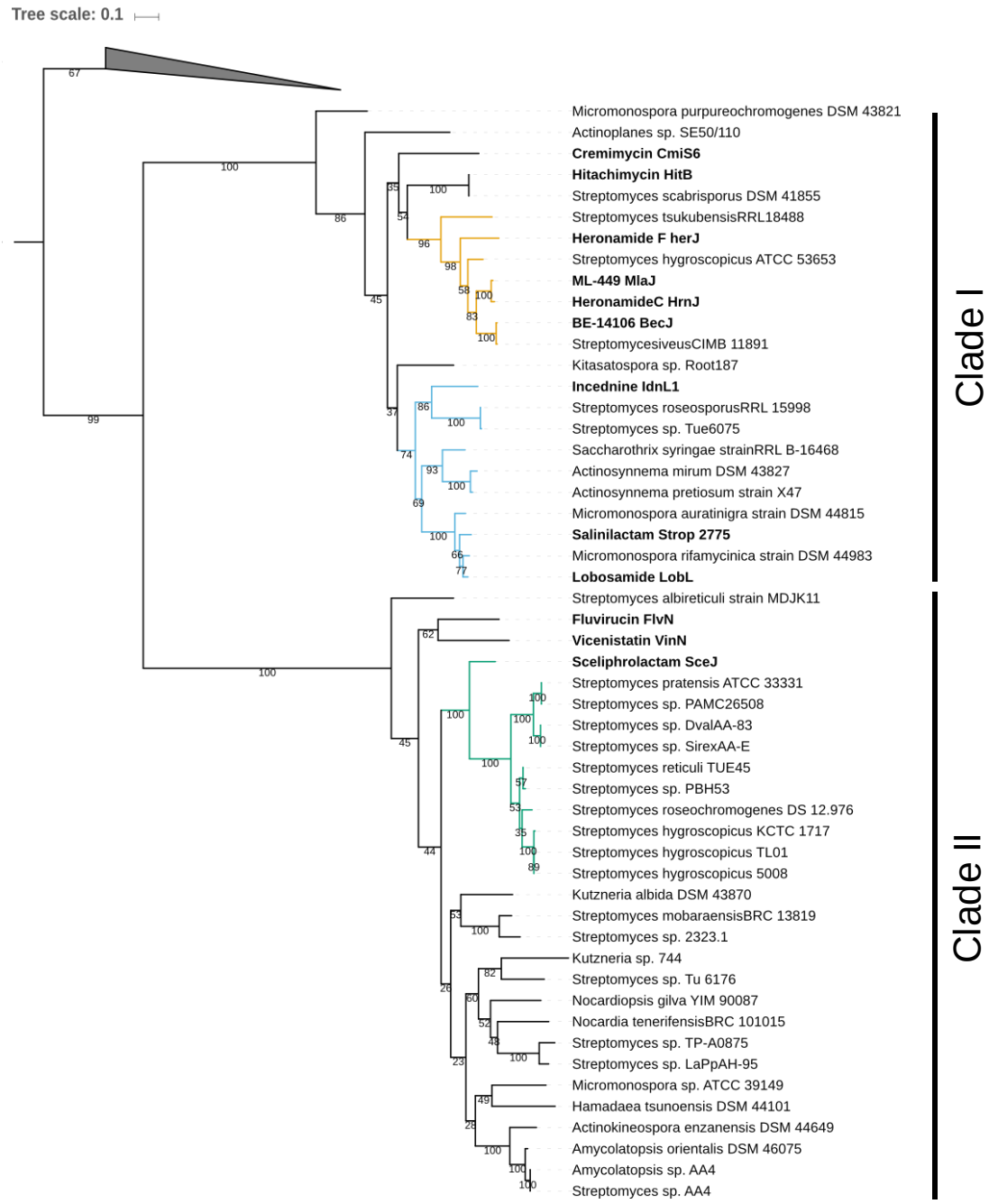


Figure 4. 40: Phylogeny of VinN-like proteins. Bootstrap support from 100 trees are shown. Clade I and Clade II are labelled, and subclades used for starter unit predictions are labelled as follows: unsaturated β-amino fatty acid activating proteins in yellow, 3-aminobutyrate in blue and L-methyl-threo-3-aspartate in green. Proteins from characterised biosynthetic gene clusters (BGCs) are shown in bold.

βPM. The other sub-clade had medium bootstrap support and contained LobL, Strop_2775 and IdnL1, which accept 3-aminobutyrate as a substrate. Seven additional sequences clustered within this sub-clade and we predicted these proteins to also activate 3-aminobutyrate. The relationship between these two subclades was poorly supported however, as was the position of several other proteins were poorly supported within Clade I; notably, CmiS6 and HitB. The position of *Micromonospora pureochromogenes* DSM43821, *Actinoplanes* sp. SE50/10 and *Kitasatospora* sp. Root187 were also poorly supported so substrate predictions could not be assigned.Clade II also contained two highly supported subclades. One contained ten Streptomyces spp. that shared a common ancestor with SceJ. The other subclade contained *Amylcolatopsis* spp. and *Actinokineospora enzanensis* DSM4469. In general, however, relationships within this clade, including the positions of VinN and FlvN, were poorly supported.

4. 2. 2. 2. VinM-like proteins

A second phylogeny was generated from the VinM-like sequences, using the same outgroup (54 sequences) (Figure 4. 8). This phylogeny showed the same well-supported clades as the VinN-like protein tree; however the two highly supported clades could not be observed. Like with VinN and FlvN, the positions of VinM and FlvM were poorly supported.

4. 2. 2. 3. VinK-like Acylytransferases

All βPM BGCs contain a conserved gene encoding a standalone VinK-like acyltransferase (AT). ATs are required to transfer the starter unit to the loading module of the PKS[16]. As they are specific to their substrates, their protein sequences should provide insight into their functions. Standalone AT protein sequences were extracted from our dataset and aligned in ClustalW[57] along with modular AT-domain sequences from BecB, CmiP4, HrnD, LobB and FlvP1 as an outgroup, yielding an alignment of 58 sequences.

All of the VinK-like Acyltransferases formed a single clade with maximum bootstrap support (Figure 4. 9). With the exception of the two BGCs, from *Kutzneria* spp., all sequences fit within two clades, analogous to those of the VinN-like proteins. Clade I contained AT domains known associated with dicarboxylate type βPMs. Although this clade was less well supported than with the VinN-like proteins supported, it contained several highly supported sub-clades. Clade II contained BGCs utilising other known starter units. Within this clade, all AT- domains incorporating hydrophobic starter units (β-amino fatty acids and β-phenylalanine) clustered in medium to well supported sub-clades.

4. 2. 2. 3. Type I Thioesterase (TE)-domains

Type I thioesterases (T1TE)-domains are found in the terminal modules of PKSs. They are responsible for release of the polyketide chain and formation of the macrocycle. Therefore, we hypothesised that relationships between TE-domain sequences could be used to predict the

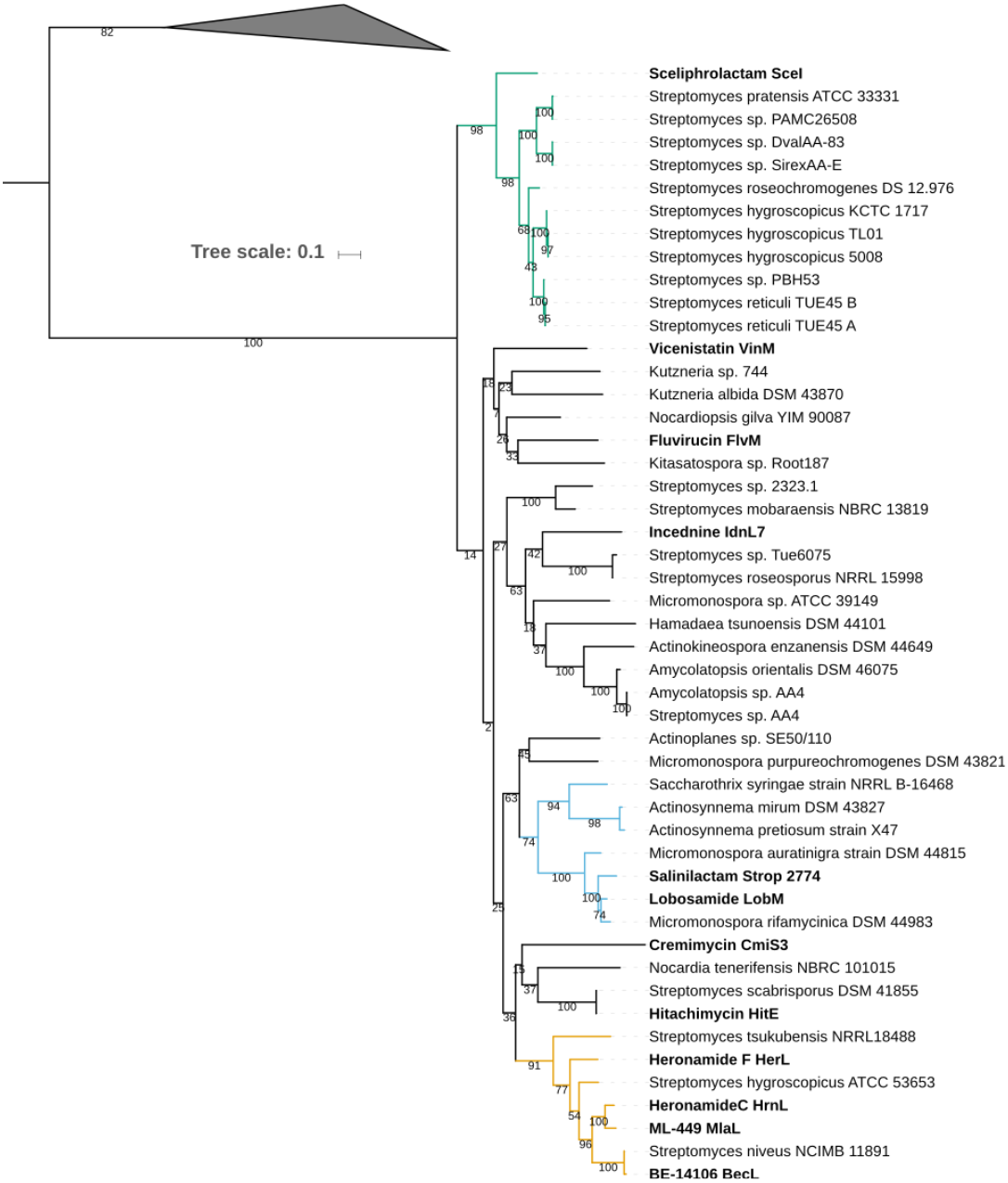


Figure 4. 41: Phylogeny of VinM-like proteins. Bootstrap support from 100 trees are shown. Well supported subclades are labelled by their associated BGCs: unsaturated β-amino fatty acid type in yellow, 3-aminobutyrate in blue and L-methyl-threo-3-aspartate in green. Proteins from characterised biosynthetic gene clusters (BGCs) are shown in bold.

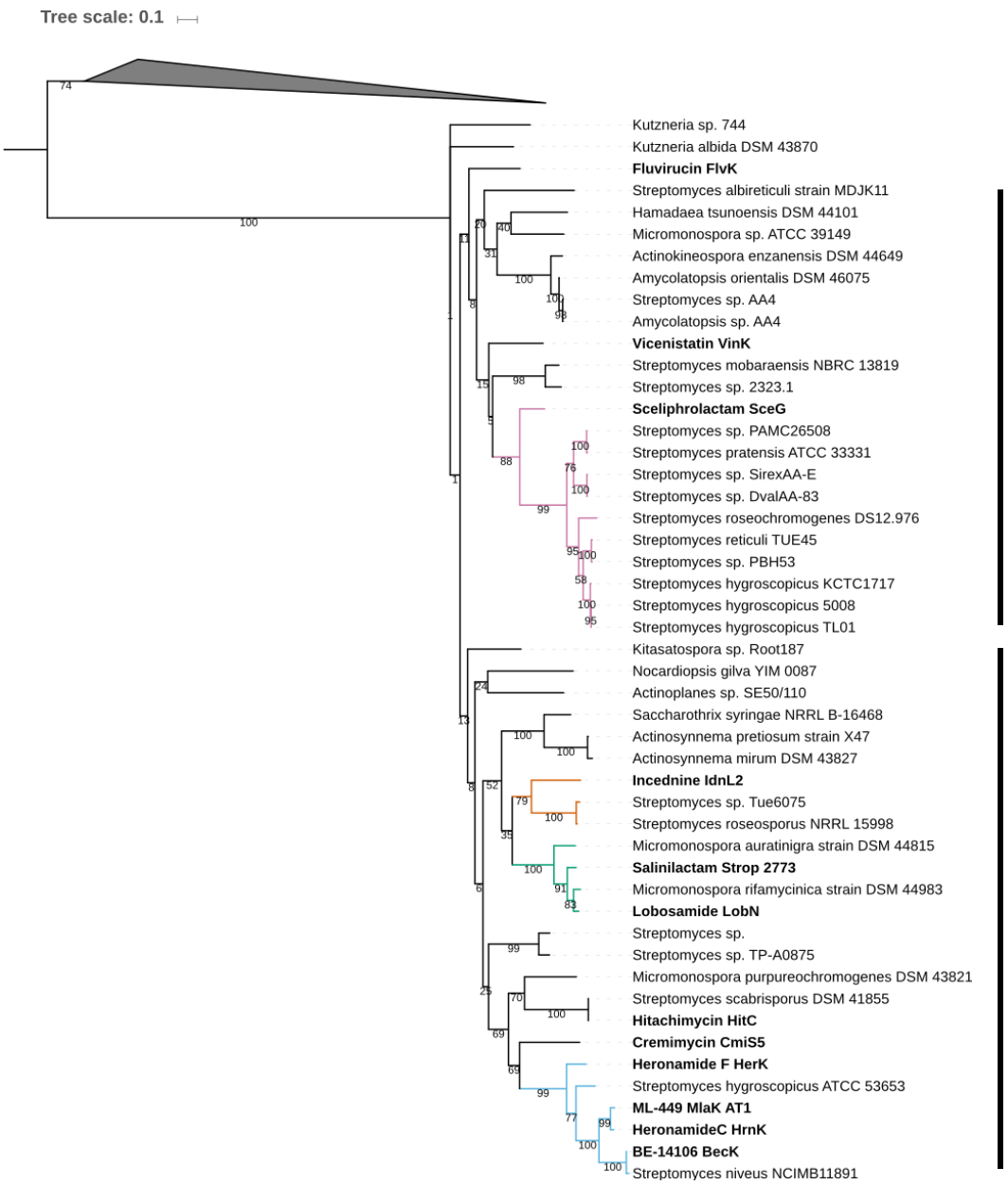


Figure 4. 42: Phylogeny of VinK-like proteins. Bootstrap support from 100 trees are shown. Clades I and II are labelled and well supported subclades are labelled by their associated BGCs: unsaturated β-amino fatty acid type in blue, 3-aminobutyrate in green and orange and L-methyl-threo-3-aspartate in purple. Proteins from characterised biosynthetic gene clusters (BGCs) are shown in bold.

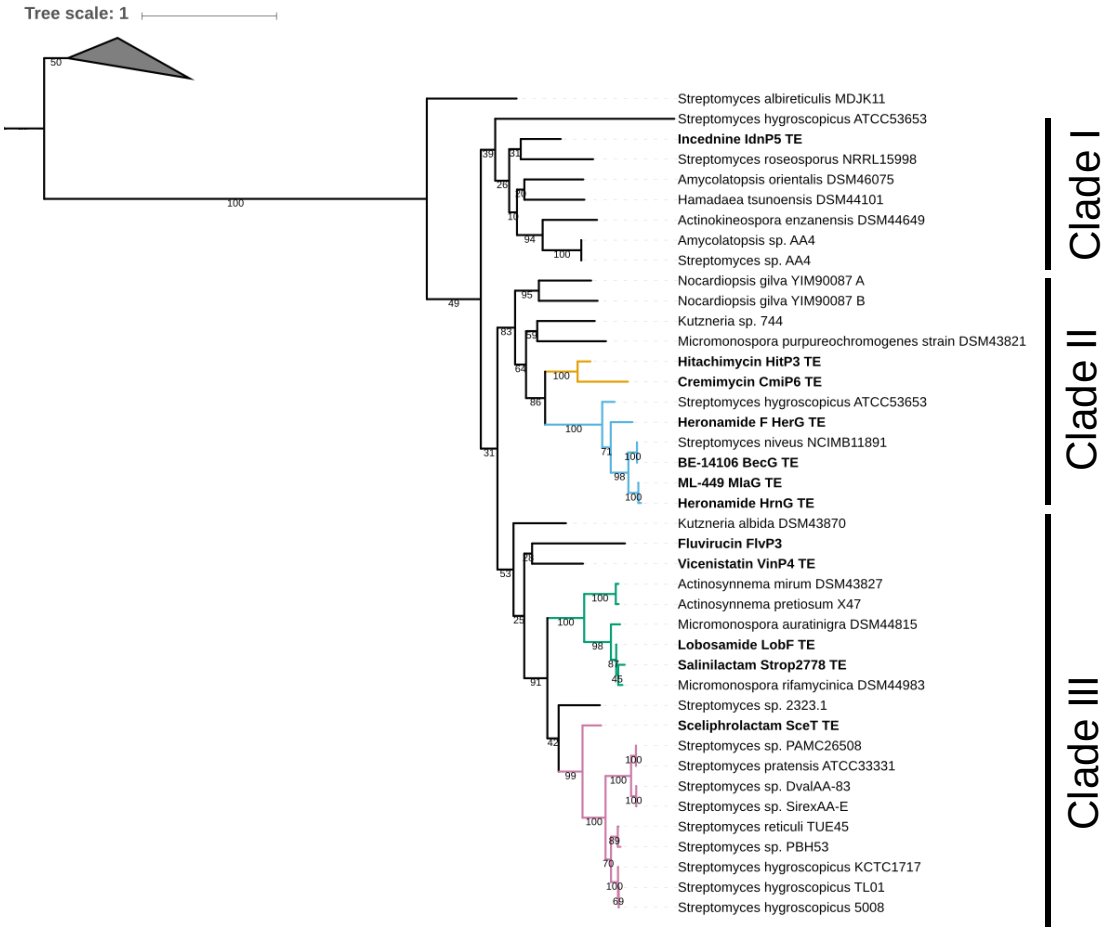


Figure 4. 43: Phylogeny of type I thioesterase domains. Bootstrap support from 100 trees are shown. Clades I, II and III are labelled. Subclades used for predictions of macrocycle size are coloured as follows: 19 (22)-membered macrolactams (see cremimycin and hitachimycin in Figure 4. 1) in orange; 20-membered macrolactams in blue and; 26-membered rings in green and purple. Domains from characterised biosynthetic gene clusters (BGCs) are shown in bold.

size of a macrocyclic product. To test this hypothesis, an alignment was generated from our dataset of PKS T1TE-domains. Type II TE (related, but distinct enzymes believed to be important in the removal of aberrant residues from the PKS[59]) sequences from previously described βPM BGCs (CmiR1, IdnB, FlvC, LobU, Strop_2673) were included in the alignment as outgroups. The resulting alignment of 50 sequences was used to build a phylogeny as above.

All T1TEs formed a single clade with very high support (Figure 4. 10). The thioesterases formed three major clades with varying levels of support. Clade II was the most strongly supported, containing highly supported subclades corresponding to the 22-membered macrolactam of hitachimycin and cremimycin, and 20-membered macrolactam of the heronamide-like βPMs. TE-domains from *Streptomyces niveus* NCIMB11891 (predicted to produce BE-14106) and *Streptomyces hygroscopicus* ATCC53653 suggesting these BGCs both encode 20-membered βPMs. Clade III is sister to clade II, but is less well supported, apparently due to the ambiguous position of the VinP4 and FlvP3 TEs. Within this clade however is a highly supported sub-clade containing all TEs from BGCs encoding 26-membered βPMs. Interestingly, there was a well-supported split in this clade between TEs from *Streptomyces* spp., including sceliphrolactam, and those from other actinobacteria (Micromonosporales and Pseudonocardiales). This is also notable because, unlike the phylogenies of VinN and VinK, these sequences are not clustered according to their associated starter unit, suggesting that phylogeny of T1TEs mirrors macrocycle size. This is further evidenced by the position of the TE-domain of IdnP5 in Clade I separate from the other TEs from 3-aminobutyrate type BGCs, meaning the TEs from 26-membered βPM BGCs are monophyletic. There are exceptions to this however, the TEs from the 20-membered βPM BGCs are paraphyletic as the TE of VinP4 is distinct from the TEs of the heronamide-like βPMs.

4. 2. 2. 5. Multiple-partioned Phylogeny

Although the individual protein alignments were useful for the prediction of specific structural features, the relationships between subclades were, in general, poorly supported. Proteins from the *vin* and *flv* BGCs were consistently positioned with low support. To examine the relationships between whole BGCs as opposed to individual proteins, alignments of VinN, VinM, VinK, VinJ and Type 1 TEs were concatenated into a single partitioned alignment. Concatenated sequences missing 2 or more of these sequences were excluded from the analysis. This yielded a final alignment of 1708 characters comprising 53 sequences. The phylogeny was built in RaxML version 8[58] using a mixed partition.

The phylogeny generated from the concatenated sequences showed overall similarity with the phylogeny of the VinN-like proteins (Figure 4. 11). It consisted of two major clades with

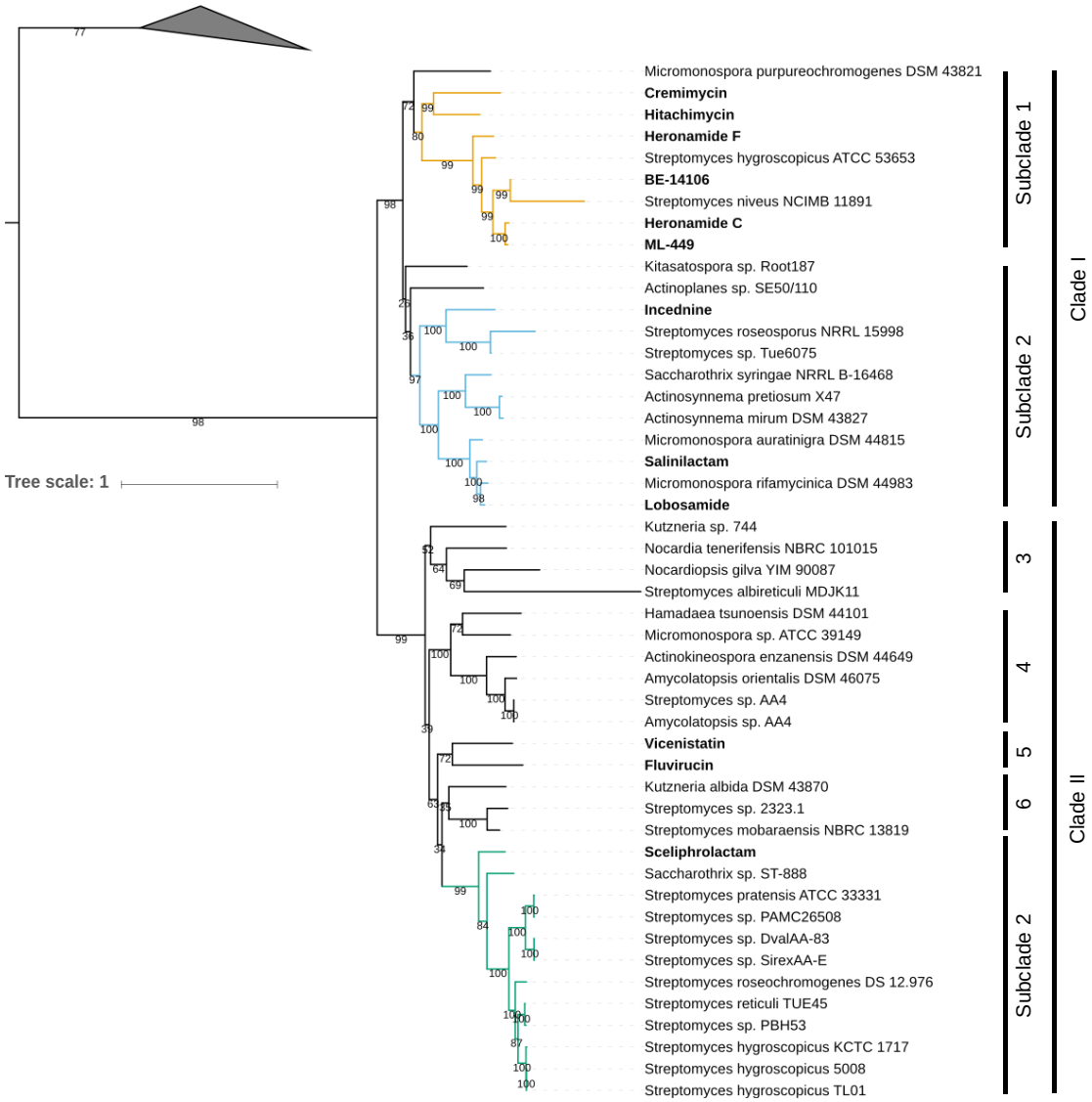


Figure 4. 44: Phylogeny of concatenated VinN, VinN, VinK, VinJ and type I thioesterase sequences. Bootstrap support from 100 trees are shown. Clades I and II and well supported subclades are labelled by their associated BGCs: large hydrophobic in blue, 3-aminobutyrate in green and orange and L-methyl-threo-3-aspartate in purple. Proteins from characterised biosynthetic gene clusters (BGCs) are shown in bold.

very high support: Clade I, representing the BGCs of hydrophobic β-amino acids and 3-aminobutyrate incorporating βPMs and; Clade II, containing L-methyl-*threo*-3-aspartate /L-aspartate incorporating βPMs. Although not all branches were highly supported, in general the support values were much higher. Furthermore, the tree clarified the relationships between certain BGCs. Importantly, all the 3-aminobutyrate associated BGCs were shown to be monophyletic, due to the inclusion of incednine. Interestingly, the *idn* BGC proteins formed a distinct subclade within this group with proteins from Streptomyces roseosporus NRRL16988 and *Streptomyces* sp. Tü6075 BGCs. Upon further inspection, these represent the only BGCs within Clade I to contain genes responsible for the biosynthesis of amino sugars (Table 4. 2). This is significant, as it suggests the utilisation of amino sugars has only evolved once in this lineage, as opposed to in Clade II where these genes are widespread. Importantly, three subclades within the phylogeny, subclades 3, 4 and 6 had no characterised members. This suggested that these could be good targets to identify novel βPMs.

4. 3 Identifying Target Macrolactams

A selection of strains from the subclades identified by our genome mining effort were aquired from the DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen) for metabolic analysis: *Actinokineospora enzanensis* DSM 44649 (subclade 4), *Hamadaea tsunoensis* DSM 44101 (subclade 4), *Kutzneria albida* DSM 43870 (subclade 6), *Nocardiopsis gilva* DSM 44841 (subclade 3) and *Streptomyces mobaraensis* DSM40847 (subclade 6). All strains were cultured on a range of solid (GYM, M1, MYM-TAP, R2, R5, and SF+M) and liquid (SM6, SM14, SM18, SV2 and TSB) media for 14 days at 30°C. 1 ml of culture was mixed with ethyl acetate (1:1). The aqueous phase was discarded, and the organic phase evaporated. The residue was then solved in methanol (200 µL) and submitted to Shimadzu LCMS platform for analysis. Metabolites with characteristics similar to identified βPMs (a molecular mass: 300-750; UV absorbance: 200 – 300 nm) were prioritised for further characterisation. Accurate masses were used to predict molecular formula and five compounds were designated as potential βPMs (Table 4. 3). The compounds of interest were identified from cultures of *Streptomyces morabensis* DSM 40847 (**8**) and *Nocardiopsis gilva* DSM 44841 (**9**-**12**) (Figure 4. 12). Unfortunately, searching the literature revealed that the identity of the *S. mobaraensis* compound of interest was likely to be Piericidin A (molar mass = 415.566 g/mol), a polyketide produced by several *Streptomyces* spp., including *S. mobaraensis* strains[60–62]. The compounds from *Nocardiopsis gilva* DSM 44841 were much more promising however. Importantly, their chromophores were remarkably similar to those of heronamide C (**7**)[52]. The empirical formulas are predicted as follows were as follows: **9**: $C_{29}H_{40}N_2O_4$; **10**: $C_{29}H_{40}N_2O_4$; **11**: $C_{23}H_{29}NO_2$ and; **12**: $C_{29}H_{27}NO_2$. These masses seem to represent a group of related compounds. **9** and **10** are potentially glycosylated congeners of **11**, whereas **12**

Strain	RT (mins)	λ max (nm)	Predicted Formula	Predicted Mass		Observed Mass		Mass Error (ppm)	
				$[M+H]^+$	$[M+Na]^+$	$[M+H]^+$	$[M+Na]^+$	$[M+H]^+$	$[M+Na]^+$
Streptomyces mobaraensis DSM40847	10.4	236	$C_{25}H_{37}NO_4$	416.2795	438.2615	416.2807	438.2595	2.882679	-4.56349
Nocardiopsis gilva DSM44841	6.4	289, 299	$C_{29}H_{40}N_2O_4$	481.3061	503.2880	481.3067	503.2894	1.246608	2.781707
	6.7	284, 299	$C_{29}H_{40}N_2O_4$	481.3061	503.2880	481.3064	503.2887	0.623304	1.390854
	7.6	287, 297	$C_{23}H_{29}NO_2$	352.2271	374.2091	352.2263	374.2069	-2.27126	-5.87907
	7.9	287, 297	$C_{23}H_{27}NO$	334.2165	356.1985	334.2157	356.1976	-2.39366	-2.52668

Table 4. 12: Putative Macrolactam Metabolites. Metabolites identified in culture extracts of Streptomyces.morabensis DSM40847 and Nocardiopsis gilva DSM44841. Retention times (RT), average mass of ionic adducts ($[M+H]^+$ and $[M+Na]^+$), absorbance maxima (λ max) and predicted empirical formulae shown.

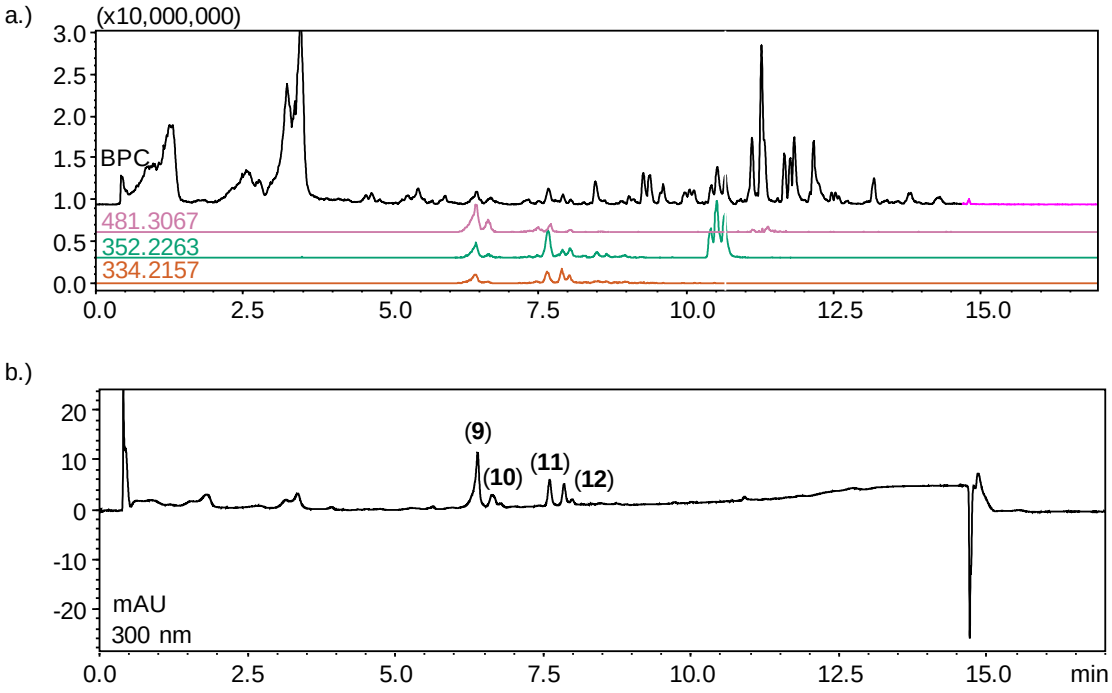


Figure 4. 45: Production of potential βPMs by Nocardiopsis gilva DSM40847. a.) LCMS spectra and b.) absorbance chromatogram of culture extracts.

appears to be a dehydrated congener. At the time of writing this thesis, the structure of these compounds has yet to be elucidated, but work is ongoing.

4.3 Conclusions

This chapter describes a sub-cluster-based approach to genome mining β-amino acid containing polyketide macrolactams (βPMs). Using this approach, public databases were surveyed for potential βPM biosynthetic gene clusters (BGCs). The number of identified clusters has been expanded significantly, and a number of non-PKS encoding BGCs were also identified. By considering the evolutionary relationships between BGCs we were able to identify the strains likely to produce the most novel βPMs. Work is ongoing to identify the products of these orphan BGCs. Additionally, new avenues for exploring the evolution of the BGCs were identified. The earliest divergence in the evolution of these BGCs splits βPMs utilising L-methyl-*threo*-3-aspartate or L-aspartate from those utilising large hydrophobic or 3-aminobutyrate. Of this later group, aminoglycosylation appears to have evolved only in a single lineage. Exploring these relationships further may provide novel insights into how BGCs evolve.

4. 4. Bibliography

(1) Lee, P.-C.; Umeyama, T.; Horinouchi, S. AfsS Is a Target of AfsR, a Transcriptional Factor with ATPase Activity That Globally Controls Secondary Metabolism in Streptomyces Coelicolor A3(2). *Mol. Microbiol.* **2002**, *43* (6), 1413–1430.

(2) Ōmura, S.; Nakagawa, A.; Shibata, K.; Sano, H. The Structure of Hitachimycin, a Novel Macrocyclic Lactam Involving β-Phenylalanine. *Tetrahedron Lett.* **1982**, *23* (45), 4713–4716.

(3) Komiyama, K.; Edanami, K.; Yamamoto, H.; Umezawa, I. Antitumor Activity of a New Antitumor Antibiotic, Stubomycin. *J. Antibiot. (Tokyo).* **1982**, *35* (6), 703–706.

(4) Komiyama, K.; Edanami, K.-I.; Tanoh, A.; Yamamoto, H.; Umezawa, I. Studies on the Biological Activity of Stubomycin. *J. Antibiot. (Tokyo).* **1983**, *36* (3), 301–311.

(5) Miyanaga, A.; Kudo, F.; Eguchi, T. Mechanisms of β-Amino Acid Incorporation in Polyketide Macrolactam Biosynthesis. *Curr. Opin. Chem. Biol.* **2016**, *35*, 58–64.

(6) Takaishi, M.; Kudo, F.; Eguchi, T. Identification of the Incednine Biosynthetic Gene Cluster: Characterization of Novel Glutamate Decarboxylase IdnL3. *J. Antibiot. (Tokyo).* **2013**, *66* (12), 691–699.

(7) Ogasawara, Y.; Kakinuma, K.; Eguchi, T. Involvement of Glutamate Mutase in the Biosynthesis of the Unique Starter Unit of the Macrolactam Polyketide Antibiotic Vicenistatin. *J. Antibiot. (Tokyo).* **2005**, *58* (7), 468–472.

(8) Takaishi, M.; Kudo, F.; Eguchi, T. A Unique Pathway for the 3-Aminobutyrate Starter

Unit from l -Glutamate through β-Glutamate during Biosynthesis of the 24-Membered Macrolactam Antibiotic, Incednine. *Org. Lett.* **2012**, *14* (17), 4591–4593.

(9) Jørgensen, H.; Degnes, K. F.; Sletta, H.; Fjærvik, E.; Dikiy, A.; Herfindal, L.; Bruheim, P.; Klinkenberg, G.; Bredholt, H.; Nygård, G.; et al. Biosynthesis of Macrolactam BE-14106 Involves Two Distinct PKS Systems and Amino Acid Processing Enzymes for Generation of the Aminoacyl Starter Unit. **2009**, *16* (10), 1109–1121.

(10) Jørgensen, H.; Degnes, K. F.; Dikiy, A.; Fjærvik, E.; Klinkenberg, G.; Zotchev, S. B. Insights into the Evolution of Macrolactam Biosynthesis through Cloning and Comparative Analysis of the Biosynthetic Gene Cluster for a Novel Macrocyclic Lactam, ML-449. *Appl. Environ. Microbiol.* **2010**, *76* (1), 283–293.

(11) Amagai, K.; Takaku, R.; Kudo, F.; Eguchi, T. A Unique Amino Transfer Mechanism for Constructing the β-Amino Fatty Acid Starter Unit in the Biosynthesis of the Macrolactam Antibiotic Cremimycin. *ChemBioChem* **2013**, *14* (15), 1998–2006.

(12) Chisuga, T.; Miyanaga, A.; Kudo, F.; Eguchi, T. Structural Analysis of the Dual-Function Thioesterase SAV606 Unravels the Mechanism of Michael Addition of Glycine to an α,β-Unsaturated Thioester. *J. Biol. Chem.* **2017**, *292* (26), 10926–10937.

(13) Miyanaga, A.; Cieślak, J.; Shinohara, Y.; Kudo, F.; Eguchi, T. The Crystal Structure of the Adenylation Enzyme VinN Reveals a Unique β-Amino Acid Recognition Mechanism. **2014**, *289* (45).

(14) Derewacz, D. K.; Covington, B. C.; McLean, J. A.; Bachmann, B. O. Mapping Microbial Response Metabolomes for Induced Natural Product Discovery. *ACS Chem. Biol.* **2015**, *10* (9), 1998–2006.

(15) Shinohara, Y.; Kudo, F.; Eguchi, T. A Natural Protecting Group Strategy To Carry an Amino Acid Starter Unit in the Biosynthesis of Macrolactam Polyketide Antibiotics. *J. Am. Chem. Soc.* **2011**, *133* (45), 18134–18137.

(16) Miyanaga, A.; Iwasawa, S.; Shinohara, Y.; Kudo, F.; Eguchi, T. Structure-Based Analysis of the Molecular Interactions between Acyltransferase and Acyl Carrier Protein in Vicenistatin Biosynthesis. *Proc. Natl. Acad. Sci.* **2016**, *113* (7), 1802–1807.

(17) Shinohara, Y.; Miyanaga, A.; Kudo, F.; Eguchi, T. The Crystal Structure of the Amidohydrolase VinJ Shows a Unique Hydrophobic Tunnel for Its Interaction with Polyketide Substrates. *FEBS Lett.* **2014**, *588* (6), 995–1000.

(18) Zhu, Y.; Zhang, W.; Chen, Y.; Yuan, C.; Zhang, H.; Zhang, G.; Ma, L.; Zhang, Q.; Tian, X.; Zhang, S.; et al. Characterization of Heronamide Biosynthesis Reveals a Tailoring Hydroxylase and Indicates Migrated Double Bonds. *Chembiochem* **2015**, *16* (14), 2086–2093.

(19) Ogasawara, Y.; Katayama, K.; Minami, A.; Otsuka, M.; Eguchi, T.; Kakinuma, K. Cloning, Sequencing, and Functional Analysis of the Biosynthetic Gene Cluster of

Macrolactam Antibiotic Vicenistatin in Streptomyces Halstedii. *Chem. Biol.* **2004**, *11* (1), 79–86.

(20) Schulz, D.; Nachtigall, J.; Geisen, U.; Kalthoff, H.; Imhoff, J. F.; Fiedler, H.-P.; Süssmuth, R. D. Silvalactam, a 24-Membered Macrolactam Antibiotic Produced by Streptomyces Sp. Tü 6392*. *J. Antibiot. (Tokyo).* **2012**, *65* (7), 369–372.

(21) Malmierca, M. G.; Pérez-Victoria, I.; Martín, J.; Reyes, F.; Méndez, C.; Olano, C.; Salas, J. A. Cooperative Involvement of Glycosyltransferases in the Transfer of Aminosugars in the Biosynthesis of the Macrolactam Sipanmycin by *Streptomyces* Sp. CS149. *Appl. Environ. Microbiol.* **2018**, AEM.01462-18.

(22) Schulz, D.; Nachtigall, J.; Geisen, U.; Kalthoff, H.; Imhoff, J. F.; Fiedler, H.-P.; Süssmuth, R. D. Silvalactam, a 24-Membered Macrolactam Antibiotic Produced by Streptomyces Sp. Tü 6392*. *J. Antibiot. (Tokyo).* **2012**, *65* (7), 369–372.

(23) Naruse, N.; Tenmyo, O.; Kawano, K.; Tomita, K.; Ohgusa, N.; Miyaki, T.; Konishi, M.; Oki, T. Fluvirucins A1, A2, B1, B2, B3, B4 and B5, New Antibiotics Active against Influenza A Virus. I. Production, Isolation, Chemical Properties and Biological Activities. *J. Antibiot. (Tokyo).* **1991**, *44* (7), 733–740.

(24) Shindo, K.; Kamishohara, M.; Odagawa, A.; Matsuoka, M.; Kawai, H. Vicenistatin, a Novel 20-Membered Macrocyclic Lactam Antitumor Antibiotic. *J. Antibiot. (Tokyo).* **1993**, *46* (7), 1076–1081.

(25) Leutou, A. S.; Yang, I.; Le, T. C.; Hahn, D.; Lim, K.-M.; Nam, S.-J.; Fenical, W. Fluvirucin B6, a New Macrolactam Isolated from a Marine-Derived Actinomycete of the Genus Nocardiopsis. *J. Antibiot. (Tokyo).* **2018**, *71* (6), 609–612.

(26) Ogasawara, Y.; Katayama, K.; Minami, A.; Otsuka, M.; Eguchi, T.; Kakinuma, K. Cloning, Sequencing, and Functional Analysis of the Biosynthetic Gene Cluster of Macrolactam Antibiotic Vicenistatin in Streptomyces Halstedii. *Chem. Biol.* **2004**, *11* (1), 79–86.

(27) Thibodeaux, C. J.; Melançon, C. E.; Liu, H. Natural-Product Sugar Biosynthesis and Enzymatic Glycodiversification. *Angew. Chemie Int. Ed.* **2008**, *47* (51), 9814–9859.

(28) Minami, A.; Eguchi, T. Substrate Flexibility of Vicenisaminyltransferase VinC Involved in the Biosynthesis of Vicenistatin. *J. Am. Chem. Soc.* **2007**, *129* (16), 5102–5107.

(29) Raju, R.; Piggott, A. M.; Conte, M. M.; Capon, R. J. Heronamides A–C, New Polyketide Macrolactams from an Australian Marine-Derived Streptomyces Sp. A Biosynthetic Case for Synchronized Tandem Electrocyclization. *Org. Biomol. Chem.* **2010**, *8* (20), 4682-4689.

(30) Booth, T. J.; Alt, S.; Capon, R. J.; Wilkinson, B. Synchronous Intramolecular Cycloadditions of the Polyene Macrolactam Polyketide Heronamide C. *Chem. Commun. (Camb).* **2016**, *52* (38), 6383–6386.

(31) Zhu, Y.; Zhang, W.; Chen, Y.; Yuan, C.; Zhang, H.; Zhang, G.; Ma, L.; Zhang, Q.; Tian, X.; Zhang, S.; et al. Characterization of Heronamide Biosynthesis Reveals a Tailoring Hydroxylase and Indicates Migrated Double Bonds. *ChemBioChem* **2015**, *16* (14), 2086–2093.

(32) Yu, P.; Patel, A.; Houk, K. N. Transannular [6 + 4] and Ambimodal Cycloaddition in the Biosynthesis of Heronamide A. *J. Am. Chem. Soc.*, **2015**, *137* (42), 13518–13523.

(33) Hoshino, S.; Okada, M.; Awakawa, T.; Asamizu, S.; Onaka, H.; Abe, I. Mycolic Acid Containing Bacterium Stimulates Tandem Cyclization of Polyene Macrolactam in a Lake Sediment Derived Rare Actinomycete. *Org. Lett.* **2017**, *19* (18), 4992–4995.

(34) Hoshino, S.; Ozeki, M.; Wong, C. P.; Zhang, H.; Hayashi, F.; Awakawa, T.; Morita, H.; Onaka, H.; Abe, I. Mirilactams C–E, Novel Polycyclic Macrolactams Isolated from Combined-Culture of Actinosynnema mirum NBRC 14064 and Mycolic Acid-Containing Bacterium. *Chem. Pharm. Bull.* **2018**, *66* (6), 660–667.

(35) Ding, N.; Han, L.; Jiang, Y.; Li, G.; Zheng, Z.; Cao, B.; Guan, P.; Mu, Y.; Lin, B.; Huang, X. Heronamides G–L, Polyene Macrolactams from *Streptomyces niveus*. *RSC Adv.* **2018**, *8* (31), 17121–17131.

(36) H. Müller , E. Bischoff , V. B. Fiedler , K. Weber , B. Fugmann , B. Rosen, Bayer AG. No Title. *Chem. Abstr.* **1994**, *121* (7441).

(37) Schulze, C. J.; Donia, M. S.; Siqueira-Neto, J. L.; Ray, D.; Raskatov, J. A.; Green, R. E.; McKerrow, J. H.; Fischbach, M. A.; Linington, R. G. Genome-Directed Lead Discovery: Biosynthesis, Structure Elucidation, and Biological Evaluation of Two Families of Polyene Macrolactams against *Trypanosoma Brucei*. *ACS Chem. Biol.* **2015**, *10* (10), 2373–2381.

(38) Thawai, C.; Kittakoop, P.; Tanasupawat, S.; Suwanborirux, K.; Sriklung, K.; Thebtaranonth, Y. Micromonosporin?A, a Novel 24-Membered Polyene Lactam Macrolide FromMicromonospora Sp. Isolated from Peat Swamp Forest. *Chem. Biodivers.* **2004**, *1* (4), 640–645.

(39) Schulz, D.; Nachtigall, J.; Riedlinger, J.; Schneider, K.; Poralla, K.; Imhoff, J. F.; Beil, W.; Nicholson, G.; Fiedler, H.-P.; Süssmuth, R. D. Piceamycin and Its N-Acetylcysteine Adduct Is Produced by Streptomyces sp. GB 4-2. *J. Antibiot. (Tokyo).* **2009**, *62* (9), 513–518.

(40) Kinsky, S. C. Effect of Polyene Antibiotics on Protoplasts of Neurospora crassa. *J. Bacteriol.* **1962**, *83* (2), 351–358.

(41) Andreoli, T. E.; Monahan, M. The Interaction of Polyene Antibiotics with Thin Lipid Membranes. *J. Gen. Physiol.* **1968**, *52* (2), 300–325.

(42) Bolard, J. How Do the Polyene Macrolide Antibiotics Affect the Cellular Membrane Properties? *Biochim. Biophys. Acta* **1986**, *864* (3–4), 257–304.

(43) Matsushima, Y.; Nakayama, T.; Fujita, M.; Bhandari, R.; Eguchi, T.; Shindo, K.; Kakinuma, K. Isolation and Structure Elucidation of Vicenistatin M, and Importance of the Vicenisamine Aminosugar for Exerting Cytotoxicity of Vicenistatin. *J. Antibiot. (Tokyo).* **2001**, *54* (3), 211–219.

(44) Shibata, K.; Satsumabayashi, S.; Sano, H.; Komiyama, K.; Zhi-Bon, Y.; Nakagawa, A.; Omura, S. Chemical Modification of Hitachimycin. II. Synthesis and Antitumor Activities of Carbonate Derivatives. *J. Antibiot. (Tokyo).* **1989**, *42* (5), 718–726.

(45) Mitchell, S. S.; Nicholson, B.; Teisan, S.; Lam, K. S.; Potts, B. C. M. Aureoverticillactam, a Novel 22-Atom Macrocyclic Lactam from the Marine Actinomycete *Streptomyces aureoverticillatus* †. *J. Nat. Prod.* **2004**, *67* (8), 1400–1402.

(46) Li, L.; Cai, Y.; Jiang, Y.; Liu, J.; Ma, J.; Yuan, C.; Mu, Y.; Han, L.; Huang, X. A Unique Macrolactam Derivative via a [4+6]-Cycloaddition from Streptomyces Niveus. *Bioorg. Med. Chem. Lett.* **2016**, *26* (6), 1599–1604.

(47) Fukuda, H.; Nishiyama, Y.; Nakamura, S.; Ohno, Y.; Eguchi, T.; Iwabuchi, Y.; Usui, T.; Kanoh, N. Synthesis and Structure-Activity Relationship of Vicenistatin, a Cytotoxic 20-Membered Macrolactam Glycoside. *Chem. - An Asian J.* **2012**, *7* (12), 2872–2881.

(48) Kojiri, K.; Nakajima, S.; Suzuki, H.; Kondo, H.; Suda, H. A New Macrocyclic Lactam Antibiotic, BE-14106. I. Taxonomy, Isolation, Biological Activity and Structural Elucidation. *J. Antibiot. (Tokyo).* **1992**, *45* (6), 868–874.

(49) Oh, D.-C.; Poulsen, M.; Currie, C. R.; Clardy, J. Sceliphrolactam, a Polyene Macrocyclic Lactam from a Wasp-Associated *Streptomyces* Sp. *Org. Lett.* **2011**, *13* (4), 752–755.

(50) Sugiyama, R.; Nishimura, S.; Matsumori, N.; Tsunematsu, Y.; Hattori, A.; Kakeya, H. Structure and Biological Activity of 8-Deoxyheronamide C from a Marine-Derived *Streptomyces* Sp.: Heronamides Target Saturated Hydrocarbon Chains in Lipid Membranes. *J. Am. Chem. Soc.* **2014**, *136* (14), 5209–5212.

(51) Chen, H.; Cai, K.; Yao, R. A New Macrolactam Derivative from the Marine Actinomycete HF-11225. *J. Antibiot. (Tokyo).* **2018**, *71* (4), 477–479.

(52) Raju, R.; Piggott, A. M.; Conte, M. M.; Capon, R. J. Heronamides A-C, New Polyketide Macrolactams from an Australian Marine-Derived Streptomyces Sp. A Biosynthetic Case for Synchronized Tandem Electrocyclization. *Org. Biomol. Chem.* **2010**, *8* (20), 4682–4689.

(53) Nishiyama, Y.; Ohmichi, T.; Kazami, S.; Iwasaki, H.; Mano, K.; Nagumo, Y.; Kudo, F.; Ichikawa, S.; Iwabuchi, Y.; Kanoh, N.; et al. Vicenistatin Induces Early Endosome-Derived Vacuole Formation in Mammalian Cells. *Biosci. Biotechnol. Biochem.* **2016**, *80* (5), 902–910.

(54) Medema, M. H.; Takano, E.; Breitling, R. Detecting Sequence Homology at the Gene Cluster Level with MultiGeneBlast. *Mol. Biol. Evol.* **2013**, *30* (5), 1218–1223.

(55) Nouioui, I.; Carro, L.; García-López, M.; Meier-Kolthoff, J. P.; Woyke, T.; Kyrpides, N. C.; Pukall, R.; Klenk, H.-P.; Goodfellow, M.; Göker, M. Genome-Based Taxonomic Classification of the Phylum Actinobacteria. *Front. Microbiol.* **2018**, *9*, 2007.

(56) Kudo, F.; Kawamura, K.; Uchino, A.; Miyanaga, A.; Numakura, M.; Takayanagi, R.; Eguchi, T. Genome Mining of the Hitachimycin Biosynthetic Gene Cluster: Involvement of a Phenylalanine-2,3-Aminomutase in Biosynthesis. *ChemBioChem* **2015**, *16* (6), 909–914.

(57) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23* (21), 2947–2948.

(58) Stamatakis, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **2014**, *30* (9), 1312–1313.

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

(60) Tamura, S.; Takahashi, N.; Miyamoto, S.; Mori, R. Isolation and Physiological Activities of Piericidin A, A Natural Insecticide Produced by Streptomyces. *Agric. Biol. Chem.* **1963**, *27* (8), 576–582.

(61) Hall, C.; Wu, M.; Crane, F. L.; Takahashi, H.; Tamura, S.; Folkers, K. Piericidin A: A New Inhibitor of Mitochondrial Electron Transport. *Biochem. Biophys. Res. Commun.* **1966**, *25* (4), 373–377.

(62) Lin, J.; Yoshida, S.; Takahashi, N. Metabolites Produced by *Streptomyces Mobaraensis*. *Agric. Biol. Chem.* **1971**, *35* (3), 363–369.

Chapter Five: Establishing the production of complex polyketides in *Streptomyces* species through the overexpression of pathway-specific activators

Contents

5. 1. Introduction

5. 1. 1. Activating Silent Biosynthetic Gene Clusters (BGCs)

As discussed in Chapter 1, the sequencing of *S. coelicolor* A3(2) revealed a wealth of unexplored specialised metabolism[1]. The sequencing of *S. avermatillis*[2] and *S. griseus*[3] genomes confirmed that *S. coelicolor* was unremarkable among the genus with respect to expansive specialised metabolism. Other prolific antibiotic producing actinobacterial genera such as *Saccharopolyspora* and *Salinospora* were also shown to have untapped metabolic potential[4,5]. In the proceeding decade, thousands of actinobacterial genome sequences have been published. Indeed, even with the vast amount of actinobacterial genomes sequenced and the number of catalogued clusters it is not uncommon to find a biosynthetic gene cluster (BGC) that is unique to a given strain, and modern predictive algorithms such as ClusterFinder[6] suggest that the extent of specialised metabolism is even larger than previously thought. Given the scale of unknown specialised metabolism it is unsurprising that last decade has seen considerable efforts to discover the extent of so-called 'cryptic' metabolism encoded by silent BGCs.

Streptomyces platensis MER 11107 is a known producer of two polyketides, the anticancer compound pladienolide B[7] and the spirotetronate spirohexenolide[8], the biosynthesis of which has yet to be elucidated. Recently, our collaborators at BioAustralis identified two additional strains of *S. platensis* (AS600 and M-5455) that also produce pladienolide and spirohexenolide. However, during the process of domestication (culturing wild strains reproducibly in the lab) stable production of these compounds was lost. This chapter describes the reactivation of these two BGCs and reestablishment of pladienolide and spirohexenolide biosynthesis.

5. 1. 1. 1. Cultivation Based Techniques

Many methods have been developed for activating silent BGCs and increasing specialised metabolite production. Perhaps the most straightforward method is through the alteration of culture conditions. This is often referred to as the One Strain Many Compounds, or (OSMAC), approach, whereby a single strain is grown under varying culture conditions, including media composition, temperature, pH and even pressure to examine the biosynthetic capacity of a given strain[9,10]. Other examples include intentionally triggering stress responses through heat, ethanol, acidic or osmotic shock[11–15]. Specific chemical elicitors may also be added to the culture media [16]. Among many others, *N*-acetyl glucosamine, chloramphenicol and rare earth elements, such scandium, can provoke the overproduction of antibiotics and activate previously dormant clusters[17–20].

Many recent efforts to activate specialised metabolism involve taking advantage of bacterial interactions. This is based on two hypotheses. Firstly, BGCs maintained in the genome

must be expressed under native conditions in order to confer a selective advantage or would be lost rapidly due to the accumulation of mutations. Hence, attempting to simulate the natural environment, where bacteria do not exist in monoculture but as consortia, may activate latent metabolism[21–23]. The second hypothesis is based on the observation that the onset of specialised metabolism is often triggered by signalling molecules and the assumption that interspecies crosstalk of such signalling factors can activate the production of metabolites not observed in laboratory monocultures. In actinomycetes, antibiotic production can be induced by γ-butyrolactones, such as A-factor (2-isocapryloyl-3 R-hydroxymethyl-γ-butyrolactone)[24]. The activation of metabolism through the coculture of different *Streptomyces* species has been shown to be a widespread phenomenom[25,26]. The development of sensitive imaging mass spectrometry techniques such as nanoDESI (nanospray desorption electrospray) have demonstrated the depth and specificity of actinobacterial interactions[21,27].

Examples of β-amino acid containing polyketide macrolactams (βPMs), discussed in Chapters Three and Four, discovered by coculture methodology are ciromicin A and B (discovered through to co-culture of *Nocardiopsis sp.* and *Rhodococcus wratislaviensis*)[28] and the niizalactams A-C (discovered through the coculture of *Streptomyces sp.* and *Tsukamurella pulmonis*)[29].

5. 1. 1. 2. Mutagenesis

Another traditional method of altering specialised metabolism is through mutagenesis. Strains are exposed to chemical mutagens such as ethyl/methyl methane sulfonate (EMS/MMS), methylnitronitrosoguanidine (MNNG), or to UV light, and the surviving progeny are subjected to metabolic analysis to select favourable phenotypes[30]. For a relevant example, this strategy was implemented during the discovery of the spirohexenolides, as will be discussed later in this chapter[8]. Another approach, developed by Ochi and co-workers in the early 2000s, involves mutation through the generation of antibiotic resistant strains. Streptomycin, gentamycin and rifampicin resistant mutants have been shown to increase antibiotic production by up to fifty-fold[31]. Resistance to these antibiotics is most commonly achieved by mutations in the *rpsL* (streptomycin) and *rpoB* (rifampicin) genes encoding the ribosomal protein S12 and RNA polymerase β-subunit respectively[32–35]. More specifically, these mutations are believed to allow the cell to bypass ppGpp-dependent upregulation of specialised metabolism[31,32]. In addition to upregulating production, this technique is also able to activate the production of natural products by previously silent BGCs[36,37].

5. 1. 1. 3 Molecular Biological Techniques

Despite their successes, the approaches described above do not guarantee the activation of specific BGCs. Hence, targeted techniques are often required. Molecular biological techniques fall in to two general categories: heterologous expression and endogenous modification. Heterologous expression involves the 'capture' of whole BGCs. This can be achieved through generation of large scale libraries from phage P1 or bacterial artificial chromosomes (PAC or BAC) or targeted methods such as TAR cloning or the assembly of PCR fragments through Gibson or yeast-mediated assembly[38–42]. The BGC of interest can then be expressed in a suitable heterologous host. For *Streptomyces* these are commonly engineered derivatives of *S. coelicolor* M145, dubbed 'superhosts'[43]. Captured BGCs are then amenable to modifications, such as gene deletions or the modification of promotors and other regulatory elements[44–47].

Alternatively, endogenous modification involves the genetic manipulation of the native producer. Typically, this involves the overexpression of transcriptional activators or the knocking-out of repressors.

5. 1. 2. Transcriptional Regulation of Specialised Metabolism in *Streptomyces*

Antibiotic production is ultimately underpinned by genetic regulation, more specifically the regulation of transcription. Transcription factors can act either as activators (by stabilising RNA polymerase-promoter binding) or repressors (by interfering with activator binding or inhibiting the recruitment of RNA polymerase by binding promotor regions[48,49]. Therefore, activation of silent BGCs, or overproduction of specific metabolites, can be achieved by overexpressing transcriptional activators[50,51] or knocking-out repressors[52–55].

A list of common transcriptional regulators involved in Actinobacterial specialised metabolism can be found in Table 5. 1 (adapted from Romero-Rodríguez et al.[56]). This chapter describes the overexpression of two pathway specific activators of the LuxR and SARP families in *Streptomyces platensis* AS600 and M-5455.

5. 1. 2. 1. LuxR Family

Many BGCs in *Streptomyces* are regulated by LuxR-like proteins, specifically of the LAL family (large ATP-binding regulators of the LuxR family). LAL-family proteins are characterised by an N-terminal ATP binding domain and a C-terminal helix-turn-helix DNA-binding domain. The first member of this family to be described was MalT from the *E.coli* maltose regulon[57]. The ability of MlaT to activate transcription is dependent upon the binding (but not hydrolysis) of ATP. The constitutive overexpression of LAL-family genes can be sufficient to activate cryptic biosynthesis[51,58]. The stambomycins (A-D) were discovered in this fashion[51]. Transcriptional

Family	Action	Regulated Functions
LuxR	Activator	Specialised metabolism
SARP	Activator	Specialised metabolism
LysR	Activator/Repressor	Specialised metabolism
MarR	Activator/Repressor	Specialised metabolism, central catabolism and oxidative stress resistance
DeoR	Repressor	Specialised metabolism and morphological development
GntR	Repressor	Specialised metabolism, central carbonmetabolism and mycelium formation
TetR	Repressor	Antibiotic Resistance

Table 5. 13: Common transcriptional regulators involved in Streptomyces specialised metabolism. Regulated functions described in different Streptomyces spp. and observed mode(s) of regulation.

analysis by RT-PCR showed that PKS encoding genes were poorly expressed under native conditions. However, following overexpression of the positive regulator transcription was activated and the products of the BGC were identified in the culture extracts. It has also been shown that related LAL-family proteins can be used to activate related BGCs from different species. It was shown that the activity of the LAL-family regulator PikD could be complemented by heterologous expression of LAL-family regulators RapH and FkbH from other macrolide producing BGCs[59]. Taking advantage of these relationships may provide a more general strategy for activating cryptic metabolism.

5. 1. 2. 2. SARP Family

The term 'SARP' (or *Streptomyces* antibiotic regulatory protein) was first coined in 1997 by Wietzorrek and Bibb to describe a family of regulatory proteins sharing a related N-terminal DNA-binding domain similar to that of OmpR[60,61]. They are proposed to bind heptameric repeats upstream of the promotor and activate transcription through the coordination of RNA polymerase[61–63]. Prototypical SARPs include ActII-ORF4[61,62,64] and DnrI[65], which are pathway-specific regulators for the production of the type II polyketides actinorhodin. and daunorubicin. However SARPs do regulate other classes of BGCs including type-I polyketide synthases (T1PKS)[66,67], β-lactams[68] and hybrid T1PKS-nonribosomal peptide synthetase BGCs[69,70] among others[71].

It is worth noting that overexpressing SARPs as a strategy for activating silent BGCs comes with caveats. Firstly, although the majority of SARPs have been shown to be pathway-specific, this is not always the case. SARPs can also regulate expression indirectly, through the activation of γ-butyrolactone production[72,73]. In *Streptomyces coelicolor* A3(2), for example, AfsR is a pleiotropic, global regulator of specialised metabolism that has also been shown to bind to targets of the primary metabolic activator PhoP [74–76]. Secondly, and perhaps due to their pleiotropic nature, SARPs are not necessarily involved in the regulation of proximal BGCs, for example the SARP TylT has no function in the regulation of tylosin production, despite being within the boundaries of the cluster[77].

5. 1. 3. Pladienolide

The pladienolides (A-G) are 12-membered macrolides produced by strains of *Streptomyces platensis*[78,79]. They contain a characteristic C18-C19 epoxide and variable acetylation and hydroxylations[79] (Figure 5. 1). Of the seven congeners described, pladienolide B (**13**) is the most bioactive. **13** is a potent modulator of hypoxia signalling through inhibition of vascular endothelial growth factor (VEGF)[78] and has been shown to reduce proliferation of a

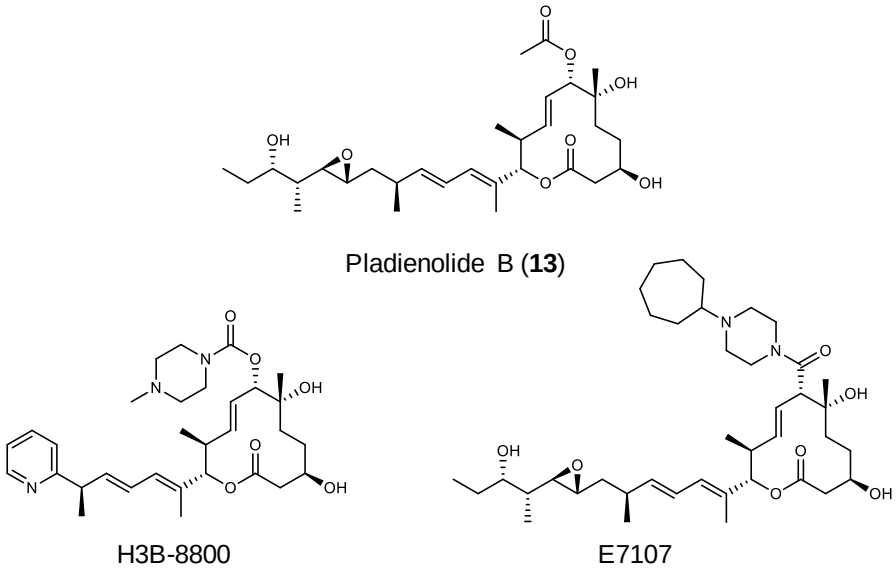


Figure 5. 46. The structure of pladienolide B and its synthetic analogues.

range of lung and breast cancer cell lines[80]. **13** inhibits cell growth by binding to the splicing factor SF3b[81–83], a hotspot for mutation in several forms of cancer[84–86].

Due to this activity, there have been several attempts to chemically diversify the pladienolides. E7107[87,88], a cycloheptanylpiperazine containing analogues of **13**, was shown to cause defects in the formation of spliceosomes by perturbing the interaction between SF3b and one of its partners U2, a small nuclear ribonucleoprotein[89]. Unfortunately, despite promising preclinical data development was suspended during phase I clinical trials due to the development of vision loss in three of sixty-six patients[90,91]. A second pladienolide analogue, H3B-8800 has recently been developed to overcome lack of preferential cytotoxicity[92] and, at the time of writing is currently recruiting for phase I clinical trials[93].

Given their value as lead compounds, there is much interest in discovering efficient routes to pladienolide production. The Pladienolides can be made synthetically, but total synthesis requires between 19 and 31 steps with a maximum yield of ~2% [94,95]. As such an efficient fermentation by a producing strain would be valuable. Two producing strains were previously identified by our collaborators at BioAustralis: *Streptomyces platensis* AS600 and *Streptomyces platensis* M-5455. However, during the domestication process these strains ceased to produce pladienolide. We sought to engineer these strains to efficiently produce pladienolide.

The pladienolide BGC was described in 2008[7] (Figure 5. 2). It consists of eight genes which putatively encode: four PKSs encoding 10 modules (PldAI-IV); post-PKS tailoring enzymes (PldBCD) and; a luxR-family transcriptional regulator (PldR). A *pldB* disrupted strain was only capable of producing 6-deoxy pladienolide B establishing its role as hydroxylase. The role of PldD is inferred as a C18-C19 epoxidase due to the similarity between *pldD* and *monC1* and *nanO* (monensin and nanchangmycin epoxidases respecitvely[4,96]). Overexpression of *pldD* has been shown to increase the production of different pladienolide analogues[97]. PldC was implicated as an *O*-acetylase due to its significant homology with Rif-Orf20[98]. Given the simple architecture of the *pld* BGC we rationalised that overexpression of the *luxR* homologue *pldR* should restore pladienolide production in *S. platensis*.

5. 1. 4. Spirohexenolides

The spirohexenolides (A (**14**) and B (**15**)) are polyketide natural products belonging to the spirotetronate family[8] (Figure 5. 3). **14** has broad anti-tumour activity, with heightened activity against leukaemia (RPMI-8226), lung cancer (HOP-92) and colon cancer (SW-620) cell lines[8]. Both spirohexenolide A and B are readily taken up by HTC-116 Human colorectal carcinoma

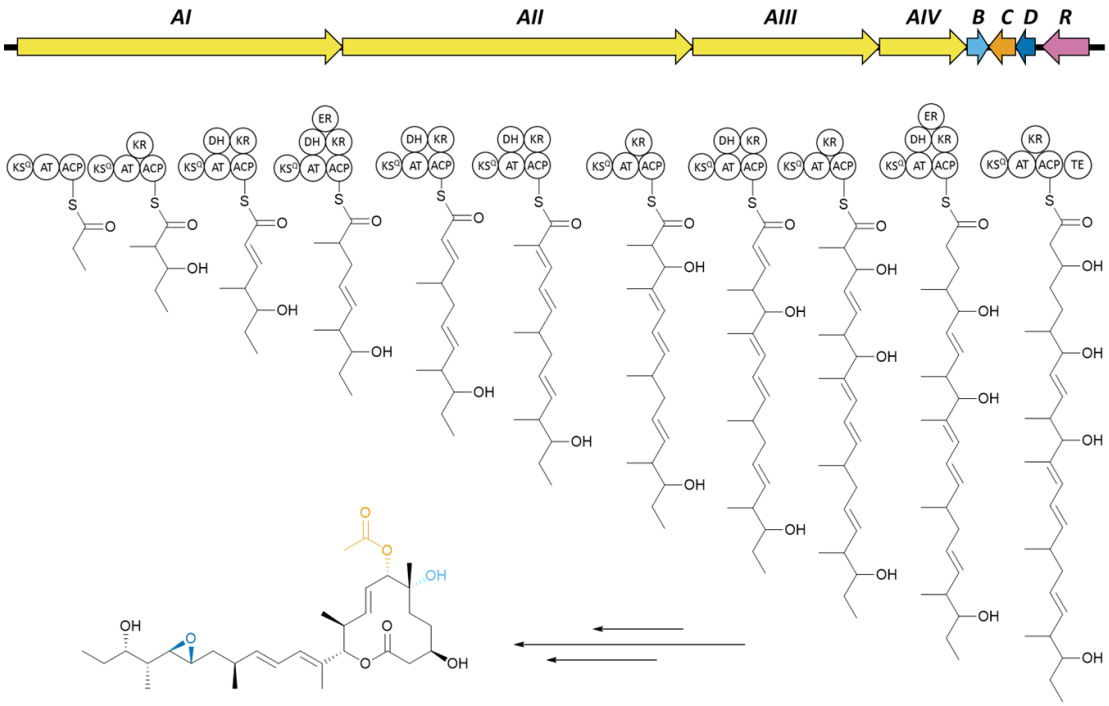


Figure 5. 47: Pladienolide Biosynthesis. Polyketide biosynthesis encoded by pldAI-IV followed by hydroxylation, acetylation, and epoxidation, encoded by pldB, pldC and pldD respectively.

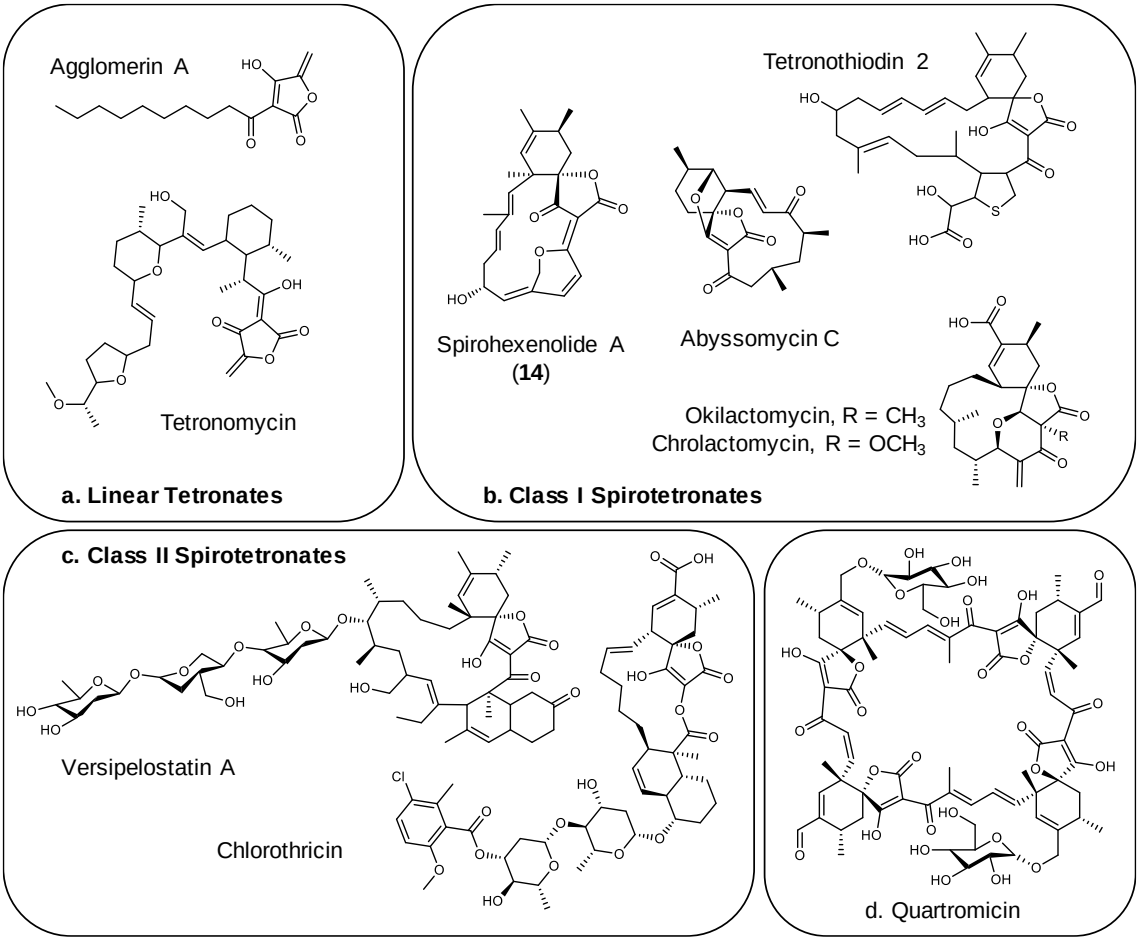


Figure 5. 48: Chemical diversity of spirotetronates. Depicted are a.) the linear tetronates, b.) the class I spirotetronates, c.) the class II spirotetronates, characterised by the decalin ring and d.) quartromicin – a spirotetronate with a unique mode of biosynthesis.

cells[8,99] where they localise to the lysosome and modulate human macrophage inhibitor factor (hMIF)[99], a protein linked to tumorigenesis in colorectal and gastric cancers[100–104].

The spirohexenolides are classified as a type I spirotetronate along with abyssomycins[105–109], okilactomycin[110,111] and chrolactomycin[112,113] (Figure 5. 3. b), sharing the cyclohexene linked tetronic acid moiety, but lacking the *trans*-decalin ring of the type II spirotetronates[114]. Although the BGC of spirohexenolide has not been described, we can predict much of its biosynthesis by looking at the BGCs of other tetronate natural products. Tetronates are the products of type I polyketide synthases (PKSs)[115–118]. Following polyketide synthesis, a C_3-unit derived from bisphosphoglycerate[116,118–120], is fused to the nascent polyketide through by a ketoacyl-ACP synthase[118]. The resulting tetronate moiety then undergoes acetylation and subsequent elimination[118,121] which generates a methylene dienophile which can then undergo [4π +2π] Diels-Alder cycloaddition with the diene of the polyketide chain to form the cyclohexene ring (Figure 5. 4)[116,122,123]. This step is particularly notable as it represents one of the few characterised examples of an enzymatically driven Diels-Alder reaction. In the abyssomycins, this reaction is catalysed by AbyU, a homodimeric, eight-stranded β-barrel[124]. Remarkably, during the biosynthesis of the pyrroindomycins, two Diels-Alder reactions occur, independently catalysed by PyrE3 and PyrI4[125].

The Spirohexenolides are further distinct from the type I spirotetronates due to the presence of a didehydropyran ring and an olefin juncture between the pyran and adjacent tetronic acid moiety[8]. However, the biosynthetic origin of these motifs is currently unknown. We sought to investigate the biosynthesis of the spirohexenolides, but unfortunately neither *S. platensis* AS600 nor M-5455 produced spirohexenolide under laboratory conditions. This is perhaps unsurprising, given that the spirohexenolides were only discovered by UV mutagenesis of the wild-type strain of *S. platensis* MJ[8]. We therefore decided to sequence the genome of *S. platensis* AS600 and M-5455 to identify the spirohexenolide BGC with the aim of gaining insight into its biosynthesis and identifying positive regulators of the BGC that may be overexpressed to activate spirohexenolide production. This would provide a system in which we could interrogate the biosynthesis *in vivo*.

5. 2. Results and Discussion

5. 2. 1. Genome Sequencing and Identification of BGCs

Genomic DNA of *S. platensis* AS600 and *S. platensis* M-5455 was sequenced using the Pacific Biosciences (PacBio) RSII platform[126]. The genome was assembled with HGAP 3.0[127]. *S. platensis* AS600 was assembled in four contigs (4.65 Mb, 3.98 Mb, 309 Kb and 22 Kb) and *S. platensis* M-5455 was assembled in two contigs (8.61 Mb and 297 Kb). Putative biosynthetic BGCs on each contig were identified using antiSMASH v 4.0[128].

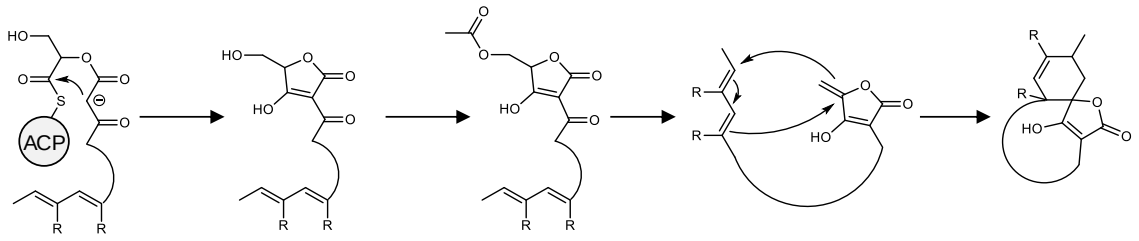


Figure 5. 49: Generalised mechanism of spirotetronate formation. Acylation and deacetylation gives rise to the reactive dienophile. This enables an enzymatically driven [4π + 2π] Diels-Alder reaction to form the mature spirotetronate.

Overall, the two genomes sequences were highly similar (99.4 % pairwise nucleotide identity). However, despite being assembled in fewer contigs, the M-5455 sequence appeared to contain many sequencing errors, resulting in a large proportion of misannotated genes. The two largest contigs of the AS600 strain were homologous to the largest M-5455 contig suggesting these contigs represent a roughly 8.6 Mb linear chromosome of *Streptomyces* platensis. In addition, the 309 Kb and the 297 Kb sequences were homologous suggesting the presence of a linear mega plasmid. The 22 Kb sequence was not homologous to any of the other sequenced contigs, and gene annotation predicted only helicase, transposase and proteins of unknown function, so this sequence was disregarded from further analysis of our strain. The chromosomal sequences contain the same natural product BGCs, which was unsurprising given the overall sequence similarity. The linear plasmid contained a single BGC encoding a putative aryl-polyene.

5. 2. 1. 1. Pladienolide BGC

The pladienolide BGCs from both strains were easily identified due to the high similarity (>95 % nucleotide similarity) with the published *pld* BGG[7]. Overall, the architecture of all three *pld* BGCs are very similar and the sequences show very high similarity overall (Figure 5. 5, Table 5. 2). The only remarkable difference is that the PKS is encoded by five genes in AS600 and M-5455 as opposed to four in the previously reported BGC. With regards to sequence similarity, the clusters are almost identical with the exception of the region internal to *pldAII* from Mer-11107 (the *pldAII-pldAIII* region from the AS600/M-5455 BGCs) corresponding to PKS module six. It is possible that such a pattern has emerged through recent recombination to either fuse or disrupt the gene of interest. Pairwise comparison of the individual *pldA* modules showed a high degree of similarity between all of the modules (99-100% identity), with exception of the module six sequences (72% identity). Interestingly, module six from Mer-11107 showed higher percentage identity with module three (87%). This hinted at a possible recombination between these modules, however self-dot plots of the whole BGC and local alignments suggested that this similarity did not correspond to the correct region. It is possible to envisage a scenario where this has been generated by multiple recombination events.

5. 2. 1. 2. Spirohexenolide BGC

Of the remaining PKS BGCs, the putative spirohexenolide cluster (referred to henceforth as the *spx* locus) was readily identified due to its similarity with the previously reported spirotetronate BGCs (Figure 5. 6, Table 5. 3). The *spx* BGC shares most similarity with the chlorothricin and malkamicin BGCs, however it is clearly distinct. The BGC consists of 5 genes encoding a 9 module PKS *spxA1-A5* and an associated type II thioesterase *spxB*, the glycerate utilisation cassette *spxC1-C4*[119] and a range of other biosynthetic genes. The *spx* BGC has four

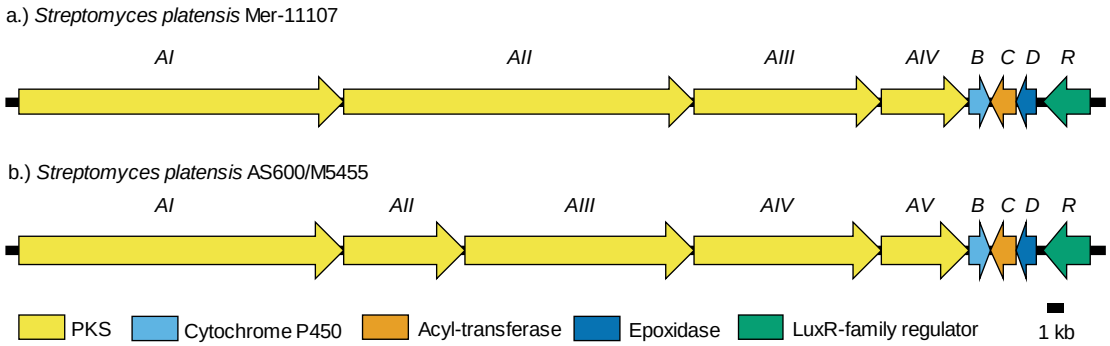


Figure 5. 50: Comparison of pladienolide BGCs. All three pld BGCs show similar overall topology. The only major difference is that the PKS encoded by Mer-1107 pldAII is encoded by two genes, pldAII and pldAIII in AS600 and M-5455.

Protein	Size (a. a.)	Proposed Function	% Similarities/ % Identities with Mer-1107 homologue
PldAI	6543	Polyketide synthase	99/100
PldAII	2373	Polyketide synthase	95/97
PldAIII	4597	Polyketide synthase	90/91
PldAIV	3712	Polyketide synthase	100/100
PldAV	1808	Polyketide synthase	100/100
PldB	399	P450 hydroxylase	99/99
PldC	407	Acyl-transferase	100/100
PldD	462	Epoxidase	100/100
PldR	902	LuxR-like regulator	100/100

Table 5. 14: Description of proteins encoded by the AS600 pladienolide BGC. Proposed function and homology to the Mer-1107 pld BGC s also shown.

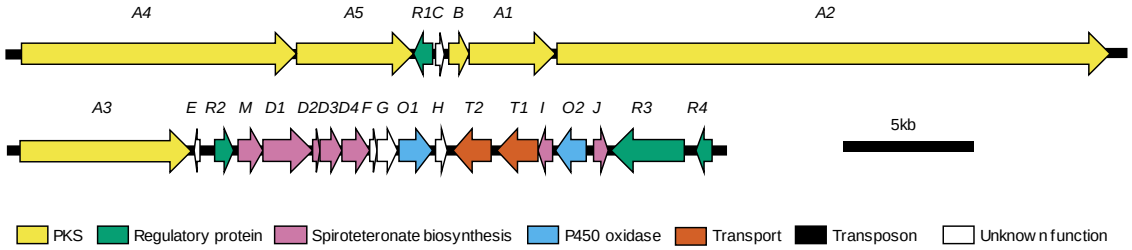


Figure 5. 51: Topology of the Spirohexenolide (spx) BGC.

Protein	Size (aa)	Proposed Function	Closest Homologue (% Similarities/ % Identities)	Closest Known Teteronate BGC Homologue (% Similarities/ % Identities)
SpxA4	3,562	Polyketide Synthase	modular polyketide synthase [Streptomyces sp. RK95-74] (54/66)	ChlA1 [Streptomyces antibioticus] (53/64)
SpxA5	1,511	Polyketide Synthase	type I polyketide synthase [Amycolatopsis sp. CA-126428] (55/64)	N/A
SpxR1	201	TetR family repressor	TetR/AcrR family transcriptional regulator [Streptomyces sp. CNQ-525] (57/75)	N/A
SpxC	108	Unknown	hypothetical protein [Actinocatenispora sera] (51/65)	N/A
SpxB	268	Type II Thioesterase	thioesterase [Streptomyces geranii] (55/63)	AbmT [Streptomyces koyangensis] (45/57)
SpxA1	1074	Polyketide Synthase	polyketide synthase [Allosalinactinospora lopnorensis] (58/70)	ChlA1 [Streptomyces antibioticus] (56/66)
SpxA2	7066	Polyketide Synthase	modular polyketide synthase [Streptomyces sp. RK95-74] (53/64)	N/A
SpxA3	2181	Polyketide Synthase	modular polyketide synthase [Streptomyces neyagawaensis] (58/69)	ChlA2 [Streptomyces antibioticus] (53/64)
SpxE	72	Unknown	hypothetical protein [Streptomyces sp. CNQ-525] (57/67)	N/A
SpxR2	254	SARP-family activator	hypothetical protein [Actinomadura sp. 5-2] (60/70)	N/A
SpxM	343	3-oxoacyl-ACP synthase	3-oxoacyl-ACP synthase [Amycolatopsis sp. CA-126428] (65/80)	AbmA1 [Streptomyces koyangensis] (61/72)
SpxD1	644	glyceryl-ACP synthase	HAD-IIIC family phosphatase [Micromonospora sp. RP3T] (62/74)	ChlD1 [Streptomyces antibioticus] (58/71)
SpxD2	75	Acyl carrier protein	acyl carrier protein [Streptomyces paucisporeus] (64/77)	AbmA3 [Streptomyces koyangensis] (53/69)
SpxD3	272	Acyltransferase	acyltransferase [Actinokineospora inagensis] (72/81)	ChlD3 [Streptomyces antibioticus] (64/75)
SpxD4	365	Hydrolase	alpha/beta hydrolase [Actinomadura chibensis] (59/73)	ChlD4 [Streptomyces antibioticus] (55/68)
SpxF	72	Ferredoxin	MULTISPECIES: ferredoxin [Actinomadura] (54/78)	N/A
SpxG	251	Unknown, putative role in pyran formation	hypothetical protein [Amycolatopsis sp. CA-126428] (40/58)	N/A
SpxO1	407	Cytochrom P450 oxidase	cytochrome P450 [Saccharothrix syringae] (59/74)	NcmO [Saccharothrix syringae] (60/75)
SpxH	152	Unknow, putative role in pyran formation	hypothetical protein [Streptoalloteichus hindustanus] (40/56)	N/A
SpxT3	484	Transport	DHA2 family efflux MFS transporter permease subunit [Gorillibacterium timonense] (56/74)	N/A
SpxT2	503	Transport	hypothetical protein KALB_5017 [Kutzneria albida DSM 43870] (67/78)	N/A
SpxI	183	Diels-Alderase	hypothetical protein [Amycolatopsis sp. CA-126428] (37/55)	N/A
SpxO2	403	Cytochrome P450 Oxidase	cytochrome P450 [Amycolatopsis sp. CA-126428] (65/79)	N/A
SpxJ	187	Diels-Alderase	hypothetical protein [Amycolatopsis sp. CA-126428] (51/68)	QmnH [Amycolatopsis orientalis] (34/48)
SpxR3	957	LuxR-like positive regulator	hypothetical protein ADK64_28580 [Streptomyces sp. MMG1121](38/52)	AmbH [Streptomyces koyangensis] (33/48)
SpxR4	217	TetR family repressor	TetR/AcrR family transcriptional regulator [Streptomyces sp. NBRC 110028] (56/70)	N/A

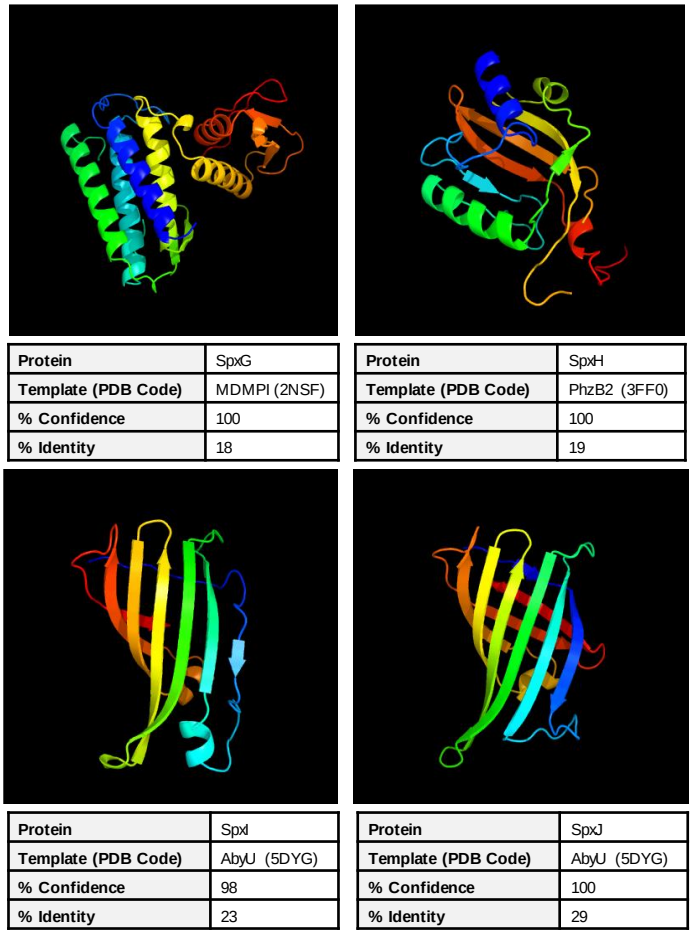
Table 5. 15: Description of the proteins encoded by the spirohexenolide (spx) BGC. Closest homologue from known spirotetronate BGCs shown where a result was in the top fifty BLASTp results.

regulatory genes within its boundaries, two positive regulators of the SARP (*spxR2*) and LuxR (*spxR3*) families, and two negative regulators belonging to the tetR family (*spxR1*, *spxR4*). The boundaries of the BGC are well defined by transposase sequences upstream of *spxA4* and a large non-coding region downstream of *spxR4*. This is supported by synteny with known spirotetronate BGCs. For six proteins (SpxC, SpxE, SpxG, SpxH, SpxI and SpxJ), all BLAST results indicated proteins with previously unknown functions. Molecular modelling with Phyre2[129] provided further insight into their functions. SpxI and SpxJ share structural homology with AbyU, making them likely candidates as Diels-Alderases (Figure 5. 7). It is surprising that the *spx* BCG contains

two Diels-Alderase candidates as this is usually only associated with Class II spirotetronate clusters (containing both spirotetronate and decalin moieties). This may be indicative of an ancestral relationship between *spx* and Class II spirotetronate BGCs.

Based on the newly sequenced BGC, we can suggest a probable mechanism for the biosynthesis of spirohexenolide (Figure 5. 8). Firstly, *spxA1-A5* encode a nine module PKS. Following polyketide synthesis, the proteins encoded by the conserved glycerate utilisation cassette (*spxD1-D4* and *spxM*) catalyse the formation of the tetronate moiety. Next, the [4π +2π] cycloaddition is catalysed by SpxI or SpxJ. The final steps are, at the time of writing, unique to spirohexenolide and so have yet to be characterised. However, we can suggest a reasonable biosynthetic hypothesis. Phylogenetic analysis of published tetronate and spirotetronate cytochrome P450s showed that *spxO1* and *spxO2* were most similar to the C-8/C-8′/C-26/C-26′ hydroxylase *QmnO*. It is reasonable to assume that one of these enzymes is responsible for C8 hydroxylation. The second is likely to be involved in the hydroxylation of the C21 methyl group to enable the formation of the didehydropyran. Four genes encoding unknown products are present in the *spx* BGC. Two, *spxG* and *spxH,* flank *spxO1* and may potentially be involved in the cyclisation reaction. Molecular modelling[129] of the encoded proteins provides some insight into their function (Figure 5. 7). *spxG* is predicted to be a four-helix bundle, with closest structural similarity to malyl-pyruvate isomerse[130] and metal dependent hydrolases[131]. It is also predicted to utilise mycothiol as a cofactor. Mycothiol dependent proteins may have a range of functions including detoxification of reactive oxygen and nitrogen species, reductions, and isomerizations[132]. Therefore, it may be possible SpxG is involved in the formation of the pyran by shifting the equilibrium of isomerisation towards the desired state (Figure 5. 8). *spxG* shares its fold with the phenazine biosynthetic proteins[133]. PhzA/B that catalyses the condensation of amino-cyclohexanone to form the pyrazine ring. As a condensation is required for the formation of pyran ring, SpxH is another reasonable target for a cyclase.

Given the clustering of *spxG*, *spxO1* and *spxH* and the lack of a homologous sub-cluster in other spirotetronate BGCs, we believe they present strong candidates for didehydropyran



Protein	SpxG
Template (PDB Code)	MDMPI (2NSF)
% Confidence	100
% Identity	18

Protein	SpxH
Template (PDB Code)	PhzB2 (3FF0)
% Confidence	100
% Identity	19

Protein	SpxI
Template (PDB Code)	AbyU (5DYG)
% Confidence	98
% Identity	23

Protein	SpxJ
Template (PDB Code)	AbyU (5DYG)
% Confidence	100
% Identity	29

Figure 5. 52: Molecular modelling of hypothetical proteins using Phyre2. SpxG and SpxH are hypothesised to have putative roles in pyran formation. SpxI and SpxJ are putative Diels-Alderases.

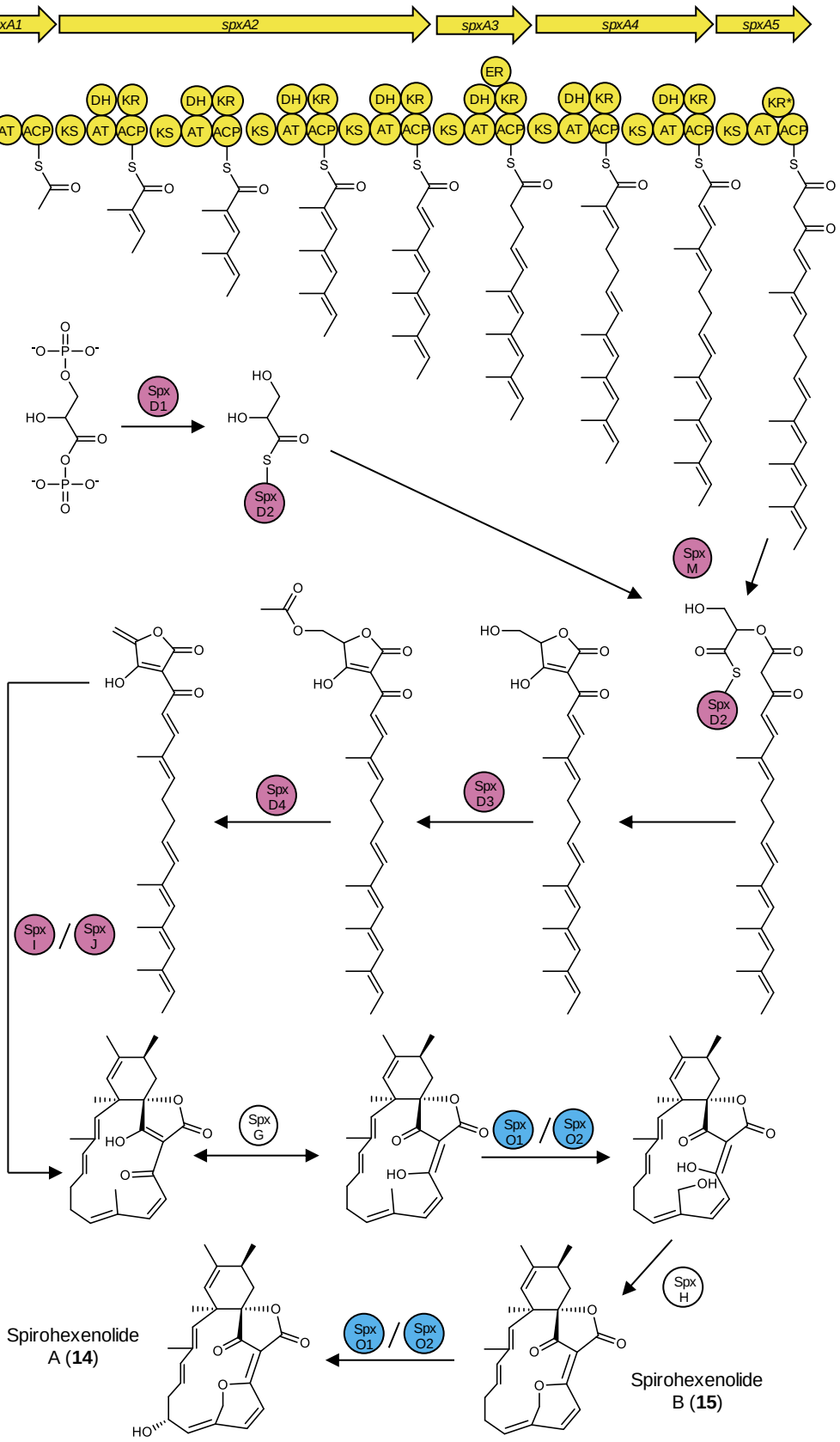


Figure 5. 53: Biosynthetic hypothesis of the spirohexenolides. Colours correspond to predicted function (see Figure 6). Where more than one protein is shown, it is not clear which is responsible for the catalysis. Proteins coloured by putative function (see Figure 5. 6).

formation. *In vivo* knock-outs and subsequent *in vitro* evaluation of these genes and their products will provide insight into their role in the biosynthesis of spirohexenolide.

5. 2. 2. Overexpression of Transcriptional Activators

The sequencing of the *Streptomyces platensis* AS600 and M-5455 genomes allowed us to identify positive transcriptional regulators from both the *pld* and *spx* BGCs. In order to constitutively express these transcriptional activators *in vivo*, several derivatives of the ActII-ORF4 expression vector pGP9 were generated. Activators *pldR*, *spxR2* and *spxR3* were cloned into the NdeI and HindIII sites of pGP9. As ActII-ORF4 is also a SARP-family protein, empty vectors were also used as a negative control. Plasmids were transformed into *S. platensis* AS600 and M-5455 by conjugation with *E. coli* and heat-shocked spores and the selection of apramycin resistant progeny.

5. 2. 3. Upregulation of Specialised Metabolism

Apramycin resistant ex-conjugants were culture on the following media: MYM, SF+M, PYE and SM18 for 7 days at 30°C. 1 cm agar plugs were mixed with EtOAc (1 ml) for 1 h. Ethyl acetate was removed *via* evaporate and the residual solved in methanol. Extracts were analysed by LC-MS to detect the presence of the desired compound.

5. 2. 3. 1. Pladienolide Production

Pladienolide B (**13**) was identified by LC-MS analysis of EtOAc culture extracts. As reported, **13** was not produced by the wild-type strains in any of the media tested. The overexpression of *pldR* restored pladienolide production to M-5455 in SM18 media. Masses for the related deoxy-pladienolide B ($[M+H]^+$ = 521.3443 (+ 2.791 ppm); $[M+Na]^+$= 543.3214 (-5.9000 ppm)) were also observed. Surprisingly, the overexpression of *pldR* appeared to have pleotropic effects and led to the production of a number of additional compounds (Table 5. 4). Given the supposed specificity of LAL-family regulators, the activation of additional metabolic pathways was surprising. However, the alteration of metabolism or regulation may have many off target effects, so further investigation would be required to determine the cause of these metabolic changes. The most conspicuous of these was **17** produced in MYM-TAP along with **18**, a predicted methyl derivative. Isolation of this compound and identification of the corresponding BGC may provide clues as to whether production is due to crosstalk between related regulatory proteins or through an indirect mechanism.

Compound	Formula	Mass					Retention Time (min)	Average Intensity							
		Predicted mass			Experimental mass			M5455+pGP9				M5455+pGP9-pldR			
		M	$[M+H]^+$	$[M+Na]^+$	$[M+H]^+$	$[M+Na]^+$		MYM-TAP	PYE	SF+M	SM18	MYM-TAP	PYE	SF+M	SM18
13	$C_{30}H_{48}O_8$	536.3349	537.3428	559.3247	537.3043	559.3214	4.91	0	0	0	0	0	0	0	414608
Deoxy-**13**	$C_{30}H_{48}O_7$	520.3359	521.3438	543.3257	521.3415	543.3238	5.19	0	0	0	0	0	0	0	2493748
16	N/A	673.3562	N/A	N/A	674.3639	696.3461	6.96	0	0	1443226	0	0	0	0	0
17	N/A	600.3521	N/A	N/A	601.3624	623.3395	4.3	0	0	0	0	53213007	0	0	0
18	N/A	586.3305	N/A	N/A	587.3354	609.3232	3.77	0	0	0	0	2908211	0	0	0
19	N/A	362.1408	N/A	N/A	363.1479	385.1313	8.64	0	0	0	0	2000572	1919625	1145327	444330

Table 5. 16: Metabolites with differential production between M-5455 over expressing pldR and M-5455 containing empty pGP9. Formula of known compounds are shown along with predicted masses. Experimental masses for both hydrogen and sodium adducts are recorded. Average intensity is a measure of peak area of sodium and hydrogen adducts averaged together across biological replicates (darker cell shading indicates greater intensity).

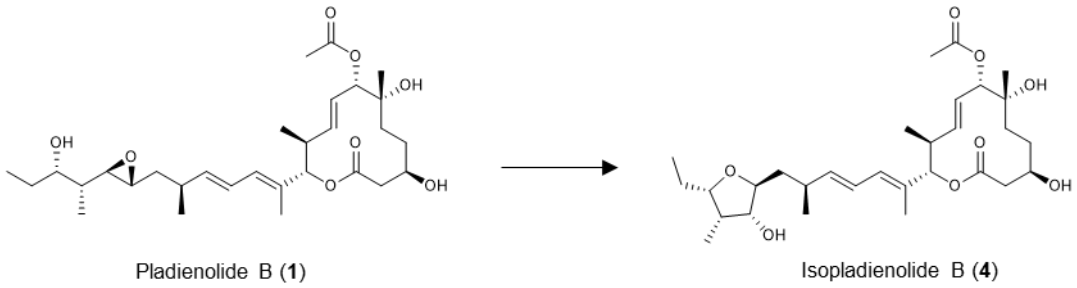
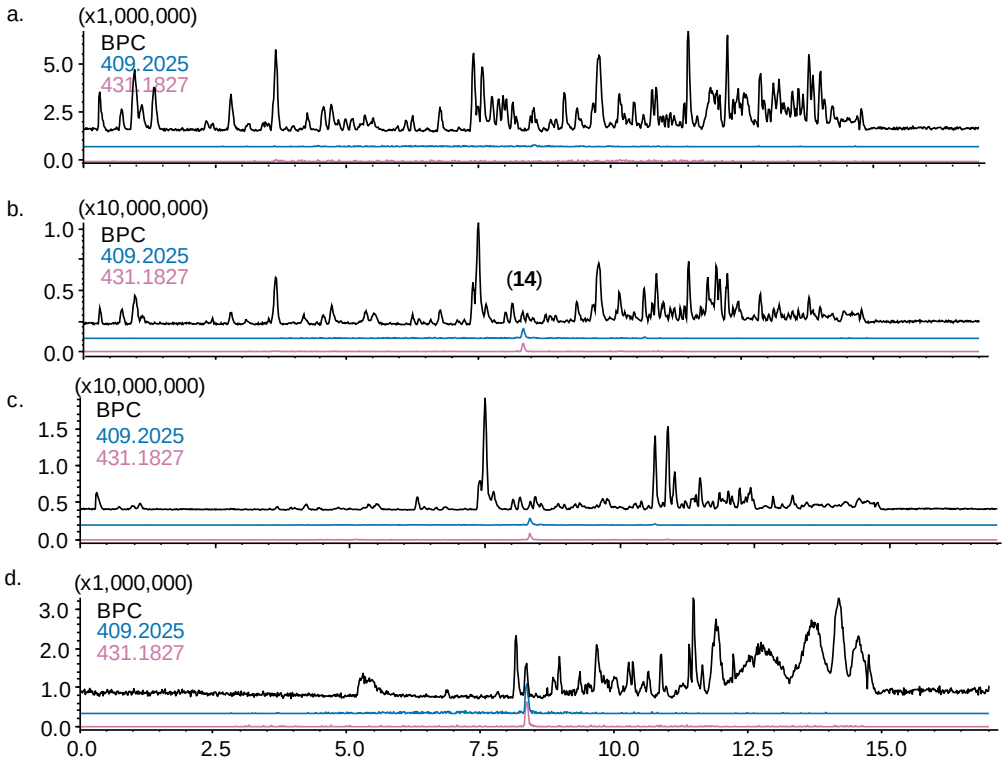


Figure 5. 54: Hypothetical conversion of pladienolide B (1) to isopladienolide B (4).



We

Figure 5. 55: Production of spirohexenolide A by Streptomyces platensis M-5455. LCMS spectra showing the base peak chromatogram (BPC) and extracted ion chromatograms for spirohexenolide A (**14**) $[M+H^+]$ = 409.2025 and $[M+Na]^+$= 431.1827. a.) Streptomyces platensis M-5455-pGP9; b.) Streptomyces platensis M-5455-pGP9-spxR2; c.) Streptomyces platensis M-5455-spxR3; and; d.) authentic standard.

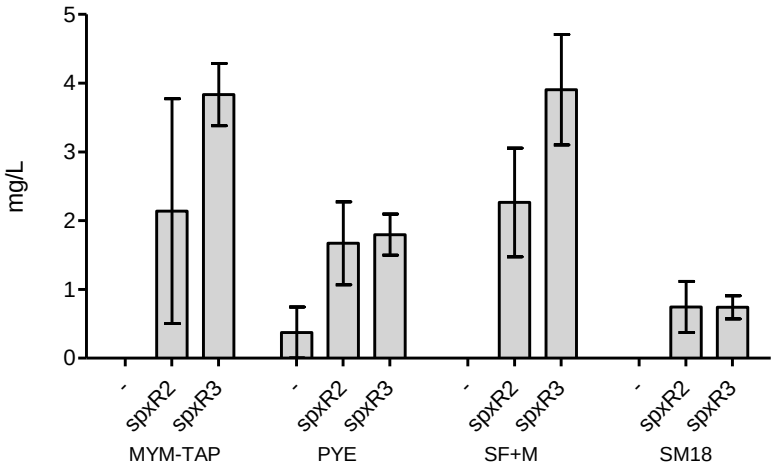


Figure 5. 56: Production of spirohexenolide A by Streptomyces platensis M-5455. Bar chart of titres calculated from standard curve of the authentic standard. Production of spirohexenolide A (**14**) is shown across four media from Streptomyces platensis M-5455-pGP9-spxR2; Streptomyces platensis M-5455-pGP9-spxR3 and; Streptomyces platensis M-5455-pGP9 (-).

provided the producing strains to BioAustralis for media optimisation, who reported stable production of the pladienolides. In addition, they noted that due to the later onset of production many co-contaminants in the purification had broken down thus aiding in the purification process. An additional isomer, isopladienolide B (**20**) was also discovered (Figure 5. 9). Isopladienolide B differs from the rest of the pladienolides as it contains, instead of the C18 -C19 epoxide, a C18-C21 ether forming a furan ring. Unfortunately, **20** showed no biological activity. Although enzymatic conversion of **13** to **20** is possible, given the inherent reactivity of epoxides, it seems more likely that this constitutes a spontaneous conversion. Nonetheless, the structure of **20** demonstrates the potential for isomerisation of the epoxide and may point to the existence of other such variants.

2.3.2 Spirohexenolide production

Spirohexenolide A (**14**) was detected by exact masses ($[M+H]^+$ = 409.2018 (+ 2.1994 ppm), $[M+Na]^+$ = 431.18 (+ 0.4638 ppm)) and compared to an authentic standard (Figure 5. 10).

2.3.2.1 Overexpression of the SARP spxR2

S. platensis M-5455 showed trace production of **14** when cultured on PYE media, however in all other media production of **14** could not be detected in the wild-type strains. Strains overexpressing *spxR2* showed a significant increase in production of **14** across all media tested. The highest predicted titre was 5.4 mg/L in MYM, however variation was quite large between biological replicates. On average, **14** was present in all media extracts in similar quantities (~2 mg/L) (Figure 5. 11).

2.3.2.1 Overexpression of the LuxR-family protein spxR3

As with the *spxR2* overexpressing strains, the pGP9-*spxR3* strains also showed increased **14** production (Figure 5. 11). Paired t-tests showed that the difference in production of **14** between *spxR2* and *spxR3* in any of the media tested. Due to the apparent variability of **14** production in *spxR2* overexpression strain, increasing the sample size may help clarify this result. Unlike strains over expressing *pldR*, the production of additional metabolites was not observed.

5. 3. Conclusion

In this Chapter, I successfully reestabilshed the production of pladienolide in domesticated strains of *Streptomyces platensis* (AS600, M5455). Due to its biological activities, pladienolide and its analogues are of significant research interest[79,87] and is of commercial value to our collaborators. During this process, we sequenced the genomes of AS600 and M-5455.

Access to the genome sequences of these strains enabled the identification of the previously undescribed spirohexenolide BGC (*spx*). As the overexpression of positive regulators proved successful to establish pladienolide production, we repeated this strategy by overexpressing *spxR2* and *spxR3* identified from the *spx* BGC to enable stable production of spirohexenolide A (**14)**. The establishment of stable **14** production provides a platform for testing the biosynthetic hypotheses outlined above. Generating gene knock-outs and complementation of genes in the putative pyran- forming cassette, *spxG*, *spxO1* and *spxH*, would allow us to test our biosynthetic hypothesis. Knockouts of *spxO1* and *spxH* should result in the production of relatively stable intermediates if our hypothesis is correct. The putative Diels-Alderases (DAses) SpxI and SpxJ are also of interest. Given that, unlike other spirotetronate BGCs containing two DAases[134], **14** does not contain a decalin ring, and therefore should only require a single enzyme to catalyse spirotetronate formation. The presence of this second DAase in the BGC is potentially interesting, either as an evolutionary intermediate, or a novel function.

This work demonstrates the importance of overexpressing positive regulators as a strategy, not only to discover novel natural products, but to promote stable production of compounds of interest. Interestingly, the overexpression of hypothetically pathway specific regulators can lead to the production of additional metabolites. The cause of these unexpected pleiotropic effects is still unknown but highlights the complex interplay of regulatory elements in specialised metabolism.

5. 4. References

(1) Bentley, S. D.; Chater, K. F.; Cerdeño-Tárraga, A.-M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; et al. Complete Genome Sequence of the Model Actinomycete Streptomyces coelicolor A3(2). *Nature* **2002**, *417* (6885), 141–147.

(2) Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Ōmura, S. Complete Genome Sequence and Comparative Analysis of the Industrial Microorganism Streptomyces Avermitilis. *Nat. Biotechnol.* **2003**, *21* (5), 526–531.

(3) Ohnishi, Y.; Ishikawa, J.; Hara, H.; Suzuki, H.; Ikenoya, M.; Ikeda, H.; Yamashita, A.; Hattori, M.; Horinouchi, S. Genome Sequence of the Streptomycin-Producing Microorganism Streptomyces Griseus IFO 13350. *J. Bacteriol.* **2008**, *190* (11), 4050–4060.

(4) Oliynyk, M.; Stark, C. B. W.; Bhatt, A.; Jones, M. A.; Hughes-Thomas, Z. A.; Wilkinson, C.; Oliynyk, Z.; Demydchuk, Y.; Staunton, J.; Leadlay, P. F. Analysis of the Biosynthetic Gene Cluster for the Polyether Antibiotic Monensin in Streptomyces cinnamonensis and Evidence for the Role of MonB and MonC Genes in Oxidative Cyclization. *Mol.*

Microbiol. **2003**, *49* (5), 1179–1190.

(5) Udwary, D. W.; Zeigler, L.; Asolkar, R. N.; Singan, V.; Lapidus, A.; Fenical, W.; Jensen, P. R.; Moore, B. S. Genome Sequencing Reveals Complex Secondary Metabolome in the Marine Actinomycete Salinispora Tropica. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (25), 10376–10381.

(6) Cimermancic, P.; Medema, M. H.; Claesen, J.; Kurita, K.; Wieland Brown, L. C.; Mavrommatis, K.; Pati, A.; Godfrey, P. A.; Koehrsen, M.; Clardy, J.; et al. Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. *Cell* **2014**, *158* (2), 412–421.

(7) Machida, K.; Arisawa, A.; Takeda, S.; Tsuchida, T.; Aritoku, Y.; Yoshida, M.; Ikeda, H. Organization of the Biosynthetic Gene Cluster for the Polyketide Antitumor Macrolide, Pladienolide, in Streptomyces platensis Mer-11107. *Biosci. Biotechnol. Biochem.* **2008**, *72* (11), 2946–2952.

(8) Kang, M.; Jones, B. D.; Mandel, A. L.; Hammons, J. C.; DiPasquale, A. G.; Rheingold, A. L.; La Clair, J. J.; Burkart, M. D. Isolation, Structure Elucidation, and Antitumor Activity of Spirohexenolides A and B. *J. Org. Chem.* **2009**, *74* (23), 9054–9061.

(9) Bode, H. B.; Bethe, B.; Höfs, R.; Zeeck, A. Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. *ChemBioChem* **2002**, *3* (7), 619.

(10) Arai, T.; Yazawa, K.; Mikami, Y. Isolation and Characterization of Satellite Antibiotics, Mimosamycin and Chlorocarcins from Streptomyces Lavendulae, Streptothricin Source. *J. Antibiot. (Tokyo).* **1976**, *29* (4), 398–407.

(11) Doull, J. L.; Singh, A. K.; Hoare, M.; Ayer, S. W. Conditions for the Production of Jadomycin B ByStreptomyces Venezuelae ISP5230: Effects of Heat Shock, Ethanol Treatment and Phage Infection. *J. Ind. Microbiol.* **1994**, *13* (2), 120–125.

(12) Hayes, A.; Hobbs, G.; Smith, C. P.; Oliver, S. G.; Butler, P. R. Environmental Signals Triggering Methylenomycin Production by Streptomyces coelicolor A3(2). *J. Bacteriol.* **1997**, *179* (17), 5511–5515.

(13) Yoon, V.; Nodwell, J. R. Activating Secondary Metabolism with Stress and Chemicals. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 415–424.

(14) Bursy, J.; Kuhlmann, A. U.; Pittelkow, M.; Hartmann, H.; Jebbar, M.; Pierik, A. J.; Bremer, E. Synthesis and Uptake of the Compatible Solutes Ectoine and 5-Hydroxyectoine by Streptomyces Coelicolor A3(2) in Response to Salt and Heat Stresses. *Appl. Environ. Microbiol.* **2008**, *74* (23), 7286–7296.

(15) Liao, Y.; Wei, Z.-H.; Bai, L.; Deng, Z.; Zhong, J.-J. Effect of Fermentation Temperature on Validamycin A Production by Streptomyces Hygroscopicus 5008. *J. Biotechnol.* **2009**, *142* (3–4), 271–274.

(16) Craney, A.; Ozimok, C.; Pimentel-Elardo, S. M.; Capretta, A.; Nodwell, J. R. Chemical

Perturbation of Secondary Metabolism Demonstrates Important Links to Primary Metabolism. *Chem. Biol.* **2012**, *19* (8), 1020–1027.

(17) Rigali, S.; Titgemeyer, F.; Barends, S.; Mulder, S.; Thomae, A. W.; Hopwood, D. A.; van Wezel, G. P. Feast or Famine: The Global Regulator DasR Links Nutrient Stress to Antibiotic Production by Streptomyces. *EMBO Rep.* **2008**, *9* (7), 670–675.

(18) Kawai, K.; Wang, G.; Okamoto, S.; Ochi, K. The Rare Earth, Scandium, Causes Antibiotic Overproduction in *Streptomyces* Spp. *FEMS Microbiol. Lett.* **2007**, *274* (2), 311–315.

(19) Tanaka, Y.; Hosaka, T.; Ochi, K. Rare Earth Elements Activate the Secondary Metabolite–biosynthetic Gene Clusters in Streptomyces Coelicolor A3(2). *J. Antibiot. (Tokyo).* **2010**, *63* (8), 477–481.

(20) Tanaka, Y.; Izawa, M.; Hiraga, Y.; Misaki, Y.; Watanabe, T.; Ochi, K. Metabolic Perturbation to Enhance Polyketide and Nonribosomal Peptide Antibiotic Production Using Triclosan and Ribosome-Targeting Drugs. *Appl. Microbiol. Biotechnol.* **2017**, *101* (11), 4417–4431.

(21) Traxler, M. F.; Watrous, J. D.; Alexandrov, T.; Dorrestein, P. C.; Kolter, R. Interspecies Interactions Stimulate Diversification of the Streptomyces coelicolor Secreted Metabolome. *MBio* **2013**, *4* (4), e00459-13.

(22) Traxler, M. F.; Kolter, R. Natural Products in Soil Microbe Interactions and Evolution. *Nat. Prod. Rep.* **2015**, *32* (7), 956–970.

(23) Behie, S. W.; Bonet, B.; Zacharia, V. M.; McClung, D. J.; Traxler, M. F. Molecules to Ecosystems: Actinomycete Natural Products In Situ. *Front. Microbiol.* **2017**, *7*, 2149.

(24) Ohnishi, Y.; Kameyama, S.; Onaka, H.; Horinouchi, S. The A-Factor Regulatory Cascade Leading to Streptomycin Biosynthesis in Streptomyces Griseus : Identification of a Target Gene of the A-Factor Receptor. *Mol. Microbiol.* **1999**, *34* (1), 102–111.

(25) Ueda, K.; Kawai, S.; Ogawa, H.-O.; Kiyama, A.; Kubota, T.; Kawanobe, H.; Beppu, T. Wide Distribution of Interspecific Stimulatory Events on Antibiotic Production and Sporulation Among Streptomyces Species. *J. Antibiot. (Tokyo).* **2000**, *53* (9), 979–982.

(26) Ueda, K.; Beppu, T. Antibiotics in Microbial Coculture. *J. Antibiot. (Tokyo).* **2017**, *70*, 361–365.

(27) Shih, C.-J.; Chen, P.-Y.; Liaw, C.-C.; Lai, Y.-M.; Yang, Y.-L. Bringing Microbial Interactions to Light Using Imaging Mass Spectrometry. *Nat. Prod. Rep.* **2014**, *31* (6), 739.

(28) Derewacz, D. K.; Covington, B. C.; McLean, J. A.; Bachmann, B. O. Mapping Microbial Response Metabolomes for Induced Natural Product Discovery. *ACS Chem. Biol.* **2015**, *10* (9), 1998–2006.

(29) Hoshino, S.; Okada, M.; Wakimoto, T.; Zhang, H.; Hayashi, F.; Onaka, H.; Abe, I. Niizalactams A–C, Multicyclic Macrolactams Isolated from Combined Culture of

Streptomyces with Mycolic Acid-Containing Bacterium. *J. Nat. Prod.* **2015**, *78* (12), 3011–3017.

(30) Baltz, R. H. New Genetic Methods to Improve Secondary Metabolite Production in Streptomyces. *J. Ind. Microbiol. Biot.* **1998**, *20* (6),360-363.

(31) Hu, H.; Ochi, K. Novel Approach for Improving the Productivity of Antibiotic-Producing Strains by Inducing Combined Resistant Mutations. *Appl. Environ. Microbiol.* **2001**, *67* (4), 1885–1892.

(32) Shima, J.; Hesketh, A.; Okamoto, S.; Kawamoto, S.; Ochi, K. Induction of Actinorhodin Production by RpsL (Encoding Ribosomal Protein S12) Mutations That Confer Streptomycin Resistance in Streptomyces Lividans and Streptomyces Coelicolor A3(2). *J. Bacteriol.* **1996**, *178* (24), 7276–7284.

(33) Hosoya, Y.; Okamoto, S.; Muramatsu, H.; Ochi, K. Acquisition of Certain Streptomycin-Resistant (Str) Mutations Enhances Antibiotic Production in Bacteria. *Antimicrob. Agents Chemother.* **1998**, *42* (8), 2041–2047.

(34) Hu, H.; Ochi, K. Novel Approach for Improving the Productivity of Antibiotic-Producing Strains by Inducing Combined Resistant Mutations. *Appl. Environ. Microbiol.* **2001**, *67* (4), 1885–1892.

(35) Okamoto-Hosoya, Y.; Xu, J.; Yao, X.; Tozawa, Y.; Lai, C.; Ochi, K. Genetic and Physiological Characterization of RpoB Mutations That Activate Antibiotic Production in Streptomyces Lividans. *Microbiology* **2002**, *148* (11), 3365–3373.

(36) Hu, H.; Zhang, Q.; Ochi, K. Activation of Antibiotic Biosynthesis by Specified Mutations in the RpoB Gene (Encoding the RNA Polymerase Beta Subunit) of Streptomyces Lividans. *J. Bacteriol.* **2002**, *184* (14), 3984–3991.

(37) Tanaka, Y.; Kasahara, K.; Hirose, Y.; Murakami, K.; Kugimiya, R.; Ochi, K. Activation and Products of the Cryptic Secondary Metabolite Biosynthetic Gene Clusters by Rifampin Resistance (RpoB) Mutations in Actinomycetes. *J. Bacteriol.* **2013**, *195* (13), 2959–2970.

(38) Gibson, D. G. Synthesis of DNA Fragments in Yeast by One-Step Assembly of Overlapping Oligonucleotides. *Nucleic Acids Res.* **2009**, *37* (20), 6984–6990.

(39) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6* (5), 343–345.

(40) Kim, J. H.; Feng, Z.; Bauer, J. D.; Kallifidas, D.; Calle, P. Y.; Brady, S. F. Cloning Large Natural Product Gene Clusters from the Environment: Piecing Environmental DNA Gene Clusters Back Together with TAR. *Biopolymers* **2010**, *93* (9), 833–844.

(41) Tang, X.; Li, J.; Millán-Aguiñaga, N.; Zhang, J. J.; O'Neill, E. C.; Ugalde, J. A.; Jensen, P. R.; Mantovani, S. M.; Moore, B. S. Identification of Thiotetronic Acid Antibiotic

Biosynthetic Pathways by Target-Directed Genome Mining. *ACS Chem. Biol.* **2015**, *10* (12), 2841–2849.

(42) Smanski, M. J.; Zhou, H.; Claesen, J.; Shen, B.; Fischbach, M. A.; Voigt, C. A. Synthetic Biology to Access and Expand Nature's Chemical Diversity. *Nat. Rev. Microbiol.* **2016**, *14* (3), 135–149.

(43) Gomez-Escribano, J. P.; Bibb, M. J. Engineering Streptomyces coelicolor for Heterologous Expression of Secondary Metabolite Gene Clusters. *Microb. Biotechnol.* **2011**, *4* (2), 207–215.

(44) Gust, B.; Chandra, G.; Jakimowicz, D.; Yuqing, T.; Bruton, C. J.; Chater, K. F. Red-Mediated Genetic Manipulation of Antibiotic-Producing Streptomyces. Adv Appl Microbiol. **2004**, *54*, 107-128.

(45) Shao, Z.; Rao, G.; Li, C.; Abil, Z.; Luo, Y.; Zhao, H. Refactoring the Silent Spectinabilin Gene Cluster Using a Plug-and-Play Scaffold. *ACS Synth. Biol.* **2013**, *2* (11), 662–669.

(46) Yamanaka, K.; Reynolds, K. a; Kersten, R. D.; Ryan, K. S.; Gonzalez, D. J.; Nizet, V.; Dorrestein, P. C.; Moore, B. S. Direct Cloning and Refactoring of a Silent Lipopeptide Biosynthetic Gene Cluster Yields the Antibiotic Taromycin A. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (5), 1957–1962.

(47) Eyles, T. H.; Vior, N. M.; Truman, A. W. Rapid and Robust Yeast-Mediated Pathway Refactoring Generates Multiple New Bottromycin-Related Metabolites. *ACS Synth. Biol.* **2018**, *7* (5), 1211–1218.

(48) Roy, S.; Garges, S.; Adhya, S. Activation and Repression of Transcription by Differential Contact: Two Sides of a Coin. *J. Biol. Chem.* **1998**, *273* (23), 14059–14062.

(49) Lee, D. J.; Minchin, S. D.; Busby, S. J. W. Activating Transcription in Bacteria. *Annu. Rev. Microbiol* **2012**, *66*, 125–152.

(50) Chiang, Y.-M.; Szewczyk, E.; Davidson, A. D.; Keller, N.; Oakley, B. R.; Wang, C. C. C. A Gene Cluster Containing Two Fungal Polyketide Synthases Encodes the Biosynthetic Pathway for a Polyketide, Asperfuranone, in *Aspergillus Nidulans*. *J. Am. Chem. Soc.* **2009**, *131* (8), 2965–2970.

(51) Laureti, L.; Song, L.; Huang, S.; Corre, C.; Leblond, P.; Challis, G. L.; Aigle, B. 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.* **2011**, *108* (15), 6258–6263.

(52) O'Rourke, S.; Wietzorrek, A.; Fowler, K.; Corre, C.; Challis, G. L.; Chater, K. F. Extracellular Signalling, Translational Control, Two Repressors and an Activator All Contribute to the Regulation of Methylenomycin Production in Streptomyces coelicolor. *Mol. Microbiol.* **2009**, *71* (3), 763–778.

(53) Gottelt, M.; Kol, S.; Gomez-Escribano, J. P.; Bibb, M.; Takano, E. Deletion of a

Regulatory Gene within the Cpk Gene Cluster Reveals Novel Antibacterial Activity in Streptomyces coelicolor A3(2). *Microbiology* **2010**, *156* (8), 2343–2353.

(54) Bunet, R.; Song, L.; Mendes, M. V.; Corre, C.; Hotel, L.; Rouhier, N.; Framboisier, X.; Leblond, P.; Challis, G. L.; Aigle, B. Characterization and Manipulation of the Pathway-Specific Late Regulator AlpW Reveals Streptomyces ambofaciens as a New Producer of Kinamycins. *J. Bacteriol.* **2011**, *193* (5), 1142–1153.

(55) Sidda, J. D.; Song, L.; Poon, V.; Al-Bassam, M.; Lazos, O.; Buttner, M. J.; Challis, G. L.; Corre, C. Discovery of a Family of γ-Aminobutyrate Ureas via Rational Derepression of a Silent Bacterial Gene Cluster. *Chem. Sci.* **2014**, *5* (1), 86–89.

(56) Romero-Rodríguez, A.; Robledo-Casados, I.; Sánchez, S. An Overview on Transcriptional Regulators in Streptomyces. *Biochim. Biophys. Acta - Gene Regul. Mech.* **2015**, *1849* (8), 1017–1039.

(57) Richet, E.; Raibaud, O. MalT, the Regulatory Protein of the Escherichia Coli Maltose System, Is an ATP-Dependent Transcriptional Activator. *EMBO J.* **1989**, *8* (3), 981–987.

(58) Thanapipatsiri, A.; Gomez-Escribano, J. P.; Song, L.; Bibb, M. J.; Al-Bassam, M.; Chandra, G.; Thamchaipenet, A.; Challis, G. L.; Bibb, M. J. Discovery of Unusual Biaryl Polyketides by Activation of a Silent Streptomyces venezuelae Biosynthetic Gene Cluster. *ChemBioChem* **2016**, *17* (22), 2189–2198.

(59) Mo, S.; Yoon, Y. J. Interspecies Complementation of the LuxR Family Pathway-Specific Regulator Involved in Macrolide Biosynthesis. *J. Microbiol. Biotechnol.* **2016**, *26* (1), 66–71.

(60) Martínez-Hackert, E.; Stock, A. M. The DNA-Binding Domain of OmpR: Crystal Structures of a Winged Helix Transcription Factor. *Structure* **1997**, *5* (1), 109–124.

(61) Wietzorrek, A.; Bibb, M. A Novel Family of Proteins That Regulates Antibiotic Production in Streptomycetes Appears to Contain an OmpR-like DNA-Binding Fold. *Mol. Microbiol.* **1997**, *25* (6), 1181–1184.

(62) Arias, P.; Fernández-Moreno, M. A.; Malpartida, F. Characterization of the Pathway-Specific Positive Transcriptional Regulator for Actinorhodin Biosynthesis in Streptomyces coelicolor A3(2) as a DNA-Binding Protein. *J. Bacteriol.* **1999**, *181* (22), 6958–6968.

(63) Tanaka, A.; Takano, Y.; Ohnishi, Y.; Horinouchi, S. AfsR Recruits RNA Polymerase to the AfsS Promoter: A Model for Transcriptional Activation by SARPs. *J. Mol. Biol.* **2007**, *369* (2), 322–333.

(64) Gramajo, H. C.; Takano, E.; Bibb, M. J. Stationary-Phase Production of the Antibiotic Actinorhodin in Streptomyces coelicolor A3(2) Is Transcriptionally Regulated. *Mol. Microbiol.* **1993**, *7* (6), 837–845.

(65) Stutzman-Engwall, K. J.; Otten, S. L.; Hutchinson, C. R. Regulation of Secondary

Metabolism in Streptomyces Spp. and Overproduction of Daunorubicin in Streptomyces peucetius. *J. Bacteriol.* **1992**, *174* (1), 144–154.

(66) Suzuki, T.; Mochizuki, S.; Yamamoto, S.; Arakawa, K.; Kinashi, H. Regulation of Lankamycin Biosynthesis in Streptomyces rochei by Two SARP Genes, *SrrY* and *SrrZ*. *Biosci. Biotechnol. Biochem.* **2010**, *74* (4), 819–827.

(67) Ye, S.; Molloy, B.; Braña, A. F.; Zabala, D.; Olano, C.; Cortés, J.; Morís, F.; Salas, J. A.; Méndez, C. Identification by Genome Mining of a Type I Polyketide Gene Cluster from Streptomyces Argillaceus Involved in the Biosynthesis of Pyridine and Piperidine Alkaloids Argimycins P. *Front. Microbiol.* **2017**, *8*, 194.

(68) Enguita, F. J.; Coque, J. J.; Liras, P.; Martin, J. F. The Nine Genes of the Nocardia Lactamdurans Cephamycin Cluster Are Transcribed into Large MRNAs from Three Promoters, Two of Them Located in a Bidirectional Promoter Region. *J. Bacteriol.* **1998**, *180* (20), 5489–5494.

(69) Mast, Y.; Weber, T.; Gölz, M.; Ort-Winklbauer, R.; Gondran, A.; Wohlleben, W.; Schinko, E. Characterization of the "pristinamycin Supercluster" of Streptomyces pristinaespiralis. *Microb. Biotechnol.* **2011**, *4* (2), 192–206.

(70) Mast, Y.; Guezguez, J.; Handel, F.; Schinko, E. A Complex Signaling Cascade Governs Pristinamycin Biosynthesis in Streptomyces pristinaespiralis. *Appl. Environ. Microbiol.* **2015**, *81* (19), 6621–6636.

(71) Garg, R. P.; Parry, R. J. Regulation of Valanimycin Biosynthesis in Streptomyces viridifaciens: Characterization of VlmI as a Streptomyces Antibiotic Regulatory Protein (SARP). *Microbiology* **2010**, *156* (2), 472–483.

(72) Kurniawan, Y. N.; Kitani, S.; Maeda, A.; Nihira, T. Differential Contributions of Two SARP Family Regulatory Genes to Indigoidine Biosynthesis in Streptomyces lavendulae FRI-5. *Appl. Microbiol. Biotechnol.* **2014**, *98* (23), 9713–9721.

(73) Ma, D.; Wang, C.; Chen, H.; Wen, J. Manipulating the Expression of SARP Family Regulator BulZ and Its Target Gene Product to Increase Tacrolimus Production. *Appl. Microbiol. Biotechnol.* **2018**, *102* (11), 4887–4900.

(74) Floriano, B.; Bibb, M. AfsR Is a Pleiotropic but Conditionally Required Regulatory Gene for Antibiotic Production in Streptomyces Coelicolor A3(2). *Mol. Microbiol.* **1996**, *21* (2), 385–396.

(75) Lee, P.-C.; Umeyama, T.; Horinouchi, S. AfsS Is a Target of AfsR, a Transcriptional Factor with ATPase Activity That Globally Controls Secondary Metabolism in Streptomyces coelicolor A3(2). *Mol. Microbiol.* **2002**, *43* (6), 1413–1430.

(76) Santos-Beneit, F.; Rodríguez-García, A.; Martín, J. F. Overlapping Binding of PhoP and AfsR to the Promoter Region of GlnR in Streptomyces coelicolor. *Microbiol. Res.* **2012**, *167* (9), 532–535.

(77) Bate, N.; Stratigopoulos, G.; Cundliffe, E. Differential Roles of Two SARP-Encoding Regulatory Genes during Tylosin Biosynthesis. *Mol. Microbiol.* **2002**, *43* (2), 449–458.

(78) Sakai, T.; Sameshima, T.; Matsufuji, M.; Kawamura, N.; Dobashi, K.; Mizui, Y. Pladienolides, New Substances from Culture of Streptomyces platensis Mer-11107. I. Taxonomy, Fermentation, Isolation and Screening. *J. Antibiot. (Tokyo).* **2004**, *57* (3), 173–179.

(79) Sakai, T.; Asai, N.; Okuda, A.; Kawamura, N.; Mizui, Y. Pladienolides, New Substances from Culture of Streptomyces platensis Mer-11107. II. Physico-Chemical Properties and Structure Elucidation. *J. Antibiot. (Tokyo).* **2004**, *57* (3), 180–187.

(80) Mizui, Y.; Sakai, T.; Iwata, M.; Uenaka, T.; Okamoto, K.; Shimizu, H.; Yamori, T.; Yoshimatsu, K.; Asada, M. Pladienolides, New Substances from Culture of Streptomyces platensis Mer-11107. III. In Vitro and in Vivo Antitumor Activities. *J. Antibiot. (Tokyo).* **2004**, *57* (3), 188–196.

(81) Kotake, Y.; Sagane, K.; Owa, T.; Mimori-Kiyosue, Y.; Shimizu, H.; Uesugi, M.; Ishihama, Y.; Iwata, M.; Mizui, Y. Splicing Factor SF3b as a Target of the Antitumor Natural Product Pladienolide. *Nat. Chem. Biol.* **2007**, *3* (9), 570–575.

(82) Yokoi, A.; Kotake, Y.; Takahashi, K.; Kadowaki, T.; Matsumoto, Y.; Minoshima, Y.; Sugi, N. H.; Sagane, K.; Hamaguchi, M.; Iwata, M.; et al. Biological Validation That SF3b Is a Target of the Antitumor Macrolide Pladienolide. *FEBS J.* **2011**, *278* (24), 4870–4880.

(83) Aouida, M.; Eid, A.; Mahfouz, M. M. CRISPR/Cas9-Mediated Target Validation of the Splicing Inhibitor Pladienolide B. *Biochim. Open* **2016**, *3*, 72–75.

(84) Yoshida, K.; Sanada, M.; Shiraishi, Y.; Nowak, D.; Nagata, Y.; Yamamoto, R.; Sato, Y.; Sato-Otsubo, A.; Kon, A.; Nagasaki, M.; et al. Frequent Pathway Mutations of Splicing Machinery in Myelodysplasia. *Nature* **2011**, *478* (7367), 64–69.

(85) Wang, L.; Lawrence, M. S.; Wan, Y.; Stojanov, P.; Sougnez, C.; Stevenson, K.; Werner, L.; Sivachenko, A.; DeLuca, D. S.; Zhang, L.; et al. *SF3B1* and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia. *N. Engl. J. Med.* **2011**, *365* (26), 2497–2506.

(86) Harbour, J. W.; Roberson, E. D. O.; Anbunathan, H.; Onken, M. D.; Worley, L. A.; Bowcock, A. M. Recurrent Mutations at Codon 625 of the Splicing Factor SF3B1 in Uveal Melanoma. *Nat. Genet.* **2013**, *45* (2), 133–135.

(87) Iwata, M.; Ozawa, Y.; Uenaka, T.; Shimizu, H.; Niijima, J.; Kanada, R. M.; Fukuda, Y.; Nagai, M.; Kotake, Y.; Yoshida, M.; et al. E7107, a New 7-Urethane Derivative of Pladienolide D, Displays Curative Effect against Several Human Tumor Xenografts. *Cancer Res.* **2004**, *64* (7 Supplement).

(88) Uenaka, T.; Iwata, M.; Ozawa, Y.; Shimizu, H.; Kotake, Y.; Mizui, Y.; Yoshimatsu, K.; Asada, M. E7107, a New 7-Urethane Derivative of Pladienolide D: Correlation of the Profile of Cell Cycle Regulatory Molecules with Tumor Cells' Response. *Cancer Res.*

2004, *64* (7 Supplement).

(89) Folco, E. G.; Coil, K. E.; Reed, R. The Anti-Tumor Drug E7107 Reveals an Essential Role for SF3b in Remodeling U2 SnRNP to Expose the Branch Point-Binding Region. *Genes Dev.* **2011**, *25* (5), 440–444.

(90) Eskens, F. A. L. M.; Ramos, F. J.; Burger, H.; O'Brien, J. P.; Piera, A.; de Jonge, M. J. A.; Mizui, Y.; Wiemer, E. A. C.; Carreras, M. J.; Baselga, J.; et al. Phase I Pharmacokinetic and Pharmacodynamic Study of the First-in-Class Spliceosome Inhibitor E7107 in Patients with Advanced Solid Tumors. *Clin. Cancer Res.* **2013**, *19* (22), 6296–6304.

(91) Hong, D. S.; Kurzrock, R.; Naing, A.; Wheler, J. J.; Falchook, G. S.; Schiffman, J. S.; Faulkner, N.; Pilat, M. J.; O'Brien, J.; LoRusso, P. A Phase I, Open-Label, Single-Arm, Dose-Escalation Study of E7107, a Precursor Messenger Ribonucleic Acid (Pre-MRNA) Splicesome Inhibitor Administered Intravenously on Days 1 and 8 Every 21 Days to Patients with Solid Tumors. *Invest. New Drugs* **2014**, *32* (3), 436–444.

(92) Seiler, M.; Yoshimi, A.; Darman, R.; Chan, B.; Keaney, G.; Thomas, M.; Agrawal, A. A.; Caleb, B.; Csibi, A.; Sean, E.; et al. H3B-8800, an Orally Available Small-Molecule Splicing Modulator, Induces Lethality in Spliceosome-Mutant Cancers. *Nat. Med.* **2018**, *24* (4), 497–504.

(93) NIH. No Title.

(94) Ghosh, A. K.; Anderson, D. D. Enantioselective Total Synthesis of Pladienolide B: A Potent Spliceosome Inhibitor. *Org. Lett.* **2012**, *14* (18), 4730–4733.

(95) Kanada, R. M.; Itoh, D.; Nagai, M.; Niijima, J.; Asai, N.; Mizui, Y.; Abe, S.; Kotake, Y. Total Synthesis of the Potent Antitumor Macrolides Pladienolide B and D. *Angew. Chemie Int. Ed.* **2007**, *46* (23), 4350–4355.

(96) Sun, Y.; Zhou, X.; Dong, H.; Tu, G.; Wang, M.; Wang, B.; Deng, Z. A Complete Gene Cluster from Streptomyces Nanchangensis NS3226 Encoding Biosynthesis of the Polyether Ionophore Nanchangmycin. *Chem. Biol.* **2003**, *10* (5), 431–441.

(97) Machida, K.; Aritoku, Y.; Nakashima, T.; Arisawa, A.; Tsuchida, T. Increase in Pladienolide D Production Rate Using a Streptomyces Strain Overexpressing a Cytochrome P450 Gene. *J. Biosci. Bioeng.* **2008**, *105* (6), 649–654.

(98) Onwueme, K. C.; Ferreras, J. A.; Buglino, J.; Lima, C. D.; Quadri, L. E. N. Mycobacterial Polyketide-Associated Proteins Are Acyltransferases: Proof of Principle with Mycobacterium Tuberculosis PapA5. *Proc. Natl. Acad. Sci.* **2004**, *101* (13), 4608–4613.

(99) Yu, W.-L.; Jones, B. D.; Kang, M.; Hammons, J. C.; La Clair, J. J.; Burkart, M. D. Spirohexenolide A Targets Human Macrophage Migration Inhibitory Factor (HMIF). *J. Nat. Prod.* **2013**, *76* (5), 817–823.

(100) Takahashi, N.; Nishihira, J.; Sato, Y.; Kondo, M.; Ogawa, H.; Ohshima, T.; Une, Y.; Todo, S. Involvement of Macrophage Migration Inhibitory Factor (MIF) in the Mechanism of

Tumor Cell Growth. *Mol. Med.* **1998**, *4* (11), 707–714.

(101) Chesney, J.; Metz, C.; Bacher, M.; Peng, T.; Meinhardt, A.; Bucala, R. An Essential Role for Macrophage Migration Inhibitory Factor (MIF) in Angiogenesis and the Growth of a Murine Lymphoma. *Mol. Med.* **1999**, *5* (3), 181–191.

(102) Ogawa, H.; Nishihira, J.; Sato, Y.; Kondo, M.; Takahashi, N.; Oshima, T.; Todo, S. AN Antibody For Macrophage Migration Inhibitory Factor Suppresses Tumour Growth and Inhibits Tumour-Associated Angiogenesis. *Cytokine* **2000**, *12* (4), 309–314.

(103) Sun, B.; Nishihira, J.; Yoshiki, T.; Kondo, M.; Sato, Y.; Sasaki, F.; Todo, S. Macrophage Migration Inhibitory Factor Promotes Tumor Invasion and Metastasis via the Rho-Dependent Pathway. *Clin. Cancer Res.* **2005**, *11* (3), 1050–1058.

(104) He, X.-X.; Chen, K.; Yang, J.; Li, X.-Y.; Gan, H.-Y.; Liu, C.-Y.; Coleman, T. R.; Al-Abed, Y. Macrophage Migration Inhibitory Factor Promotes Colorectal Cancer. *Mol. Med.* **2009**, *15* (1–2), 1–10.

(105) Bister, B.; Bischoff, D.; Ströbele, M.; Riedlinger, J.; Reicke, A.; Wolter, F.; Bull, A. T.; Zähner, H.; Fiedler, H.-P.; Süssmuth, R. D. Abyssomicin C—A Polycyclic Antibiotic from a MarineVerrucosispora Strain as an Inhibitor of Thep-Aminobenzoic Acid/Tetrahydrofolate Biosynthesis Pathway. *Angew. Chemie Int. Ed.* **2004**, *43* (19), 2574–2576.

(106) Niu, X.-M.; Li, S.-H.; Görls, H.; Schollmeyer, D.; Hilliger, M.; Grabley, S.; Sattler, I. Abyssomicin E, a Highly Functionalized Polycyclic Metabolite from Streptomyces Species. *Org. Lett.* **2007**, *9* (13), 2437–2440.

(107) Keller, S.; Nicholson, G.; Drahl, C.; Sorensen, E.; Fiedler, H.-P.; Süssmuth, R. D. Abyssomicins G and H and Atrop-Abyssomicin C from the Marine Verrucosispora Strain AB-18-032†. *J. Antibiot. (Tokyo).* **2007**, *60* (6), 391–394.

(108) Wang, Q.; Song, F.; Xiao, X.; Huang, P.; Li, L.; Monte, A.; Abdel-Mageed, W. M.; Wang, J.; Guo, H.; He, W.; et al. Abyssomicins from the South China Sea Deep-Sea Sediment *Verrucosispora* Sp.: Natural Thioether Michael Addition Adducts as Antitubercular Prodrugs. *Angew. Chemie Int. Ed.* **2013**, *52* (4), 1231–1234.

(109) Tu, J.; Li, S.; Chen, J.; Song, Y.; Fu, S.; Ju, J.; Li, Q. Characterization and Heterologous Expression of the Neoabyssomicin/Abyssomicin Biosynthetic Gene Cluster from Streptomyces koyangensis SCSIO 5802. *Microb. Cell Fact.* **2018**, *17* (1), 28.

(110) Imai, H.; Kaniwa, H.; Tokunaga, T.; Fujita, S.; Furuya, T.; Matsumoto, H.; Shimizu, M. Okilactomycin, a Novel Antibiotic Produced by a Streptomyces Species. II. Structure Determination. *J. Antibiot. (Tokyo).* **1987**, *40* (11), 1483–1489.

(111) Zhang, C.; Ondeyka, J. G.; Zink, D. L.; Basilio, A.; Vicente, F.; Salazar, O.; Genilloud, O.; Dorso, K.; Motyl, M.; Byrne, K.; et al. Discovery of Okilactomycin and Congeners from Streptomyces Scabrisporus by Antisense Differential Sensitivity Assay Targeting

Ribosomal Protein S4. *J. Antibiot. (Tokyo).* **2009**, *62* (2), 55–61.

(112) Nakai, R.; Kakita, S.; Asai, A.; Chiba, S.; Akinaga, S.; Mizukami, T.; Yamashita, Y. Chrolactomycin, a Novel Antitumor Antibiotic Produced by Streptomyces sp. *J. Antibiot. (Tokyo).* **2001**, *54* (10), 836–839.

(113) Iorio, M.; Maffioli, S. I.; Gaspari, E.; Rossi, R.; Mauri, P.; Sosio, M.; Donadio, S. Chrolactomycins from the Actinomycete Actinospica. *J. Nat. Prod.* **2012**, *75* (11), 1991–1993.

(114) Lacoske, M. H.; Theodorakis, E. A. Spirotetronate Polyketides as Leads in Drug Discovery. *J. Nat. Prod.* **2015**, *78* (3), 562–575.

(115) Jia, X.-Y.; Tian, Z.-H.; Shao, L.; Qu, X.-D.; Zhao, Q.-F.; Tang, J.; Tang, G.-L.; Liu, W. Genetic Characterization of the Chlorothricin Gene Cluster as a Model for Spirotetronate Antibiotic Biosynthesis. *Chem. Biol.* **2006**, *13* (6), 575–585.

(116) Hua Zhang; Jess A. White-Phillip; Charles E. Melançon, I.; Hyung-jin Kwon; Wei-luen Yu, and; Liu, H. Elucidation of the Kijanimicin Gene Cluster: Insights into the Biosynthesis of Spirotetronate Antibiotics and Nitrosugars. *J Am Chem Soc.* **2007,** *129* (47), 14670-14683.

(117) Demydchuk, Y.; Sun, Y.; Hong, H.; Staunton, J.; Spencer, J. B.; Leadlay, P. F. Analysis of the Tetronomycin Gene Cluster: Insights into the Biosynthesis of a Polyether Tetronate Antibiotic. *ChemBioChem* **2008**, *9* (7), 1136–1145.

(118) Kanchanabanca, C.; Tao, W.; Hong, H.; Liu, Y.; Hahn, F.; Samborskyy, M.; Deng, Z.; Sun, Y.; Leadlay, P. F. Unusual Acetylation-Elimination in the Formation of Tetronate Antibiotics. *Angew. Chemie Int. Ed.* **2013**, *52* (22), 5785–5788.

(119) Fang, J.; Zhang, Y.; Huang, L.; Jia, X.; Zhang, Q.; Zhang, X.; Tang, G.; Liu, W. Cloning and Characterization of the Tetrocarcin A Gene Cluster from Micromonospora Chalcea NRRL 11289 Reveals a Highly Conserved Strategy for Tetronate Biosynthesis in Spirotetronate Antibiotics. *J. Bacteriol.* **2008**, *190* (17), 6014–6025.

(120) Sun, Y.; Hong, H.; Gillies, F.; Spencer, J. B.; Leadlay, P. F. Glyceryl-S-Acyl Carrier Protein as an Intermediate in the Biosynthesis of Tetronate Antibiotics. *ChemBioChem* **2008**, *9* (1), 150–156.

(121) Wu, L.-F.; He, H.-Y.; Pan, H.-X.; Han, L.; Wang, R.; Tang, G.-L. Characterization of QmnD3/QmnD4 for Double Bond Formation in Quartromicin Biosynthesis. *Org. Lett.* **2014**, *16* (6), 1578–1581.

(122) Gottardi, E. M.; Krawczyk, J. M.; von Suchodoletz, H.; Schadt, S.; Mühlenweg, A.; Uguru, G. C.; Pelzer, S.; Fiedler, H.-P.; Bibb, M. J.; Stach, J. E. M.; et al. Abyssomicin Biosynthesis: Formation of an Unusual Polyketide, Antibiotic-Feeding Studies and Genetic Analysis. *Chembiochem* **2011**, *12* (9), 1401–1410.

(123) Gui, C.; Zhang, S.; Zhu, X.; Ding, W.; Huang, H.; Gu, Y.-C.; Duan, Y.; Ju, J.

Antimicrobial Spirotetronate Metabolites from Marine-Derived *Micromonospora Harpali* SCSIO GJ089. *J. Nat. Prod.* **2017**, *80* (5), 1594–1603.

(124) Byrne, M. J.; Lees, N. R.; Han, L.-C.; van der Kamp, M. W.; Mulholland, A. J.; Stach, J. E. M.; Willis, C. L.; Race, P. R. The Catalytic Mechanism of a Natural Diels–Alderase Revealed in Molecular Detail. *J. Am. Chem. Soc.* **2016**, *138* (19), 6095–6098.

(125) Tian, Z.; Sun, P.; Yan, Y.; Wu, Z.; Zheng, Q.; Zhou, S.; Zhang, H.; Yu, F.; Jia, X.; Chen, D.; et al. An Enzymatic [4+2] Cyclization Cascade Creates the Pentacyclic Core of Pyrroindomycins. *Nat. Chem. Biol.* **2015**, *11* (4), 259–265.

(126) Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **2009**, *323* (5910), 133–138.

(127) Chin, C.-S.; Alexander, D. H.; Marks, P.; Klammer, A. A.; Drake, J.; Heiner, C.; Clum, A.; Copeland, A.; Huddleston, J.; Eichler, E. E.; et al. Nonhybrid, Finished Microbial Genome Assemblies from Long-Read SMRT Sequencing Data. *Nat. Methods* **2013**, *10* (6), 563–569.

(128) Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran, H. G.; de los Santos, E. L. C.; Kim, H. U.; Nave, M.; et al. AntiSMASH 4.0—improvements in Chemistry Prediction and Gene Cluster Boundary Identification. *Nucleic Acids Res.* **2017**, *45* (W1), W36–W41.

(129) 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.* **2015**, *10* (6), 845–858.

(130) Wang, R.; Yin, Y.-J.; Wang, F.; Li, M.; Feng, J.; Zhang, H.-M.; Zhang, J.-P.; Liu, S.-J.; Chang, W.-R. Crystal Structures and Site-Directed Mutagenesis of a Mycothiol-Dependent Enzyme Reveal a Novel Folding and Molecular Basis for Mycothiol-Mediated Maleylpyruvate Isomerization. *J. Biol. Chem.* **2007**, *282* (22), 16288–15294.

(131) Rajan, S. S.; Yang, X.; Shuvalova, L.; Collart, F.; Anderson, W. F. YfiT from Bacillus Subtilis Is a Probable Metal-Dependent Hydrolase with an Unusual Four-Helix Bundle Topology. *Biochemistry* **2004**, *43* (49), 15472–15479.

(132) Newton, G. L.; Buchmeier, N.; Fahey, R. C. Biosynthesis and Functions of Mycothiol, the Unique Protective Thiol of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **2008**, *72* (3), 471–494.

(133) Ahuja, E. G.; Janning, P.; Mentel, M.; Graebsch, A.; Breinbauer, R.; Hiller, W.; Costisella, B.; Thomashow, L. S.; Mavrodi, D. V.; Blankenfeldt, W. PhzA/B Catalyzes the Formation of the Tricycle in Phenazine Biosynthesis. *J. Am. Chem. Soc.* **2008**, *130* (50), 17053–17061.

(134) Hashimoto, T.; Hashimoto, J.; Teruya, K.; Hirano, T.; Shin-ya, K.; Ikeda, H.; Liu, H.;

Nishiyama, M.; Kuzuyama, T. Biosynthesis of Versipelostatin: Identification of an Enzyme-Catalyzed [4+2]-Cycloaddition Required for Macrocyclization of Spirotetronate-Containing Polyketides. *J. Am. Chem. Soc.* **2015**, *137* (2), 572–575.

Chapter Six: General Discussion

Contents

6. 1. The Generation of Diversity by Biosynthetic Assembly Lines

Actinobacteria have a wealth of specialised metabolism and are capable of producing structurally diverse molecules with a variety of biological activities. This thesis has explored a subset of this diversity focusing on the modular type-I polyketide synthase (T1PKS) and the non-ribosomal peptide synthetase (NRPS) biosynthetic assembly-lines. In broad terms, diversity can be introduced to assembly-line products during three stages: i) the supply of precursors, ii) assembly-line biosynthesis and, iii) post-assembly line tailoring reactions. Therefore, it can be helpful to consider the evolution of BGCs in terms of the interactions between functional sub-clusters consisting of one or more genes. The concept of 'sub-cluster sharing' was demonstrated by Medema et al., through a systematic analysis of BGCs[1]. They showed that functional sub-clusters were present among distantly related natural products. As an illustrative example they point to rubradirin, the biosynthesis of which is composed almost entirely of conserved sub-clusters that can be found in other biosynthetic pathways (Figure 6. 1). This concept has profound implications on genome mining and understanding the evolution of natural product BGCs.

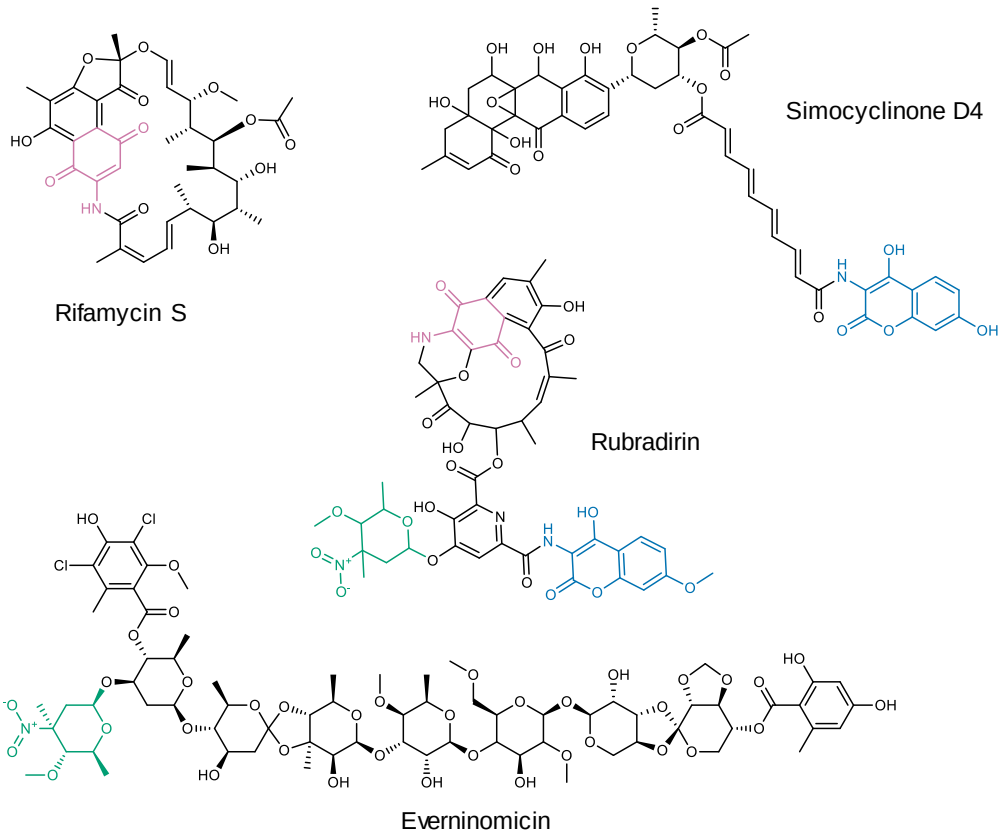


Figure 6. 57: Conserved sub-structures in the biosynthesis of rubradirin. Coloured structural features correspond to conserved biosynthetic sub-clusters sharing an evolutionary origin. Rubradirin is composed of several sub-clusters from disperate loci (purple: rifamycin; blue simocyclinone and; green: everninomicin).

6. 2. Genome Mining

In Chapter Four, conserved genes from the vicenistatin (*vin*) and heronamide C (**7**)(*hrn*) BGCs were used to mine publicly available genomes for β-amino acid containing polyketide macrolactams (βPMs). This method of genome mining also serves to provide a comprehensive survey of the diversity of the βPM family. This approach identified 40 putative βPM-encoding BGCs that have yet to be characterised experimentally, the vast majority of which are likely to produce as of yet uncharacterised products. In addition, by using a phylogenetic approach we were able to prioritise clades with no characterised members (subclades 3, 4 and 6, Figure 4.11). Although the structures have yet to be elucidated, we have identified target compounds from *Nocardiopsis gilva* DSM 44841 (**9**-**12**) that are likely to be novel βPMs. Furthermore, by considering only the genes responsible for the incorporation of the β-amino acid, the results were decoupled from other biosynthetic steps, such as polyketide biosynthesis, facilitating the discovery of a new putative family of BGCs that shared homologues to the AMP-binding proteins VinN and VinM but were unlikely to be involved in the biosynthesis of polyketides. Despite their conservation and apparent involvement in specialised metabolism, these BGCs were inconsistently recognised as BGCs by antiSMASH[2] demonstrating the value of sub-cluster mining as method to identify 'unknown unknowns.' Going forward, the identification of the BGC products from this new family would validate the usefulness of this approach. This can hopefully be achieved by a combination of BGC disruption and metabolic screening.

Within the BGCs explored in this thesis, there are several candidate sub-clusters to apply method of genome mining. The genes involved in tetronate and spirotetronate formation are well characterised and highly conserved making them promising targets for this approach. In addition, the putative pyran forming cassette *spxIO2J* could be used to discover more examples utilising this biosynthesis that, as of yet, has only observed in the spirohexenolides. A more general approach could be to mine for sub-clusters involved in the biosynthesis of modified sugars. For example, amino sugars are incorporated into a wide range of natural products[3,4]. The βPM Sipanmycin was discovered through a similar strategy[3]. 71 *Streptomyces* strains were screened *via* PCR for the presence of NDP-D-glucose synthase and NDP-D-glucose 4,6-dehydratase. Metabolic screening of positive strains led to the discovery of lobophorins, vicenistatin, chromomycins, benzanthrins and the novel macrolactams sipanmycin A and B (Figure 6. 2). The diversity of natural products identified by this screen demonstrate the utility of sugar biosynthesis as a mining target. *In silico* analysis of the public databases would therefore provide an opportunity to upscale this process while providing data about the broad distribution of these genes on the generic level or higher. Indeed, any genes involved in the biosynthesis of biosynthetic precursors can provide targets for this strategy. For example, sub-clusters involved

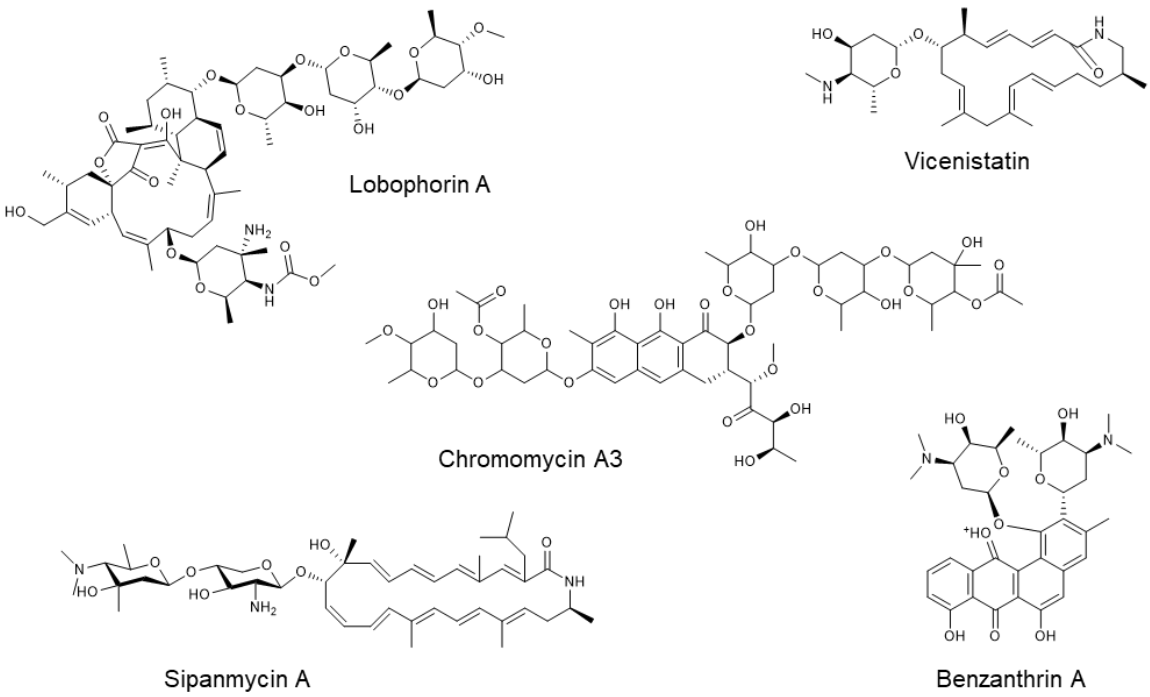


Figure 6. 58: Amino sugars in polyketide natural products. Compounds discover by Malmierca et al.[3] by targeted genome mining of NDP-D-glucose synthase and NDP-D-glucose 4,6-dehydratase in Streptomyces spp..

in the biosynthesis of non-canonical amino acids such as L-ornithine or L-*allo*-isoleucine sub-clusters as observed in the wollamide (**3** and **4**) *wol* BGC.

It may be possible to use genome mining to identify other bifurcated assembly-lines. EvoMining or phylogenomics[5,6] describes the concept of searching for gene duplications or expansion of gene families to identify BGCs. Using duplication as a tool to directly probe assembly-line proteins is problematic due to their repetitive sequences. As observed in the *wol* BGC however, it may not be the case that only the assembly line proteins are duplicated. We have proposed that the MbtH-like genes *wolF1* and *wolF2* have arisen during the same event that duplicated the NRPS genes *wolG1* and *wolG2*, therefore searching for multiple MbtH homologues within close proximity in a genome could be a viable strategy. However, given the low sequence homology among functional MbtH-like proteins[7], this approach would require considerably more optimisation than the previous examples.

6. 3. The Evolution of Biosynthetic Assembly-Lines

The discovery of the *wol* BGC is a significant piece of evidence in understanding the evolution of biosynthetic assembly-lines and expands the current model. In addition, new experimental results, particularly from the field of synthetic biology have provided further insight into these processes.

6. 3. 2. Precursor Supply

Most assembly-lines incorporate substrates that are absent from, or only produced in small quantities during primary metabolism[8,9]. Unfortunately, very little is known about the evolution of precursor biosynthesis or how biosynthetic genes are recruited to a BGC. In the biosynthesis of the desotamides, L-*allo*-isoleucine is converted from L-isoleucine by DsaD and DsaE[10]. In the *wol* BGC, these genes were also accompanied by genes involved in the biosynthesis of L-ornithine. However, our evolutionary hypothesis for the evolution of these two BGCs (Figure 2. 12) predicts that this difference in gene content is a result of gene loss. As such, the *wol* and *dsa* BGCs alone cannot provide a useful model for studying this process. However, BLAST[11] searches revealed a close relation between the NRPS encoding genes for *wol* and *dsa* and the recently sequenced BGC for the hexapeptides ulleungmycin[12]. Ulleungmycin, and the related longicatenamycin[13] contain residues derived from the non-proteinogenic amino acids L-homoleucine and L-*threo*-β-hydroxy—asparagine (Figure 6. 3)[12,13]. Further examination of this family could provide more insight into how precursors are recruited to assembly-line and importantly BGCs and how assembly-lines adapt to incorporate new monomers.

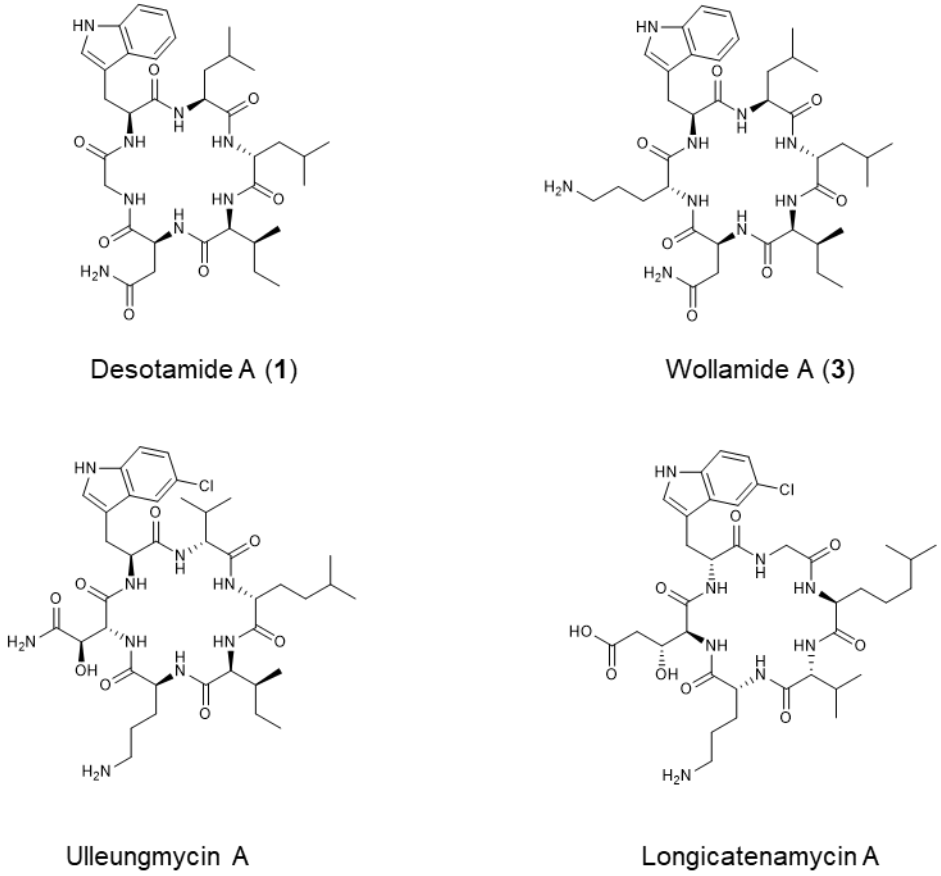


Figure 6. 59: The structures of desotamide related hexapeptides. Showing desotamide A (**1**), wollamide A (**3**), ulleungmycin A and longicatenamycin A.

6. 3. 3 Assembly-Line Expansion and Contraction

The modular nature of biosynthetic assembly-lines raises important questions about their evolutionary origin. The current model of assembly-line evolution is based on two main observations. Firstly, phylogenetic studies of biosynthetic domains show that, in the majority of cases, domains are more closely related to homologues within the same BGC than homologues from other BGCs[1,14]. This is most commonly interpreted as the result of multiple duplications of biosynthetic modules during the evolution of the BGC, however, there is also evidence that concerted evolution (the homogenisation of paralogous sequences through homologous recombination) has contributed to this phenomenon[15]. The second common observation is that coding regions responsible for substrate specificity are particularly amenable to recombination[16,17]. Hence a model of intragenic duplication and recombination is the prevailing model of assembly line biosynthesis. Recent experiments in the rapamycin BGC (*rap*) provide support for this hypothesis. Through the technique of 'accelerated evolution' Wlodek et al., were able to increase the rate of recombination within the *rap* PKS[18], generating several functional assembly-lines producing rapamycin analogues. Although the majority of functional BGCs produced analogues with reduced ring size, in one example the assembly-line was extended by an extra module. Examination of the sequence showed that this was caused by duplication of modules from within the *rap* PKS, providing strong evidence for the feasibility of such a mechanism in Nature.

The *wol* BGC provides strong evidence for an alternate mechanism involving intergenic duplication. Rather than extending the PKS by one module, intergenic duplication results in genetic redundancy. It is likely that such functional redundancy lowers the selective pressure on one of the genes, allowing refunctionalisation to occur without the potentially high fitness costs associated with perturbing the original assembly-line. The refunctionalisation of one of the modules would result in a bifurcated assembly-line as observed in the *wol* BGC. Furthermore, breakpoint analysis of recombination identified a previously unobserved relationship between the substrate binding pocket and the P-loop of adenylation domains.

6. 4. Engineering NRPS Assembly-lines

The *de novo* design and reprogramming of synthetic NRPSs is a major goal in synthetic biology, but efforts are hampered by limited understanding of their structural-functional relationships, manifesting in low yields and a lack of diversity among synthetic products. On the contrary, evolution has repeatedly produced efficient NRPSs capable of producing diverse and structurally complex products. Our analysis of the *wol* and *dsa* BGCs can begin to provide insights into how Nature has refunctionalised BGCs that may be applicable in the lab. Much of our analysis focused on the evolution of the adenylation domain, however comparison of *wolG2* and

dsaG allowed us to identify differences between residues in the condensation domain. Future work will involve probing the importance of these mutations on the activity of the condensation domains.

The wollamides (**3** and **4**) are important leads as antitubercular drugs, and as such we aim to produce a strain capable of producing **3** and **4** in high titres. One approach, described in Chapter 2, is knocking out *dsaG* in Streptomyces MST-70754 and inserting pGP9-*wolG2*. Another option is the expression of the entire assembly-line in a heterologous host. In Chapter 2, I described pBO1, an integrative *Streptomyces* expression vector capable of propagating in both *E. coli* and *S. cerevisae* designed for the purpose of assembling and overexpressing large genes, such as those encoding NRPSs. By changing the intergrase and selective marker of pBO1 it will be possible to generate a suite of plasmid for the expression of entire assembly-lines heterologously. Importantly, this will enable the fermentation of the wollamides in a clean metabolic background and provide a platform for further modification.

6. 5. Conclusions

This thesis has explored a range of different BGCs involved in the biosynthesis of actinobacterial specialised metabolites. The discovery of the *wol* BGC has expanded the current model of assembly-line evolution by providing the first evidence of intergenic duplication and recombination. Studies of the β-amino acid containing polyketide macrolactam (βPM) heronamide C (**7**) demonstrated spontaneous thermal and photochemical routes to generating structural diversity through respective [6π + 6π] and [6π + 4π] cycloadditions. Further examination of the βPM family using sub-cluster mining allowed strains to be prioritised for metabolic analysis, leading to the identification of candidate βPMs from *Nocardiopsis gilva* DSM44841. Finally, through the overexpression of LAL-family and SARP-family transcriptional activators I was able to activate the biosynthesis of the polyketide macrolide pladienolide B and spirotetronate spirohexenolide and active their expression in strains of *Streptomyces platensis*.

6. 6. Bibliography

(1) Medema, M. H.; Cimermancic, P.; Sali, A.; Takano, E.; Fischbach, M. A. A Systematic Computational Analysis of Biosynthetic Gene Cluster Evolution: Lessons for Engineering Biosynthesis. *PLoS Comput. Biol.* **2014**, *10* (12), e1004016.

(2) Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran, H. G.; de los Santos, E. L. C.; Kim, H. U.; Nave, M.; et al. AntiSMASH 4.0—improvements in Chemistry Prediction and Gene Cluster Boundary Identification. *Nucleic Acids Res.* **2017**, *45* (W1), W36–W41.

(3) Malmierca, M. G.; González-Montes, L.; Pérez-Victoria, I.; Sialer, C.; Braña, A. F.;

García Salcedo, R.; Martín, J.; Reyes, F.; Méndez, C.; Olano, C.; et al. Searching for Glycosylated Natural Products in Actinomycetes and Identification of Novel Macrolactams and Angucyclines. *Front. Microbiol.* **2018**, *9*, 39.

(4) Ogasawara, Y.; Katayama, K.; Minami, A.; Otsuka, M.; Eguchi, T.; Kakinuma, K. Cloning, Sequencing, and Functional Analysis of the Biosynthetic Gene Cluster of Macrolactam Antibiotic Vicenistatin in Streptomyces halstedii. *Chem. Biol.* **2004**, *11* (1), 79–86.

(5) Medema, M. H.; Fischbach, M. A. Computational Approaches to Natural Product Discovery. *Nat. Chem. Biol.* **2015**, *11* (9), 639–648.

(6) Cruz-Morales, P.; Kopp, J. F.; Martínez-Guerrero, C.; Yáñez-Guerra, L. A.; Selem-Mojica, N.; Ramos-Aboites, H.; Feldmann, J.; Barona-Gómez, F. Phylogenomic Analysis of Natural Products Biosynthetic Gene Clusters Allows Discovery of Arseno-Organic Metabolites in Model Streptomycetes. *Genome Biol. Evol.* **2016**, *8* (6), 1906–1916.

(7) Baltz, R. H. Function of MbtH Homologs in Nonribosomal Peptide Biosynthesis and Applications in Secondary Metabolite Discovery. *J. Ind. Microbiol. Biotechnol.* **2011**, *38* (11), 1747–1760.

(8) Walsh, C. T.; O'Brien, R. V; Khosla, C. Nonproteinogenic Amino Acid Building Blocks for Nonribosomal Peptide and Hybrid Polyketide Scaffolds. *Angew. Chem. Int. Ed. Engl.* **2013**, *52* (28), 7098–7124.

(9) Ray, L.; Moore, B. S. Recent Advances in the Biosynthesis of Unusual Polyketide Synthase Substrates. *Nat. Prod. Rep.* **2016**, *33* (2), 150–161.

(10) Li, Q.; Qin, X.; Liu, J.; Gui, C.; Wang, B.; Li, J.; Ju, J. Deciphering the Biosynthetic Origin of l - *Allo* -Isoleucine. *J. Am. Chem. Soc.* **2016**, *138* (1), 408–415.

(11) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410.

(12) Son, S.; Hong, Y.-S.; Jang, M.; Heo, K. T.; Lee, B.; Jang, J.-P.; Kim, J.-W.; Ryoo, I.-J.; Kim, W.-G.; Ko, S.-K.; et al. Genomics-Driven Discovery of Chlorinated Cyclic Hexapeptides Ulleungmycins A and B from a Streptomyces Species. *J. Nat. Prod.* **2017**, *80* (11), 3025–3031.

(13) Shiba, T.; Mukunoki, Y. The Total Structure of the Antibiotic Longicatenamycin. *J. Antibiot. (Tokyo).* **1975**, *28* (8), 561–566.

(14) Jenke-Kodama, H.; Börner, T.; Dittmann, E. Natural Biocombinatorics in the Polyketide Synthase Genes of the Actinobacterium Streptomyces Avermitilis. *PLoS Comput. Biol.* **2006**, *2* (10), e132.

(15) Liao, D. Concerted Evolution: Molecular Mechanism and Biological Implications. *Am. J. Hum. Genet.* **1999**, *64* (1), 24–30.

(16) Tanabe, Y.; Kaya, K.; Watanabe, M. M. Evidence for Recombination in the Microcystin

Synthetase (Mcy) Genes OfToxic Cyanobacteria Microcystis Spp. *J. Mol. Evol.* **2004**, *58* (6), 633–641.

(17) Shishido, T.; Kaasalainen, U.; Fewer, D. P.; Rouhiainen, L.; Jokela, J.; Wahlsten, M.; Fiore, M.; Yunes, J.; Rikkinen, J.; Sivonen, K. Convergent Evolution of [D-Leucine1] Microcystin-LR in Taxonomically Disparate Cyanobacteria. *BMC Evol. Biol.* **2013**, *13* (1), 86.

(18) Wlodek, A.; Kendrew, S. G.; Coates, N. J.; Hold, A.; Pogwizd, J.; Rudder, S.; Sheehan, L. S.; Higginbotham, S. J.; Stanley-Smith, A. E.; Warneck, T.; et al. Diversity Oriented Biosynthesis via Accelerated Evolution of Modular Gene Clusters. *Nat. Commun.* **2017**, *8* (1), 1206.

Chapter Seven: Materials and Methods

Contents

7. 1 General Methods

7. 1. 1. Reagents and Suppliers

Reagents and chemicals were purchased from: Fisher Scientific, Holland and Barret, Merck, Sigma-Aldrich and Thermo Scientific (Oxoid). All solvents were supplied by Fisher Scientific and were HPLC-grade or LCMS-grade where appropriate.

7. 1. 2. Microbial Strains

A list of bacterial strains used in this study can be found in Table 7. 17. *Streptomyces* spp. were maintained at 30°C on SF+M agar, with the exception of *Hamadaea tsunoensis* DSM44101 which was maintained on M1 agar. For the production of specialised metabolites, see the specific chapter for culture media. Strains of *E. coli* were maintained at 37°C on LB agar. S. cerevisiae CEN.PK 2-1 was maintained on YPAD agar at 30°C. Glycerol stocks were generated from a 1:1 mixture of liquid culture and 40% glycerol and stored at -80°C. *Streptomyces* spore stocks were generated using a modified version of the protocol published by Kieser et al.[1]. Spores were harvested in 5 ml sterile water and filtered through sterile cotton wool. Filtrate was centrifuged, had the supernatant removed and was concentrated in 100 µl 20% glycerol.

7. 1. 3. Culture Media

Table 7. 2 contains a list of media used in this study. Where solid agar is used, Difco-Bacto agar 10g/L added to the media prior to autoclaving. All media were made use distilled water, except for MYM-TAP, SM6 and SM18, where tap water was used. For R2 and R5 2 ml trace element solution ($ZnCl_2$ 40 mg/L, $FeCl_3 \cdot 6H_2O$ 200 mg/L, $CuCl_2 \cdot 2H_2O$ 10 mg/L, $MnCl_2 \cdot 4H_2O$ 10 mg/L, $Na_2B_4O_7 \cdot 10H_2O$ 10 mg/L, $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$ 10 mg/L) was added after autoclaving. Stock and final concentrations of antibiotics are found in Table 7. 3.

7. 1. 4. Plasmids

Table 7. 4 contains a list of plasmids used in this study.

7. 1. 5. Oligonucleotides

Oligonucleotides used in this study are listed in Table 7. 5.

Strain	Description	Reference
Actinokineospora enzanensis DSM 44649	Putative βPM producing strain	DSM
Escherichia coli ET12567 puz8002	Conjugation strain; F- dam-13::Tn9 dcm-6 hsdM hsdR zjj-202::Tn10 recF143 galK2 galT22 ara-14 lacY1 xyl-5 leuB6 thi-1 tonA31 rpsL136 hisG4 tsx-78 mtl-1 glnV44, subsiquently transformed with plasmid puZ8002	Castellani and Chalmers
Escherichia coli NiCo21(DE3)	Protein expression strain; can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5	New England Biolabs
Escherichia coli NiCo21(DE3); peT28a(+)-orf6595a	Protein expression strain expressing Orf6595A; can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-;B;int::(lacI::PlacUV5::T7 gene1) i21 Δnin5; peT28a(+)-orf6595A	This study
Escherichia coli NiCo21(DE3); peT28a(+)-orf6595a; pCDFDuet-1-wolF2	Protein expression strain expressing Orf6595A and WolF2; can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-;B;int::(lacI::PlacUV5::T7 gene1) i21 Δnin5; peT28a(+)-orf6595A; pCDFDuet-1-wolF2	This study
Escherichia coli NiCo21(DE3); peT28a(+)-wolG1A2	Protein expression strain expressing WolG2A1; can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-;B;int::(lacI::PlacUV5::T7 gene1) i21 Δnin5; peT28a(+)-wolG1A2	This study
Escherichia coli NiCo21(DE3); peT28a(+)-wolG1A2; pCDFDuet-1-wolF2	Protein expression strain expressing WolG2A1 and WolF2; can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-;B;int::(lacI::PlacUV5::T7 gene1) i21 Δnin5; peT28a(+)-wolG1A2; pCDFDuet-1-wolF2	This study

Table 7. 17: List of microbial strains used in this study (including genotype).

Strain	Description	Reference
Escherichia coli NiCo21(DE3); peT28a(+)-wolG2A2	Protein expression strain expressing WolG2A2; can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-;B;int::(lacI::PlacUV5::T7 gene1) i21 Δnin5; peT28a(+)-wolG2A2	This study
Escherichia coli NiCo21(DE3); peT28a(+)-wolG2A2; pCDFDuet-1-wolF2	Protein expression strain expressing WolG2A2 and WolF2; can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-;B;int::(lacI::PlacUV5::T7 gene1) i21 Δnin5; peT28a(+)-wolG2A2; pCDFDuet-1-wolF2	This study
Escherichia coli S17 λpir	Conjugation Strain; TpR SmR recA, thi, pro, hsdR-M+RP4: 2-Tc:Mu: Km Tn7 λpir	Simon et al., 1983
Escherichia coli DH5α	Assembly strain; MATa; his3D1; leu2-3_112; ura3-52; trp1-289; MAL2-8c; SUC2	Hanahan, 1985
Hamadaea tsunoensis DSM 44101	Putative βPM producing strain	DSM
Kutzneria albida DSM 43870	Putative βPM producing strain	DSM
Nocardiopsis gilva DSM 44841	Putative βPM producing strain	DSM
Saccharomyces cerevisiae CEN.PK 2-1C	Assembly strain; MATa; his3D1; leu2-3_112; ura3-52; trp1-289; MAL2-8c; SUC2	Euroscarf
Streptomyces mobaraensis DSM40847	Putative βPM producing strain	DSM
Streptomyces platensis AS600	Desotamide/spirohexenolide producing strain.	Kang et al., 2009
Streptomyces platensis M-5455	Desotamide/spirohexenolide producing strain.	BioAustralis
Streptomyces platensis M-5455; pGP9-pldR	Desotamide/spirohexenolide producing strain transformed with pGP9-pldR to activate pladienolide production.	This study
Streptomyces platensis M-5455; pGP9-spxR2	Desotamide/spirohexenolide producing strain transformed with pGP9-spxR2 to upregulate spirohexenolide production.	This study
Streptomyces platensis M-5455; pGP9-spxR3	Desotamide/spirohexenolide producing strain transformed with pGP9-spxR3 to upregulate spirohexenolide production.	This study
Streptomyces sp. CMB-0406	Heronamide C producing strain	Raju et al., 2010
Streptomyces sp. MST-110588	Wollamide and desotamide producing strain	Khalil et al., 2014
Streptomyces sp. MST-70754	Desotamide producing strain	BioAustralis
Streptomyces sp. MST-70754; pBO1-wolG2	Desotamide producing strain; transformed with pBO1-wolG2 to produce wollamide	This study

Table 7. 1 (cont.).

Media	Reagent	Supplier/Brand	Quantity
GYM	Glucose	Fisher	4.00 g/L
	Yeast Extract	Merck	4.00 g/L
	Malt Extract	Oxoid	10.00 g/L
	Calcium Carbonate	Fisher	2.00 g/L
IOM	Starch	Difco.	5.00 g/L
	Yeast Extract	Merck	2.00 g/L
	Soy Peptone	Fluka	1.00 g/L
	Instant Ocean	Aquarium Systems	16.50 g/L
LB	LB Broth	Miller	25.00 g/L
M1	Soy Peptone	Fluka	5.00 g/L
	Yeast Extract	Merck	3.00 g/L
MYM-TAP	Maltose	Sigma-Aldrich	4.00 g/L
	Yeast extract	Merck	4.00 g/L
	Malt Extract	Oxoid	10.00 g/L
PYE	Meat Extract	Fluka	0.40 g/L
	Sodium Acetate	Fisher	0.20 g/L
	Yeast Extract	Merk	0.20 g/L
	Tryptone	Difco.	0.40 g/L
R2	Sucrose	Fisher	103.00 g/L
	Potassium sulphate	Sigma-Aldrich	0.25 g/L
	Magnesium chloride	Fisher	10.12 g/L
	Casamino Acid	Sigma-Aldrich	0.10 g/L
	Glucose	Fisher	10.00 g/L
R5	Sucrose	Fisher	103.00 g/L
	Potassium sulphate	Sigma-Aldrich	0.25 g/L
	Magnesium chloride	Fisher	10.12 g/L
	Casamino Acid	Sigma-Aldrich	0.10 g/L
	Glucose	Fisher	10.00 g/L
	Trace Elements	-	2 mL
	Yeast Extract	Merck	5.00 g/L
	TES	Sigma-Aldrich	5.73 g/L
SF + M	Soya Flour	Holland & Barret	15.00 g/L
	Mannitol	Sigma-Aldrich	15.00 g/L
SM6	Corn Steep Liquor	Sigma-Aldrich	40.00 g/L
	Maltodextrin	Sigma-Aldrich	20.00 g/L
	Sodium Chloride	Fluka	2.50 g/L
	Magnesium Sulphate	Fisher	0.50 g/L
SM14	Glucose	Fisher	10.00 g/L
	Bactosoytone	Difco.	20.00 g/L
	Meat Extract	Fluka	5.00 g/L
	Sodium Chloride	Fluka	5.00 g/L
	Zinc sulphate dihydrate	Fisher	0.01 g/L

Table 7. 18: Media used in this study.

Media	Reagent	Supplier/Brand	Quantity
SM18	Glucose	Fisher	15.00 g/L
	Soluble Starch	Difco.	25.00 g/L
	Pharmedia	Pharmedium	25.00 g/L
	Cane Mollasses	Holland & Barret	20.00 g/L
	Calcium Carbonate	Fisher	8.00 g/L
SV2	Glucose	Fisher	15.00 g/L
	Glycerol	Fisher	15.00 g/L
	Soy Peptone	Fluka	15.00 g/L
	Sodium Chloride (NaCl)	Fluka	3.00 g/L
	Calcium Carbonate	Sigma-Aldrich	1.00 g/L
YPAD	Yeast Extract	Merk	10.00 g/L
	Peptone	Formedium	20.00 g/L
	Glucose	Fisher	20.00 g/L
	Adenine	Sigma-Aldrich	0.02 g/L

Table 7.2 (cont.).

Antibiotic	**Solvent**	**Stock Concentration (mg/mL)**	**Final concentration (µg/mL)**
Apramycin	Water	50	50
Carbenicillin	Water	100	100
Chloramphenicol	Ethanol	25	25
Kanmycin	Water	50	50
Streptomycin	Water	100	100

Table 7. 19: Antibiotics used in this study.

Plasmid	Description	Reference
pET28a(+)	*E. coli* protein expression vector	Novagen
pGP9	*Streptomyces* expression vector; propagates E.coli for cloning/conjugation	Zhang et al., 2009
pFF62A	Yeast shuttle vector used to construct pBO1	Schimming et al., 2014
pBO1	*Streptomyces* expression vector; propagates E.coli for cloning/conjugation; propagates in S.	This study
pBO1-wolG2	pBO1 with the gene encoding the NRPS wolG2 cloned into the NdeI start site	This study
pET28a(+)-wolG2A2	*E. coli* protein expression vector with the adenylation-domain encoding region wolG1A2 cloned into NdeI and XhoI sites	This study
pET28a(+)-wolG1A2	*E. coli* protein expression vector with the adenylation-domain encoding region wolG1A2 cloned into NdeI and XhoI sites	This study
pET28a(+)-orf6595A	*E. coli* protein expression vector with the adenylation-domain encoding region wolG1A2 cloned into NdeI and XhoI sites	This study
pGP9-pldR	*Streptomyces* expression vector; pldR cloned into NdeI and HindIII sites	This study
pGP9-spxR2	*Streptomyces* expression vector; spxR2 cloned into NdeI and HindIII sites	This study
pGP-spxR3	*Streptomyces* expression vector; spxR3 cloned into NdeI and HindIII sites	This study

Table 7. 20: Plasmids used in this work.

Primer	Sequence	Annealing Temperature (°C)	Description
PGP9_SEQ_F1	gagcggcggtcgaagggagatg	50 °C	Sequencing and colony PCR of pGP9 and pBO1 constructs
PGP9_SEQ_R1	cgagcgttctgaacaaatccag		
PLA_LUX_F2	ttaattaaggaggacacatatgcatctcttcgggcgggac	65 °C	Cloning of pldR into pGP9
PLA_LUX_R2	ccaagctcagctaattaagcttcatgccgtctggggcag		
SARP_2_F2	gcatcgattaattaaggaggacacatatgatgcggtacgaactcctg	60 °C	Cloning of spxR2 into pGP9
SARP_2_R1	ttagtgatggtgatgagatctggtctagaaagtggcgctaaagcc		
LUXR_4F	taaggaggacacatatggtgctcgggtcgcttcga	69 °C	Cloning spxR3 into pGP9
LUXR_4R	tcagctaattaagctttcagtcgagcctcgctgg		
pBO1_pGP_F1	aatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcacttcctcgctcactgactcg	70 °C	Assembly of pBO1
pBO1_pGP_R1	agcagcaccatatgatcacgtttttcattcggatctttaaacagtgcgctctgatagctcagataatgattatccggcagcagaaccggaccatcaccaacgcgttggccgattcattaa		
pBO1_pFF_F1	agaacgctcggttgccgccgggcgtttttttattggtgagaatccaagctagaaatctgcattaatgaatcggccaacgcgttggtgatggtccggttctg	70 °C	Assembly of pBO1
pBO1_pFF_R1	ccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagtgcctcggtgagttttctcc		
pBO1_wolG2_F1	ttcgagcctccttcgagccacggggccgacgatgacgacgaccaccggacgaacgcatcgattaattaaggaggacacatatggacgcggcggcatcac	70 °C	Cloning of wolG2 into pBO1
pBO1_wolG2_R1	ttcgagcctccttcgagccacggggccgacgatgacgacgaccaccggacgaacgcatcgattaattaaggaggacacatatggacgcggcggcatcac		
pCDF_wolF2_F1	aagtataagaaggagatatacatatgtcgaacccgttcg	60 °C	Cloning of wolF2 into pCDF-duet-1
pCDF_wolF2_R1	ctcgagtctggtaaagaaacggtaccgtgtgtacgagtggtg		
pET_WolG1_F1	catcaccacagccaggatccgaatccgacgtacgcgcagctcaacga	62 °C	Cloning of wolG1A2 into pET28a(+)
pET_WolG1_R1	cttaagcattatgcggccgcaagcttttagatcacgcacgcctgggcca		
pET_WolG2_F1	catcaccacagccaggatccgaatccgaccttcgcggagctggaggc	62 °C	Cloning of wolG2A2 into pET28a(+)
pET_WolG2_R1	cttaagcattatgcggccgcaagcttttacaccacggcctgcgccacgt		
pET_6595_F1	catcaccacagccaggatccgaattcgacctacgccgaactggaagc	60 °C	Cloning of orf6595A into pET28a(+)
pET_6595_R1	cttaagcattatgcggccgcaagcttttaggcggccagctccacaccct		

Table 7. 21: List of oligonucleotides used in this study. Annealing temperature for primer pairs used in PCR reactions included.

7. 2 DNA Extraction and Sequencing

7. 2. 1. Genomic DNA Extraction

High molecular weight genomic DNA was extracted according to the salting out procedure described by Kieser et al.[1] with the modifications as described here: wet mycelium (0.5 mL) from a 30 h old SV2 culture was washed with 10% sucrose (10 mL) before resuspension in SET buffer (5 mL; 75 mM NaCl, 25 mM EDTA, 20 mM TrisHCl pH 8.0) to which lysozyme (200 µL; 50 mg/mL) and ribonuclease A (15 µL; 10 mg/mL) were added. The cells were incubated overnight at 37°C; fresh lysozyme (300 µL) was added after ca. 17 h followed by an additional 2 h incubation. The subsequent steps were performed according to Kieser et al.[1].

7. 2. 2. DNA Sequencing

Sanger sequencing was performed using Mix2Seq barcoded tubes (Eurofins Genomics). Sequences were view in Chromas v 2. 6. 5 (Technelysium).

Genomic DNA of *Streptomyces sp.* MST-110588, *Streptomyces sp.* CMB-0406, *Streptomyces platensis* AS600 and *Streptomyces platensis* M-5455 was sequenced with Pacific Biosciences (PacBio) RSII SMRT technology (commissioned to The Genome Analysis Centre (TGAC) Norwich, UK)[2] and assembled *via* the HGAP2.0 pipeline. Genomic DNA of *Streptomyces* sp. MST-70754 was sequenced with Illumina MiSeq using paired end sequencing and Nextera Mate Pair library preparation and assembled in house (commissioned to the University of Cambridge DNA sequencing facility).

7. 3 Molecular Cloning and Transformation

7. 3. 1 Molecular cloning

For cloning purposes, polymerase chain reaction (PCR) was performed using Q5 high-fidelity DNA polymerase in 25 – 50 µL aliquots supplemented according to the manufacturers protocol supplemented with 3% DMSO. For screening purposes, PCR was performed using GoTaq Green Master Mix (Promega) in 10ul reactions. For colony PCR, sterile toothpicks were pressed against colonies and then into 50 µl sterile ddH_2O. H_2O sample was boiled for 1 min and 1 µl used in the reaction. Thermal cycling was carried out according to manufacturer's protocol with varying the annealing temperature for different primer pairs (Table 7. 5) in a Bio-Rad T100 thermal cycler. PCR products were run on TBE (Trizma base 10.8 g/L, boric acid 5.5 g/L, EDTA 0.9 g/L) agarose (0.6 – 1.2 %) gels supplemented with 1 µL ethidium bromide. Gel electrophoresis was carried out between 50 - 100 V for 15 -40 mins and bands were visualised under UV-light. The sizes of PCR products were confirmed by comparison to 1 kb ladder (New England Biolabs). For molecular cloning, PCR products were purified using Wizard SV gel and PCR clean up system (Promega) according to manufacturer's protocols.

Molecular cloning was performed *via* restriction enzyme digestion and ligation or Gibson Assembly. Gibson assembly was performed in 10 µl reactions using Gibson Assembly MasterMix (New England Biolabs) according to the manufacturers protocols. For restriction digestion and ligation, PCR products were digested with high-fidelity restriction nucleases overnight at 37°C. DNA was then treated with alkaline phosphatase prior to purification with Wizard SV gel and PCR clean up system (Promega) according to manufacturer's protocols. Ligations were performed overnight using T4 DNA ligase (New England Biolabs) according to the manufacturers protocols. Ligations and Gibson assemblies were subsequently transformed into electrocompetent *E. coli* DH5α using a BioRad GenePulser (2.5 kV, 200Ω, 25 µFD).

7. 3. 2 Generation and Transformation of *Saccharomyces cerevisiae*

Frozen chemically competent *S. cerevisiae* CEN.PK 2-1C were generated as per Gietz et al.[3] and were transformed using the lithium acetate, salmon sperm carrier, polyethylene glycol method[4]. Plasmids were extracted using Zymoprep yeast plasmid miniprep kit (Zymo Research), according to the manufacturers protocols and transformed into electrocompetent *E. coli* DH5α.

7. 3. 3 Conjugation of *Streptomyces* spp.

Plasmids were extracted from *E. coli* DH5α using Wizard Plus SV miniprep system (Promega) and transformed into electrocompetent *E. coli* ET12567 pUZ8002 and *E. coli* S17 λpir. The resulting strains were used for the conjugation. Conjugation of *Streptomyces* spores was carried out according to Kieser et al.[1].

7. 4 Protein Expression, Purification and Activity Assays

7. 4. 1 Expression and Purification of Adenylation-Domain Proteins

Strains were grown from overnight cultures in 1 L LB to OD600 ~0.5 prior to induction with IPTG (1mM). Following induction, cultures were grown overnight at 18°C. Cultures were centrifuged at 9000 RPM for 20 minutes and the supernatant was discarded. Pellets were resuspended in 15-20 ml P-Buffer (1M K_2HPO_4 9.4 %; 1M KH_2PO_4, 0.6 %, pH 8.0). Cells were homogenised using an Avestin EmulsiFlex-B15 (40 bar) and the supernatant was removed and filtered. Proteins were purified on an ÄKTA pure chromatography system (GE Healthcare using a His-Trap 1ml column. The column was equilibrated with P-buffer until a stable UV signal was established. The equilibrated column was sequentially washed with 5 column volumes of P-buffer containing 10 mM, 20 mM and 50 mM. imidazole. Protein was eluted with 24 column volumes of P-buffer containing of 250 mM imidazole. Fractions with an UV absorbance at 280 nm were pooled. The presence of the correct MW protein was assessed by SDS-PAGE using 12 % RunBlue SDS protein gels (Expedeon). Protein concentration was determined by Bradford assay[5] (BIORAD).

7. 4. 2 Hydroxylamine Trapping Assay

Hydroxylamine trapping assays were performed using the method described by Kadi and Challis[6] with minor modifications. 30 µL 50 mM each enzyme added to assay solution (15 mM $MgCl_2$, 2.25 mM ATP, 150 mM hydroxylamine, 3 mM amino acid). Samples were incubated for 2 h prior to the addition of stopping solution (10 % (w/v) $FeCl_3 \cdot 6H_2O$, 3.3% TCA in 0.7 *M* HCl). A540 was measured using a BioMate 3 spectrophotometer (Thermospectronic).

7. 5. Analytical Chemistry

7. 5. 1 High Performance Liquid Chromatography (HPLC) and Liquid Chromatography Mass Spectrometry (LCMS)

Analytical HPLC was performed on an Agilent 1100 series HPLC system with a Gemini-NX C_{18} 110A column (150 × 4.6 mm, 3 µm; Phenomenex) using a two-step water-MeOH gradient (10-50% MeOH over 3 min followed by 50-100% over 15 min with a 2 min hold at 100%). LCMS was performed on a Shimadzu LC–MS platform (equipped with a NexeraX2 liquid chromatograph (LC30AD), a Prominence photo diode array detector (SPD-M20A) and an LCMS-IT-TOF mass spectrometer) with chromatography over a Kinetex C_{18} 100A column (100 × 2.1 mm, 2.6 µm; Phenomenex) using a water-MeOH gradient (20-100% MeOH over 12 min with 1 min hold at 100%).

7. 5. 2 Purification of Heronamide C

For large scale purification six SV2 cultures (400 mL in 2 L Erlenmeyer flasks) were grown as described above. The resulting broth (2.4 L) was extracted with an equal volume of EtOAc. The phases were separated by centrifugation at 5000 rpm for 15 min and the solvent removed under reduced pressure to yield a yellow extract (1.8 g). This was dissolved in excess methanol (50 mL) and triturated with an equal volume of water. The insoluble fraction (67 mg) was separated and dissolved in methanol and then fractionated by preparative HPLC (see method below) to yield heronamide A (**5**) (1.2 mg), heronamide B (**6**) (0.3 mg) and heronamide C (**7**) (4.8 mg).

Preparative HPLC was performed on an UltiMate 3000 LC system with a Gemini-NX C_{18} 110A column (150 × 21.2 mm, 5 µm; Phenomenex) using a two-step MeOH gradient (10-50% MeOH over 5 min followed by 50-100% over 15 min and a 2 min hold at 100%). Typical injection volumes were between 1-1.75 mL and fractions were collected at 10.1-10.4 min **5**, 11.5 – 12 min **7** and 13-13.4 min **6**. LCMS was performed on a Shimadzu LC–MS platform (equipped with a NexeraX2 liquid chromatograph (LC30AD), a Prominence photo diode array detector (SPD-M20A) and an LCMS-IT-TOF mass spectrometer) with chromatography over a Kinetex

C_{18} 100A column (100 × 2.1 mm, 2.6 μm; Phenomenex) using a water-MeOH gradient (20-100% MeOH over 12 min with 1 min hold at 100%).

7. 5. 3 Transformation of Heronamide C to Heronamide A

7 was dissolved in methanol containing 10% DMSO at a final concentration of 225 μM while minimizing exposure to light. Aliquots (1 mL) were then incubated at 4°C, 30°C and 60°C for 7 days with all light excluded. Samples were then analyzed by HPLC using the method described above and the molar yields for each component calculated by comparison to UV calibration curve generated with isolated material.

7. 5. 4 Photochemistry of Heronamide C

7 was dissolved in methanol at a final concentration of 225 μM while minimizing the exposure to light. Aliquots (100 μL) were then placed into borosilicate vials and incubated under ambient conditions either exposed to bright sunlight by standing on a bench or placed in a box so that light was excluded and located adjacent to the first sample. After 60 min the samples were analyzed by LCMS and compared to calibration curves of authentic material in order to quantify the species present. In a second set of experiments aliquots of **7** in methanol (1 mL, 225 μM) were placed into quartz cuvettes and exposed to varying wavelengths of UV radiation (330, 375, 405 nm) for a range of exposure times (10, 20, 40, 80, 160 s) using an Atlas Photonics LUMOS 43 photoreactor. The photon flux at each wavelength was calculated using ferioxilate chemical actinometry to allow for normalization across wavelengths. From these data, the half-life, rate constant and quantum yields were determined, the latter using the methods described below:

$$\Phi_{apparent} = \frac{N_{mole}}{N_{photons}}$$

The apparent quantum yield ($\Phi_{apparent}$) was calculated at different time points with the above equation, where N_{mole} represents the number of moles of heronamide B **6** at a given time point. $N_{photons}$ represents the number of photons absorbed by heronamide C **7** and was calculated by the following equation.

$$N_{photons} = q\ \chi\ t$$

Where q is the number of photons delivered to the cuvette per second, χ is the proportion of light absorbed by the sample and t is the exposure time.

7. 6 *In Silco* Analyses

Nucleotide and protein sequences were viewed in BioEdit[7] or Geneious (Geneious).

7. 6. 1. Analysis of Genomic DNA and BGCs

Genomic DNA sequences were annotated in prodigal[8] to identify protein coding sequences, and RAST[9] to assign putative gene functions. Detection of specialised metabolic BGCs and annotation of gDNA was performed in antiSMASH v3.0[10] or v4.0[11] including ClusterFinder[12], ClusterBlast and whole-genome PFAM analysis[10]. Adenylation domain specificities were performed by NRPSpredictor2[13] or based on the specificity codes described by Stachelhaus[14] and Minowa[15], as implemented in antiSMASH[11]. Acyltransferase-domain specificities were predicted by the specificity codes described Minowa[15], as implemented in antiSMASH[11]. Gene homologues were identified from translated protein sequences using BLASTp[16]. Genes with no predicted sequence homology to proteins with known function were further analysed *via* remote homology detection using Phyre2[17].

7. 6. 2. Detection of Recombination Events

Adenylation domain coding nucleotide sequences of *Streptomyces* sp., annotated as above, were aligned using ClustalW (cost matrix: BLOSUM, gap open cost: 10, gap extend cost: 0.1)[18,19]. Recombination detection was performed using RDP[20], GENECONV[21], BOOTSCAN[22], MAXCHI[23], CHIMERA[24], SISCAN[25] and 3SEQ[26] as implemented in RDP4[27]. Using default settings. Breakpoints were plotted on translated protein sequences by manual annotation. To calculate D_S/D_N ratios, codon alignments were generated using TranslatorX[28] and the resulting alignments were submitted to Synonymous Non-synonymous Analysis Program (SNAP)[29].

7. 6. 3 Sub-cluster Genome Mining

Sub-cluster genome mining was performed in MultiGeneBlast (MGB) v.1.1.4[30]. The Database was constructed from the GenBank bacterial sub-division (BCT)[31] on 29/03/16 and updated on 14/03/17 and 22/07/18. Architecture searches were performed using the sequences of VinN, VinM, VinK and VinJ (AB086653.1)[32] and HrnI, HrnJ, HrnK, HrnS, HrnL, HrnU (LT548273.1)[33] with a maximum distance between loci as 100 Kb, all other parameters set to default. Genomes were downloaded and submitted to antiSMASH with clusterBlast[11]. Homologous BGCs from clusterBlast that were not identified by MGB were also included. A blast database of all predicted BGCs was generated in Geneious.

7. 6. 4 Phylogenetics

Protein sequence homologues of VinN, VinM, VinK and VinJ were identified by BLAST[16] using the custom database described above. Protein sequences and aligned in ClustalW (cost matrix: BLOSUM, gap open cost: 10, gap extend cost: 0.1)[18,19]. Gaps at the start or end of the sequence alignment were treated as missing data ('X'). A concatenated alignment (1708 characters) was also generated by combining data from all five individual alignments. OTUs with fewer than 3 identified homologues were excluded from the analysis. Missing sequences were regarded as missing data ('X'). Phylogenetic trees were generated using RAxML version 8[34] using

LG and WAG substitution models. For the concatenated alignment each protein sequence was regarded as an individual partition. Trees were visualised in FigTree v1.4.3[35] and Interactive tree of life (iTOL) v3[36].

7. 6. 5 Modelling Protein Structure

Protein structures were modelled using remote homology detection in Phyre2 (intensive mode)[17] and were visualised in PyMol v2.0[37].

7. 7. Bibliography

(1) Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A.; John Innes Foundation. *Practical Streptomyces Genetics*; John Innes Foundation, 2000.

(2) Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **2009**, *323* (5910), 133–138.

(3) Gietz, R. D.; Schiestl, R. H. Frozen Competent Yeast Cells That Can Be Transformed with High Efficiency Using the LiAc/SS Carrier DNA/PEG Method. *Nat. Protoc.* **2007**, *2* (1), 1–4.

(4) Gietz, R. D.; Schiestl, R. H. Quick and Easy Yeast Transformation Using the LiAc/SS Carrier DNA/PEG Method. *Nat. Protoc.* **2007**, *2* (1), 35–37.

(5) Bradford, M. M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **1976**, *72* (1–2), 248–254.

(6) Kadi, N.; Challis, G. L. Chapter 17 Siderophore Biosynthesis. In *Methods in enzymology*; 2009; Vol. 458, pp 431–457.

(7) Hall. BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*.

(8) Hyatt, D.; Chen, G.-L.; LoCascio, P. F.; Land, M. L.; Larimer, F. W.; Hauser, L. J. Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinformatics* **2010**, *11* (1), 119.

(9) Aziz, R. K.; Bartels, D.; Best, A. A.; DeJongh, M.; Disz, T.; Edwards, R. A.; Formsma, K.; Gerdes, S.; Glass, E. M.; Kubal, M.; et al. The RAST Server: Rapid Annotations Using Subsystems Technology. *BMC Genomics* **2008**, *9* (1), 75.

(10) Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Muller, R.; Wohlleben, W.; et al. AntiSMASH 3.0--a Comprehensive Resource for the Genome Mining of Biosynthetic Gene Clusters. *Nucleic Acids Res.* **2015**, gkv437-.

(11) Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran,

H. G.; de los Santos, E. L. C.; Kim, H. U.; Nave, M.; et al. AntiSMASH 4.0—improvements in Chemistry Prediction and Gene Cluster Boundary Identification. *Nucleic Acids Res.* **2017**, *45* (W1), W36–W41.

(12) Cimermancic, P.; Medema, M. H.; Claesen, J.; Kurita, K.; Wieland Brown, L. C.; Mavrommatis, K.; Pati, A.; Godfrey, P. A.; Koehrsen, M.; Clardy, J.; et al. Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. *Cell* **2014**, *158* (2), 412–421.

(13) Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. NRPSpredictor2--a Web Server for Predicting NRPS Adenylation Domain Specificity. *Nucleic Acids Res.* **2011**, *39* (Web Server issue), W362-7.

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

(15) Minowa, Y.; Araki, M.; Kanehisa, M. Comprehensive Analysis of Distinctive Polyketide and Nonribosomal Peptide Structural Motifs Encoded in Microbial Genomes. *J. Mol. Biol.* **2007**, *368* (5), 1500–1517.

(16) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410.

(17) 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.* **2015**, *10* (6), 845–858.

(18) 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.* **1994**, *22* (22), 4673–4680.

(19) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23* (21), 2947–2948.

(20) Martin, D.; Rybicki, E. RDP: Detection of Recombination amongst Aligned Sequences. *Bioinformatics* **2000**, *16* (6), 562–563.

(21) Padidam, M.; Sawyer, S.; Fauquet, C. M. Possible Emergence of New Geminiviruses by Frequent Recombination. *Virology* **1999**, *265* (2), 218–225.

(22) Salminen, M. O.; Carr, J. K.; Burke, D. S.; McCutchan, F. E. Identification of Breakpoints in Intergenotypic Recombinants of HIV Type 1 by Bootscanning. *AIDS Res. Hum. Retroviruses* **1995**, *11* (11), 1423–1425.

(23) Smith, J. M. Analyzing the Mosaic Structure of Genes. *J. Mol. Evol.* **1992**, *34* (2), 126–129.

(24) Posada, D.; Crandall, K. A. Evaluation of Methods for Detecting Recombination from DNA Sequences: Computer Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (24), 13757–13762.

(25) Gibbs, M. J.; Armstrong, J. S.; Gibbs, A. J. Sister-Scanning: A Monte Carlo Procedure for Assessing Signals in Recombinant Sequences. *Bioinformatics* **2000**, *16* (7), 573–582.

(26) Boni, M. F.; Posada, D.; Feldman, M. W. An Exact Nonparametric Method for Inferring Mosaic Structure in Sequence Triplets. *Genetics* **2007**, *176* (2), 1035–1047.

(27) Martin, D. P.; Murrell, B.; Golden, M.; Khoosal, A.; Muhire, B. RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes. *Virus Evol.* **2015**, *1* (1), vev003.

(28) Abascal, F.; Zardoya, R.; Telford, M. J. TranslatorX: Multiple Alignment of Nucleotide Sequences Guided by Amino Acid Translations. *Nucleic Acids Res.* **2010**, *38* (suppl_2), W7–W13.

(29) Korber, B. HIV Signature and Sequence Variation Analysis. In *Computational and Evolutionary Analysis of HIV Molecular Sequences*; 2000; pp 55–72.

(30) Medema, M. H.; Takano, E.; Breitling, R. Detecting Sequence Homology at the Gene Cluster Level with MultiGeneBlast. *Mol. Biol. Evol.* **2013**, *30* (5), 1218–1223.

(31) Benson, D. A.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Sayers, E. W. GenBank. *Nucleic Acids Res.* **2015**, *43* (D1), D30–D35.

(32) Ogasawara, Y.; Katayama, K.; Minami, A.; Otsuka, M.; Eguchi, T.; Kakinuma, K. Cloning, Sequencing, and Functional Analysis of the Biosynthetic Gene Cluster of Macrolactam Antibiotic Vicenistatin in Streptomyces Halstedii. *Chem. Biol.* **2004**, *11* (1), 79–86.

(33) Booth, T. J.; Alt, S.; Capon, R. J.; Wilkinson, B.; Liu, H.-W.; Zhang, G.; Ma, L.; Zhang, Q.; Tian, X.; Zhang, S.; et al. Synchronous Intramolecular Cycloadditions of the Polyene Macrolactam Polyketide Heronamide C. *Chem. Commun. (Camb).* **2016**, *52* (38), 6383–6386.

(34) Stamatakis, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **2014**, *30* (9), 1312–1313.

(35) FigTree v.1.4.3 http://tree.bio.ed.ac.uk/software/figtree/ (accessed Sep 27, 2018).

(36) Letunic, I.; Bork, P. Interactive Tree of Life (ITOL) v3: An Online Tool for the Display and Annotation of Phylogenetic and Other Trees. *Nucleic Acids Res.* **2016**, *44* (W1), W242-5.

(37) Schrödinger, L. PyMOL The PyMOL Molecular Graphics System, Version 2.0.

Appendix 1: Protein Purification

Contents

A1. 1. Purification of WolG1A2

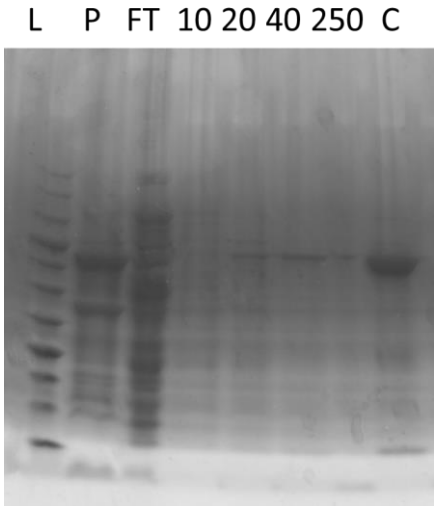


Figure A1. 60: SDS-PAGE of initial WolG1A2 purification fractions. From left to right: NEB Blue Protein Standard Broad Range, pellet (P), column flow through (FT), 10 mM imidazole (10), 20 mM imidazole (20), 40 mM imidazole (40), 250 mM imidazole (250), concentrated sample post buffer exchange (C).

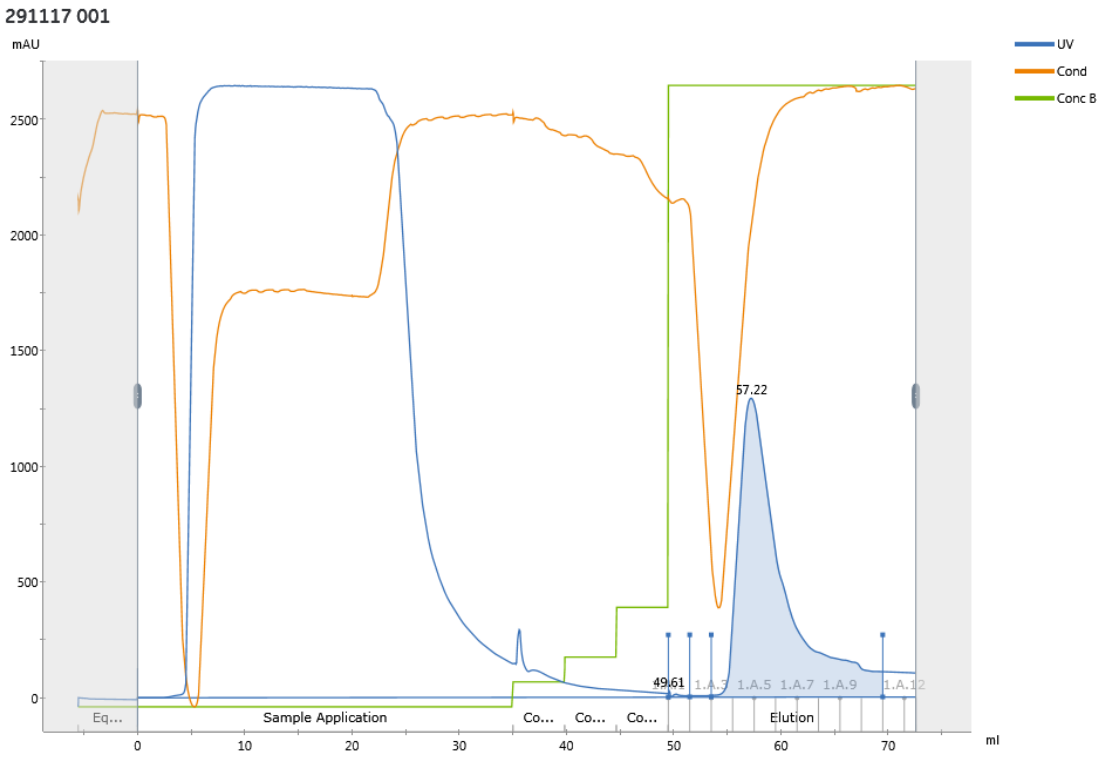


Figure A1. 61: Output from the ÄKTA pure chromatography system from the purification of WolG1A2. UV recorded at 280 nm (blue), percentage of 250 mM imidazole (0%, 4%, 8%, 16%, 100%, green) and conductivity (mS/cm, orange).

A1. 2. Purification of WolG2A2

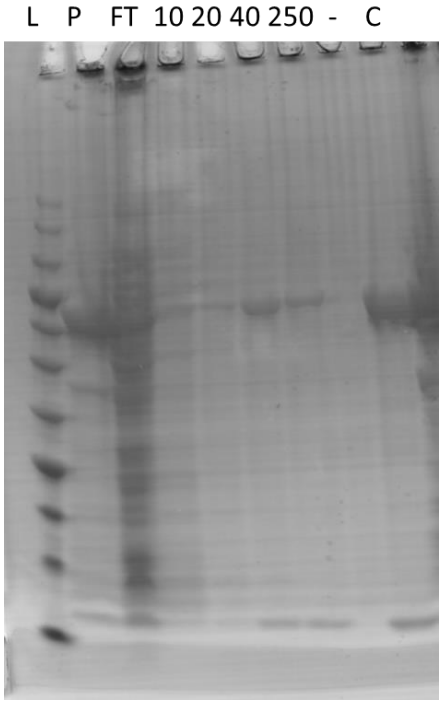


Figure A1. 62: SDS-PAGE of initial WolG2A2 purification fractions. From left to right: NEB Blue Protein Standard Broad Range, pellet (P), column flow through (FT), 10 mM imidazole (10), 20 mM imidazole (20), 40 mM imidazole (40), 250 mM imidazole (250), empty lane (-), concentrated sample post buffer exchange (C).

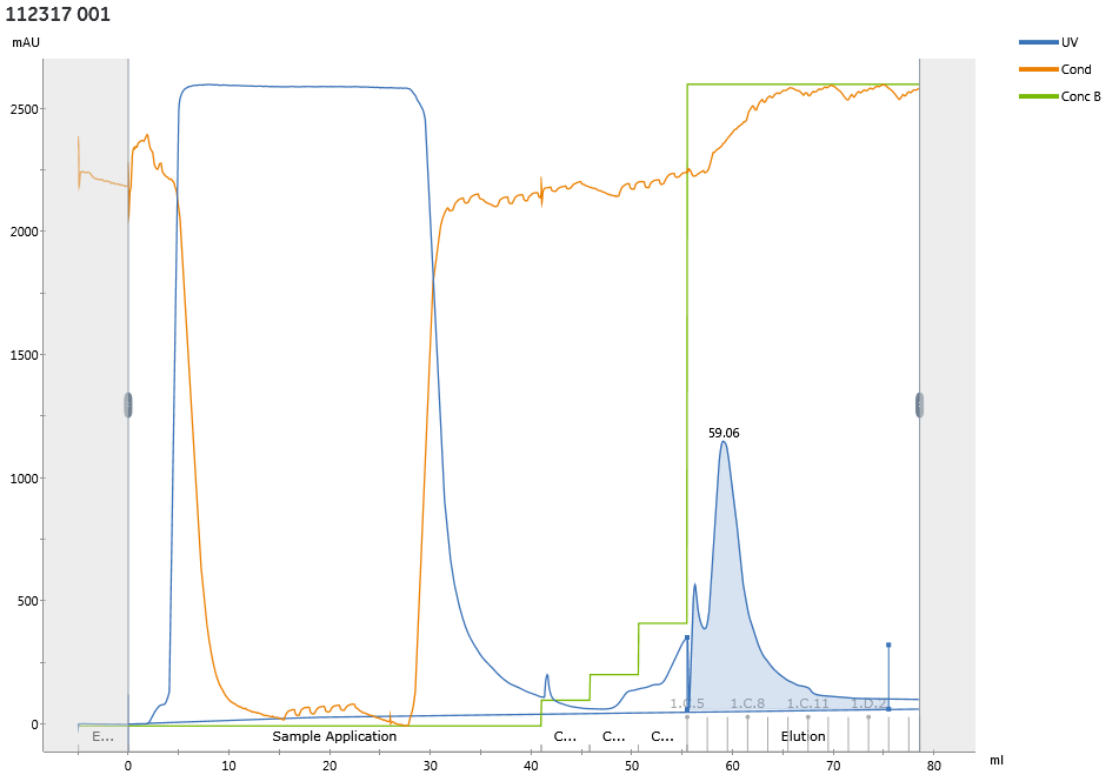


Figure A1. 63: Output from the ÄKTA pure chromatography system from the purification of WolG2A2. UV recorded at 280 nm (blue), percentage of 250 mM imidazole (0%, 4%, 8%, 16%, 100%, green) and conductivity (mS/cm, orange).

A1. 3. Purification of ORF6595A

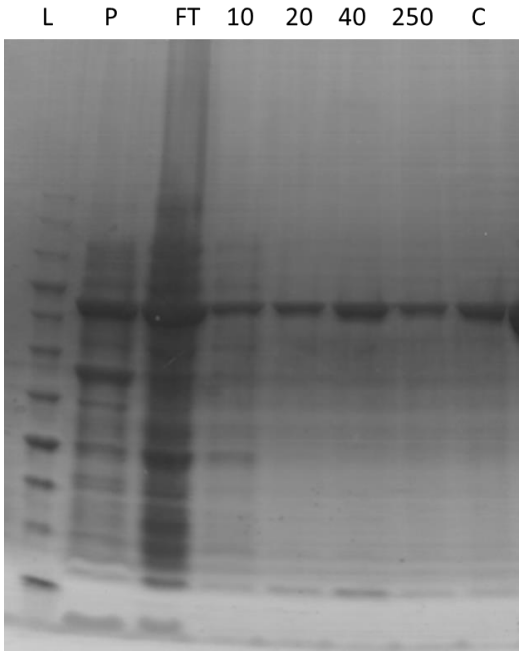


Figure A1. 64: SDS-PAGE of initial Orf6595A purification fractions. From left to right: NEB Blue Protein Standard Broad Range, pellet (P), column flow through (FT), 10 mM imidazole (10), 20 mM imidazole (20), 40 mM imidazole (40), 250 mM imidazole (250), empty lane (-), concentrated sample post buffer exchange (C).

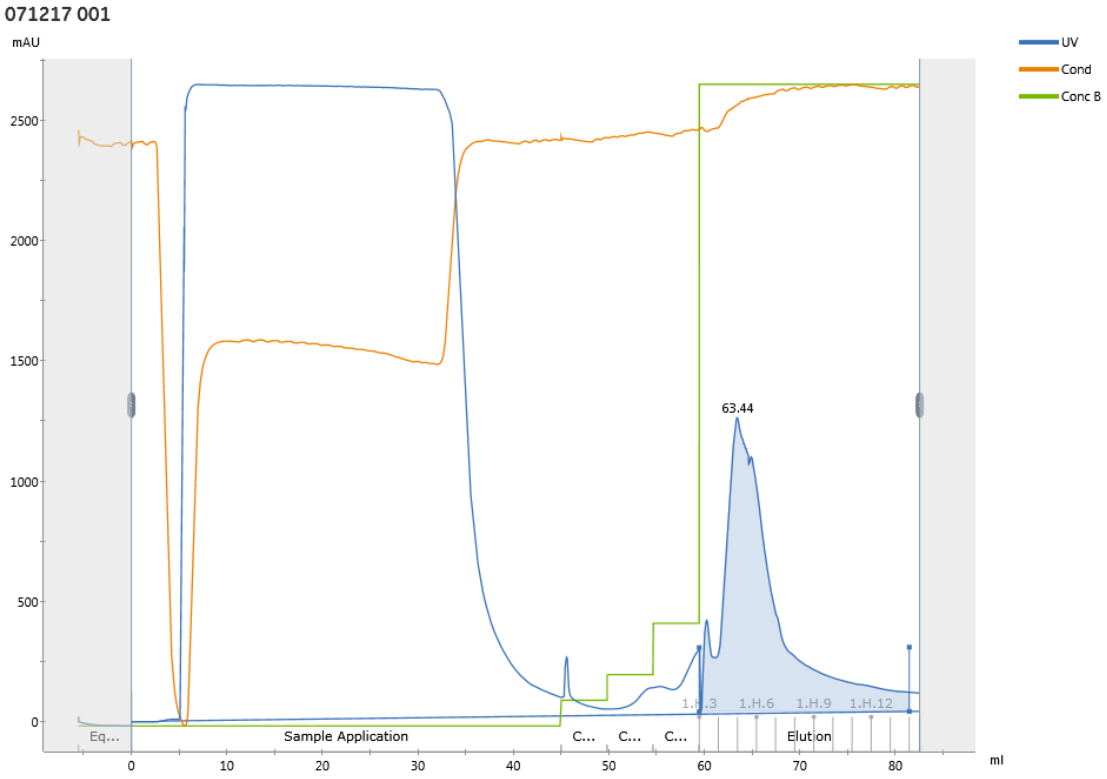


Figure A1. 65: Output from the ÄKTA pure chromatography system from the purification of Orf6595A. UV recorded at 280 nm (blue), percentage of 250 mM imidazole (0%, 4%, 8%, 16%, 100%, green) and conductivity (mS/cm, orange).

Appendix 2: MS/MS Tables

Contents

A2. 1 Desotamide A and Wollamide A

A2. 1. 1 MS/MS of Desotamide A from *Streptomyces sp.* MST-110588 (Authentic Standard)

Precursor	719.3809						
Base Peak	m/z 674.3701 (Inten : 312,538)						
m/z	Rel. Inten.	m/z	Rel. Inten.	m/z	Rel. Inten	m/z	Rel. Inten.
704.3695	8.07	566.3584	3.3	464.2662	34.79	377.1637	4.72
703.3703	44.84	565.3525	15.34	458.2888	2	368.685	5.79
702.3667	99.59	562.2903	9.77	458.1926	3.31	368.1821	6.51
701.3828	17.17	561.2852	39.47	449.2793	1.3	364.1058	1.46
693.4064	6.26	560.3128	5.92	449.2049	8.01	363.2062	2.26
692.3983	25	546.2981	2.97	448.2964	10.56	363.1117	13.51
691.3983	60.09	545.2638	1.52	448.2013	45.02	360.1747	5.26
684.3642	1.16	544.2626	12.64	447.2288	2.16	359.6759	6.03
676.3824	8.36	543.2749	8.69	446.2574	3.8	359.1769	6.83
675.376	52.36	534.3117	1.34	435.2739	2.18	357.1878	2.34
674.3701	100	533.3092	8.57	434.2539	3.12	352.1467	10.08
673.3844	4.84	532.3282	1.46	432.2758	5.85	351.1821	19.36
658.3562	2.93	521.3338	1.83	432.1999	2.26	346.1737	1.04
657.3559	6.34	520.3313	15.14	431.2664	16.21	346.0813	5.33
656.359	5.77	518.2691	1.29	431.1837	11.81	345.1938	1.4
646.3817	5.74	516.2829	5.27	430.1971	9.01	345.1082	1.4
634.3777	3.72	515.298	2.62	429.2278	1.53	336.1164	6.6
632.3924	5.83	510.2741	1.16	421.2306	1.71	335.1156	43.67
629.3522	2.38	506.3163	9.45	420.2284	7.58	334.2514	8.29
618.348	7.31	505.3155	37.62	419.245	2.01	320.196	3.05
617.3472	16.21	493.224	2.35	413.2577	2	319.2188	1.83
605.3451	1.77	492.2592	1.34	407.249	14.07	318.0932	10.96
604.3425	7.79	491.2495	1.98	405.2811	1.22	317.1089	4.65
603.3363	13.97	490.2472	8.42	393.2368	1.09	294.6466	1.1
600.3072	3.7	488.2949	1.78	392.2324	18.57	294.163	1.22
589.2884	3.68	487.2902	1.1	391.2752	3.23	282.6537	1.46
588.3205	6.03	477.2984	1.46	389.2716	4.74	266.1904	2.81
587.335	6.67	476.2905	8.45	380.2564	10.71	266.0931	5.5
579.314	6.96	476.1982	6.49	380.1486	2.07	248.084	1.09
578.3221	32.1	475.217	3	379.2711	36.71	238.0962	1.64
577.3522	24.01	466.2338	1.77	379.1784	5.18		
575.3283	1.05	465.2409	24.58	378.1864	1.38		

A2. 1. 2 MS/MS of Desotamide A from *Streptomyces sp.* MST-70754

Precursor	719.3919										
Base Peak	m/z 702.3649 (Inten : 517,946)										
m/z	Rel. Inten	Present in WT	m/z	Rel. Inten.	Present in WT	m/z	Rel. Inten.	Present in	m/z	Rel. Inten.	Present in WT
704.3674	5.68	TRUE	560.302	0.96	TRUE	457.2062	0.2	FALSE	361.2518	0.2	TRUE
703.3691	35.53	TRUE	555.2996	0.2	TRUE	453.275	0.19	FALSE	361.169	0.53	TRUE
702.9083	0.19	TRUE	551.3363	0.52	TRUE	452.2744	0.24	FALSE	360.2242	0.28	TRUE
702.8588	0.58	TRUE	548.3202	0.19	FALSE	449.2918	0.19	TRUE	360.1725	0.91	TRUE
702.3649	100	TRUE	547.3431	0.49	TRUE	449.2021	5.58	TRUE	360.1179	0.19	FALSE
701.3806	21.35	TRUE	546.3	0.68	TRUE	448.2908	2.81	TRUE	359.6661	0.8	TRUE
700.4125	0.19	FALSE	545.301	0.24	TRUE	448.1998	27.86	TRUE	359.1678	1.28	TRUE
693.4011	3.11	TRUE	545.2712	0.44	TRUE	447.223	0.77	TRUE	357.1853	0.98	TRUE
692.3959	18.5	TRUE	544.2654	6.96	TRUE	446.2716	0.58	TRUE	352.1447	4.79	TRUE
692.0191	0.19	FALSE	543.2759	4.72	TRUE	446.2322	0.19	FALSE	351.1793	10.57	TRUE
691.3978	54.94	TRUE	542.3296	0.28	TRUE	445.1687	0.19	TRUE	349.1318	0.47	FALSE
684.354	1.35	TRUE	535.3245	0.2	FALSE	443.2518	0.19	FALSE	347.1024	0.19	TRUE
676.3852	2.02	TRUE	534.2998	0.19	TRUE	435.292	0.38	TRUE	346.081	2.19	TRUE
675.3761	36.7	TRUE	533.309	4.76	TRUE	435.2263	0.19	FALSE	345.6614	0.19	FALSE
675.093	0.29	FALSE	532.3267	0.24	TRUE	434.2697	1.01	TRUE	345.2452	0.19	TRUE
674.8507	0.58	TRUE	529.2445	0.19	TRUE	434.176	0.19	FALSE	345.1759	0.2	TRUE
674.3709	97.14	TRUE	528.2685	0.19	TRUE	433.2111	0.43	TRUE	345.098	0.39	TRUE
673.3866	4.51	TRUE	526.3084	0.19	FALSE	432.2767	2.35	TRUE	344.2233	0.19	TRUE
658.3846	1.74	TRUE	521.3338	1.44	TRUE	432.2366	0.62	TRUE	337.1296	0.2	FALSE
657.3517	4.83	TRUE	520.3285	9.79	TRUE	431.2671	8.62	TRUE	336.1204	2.05	TRUE
656.3686	4.61	TRUE	519.2698	0.19	TRUE	431.1797	6.23	TRUE	335.2086	0.92	TRUE
648.3605	2.42	TRUE	518.2819	0.19	TRUE	430.2585	0.2	FALSE	335.114	24.94	TRUE
647.6018	0.19	FALSE	518.2604	0.19	TRUE	430.1885	4.43	TRUE	334.2486	2.03	FALSE
647.4119	0.19	TRUE	517.2735	0.28	TRUE	429.239	0.34	TRUE	333.1675	0.2	FALSE
646.3779	3.36	TRUE	516.2884	1.96	TRUE	421.212	0.48	TRUE	322.1582	0.48	TRUE
635.3536	0.24	TRUE	515.2962	0.87	TRUE	420.2214	2.35	TRUE	321.2881	0.19	FALSE
634.3823	2.53	TRUE	509.2942	0.43	TRUE	419.2526	0.62	FALSE	321.1878	0.28	TRUE
633.4009	0.2	FALSE	506.3138	4.39	TRUE	418.2497	0.19	TRUE	320.1928	1.19	TRUE
632.4008	2.07	TRUE	505.3138	27.36	TRUE	417.2679	0.2	FALSE	319.2124	0.29	TRUE
629.3109	0.19	FALSE	504.1923	0.19	TRUE	417.2392	0.48	TRUE	319.168	0.19	TRUE
622.403	0.2	FALSE	498.2715	0.39	TRUE	414.2529	0.19	TRUE	319.1214	0.68	TRUE
620.3567	0.66	FALSE	497.2309	0.29	FALSE	408.256	0.19	TRUE	319.0914	0.24	TRUE
618.3586	2.54	FALSE	494.1985	0.19	TRUE	407.2465	4.8	TRUE	318.1855	0.2	TRUE
618.3238	0.77	TRUE	493.2864	0.19	TRUE	405.2834	0.2	TRUE	318.0868	3.45	TRUE
617.3456	10	TRUE	493.2391	0.34	TRUE	405.2458	0.28	TRUE	317.1009	1.58	TRUE
616.3899	0.19	TRUE	492.2609	1.01	TRUE	402.2017	0.24	TRUE	308.0505	0.19	TRUE
613.3711	0.19	FALSE	491.2557	1.43	FALSE	397.1738	0.2	TRUE	307.1343	0.39	TRUE
606.3169	0.34	TRUE	490.2502	3.71	FALSE	393.2423	0.2	TRUE	305.1593	0.19	FALSE
605.3235	1.69	TRUE	488.3058	0.19	TRUE	392.2302	10.17	TRUE	300.177	0.19	TRUE
604.3474	5.45	TRUE	487.3039	0.39	TRUE	391.2628	1.06	TRUE	291.067	0.19	FALSE
603.3291	10.64	TRUE	483.3505	0.19	FALSE	390.2973	0.19	FALSE	283.1481	0.2	TRUE
601.3131	0.19	FALSE	479.2644	0.28	FALSE	389.2732	3.31	TRUE	282.6303	0.2	TRUE
600.298	0.71	TRUE	478.2945	0.75	TRUE	387.1688	0.24	TRUE	279.1595	0.19	FALSE
592.3318	0.71	FALSE	478.261	0.68	TRUE	381.1278	0.19	TRUE	276.0509	0.2	TRUE
591.3353	0.2	FALSE	477.2887	0.19	TRUE	380.2648	2.92	TRUE	267.1803	0.19	FALSE
590.2696	0.19	TRUE	477.2011	0.67	TRUE	380.2238	0.59	FALSE	266.1872	1.61	TRUE
589.8731	0.19	FALSE	476.2837	4.76	TRUE	380.1401	0.52	TRUE	266.0879	2.76	TRUE
589.2804	0.43	TRUE	476.1922	2.52	TRUE	379.2705	20.59	TRUE	262.1268	0.19	TRUE
588.3207	2.35	TRUE	475.202	1.34	TRUE	379.1729	2.63	TRUE	249.5936	0.19	FALSE
587.3331	5.84	TRUE	474.2408	0.82	TRUE	378.1833	0.58	TRUE	238.0936	0.76	TRUE
579.3197	3.43	TRUE	466.2217	0.97	TRUE	377.1524	2.07	TRUE	237.0656	0.19	FALSE
578.32	25.34	TRUE	465.2652	5.44	FALSE	375.1778	0.19	TRUE	234.1232	0.19	TRUE
577.3538	20.43	TRUE	465.226	9.52	TRUE	372.1843	0.19	FALSE	233.0824	0.2	TRUE
575.3611	0.19	FALSE	464.6663	0.19	FALSE	368.6883	1.18	TRUE	231.1424	0.19	FALSE
575.2916	0.62	TRUE	464.2667	21.39	TRUE	368.1797	1.86	TRUE	129.8562	0.19	FALSE
571.2816	0.53	TRUE	462.2965	0.19	FALSE	366.2512	0.19	FALSE	94.8875	0.19	FALSE
566.3645	1.28	TRUE	462.2184	1.76	TRUE	366.162	0.28	FALSE	91.1005	0.19	FALSE
565.3522	14.17	TRUE	460.3029	0.19	FALSE	365.2285	0.84	TRUE	82.5667	0.2	FALSE
563.297	0.19	TRUE	459.1829	0.19	TRUE	363.2101	2.28	TRUE			
562.287	6.19	TRUE	458.2789	0.39	TRUE	363.1094	6.96	TRUE			
561.286	26.32	TRUE	458.164	0.2	TRUE	362.1389	0.2	TRUE			

A2. 1. 3. MS/MS of Wollamide A from *Streptomyces sp.* MST-110588

Precursor	776.4387		
Base Peak	m/z 731.4297 (Inten : 391,073)		
m/z	Rel. Inten.	m/z	Rel. Inten.
764.82	3.27	506.2613	3.27
760.4408	3.27	505.9675	3.27
759.4236	7.46	505.2648	4.94
758.4512	4.11	504.2903	3.27
748.4622	31.66	487.2599	4.94
742.4242	4.11	486.2514	3.27
741.4123	20.41	477.2771	4.94
732.4354	19.15	476.2858	14.93
731.4297	100	460.2673	4.11
715.3927	3.27	435.2924	3.27
714.42	3.27	434.2305	4.11
713.3982	3.27	431.2622	5.78
705.4601	3.27	408.2311	11.57
689.4436	4.94	407.2401	5.78
662.3846	3.27	392.2172	3.27
661.4126	11.55	391.2018	4.94
660.391	4.11	389.2854	3.27
657.3406	4.94	379.2768	20
645.3897	3.27	374.1957	3.27
645.3186	3.27	364.2019	4.11
635.367	4.11	357.1796	3.27
634.4153	12.39	357.1443	3.27
618.3558	8.22	339.1837	3.27
617.3516	30.38	334.2541	3.27
591.3241	4.11	320.1991	3.27
590.3581	3.27	288.1055	3.27
587.5059	3.27	269.3656	3.27
566.3562	3.27	266.1907	3.27
566.3118	3.27	231.2586	3.27
565.3564	22.5	155.5566	3.27
547.295	3.27	72.8127	3.27
522.3349	7.38	72.6695	3.27
521.3231	24.24	60.8087	3.27
520.3346	19.15		

A2. 1. 4 Wollamide A from *Streptomyces sp.* MST-70754 pBO1-*wolG2*

Precursor	776.4479				
Base Peak	m/z 617.3465 (Inten : 718,978)				
m/z	Rel. Inten.	Present in WT	m/z	Rel. Inten.	Present in WT
760.4328	5.41	TRUE	476.2957	1.62	TRUE
759.4252	19.05	TRUE	470.2486	1.92	FALSE
758.4491	1.31	FALSE	461.2436	5.09	FALSE
741.4197	1.28	TRUE	460.2593	14.73	TRUE
663.3727	4.93	FALSE	453.2585	4.6	FALSE
662.3732	15.35	TRUE	452.2664	1.76	FALSE
635.3727	16.6	TRUE	435.2679	5.59	TRUE
634.3741	36.99	TRUE	434.2538	1.61	FALSE
619.3776	5.58	FALSE	433.2207	3.99	FALSE
618.357	38.52	TRUE	431.2751	1.28	TRUE
617.3465	100	FALSE	418.2142	6.34	FALSE
603.341	17.4	FALSE	408.2509	2.39	FALSE
602.3658	13.94	FALSE	407.2438	16.93	TRUE
599.336	1.61	FALSE	401.1997	1.91	FALSE
591.3709	2.26	FALSE	392.2725	1.76	FALSE
590.3652	4.45	TRUE	392.2335	9.02	TRUE
586.3046	4.13	FALSE	391.2681	14.37	FALSE
576.3741	1.28	FALSE	379.2688	9.07	TRUE
575.3733	7.83	FALSE	374.1793	1.6	TRUE
573.3468	4.43	FALSE	365.2038	1.13	FALSE
567.3633	1.43	FALSE	364.1975	22.11	TRUE
566.3527	16.62	TRUE	347.1765	11.97	FALSE
565.3495	53.07	TRUE	345.1888	1.28	FALSE
561.2791	1.46	FALSE	339.1881	1.3	TRUE
547.3364	7.4	TRUE	335.2062	6.73	FALSE
546.3054	7.21	FALSE	322.1544	1.43	FALSE
543.2752	7.29	FALSE	320.1969	2.09	TRUE
521.3309	10	TRUE	319.207	1.44	FALSE
520.3268	41.47	TRUE	278.1905	1.45	FALSE
505.3137	11.22	TRUE	276.2104	1.12	FALSE
488.2852	6.3	FALSE	266.1888	2.71	TRUE
478.284	6.76	FALSE	251.1125	2.73	FALSE
477.2815	37.05	TRUE	234.0773	1.13	FALSE

A2. 2 The Heronamides

A2. 2. 1 MS/MS of Heronamide C from *Streptomyces sp.* CMB-0406 (Authentic Standard)

Precursor	450.3022				
Base Peak	m/z 340.1920 (Inten : 110,781)				
m/z	Rel. Inten.	m/z	Rel. Inten.	m/z	Rel. Inten.
449.2482	0.79	308.191	0.8	219.1174	1.57
443.8092	1.15	302.2497	12.83	218.1206	1.37
440.7496	0.78	300.2452	0.79	215.1734	0.79
433.3023	1.16	298.2287	1.55	214.1623	0.79
433.2634	0.79	297.1506	0.8	211.1025	0.79
432.2925	31.99	296.2029	0.8	209.1356	0.79
422.3024	7.51	294.1919	0.79	207.1108	1.58
414.2857	8.19	292.172	0.79	201.1683	0.79
404.2927	3.33	288.2004	0.79	198.1375	0.79
393.2358	0.79	288.1371	1.16	197.128	1.15
390.3008	0.79	286.1775	1.95	195.122	0.97
380.2256	1.38	284.2246	0.79	194.1487	1.96
373.6969	0.79	284.1617	1.16	193.1028	1.94
372.256	0.82	281.167	0.79	191.1251	0.79
368.2543	0.79	274.252	6.49	190.1594	0.79
366.1992	0.79	272.1396	0.79	187.162	1.18
364.2743	0.79	268.1696	1.15	185.1009	0.78
362.2036	0.79	267.2083	0.82	183.1142	0.79
354.252	0.8	263.1384	0.79	181.1006	1.96
351.2219	0.78	258.1848	0.79	179.0857	0.79
341.1926	3.5	253.1299	0.79	176.103	0.79
340.192	100	247.1418	0.8	173.1453	0.79
326.2072	4.52	242.1315	0.79	170.8948	0.8
324.2007	0.79	239.1873	1.57	166.1251	0.79
322.1785	7.44	238.15	1.16	165.069	1.16
312.2002	1.96	237.1293	0.8	155.0649	0.79
310.6599	1.15	229.113	0.79	151.077	1.56
310.2162	0.79	225.1636	0.79	144.072	1.18

A2. 2. 2 MS/MS of Heronamide C from *Streptomyces sp.* CMB-0406

Precursor	450.2978							
Base Peak	m/z 340.1930 (Inten : 103,445)							
m/z	Rel. Inten.	Present in Standard	m/z	Rel. Inten.	Present in Standard	m/z	Rel. Inten.	Present in Standard
452.3087	15.13	FALSE	300.2322	1.07	TRUE	223.1508	2	FALSE
433.2988	24.62	TRUE	294.1863	5.6	TRUE	223.1141	1.48	FALSE
432.2928	69.21	TRUE	287.1855	1.94	FALSE	220.1707	1.08	FALSE
423.314	10.73	FALSE	286.1873	17.45	TRUE	219.1203	1.06	TRUE
422.3087	22.05	TRUE	285.229	8.95	FALSE	218.155	1.08	FALSE
415.2941	3.43	FALSE	284.2365	14.84	TRUE	218.1206	1.81	TRUE
414.2846	6.42	TRUE	284.1602	1.26	TRUE	214.1241	1.07	FALSE
405.2897	2.36	FALSE	282.1874	1.82	FALSE	211.1054	1.07	TRUE
404.3009	7.43	TRUE	275.266	3.15	FALSE	204.1315	1.45	FALSE
388.2618	1.8	FALSE	274.2558	32.15	TRUE	201.171	1.81	TRUE
380.2512	2.59	FALSE	271.1435	2.22	FALSE	200.1538	1.07	FALSE
380.2266	1.64	TRUE	270.1463	2.2	FALSE	195.1195	1.8	TRUE
362.223	1.81	TRUE	269.1815	1.07	FALSE	193.1082	1.81	TRUE
355.244	1.28	FALSE	268.1726	1.82	TRUE	187.152	1.48	TRUE
341.1939	10.16	TRUE	267.2135	19	TRUE	185.1289	2.18	TRUE
340.193	100	TRUE	261.1332	1.08	FALSE	183.1169	1.81	TRUE
337.2453	1.08	FALSE	258.2344	2.38	TRUE	181.1069	1.27	TRUE
326.2153	5.07	TRUE	253.1774	1.07	FALSE	178.159	1.6	FALSE
323.1888	1.07	FALSE	251.1338	1.82	FALSE	171.123	1.48	FALSE
322.1843	15.65	TRUE	239.1758	1.07	TRUE	167.0907	1.07	FALSE
312.1983	7.76	TRUE	239.1428	1.28	FALSE	159.1233	1.07	FALSE
304.2042	1.08	FALSE	235.1546	3.21	FALSE	159.088	1.47	FALSE
303.2532	6.07	FALSE	233.1325	1.78	FALSE	157.1023	1.07	FALSE
303.1803	1.07	FALSE	230.1546	1.07	FALSE			
302.2497	40.51	TRUE	225.1343	2.35	FALSE			

A2. 2. 3 MS/MS of Heronamide B from *Streptomyces sp.* CMB-0406

Precursor	450.3014				
Base Peak	m/z 340.1927 (Inten : 71,462)				
m/z	Rel. Inten.	Present in Standard	m/z	Rel. Inten.	Present in Standard
452.294	1.18	FALSE	274.2552	27.73	TRUE
433.2875	8.84	TRUE	272.158	1.16	FALSE
432.291	56.57	TRUE	270.1923	1.61	FALSE
423.3138	1.4	FALSE	270.1514	1.21	FALSE
422.3123	16.83	TRUE	268.2184	2.73	TRUE
414.2856	1.76	TRUE	267.2266	4.28	TRUE
404.2915	3.46	TRUE	257.2297	1.18	FALSE
380.2548	1.18	FALSE	247.1154	1.18	TRUE
380.2184	1.18	TRUE	244.1411	1.18	FALSE
376.2645	1.62	FALSE	236.1137	1.18	FALSE
341.1958	4.04	TRUE	233.1851	1.18	FALSE
340.1927	100	TRUE	232.1751	1.98	FALSE
335.2274	1.18	FALSE	218.1825	1.18	FALSE
334.2378	1.62	FALSE	218.119	2.57	TRUE
326.2111	2.17	TRUE	207.1136	1.99	TRUE
322.1806	7.83	TRUE	204.1527	1	FALSE
312.197	4.68	TRUE	201.1181	1.18	FALSE
303.2501	1.36	FALSE	200.1142	1.18	FALSE
302.2513	21.2	TRUE	197.1319	2.91	TRUE
298.2157	1.18	TRUE	192.1477	1.18	FALSE
287.1912	1.18	FALSE	190.1235	1.98	FALSE
286.1835	9.29	TRUE	183.1272	2.16	TRUE
285.2348	3.47	FALSE	178.1624	1.39	FALSE
284.245	6.23	TRUE	176.118	1.18	TRUE
282.1639	1.18	FALSE	157.0987	1.17	FALSE
277.15	1.18	FALSE			

A2. 2. 4 MS/MS of Heronamide A from *Streptomyces sp.* CMB-0406 (Authentic Standard)

Precursor	466.2946				
Base Peak	m/z 356.1864 (Inten : 471,879)				
m/z	Rel. Inten.	m/z	Rel. Inten.	m/z	Rel. Inten.
449.2916	2.49	320.1698	1.01	226.1287	0.44
448.2878	7.87	316.2299	0.48	225.1216	0.41
439.302	0.41	312.1757	0.44	224.1112	0.34
438.3057	0.34	311.1906	0.36	221.1312	0.37
431.2782	2.1	310.1781	0.83	216.1015	0.41
430.2744	1.53	296.1868	0.28	214.1259	0.21
421.3037	0.24	294.1484	0.31	212.1089	0.38
420.2951	0.67	292.172	0.27	211.1161	0.48
402.2749	0.34	290.1511	0.28	209.1289	0.34
376.2612	0.62	285.1727	0.27	208.1158	0.41
371.2395	0.51	284.164	0.76	207.1126	1.04
370.2392	0.89	282.1654	0.41	206.1196	0.92
358.1936	1.52	274.1396	0.34	204.1021	0.48
357.542	0.41	268.1355	0.65	200.1066	0.27
357.354	0.21	266.1565	0.7	199.1052	0.27
357.1896	34.62	258.1549	0.34	198.0933	0.38
356.7313	0.34	256.1676	0.37	196.1684	0.27
356.5367	0.91	256.1409	0.21	195.122	0.34
356.462	0.24	255.1519	0.53	194.1101	0.27
356.3515	0.47	253.122	0.44	192.1014	0.27
356.1864	100	252.1664	0.34	188.1025	0.27
353.2348	0.6	249.125	0.34	186.0919	0.62
352.2314	0.61	242.1502	0.77	185.1262	0.41
339.1885	1.58	242.1242	0.89	183.1178	0.24
338.1757	4.23	234.1044	0.43	176.1389	0.83
334.2128	0.37	232.1296	0.34	169.0952	0.28
329.2053	1.02	230.1153	0.64	167.0881	0.34
328.1905	1.66	228.1032	2.24	160.0811	0.4

A2. 2. 5 MS/MS of Heronamide A from *Streptomyces sp.* CMB-0406 (Authentic Standard)

Precursor	466.2926							
Base Peak	m/z 356.1891 (Inten : 206,833)							
m/z	Rel. Inten.	Present in Standard	m/z	Rel. Inten.	Present in Standard	m/z	Rel. Inten.	Present in Standard
449.2903	3.78	TRUE	298.2231	0.59	FALSE	226.1294	0.50	TRUE
448.2901	18.31	TRUE	296.1677	0.32	TRUE	225.1314	1.01	TRUE
438.3097	0.59	TRUE	294.1854	0.47	TRUE	224.1135	1.49	TRUE
431.2975	0.32	TRUE	294.1516	0.32	TRUE	223.1507	0.53	TRUE
430.2800	3.91	TRUE	292.1751	1.11	TRUE	222.1466	0.47	TRUE
420.2946	1.89	TRUE	290.2526	0.55	FALSE	222.1080	0.35	TRUE
406.2747	0.47	FALSE	286.1849	0.47	TRUE	221.1364	0.75	TRUE
402.2895	0.7	TRUE	284.1644	0.99	TRUE	219.1271	0.36	TRUE
396.2630	0.59	FALSE	283.1907	0.35	FALSE	218.1312	0.89	TRUE
396.2134	0.43	TRUE	282.1851	0.94	TRUE	216.1082	1.63	TRUE
394.2807	0.55	FALSE	281.2098	0.32	FALSE	215.1288	0.47	FALSE
378.2446	0.36	TRUE	281.1889	0.62	FALSE	214.1247	0.46	TRUE
378.2083	0.47	TRUE	280.2170	0.64	FALSE	213.1139	0.55	TRUE
376.2673	0.47	TRUE	278.1915	0.41	TRUE	211.1130	0.64	TRUE
370.2380	1.32	TRUE	278.1531	0.55	TRUE	209.1356	0.58	TRUE
368.2286	0.53	FALSE	277.1660	0.35	FALSE	208.1275	0.58	TRUE
358.1907	0.95	TRUE	276.1540	0.50	FALSE	207.1217	0.76	TRUE
357.1927	24.99	TRUE	272.1436	0.35	FALSE	206.1207	1.72	TRUE
356.7383	0.84	TRUE	270.1935	0.47	TRUE	205.1036	0.5	FALSE
356.5415	1.01	TRUE	268.1766	0.35	TRUE	204.1246	0.58	TRUE
356.4637	0.67	TRUE	267.1672	0.35	FALSE	202.1183	0.80	FALSE
356.3501	0.95	TRUE	266.1553	1.85	TRUE	200.1127	1.08	TRUE
356.1891	100	TRUE	261.1657	0.72	FALSE	199.1067	0.56	TRUE
354.2449	0.67	FALSE	260.1357	0.35	TRUE	198.1259	0.35	TRUE
353.2165	0.88	TRUE	257.1537	0.63	FALSE	198.0984	0.47	TRUE
352.2259	1.41	TRUE	256.1396	0.91	TRUE	197.1228	1.02	FALSE
350.2301	1.12	FALSE	254.1217	0.32	TRUE	197.1030	1.13	FALSE
342.2052	0.83	FALSE	253.1955	0.3	FALSE	195.1243	1.53	TRUE
341.1914	0.73	FALSE	253.1658	0.41	FALSE	194.1557	0.88	FALSE
340.1947	8.48	TRUE	253.1292	0.56	TRUE	194.1046	0.32	TRUE
339.1868	0.47	TRUE	251.1426	0.32	FALSE	193.1058	1.59	TRUE
338.1797	5.75	TRUE	249.1264	1.67	TRUE	192.0966	0.47	TRUE
334.2112	1.33	TRUE	244.1380	0.41	FALSE	188.1113	0.89	TRUE
328.1942	2.15	TRUE	242.1593	1.83	TRUE	185.1027	0.7	TRUE
324.2088	0.73	TRUE	242.1180	0.61	TRUE	183.1213	1.35	TRUE
323.2163	0.32	FALSE	239.1373	0.53	TRUE	182.1114	0.38	TRUE
322.1800	3.73	TRUE	236.1522	0.35	FALSE	181.1021	1.46	FALSE
321.1855	0.32	FALSE	235.1514	0.6	FALSE	178.0773	0.3	TRUE
320.1652	1.76	TRUE	235.1176	0.90	TRUE	176.1538	0.35	FALSE
318.2484	1.14	FALSE	234.1144	1.18	TRUE	175.0894	0.64	FALSE
312.1732	0.82	TRUE	233.1300	1.12	FALSE	174.0960	1.01	FALSE
310.1895	1.83	TRUE	232.1351	0.41	TRUE	173.1347	0.36	TRUE
308.1662	0.35	FALSE	230.1512	0.58	FALSE	169.1067	0.38	TRUE
306.1919	0.42	TRUE	230.1204	1.82	TRUE	167.0948	0.32	TRUE
304.1855	0.82	FALSE	229.1076	0.64	TRUE	157.1053	0.35	TRUE
302.1815	2.30	TRUE	228.1301	0.83	TRUE	155.0959	0.32	FALSE
300.2316	2.28	FALSE	228.1038	2.96	TRUE			
300.1651	0.53	FALSE	227.1424	0.56	TRUE			

A2. 3. Spirohexenolide A

A2. 3. 1 MS/MS of Spirohexenolide A from *Streptomyces sp.* AS600 (Authentic Standard)

Precursor	409.2		
Base Peak	m/z 391.1933 (Inten : 22,120)		
m/z	Rel. Inten.	m/z	Rel. Inten.
392.1972	3.86	223.1442	1.65
391.1933	100	217.0907	2.76
373.183	13.5	211.1498	3.28
363.1979	6.31	207.1197	4.68
355.6893	1.66	197.134	3.58
327.1261	41.42	195.1099	4.97
309.1216	4.39	194.1042	2.2
308.1517	1.67	191.0409	4.69
307.1209	3.3	183.1136	1.66
297.1583	3.58	181.0996	1.67
293.1112	1.67	178.0869	1.67
283.1206	3.87	175.1475	5.24
271.1485	2.2	173.0319	1.65
271.0992	3.87	163.043	3.3
261.0903	2.2	163.0251	1.64
259.0846	1.65	161.0974	3.84
257.1413	2.21	159.1145	2.21
251.178	1.65	155.0809	1.65
250.1295	2.22	149.1333	5.51
243.0647	2.75	147.0773	19.4
237.1717	4.4	145.0605	1.66
237.1342	3.82	143.0899	2.21
233.0438	46.6	114.9079	2.2
225.09	2.17	77.7976	2.2

A2. 3. 2 MS/MS of Spirohexenolide A from *Streptomyces sp.* M-5455 pGP9-*spxR3*

Precursor	409.2015	
Base Peak	m/z 391.2095 (Inten : 9,728)	
m/z	Rel. Inten.	Present in Standard
391.2095	100	TRUE
250.1295	68.42	TRUE
231.063	68.42	TRUE
189.0816	68.42	TRUE
171.7337	65.79	FALSE
54.3547	65.79	FALSE