Characterising and engineering enzymes involved in antimycin and surugamide biosynthesis

Asif Fazal

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds
Faculty of Biological Sciences
School of Molecular and Cellular Biology

March 2021

Declaration

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Asif Fazal to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgements

Firstly, I would like to thank my supervisor, Dr Ryan Seipke, for his support and encouragement over the course of the PhD. A large number of things stand out, especially some character aspects I hope to take into my own scientific career, but I will always remember the super-early morning chats and the immense knack for a great turn of phrase. The humour and the gags weren't too bad either. Many thanks also to my secondary supervisor Mike Webb, as well as Glyn Hemsworth, for all their help at various points and in various different projects; their help was often key to the success of an experiment. I would also like to thank all current and previous members of the Seipke group, as well as the O' Neill and McDowall groups, for being a great group of people to be around. Particular thanks for Divya Thankachan and Dan Van, who were part of the group for the majority of my time in the lab, and who always made doing science that little bit more fun. Thanks also to the various facility and lab managers for their help and expertise in the use of, often new, techniques and machinery. Special mentions to the Back to the Future crew (Ryan and Dan; "you mean to tell me I got a mention in the acknowledgements"); the Nandos crew (Divya and Dan; all the table marking cockerel sticks); and the Flags of the Flagship (Amy, Sophie, Dave, James, Johnathan). Thanks also to the Wellcome Trust and the University of Leeds for funding.

I would also like to thank my family (Mum and Dad, Adil and Riaz) for being the absolute best foundation upon which everything else in my life is based. They were always there for help, support, advice; and a lot of times there was nothing better than going home for the weekend to just chill. Thanks as well to Nikita(aaaaa) for letting me be me, and putting up with endless ramblings about natural product biosynthesis and why it's so cool. Finally, I would like to give extra gratitude to God for providing the morals and beliefs in guiding me through the past years. Every struggle and tough time was made that bit easier knowing there's always an option if your heart is in the right place.

Abstract

Increased antimicrobial resistance combined with the lack of new antibiotics, is leading towards a return to a pre-antibiotic world and immense clinical strain. The vast majority of clinically utilised antibiotics are derived from secondary metabolites produced by *Streptomyces* species. Work into the discovery and biosynthesis of these secondary metabolites has undergone a recent resurgence, particularly due to the observation that most organisms harbour an increased number of biosynthetic gene clusters within their genomes than originally postulated. The often-novel enzymology involved in natural product biosynthesis has also generated interest, with understanding their enzymology a key milestone toward the goal of bioengineering novel natural products.

This work was centred around three projects based on the biosynthetic pathways producing the antimycins and surugamides. The antimycins are potent bioactive compounds, chemically composed of a central multi-substituted dilactone ring, bound to a rare 3-formamidosalicylate moiety. The biosynthetic steps involved in compound formation are relatively well characterised. In this work, one uncharacterised aspect of the biosynthesis, the function of a standalone ketoreductase, was studied in detail, with particular emphasis on how the structure of the protein mediates specific spatiotemporal catalysis. Engineering of an antimycin pathway acyltransferase domain, utilising a novel methodology, also gave insights into pathway functionality and potentially yielded novel antimycin-type compounds, demonstrating the utility of engineering workflows in the creation of new metabolites. The surugamides are non-ribosomal peptides, whose biosynthesis employs a novel strategy for chain termination and peptide cyclisation. In this work, SurE was found to be a standalone cyclase, and functional characterisation of the enzyme and its substrate scope gave insights into the nature of this versatile biocatalyst. The insights gained in understanding the enzymology of secondary metabolite biosynthesis, and the subsequent engineering of those enzymes, would be valuable in future attempts to create novel bioactive compounds.

Table of Contents

List of Tables

List of Figures

Abbreviations

x *g*	Times gravity (relative centrifugal force)
APS	Ammonium persulfate
aTC	Anhydrotetracycline
A domain	Adenylation domain
ACP	Acyl carrier protein
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
AMPA	3-amino-2-methylpropionic acid
AMR	Antimicrobial resistance
AT domain	Acyltransferase domain
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
BAC2H	Bacterial adenylate cyclase 2 hybrid
BGC	Biosynthetic gene cluster
C domain	Condensation domain
CAT	Chloramphenicol acetyltransferase
CCR	Crotonyl-coenzyme A carboxylase-reductase
CD	Circular dichroism
CPEC	Circular polymerase extension cloning
CoA	Coenzyme A
COMM	Communication mediating
CRISPR	Clustered regularly interspaced short palindromic repeats
C$_T$ domain	Condensation-like domain
CYP450	Cytochrome P450
DCM	Dichloromethane
dEB	Deoxyerythronolide B
DEBS	6-deoxyerythronolide B synthase
DH domain	Dehydratase domain
DHB	Dihydroxybenzoate
DIC	Diisopropylcarbodiimide
DIPEA	Diisopropylethylamine

DMF	*N,N*-Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
E domain	Epimerase domain
ECF	Extracytoplasmic function
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
ER domain	Enoylreductase domain
FAS	Fatty acid synthase
FLP	Flippase
FRT	Flippase recognition target
FSA	Formamidosalicylic acid
GABA	Gamma-aminobutyric acid
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HFIP	Hexafluoroisopropanol
HPLC	High performance liquid chromatography
HOBt	Hydroxybenzotriazole
HOCl	Hypochlorous acid
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β- D-1-thiogalactopyranoside
ISP2	International *Streptomyces* project medium 2
ITC	Isothermal titration calorimetry
$\boldsymbol{K_D}$	Dissociation constant
KR domain	Ketoreductase domain
KS domain	Ketosynthase domain
LB	Lennox broth
LC-HRESIMS	Liquid chromatography-high resolution electrospray ionisation mass spectrometry
LC-HRMS	Liquid chromatography-high resolution mass spectrometry
LC-MS	Liquid chromatography-mass spectrometry
mCoA	Malonyl-coenzyme A
MDR	Medium chain dehydrogenase/reductase

MeOH	Methanol
mmCoA	Methylmalonyl-coenzyme A
MT domain	Methyltransferase domain
m/z	Mass to charge ratio
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide, reduced form
NADP	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate, reduced form
NMR	Nuclear magnetic resonance
No-SCAR	Scarless Cas9 assisted recombineering
NRP	Non-ribosonal peptide
NRPS	Non-ribosomal peptide synthetase
NTA	Nitriloacetic acid
OD	Optical density
ONPG	*o*-nitrophenyl β-D galactopyranoside
Oxyma	Ethyl cyanohydroxyiminoacetate
PAC	P1-based artificial chromosome
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PEGA	Acrylamide polyethylene glycol
PK	Polyketide
PKS	Polyketide synthase
PLP	Pyridoxal 5'-phosphate
Ppant	4'-phosphopantetheinyl
PP$_i$	Inorganic pyrophosphate
PPI	Protein-protein interaction
PPTase	Phosphopantetheinyl transferase
PyBOP	Benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate
R domain	Reductase domain

RiPP	Ribosomally encoded and post-translationally modified peptides
RNA	Ribonucleic acid
RNAse A	Bovine pancreatic ribonuclease A
rpm	Revolutions per minute
SAM	*S*-adenosylmethionine
SDR	Short chain dehydrogenase/reductase
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SFM	Soy flour mannitol
sgRNA	Single guide RNA
SNAC	N-acetylcysteamine
SPE	Solid phase extraction
SPPS	Solid-phase peptide synthesis
T domain	Thiolation domain
TAE	Tris acetate ethylenediaminetetraacetic acid
TAR	Transformation associated recombination
TE domain	Thioesterase domain
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TIPS	Triisopropylsilane
TSB	Tryptic soy broth
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
WT	Wildtype
YCC	Acyl-coenzyme A carboxylase

Chapter 1 – Introduction

1.1 Secondary metabolites as a source of varied drugs

The discovery, synthesis, and formulation of compounds as drugs targeting specific cellular and molecular constructs is a historically important achievement, with the medical, monetary, and societal value of such compounds becoming increasingly high. The use of antibiotics has mediated the immediate successful treatment of previously dangerous, disease-causing microorganisms, and therefore is the most successful form of chemotherapeutic treatment used against the diseases that affect humans (Katz and Baltz, 2016; Gould, 2016). The clinical and prophylactic availability of antibiotics has led to dramatic reductions in rates of mortality from diseases caused by common bacterial species, some of which now pose little threat to human life due to the presence of effective treatment options (Ventola, 2015).

Antibiotics came to the forefront mainly during the golden era of antibiotic discovery (1930s – 1950s) when the vast majority of clinically important antibiotics were discovered (Davies, 2006). Although antibiotics were used previously, the discovery of penicillin laid the foundation for their widespread use in treating bacterial infection (Hutchings et al., 2019). Previous to the widespread production of penicillin, Prontosil was readily used as a chemotherapeutic agent targeting Gram-positive bacteria. Prontosil owed its discovery to the groundwork carried out by Paul Ehrlich in the early 1900s (Bosch and Rosich, 2008). Ehrlich, based on the observation of the binding of synthetic dyes to specific microorganisms, believed in the idea of a 'magic bullet' which could theoretically target disease causing bacteria, whilst leaving the host unaffected. Ehrlich subsequently carried out a large-scale screen of synthetic dye-like compounds against the syphilis causing spirochete *Treponema pallidium,* leading to the discovery of Salvarsan in 1909 (Gelpi et al., 2015). Later, in 1928, Alexander Fleming noted a strain of *Penicillium notatum* was able to inhibit the growth of *Staphylococcus aureus* by production and secretion of a lytic compound, which he named penicillin (Fleming, 1929). Purification of the active compound, however, proved to be a difficult and arduous task, finally accomplished between 1938 and 1940 by Howard Florey and Ernst Chain, with scale up procedures and access to copious amounts of penicillin achieved by 1945 (Abraham et al., 1941). The vast majority of antibiotic classes discovered during the golden era, however, were not created synthetically, but by discovery and extraction of bioactive compounds from actinomycetes. Selman Waksman, a soil microbiologist, believed that actinomycete bacteria

held particular promise as producers of bioactive molecules due to their ability to survive and thrive in habitats and niches where pathogenic bacteria were generally prevalent (Waksman et al., 2010). Waksman subsequently instigated a large-scale bioactivity-based screening programme where actinomycetes were tested for their ability to inhibit the growth of Gram-negative bacteria. This led to the discovery of streptomycin in 1943, an aminoglycoside drug active against *Mycobacterium tuberculosis*. Due to the unexplored nature of many soil-dwelling bacterial species, as well as the use of high-throughput screening platforms, the discovery of the antibacterial classes; β-lactams, tetracyclines, aminoglycosides, and glycopeptides (Figure 1.1), rapidly followed the discovery of the initial effective agents (Aminov, 2010). The techniques developed and utilised during the golden era, particularly screening platforms, are still used presently, and the difficulties faced by past pioneering researchers, including the discovery, extraction, and purification of novel compounds, still plague antibiotic discovery (Silver, 2011).
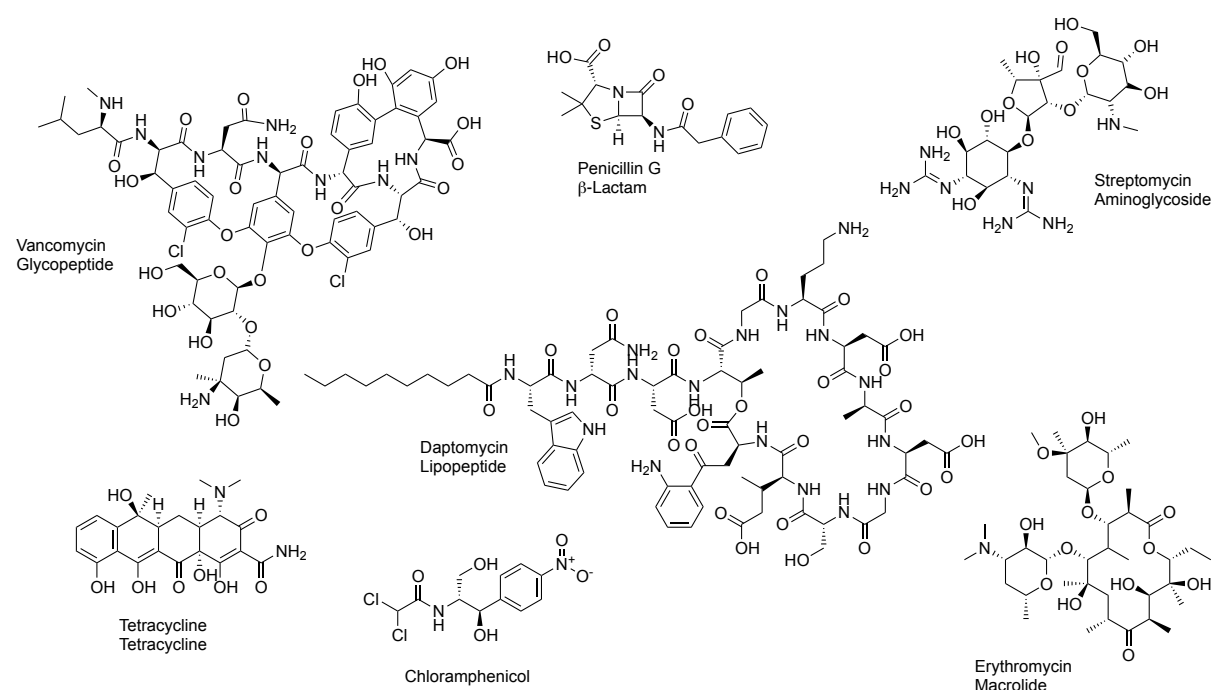
Figure 1.1: A selection of natural products produced by bacterial or fungal species and used clinically, either presently or historically, as antibiotics. The structural class to which each compound belongs is noted below the name of individual compounds.

The difficulties faced in the pursuit of novel antibacterial compounds in the modern era extend to the rediscovery of known, active compounds, and more worryingly, the rampant spread of antimicrobial resistance (AMR) (Kapoor et al., 2017; Zaman et al., 2017). Bacteria resistant to clinically used antibiotics were first identified in the late 1930s, quickly after the clinical

introduction of sulfonamide antibiotics (Prescott, 2014). The presence of penicillin degrading enzymes, penicillinases or β-lactamases, within certain bacterial species was confirmed in 1940, before the widespread use of penicillin as a therapeutic, corroborating the fact that resistance mechanisms to antibacterial agents exist naturally within bacterial populations (Kong et al., 2010; Bush, 2018). Streptomycin resistant *Mycobacterium tuberculosis* strains were identified soon after the introduction of the antibiotic, and often in patients during courses of treatment. Resistance mechanisms to the vast majority of clinically used antibiotic drug classes used today have arisen in swathes of bacterial populations, including the commonly used β-lactams, glycopeptides, macrolides, rifamycins, and lipopetides, amongst others (Munita and Arias, 2016). These resistance mechanisms, ranging from cellular target alteration (such as modification or protection) to antibiotic modification (such as hydrolysis, chemical adjustment, or efflux) become more widespread upon increased use of the antibacterial agent, necessitating altered courses of treatment and, ideally, the discovery and use of novel antibiotics with reduced or no detectable levels of resistance. AMR is predicted to cause increased numbers of deaths per year, potentially reaching ~ 10 million deaths per year from 2050 onwards (O' Neill, 2016). Therefore, the discovery of novel antibiotics with new modes of action and new cellular targets, is a key research focus in continuing efforts to combat the adverse effects caused by bacterial infection.

During the golden era of antibiotic discovery, the majority of discovered bioactive antimicrobial agents were isolated from bacteria, or other *in vivo* sources such as fungi. These natural products have continued to provide a rich source of clinically valuable drugs targeting a variety of disease-causing organisms or cellular targets, and have continually been mined for their therapeutic effects (Newman and Cragg, 2016), with the search for novel natural products coming full circle after the problems faced in the rediscovery of known molecules and the overall failure of combinatorial chemistry platforms in yielding active and effective compounds (Ribeiro da Cunha et al., 2019). Sequencing of the first *Streptomyces* genome, that of *Streptomyces coelicolor* A3(2), opened up new avenues of genome mining due to the presence of a larger number of biosynthetic gene clusters than previously known and linked to the production of bioactive compounds (Bentley et al., 2002).

The naturally produced and isolated molecules are also termed secondary metabolites, and are generally thought to be non-essential for the ultimate survival of the producing organism, but

confer that same organism with definite survival advantages (van der Meij et al., 2017). These selective advantages range primarily from increasing adaptation to a niche, such as the use of siderophores to scavenge iron, to providing defensive or offensive compounds to reduce competition, such as the production of antimicrobial compounds. The specificity of such secondary metabolites makes the compounds extremely valuable in a medical and clinical sense, with the range of bioactivities observed also not limited to antibiosis. Remarkably, isolated bacterial natural products have also been clinically used as anticancer drugs (such as daunorubicin) and immunosuppressants (such as tacrolimus) amongst other antimicrobial compounds (such as insecticides and fungicides) (Newman and Cragg, 2016). The relatively widespread nature and varied bioactivities of secondary metabolites, as well as the large presence of silent clusters within bacterial genomes, has led to a resurgence in their interest and subsequently in efforts to find novel natural products, active against a wide range of medically important targets.

1.2 Actinobacteria

Actinobacteria represent one of the largest phyla within the Bacterial kingdom, and are generally filamentous, Gram-positive bacteria, with high genomic GC content (>70%). Habitats occupied by the Actinobacteria range vastly from soil environments, where they are generally ubiquitous, to marine or much more arid environments (Barka et al., 2016). The majority of Actinobacteria have typical Gram-positive cell structures, with a cell membrane and thick external peptidoglycan layer, and are also generally unicellular, branching organisms with relatively distinct growth phases. The bacterial species falling into the phylum are, however, best known for their prolific production of secondary metabolites or natural products, which have been found to have an extremely wide range of uses (Manivasagan et al., 2014). Isolated and purified metabolites from the bacteria have been externally and clinically used as antibiotics, antifungals, antivirals, antiparasitic drugs, anticancer drugs, immunosuppressive therapies, insecticides, and herbicides, as well as, as pigments, probiotics, and biosurfactants. The phylum contains many genera, including *Streptomyces*, *Salinispora*, and *Saccharopolyspora*, amongst many others, which are relatively widely studied, predominantly due to their production of natural products (Baltz, 2008). In total, Actinobacteria are responsible for the initial production and isolation of over two thirds of the clinically relevant antibiotics used today, and the looming threat of multi-drug resistant bacterial strains has led

to a resurgence in the mining of this talented phylum, for the discovery of unexplored niches and novel compounds (Clardy et al., 2006).

1.3 *Streptomyces*

Streptomyces are the largest genus of Actinobacteria comprising over 500 individual species and are well known for their ubiquity in soil environments. The bacteria are sporulating saprophytes and are therefore heavily involved in nutrient decomposition and carbon and nitrogen cycling in their respective niche. The non-motility of *Streptomyces* means they are exposed to a variety of environmental conditions, some of which are non-favourable or particularly harmful (Bobek et al., 2017). The ability of *Streptomyces* bacteria to survive and thrive in both arid and wet conditions, such as desert and marine environments, is testament to their ubiquity, adaptability, and genomic plasticity (Chater, 2016). The most notable aspect of *Streptomyces* biology, however, is the widespread production of secondary metabolites, not only important for self-preservation and survival, but which also provides an incredibly important source of bioactive compounds for clinical use.

Streptomyces generally have a complex developmental life cycle (Figure 1.2), beginning with a free spore germinating under favourable nutrient and environmental conditions to yield germ tubes (Flärdh and Buttner, 2009; Jones and Elliot, 2018). The germ tubes grow by tip extension and hyphal branching to form multigenomic hyphae, beneath the level of the solid medium upon which the bacteria are actively growing. These vegetative or substrate mycelia also release enzymes such as cellulases, which aid in nutrient access and scavenging by breakdown of soil substituents. Due to the wide variety of niches inhabited by *Streptomyces* species, they have evolved to respond quickly to changes in environmental conditions; for example, in response to nutrient deprivation, as well as other stress signals, *Streptomyces* switch to a distinct phase of aerial growth (Kelemen et al., 2001). Unigenomic hyphae, capable of breaking the surface tension, grow into the air and initially form non-branching spore compartments. Spore chains then septate and form individual spores, which at the end of the cycle are capable of dispersal to new environments, and re-initiation of the growth cycle. Accompanying the switch in morphological development, triggered by sensing of stress indicators, is the production of secondary metabolites, such as antibiotics, which serve to reduce the bacterial competition present within the environment.
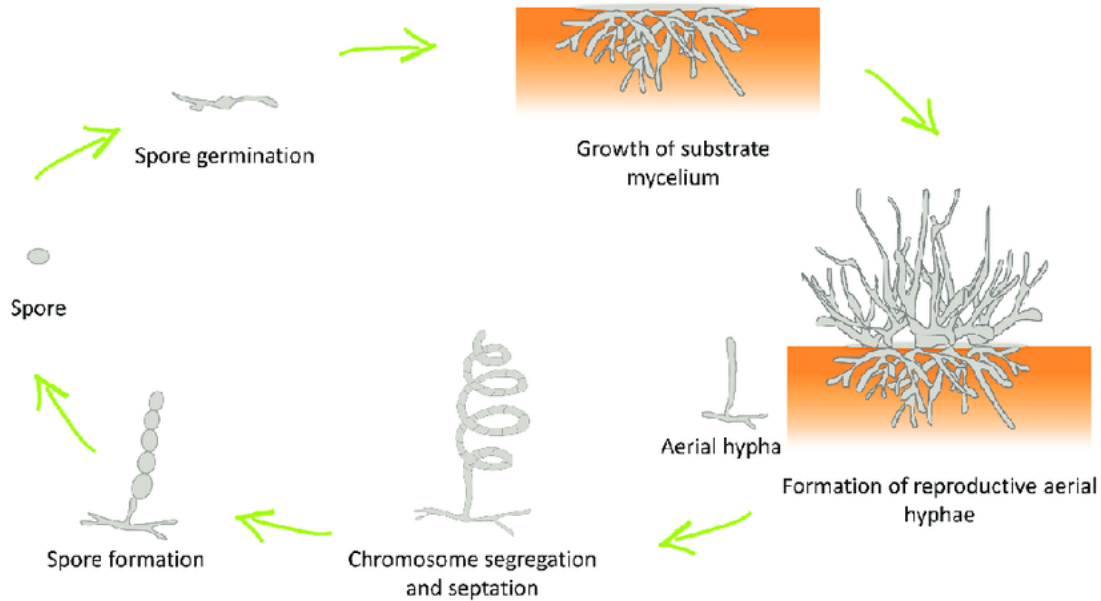
Figure 1.2: Schematic representation of the life cycle of a typical streptomycete (Law et al., 2019). Free spores germinate under favourable conditions, subsequently forming germ tubes and extensive substrate mycelia. Upon the sensing of certain signals, morphological differentiation is induced and aerial hyphae are produced. Aerial hyphae then septate to form long chains of spores, which are released and dispersed to start new growth cycles.

In natural environments, *Streptomyces* produce defence compounds, such as antibiotics, and other secondary metabolites, likely to ward off and outcompete other species present within the local environment. Their production is usually temporally correlated with the growth of aerial hyphae and the concomitant breakdown of the substrate mycelium (Bibb, 2005). Therefore, the roles of primary metabolism, morphogenetic differentiation, and secondary metabolism are intricately linked and make important and obvious contributions to the successful growth and completion of the growth cycle. Genomic mapping studies, initiated in the 1950s, followed by multiple gene cloning experiments in the 1980s and 1990s, culminated in the complete sequencing of the *S. coelicolor* genome in 2002, which showed the nature of many of the specialised metabolite production genes (Hopwood, 1967; Bentley et al., 2002). The complete sequence showed the genome contained an unprecedented number of regulatory genes, important for organism-level responses to external stimuli. Many of these regulatory systems also control the production of secondary metabolites through control of the production of the protein systems that synthesise the bioactive compound. Genome analysis revealed that many of the genes involved in the production of a natural product are clustered together, often

in extremely large biosynthetic gene clusters (BGCs), which themselves contain multiple individually regulated operons. Many of the gene clusters were also observed to contain genes typically involved in primary metabolism, indicating the intricate link between primary and secondary metabolic processes.

The sequencing of the *Streptomyces coelicolor* genome, and the availability of data generated from sequencing studies of other *Streptomyces* species, has intensified and rapidly increased the search for clusters directing the synthesis of novel bioactive compounds, or new natural product classes. The number of distinct classes of secondary metabolites biosynthesised by Actinobacteria is extremely varied, and the compound members of each class show similarly high levels of diversity and wide usage of chemical space (Dobson, 2004; Lachance et al., 2012). Natural product classes range from large, individual, and free standing, assembly line-like systems, such as non-ribosomal peptide synthetases and polyketide synthases, to systems that explicitly require other cellular machinery, such as the ribosomally encoded and post-translationally modified peptides (RiPPs) which, as the name suggests, necessitate peptide synthesis by the ribosome before formation of the final compound (Scott and Piel, 2019). Nonetheless, the multitude of secondary metabolites produced by *Streptomyces* have made large impacts on modern medicine, agriculture, and other endeavours, and have combined to contribute greatly to human lifestyle. The two most populous and well-studied classes of produced natural products, the non-ribosomal peptides and polyketides, are discussed in detail below.

1.4 Non-ribosomal peptide synthetases

1.4.1 Biosynthetic gene clusters of non-ribosomal peptide synthetases

Non-ribosomal peptide synthetases are large assembly line-like, multienzyme machineries, with the biosynthetic capability of producing a wide variety of structurally diverse compounds (Finking and Marahiel, 2004). Ribosomal synthesis of peptides and proteins is an essential process and mediates the proper function of every cellular pathway, including metabolic and biosynthetic processes that are reliant upon ribosomally synthesised proteins. Various other methods for the creation and synthesis of peptide or amide bonds are used within cells, such as protein or peptide ligase enzymes, however, the major biosynthetic strategy for bioactive peptide creation in bacteria and fungi is through the action of non-ribosomal peptide synthetases (NRPSs) (Sieber and Marahiel, 2003). The assembly line synthetic rationale

utilised by NRPSs allows the biosynthetic machinery to operate with less restrictions than other major cellular pathways, observable in the structural diversity of the final compounds produced.
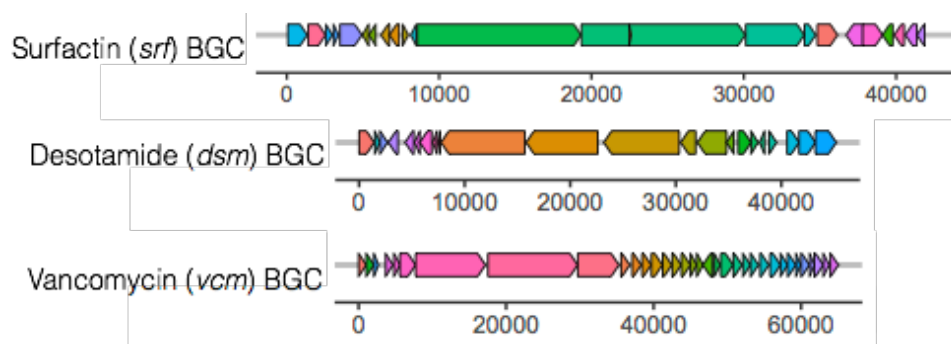
Figure 1.3: Examples of NRPS BGCs, including the clusters responsible for the production of the antibacterial biosurfactant surfactin; the antibacterial cyclic peptide desotamide; and the potent glycopeptide antibiotic vancomycin. Each arrow represents a gene in the proposed cluster, whose product contributes to the formation of the final compound. The x-axis scale represents the length of each BGC in DNA base pairs.

The totality of the enzymes required for complete compound production are predominantly encoded by genes held within a contiguous region of the genome, termed a biosynthetic gene cluster (BGC). Examples of NRPS BGCs are shown in Figure 1.3. The nature and length of NRPS BGCs varies widely between clusters, with the genes held within a BGC dependent upon the synthesised final compound; for example, an NRPS producing a toxic peptide would require a self-resistance mechanism, which would usually be a protein, separately encoded within the specialised metabolite BGC (Yan et al., 2020). As NRPSs are arranged as relatively lengthy assembly lines, the proteins forming the synthetase are also large, and usually composed of multiple, physically connected, enzymes. The synthetases themselves, however, are often not functional in the creation of mature peptides in a cellular environment, and require the function of a variety of other proteins, such as precursor-forming enzymes and phosphopantetheinyl transferases for the conversion of carrier proteins from *apo* to *holo* form, as well as other enzymes that actively contribute to biosynthetic processes, either upstream or downstream of peptide synthesis (Süssmuth and Mainz, 2017).

Since the discovery of the existence of non-ribosomal peptides (NRPs) in the 1960s (Mach et al., 1963), and the subsequent observation of the organisation of the required genes in BGCs,

a lot of study time has been devoted into deciphering the functionality of the pathways that synthesise the chemically varied compounds (Felnagle et al., 2008). A well-studied NRPS system is the pathway that synthesises the surfactins, lipopetides that have been observed to have a range of clinically and commercially useful bioactivities (Kluge et al., 1988; Peypoux et al., 1999). Surfactins A-D, which are distinguished by the length of the fatty acid chain ranging from seven to ten carbon units in the alkyl chain, are synthesised by an NRPS system comprised of three multi-modular polypeptides, therefore encoded by three large genes in the BGC; *srfA-A*, *srfA-B*, *srfA-C*. Other genes within the comparatively short BGC are important in initiation of lipopeptide synthesis and priming of enzymatic domains, in order to enable efficient biosynthesis to occur (Steller et al., 2004; Kraas et al., 2010). In any case, the genes held within a cluster have evolved to be efficiently coregulated and to produce proteins as spatiotemporally required by the assembly line system, so that bioactive compound formation occurs efficiently within the cell.

1.4.2 Modular assembly line logic of NRPS biosynthesis

The assembly line-like biosynthetic logic utilised by NRPS systems lends itself to proteins having relatively defined and regimented enzymatic roles. This is visible in the organisation of NRPS pathways into distinct modules, each composed of multiple catalytically active domains (Figure 1.4). Modules are collections of two or more individual, catalytic domains, which each carry out an enzymatic reaction in the synthesis of the mature peptide (Schwarzer et al., 2003). The biosynthetic modules are responsible for the addition of a single amino acid onto the growing peptide, and are predominantly composed of the condensation (C), adenylation (A), and peptidyl carrier protein (PCP) domains (often termed thiolation or T domains). In the usage of modules to synthesise peptides, NRPS systems obey the collinearity rule, which states that, for example, a hexapeptide mature compound produced by an NRPS system requires six modules in order to form the peptide (Fischbach and Walsh, 2006). In comparison to ribosomal synthesis of peptides and proteins, the flexibility of modules in permitting the inclusion and recruitment of alternative or extra catalytic domains, allows disparate chemical reactions to occur leading to the biosynthesis of peptides with varied functional groups.
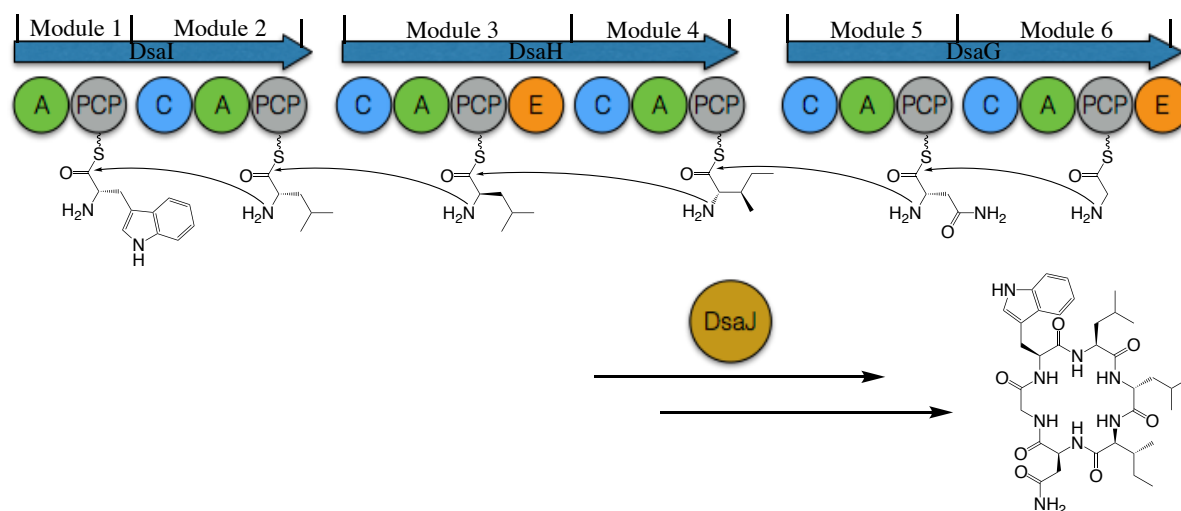
Figure 1.4: The desotamide pathway as an illustration of the modularity of type A NRPS systems. The pathway is composed of six modules, the first serving as the loading module and the rest acting as extension modules, each responsible for the addition of one amino acid. Modules are formed minimally of condensation (carry out condensation reaction and join upstream intermediate with loaded amino acid), adenylation (select and activate amino acid substrates), and peptidyl carrier protein (act as proteinaceous tethers for amino acids and substrates) domains.

Although the vast majority of known and well-studied NRPS systems use a modular biosynthetic logic to synthesise diverse peptides, often termed as type A NRPS systems, other pathways exist, which utilise variations of the canonical, collinear modular system. An overview of the three major types of NRPS systems is shown in Figure 1.5. The products of type B NRPS systems, such as enterobactin, gramicidin S, and the quinoxaline antibiotics, were observed to regularly display two- or three-fold symmetry in the final compounds, predominantly because the mature products were formed from two or three copies of a single peptide, joined by a macrocyclisation reaction (Hoyer et al., 2007). The biosynthesis of these peptides is made possible by the action of type B NRPS pathways which utilise the modular system for the synthesis of NRPs, but disobey the collinearity rule by iterative usage of the set of modules comprising the system in order to form further identical peptides, whilst the primarily synthesised peptide remains attached to the terminal module, usually covalently bound to a thioesterase (TE) domain (Jaremko et al., 2020). Upon completion of the next round of peptide synthesis, the two identical peptides are macrocyclised and either released to form the final compound, or covalently retained for the synthesis and adjoining of further peptides to the intermediate. For example, in the biosynthesis of enterobactin (Figure 1.5), 2,3-

dihydroxybenzoate (2,3-DHB) is synthesised and loaded onto the first module by gene cluster-encoded enzymes, followed by peptide bond formation with an activated Ser by the second module. The dipeptide is held on a catalytic residue of the TE domain, therefore allowing further peptide synthesis by the preceding modules, which synthesise two further dipeptides before macrocyclisation yielding the catecholamide and trilactone ring-containing siderophore (Gehring et al., 1997). Type C NRPSs, such as those that synthesise the potent antibacterial drug capreomycin and the anticancer bleomycins, are non-linear NRPSs that can also have varied module organisations and utilise certain catalytic domains multiple times. Multiple modules within type C pathways can be unorthodox, by missing normally required domains or containing altered or unusual domains (Keating et al., 2000; Barkei et al., 2009). Generally, however, type C NRPS systems are less studied and require extended experimental effort to determine the functionality of the pathway due to the peculiarities driving secondary metabolite biosynthesis.
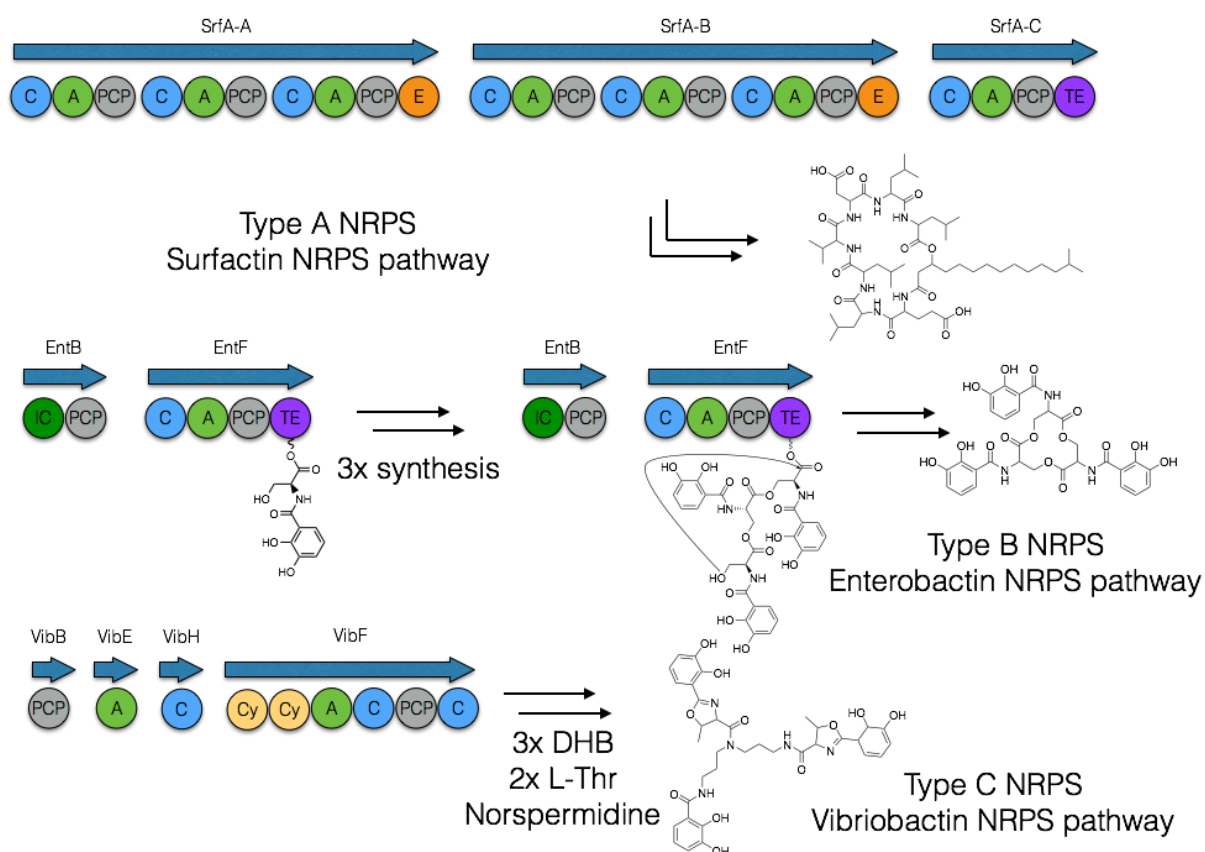
Figure 1.5: Examples of the three different NRPS biosynthetic logics employed by systems to synthesise complex, bioactive molecules.

1.4.3 Core catalytic domains in NRP biosynthesis

1.4.3.1 Adenylation (A) domains

The first step in NRP biosynthesis is the selection and activation of amino acid substrates, followed by the covalent tethering of residues to the NRPS assembly line. This essential function is carried out by adenylation (A) domains using a distinct two-step mechanism, as illustrated in Figure 1.6. Initially, the required amino acid is selected from the cellular pool and subsequently adenylated by a catalytic reaction with adenosine triphosphate (ATP). The adenylation reaction occurs at the α-carboxy group of the amino acid, and results in the release of inorganic pyrophosphate (PPi) and the formation of an aminoacyl-AMP, which is held in the active site of the A domain to avoid unwanted hydrolysis of the phosphodiester bond and breakdown of the activated conjugate (Walsh, 2016). The A domain can then undergo a conformational alteration, allowing the 4'-phosphopantetheinyl (Ppant) arm, covalently attached to the peptidyl carrier protein (PCP) domain, to enter the active site, a process which is followed by catalytic loading of the amino acid onto the Ppant arm (Quadri et al., 1998). This reaction completes the catalytic cycle of the A domain and results in release of adenosine monophosphate (AMP), liberated from the aminoacyl-AMP conjugate, and conversion of the *holo*-PCP domain to the amino acid loaded form.
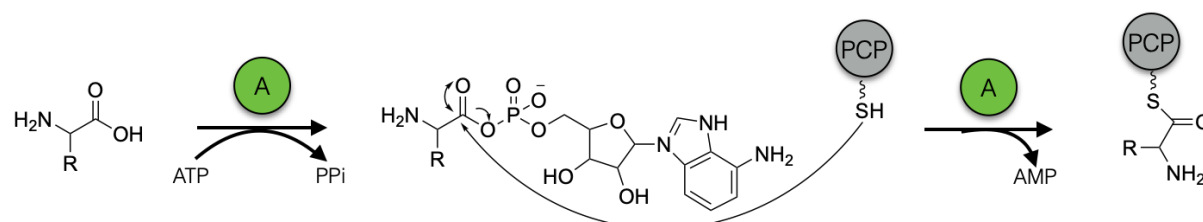
Figure 1.6: The basic catalytic cycle of the module encoded adenylation domains within NRPS pathways. ATP is used to activate the selected amino acids, followed by controlled hydrolysis of the phosphodiester bond and loading of the downstream PCP domain.

A domains, much like the functionality of total NRPS systems, employ a structurally and dynamically exquisite mechanism, and one that allows efficient catalysis with minimal chances of wasted enzymatic activity. Most A domains are ~ 60 kDa in size and are composed of two distinct subdomains; a much larger core domain and a relatively small C-terminal subdomain. The core domain contains the catalytic residues, however, both subdomains are important in substrate recognition with residues lining the amino acid binding pocket lying at the interface of the two subdomains (Challis et al., 2000). Initial studies into the amino acid specificity

conferring code of A domains, carried out on the Phe activating domain of the linear gramicidin synthetase pathway (Conti et al., 1997; Stachelhaus et al., 1999), revealed the presence of ten amino acid residues surrounding the bound amino acid substrate in the crystal structure, which play roles of differing importance in the selection of varying amino acids. For example, the lack of a side chain in Gly leads to less penetration of the A domain binding pocket and larger residues are present at the front of the binding pocket to stabilise the selected amino acid. Whereas the opposite is true for the selection of larger amino acids, such as those with aromatic side chains. The selection and substrate activation carried out by A domains plays an important role in the structure and composition of the final natural product produced, therefore studies have focussed on providing detail on its mode of action, with an eye to aid compound bioengineering (Stanišić and Kries, 2019). A number of these studies have focussed on the structural mechanism employed by A domains in the adenylation and PCP domain loading reactions. Structural studies of domains, didomains, and complete modules revealed the dynamic nature of NRPS modules, particularly in the activation and loading half-reactions (Tanovic et al., 2008; Marahiel, 2016). Upon adenylation of a selected amino acid by the A domain, a large conformational change in the smaller subdomain leads to a change in mode of catalysis, by translocation of the catalytic Lys involved in adenylation, and allows the lengthy Ppant arm of the PCP domain, through a concerted conformational change of its own, to enter the active site of the A domain and undergo the loading reaction (Figure 1.7). This extremely dynamic mechanism is seemingly typical of A domains in NRPS systems, and highlights the evolved nature of NRPS systems in achieving functionality and efficiency in non-ribosomal peptide formation (Reimer et al., 2019).

A

B
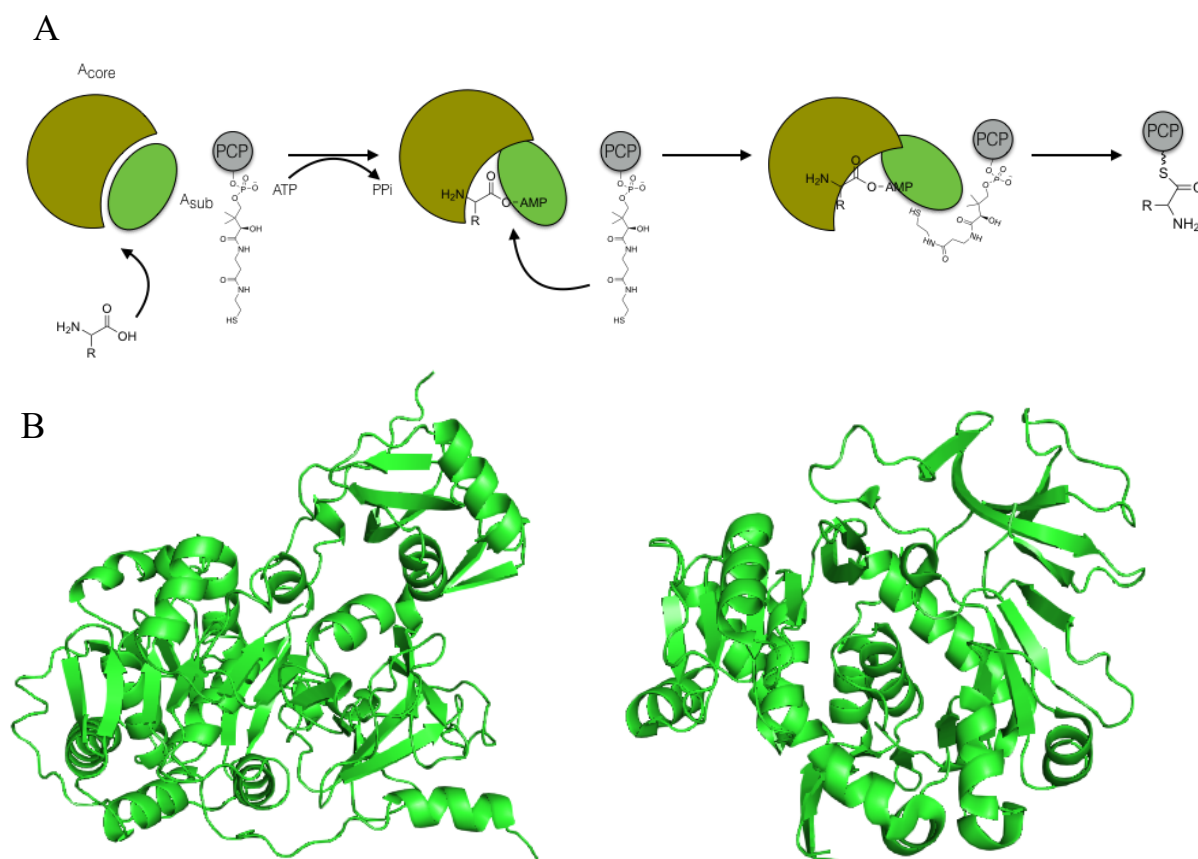
Figure 1.7: A) An overview of the dynamic mechanism employed by NRPS A domains in the selection and activation of amino acids. Concerted conformational changes in both A_{core} and A_{sub} domains allow amino acid entry and adenylation, as well as entry of the 4'-Ppant extension of the PCP domain and its subsequent loading. B) Solved structures of NRPS A domains PheA (PDB ID: 1AMU) and SidNA3 (PDB ID: 3ITE), illustrating the N-terminal and C-terminal domains (upper right of structures) (Conti et al., 1997; Lee et al., 2010).

A domains, however, are not limited to only carrying out adenylation and loading within the context of being embedded within an assembly line. Whilst the catalytic activity remains similar, examples of standalone A domains have been observed, such as in the vibriobactin biosynthetic pathway, and A domains activating alternative substrates, other than amino acids, also naturally exist (Magarvey, Ehling-Schulz, et al., 2006). Assembly line-embedded A domains can also play a role in the recruitment of other proteins to the NRPS assembly line, either to increase the efficiency of peptide synthesis or add structural diversity to the chemical scaffolds produced. The most commonly recruited proteins are the MbtH-like proteins, which function in the stabilisation of A domains, whilst also allosterically increasing the catalytic efficiency at which adenylation and loading reactions occur (Felnagle et al., 2010). A domains, therefore, have a wider role in the biosynthesis of NRPs, but their role as core NRPS domains

in the selection and activation of amino acid substrates is pivotal in the efficient synthesis of NRPs.

1.4.3.2 Peptidyl carrier protein (PCP) domains

PCP domains, with a mass of ca. 9 – 10 kDa, are the smallest domains regularly involved in synthesis of core NRP chemical scaffolds. The domains act as protein tethers to which intermediates and substrates are covalently attached, primarily to avoid dissolution of the synthesised compound and also to provide ease and directionality to the synthetic procedure. PCP domains are post-translationally modified by a phosphopantetheinyl transferase (PPTase) enzyme, which covalently attaches the ~ 18 Å Ppant arm onto a conserved Ser residue to yield a *holo* PCP domain, capable of thioester bond formation and amino acid loading by the upstream A domain (Figure 1.8) (Beld et al., 2013). The catalytic cycle of an NRPS elongation module revolves around substrate binding to the PCP domain. Initially, the PCP domain is loaded with a specific amino acid by the A domain, after which the PCP domain enters the active site of the upstream condensation (C) domain. The C domain catalyses a condensation reaction between the upstream peptidic intermediate, held on the PCP domain from the previous module, and the newly donated amino acid to result in an elongated peptide bound to the PCP and completion of elongation for the NRPS module (Reimer et al., 2018).
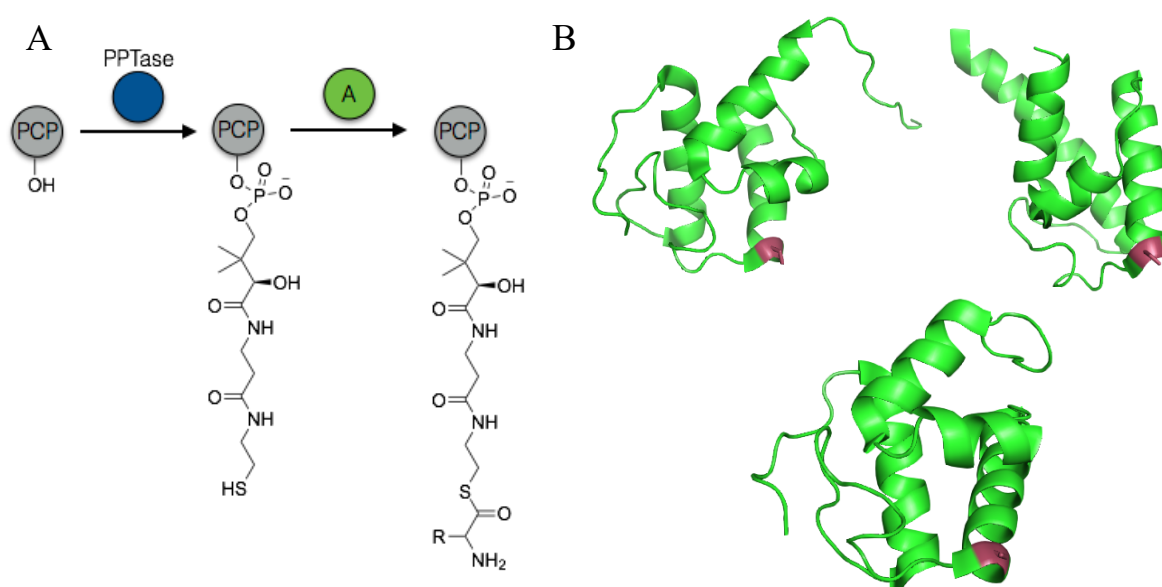
Figure 1.8: A) Overview of the transformation of PCP domains from *apo* to *holo* to amino acid loaded forms. The attachment of the 4'-Ppant group to a conserved PCP domain serine residue is obligately required for amino acid loading and peptide formation. B) Solved structures of

PCP domains; clockwise from top-left: Structure of PCP7tei, the final PCP domain in the teicoplanin biosynthetic system (PDB ID: 2MR7) (Haslinger et al., 2015). Structure of BlmI, a typeII PCP from the bleomycin biosynthetic pathway (PDB ID; 4NEO) (Lohman et al., 2014). Structure of the PCP from module 5 of tyrocidine synthetase (PDB ID: 2JGP) (Samel et al., 2007). The location of the conserved Ser residue, the site of Ppant attachment, is indicated by maroon colouring.

PCP domains, through their inherent role in biosynthesis, are incredibly conformationally dynamic. This is especially true for the *holo* or loaded form when the lengthy prosthetic arm is attached and capable of extending into the catalytic centres of nearby domains. The ability of the PCP domain to extend the attached Ppant arm into the active site of upstream A and C domains, as well as the downstream C domain, is well understood and mechanistically accepted (Strieker et al., 2010; Tarry et al., 2017). However, the capacity and potential capability of PCP domains to select a specific catalytic centre at a specific time point within the biosynthetic cycle remains poorly understood, and the successful delivery of substrates and intermediates to the required domains has been attributed to a combination of domain movement and competitive binding. Structural studies have revealed the multiple conformations adopted by PCP domains in order to deliver biosynthetic intermediates to the required domains, however, more biochemical studies are required in order to fully understand the complex mechanistic roles played these core domains in the biosynthesis of structurally complex NRPs.

1.4.3.3 Condensation (C) domains

C domains are the core enzymatic component involved in assembly and biosynthesis of NRP intermediates. As stated titularly, C domains catalyse a condensation reaction between a donor peptide, tethered to an upstream PCP domain, and an acceptor amino acid, covalently attached to the downstream PCP (Figure 1.9). The condensation reaction results in amide bond formation and extension of the peptidic intermediate, with concomitant transfer of the formed peptide from the upstream to the downstream PCP domain and release of a water molecule (Bloudoff and Schmeing, 2017). C domains form a large part of the core NRPS module and usually have a mass of ~ 50 kDa, as well as a core catalytic motif of HHxxxDG. The second His residue of the motif has been shown to be important in condensation, and mutational analysis of the residues forming the motif showed that alteration of the His residue disrupts peptide formation (Stachelhaus et al., 1998). The complete mechanism of the catalysed

condensation reaction is not known and various roles for the conserved residues have been postulated. Despite the catalytic motif being highly conserved, the residues could play varied roles depending on the specific nature of the C domain, ranging from importance in structural stability to active chemical roles in catalysis.
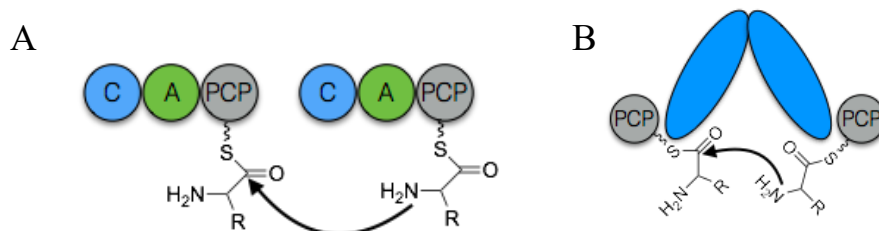
Figure 1.9: A) Mechanism of condensation domain mediated action in the extension of peptide intermediates. B) An illustration of the V-shaped, two-lobed architecture of condensation domains and the interaction with upstream and downstream PCPs. The amine of the acceptor amino acid attacks the carbonyl of the donor peptide, leading to the translocation of the peptide onto the downstream PCP.

Structural studies of C domains have demonstrated the intricate nature of C domain mediated catalysis and highlighted the structural features which allow condensation reactions between two protein tethered substrates to occur. As C domains interact with multiple different protein partners, including downstream A and PCP domains, as well as an upstream PCP domain, deconvoluting the structural basis for all of these interactions remains difficult, although models based on crystal structures have shed light onto the catalytic functionality of C domains (Keating et al., 2002; Bloudoff et al., 2013). The domains are pseudodimeric and composed of two subdomains, N-terminal and C-terminal, whilst the subdomains each resemble the chloramphenicol acetyltransferase (CAT) type proteins (Figure 1.10). The proteins form conventional 'V' shapes, with the conserved catalytic motif residing close to the interface between the two parts of the protein. The structure and shape of the protein allows approach from both sides by donor and acceptor PCP domains, and both substrates, via the Ppant extensions, can enter the active site and undergo condensation and peptide extension, yielding the subsequently required and translocated peptide intermediate (Figure 1.9). Early studies showed the potential for condensation domains to provide an extra layer of substrate selection, beyond the initial A domain mediated amino acid selection, by catalysis of specific substrates, and only those required in the biosynthesis of the specific NRP. However, certain C domains have been shown to be promiscuous in their usage of peptides and amino acids, whilst retaining the capability of producing extended peptides, often with altered composition (Schoppet et al.,

2019). This, coupled with the mechanistic uncertainty and incomplete detail of protein interaction formation, show that studies into the functionality of C domains are yet required to understand, and subsequently alter, the catalytic capability of the proteins catalysing the central coupling reaction in NRP biosynthesis.
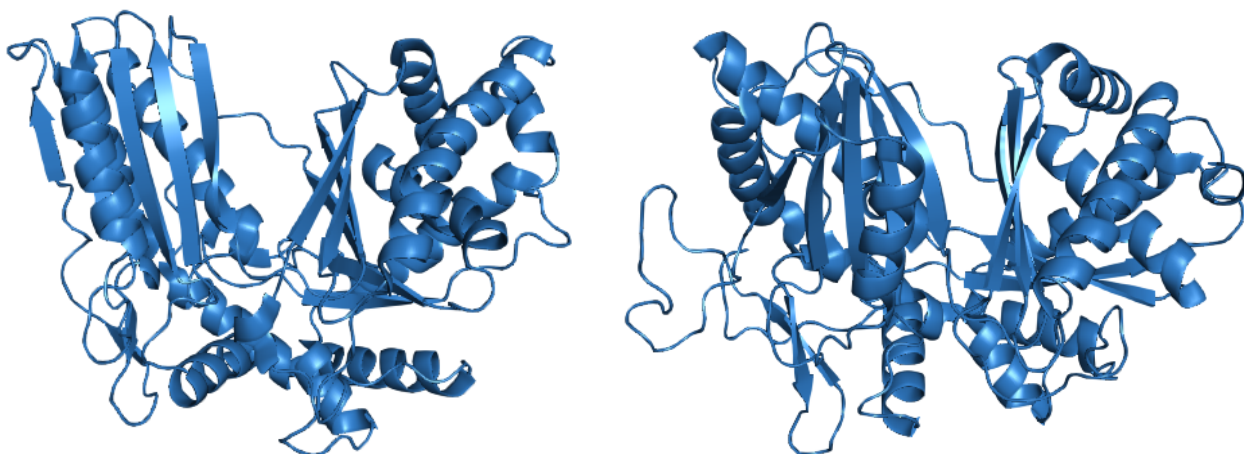
Figure 1.10: Crystal structures of VibH, a condensation domain from vibriobactin biosynthesis, and CDA-C1, the C domain from the first module of calcium-dependent antibiotic biosynthesis, respectively. The structures illustrate the V-shaped, pseudodimeric structures of the domains, reminiscent of chloramphenicol acetyltransferase enzymes, which allows binding of donor PCP domain from the top side and acceptor PCP domains from the bottom (Keating et al., 2002; Bloudoff et al., 2013).

1.4.4 Accessory catalytic domains in NRP biosynthesis

The products produced by NRPS biosynthetic systems are rarely constructed of solely L-amino acids connected by peptide bonds, as is observed in the ribosomal synthesis of peptides and proteins. Structural diversity in the chemical scaffolds of NRPs can be introduced at various points along the length of the peptide, and these diverse chemical motifs are commonly responsible for the bioactivity displayed against various cellular targets (Caboche et al., 2010). The core modules of NRPS pathways function to select amino acid substrates and catalyse peptide bond formation between tethered intermediates. Although the function of core domains and modules can result in the synthesis of peptides with altered chemical moieties, usually through the A domain mediated selection of variant substrates, the vast majority of structural diversity observed in NRPs is introduced by enzymes not forming part of the core module, also designated as accessory domains. A number of accessory domains have been discovered as part of NRPS systems and the selection of proteins carry out varied chemical reactions on the multitude of functional groups existing along the peptide backbone and within amino acid side

chains. The alterations made on NRPs range from relatively simple chemical modifications such as methylations, to more complex reactions such as reductions, oxidations, and cyclisations. The majority of these chemical alterations are carried out by *cis*-encoded domains, of which the most common is the epimerase (E) domain, but examples of *trans*-acting and post assembly line enzymes are commonly observed.

Most NRPs contain at least one D-amino acid, which confers multiple benefits onto the characteristics of the final peptide, chiefly a reduction in the rate of proteolysis of the mature, released peptide (Hamamoto et al., 2002). A minority of A domains are capable of directly adenylating and activating D-amino acids, however, the vast majority of D-configured residues are installed with the aid of NRPS module-integrated E domains. E domains are modified condensation domains, which retain the catalytically conserved residues as well as the overall pseudodimeric protein conformation, albeit whilst usually only retaining <20% sequence identity. Early structural and biochemical studies into various E domains from the tyrocidine and gramicidin pathways were able to highlight the similarities and, importantly, the differences between epimerase and condensation catalysing domains (Stachelhaus and Walsh, 2000; Linne et al., 2001; Stein et al., 2005). E domains contain the HHxxDG and EGHGRE conserved motifs, which contain the catalytic His and Glu residues (Figure 1.11). These motifs are held in close proximity at the central cleft of the protein, where they would be able to interact with PCP-tethered substrates and carry out the postulated acid-base mechanism for epimerisation at the α-carbon of the amino acid substrate. E domains, however, contain an insertion of multiple amino acid residues located near the entry site of the acceptor PCP, which occlude binding at that site and limit E domains to functionality on a single peptide substrate, donated from the upstream PCP (Samel et al., 2014). The catalytic activity of E domains could hypothetically occur on aminoacyl or peptidyl substrates, i.e. directly after adenylation and attachment of an amino acid to the PCP, or post condensation when an extended peptide has been formed. Biochemical studies have shown E domains behave variably in this regard, dependent upon the nature of the domain itself, however, the accepted convention for elongation modules is that the PCP will deliver its aminoacyl substrate to the upstream C domain for peptide elongation, before presentation to the downstream E domain for conversion of the C-terminal amino acid to the D-configuration (Figure 1.11) (Linne and Marahiel, 2000). In this way directionality of biosynthesis is not lost and efficiency in NRP production is retained.

A

B
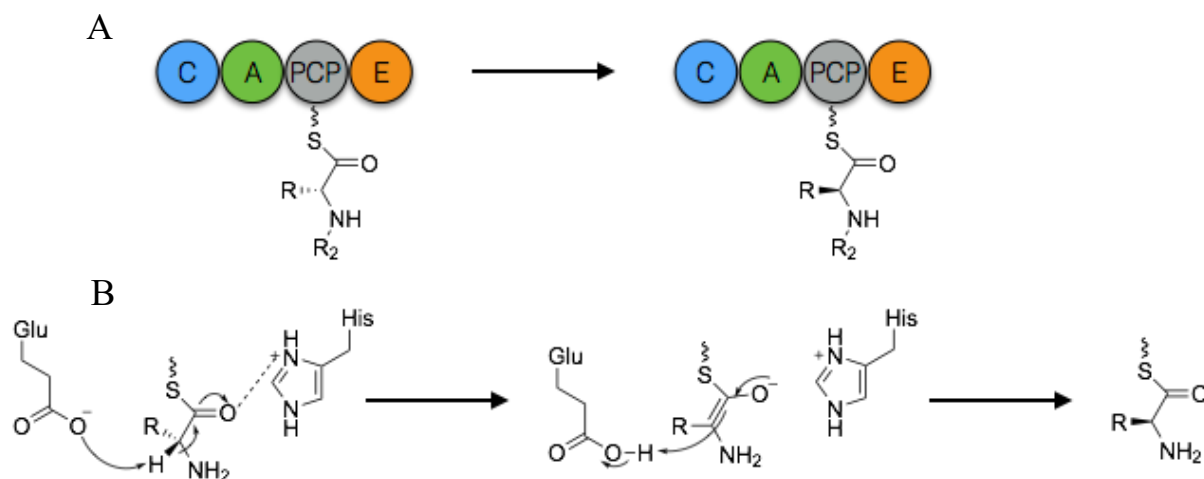
Figure 1.11: A) Basic overview of the effect of the presence of E domains within NRPS modules. Epimerisation at the alpha carbon of the upstream intermediate is catalysed prior to continuation of NRP synthesis. B) Proposed catalytic mechanism of epimerisation. A catalytic Glu residue is predicted to abstract the alpha proton, with stabilisation of the formed enolate by a His residue. Collapse of the enolate intermediate and re-protonation from the opposing side leads to formation of the epimerised product.

1.4.5 Assembly line release of NRPs

Chain termination and assembly line release of products from NRPS assembly lines is a crucial and essential process in biosynthesis, ensuring the production of mature, bioactive peptides and allowing further cycles of catalysis to occur on NRPS megasynthases, without extended times of stalling. The most common method of assembly line release observed within NRPS systems is through the action of a thioesterase (TE) domain. TE domains are usually located at the extreme C terminus of NRPS pathways, and are ~ 30 kDa proteins belonging to the α/β-hydrolase superfamily, commonly also containing proteins with lipase and protease activity (Shaw-Reid et al., 1999). The conserved Ser-His-Asp catalytic triad of the TE domain is responsible for the enzymatic activity of termination domains, which typically occurs in a two-step process (Figure 1.12). The Asp and His residues act as general acids and bases, resulting in deprotonation of the Ser residue, subsequently allowing nucleophilic attack of the Ser hydroxyl oxygen into the carbonyl of the thioester bond, tethering the linear peptide to the terminal PCP domain. The translocation of the peptide onto the Ser hydroxyl yields a peptidyl-*O*-Ser oxoester bond on the TE domain, which can itself be attacked by a nucleophile to release the peptide from the TE domain. The attacking nucleophile, depending on the TE domain, can be recruited from the external, cellular environment, usually water from the surrounding bulk

solvent, leading to a linear hydrolysed product, but can also be an internal nucleophile, already present on the peptidyl intermediate. In this scenario, nucleophiles such as, but not limited to, hydroxyl or amino groups present within amino acid side chains of the peptidyl intermediate, or the N-terminal amino group, can attack the oxoester bond as part of the peptidyl-*O*-Ser intermediate and result in the formation of a macrocyclic structure (Trauger et al., 2000). The structures of multiple TE domains have been solved and show how the topology and conformation of the protein mediate its activity (Bruner et al., 2002; Frueh et al., 2008). The protein is formed of an alternating β/α/β motif, with an open hydrophobic channel at the centre of which is the catalytic triad. Two α helices form a lid region at the top of the substrate binding channel, which in the crystal structure of the TE domain from the erythromycin synthesising system (6-deoxyerythronolide B synthase, DEBS) were shown to be somewhat controlling of substrate binding (Tsai et al., 2001; Tsai et al., 2002). PCP domains are capable of interacting with TE domains at the N-terminus of the protein and delivering the tethered intermediate to the substrate binding channel, via extension of the Ppant arm. Catalysis can then occur, releasing the mature peptide and priming the NRPS system for further NRP biosynthesis.
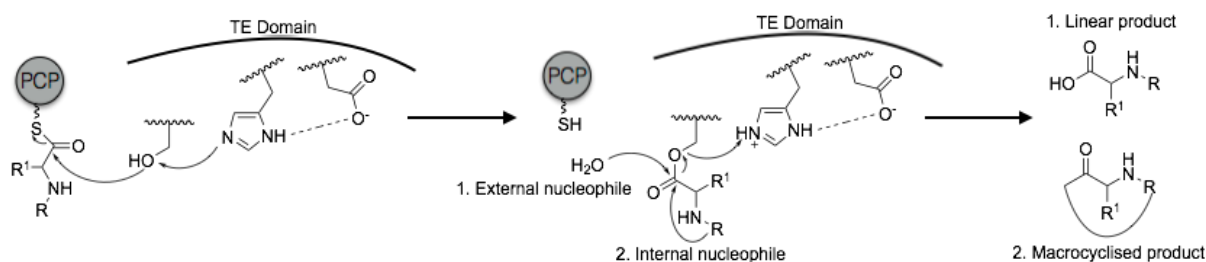
Figure 1.12: The catalytic mechanism employed by thioesterase domains in the release of peptides from PCP domains. TE domains use either external or internal nucleophiles in order to catalyse the formation of linear or macrocyclised products, respectively.

TE domains are also capable of catalysing a multitude of different reactions, including carbon-carbon bond formation and oligomerisation, all resulting in chain release from the biosynthetic system. However, TE domains are not the sole protein responsible for chain termination in NRPS systems, and other domains have been observed to carry out offloading, resulting in varied chemical and functional groups at the C-terminus of the synthesised peptide. Reductase (R) and condensation-like (C_T) domains are the two most commonly observed alternative domains that catalyse peptide release and termination of NRP biosynthesis (Du and Lou, 2010). An R domain as part of an NRPS system was first described for the primary metabolic pathway

leading to the biosynthesis of the natural amino acid Lys, followed by subsequent observations in the secondary metabolic pathways of saframycin and linear gramicidin A biosynthesis (Ehmann et al., 1999; Schracke et al., 2005). The domains usually carry out two or four-electron reductions leading to the formation of aldehydes or primary alcohols respectively, but can also carry out other, varied chemical reactions. R domains are members of the short chain dehydrogenase/reductase (SDR) enzyme superfamily and possess a Rossmann fold super-secondary structural motif. The structure contains a characteristic nucleotide binding motif, allowing the use of NADP(H) as a cofactor for reductive release of peptidyl intermediates. The linear gramicidin A synthetase is composed of four NRPS polypeptides, the last of which (LgrD) terminates in an R domain, able to catalyse release of an aldehyde terminating linear gramicidin A peptide. A further standalone R domain catalyses further reduction of the aldehyde functionality, resulting in formation of the alcohol, the final product of the NRPS pathway. The mechanism of R domain mediated peptide termination is also thought to differ from the mechanism employed by canonical TE domains. R domain offloading is posited to occur directly from the PCP, via reduction of the thioester bound substrate, rather than the two-step process of translocation followed by offloading utilised by TE domains (Barajas et al., 2015).

C_T domains are the evolutionarily preferred method of peptide offloading and macrocyclisation used in many fungal NRPS systems, and resemble C domains used in elongation modules in their structure and basic functionality (Gao et al., 2012). The major dissimilarity between C_T and C domains is in the structure of two of the N-terminal α-helices, which adopt distinct conformations resulting in the occlusion of the acceptor binding site, much like the blocking of one of the substrate binding sites in canonical E domains. The typical protein structure of C_T domains, whilst permitting substrate access from the donor site, also restricts access by external nucleophiles from bulk solvent, therefore commonly leading to macrocyclic products, via intramolecular cyclisation. This was first observed upon *in vitro* characterisation of the biosynthetic pathway leading to the production of fumiquinazoline F, a fungal natural product, where the C_T domain was responsible for the first cyclisation and concomitant release of the intermediate from the final PCP domain (Ames et al., 2010). A further example of C domain catalysed product release is in the formation of the *Streptomyces* antitumour antibiotic C-1027, where the C_T domain catalyses attachment of a modified β-tyrosine to the enediyne core of the compound (Lin et al., 2009). The biosynthetic pathway, whilst not completely NRPS

dependent, utilises standalone C, A, and PCP domains to recruit various modifying enzymes and catalyse the joining of the main pharmacophores to form the final product. The action of the C_T domain also results in offloading of C-1027 from the PCP domain and was additionally found to explicitly require the PCP domain for functionality as well as possessing the innate ability to catalyse the formation of both ester and amide bonds, highlighting the potential for chain termination enzymes in engineering and biocatalytic applications. NRPS systems have also adapted to the use of various other methods and mechanisms for chain termination and release, including examples of spontaneous release, showing the chemical and mechanistic variability employed by NRPS systems in the biosynthesis of a multitude of peptidic scaffolds.

1.4.6 Precursor biosynthesis

NRPS systems are capable of activating and utilising varied precursors for subsequent incorporation into final products. The precursors used during NRP biosynthesis range from stereochemically altered amino acids, to unnatural modified and β amino acids, however, the vast majority of amino acids activated by A domains are α amino acids derived from the primary metabolic pathways that synthesise the amino acids for ribosomal protein synthesis, as well as for other cellular pathways (Wilkinson and Micklefield, 2009, p.14). NRPS biosynthetic gene clusters have been observed to encode the required proteins for the totality of the necessary cellular pathways, required for complete production of NRPs. These systems include the biosynthetic pathway itself, as well as proteins responsible for regulation of the cluster, *in trans* modification of chemical moieties, and transport of the final compound. Due to the chemical variability of NRPs and the lack of cellular pathways synthesising the varied precursors required, NRPS BGCs often encode the proteins and enzymatic systems required for production of the altered substrate. These enzymatic systems vary dramatically dependent upon the nature of the required precursor, but usually take products or shunt metabolites from primary metabolism and catalytically transform them to the necessitated compound, usually through a multi-step procedure. A number of these precursor biosynthetic pathways have been discovered, and a variety of these are described in sections 1.7 and 5.1.3. The unearthing of these systems has yielded novel and interesting enzymes, and many more precursor biosynthesis proteins are yet to be discovered. With increasing interest in NRPS systems and the compounds produced, and access to genomic information ever increasing, higher numbers of precursor biosynthetic proteins should be revealed, leading to an expanded set of fascinating enzymes available for *in vitro* and *in vivo* biocatalytic applications.

1.5 Polyketide synthases

1.5.1 Biosynthetic gene clusters encoding polyketide synthases

Polyketide synthases (PKSs), much like NRPSs, are multicatalytic machineries carrying out a diverse set of chemical reactions towards the biosynthesis of an extremely varied range of natural products. PKSs are also evolutionarily, and therefore mechanistically, related to fatty acid synthases (FASs), and function in an assembly line-like manner with intermediates passed along the synthetic machinery until a mature compound is formed and released (Staunton and Weissman, 2001). Similar to NRPS biosynthetic systems, the proteins required and involved in the biosynthesis of the final polyketide are colocalised within the genome within a BGC. The BGCs usually contains numerous genes, as illustrated in Figure 1.13, with the largest and most prominent being the genes encoding the PKS assembly lines itself. Other genes in the BGC can, when expressed, result in the production of transporters and regulators, as well as, commonly, modification enzymes, such as glycosyl transferases and methyltransferases, amongst a multitude of others. Possibly the most well studied PKS pathway is the system leading to the biosynthesis of 6-deoxyerythronolide B (6-dEB) (Caffrey et al., 1992). The DEBS system contains three large genes (*eryAI*, *eryAII*, *eryAIII*), each ca. 10 kb in length, and individually encoding the formation of two PKS modules, giving a hexamodular biosynthetic system. The action of the modular PKS system yields the 6-dEB aglycone, which is then transformed into erythromycin by the action of further enzymes, efficiently produced by BGC encoded genes, as spatiotemporally required for the assembly line.
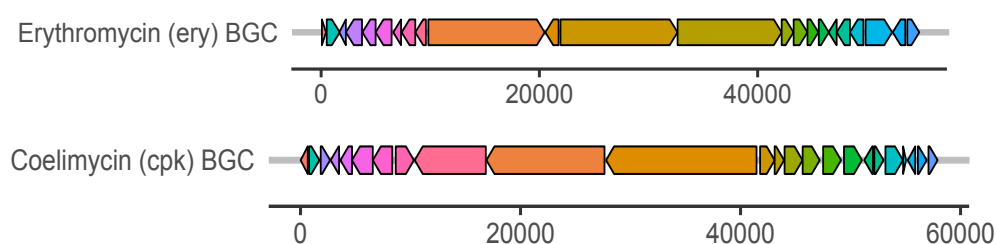
Figure 1.13: Examples of PKS BGCs, including the clusters responsible for the production of the antimicrobials erythromycin and coelimycin. Each arrow represents a gene in the proposed cluster, whose product contributes to the formation of the final compound. The x-axis scale represents the length of each BGC in DNA base pairs.

1.5.2 Biosynthetic logic of PKS assembly lines

Most PKS pathways utilise a modular biosynthetic logic, with each module composed of two or more individual catalytic domains (Figure 1.14). In much the same way that NRPSs synthesise a wide variety of bioactive peptides, PKS modules are each responsible for the addition of one substrate, an acyl-coenzyme A (CoA) derivative, usually malonyl or methylmalonyl-CoA. PKS elongation modules are frequently composed of ketosynthase (KS), acyltrasferase (AT), and acyl carrier protein (ACP) domains, which each play important catalytic roles in the elongation of the polyketide intermediate and formation of the final product (Khosla et al., 1999). A variety of accessory catalytic domains are also frequently observed as part of PKS modules, predominantly ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) domains, and these enzymes are capable of causing marked chemical variations in the polyketide product. The collinearity rule is also observed for the majority of PKS pathways, where the number of biosynthetic modules is directly related to the number of substrates used, and therefore corresponds to the nature and length of the produced natural product (Callahan et al., 2009).
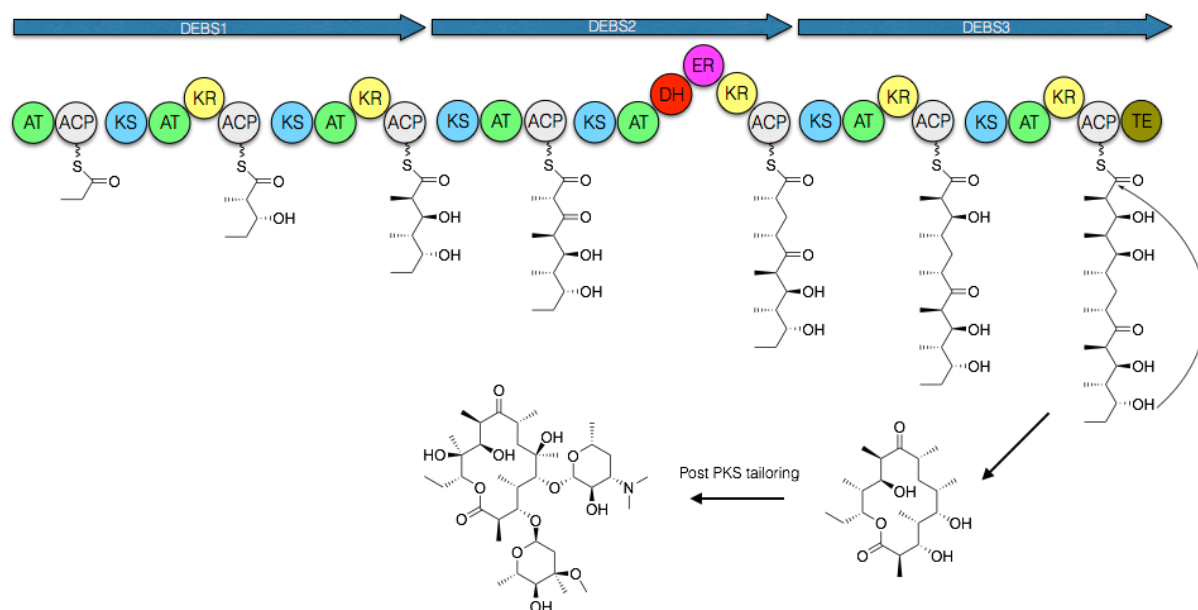
Figure 1.14: Overview of the erythromycin biosynthetic pathway, illustrating the functionality of the large PKS machinery in the sequential addition of two carbon units. Module encoded tailoring domains alter the β-position of intermediates, before a TE domain releases 6-deoxyerythronolide B from the assembly line by cyclisation. Post PKS tailoring enzymes act on the released compound to result in the final formation of erythromycin.

A further similarity of PKSs to NRPS systems is in the variability of the systems observed in nature. A large proportion of the pathways known and studied to date are modular type I PKSs observing the collinearity rule, i.e. the PKS modules are arranged in the canonical assembly

line-like manner and each module is used once in the addition of a substrate, followed by chain termination by a specialised C-terminally located domain. Two other biosynthetic methodologies are also commonly employed by PKS systems in the synthesis of varied compounds. Type II PKSs contain a set of domains which act iteratively, generally towards the formation of aromatic polyketide products (Hertweck et al., 2007). Type II PKSs utilise KS domain subunits and employ a thiotemplating mechanism through the use of ACP domains to form poly-β-keto chains, which are subsequently cyclised and aromatised by distinct enzymes before potential further derivatisation by tailoring enzymes, yielding diverse chemical and reactive handles contributing to increasingly potent therapeutic properties (Z. Zhang et al., 2017). The iterative action of the domains forming type II PKSs, and the similarity of the substrates initially activated, are the primary differentiating factors between the functionalities of type I and type II PKS systems, both of which have yielded important compounds for research and clinical use. Type III PKSs, also known as chalcone synthase-like PKSs, are homodimeric enzymes, essentially formed solely of KS domains (Austin and Noel, 2003). These PKSs are also iteratively acting but are ACP-independent and act directly on CoA-bound substrates. KS domains catalyse multiple, iterative rounds of joining between acyl-CoA precursors which are then released and cyclised to form aromatic compounds. The presence of PKSs utilising multiple different biosynthetic strategies demonstrates the evolutionary variability of PKSs in the creation of a diverse set of interesting, bioactive compounds (Shen, 2003).

1.5.3 Core catalytic domains in PK biosynthesis

1.5.3.1 Acyltransferase (AT) domains

AT domains strongly influence the chemical nature of the products made by PKSs, due to its functionality of selection and activation of acyl-CoA extender units, which, by the catalytic action of the remainder of the assembly line, become integral parts of the final structural scaffold. The overall enzymatic mechanism employed by AT domains occurs via a two-step process: activation and formation of an acyl-enzyme intermediate; and subsequent loading of the downstream ACP domain (Figure 1.15) (Dunn et al., 2013). Initially, an acyl-CoA substrate is selected from the cellular pool and the AT domain covalently binds to the precursor through an active site Ser residue to form an acyl-O-AT intermediate. The formation of the covalently

bound complex results in the release of coenzyme A, whilst the activated substrate is generally shielded from bulk solvent in an attempt to avoid futile hydrolysis, and release of the required substrate. The catalytic cycle of the AT domain is complete upon transfer of the activated substrate onto the terminating thiol group of the *holo*-ACP domain, which then allows catalysis by the remainder of the module-resident domains and elongation of the upstream polyketide intermediate.
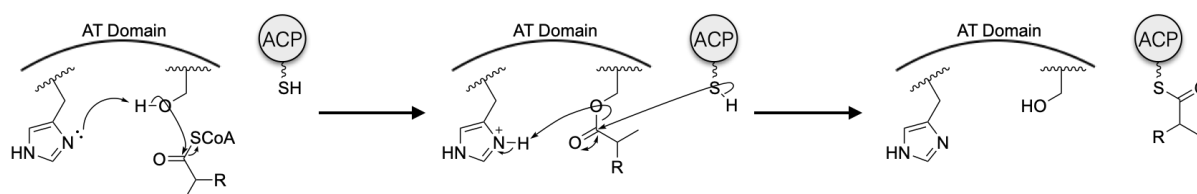
Figure 1.15: Catalysis by AT domains in the loading of ACP domains with required acyl-CoA substrates. The AT domain activates the substrate, leading to a Ser-tethered intermediate and release of CoA, followed by loading onto a *holo* ACP domain by nucleophilic attack of the Ppant terminating thiol group.

As AT domain mediated substrate selection is one of the most important aspects in controlling the diversity of biosynthesised polyketides, a number of studies have been devoted to examining the structural and mechanistic rationale of the domain and how these may be manipulated to form novel biosynthetic pathways and novel natural products (Ruan et al., 1997; Reeves et al., 2001). A variety of domain, didomain, and modular structures of PKSs have been solved, with the majority from the DEBS pathway and including an impactful structure of the complete fifth module involved in pikromycin biosynthesis (Dutta et al., 2014). AT domains are generally larger domains, usually composed of ~ 400 amino acid residues, and were shown to form extensive protein contacts with both KS and ACP domains, with interactions with carrier proteins found to be increasingly important. Kinetic analysis of AT domain mediated activation of methylmalonyl-CoA (mmCoA), by the domain from module three of the DEBS system, showed that efficient transfer of an activated substrate only occurred in the presence of the cognate *holo*-ACP domain, and not in the presence of another ACP from the same system, instead resulting in eventual hydrolysis of the acyl group from the active site of the AT domain (Dunn et al., 2013). Whilst the molecular detail surrounding substrate selection by AT domains is not completely resolved, studies have focussed on various conserved stretches of amino acids within the protein, with the most notable being motif III, albeit as a small stretch of residues in a complex substrate selecting enzyme (Keatinge-Clay et al., 2003; Oefner et al., 2006). Motif III resides approximately 100 residues downstream of the active site Ser and

generally forms part of the wall of the binding pocket. The motif is often conserved in domains that accept mmCoA as the amino acids YASH (Tyr, Ala, Ser, His) and in domains that activate and utilise malonyl-CoA (mCoA) as a HAFH motif (His, Ala, Phe, His) (Smith and Tsai, 2007). Interchanging of these motifs in the DEBS system has been shown to alter the substrate usage *in vivo*, however, the generalisable nature of this motif switch as a way of engineering AT domains is unknown, mainly due to the lack of kinetic data surrounding the functionality of mutant domains. Nevertheless, efficient and functional AT domain engineering remains an extremely sought-after outcome and a multitude of methodologies have been tested in this regard, including domain swaps, module swaps, mutagenesis, and *in trans* complementation. As the molecular toolbox for engineering AT domains, and PKS modules grows, the engineered biosynthesis by PKS pathways should ideally allow access to novel polyketide scaffolds, useful in the pursuit of new clinically and therapeutically relevant compounds.

1.5.3.2 Acyl carrier protein (ACP) domains

ACP domains share large amounts of functional identity with NRPS PCP domains, being the carrier proteins that tether substrates and intermediates to the assembly lines via a thiotemplated mechanism. ACPs are also small helical bundle proteins, with a mass of approximately 9 – 10 kDa, that are post-translationally modified by a PPTase. The addition of a thiol terminating Ppant arm onto a conserved ACP Ser residue by PPTase enzymes, allows the ensuing polyketide synthesis to occur (Beld et al., 2013). ACP domains are essential for the function of PKSs and are an inherent part of the mechanism employed by PKS modules. AT domains select and activate malonyl-CoA substrates, with the second half-reaction leading to the loading of the selected substrate onto the lengthy Ppant arm of the ACP domain. The upstream KS domain can then catalyse a decarboxylative Claisen reaction, causing the elongation of the polyketide intermediate and translocation of the complete chain onto the ACP (Charkoudian et al., 2011). Whilst the ACP itself is not enzymatically active, the domain plays a crucial role in substrate tethering and transfer to catalytic centres, without which the PKS machinery would not be functional in the creation of diverse polyketide products.

The structure and dynamic nature of ACP domains has been the topic of interest of numerous different studies. Structures of ACP domains have been solved by multiple techniques, including X-ray crystallography, electron microscopy (EM), and nuclear magnetic resonance (NMR) imaging, and as part of a multitude of different protein complexes (Alekseyev et al.,

2007; Keatinge-Clay, 2012). The structurally solved protein complexes ranged in size and modular composition from partially reducing complete modules to didomain constructs, and solely ACP domains. The carrier proteins themselves have also been visualised in *apo*, *holo*, and loaded forms, each depending on the nature and presence of the Ppant arm. Overall, ACP domains are small, simple four-helix bundle proteins, composed of three major α-helices and another smaller, perpendicularly located helix, important in the helical packing of the protein. However, upon posttranslational modification of a conserved Ser residue, residing within a well-conserved GxDSL motif, by a PPTase enzyme, the *holo*-ACP domain becomes extremely structurally and conformationally dynamic. The dynamics generally involve movement of the whole protein by occupation of multiple modular sites, dependent upon the required reaction location and necessitated by the polyketide structure and the domains within a module (Li et al., 2003; Whicher et al., 2014). Molecular level rotation and extension of protein motifs allows extension of the Ppant arm into the active sites of other domains, allowing reactions to occur on the covalently linked intermediate, and therefore, mediating the overall construction of polyketides by the megasynthases. ACPs, therefore, by their structural simplicity but complex conformational flexibility, play a role of utmost importance in the function of PKSs.

1.5.3.3 Ketosynthase (KS) domains

KS domains are the main catalytic enzymes of PKS modules and during elongation, catalyse the joining of the upstream intermediate with the downstream, activated substrate. In contrast to the mechanism employed by NRPS C domains, where both upstream and downstream substrates enter the C domain active site simultaneously, KS domains initially become covalently attached to the upstream substrate through an active site Cys residue (Figure 1.16). The polyketide intermediate on the upstream ACP is attacked by the KS domain Cys residue, activated by a general acid-base mechanism of a nearby His residue, causing translocation of the intermediate onto the KS domain (Robbins et al., 2016). Once an intermediate is bound in the KS domain active site, the downstream substrate, held on the module-resident ACP, is then able to enter the KS domain active site and undergo a decarboxylative Claisen condensation. The reaction, again thought to be mediated by general acid-base catalysis by active site His residues, results in loss of carbon dioxide and extension of the backbone polyketide scaffold by two carbon units. Concomitant with the extension of the intermediate is the translocation of the complete scaffold from the KS domain to the downstream ACP. Further reactions,

predominantly catalysed by module encoded enzymes, can then be carried out to determine the oxidative state and stereochemistry of the newly added β-carbon.
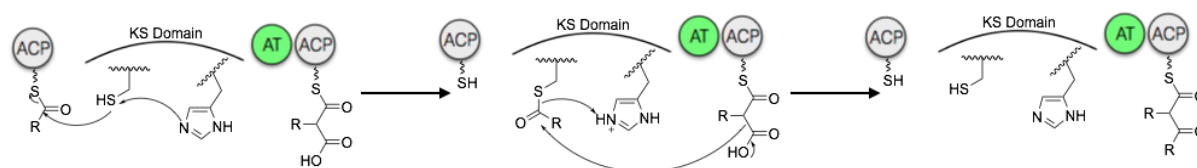
Figure 1.16: Mechanism of catalysis by KS domains in the initial reaction with an upstream ACP appended substrate, which results in attachment onto the KS domain catalytic cysteine. Subsequent decarboxylative Claisen condensation with a compound tethered to the downstream ACP results in translocation and extension of the polyketide chain.

PKS modules have been known to be homodimeric proteins for numerous years, however, the overall structural topology adopted by the modules has been debated. Recent studies have suggested a horseshoe shaped homodimer, with the KS domain forming the major dimerisation interface (Dutta et al., 2014). N-terminal docking domains also play a major role in dimerisation by formation of a coiled-coil interface, whilst the KS domain forms a characteristic αβαβα fold, with β strands from each monomer held together to form the KS dimer (Gay et al., 2014). The domain also forms important interactions with other domains, notably covalently linked to the AT domain and requiring precise interactions with the upstream and downstream ACP domains to mediate catalysis and polyketide extension. The post-AT linker has been shown to be extremely flexible and mobile, with the ability to reach and interact with the KS domain in the absence of the ACP itself (Chen et al., 2007). Therefore, PKS module activity is predominantly centred around delivery of substrates and intermediates to the KS domain, where catalysis of chain elongation occurs, highlighting the importance of the primary catalytic domain in PKS functionality and polyketide formation.

1.5.4 Accessory catalytic domains in PK biosynthesis

The core PKS module is composed of the KS, AT, and ACP domains, and are proficient and effective in the formation of basic polyketide scaffolds. However, the vast majority of polyketides and PKS-synthesised products are composed of saturated, reduced, or modified carbon centres. The alteration of these atoms is carried out by module-encoded domains, not essential for the initial extension of the polyketide chain. These accessory domains, the ketoreductase, dehydratase, and enoyl reductase domains, are usually functional at the β-keto

group of the growing chain and alter the oxidation state at that position. These relatively common domains are discussed below. Other domains, however, are also observed and functional within the confines of a PKS module (Hertweck, 2009). These domains occur less frequently within PKS systems but are capable of carrying out more varied modifications upon the growing polyketide chain. *O-* and *C-*methyltransferase (MT) domains are examples of such domains and are catalytically active in the *S-*adenosylmethionine (SAM) dependent methylation of a β-hydroxy group or α-positioned saturated carbon, respectively. These varied modifications yield polyketides with diverse chemical moieties and allow the exploration of further available chemical space, thereby potentially giving compounds with new and interesting bioactivities.

1.5.4.1 Ketoreductase (KR) domains

Upon KS domain activity and elongation of the growing polyketide chain, the first accessory, reductive domain that can then function on the intermediate, is the KR domain. Prior to the identification of PKS pathways as the synthesising machinery for polyketide compounds, Celmer noticed a strong homology between certain positions of macrolide antibiotics and the stereochemistry at the specified position, therefore, suggesting a genetic origin for the stereochemical control (Siskos et al., 2005). Study of the mechanism of the domains involved in PKS biosynthesis indicated that KR domains were also capable of controlling the stereochemistry at the α position (those positions that are stereogenic and not synthesised from malonyl-CoA), as well as at the newly formed β-hydroxy position (Holzbaur et al., 1999). These studies demonstrated an interesting epimerase activity of KR domains, as well as the titular ketoreductase activity.

B
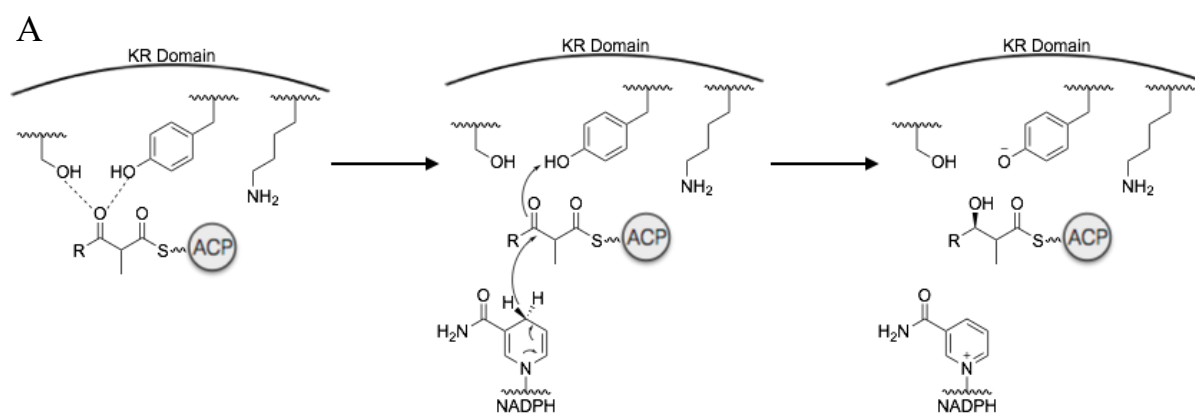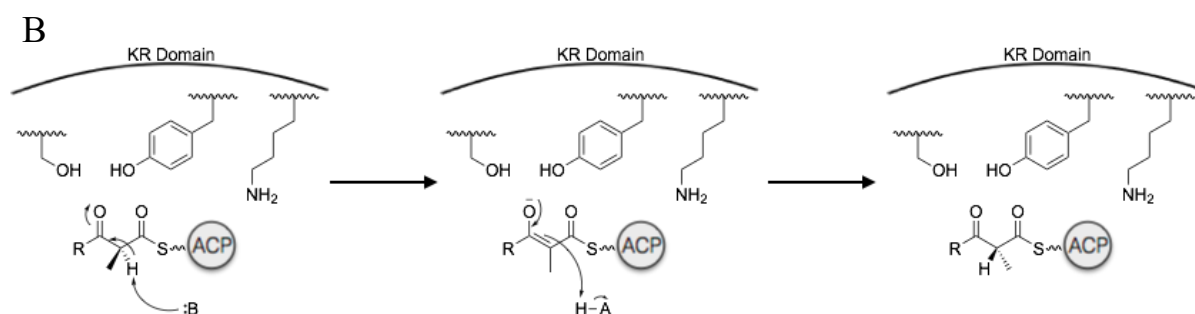
Figure 1.17: A) Mechanism of ketoreduction carried out by KR domains. The newly formed substrate on the downstream ACP is stabilised and activated by KR domain active site residues before receiving a hydride from the NADPH cofactor, resulting in stereospecific reduction of the β-keto group. B) General mechanism of epimerisation utilised by KR domains, where B represents a general base and A represents a general acid. Deprotonation at the α position followed by reprotonation from the opposing side results in formation of the stereoisomer.

KR domains are members of the short chain dehydrogenase/reductase (SDR) family of enzymes, and therefore, contain a Ser-Tyr-Lys active site catalytic triad (Filling et al., 2002). The conserved Ser and Tyr residues cooperate to hydrogen bond to the oxygen atom of the carbonyl at the β position, and the Tyr side chain also acts as a general acid, which activates the carbonyl for reductive, nucleophilic attack. The activated substrate can then receive a hydride from the active site bound nicotinamide adenine dinucleotide phosphate (NADPH) cofactor, followed by abstraction of a proton from the active site Tyr, which is itself stabilised by the conserved Lys residue, resulting in formation of the hydroxyl group (Figure 1.17) (Reid et al., 2003). Dependent upon the side of hydride donation (*re*, top or *si*, bottom) and subsequent hydroxyl formation, intermediates with different stereochemistries (*R* or *S*, respectively) at the β position are created by KR domains. The resultant stereochemistry determines the type of KR domain; through standard nomenclature, A-type KRs result in (*S*) and B-type KRs in (*R*) stereocentres (Caffrey, 2003; Zheng et al., 2010). In addition, C-type KR domains are reductase inactive. KRs can also be selective for defined α-configured substrates, where an extra level of functional complexity is provided by the epimerase activity of the domains. Interestingly, the catalysis of epimerisation at the α position by KR domains was found to be reliant upon the same conserved Tyr and Ser residues as utilised in the reductase activity (Figure 1.17) (Garg et al., 2014; Xie et al., 2016). Similar to ketoreduction, the Tyr and Ser residues hydrogen bond to the β-carbonyl, stabilising the oxygen atom to take a negative charge in the ensuing enolate formation. An unknown base abstracts a proton at the α position resulting in enolate formation, followed by collapse of the enolate intermediate and reformation of the carbonyl.

Reprotonation at the α position, however, occurs from the opposite side to abstraction, resulting in formation of the epimerised product. The mechanistic and functional diversity of KR domains results in the production of varied polyketide products, adding further to the diversity of the biosynthesised compounds.

1.5.4.2 Dehydratase (DH) domains

The action of KR domains on polyketide intermediates results in the formation of ACP bound, β-hydroxy compounds, with the reductase domain also defining the stereochemical outcome. The next domain active in a fully reducing module is the dehydratase (DH) domain. DH domains catalyse the *syn*-coplanar elimination of water, thereby generating a double bond between the α and β positions (Figure 1.18). Similar to DH domains of FASs, PKS DH domains all adopt a characteristic double hotdog fold and contain an invariant catalytic dyad composed of Asp and His (Keatinge-Clay, 2008; Akey et al., 2010). The His-Asp catalytic dyad is situated at the end of a Ppant binding tunnel, the site of ACP-mediated substrate donation, and a further hydrophobic binding pocket accommodates the remainder of the polyketide substrate beyond the β-hydroxy group. The mechanism of dehydration proceeds through α abstraction of a proton by the catalytic His residue, followed by donation of a proton by the formed conjugate acid to the β-hydroxyl, leading to elimination of water and generation of the α,β-unsaturated thioester. The conserved Asp residue, initially thought to participate in acidic proton donation, likely orients the hydroxyl group in a favourable position for olefin formation (Xie and Cane, 2018). The stereochemical outcome post DH domain function is not thought to be genetically determined, as is the case for KR domain functionality. A lack of conserved motifs in *cis* or *trans* olefin generating domains suggest the stereospecificity in product formation is controlled elsewhere. Initial evidence in this regard suggested the geometry of the double bond was inherently controlled by the preceding KR domain; i.e. A-type KRs produced (*S*)-configured alcohols, which subsequently produced *cis*-olefins upon DH domain action, and vice versa for B-type KRs (Keatinge-Clay, 2016). Exceptions to this rule, however, have been observed, further obscuring the functionality of the primary olefin creating domain observed in many PKS systems.
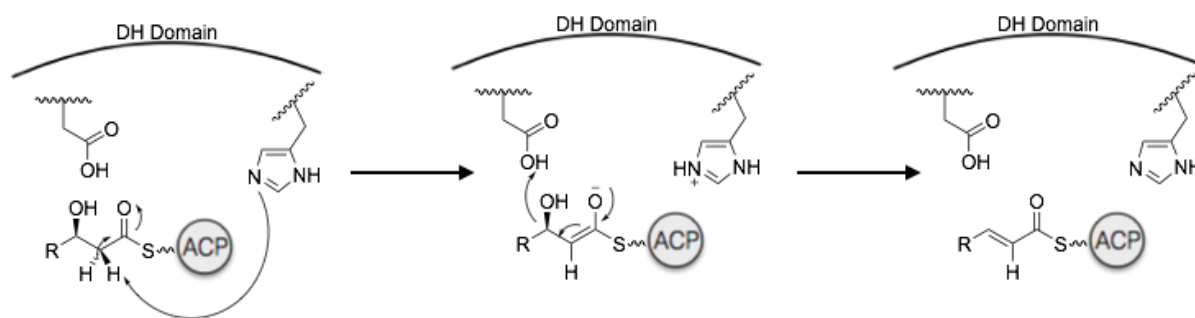
Figure 1.18: Mechanism of DH domain mediated dehydration of a β-hydroxythioester. A catalytic His residue deprotonates the α carbon, followed by collapse of the enolate intermediate and loss of a water molecule.

1.5.4.3 Enoyl reductase (ER) domains

The final step in formation of a fully reduced polyketide is enoyl reduction by ER domains. The overall reaction catalysed by ER domains is the reduction of the α,β-unsaturated double bond, with concomitant determination of the stereochemistry at the α position, should any substituents be present. Bacterial PKS ER domains belong to the medium chain dehydrogenase/reductase (MDR) family of enzymes, which contain two characteristic subdomains. Similar to SDR enzymes, a Rossmann fold motif mediates nucleotide cofactor binding, with the second MDR subdomain binding requisite substrate molecules (Ames et al., 2012). Although catalytic residues are not categorically determined for ER domains, invariant Lys and Asp residues have been observed in numerous ERs, which could play a part in proton donation and/or substrate stabilisation. Mutation of the conserved Lys residue has been shown to severely diminish enzymatic activity, suggesting a role in catalysis for this amino acid (Kwan and Leadlay, 2010). The overall mechanism of ER activity begins with substrate binding, followed by hydride transfer from the NADPH cofactor to the β position. Protonation of the formed carbanion results of generation of the saturated species (Figure 1.19). The stereochemistry of any α substituents is determined, at least in part, by a Tyr residue observed in ~ 50% of ER domains (Zhang et al., 2018). In these domains the aromatic amino acid has been shown to be important in formation of the (S)-configured product, however, alteration of corresponding residues to Tyr in other ER domains did not result in switching of stereochemistry, showing that other amino acids also participate in stereocontrol of enoylreduction.
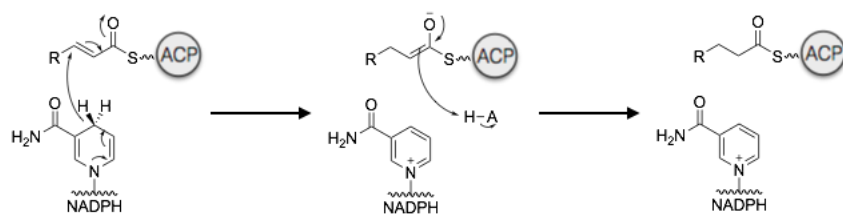
Figure 1.19: Basic catalysis of reduction carried by enoylreductase domains in the formation of saturated thioester products. A represents a general acid.

1.5.5 Canonical assembly line release mechanisms of PKSs

Upon biosynthesis of a complete polyketide chain by a PKS megasynthase, the scaffold must be offloaded to yield the mature compound. NRPS and type I PKS systems share many similarities in the mechanisms used for offloading and compound release. In both cases, type I TE domains are the most commonly observed domain responsible for termination of biosynthesis, with reductive and condensation domain mediated release also employed by both types of synthetic system (Du and Lou, 2010). TE domains of PKS systems, therefore, closely resemble their NRPS counterparts, and are members of the α/β hydrolase superfamily with a β/α/β structural topology. The domains also contain the invariant Ser/His/Asp catalytic triad and the general mechanism of nucleophilic attack by the Ser hydroxyl, mediated by general acid-base catalysis carried out by the His and Asp residues, causing translocation of the polyketide chain onto the TE domain via an acyl-*O*-Ser linkage. A similar nucleophilic attack by an intramolecular or external nucleophile leads to the formation of the final product (Figure 1.20) (Gokhale et al., 1999). PKS TE domains are therefore also capable of forming a variety of linear, hydrolysed polyketides, as well as macrocyclised products. TE domains have also been found to be able to catalyse varied chemical reactions such as carbon-carbon bond forming Claisen-like cyclisations, as well as Dieckmann condensations (Korman et al., 2010).
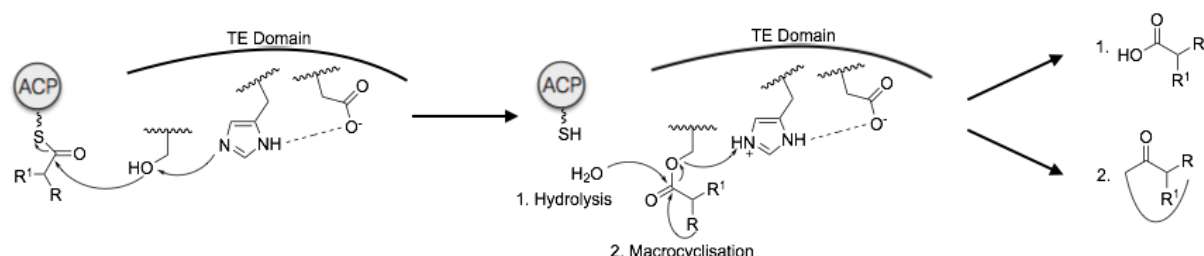
Figure 1.20: The catalytic mechanism utilised by TE domain to achieve the offloading of polyketides. The terminal PKS intermediate is transferred onto a TE domain catalytic Ser

residue, after which internal or external nucleophiles can attack at the carbonyl, releasing the product to yield either a linear or cyclic product, respectively.

In much the same way as NRPS systems, PKSs also employ a variety of offloading domains away from the commonly observed, canonical TE mechanism. Reductase domains function similarly in PKS systems as in NRPS pathways, whilst a study on the enacyloxin producing PKS system showed a dual transacylation offloading mechanism involving a condensation inactive KS^0 domain, and a C domain resembling those observed as part of core NRPS modules (Masschelein et al., 2019). Other, rarer examples of polyketide offloading domains are the lactamases, specialised AT domains, and pyridoxal 5'-phosphate (PLP) dependent enzymes observed in few fungal and bacterial pathways (Gerber et al., 2009; Xie et al., 2009). These varied domains utilise differing mechanisms and often result in condensation with external molecules or the generation of novel chemical structures from the polyketide scaffold, therefore extending the diversity of compounds produced and the chemical space occupied, helpful in the targeting of a heterogeneous set of intracellular molecular constructs.

1.6 Standalone enzymes

1.6.1 Standalone enzymes of NRPS biosynthetic systems

The chemical diversity of NRP scaffolds, whilst greatly dependent on the NRPS systems that synthesise them, is also derived from the action of a series of standalone or *trans*-acting enzymes that play significant roles at almost every stage of the biosynthetic process. These individual domains contribute to the biosynthesis of bespoke precursors, modification of chemical intermediates whilst assembly line tethered, as well as chain termination and post-assembly line tailoring steps, further diversifying the compounds formed (Walsh et al., 2001).

The majority of peptides biosynthesised by NRPS systems are composed primarily of natural amino acids, synthesised as part of primary metabolic pathways for use in a variety of cellular processes, chiefly the formation of proteins by ribosomes. However, as mentioned previously, NRPs are often heavily modified and include diverse chemical motifs, not found in the synthesis of peptides by other routes. One way of incorporating diversity into the final scaffolds is the activation and usage of bespoke precursors. NRPS BGCs normally contain the requisite genes encoding for the enzymatic systems that possess the catalytic activity for the creation of the altered precursor from available, primary metabolic compounds (Süssmuth and Mainz,

2017). The biosynthetic systems producing a large proportion of the complex precursors used in many NRPS pathways are unstudied, however, a few well characterised examples involving the use of standalone enzymes in the creation of these precursors exist (discussed further in section 5.1.3).

Specialised standalone enzymes can also be functional during biosynthesis of the main peptide chain by the assembly line NRPS. In this context, varied chemical groups are enzymatically added to a PCP-bound precursor, prior to C domain action and elongation of the growing peptide chain. Prominent examples of such domains are hydroxylases, halogenases, and methyltransferases. As occurs in the regular function of NRPS modules, an amino acid is adenylated by an A domain, and loaded onto the downstream PCP. Following loading of the PCP, the bound amino acid can be chemically modified, resulting in functional groups, usually important for the bioactivity of the final compound, being added at varied positions of the amino acid. Hydroxylation of the β-carbon of amino acids is a hallmark of many NRPs, and is relatively widely observed (Yin and Zabriskie, 2004; Haltli et al., 2005; Strieker et al., 2009). Overall, hydroxylase activity proceeds via target amino acid binding within the active site, alongside α-ketoglutarate and an Fe(II) ion. A series of recombination and decomposition reactions lead to the activation of the β-methylene group, allowing subsequent stereospecific transfer of a generated hydroxyl group, and formation of the β-hydroxylated amino acid (Figure 1.21). Halogenase enzymes function through a different, complex catalytic route, and generally result in the chlorination of aromatic amino acids such as Tyr or Trp (Zehner et al., 2005; van Pée and Patallo, 2006). Trp halogenases from these systems utilise a reduced flavin cofactor, produced by a flavin reductase as part of the two-component halogenase, molecular oxygen, and a halide ion (Cl-) to generate hypochlorous acid (HOCl) (Yeh et al., 2007). The active site HOCl is captured by a neighbouring Lys residue to generate a stable, covalent N-Cl bond on the ε-NH2, which then acts as the chlorinating agent upon entry of Trp into the active site (Figure 1.21). Structural analysis of PrnA (pyrrolnitrin) and RebH (rebeccamycin) enzymes highlighted the presence of two distinct binding modules specialised for binding the flavin cofactor and the Trp substrate, respectively (Dong et al., 2005). The binding modules were separated by a 10 Å tunnel through which the generated HOCl travels before capture by the active site Lys and delivery to the bound Trp. This represented a novel chlorination mechanism and the expansion of the mechanistic repertoire employed by natural product biosynthetic

systems. These alterations add further diversity to the NRPs produced, as well as potentially providing novel enzymes for combinatorial biosynthetic applications.

A

B
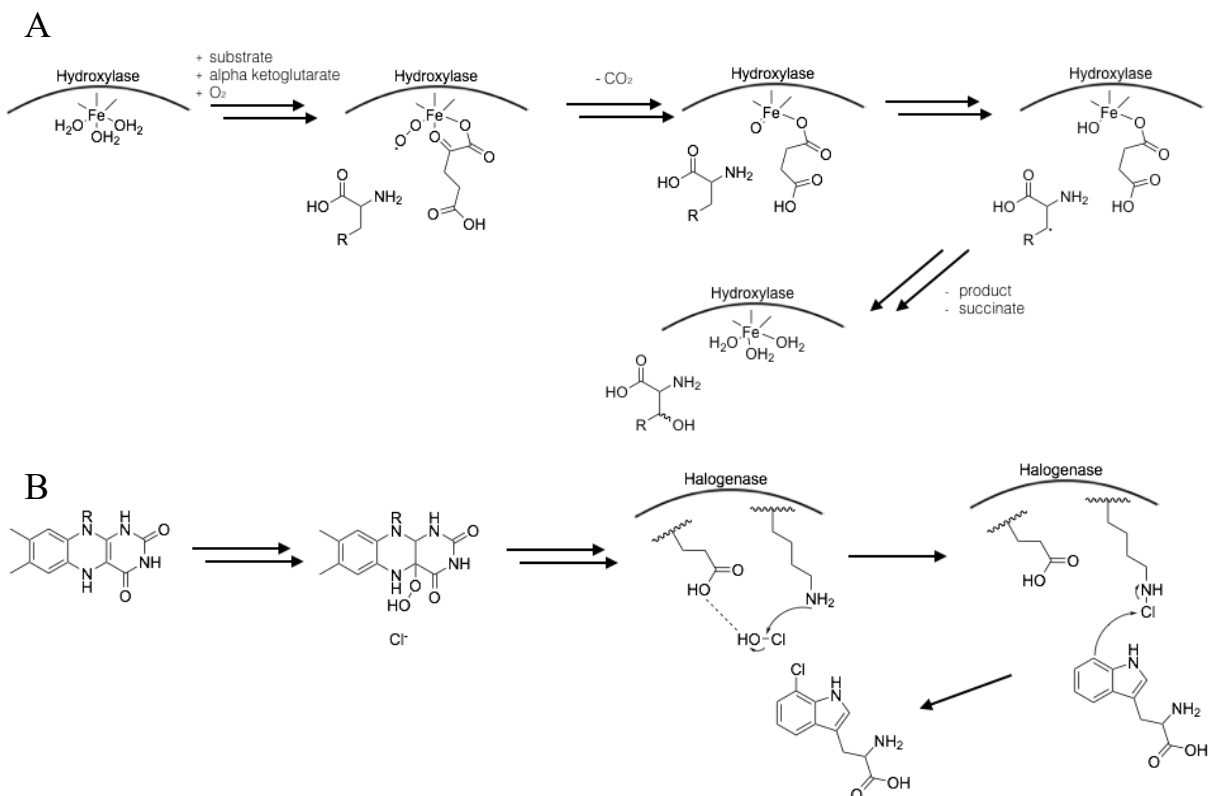
Figure 1.21: A) Overview of the recombinations and decompositions involved in the hydroxylation of amino acid substrates by iron- α-ketoglutarate dependent dioxygenase enzymes. The mechanism centres on a complexed Fe(II) ion and involves the conversion of α-ketoglutarate to succinate and oxygen to carbon dioxide, with concomitant β-hydroxylation of the substrate. B) Overview of the mechanism of halogenation of aromatic amino acids, such as Trp. FAD is reduced, usually by a coupled enzyme partner, followed by formation of a HOCl intermediate, which acts as the chlorinating agent.

A final enzymatic methodology employed by NRPS systems in the diversification of chemical scaffolds is the late-stage modification of compounds released from assembly lines. Often, if a late-stage tailoring enzyme is present within a pathway, the catalysed modification is required for bioactivity. One of the most commonly observed post-NRPS tailoring enzymes are glycosyltransferases, as utilised in the biosynthesis of glycopeptide antibiotics such as vancomycin and teicoplanin (Li et al., 2004). The glycosyltransferases GtfD and GtfE construct the L-vancosaminyl-1,2- D-glucosyl disaccharide by sequential addition to the phenolic hydroxyl of the cross-linked vancomycin core (Losey et al., 2001; Mulichak et al., 2004). The glycosyltransferases generally adopt a bi-lobed structure with aglycone and sugar binding

domains. The enzyme activates the electrophilic sugar and binds the nucleophilic moiety of the synthesised aglycone, catalysing transfer of the sugar, glycosylation of the peptide, and further diversification of the chemical scaffold.

1.6.2 Standalone domains involved in polyketide biosynthesis

Polyketides are generally heavily modified scaffolds, with the majority of the chemical modifications carried out by accessory domains forming part of the PKS modules. The majority of the accessory domains (KR, DH, ER) alter the oxidation state at the β position, with other catalytic activities, such as methyl transfer, also observed on occasion. Further diversity in the metabolites produced by PKS systems is added by standalone domains, which, similar to NRPS pathways, function in precursor biosynthesis and, commonly, in post-PKS tailoring (Olano et al., 2010).

Most PKS extender modules, and the AT domains encoded within them, activate and utilise mCoA or mmCoA, and these molecules, upon decarboxylative condensation, are used as the acceptors in the extension of the growing polyketide chain by two carbon units. Varied precursors, however, can often be utilised at specific points within polyketide biosynthesis, with the diverse substrates usually synthesised and provided by crotonyl-CoA carboxylase-reductase (CCR) enzymes. CCRs were initially observed to catalyse the final step in butyryl-CoA formation and subsequently, the reductive carboxylation of crotonyl-CoA to ethylmalonyl-CoA within central metabolic pathways (Wallace et al., 1995; Erb et al., 2009). The enzymes were then observed within secondary metabolic pathways, initially the pathway synthesising monensin A, but subsequently found in multiple other PKS and hybrid pathways (Liu and Reynolds, 2001). CCRs are members of the MDR family of enzymes and are capable of carrying out carboxylation and reduction reactions, with CO_2 and NADPH the usually preferred cofactors for each reaction, respectively. Since the initial discovery of CCRs, multiple isoforms have been observed in PKS pathways, functional on varying CoA substrates and providing diverse alkylmalonyl-CoA products for use by extension modules (Figure 1.22) (Erb et al., 2007; Wilson and Moore, 2012). The overall mechanism of CCR function involves hydride donation from NADPH to the β position of the α,β-unsaturated compound, followed by carboxylation at the α position, resulting in the formation of the alkylmalonyl-CoA product. CCRs have also, interestingly, been observed to be able to catalyse enoyl reduction in the absence of CO_2 as the reactive species for carboxylation, however, overall reductive

carboxylation is the preferred mechanism of action by the enzymes. A further enzymatic route towards the formation of diverse alkylmalonyl-CoA precursors was recently discovered, and involves the use of acyl-CoA carboxylase (YCC) enzymes. This group of enzymes, initially observed to contribute to diverse precursor formation in the stambomycin PKS pathway, also catalyse the formation of the commonly utilised mCoA and mmCoA substrates (Ray et al., 2016). YCCs utilise ATP and HCO_3^- to catalyse the carboxylation of CoA thioesters to varied alkylmalonyl-CoAs; pentylmalonyl-CoA is formed by the YCC enzyme for use in the stambomycin biosynthetic pathway. These enzymes give rise to varied precursors for use in PKS extender modules, thereby diversifying the polyketide products formed.
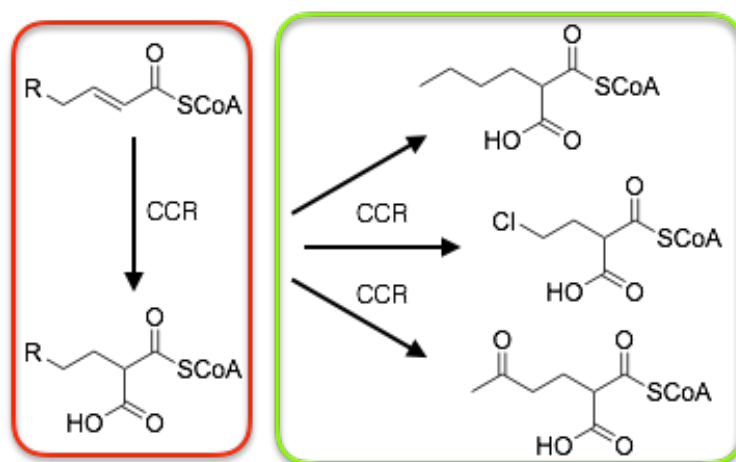
Figure 1.22: The overall activity of crotonyl-CoA carboxylase-reductase enzymes in the creation of varied precursors for use in PKS pathways. The overall reaction involves reduction of the α,β-unsaturated alkene, with carboxylation at the α position. The red box illustrates the overall reaction catalysed by CCR enzymes and the green box shows a few of the varied acyl-CoAs observed to have been originated from the activity of PKS pathway specific CCR enzymes.

Post-PKS tailoring enzymes are prolific among many PKS systems and, due to the diversity of proteins observed, are capable of adding a range of chemical moieties and functional groups. Standalone reductases with mechanisms similar to the module encoded KR domains are often observed as post-PKS enzymes, whilst oxygenases, such as cytochrome P450 (CYP450) mono- or dioxygenases, are also widely prevalent (Rix et al., 2002). The erythromycin and pikromycin biosynthetic pathways both contain late-stage, post-PKS CYP450 enzymes. These CYP450s are specialised in binding macrolide compounds and activating particular carbon atoms within the macrocycle, subsequently catalysing stereospecific hydroxylation at that position (Cupp-

Vickery and Poulos, 1995). A number of CYP450s have also been shown to have relaxed substrate specificity, useful in combinatorial biosynthesis of new natural products. Transferase enzymes are also extremely common in PKS systems and include a variety of enzymes. Examples of glycosyltransferases, methyltransferases, acyltransferases, prenyltransferases, and aminotransferases have all been relatively widely reported (Tang and McDaniel, 2001; Skiba et al., 2016). Many clinically important PKS synthesised products are glycosylated (erythromycin, doxorubicin, avermectin) and the glycosyltransferases, functionally similar to their NRPS counterparts, are therefore important for ultimate compound activity. The remaining transferases are capable of adding various groups to polyketide scaffolds and the catalytic mechanisms of many of the enzymes have been exploited for combinatorial biosynthetic applications, often yielding novel products with interesting applications.

1.7 The biosynthesis of antimycins

1.7.1 Introduction to antimycins

The antimycins are a family of compounds with potent bioactivity against fish, fungi, insects, and nematodes. Recently, however, much attention has focussed on the anticancer properties of various derivatives of naturally produced antimycins, where the mode of action is the potent and selective inhibition of the anti-apoptotic proteins Bcl-2 and Bcl-x_L, thereby promoting programmed cell death (Seidel et al., 2019). The mode of action of antimycins is the inhibition of the cytochrome c reductase complex in the electron transport chain, causing respiration to cease prematurely (Kröger et al., 1973). Chemically, antimycins are a large group of compounds all containing a rare 3-formamidosalicylic acid (3-FSA) moiety and a central lactone ring, with the ring ranging in size from 9 to 18 atoms (9, 12, 15, or 18 membered rings) (Figure 1.23) (Liu et al., 2016). The antimycins themselves are the simplest of the group of compounds, comprising a nine-membered dilactone scaffold with two variable groups at positions 7 and 8 of the large ring. The major uses of antimycins are as commercial piscicides and as a research tool to study the function of cytochromes, as well as the Bcl-2 and Bcl-x_L proteins (Tzung et al., 2001).
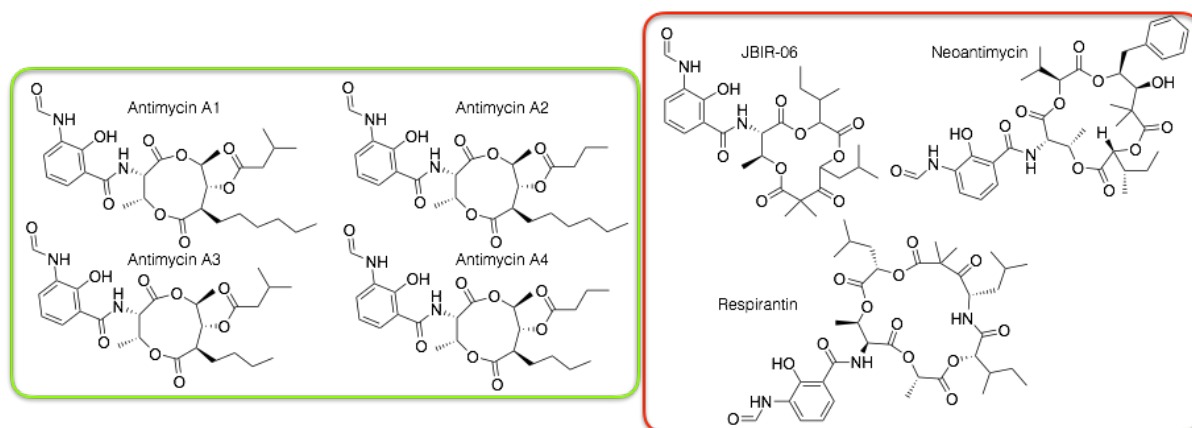
Figure 1.23: Chemical structures of antimycin-type depsipeptides. Structures in the green box are illustrative of the four most commonly observed nine-membered ring compounds biosynthesised *in vivo*; antimycins A1 – A4. The structures in the red box show a selection of ring expanded antimycin-type depsipeptides, including the 12-membered ring JBIR-06; the 15-membered ring containing neoantimycin; and respirantin, which features an 18-membered central ring.

The antimycin biosynthetic gene cluster encodes the majority of the necessary proteins required for production of the compounds, including regulators, starter unit biosynthetic genes, and the NRPS and PKS megasynthases. The BGCs within *Streptomyces* species are either S-, I-, or L-form and therefore composed of 15, 16, or 17 genes respectively (Seipke and Hutchings, 2013a). The *antP* and *antQ* genes, encoding a kynureninase and a phosphopantetheinyl transferase respectively, are not found in S-form clusters and are likely compensated for by the use of enzymes from separate pathways. The gene cluster is arranged into four polycistronic operons: *antAB*, *antCDE*, *antFG*, and *antHIJKLMNO*, with each operon important in the functionality of the complete pathway and the biosynthesis of antimycins.

1.7.2 Regulation of the *ant* biosynthetic gene cluster

The production of secondary metabolites, particularly in species of *Streptomyces*, is closely related to development and morphological differentiation. Correct temporal production of the proteins required for biosynthesis is essential for the synthesis of the natural product; for example, the production of an NRP is dependent upon the presence of the NRPS proteins, and any other proteinaceous factors required for biosynthesis, being correctly present in time and space. Therefore, regulation of the cluster and gene expression is of pivotal importance and cluster encoded regulators generally play key roles in compound production.
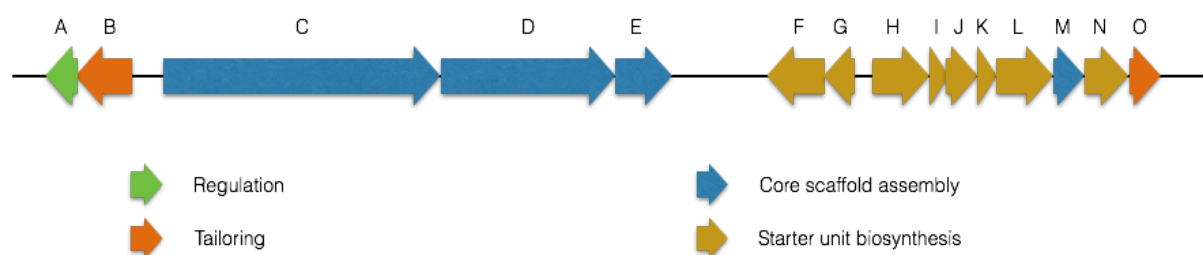
Figure 1.24: Schematic representation of an S-form *ant* biosynthetic gene cluster, such as the cluster found in various genomes of *S. albus*. The pathway function of each gene product is indicated by the colour and gene names are shown above the representation of each gene. I- and L-form BGCs are also found within various other *Streptomyces* spp. and contain either or both of *antP* (kynureninase encoding) and *antQ* (phosphopantetheinyl transferase encoding).

Early on, the regulation of antimycins was noted as unusual. Expression of the BGC (Figure 1.24) occurs extremely early during growth, conversely to the expected dogma of secondary metabolite production correlating with aerial growth (Seipke et al., 2014). In addition, antimycins were not produced until later in the growth cycle, exhibiting a lag between BGC expression and detection of the compounds. Unusually, the gene cluster encodes an extracytoplasmic function (ECF) RNA polymerase sigma (σ) factor, σ^{AntA}. Although ECF σ factors are normally observed in the regulation of genes required for stress responses or differentiation, σ^{AntA} controls the expression of *antFG* and *antHIJKLMNO*, with degradation of the regulatory sigma factor unusually controlled by the ClpXP protease, rather than a cognate anti-σ factor (Bilyk et al., 2020). The remaining two operons (*antAB* and *antCDE*), containing the genes for the NRPS and PKS synthases as well as the gene for σ^{AntA} itself, were regulated in S-form *ant* clusters by the LuxR-type regulator FscRI, a regulator from the BGC for the polyene antifungal, candicidin (McLean et al., 2016). Therefore, FscRI, which indirectly regulates its own production, activates the expression of *antBA* and *antCDE*, followed by σ^{AntA} mediated activation of the expression of *antGF* and *antHIJKLMNO*, producing the remainder of the proteins required for the production of antimycins. σ^{AntA} can then be directly proteolysed by the ClpXP protease, as shown by *in vitro* proteolysis and *in vivo* strategies, in a novel methodology for the control of cellular sigma factor levels.

1.7.3 Biosynthesis of antimycin-type depsipeptides

Antimycins are produced by a hybrid NRPS-PKS system, with extensive pre-assembly line biosynthesis of the rare 3-FSA starter unit. Biosynthesis is initiated by the enzymatic breakdown of the amino acid Trp, by the pathway specific Trp-2,3-dioxygenase, AntN. The catalytic action of AntN results in the transfer of molecular oxygen and opening of the indole ring, yielding *N*-formyl- L-Kyn (Thackray et al., 2008). A kynureninase, either pathway encoded in L- and certain I-form gene clusters or from primary metabolism, then converts the Kyn into anthranilate, an *ortho*-substituted benzene ring with a carboxylic acid and an amine. AntG is a free-standing carrier protein, resembling an NRPS PCP domain, that acts as a shuttle between the precursor biosynthetic enzymes and the main hybrid assembly line (Liu et al., 2015). The carrier protein is loaded by the action of the pathway acyl-CoA ligase, AntF, which activates the formed anthranilate and catalyses the thioester bond formation required for loading the carrier protein. A chemically interesting step in the biosynthesis of the antimycin precursor is the formation of 3-aminosalicylate from the AntG loaded anthranilate. The step involves a multicomponent oxygenase enzyme complex, AntHIJKL, which catalyses an unprecedented 1,2 shift with concomitant oxidation (Schoenian et al., 2012). This occurs through the formation of an epoxide intermediate, which breaks down to cause a shift of the thioester and hydroxylation at C2, forming 3-aminosalicylate. AntO is a lipase homologue responsible for the installation of the formyl group at the free amine of the 3-aminosalicylate unit, forming the 3-FSA observed in mature antimycins. AntO is also thought to play a role in deformylation of *N*-formyl-L-Kyn prior to the action of the kynureninase to result in anthranilate formation. Subsequent to the action of AntO, the 3-formamidosalicyloyl-*S*-AntG complex is presented to the first module of the NRPS AntC, commencing biosynthesis of the dilactone ring (Yan et al., 2012).
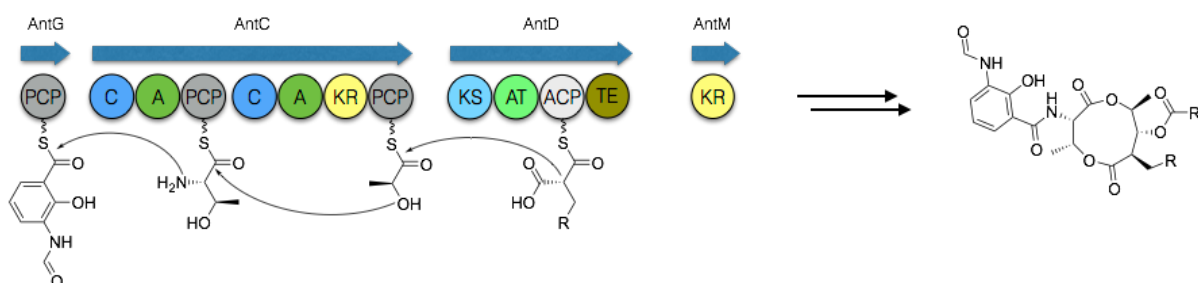
Figure 1.25: The biosynthetic pathway involved in the formation of antimycin-type depsipeptides. Starter unit biosynthetic proteins synthesise the 3-FSA unit, before the large assembly line NRPS and PKS proteins continue the biosynthesis, forming the dilactone ring. Post assembly line tailoring reactions complete the formation of the antimycins.

The main assembly line involved in the biosynthesis of antimycins is composed of two NRPS modules and a single PKS module, whose catalysis is followed by the action of a number of tailoring enzymes (Figure 1.25) (Liu et al., 2016). The first NRPS module of AntC has the domain organisation C-A-PCP and functions by the A domain mediated selection of Thr, followed by loading of the modular PCP with the selected amino acid. The C domain then catalyses the condensation of the activated Thr with the AntG tethered 3-aminosalicylate starter unit. Module 2 of the hybrid antimycin forming pathway is composed of C, A, KR, and PCP domains, and utilises a quirk of NRPS biosynthesis with its capability in activating an α-keto acid. The A domain selects pyruvate, a central component of a number of primary metabolic pathways including glycolysis, and loads the acid onto the PCP domain. The KR domain, not commonly observed as modular parts of NRPS synthases, catalyses the stereospecific conversion of the pyruvate keto functionality to a secondary alcohol, which can then be used as the nucleophile in the condensation reaction catalysed by the C domain. The condensation with the intermediate held on the PCP of module one causes formation of a depsipeptide, extension of the growing chain, and concomitant translocation of the complete intermediate to the second PCP domain (Sandy et al., 2012). The PKS, AntD, is formed of the domains KS-AT-ACP-TE. The AT domain, normally capable of activating malonyl- or methylmalonyl-CoA in canonical PKS systems, can activate a multitude of carboxylated acyl-CoA derivatives, leading to diversity in the compounds produced. The varied acyl-CoA precursors are provided by AntE, a crotonyl-CoA reductase capable of accepting extended and varied derivatives of crotonyl-CoA as substrates, thereby yielding alkylated malonyl-CoA variants (Zhang et al., 2015). The KS domain catalyses the decarboxylative condensation between the ACP tethered acyl-CoA and the intermediate held on the PCP of the preceding NRPS module. AntM subsequently catalyses reduction at the β-keto group of the terminal intermediate, before the module encoded TE domain promotes macrolactonisation, by attack of the Thr hydroxyl into the acyl-*O*-TE ester bond, releasing the dilactone.

Upon release from the terminal ACP of the assembly line, the antimycin scaffold is modified by one tailoring enzyme, AntB, which is encoded by every antimycin BGC. AntB is a standalone acyltransferase, which catalyses an esterification reaction at the C8 position of the antimycin dilactone ring (Sandy et al., 2013). The enzyme has been shown to be extremely promiscuous in its substrate tolerance and acceptance, and therefore is capable of activating a

variety of acyl-CoAs, subsequently generating diversity at the C8 position of the macrocyclic ring. The action of the tailoring domain AntB completes the biosynthesis of the diverse antimycins, which are then transported out of the cell into the surrounding bacterial habitat.

1.8 The biosynthesis of surugamides

1.8.1 Introduction to surugamides

The surugamides are a collection of structurally and chemically similar peptides, biosynthesised by NRPS systems and isolated from various *Streptomyces* species, both marine and terrestrial. The majority of the surugamides are cyclic and octapeptidic and the first compounds, surugamides A-E, were discovered in 2013 from *Streptomyces* sp. JAMM992 (Takada et al., 2013). Later, surugamide F, a linear decapeptide, was also found to be produced by strains containing the BGC (Ninomiya et al., 2016). Surugamide A, the prototypical member of the cyclic surugamides formed, was initially found to have good bioactivity against Gram-positive indicator organisms, and further studies have shown a range of other promising bioactivities (Takada et al., 2013; Matsuda, Kuranaga, et al., 2019). The surugamides have been shown to have anticancer activities by the inhibition of bovine cathepsin B, often involved in tumour metastasis, whilst a surugamide derivative, acyl-surugamide A, possesses anti-fungal activity. The molecular mode of action of the surugamides exhibiting bioactivity is unknown, however, the varied bioactivities indicate multiple potential molecular targets. Chemically, the surugamides are polypeptides, connected by backbone amide bonds, and containing many non-natural amino acids, including stereochemically altered residues and amino acids with additional chemical groups added (Figure 1.26). The majority of the surugamides, except surugamide F, are head to tail cyclised post biosynthesis, yielding macrocyclic compounds (Kuranaga, Fukuba, et al., 2018).

Surugamide A   Surugamide F
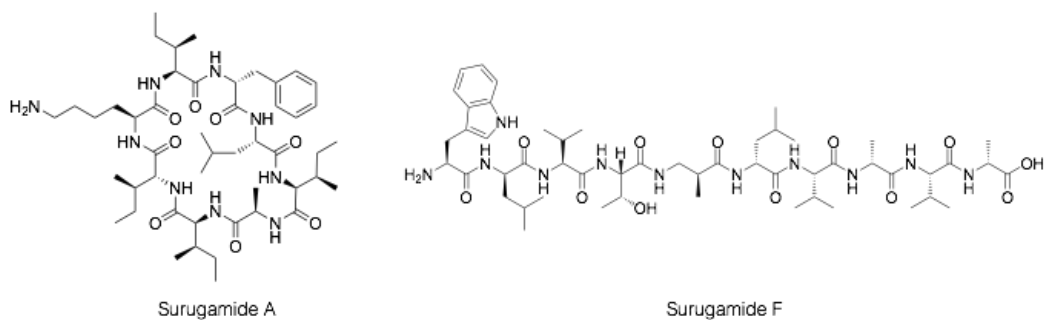
Figure 1.26: Chemical structures of the dominant surugamides produced *in vivo*; surugamide A, a cyclic octapeptide, and surugamide F, a linear decapeptide. Surugamides B – E are also produced in varying quantities, and are shunt, or minor, products of the surugamide A biosynthetic system, where one of the multiple Ile residues is substituted for a Val residue.

The surugamide BGC, much like many other specialised metabolite clusters, encodes the necessary genes required for production of the compound, as well as any other required processes, such as metabolite transport (Ninomiya et al., 2016). The *sur* BGC contains genes encoding regulators, transporters, and a number of genes for hypothetical proteins, whose function is unknown. The large NRPS genes form the majority of the cluster and encode the machinery required to form the multiple surugamide peptides. A number of other biosynthetically important genes make up the remainder of the cluster and ensure the smooth running of the NRPS system, and efficient production of the surugamides.

1.8.2 Regulation of the *sur* biosynthetic gene cluster

Regulation of the *sur* BGC (Figure 1.27) is, on the whole, relatively uncharacterised, and the specific pathways and mechanisms that lead to activation of the cluster and the production of surugamides are mostly unknown. Recently however, SurR was found to be a pathway specific transcriptional regulator of the GntR family, which was capable of modulating the level of production of the surugamides (Xu et al., 2017). Exposure of *Streptomyces albus* J1074, which contains the *sur* BGC, to ivermectin or etoposide resulted in the 2-fold downregulation of *surR* expression and subsequent increase in the surugamide titre. Deletion of *surR* had a similar effect and increased production titres, confirming the role of the encoded protein as a repressor of the *sur* BGC. Further regulatory proteins are encoded within the *sur* BGC, and further study is required to elucidate the mechanistic basis of regulation carried out by SurR and the remaining pathway specific regulators.
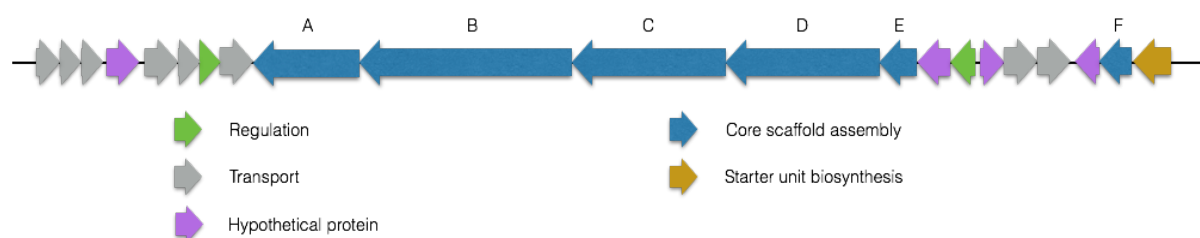
Figure 1.27: Schematic representation of the *sur* BGC. The functions of the gene products are denoted by the colours of the genes. The main scaffold assembly proteins (four NRPS genes, cyclase, and type II thioesterase) are noted with the gene names above the representative genes.

1.8.3 Biosynthesis of surugamides

The biosynthesis of the surugamides is carried out by dedicated non-ribosomal peptide synthetase assembly lines. The *sur* BGC contains four NRPS genes; *surA*, *surB*, *surC*, and *surD*, which, in total, correspond to eighteen biosynthetic NRPS modules. Initially a peptide with eighteen amino acids would be expected from such a system, if the standard collinearity principle was obeyed. The surugamide system, however, employs a unique organisation of the encoded NRPS proteins; SurA and SurD form one NRPS system resulting in the formation of the octapeptidic surugamides, whilst SurB and SurC form a separate system containing ten modules and forming the linear decapeptide surugamide F (Figure 1.28) (Ninomiya et al., 2016; Kuranaga, Matsuda, et al., 2018). The synergy and interplay between the two compounds, as well as the NRPS systems synthesising them, is largely unknown.
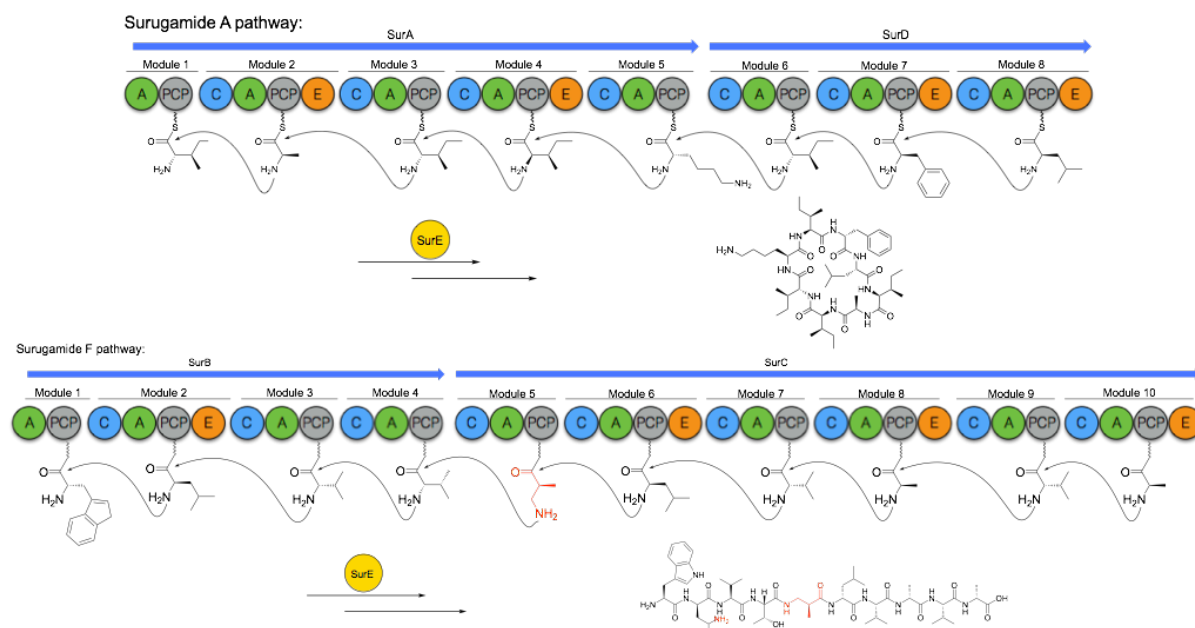
Figure 1.28: Overview of the biosynthesis of surugamide A and surugamide F. The amino acid accepted by each A domain is shown attached the PCP domain of the relevant module. The NRPS proteins work canonically, following the collinearity principle, in the biosynthesis of peptides, followed by SurE cyclase mediated release of peptides. Structures in red are atypical amino acids.

The biosynthetic pathway synthesising surugamide A (SurA + SurD) contains eight modules, the first of which is composed of A and PCP domains. The remaining seven extension modules all contain the core set of NRPS domains; C, A, and PCP. Modules 2, 4, 7, and 8 all terminate in E domains, which alter the stereochemistry of activated and tethered amino acids at the α position. The A domain of the first module activates Ile and tethers it to the PCP domain, whilst the second module adenylates Ala. The C domain of module 2 catalyses condensation of the Ala residue with the Ile residue bound to the upstream PCP, causing chain elongation and translocation. The Ala residue is then epimerised by the downstream E domain, resulting in the formation of an Ile- D-Ala dipeptide. Biosynthesis continues with the activation of Ile by module 3 and C domain catalysed elongation, followed by the addition of another Ile residue by module 4. The Ile added by module 4 is epimerised by the module encoded E domain, and biosynthesis continues and concludes with the activation and addition of Lys, Ile, Phe (epimerised to D-Phe), and Leu (epimerised to D-Leu), by modules 5 – 8, respectively. Upon formation of the terminal linear intermediate on the PCP of module 8, the mature surugamide A is released from the assembly line by macrocyclic release, causing head to tail cyclisation of the compound, and re-priming of the assembly line for further compound formation (Thankachan et al., 2019; Zhou et al., 2019).

Surugamide F is synthesised by the action of the NRPS assembly line formed by SurB and SurC, a system composed of ten modules. The first module serves as the initiation module, so contains only A and PCP domains, and activates Trp before tethering to the modular PCP. Leu is adenylated by the A domain of module 2, a process which precedes the C domain mediated condensation of the bound Leu with the Trp residue activated by the first module. The E domain subsequently epimerises the Leu residue, resulting in the formation of a PCP bound Trp- D-Leu dipeptide. Val and Thr are added by the action of modules 3 and 4, before the A domain of module 5 activates the unnatural amino acid 3-amino-2-methylpropionic acid (AMPA), a β-amino acid with a methyl side chain. The non-natural AMPA is, however, utilised in the same way by the assembly line; following its activation and PCP tethering by the A domain, the C domain catalyses chain elongation with the peptide held on the upstream PCP. Modules 6 – 10 activate and add Leu (epimerised to D-Leu), Val, Ala (epimerised to D-Ala), Val, and Ala (epimerised to D-Ala), respectively, leading to formation of the decapeptide bound to the terminal PCP domain. Offloading of surugamide F was thought to occur by a simple hydrolytic mechanism, resulting in release of the linear peptide. However, the macrocyclic release

mechanism used to form surugamide A, suggests the potential of a more complex release mechanism or the utilisation of different release factors for the different surugamide peptides (Matsuda, Kobayashi, et al., 2019).

1.9 Aims and objectives

Outline of the major aims of this thesis:

Data generated during this PhD project spans two major biosynthetic pathways and three results chapters:

Characterisation of antimycin biosynthesis and bioengineering of the assembly line for the creation of novel antimycin-type compounds:

- Chapter 3: CRISPR/Cas9 engineering of polyketide synthase substrate utilisation
- Chapter 4: The structure and function of a standalone β-ketoreductase that acts concomitantly with biosynthesis of the antimycin core scaffold

Characterisation of the offloading enzyme in surugamide biosynthesis:

- Chapter 5: A standalone cyclase offloading strategy for assembly line release of surugamides

Overall, the aim of this PhD thesis was to address fundamental questions regarding the basic biology of the biosynthesis of polyketide and non-ribosomal peptide scaffolds. The systems of interest are the antimycin and surugamide pathways, whose biosynthetic gene clusters can both be found in the strain of interest, *Streptomyces albus* S4. *S. albus* S4 was first isolated from the nests of leaf cutting ants, which, interestingly, employ antibiotic producing Actinobacteria to protect their fungal cultivar from invading fungal species (Seipke et al., 2011). The antimycin and surugamide biosynthetic systems have been variably studied for numerous years, however, questions remain regarding the specific biosynthetic mode of action of proteins involved in the pathway. An understanding of the functionality of the enzymes involved in biosynthetic pathways would contribute to the ultimate goal of combinatorial biosynthesis and using engineered systems for the production of novel natural products, with new or interesting clinically relevant activities.

The antimycin biosynthetic pathway is composed of two NRPS modules and a terminating PKS module. The PKS module is composed of canonical KS, AT, and ACP domains, as well as the terminating thioesterase domain, and the AT domain has been shown to be promiscuous

in its substrate acceptance and usage, therefore, resulting in the formation of antimycins with varied alkyl groups at the C7 position of the macrocyclic ring. Numerous studies have focussed on increasing the diversity of substrates utilised by AT domains, either by mutation or enzymatic swaps, in attempts to produce metabolites with altered groups at specific positions of the final products. The aims of this study were two-fold. Initially, a generalisable CRISPR/Cas9 platform for the creation of site-directed mutants on cloned BGCs was developed, primarily for targeting the AT domain active site for mutagenesis. Secondly, specific mutants created in the AT domain motif III region were evaluated for their selective incorporation of alkylmalonyl-CoA precursors, and therefore, the production of certain antimycin-type depsipeptides, potentially not naturally observed as products from the antimycin biosynthetic pathway. The antimycin BGC also encodes a standalone ketoreductase, AntM, which, in previously carried out *in vitro* biosynthesis reactions, was shown to be essential for the production of antimycins of any variety. The goal was therefore to characterise the AntM ketoreductase both *in vivo* and *in* vitro to provide a robust characterisation of its biochemical functionality within the context of antimycin biosynthesis.

The surugamide BGC is also found in the genome of *S. albus* S4 and is responsible for the production of two major peptide products, surugamide A and surugamide F. Surugamide A is a cyclic octapeptide, produced by the NRPS megasynthases SurA and SurD, and surugamide F is a linear decapeptide produced by the NRPSs SurB and SurC. NRPSs canonically contain a thioesterase domain as part of the synthetase, usually at the extreme C terminus, which is responsible for offloading the peptide chain and forming the mature peptide product. Examination of the SurC and SurD NRPSs showed the lack of a TE domain, with both proteins terminating in epimerase domains, not known to possess peptide offloading activity. This study, therefore, aims to determine the offloading factor responsible for the creation of mature surugamide A and F. Also of interest is the fact that the two major surugamide peptides are produced with differing structural outcomes; surugamide A is cyclic and octapeptidic and surugamide F is linear and formed of ten amino acids. The goal of this project was to characterise any offloading domain for substrate acceptance and promiscuity, as well as analysing the enzymatic activity of the domain.

Chapter 2 – Materials and Methods

2.1 Instrumentation and Equipment

Bacterial cell cultures were incubated in either a Forma Scientific Orbital Shaker at 30°C or a New Brunswick Scientific Excella E25 shaker incubator set to 37°C. Agar plates with spread bacterial cultures were incubated at either 30°C or 37°C in a Memmert incubator and a Leec Classic Incubator, respectively. All centrifugation steps were undertaken in a Prism Microcentrifuge or a Sigma 3-18K Centrifuge, whilst all weighing and balancing was carried out on A&D GX-2000 and Denver Instrument SI-64 scales. Cell culture optical densities were measured on an Amersham Biosciences Spectrophotometer, and subsequent bacterial cell stocks were stored at -80°C in a Sanyo Ultra-Low VIP Freezer. All temporal incubations, such as restriction digests and heat shocks, were carried out in Grant Water Baths set to the required temperature, or in the incubators mentioned above. Polymerase Chain Reactions (PCRs) and Circular Polymerase Extension Cloning (CPEC) reactions were accomplished in a Bio-Rad T100 thermal cycler and DNA products were analysed using agarose gel electrophoresis in Bio-Rad gel tanks. UV visualisation of agarose gels occurred in a Syngene Genius Geldoc system or on a Vernier BlueView transilluminator, DNA bands were extracted using a Qiagen Gel Extraction kit, and quantitation of DNA products was undertaken on a Geneflow Nanophotometer. DNA, plasmids and cosmids were introduced into bacterial cells by heat shock, or electroporation using Cell Projects Electroporation Cuvettes and a Bio-Rad Gene Pulser X Cell electroporator. General samples were mixed by vortex using a Fisons Whirlimixer and the pH checked by an Orion 410A pH meter. Antibiotic stock solutions in water were sterilised by passing through a Sartorius Solutions filter steriliser and a general sterile environment was provided by a Bio48 Faster Safety Cabinet. Previous to protein purification, bacterial cells were lysed using a Constant Systems Cell Disruptor or a Bandelin Sonopuls sonicator, gel filtration chromatography was undertaken on a Bio-Rad FPLC system, or, primarily, on a GE Healthcare ÄKTA Pure system, whilst proteins were analysed by SDS-PAGE using a Bio-Rad Mini Protean Tetra Cell electrophoresis kit and imaged on a Bio-Rad Gel Doc XR gel imaging system, or a generic white light box. Mass spectrometric analysis of protein samples and chemical extracts was accomplished using a Bruker MaXis Impact mass spectrometer with a Thermo Scientific Dionex Ultimate 3000 HPLC or an MSQ Plus LCMS machine. HPLC analyses were carried out on an ECOM system (ECP2010 pump, ECD2800 detector, HT4000L autosampler, Clarity software).

2.2 General and Biological Materials

2.2.1 General Materials

All general materials were purchased from Merck (Sigma-Aldrich) unless otherwise stated. Solvents for chemical synthesis and extractions were also purchased from Merck (Sigma-Aldrich). Materials required for molecular biology and DNA manipulation were purchased from New England Biolabs. Agar and SYBR Safe agarose gel stain were purchased from Thermo Fisher Scientific. InstantBlue SDS-PAGE gel stain was purchased from Expedeon and Bradford Reagent from Bio-Rad. Ni-NTA resin and IPTG were purchased from Generon. A 50x stock of Tris acetate EDTA (TAE) buffer was purchased from Severn Biotech Ltd. Auto-induction media used for protein expression were purchased from Formedium. Agarose for analytical DNA gels was purchased from Bioline and acrylamide/bis-acrylamide for SDS-PAGE gels was purchased from Bio-Rad. Laemmli buffer for SDS-PAGE sample preparation was also purchased from Bio-Rad. All water used was deionised unless otherwise stated, and nuclease free water was purchased from Invitrogen.

2.2.2 Specific Kits and Enzymes

Miniprep kits for plasmid and cosmid DNA isolation and purification were purchased from New England Biolabs (Monarch plasmid miniprep kits) and QIAquick gel extraction kits were purchased from Qiagen. All restriction enzymes, PCR reagents, and other materials required for molecular biology (such as ligases, phosphatases etc.) were purchased from New England Biolabs, apart from *Taq* polymerase for diagnostic PCR reactions, which was purchased from Promega (GoTaq G2 DNA polymerase).

2.2.3 Growth Media

Lennox broth or agar (LB): 10 g tryptone, 5 g yeast extract, and 5 g NaCl were dissolved in 1 L dH$_2$O and sterilised by autoclaving (15 minutes at 121 °C). For the preparation of Lennox agar, 15 g agar was added prior to autoclaving. When using hygromycin as a selective agent, a no-salt variant of LB was utilised where salt was omitted from the prepared medium.

Soy flour mannitol medium (SFM): 20 g soy flour and 20 g mannitol in 1 L warm tap water was sterilised by autoclaving. SFM agar was prepared by the addition of 20 g agar to the mixture prior to autoclaving.

Tryptic soy broth (TSB): 30 g/L tryptic soy broth (Sigma) was dissolved in dH_2O and sterilised by autoclaving.

International Streptomyces *project medium 2 (ISP2):* 38 g/L ISP2 was added to dH_2O and sterilised by autoclaving to prepare solid ISP2 medium. In order to prepare liquid medium the components of ISP2 (4 g yeast extract, 10 g malt extract, 4 g dextrose) were added separately to 1 L dH_2O and sterilised.

Auto induction medium – super broth: 74.85 g/L auto induction medium was added to 1 L dH_2O, usually prepared in 500 mL volumes, and sterilised by autoclaving.

2.2.4 Antibiotics

Table 1: Antibiotics utilised. All stocks were prepared at 1000x concentration and diluted into samples to obtain the required working concentrations.

Antibiotic	Working concentration (µg/mL)	Solvent used	Supplier
Apramycin	50	Water	Sigma
Carbenicillin	50	Water	Formedium
Chloramphenicol	25	EtOH	Formedium
Hygromycin B	50	PBS	Invitrogen
Kanamycin	50	Water	Sigma
Nalidixic acid	25	0.3 M NaOH	Formedium
Spectinomycin	50	Water	Sigma
Streptomycin	50	Water	Sigma

2.2.5 Buffers

All buffers provided as part of kits, such as miniprep or gel extraction kits, were used without modification and as suggested in manufacturers recommendations. All other buffers utilised are described below.

Lysis buffer: 50 mM HEPES, 500 mM NaCl, 10 mM imidazole, 1 mM EDTA, one protease inhibitor cocktail tablet (/L, Sigma), DNAse I (arbitrary mass), pH 7.5

Ni-NTA wash buffer: 50 mM HEPES, 200 mM NaCl, 1 mM EDTA, 50 mM imidazole, pH 7.5

Ni-NTA elution buffer: 50 mM HEPES, 200 mM NaCl, 1 mM EDTA, 300 mM imidazole, pH 7.5

Gel filtration buffer: 50 mM HEPES, 200 mM NaCl, pH 7.5

Storage buffer: 50 mM HEPES, 200 mM NaCl, 10% v/v glycerol, pH 7.5

SDS-PAGE running buffer (5x): 125 mM Tris-base, 960 mM glycine, 0.5% w/v SDS

Buffer P1: 50 mM Tris-Cl, 10 mM EDTA, 100 µg/mL RNAse A, pH 8.0

2.2.6 Bacterial Strains

Table 2: Bacterial strains used and generated in the course of this study.

Bacterial strain	Description	Reference
E. coli strains		
XL10-Gold	General cloning host	Agilent Technologies
BL21-Gold (DE3)	Protein production host	Agilent Technologies
Rosetta (DE3)	Protein production host	Novagen
ET12567	Non-methylating host for conjugal transfer of DNA into _Streptomyces_ (_dam, dcm, hsdM_); Cam^R	(MacNeil et al., 1992)
GB05-red	RecET recombineering host	(Fu et al., 2012)
BW25113	General cloning and CRISPR host	CGSC
DH5α	FLP recombination host	CGSC
Streptomyces strains		
S. albus S4	_Streptomyces albus_ S4 wildtype	(Barke et al., 2010)

S4 Δ5	Derivative of *S. albus* S4 harboring a complete pathway deletion for the antimycin BGC and mutations that abrogate biosynthesis of candicidin, surugamides, fredericamycin and albaflavenone	(Fazal, Thankachan, et al., 2020)
Δ5 *attB* ΦC31::Cos213	S4 Δ5, *Cosmid 213* insertion, Apr^R	This study
Δ5 *attB* ΦC31::Cos213- HAFH	S4 Δ5, *Cosmid 213* with a HAFH mutation in the AT domain of the antimycin BGC, Apr^R	This study
Δ5 *attB* ΦC31::Cos213- YASH	S4 Δ5, *Cosmid 213* with a YASH mutation in the AT domain of the antimycin BGC, Apr^R	This study
Δ5 *attB* ΦC31::Cos213- AAPH	S4 Δ5, *Cosmid 213* with a AAPH mutation in the AT domain of the antimycin BGC, Apr^R	This study
Δ5 *attB* ΦC31::Cos213- Δ*antB*	S4 Δ5, *Cosmid 213 antB* mutant, Apr^R	This study
Δ5 *attB* ΦC31::Cos213- HAFH-Δ*antB*	S4 Δ5, *Cosmid 213 antB* mutant, with a HAFH mutation in the AT domain of the antimycin BGC, Apr^R	This study
Δ5 *attB* ΦC31::Cos213- YASH-Δ*antB*	S4 Δ5, *Cosmid 213 antB* mutant, with a YASH mutation in the AT domain of the antimycin BGC, Apr^R	This study
Δ5 *attB* ΦC31::Cos213- AAPH-Δ*antB*	S4 Δ5, *Cosmid 213 antB* mutant, with a AAPH mutation in the AT domain of the antimycin BGC, Apr^R	This study
Δ5 *attB* ΦC31::Cos213- Δ*antM*	S4 Δ*5 Cos 213 antM* mutant, Apr^R	This study
Δ5 *attB* ΦC31::Cos213- Δ*antM attB* ΦBT1::pIJ10257-*antM*	S4 Δ*5 Cos 213 antM* mutant complemented with *antM* expressed from the *ermE** promoter; Apr^R, Hyg^R	This study
Δ5 *attB* ΦC31::Cos213- Δ*antM attB* ΦBT1::pIJ10257-*antM*-WW	S4 Δ*5 Cos 213 antM* mutant complemented with *antM,* with Y185W, F223W mutations, expressed from the *ermE** promoter; Apr^R, Hyg^R	This study
Δ5 *attB* ΦC31::Cos213- Δ*antM attB* ΦBT1::pIJ10257-*antM*-WA	S4 Δ*5 Cos 213 antM* mutant complemented with *antM,* with Y185W, F223A mutations, expressed from the *ermE** promoter; Apr^R, Hyg^R	This study
Δ5 *attB* ΦC31::Cos213- Δ*antM attB* ΦBT1::pIJ10257-*antM*-WF	S4 Δ*5 Cos 213 antM* mutant complemented with *antM,* with the Y185W mutation, expressed from the *ermE** promoter; Apr^R, Hyg^R	This study
Δ5 *attB* ΦC31::Cos213- Δ*antM attB* ΦBT1::pIJ10257-*antM*-AW	S4 Δ*5 Cos 213 antM* mutant complemented with *antM,* with Y185A, F223W mutations, expressed from the *ermE** promoter; Apr^R, Hyg^R	This study

Δ5 *attB* ΦC31::Cos213- Δ*antM attB* ΦBT1::pIJ10257-*antM*-AA	S4 Δ*5 Cos 213 antM* mutant complemented with *antM,* with Y185A, F223A mutations, expressed from the *ermE** promoter; Apr^R, Hyg^R	This study
Δ5 *attB* ΦC31::Cos213- Δ*antM attB* ΦBT1::pIJ10257-*antM*- AF	S4 Δ*5 Cos 213 antM* mutant complemented with *antM,* with the Y185A mutation, expressed from the *ermE** promoter; Apr^R, Hyg^R	This study
Δ5 *attB* ΦC31::Cos213- Δ*antM attB* ΦBT1::pIJ10257-*antM*-YW	S4 Δ*5 Cos 213 antM* mutant complemented with *antM,* with the F223W mutation, expressed from the *ermE** promoter; Apr^R, Hyg^R	This study
Δ5 *attB* ΦC31::Cos213- Δ*antM attB* ΦBT1::pIJ10257-*antM*-YA	S4 Δ*5 Cos 213 antM* mutant complemented with *antM,* with the F223A mutation, expressed from the *ermE** promoter; Apr^R, Hyg^R	This study
S4 Δ*surE*	S4, *surE* mutant, Apr^R	(Thankachan et al., 2019)

2.2.7 Plasmids and Cosmids

Table 3: Cosmids and plasmids used and constructed in the course of this study.

Cosmid/Plasmid	Description	Reference
Cosmids		
Supercos1	Cosmid backbone	Stratagene
Cos213 ΦC31Apr	Supercos1 derivative spanning the entire antimycin gene cluster; Apr^R	This study
Cos213 ΦC31Apr- HAFH	Cos213 derivative with HAFH mutation in *antD* AT domain; Apr^R	This study
Cos213 ΦC31Apr- YASH	Cos213 derivative with YASH mutation in *antD* AT domain; Apr^R	This study
Cos213 ΦC31Apr- AAPH	Cos213 derivative with AAPH mutation in *antD* AT domain; Apr^R	This study
Cos213 ΦC31Apr Δ*antB*	Cos213 derivative with *antB* deletion; Apr^R	This study
Cos213 ΦC31Apr- HAFH Δ*antB*	Cos213 derivative with HAFH mutation in *antD* AT domain and *antB* deletion; Apr^R	This study
Cos213 ΦC31Apr- YASH Δ*antB*	Cos213 derivative with YASH mutation in *antD* AT domain and *antB* deletion; Apr^R	This study
Cos213 ΦC31Apr- AAPH Δ*antB*	Cos213 derivative with AAPH mutation in *antD* AT domain and *antB* deletion; Apr^R	This study
Cos213 ΦC31Apr Δ*antM*	Cos213 derivative with *antM* deletion; Apr^R	This study
Plasmids		

pUZ8002	Encodes the conjugation machinery for mobilisation of plasmids from *E. coli* to *Streptomyces*; Kan^R	(MacNeil et al., 1992)
pBT340	Encodes the FLP recombination machinery and has a temperature sensitive origin of replication, Cam^R	(Datsenko and Wanner, 2000)
pET28a	Commercial protein production vector, Kan^R	Agilent Technologies
pIJ10257	pMS81 derivative containing *ermE**p, integrates into the ΦBT1 *attB* site in *Streptomyces*; Hyg^R	(Hong et al., 2005)
pIJ773	ReDirect PCR template for the *aac(3)IV + oriT* cassette	(Gust et al., 2003)
pKT25	pSU40 derivative containing the T25 portion of the *B. pertussis* adenylate cyclase and a C-terminal multiple cloning site; Kan^R	(Karimova et al., 1998)
pUT18	pUC19 derivative containing the T18 portion of the *B. pertussis* adenylate cyclase and an N-terminal multiple cloning site; Carb^R	(Karimova et al., 1998)
pCRISPomyces- 2	Streptomyces CRISPR/Cas9 editing vector, Apr^R	(Cobb et al., 2015)
pET28a- *TEV-MBP*	pET28a derivative containing the *TEV* and *MBP* genes cloned into the HindIII-NotI restriction sites; Kan^R	This study
pET28a- *antM*	pET28a derivative containing the *antM* gene cloned into the NdeI-HindIII restriction sites; Kan^R	This study
pET28a- *antM*- WW	pET28a derivative containing the *antM* gene, with the Y185W, F223W mutations, cloned into the NdeI-HindIII restriction sites; Kan^R	This study
pET28a- *antM*- WF	pET28a derivative containing the *antM* gene, with the Y185W mutation, cloned into the NdeI-HindIII restriction sites; Kan^R	This study
pET28a- *antM*- YW	pET28a derivative containing the *antM* gene, with the F223W mutation, cloned into the NdeI-HindIII restriction sites; Kan^R	This study
pET28a- *antM*- AA	pET28a derivative containing the *antM* gene, with the Y185A, F223A mutations, cloned into the NdeI-HindIII restriction sites; Kan^R	This study

pIJ10257- *antM*	pIJ10257 derivative containing the *antM* coding sequence cloned into the NdeI-HindIII sites; Hyg[R]	This study
pIJ10257- *antM*- WW	pIJ10257 derivative containing the *antM* coding sequence, with Y185W, F223W mutations, cloned into the NdeI-HindIII sites; Hyg[R]	This study
pIJ10257- *antM*- WA	pIJ10257 derivative containing the *antM* coding sequence, with Y185W, F223A mutations, cloned into the NdeI-HindIII sites; Hyg[R]	This study
pIJ10257- *antM*- WF	pIJ10257 derivative containing the *antM* coding sequence, with Y185W mutation, cloned into the NdeI-HindIII sites; Hyg[R]	This study
pIJ10257- *antM*- AW	pIJ10257 derivative containing the *antM* coding sequence, with Y185A, F223W mutations, cloned into the NdeI-HindIII sites; Hyg[R]	This study
pIJ10257- *antM*- AA	pIJ10257 derivative containing the *antM* coding sequence, with Y185A, F223A mutations, cloned into the NdeI-HindIII sites; Hyg[R]	This study
pIJ10257- *antM*- AF	pIJ10257 derivative containing the *antM* coding sequence, with Y185A mutation, cloned into the NdeI-HindIII sites; Hyg[R]	This study
pIJ10257- *antM*- YW	pIJ10257 derivative containing the *antM* coding sequence, with F223W mutation, cloned into the NdeI-HindIII sites; Hyg[R]	This study
pIJ10257- *antM*- YA	pIJ10257 derivative containing the *antM* coding sequence, with F223A mutation, cloned into the NdeI-HindIII sites; Hyg[R]	This study
pCDF-Duet- *antD*	pCDF-Duet derivative containing the *antD* gene cloned into the EcoRI-NotI restriction sites; Spc[R]	(Sandy et al., 2012)
pKT25- *antC M2*	pKT25 derivative with the *antC* module 2 gene cloned downstream of the adenylate cyclase T25 portion by Gibson Assembly; Kan[R]	This study
pUT18- *antD*	pUT18 derivative with the *antD* gene cloned upstream of the adenylate cyclase T18 portion by Gibson Assembly; Carb[R]	This study
pKT25- *antM*	pKT25 derivative with the *antM* gene cloned downstream of the adenylate cyclase T25 portion by Gibson Assembly; Kan[R]	This study

pKT25- *antD*	pKT25 derivative with the *antD* gene cloned downstream of the adenylate cyclase T25 portion by Gibson Assembly; Kan^R	This study
pUT18- *antM*	pUT18 derivative with the *antM* gene cloned upstream of the adenylate cyclase T18 portion by Gibson Assembly; Carb^R	This study
pET28a- *surE*	pET28a derivative containing the *surE* gene cloned into the NdeI-HindIII restriction sites; Kan^R	This study
pET28a- *TEV-MBP- surE*	pET28a- *TEV-MBP* derivative containing the *surE* gene cloned into the NgeI-HinDIII restriction sites; Kan^R	This study
pET28a- *surE Mut*	pET28a- *surE* derivative containing a Ser to Ala active site mutation; Kan^R	This study
pET28a- *surF*	pET28a derivative containing the *surF* gene cloned into the NdeI-HindIII restriction sites; Kan^R	This study
pET28a- *surD mod3*	pET28a derivative containing the *surD* NRPS module 3 gene cloned into the NdeI-HindIII restriction sites; Kan^R	This study
pET28a- *surD A-PCP-E*	pET28a derivative containing the *surD* NRPS module 3 A-PCP-E domain genes cloned into the NdeI-HindIII restriction sites; Kan^R	This study
pET28a- *surD A-PCP*	pET28a derivative containing the *surD* NRPS module 3 A-PCP domain genes cloned into the NdeI-HindIII restriction sites; Kan^R	This study
pET28a- *surD PCP-E*	pET28a derivative containing the *surD* NRPS module 3 PCP-E domain genes cloned into the NdeI-HindIII restriction sites; Kan^R	This study
pET28a- *surD PCP*	pET28a derivative containing the *surD* NRPS module 3 PCP domain gene cloned into the NdeI-HindIII restriction sites; Kan^R	This study
pET28a- *surC mod6*	pET28a derivative containing the *surC* NRPS module 6 gene cloned into the NdeI-HindIII restriction sites; Kan^R	This study
pET28a- *surC A-PCP-E*	pET28a derivative containing the *surC* NRPS module 6 A-PCP-E domain genes cloned into the NdeI-HindIII restriction sites; Kan^R	This study

pET28a- *surC A-PCP*	pET28a derivative containing the *surC* NRPS module 6 A-PCP domain genes cloned into the NdeI-HindIII restriction sites; Kan[R]	This study
pET28a- *surC PCP-E*	pET28a derivative containing the *surC* NRPS module 6 PCP-E domain genes cloned into the NdeI-HindIII restriction sites; Kan[R]	This study
pET28a- *surC PCP*	pET28a derivative containing the *surC* NRPS module 6 PCP domain gene cloned into the NdeI-HindIII restriction sites; Kan[R]	This study
pCRISPomyces-2- *surd PCP-E-LD*	pCRISPomyces-2 derivative containing a Landing domain (post E domain) spacer sequence and HDR repair arms to delete *surD* PCP-E-LD domains; Apr[R]	This study
pCRISPomyces-2- *surd E-LD*	pCRISPomyces-2 derivative containing a Landing domain (post E domain) spacer sequence and HDR repair arms to delete *surD* E-LD domains; Apr[R]	This study
pCRISPomyces-2- *surd E*	pCRISPomyces-2 derivative containing a Landing domain (post E domain) spacer sequence and HDR repair arms to delete *surD* E domain; Apr[R]	This study
pCRISPomyces-2- *surd LD*	pCRISPomyces-2 derivative containing a Landing domain (post E domain) spacer sequence and HDR repair arms to delete *surD* LD domain; Apr[R]	This study

2.2.8 Oligonucleotide Primers and other synthetic DNA

All primers were purchased from Integrated DNA Technologies and were designed to manufacturer instructions and to give favourable PCR or sequencing reaction conditions.

Table 4: Oligonucleotide PCR primers used in the course of this study:

Primer name	Sequence (5' → 3')	Description
AF001	gagcacgtgcgagtgggcgggtttttagagctagaaatagcaag	Clone sgRNA into pKD plasmid via CPEC, fwd
AF002	ccgcccactcgcacgtgctcgtgctcagtatctctatcactga	Clone sgRNA into pKD plasmid via CPEC, rev
AF003	ccaattgtccatattgcatca	Amplification of pKD plasmid and sgRNA, fwd
AF004	tttataacctccttagagctcga	Amplification of pKD plasmid and sgRNA, rev

AF005	caacagccccgtggcgtgca	Amplification of Ant repairing G Block, fwd
AF006	agtcgcccgtcacgttggtg	Amplification of Ant repairing G Block, rev
AF007	cagtgaatggggggtaaatgg	sgRNA confirmation in pKD plasmid, fwd
AF008	gcctgcagtctagactcgag	sgRNA confirmation in pKD plasmid, rev
AF009	agctttcgctaaggatgattt	Sequencing of pKD for specific sgRNA cloning confirmation
AF010	tagacatatggagaatttgtattttcagggtatgaccaccacccagcggca	Amplification of a*ntM* from cosmid for cloning into plasmid, with restriction and TEV sites, fwd
AF011	atcaagctttcacagaccgaggccgacac	Amplification of a*ntM* from cosmid for cloning into plasmid, rev
AF017	tgttccatttcgtcgattacgt	PCR checking of correct AT domain mutation after CRISPR, fwd
AF018	tccccccttttccaccccttcccacgaggagagtcccatgattccggggatccgtcgacc	In frame deletion of a*ntM* by ReDirect protocol, fwd
AF020	gcccgggtgcacacagaaca	PCR checking of correct AT domain mutation after CRISPR, rev
AF021	tagggacagaccggtgtggggcgcggggacggccggtcatgtaggctggagctgcttc	In frame deletion of *antM* by ReDirect protocol, rev
AF022	aactgccgcgccaaggtcgt	Confirmation of *antM* replacement with Hyg cassette by ReDirect protocol, fwd
AF023	tgttgcagggtgttgaggac	Confirmation of *antM* replacement with Hyg cassette by ReDirect protocol, rev
AF030	catccggctcatcgcctcct	Amplification of G Block repairing CRISPR cut and mutation to YASH motif, fwd
AF031	atgccgtcggcgaaccggac	Amplification of G Block repairing CRISPR cut and mutation to YASH motif, rev
AF036	tagacatatggatgacccgcaggccctgcc	Amplification of *antD* for cloning, fwd
AF037	tctaaagctttcagccggaggccgggcgca	Amplification of *antD* for cloning, rev
AF043	taatacgactcactataggg	T7 promoter, sequencing primer
AF044	gctagttattgctcagcgg	T7 terminator, sequencing primer

AF045	tagaaagcttggatgacccgcaggccctgcc	Subclone *antD* into pUT18 (Bac 2 Hybrid), fwd
AF046	tctagaattcatgccggaggccgggcgcaggg	Subclone *antD* into pUT18 (Bac 2 Hybrid), rev
AF047	tagatctagatacggcgtccgagcgcccctc	Subclone *antC M2* into pKT25 (Bac 2 Hybrid), fwd
AF048	tctaggatcctccatggcgtcggtccttcc	Subclone *antC M2* into pKT25 (Bac 2 Hybrid), rev
AF049	gtgatgtgcagcagcaggtc	Confirmation of *antB* deletion by ReDirect protocol, fwd
AF050	cggccggattcatggaagtc	Confirmation of *antB* deletion by ReDirect protocol, rev
AF051	tgccaagatcggggcctttcgctgtctccctcgccgtgattccgggg atccgtcgacc	In frame deletion of *antB* (ReDirect protocol), fwd
AF052	cttgcggggaggacggtgcgggcgggcggtccgcggtcatgtagg ctggagctgcttc	In frame deletion of *antB* (ReDirect protocol), rev
AF055	tagaaagcttgaaaacctttacttccagggcaaaactgaagaaggtaa act	Creation of C-terminal MBP-TEV pET28a plasmid, with N-term His and multiple cloning site, fwd
AF056	tctagcggccgctcaagtctgcgcgtctttcaggg	Creation of C-terminal MBP-TEV pET28a plasmid, with N-term His and multiple cloning site, rev
AF057	tagacatatgaaactcctgatcatcggcgg	Subclone *natE* gene into pET28a, fwd
AF058	tctaaagctttcactgcggctgctttccct	Subclone *natE* gene into pET28a, rev
AF075	ggcgggctgcagggtcgactacggcgtccgagcgcccc	Subclone half of *antC M2* into pKT25 (Bac 2 Hybrid), fwd
AF076	cggcgaggagcccgaacgcgaagttcggggc	Subclone half of *antC M2* into pKT25 (Bac 2 Hybrid), rev
AF077	cgcgttcgggctcctcgccgagcaggcc	Subclone second half of *antC M2* into pKT25 (Bac 2 Hybrid), fwd
AF078	tagttacttaggtacccgggttccatggcgtcggtccttc	Subclone second half of *antC M2* into pKT25 (Bac 2 Hybrid), rev
AF079	ctatgaccatgattacgccagatgacccgcaggccctg	Subclone half of *antD* into pUT18 (Bac 2 Hybrid), fwd
AF080	cagcgcctcgtcgatctccatcaccccgg	Subclone half of *antD* into pUT18 (Bac 2 Hybrid), rev
AF081	tggagatcgacgaggcgctgccggtggtc	Subclone second half of *antD* into pUT18 (Bac 2 Hybrid), fwd

AF082	cgtggcctcgctggcggctgcgccggaggccgggcgcag	Subclone second half of *antD* into pUT18 (Bac 2 Hybrid), rev
AF109	tagacatatggtgccgatcgaacgcatcaa	Subclone *surE* from *S. albus* S4 into pET28a, fwd
AF110	tctagaattctcaggcgcgctgcgcgaaga	Subclone *surE* from *S. albus* S4 into pET28a, rev
AF111	tctagaattcggcgcgctgcgcgaagaagt	Subclone *surE* from *S. albus* S4, without a STOP codon, into pET28a-*TEV-MBP*, rev
AF112	ctgcagggtcgactctagagatgaccaccacccagcgg	Subclone *antM* into pKT25 (Bac 2 Hybrid), fwd
AF113	tagttacttaggtacccgggtcacagaccgaggccgac	Subclone *antM* into pKT25 (Bac 2 Hybrid), rev
AF114	gttcgccattatgccgcatc	Sequencing of pKT25 plasmid
AF115	ggatgtgctgcaaggcgatt	Sequencing of pKT25 plasmid
AF116	gtgtggaattgtgagcggat	Sequencing of pUT18 plasmid
AF117	ttccacaacaagtcgatgcg	Sequencing of pUT18 plasmid
AF118	tagacatatgaccgggccggtcgccaccac	Subclone *surD M3+E* from *S. albus* S4 into pET28a, fwd
AF119	tctaaagctttcacttctccaggcaagccg	Subclone *surD M3+E* from *S. albus* S4 into pET28a, rev
AF120	tctaaagctttcagtccgcctcggcggcgaggc	Subclone *surD M3* from *S. albus* S4 into pET28a, rev
AF121	tagacatatggaacgcgccctgtgccagat	Subclone *surD M3 PCP* from *S. albus* S4 into pET28a, fwd
AF128	tagacatatggtcaccccggcgccgcgcga	Subclone *surC M6+E* from *S. albus* S4 into pET28a, fwd
AF129	tctaaagcttctagctgcggcaggcgctgc	Subclone *surC M6+E* from *S. albus* S4 into pET28a, rev
AF130	tctaaagctttcagtcgtcggccacctggccgt	Subclone *surC M6* from *S. albus* S4 into pET28a, rev
AF131	tagacatatgggccacgtctacacggcccc	Subclone *surC M6 PCP* from *S. albus* S4 into pET28a, fwd
AF132	tagacatatgaccaccacccagcggcacca	Subclone *antM* into pIJ10257, fwd
AF137	tagacatatgtcgtacgccgaggtggacgc	Subclone *surD M3 A domain* (A-PCP, A-PCP-E) into pET28a, fwd
AF138	tagacatatgacctacgcggagctgggccg	Subclone *surC M6 A domain* (A-PCP, A-PCP-E) into pET28a, fwd

AF139	ctactccgactaccgcctct	Mutagenesis of pET28a-*surE,* vector portion, fwd
AF140	gaaggggaaacccgtccg	Mutagenesis of pET28a-*surE,* vector portion, rev
AF141	cccggacgggtttcccttcggcGCTgtgaccaagttcct	Mutagenesis of pET28a-*surE,* insert portion including mutation, fwd
AF142	agaggcggtagtcggagtagctctcgcgggtcaggaagag	Mutagenesis of pET28a-*surE,* insert portion, rev
AF143	tagacatatggagaatttgtattttcagggtgtgggtgccgagggggcggaa	Subclone *surE* into pET28a with TEV site, fwd
AF144	tctaaagctttcagagccggtgcatggccc	Subclone *surE* into pET28a with TEV site, rev
AF166	ctgagcaacccggagatccg	Mutagenesis of pET28a-*antM,* vector portion, fwd
AF167	gacgagtgcggttctgccct	Mutagenesis of pET28a-*antM,* vector portion, rev
AF168	agggcagaaccgcactcgtcaccTTcggcagccggggc	Mutagenesis of pET28a-*antM,* insert portion including mutation 1, fwd
AF169	agggcagaaccgcactcgtcaccggcggcagccggTTcatcggc	Mutagenesis of pET28a-*antM,* insert portion including mutation 2, fwd
AF170	agggcagaaccgcactcgtcaccTTcTTcagccggTTcatcTTccgcgcc	Mutagenesis of pET28a-*antM,* insert portion including mutation 3, fwd
AF171	cggatctccgggttgctcagcgtcgggttg	Mutagenesis of pET28a-*antM,* insert portion, rev
AF172	cggatctccgggttgctcagcgtcgggttgatctcgTActcggtga	Mutagenesis of pET28a-*antM,* insert portion including mutation 9, rev
AF173	cggatctccgggttgctcagcgtcgggttgaActcggtctc	Mutagenesis of pET28a-*antM,* insert portion including mutation 10, rev
AF174	cggatctccgggttgctcagcgtcgggttgaActcgTActcggtga	Mutagenesis of pET28a-*antM,* insert portion including mutation 11, rev
AF175	agggcagaaccgcactcgtcaccggcggca	Mutagenesis of pET28a-*antM,* insert portion, fwd
AF176	acgcagaccgcctacccggc	Mutagenesis of pET28a-*antM,* vector portion, fwd
AF177	gccggcgttgttgaccagga	Mutagenesis of pET28a-*antM,* vector portion, rev
AF178	tcctggtcaacaacgccggcatcTAcctgccccgtccc	Mutagenesis of pET28a-*antM,* insert portion including mutation 4, fwd

AF179	tcctggtcaacaacgccggcatcaccctgGcccgtccc	Mutagenesis of pET28a-*antM,* insert portion including mutation 5, fwd
AF180	tcctggtcaacaacgccggcatcTAcctgGcccgtccc	Mutagenesis of pET28a-*antM,* insert portion including mutation 6, fwd
AF181	gccgggtaggcggtctgcgtcgcggcggag	Mutagenesis of pET28a-*antM,* insert portion, rev
AF182	gccgggtaggcggtctgcgtcgcgAAggaggaga	Mutagenesis of pET28a-*antM,* insert portion including mutation 7, rev
AF183	gccgggtaggcggtctgcgtGAAgAAggaggaga	Mutagenesis of pET28a-*antM,* insert portion including mutation 8, rev
AF184	tcctggtcaacaacgccggcatcaccctgc	Mutagenesis of pET28a-*antM,* insert portion, fwd
AF199	tctaaagcttgagccggtgcatggccc	Subclone *surE,* without a STOP codon, into pET28a-*TEV-MBP*, rev
AF216	ACGCcgccgcccccgaccaggaag	Protospacer sequence to make CRISPR deletions in *surD*, fwd
AF217	AAACcttcctggtcggggggcggcg	Protospacer sequence to make CRISPR deletions in *surD*, rev
AF218	ACGCatcgccgccatggtccaaca	Protospacer sequence (2) to make CRISPR deletions in *surD*, fwd
AF219	AAACtgttggaccatggcggcgat	Protospacer sequence (2) to make CRISPR deletions in *surD*, rev
AF220	tgccgccgggcgttttttatctagactcactgaccggcgcgggcg	Upstream region of *surD M3 PCP* for deletion of region of *surD* (PCP-E-LD), fwd
AF221	tgtacctactgcaatgccatcgtgtcggtgcgcggggcca	Upstream region of *surD M3 PCP* for deletion of region of *surD* (PCP-E-LD), rev
AF222	atggcattgcagtaggtacatgaggcggggggcggggg	Downstream region of *surD M3* Landing domain for deletion of region of *surD* (PCP-E-LD, E-LD, LD), fwd
AF223	ctttttacggttcctggcctctagaggcgatcgccgtacggaccg	Downstream region of *surD M3* Landing domain for deletion of region of *surD* (PCP-E-LD, E-LD, LD), rev
AF224	tgccgccgggcgttttttatctagacgtcggccagaccgtgg	Upstream region of *surD M3* Landing domain for deletion of region of *surD* (LD), fwd

AF225	tgtacctactgcaatgccatggcctccagggcgacgggggt	Upstream region of *surD M3* Landing domain for deletion of region of *surD* (LD), rev
AF226	tgccgccgggcgttttttatctagacaccacgctgttcaccggcg	Upstream region of *surD M3* E domain for deletion of region of *surD* (E, E-LD), fwd
AF227	tgtacctactgcaatgccatggcctcggactcctcccc	Upstream region of *surD M3* E domain for deletion of region of *surD* (E, E-LD), rev
AF228	atggcattgcagtaggtacacccgccacccccggctg	Downstream region of *surD M3* E domain for deletion of region of *surD* (E), fwd
AF229	ctttttacggttcctggcctctagacgtctcggcgggcctgttcg	Downstream region of *surD M3* E domain for deletion of region of *surD* (E), rev
AF250	gagaccgagatcaacccgac	Mutagenesis of pET28a-*antM,* vector portion, fwd
AF251	ggtctgcgtcgcggcggagg	Mutagenesis of pET28a-*antM,* vector portion, rev
AF252	cctccgccgcgacgcagaccgccGCcccggcgatcg	Mutagenesis of pET28a-*antM,* insert portion with Y185A mutation, fwd
AF253	cctccgccgcgacgcagaccgcctGGccggcgatcg	Mutagenesis of pET28a-*antM,* insert portion with Y185W mutation, fwd
AF254	gggttgatctcggtctcggtgGCgccgggcgcg	Mutagenesis of pET28a-*antM,* insert portion with F223A mutation, rev
AF255	gggttgatctcggtctcggtCCagccgggcgcg	Mutagenesis of pET28a-*antM,* insert portion with F223W mutation, rev
AF256	cctccgccgcgacgcagaccgccttaccggcgatcg	Mutagenesis of pET28a-*antM,* insert portion, fwd
AF257	gggttgatctcggtctcggtgaagccgggcgcg	Mutagenesis of pET28a-*antM,* insert portion, rev
AF258	cctccgccgcgacgcagaccgcctTcccggcgatcgtcgcctactcgatg	Mutagenesis of pET28a-*antM,* insert portion with Y191F mutation, fwd
AF261	caactgttgggaagggcgatc	Sequencing pCRISPomyces-2 plasmid
AF-GB01	caacagccccgtggcgtgcacggtcgccggcgacaccgacgcggt cgaccggctggaggcggaactgacccgccgcgatgttccatttcgtc gattacgtatgcctgccgcccctcactctcatgtacttgatcctatccttg	Sequence of Gblock used to repair CRISPR cut and introduce AAPH motif into AT domain sequence

	aaagtttcgccggtcacctgcgcacgctgaccctgcggccgccgcgc atcccgtacgtcaccaacgtgacgggcgact	
AF-GB02	caacagccccgtggcgtgcacggtcgccggcgacaccgacgcggt cgaccggctggaggcggaactgacccgccgcgatgttccatttcgtc gattacgtatgcctcatgcctttcactctcatgtacttgatcctatccttga aagtttcgccggtcacctgcgcacgctgaccctgcggccgccgcgca tcccgtacgtcaccaacgtgacgggcgact	Sequence of Gblock used to repair CRISPR cut and introduce HAFH motif into AT domain sequence
AF-GB03	catccggctcatcgcctcctcgggcggtgccaccgtcggggtcgcgg cctccgtcgacaccgttcttccactgctgggcgaaggtctgtctctggc ggcggttaacagcccggttgcgtgtactgttgcaggtgataccgatgc ggttgatcgtctggaagcggaactgacccgtcgtgatgttccgtttcgt cgtctgcgtatgccgtacgcctcccactctcatgttctggatccgattct ggaaagctttgcaggtcatctgcgtactctgaccctgcgtccgccgcg tattccgtatgttaccaacgttactggtgattgggcaaccgatgcacag gcgacggaggtcggtcactgggtggaccacacccggcgcaccgtc cggttcgccgacggcat	Sequence of Gblock used to repair CRISPR cut and introduce YASH motif into AT domain sequence

2.3 Growth and Storage of Bacterial Species'

2.3.1 *E. coli* Culture Conditions

All *E. coli* strains were cultured at 37 °C in the selected medium, with required antibiotic selection (as described above). Bacteria were cultured for ~ 16 hours whilst shaking at 180 rpm.

2.3.2 *Streptomyces* Culture Conditions

All *Streptomyces* strains were propagated on SFM agar at 30 °C for seven days, in the presence of the appropriate antibiotic selection (Kieser et al., 2000). Growth of *Streptomyces* strains for specialised metabolite production was tailored in each individual scenario with regards to media condition and length of incubation (developed in more detail in section 2.6.1).

2.3.3 Preparation of *E. coli* Glycerol Stocks

All *E. coli* strains, containing required and constructed plasmids of interest, were stocked and maintained at -80 °C for use at later dates. For the creation of glycerol stocks, *E. coli* bacterial strains were cultured overnight (~ 16h, 37 °C, 180 rpm rotation), after which 500 μL of the

culture was mixed with 500 µL of 50 % v/v glycerol. The mixed glycerol stock was flash frozen in a dry ice-ethanol bath and stored at -80 °C.

2.3.4 Preparation of *Streptomyces* Spore Stocks

Spore stocks of required *Streptomyces* strains were prepared and maintained at both -80 °C, for long term storage, and -20 °C, for more immediate usage and applications. Initially SFM plates were inoculated with concentrated spores or cultured mycelia of the required strain. The plates were incubated at 30 °C for 7 days. After full visible sporulation of the cultured streptomycete 2 mL of sterile 20 % v/v glycerol was added to each plate and the spores were scraped from the plate with a sterile cotton bud. The mixture of glycerol and spores was pipetted into sterile tubes; 1 mL into a tube for storage at -80 °C and lower volumes of ~ 200 µL into tubes for storage at -20 °C and subsequent usage. Agar plates were then either discarded or used for the extraction of specialised metabolites, as required.

2.3.5 Preparation of Chemically Competent *E. coli* Cells

LB medium (50 mL) was inoculated with 500 µL of an overnight starter culture of the *E. coli* strain for which competent cells were required. The larger sub-culture was incubated with shaking (180 rpm) at 37 °C until $OD_{600} = 0.4 - 0.6$. The cells were subsequently rested on ice for 20 minutes and then harvested by centrifugation (4000 x g, 4 °C, 5 mins). The spent medium was decanted and the cell pellet was resuspended in 5 mL ice cold 0.1 M $CaCl_2$ and rested on ice for 20 minutes. The cells were then collected by centrifugation (as above) and the supernatant decanted before resuspension of the cells in 2.5 mL ice cold 0.1 M $CaCl_2$ containing 15 % v/v glycerol. The cell suspension was divided into 100 µL aliquots, flash frozen in a dry ice-ethanol bath, and stored at -80 °C for subsequent transformation.

2.3.6 Preparation of Electrocompetent *E. coli* Cells

For the preparation of electrocompetent cells a similar process to the preparation of chemically competent cells was followed until cells reached the required optical density. LB medium (50 mL, containing antibiotics as required) was inoculated with an overnight starter culture of the required strain (500 µL). The larger culture was incubated (37 °C, 180 rpm) until the OD_{600} reached 0.4 – 0.6 and cells were harvested by centrifugation (4000 x g, 4 °C, 5 mins). The medium was decanted and the pellet gently resuspended in sterile, ice cold 10 % v/v glycerol. The process of resuspension in 10 % v/v glycerol and collection by centrifugation (as above)

was repeated before cells were again resuspended in 10 % v/v glycerol and divided into 100 μL aliquots. These aliquots were then flash frozen in a dry ice-ethanol bath and stored at -80 °C for subsequent transformation.

2.3.7 Transformation of Competent *E. coli*

For the transformation of chemically competent *E. coli* (as prepared above), cells were mixed with the plasmid or cosmid of interest (3 μL for purified DNA stocks, 20 μL for ligation reactions) under sterile conditions and incubated on ice for 20 – 30 minutes. The mixture of cells and DNA was transferred to a 42 °C water bath for ~ 45 seconds and then returned to ice. 900 μL of sterile medium was added to the heat shocked mixture and cells were incubated at 37 °C for 1 hour, with shaking (180 rpm). Post recovery, the bacterial cells were collected by low speed centrifugation in a microcentrifuge, resuspended in a minimal volume of sterile medium (~ 100 μL) and spread onto warmed agar plates, containing the required antibiotic selective agent. Dried plates were then incubated overnight at 37 °C.

For the transformation of electrocompetent cells (as prepared above), cells were thawed and mixed with the plasmid or cosmid of interest (3 μL purified DNA) under sterile conditions, and transferred into a cold, 0.2 cm electroporation cuvette. Cells were pulsed in a Bio-Rad gene pulser using the following conditions: 200 Ω resistance, 25 μF capacitance, and 2.5 kV of voltage, giving an expected time constant of 4.5 – 4.9 ms. Post electroporation, 900 μL of sterile medium was added and cells were incubated at 37 °C for one hour, after which cells were collected by low speed centrifugation in a microcentrifuge. The cell pellet was resuspended in a minimal volume of medium (~ 100 μL) and spread onto warmed agar plates with the required antibiotic selection. Plates were again incubated overnight at 37 °C.

2.3.8 Intergeneric Conjugation into *Streptomyces*

Intergeneric conjugation of circular DNA from *E. coli* into *Streptomyces* was carried out following an established and well documented protocol (Kieser et al., 2000). Briefly, the plasmid or cosmid of interest was transformed into the de-methylating strain of *E. coli* ET12567/pUZ8002. A single colony was cultured overnight in 10 mL of LB medium (37 °C, 180 rpm) containing the requisite antibiotics (kanamycin, chloramphenicol, and the plasmid/cosmid selective marker). Two hundred microlitres of the culture was used to inoculate 20 mL fresh medium with the same antibiotic selections. This sub-culture was subsequently allowed to grow until OD_{600} 0.4 – 0.6 was reached, after which the 10 mL

overnight culture and the 20 mL sub-culture were pooled and combined. The pooled culture was centrifuged to obtain a cell pellet, which was then resuspended in 10 mL fresh LB medium and centrifuged, in order to wash away remaining antibiotics. The wash step was repeated, followed by resuspension of the cell pellet in 1 mL fresh medium. The concentrated and washed cell pellet was then mixed with 200 – 300 μL *Streptomyces* spores and centrifuged to pellet the mixture. The mixed cells were resuspended in 300 μL LB medium and plated onto SFM lacking any selective agents. Agar plates were incubated at 30 °C for 16 – 18 hours, after which the plates were overlaid with 1 mL sterile water containing nalidixic acid (20 μL) and the desired antibiotic (50 μL apramycin or hygromycin). Overlaid plates were incubated at 30 °C for 2 – 5 days, until the appearance of single colonies of *Streptomyces.*

2.4 DNA Isolation, Purification, and Methodologies

2.4.1 Isolation of Plasmid/Cosmid DNA from *E. coli*

LB medium (10 mL, containing requisite antibiotic selection) was inoculated with a single colony or glycerol stock of the *E. coli* strain containing the desired plasmid or cosmid of interest. The culture was grown overnight (37 °C, 180 rpm) after which cells were harvested by centrifugation (4000 x *g*, 5 mins). Plasmid/cosmid DNA was harvested using a Monarch plasmid miniprep kit (New England Biolabs) as per the manufacturer's instructions and recommendations. DNA concentration quantitation, if required, was carried out using a Geneflow nanophotometer.

2.4.2 Extraction of *Streptomyces* Genomic DNA

LB medium (10 mL, containing the requisite antibiotics or without selection for wild-type *Streptomyces* spp.) was inoculated with spores of the streptomycete from which genomic DNA was required. Cultures were incubated with shaking for 24 – 48 hours (30 °C, 180 rpm) and cells were subsequently pelleted by centrifugation (4000 x *g*, 15 mins). Pelleted cells were resuspended in buffer P1 (50 mM Tris-Cl, 10 mM EDTA, 100 μg/mL RNAse A, pH 8.0) and lysozyme was added at a final concentration of 10 mg/mL. Resuspended cells were incubated for ~ 2 hours at 37 °C, after which SDS and NaCl were added to final concentrations of 1 % w/v and 1.25 M, respectively, and the incubation period continued on ice for an additional 10 minutes. The resultant lysate was extracted twice with equal volumes of phenol: chloroform: isoamyl alcohol (25: 24:1) after which the sample was ethanol precipitated to concentrate the

DNA. 0.1 volumes of 3 M sodium acetate was added, followed by 2.5 volumes of 100 % v/v ethanol. The sample was incubated at -20 °C for at least 30 minutes, after which samples were centrifuged (15900 x *g*, 15 mins) to pellet the DNA. The resultant pellet was washed twice with 500 μL of 70 % v/v ethanol and the pellet was air dried for one minute. To elute DNA and obtain a concentrated genomic DNA stock solution, 100 μL pure water was used to solubilise the pellet. Quantity and quality of the DNA solution was ascertained using a nanophotometer.

2.4.3 DNA Amplification by Polymerase Chain Reaction (PCR)

PCR mixtures were prepared on ice according to the component table below, and contained template DNA, primers (Integrated DNA Technologies), dNTPs (New England Biolabs), a DNA polymerase enzyme; either Q5 high fidelity DNA polymerase (New England Biolabs) for preparing insert for cloning, or GoTaq G2 DNA polymerase (Promega) for diagnostic purposes, and PCR reaction buffers. Annealing temperatures for primers were calculated using online analysers (New England Biolabs) and optimised experimentally. Thermocycling was carried out in a T100 thermal cycler (Bio-Rad) and PCR products were either analysed by gel electrophoresis, used for other protocols, or stored at -20 °C for future use. Colony PCR reactions were carried out using GoTaq G2 DNA polymerase (and the corresponding required conditions) with no input template DNA. However, a bacterial colony was scraped and resuspended in the premixed PCR solution and the same thermocycling conditions were utilised.

Table 5: PCR mixture contents for Q5 polymerase:

Component	Concentration	Volume (μL)
Q5 reaction buffer	5x	10.0
Q5 high GC enhancer	5x	10.0
DMSO	-	1.5
dNTPs	10 mM (2.5 mM each dNTP)	4.0
Forward primer	20 μM	1.25
Reverse primer	20 μM	1.25
Template DNA	Varying	To 20 – 100 ng
Q5 polymerase	2,000 U/mL	0.5
Water	-	To 50.0

Table 6: PCR mixture contents for GoTaq G2 polymerase:

Component	Concentration	Volume (μL)

GoTaq reaction buffer	5x	4.0
DMSO	-	1.0
dNTPs	10 mM (2.5 mM each dNTP)	1.6
Forward primer	20 μM	0.5
Reverse primer	20 μM	0.5
Template DNA	Varying	To 20 – 100 ng
GoTaq G2 polymerase	2,000 U/mL	0.5
Water	-	To 20.0

Table 7: Thermocycling conditions; conditions shown are for Q5 polymerase, conditions in brackets are alterations required for use of GoTaq G2 DNA polymerase:

Temperature (°C)	Time (seconds)	Purpose	Cycles
98 (95)	30 (120)	Initial denaturation	1
98, 50 – 72, 72 (95, 42 – 65, 72)	10, 30, 30 / kilobase (30, 30, 60 / kilobase)	Cycles of denaturation, primer annealing, and extension	30
72	120 (300)	Final extension	1
10	∞	Infinite hold	1

2.4.4 Agarose Gel Electrophoresis and Extraction of DNA

DNA fragments were analysed by agarose gel electrophoresis to determine size, purity, and relative concentration. All agarose gels were made as required and concentration of agarose selected was dependent on expected size of DNA fragments to be analysed. Therefore, agarose gels ranging from 0.6 to 2.0 % w/v for, for example, cosmids and short DNA fragments, respectively. Volumes of agarose gels were 50 or 100 mL, dependent upon number of samples to be analysed. Powdered agarose was weighed out to the required concentration and added to the required volume of TAE buffer. The mixture was boiled in a microwave for ~ 1 minute to dissolve the agarose and 2 μL of SybrSafe was added to the cooled solution, after which it was poured into a suitably sized gel cast with a comb containing the requisite number of sample wells. The gel, solidified after ~ 15 minutes at room temperature, was loaded into a Bio-Rad gel tank and submerged in TAE buffer. Samples were loaded into the formed wells, alongside a DNA marker ladder, and were ran through the gel at a constant voltage of 100 V for ~ 45 minutes, or until samples had travelled and separated suitably through the gel. After completion of the run, agarose gels were analysed using a transilluminator and subsequently imaged.

Samples for electrophoresis were prepared in microcentrifuge tubes with 2 µL of sample loading buffer and 10 µL of DNA sample. 10 µL of each sample was loaded into the gel, with 5 µL of 1 kb or 1 kb plus ladder (New England Biolabs) used as the molecular marker for relative size determination.

2.4.5 Sanger Sequencing of DNA Constructs

The integrity, quality, and sequence of cloned DNA, within plasmids, cosmids, and purified PCR products, was confirmed by Sanger sequencing of the construct of interest. Samples were prepared as per company instructions (5 µL template DNA at required concentration, 5 µL primer (5 µM)) and sent to GATC Biotech (Eurofins Genomics) for sequencing. Sequences were checked and analysed using various computer programmes, primarily the A Plasmid Editor tool (ApE, University of Utah).

2.4.6 Restriction Digestion

Restriction digests were primarily carried out for the preparation of digested vectors and inserts, containing sticky ends, for ligation reactions. Diagnostic digests were also carried out for the purpose of initial DNA sequence confirmation and DpnI digests were utilised to remove bacteria-derived circular DNA elements, post PCR amplification of the required sections of DNA. Insert/vector digestions, as well as DpnI digests, were carried out in a final volume of 50 µL (as shown below), whilst diagnostic digests were scaled to a final volume of 10 µL. Digestion reactions were incubated at 37 °C for one hour followed by analysis of the products by agarose gel electrophoresis.

Table 8: Insert/vector digestion:

Component	Volume (µL)
Insert/vector DNA	35
10x CutSmart buffer	5
Restriction enzyme 1 (e.g. NdeI)	2
Restriction enzyme 2 (e.g. HindIII)	2
Deionised water	6

Table 9: DpnI digest:

Component	Volume (µL)
PCR product DNA	40
10x CutSmart buffer	5

DpnI	2
Deionised water	3

2.4.7 Ligation Reactions

Ligations of plasmid and insert were carried out using T4 DNA ligase (New England Biolabs) and according to manufacturer's instructions. Briefly, ligations were prepared on a 20 µL scale using a 3:1 molar ratio of insert to vector and containing 1x ligation buffer (final concentration) and 400 units of T4 DNA ligase. Reactions were typically incubated at room temperature overnight and subsequently the entire mixture was transformed into chemically competent XL10-Gold cells.

2.4.8 Gibson Assembly of DNA Fragments

Gibson assembly of overlapping DNA fragments was carried out using the NEBuilder HiFi DNA assembly kit (New England Biolabs) and according to manufacturer's instructions. DNA fragments with at least 20 bp overlaps were amplified by PCR and purified. The fragments were then assembled using a 2:1 insert to vector ratio in a total volume of 20 µL, half of which was the NEBuilder assembly master mix at a final 1x concentration. Assembly mixtures were incubated at 50 °C for 30 minutes in a thermal cycler, after which the mixture was transformed into chemically competent XL10-Gold cells and spread onto selective agar plates.

2.4.9 Circular Polymerase Extension Cloning (CPEC)

CPEC was used as a methodology for creating circular plasmids from two linear DNA fragments, containing overlapping regions at both ends of both fragments. Overlapping DNA fragments were amplified using PCR and purified prior to the occurrence of the CPEC reaction. Circular plasmids were created from overlapping fragments in a CPEC reaction, carried out with Q5 DNA polymerase (New England Biolabs) in rounds of melting, annealing, and extension, much like conditions utilised during PCR. The CPEC process, however, does not result in the amplification of DNA and resultant circular DNA concentration was dependent upon the concentration of input, overlapping linear fragments. Reactions were carried out in a T100 thermal cycler, with contents and conditions as shown below, after which the CPEC mixtures were transformed into XL10-Gold cells and plated onto selective medium.

Table 10: Composition of CPEC reactions:

Component	Concentration	Volume (µL)

Linear DNA 1	Varying	100 – 200 ng
Linear DNA 2	Varying	100 – 200 ng
Q5 DNA polymerase	2,000 U/mL	0.5
Q5 reaction buffer	5x	4.0
dNTPs	10 mM (2.5 mM each dNTP)	1.6
Deionised water	-	To 20.0

Table 11: Thermocycling conditions for CPEC reactions:

Temperature (°C)	Time (seconds)	Purpose	Cycles
98	30	Initial denaturation	1
98, 55, 72	10, 30, 30 / kilobase	Cycles of linear DNA extension and circular DNA formation	15
72	300	Final extension	1
10	∞	Infinite hold	1

2.4.10 CRISPR/Cas9 Mediated Mutation of Cosmid Constructs

CRISPR/Cas9 methodologies are well documented for mutating and modifying the *E. coli* genome. However, in this study, mutations were required in a cosmid construct held in an *E. coli* host strain. Therefore, an existing methodology was slightly modified to allow targeted mutation of the cosmid construct.

Initially, a cosmid gene specific single guide RNA (sgRNA) was cloned into the plasmid pKDsgRNA (Reisch and Prather, 2015), whilst the cosmid itself was transformed into *E. coli* BW25113. The sgRNA would target the Cas9 enzyme to cosmid, in order to cause a specific double strand break, and was cloned by CPEC into the requisite plasmid. Subsequently, a Cas9 containing plasmid (pCas9-Cr4, (Reisch and Prather, 2015)) was transformed into the *E. coli* strain containing the cosmid, followed by a second transformation of the sgRNA containing plasmid. The *E. coli* strain containing the three circular DNA constructs was confirmed by antibiotic selection and growth on selective medium. This strain was then grown to OD_{600} 0.4 – 0.6 in liquid medium with arabinose induction of the λ RED genes held on the sgRNA plasmid, followed by transformation of a linear, double-stranded DNA repair template. This template contained homology to the region where the double strand break would occur as well as the required mutation, so would be able to repair the break caused by the Cas9 enzyme and concomitantly introduce the desired mutation. Transformed cells were plated onto selective medium, as well as anhydrotetracycline (aTC) to induce expression of the Cas9 gene.

Successfully mutated cosmids were initially identified by colony PCR and isolated by sequential rounds of miniprep and subsequent transformation to obtain a pure stock of mutated cosmids. The ultimate identity of the mutated cosmid was confirmed by sequencing, followed by mobilisation of the cosmid into *Streptomyces* by intergeneric conjugation and confirmation by selective growth and PCR.

2.4.11 ReDirect Recombineering

Mutant *Streptomyces* strains were predominantly created by first mutating a cosmid DNA construct in *E. coli*, using the ReDirect PCR targeting protocol, followed by mobilisation of the mutagenised cosmid into *Streptomyces*. The ReDirect PCR targeting protocol was carried out as per published protocols and methodologies (Gust et al., 2003). A gene deletion cassette consisting of an *oriT* and the hygromycin or apramycin resistance genes was PCR amplified with primers containing 40 nucleotides of homology that included the start and stop codons of the gene to be deleted, as well as 37 nucleotides upstream or downstream. The gel purified PCR products were then used to delete the requisite gene from cosmid constructs within *E. coli*. Briefly, *E. coli* GB05-RED cells were initially transformed with the cosmid. The cells containined a chromosomal integration of the λ RED genes, facilitating later uptake of linear dsDNA. Cosmid-containing cells were cultivated to OD_{600} 0.4 – 0.6 with concomitant arabinose induction of the λ RED genes and subsequently transformed with the disruption cassette. Cells were then spread onto antibiotic-containing agar plates to select for the incoming disruption cassette. Successfully mutagenised cosmids were confirmed by PCR and then mobilised into *Streptomyces* by intergeneric conjugation, followed by confirmation of successful chromosomal integration by PCR using *Streptomyces* genomic DNA as template. Often, the deleted gene was part of a larger operon, meaning its deletion had the potential to cause polar mutations and affect the expression of downstream genes. In these scenarios, the disruption cassette was removed, leaving an 81 bp scar in place of the disruption cassette, and in frame with the remainder of the operon. This excision of the cassette was mediated by the FLP recombinase system, as well as the presence of FLP recognition target (FRT) sequences within the cassette itself. Briefly, *E. coli* DH5α / pBT340 was transformed with the mutagenized cosmid and incubated at 43 °C overnight. This induced synthesis of the FLP recombinase, and subsequent excision of the disruption cassette from the cosmid and curing of pBT340, resulting in the *E. coli* strain containing the mutagenised cosmid harbouring the 81 bp scar sequence. Successful FLP recombinants were identified by replica plating on

appropriate antibiotic selection and the integrity of the scar sequence was confirmed by PCR, followed by mobilisation of the cosmid into *Streptomyces* by intergeneric conjugation.

2.4.12 CRISPR/Cas9 Mediated Deletions in the *Streptomyces* Genome

The pCRISPomyces-2 system was used to delete genes within the *Streptomyces* genome, utilising a protocol previously described (Cobb et al., 2015). Briefly, double stranded sgRNA sequences were generated by annealing oligonucleotide primers containing the protospacer of interest, followed by introduction of the sgRNA into the pCRISPomyces-2 plasmid by Golden Gate Assembly utilising the vector BbsI restriction site (as described in the referenced protocol). Alongside this, 1 kbp homology arms, flanking the genes to be deleted, were amplified by PCR, purified, and subsequently cloned into the pCRISPomyces-2 plasmid XbaI site by Gibson Assembly (NEBuilder HiFi DNA Assembly Kit). The resultant plasmid was then mobilised into *Streptomyces* by conjugation, where CRISPR/Cas9-mediated deletion of the desired *Streptomyces* gene would occur. The temperature sensitive pCRISPomyces-2 plasmid was then cured from *Streptomyces* strains by passage in LB medium at 37 °C, followed by cultivation on SFM agar at 37 °C. Apramycin sensitivity of the resultant spores was then tested by replica plating, followed by identification, analysis, and isolation of those cultured stocks containing the requisite deletion. These deletion mutant strains, post-PCR confirmation of the deletion, could then be used in downstream applications.

2.5 Production and Purification of Proteins

2.5.1 IPTG Induction of Gene Expression

After creation and cloning of genes into protein production vectors, and subsequent transformation of *E. coli* strains optimised for protein production (such as BL21-Gold DE3), cultures were prepared to test and optimise levels of gene expression and protein production. Rich culture medium (usually LB) with antibiotic selection was prepared in either small volumes (50 mL) or for larger cultures (500 mL) and inoculated with 1 % v/v of an overnight starter culture of the *E. coli* strain (e.g. 5 mL overnight culture in 500 mL sterile medium). These cultures were incubated (37 °C, 180 rpm) until OD_{600} was equal to 0.6, followed by cooling of the cultures to room temperature. Isopropyl β- D-1-thiogalactopyranoside (IPTG) was added from a 1 M stock solution to a final concentration of 0.1 – 1 mM. The concentration of IPTG used for larger scale cultures was determined empirically by multiple small-scale

trials, in which the levels of protein production were tested and relatively quantitated against each other to determine the optimal concentration. Following the addition of IPTG to the growth medium, bacterial cultures were incubated at 16 °C overnight with shaking (180 rpm). Cells were harvested (4000 x g, 10 minutes, 4 °C) after overnight growth (~ 16 hours) and either stored as a bacterial cell pellet at -80 °C, or used for downstream protein purification applications.

2.5.2 Auto-Induced Expression of Genes

Auto-induction medium was prepared by suspension of pre-formed powder (Formedium) into deionised water, followed by sterilisation by autoclaving. Cultures were set up by inoculation of medium with 1 % v/v of an overnight starter culture of the *E. coli* strain containing the vector encoding production of the desired protein. After inoculation, auto-induction media cultures, containing antibiotics selecting for the desired *E. coli* strain and desired vector, were incubated at 16 °C for 48 – 72 hours, with shaking at 180 rpm. The timing of incubation was determined empirically with previous small-scale tests, with samples taken at various time points to determine the time of optimal protein production. After incubation of the cultures for the optimal length of time, the bacterial cells were collected by centrifugation (4000 x g, 10 minutes, 4 °C) and stored at -80 °C for later use, or utilised in downstream protein purification applications.

2.5.3 Ni^{2+} Affinity Chromatography for Protein Purification

Cell pellets generated from small- and large-scale cultures were predominantly used for the purification of protein for downstream assays. Bacterial cells were lysed by initial resuspension of the collected pellets in lysis buffer, followed by sonication of the resuspended samples in 50 mL batches using an ultrasonic probe (40 % power, 2 x 90 seconds, 1 pulse s^{-1}). The lysates were clarified by centrifugation (25000 x g, 30 minutes, 4 °C) and stored on ice for subsequent protein purification. Three millilitres of Ni-NTA resin (Generon) was loaded into a Bio-Rad Econo Pac column and equilibrated with five column volumes of lysis buffer. The clarified lysate was then loaded onto the column and drawn through, either by gravity flow or a peristaltic pump for larger volumes of lysate. A minimum of 5 column volumes of Ni-NTA wash buffer was applied to column and the flow through collected. The collected samples were checked for release of protein from the column by Bradford assay and the principles of dye binding using the Bio-Rad protein assay dye reagent. Ni-NTA wash buffer was usually added

until no protein was released from the column, except in those cases where the concentration of imidazole within the Ni-NTA wash buffer was shown to release the protein of interest from the column, determined empirically in each case. Following washing of the column, proteins were eluted from the resin by application of Ni-NTA elution buffer in 5 mL volumes, with each fraction collected separately. Ni-NTA elution buffer was applied until the Bradford dye binding assay showed no protein actively being eluted from the column. Collected flow through, wash, and elution fractions were subsequently analysed by SDS-PAGE to determine the and identify and identify the protein of interest.

2.5.4 Gel Filtration Chromatography

Gel filtration chromatography was carried out prior to many downstream, assay applications to supply purer samples of protein for use. Concentrated fractions containing the protein of interest post Ni-NTA affinity chromatography, were pooled and concentrated to a volume of 1 – 2 mL. Alongside the concentration of the protein sample, a gel filtration chromatography column (usually a column with HiLoad Superdex 200 26/600 prep grade resin) was fitted to an ÄKTA Pure system (GE Healthcare) and equilibrated, first with two column volumes of deionised and degassed water, followed by two column volumes of filtered and degassed gel filtration buffer. A pressure-controlled flow rate of 2.6 mL/min was utilised, whilst the whole system was maintained at 4 °C. The sample was then loaded onto the column, followed by separation and elution of proteins by the application of one column volume of gel filtration buffer. Elutions were collected in 3 mL volumes and analysed for the presence of the desired protein by SDS-PAGE. Samples containing the pure protein of interest were again pooled and concentrated, followed by immediate downstream use or storage at -80 °C.

2.5.5 SDS-PAGE Analysis of Proteins

Sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse protein of interest based on the respective size of the protein. SDS-PAGE gels were made according to the following protocol.

Table 12: Resolving gel for SDS-PAGE gels:

Component	Concentration	Volume for 10 % w/v gel (mL)	Volume for 15 % w/v gel (mL)
Deionised water	-	1.773	1.148
Acrylamide/bis solution (37.5: 1)	40 % v/v	1.250	1.875

Tris (pH 8.8)	1 M	1.875	1.875
SDS	10 % w/v	0.050	0.050
Ammonium persulfate (APS)	10 % w/v	0.050	0.050
TEMED	-	0.005	0.005

Table 13: Stacking gel for SDS-PAGE gels:

Component	Concentration	Volume (mL)
Deionised water	-	1.328
Acrylamide/bis solution (37.5: 1)	40 % v/v	0.255
Tris (pH 6.8)	1 M	0.375
SDS	10 % w/v	0.020
Ammonium persulfate (APS)	10 % w/v	0.020
TEMED	-	0.002

Gel mixtures, as listed above, were prepared in universal tubes and APS and TEMED were added last, prior to pouring of the liquid mixture between glass plates. Isopropanol was added above the resolving gel before it was allowed to set at room temperature. The process was repeated for the stacking gel, layered upon the resolving gel and allowed to solidify, after the addition of a comb at the top of the glass plates, designed to impregnate sample wells into the gel. Samples were prepared for electrophoresis by addition of 5 μL of Laemmli sample buffer (Bio-Rad) to 15 μL of protein sample, followed by boiling of the sample at 95 °C for 10 minutes. Samples were loaded onto the gel, alongside a molecular marker ladder, held in a gel tank and filled with 600 mL of 1x SDS-PAGE running buffer per gel. Samples were electrophoresed at 180 V for 45 – 55 minutes, followed by release of the gel and staining with InstantBlue (Expedeon) with gentle rocking overnight for visualisation of protein bands.

2.5.6 General Protein Handling, Analysis, and Storage Techniques

Protein samples were concentrated down to the volumes and concentrations required for downstream applications and assays, such as further purification by gel filtration chromatography or other *in vitro* biochemical analyses to further understand protein functionality. Concentration of samples was carried out using Amicon Ultra centrifugal filter units (Merck, 10 kDA MW cut-off) where the samples was applied to the filter and centrifuged (4500 x *g*, 12 °C, 10 minutes, as many times as required). Sample was constantly applied and resuspended to avoid precipitation of the sample or sticking of protein to the cellulose

membranes. Samples were centrifuged and concentrated until the required concentration and/or volume was obtained.

Quantitation of protein samples was carried out using a Geneflow nanophotometer, using the principles of the Beer-Lambert law and calculated values for the protein molecular weight and molar extinction coefficient for a more accurate estimate of the concentration of the sample. Initially, 2 µL of buffer was used to give a zero or blank reading on the nanophotometer, followed by measurement of 2 µL of the protein sample (within the same buffer system as used to blank the photometer). The nanophotometer yielded an output absorbance (A_{280}) reading as well as concentration in mg/mL. These values were used to calculate molar concentrations of protein as required.

Protein samples often required desalting or dialysis in order to exchange buffer systems for downstream applications. Desalting was carried out using PD10 desalting columns (GE Life Sciences) and as per manufacturer's instructions. Briefly, the ethanol storage solution was decanted and the column was equilibrated with 25 mL of elution buffer by gravity flow. After equilibration, 2.5 mL of protein sample (sample diluted to 2.5 mL if lower volume) was applied to the column and allowed to run into the column by gravity flow. Samples were eluted from desalting columns by application of 3.5 mL of elution buffer and the flow through was collected, giving a buffer exchanged sample of protein. Dialysis of protein samples was carried out using pre-soaked snakeskin dialysis tubing (Scientific Laboratory Supplies) into which the protein sample was applied. The ends of the tubing were clipped shut to avoid sample loss, and the sample-containing tubing was placed in a beaker filled with the required solvent (at least 300x the sample volume of buffer was used). The tubing was stirred in the buffer overnight at 4 °C, with occasional rotation of the buffer with a freshly prepared solution.

Proteins were usually prepared on a large scale and a large amount of the sample prepared was not immediately required for downstream biochemical applications. Therefore, the protein not rapidly required was stored for future use, usually in a buffer solution containing 10 % v/v glycerol. The total remaining sample was exchanged into a buffer containing 10 % v/v glycerol and split into 100 or 200 µL aliquots (dependent upon protein concentration). The aliquots were flash frozen in an ethanol-dry ice bath and stored at -80 °C for future use.

2.6 Generation and Analysis of Chemical Extracts

2.6.1 Bacterial Growth for Extract Preparation

Optimal growth conditions and media were established empirically in each case to determine the most favourable conditions for specialised metabolite production. Commonly SFM, ISP2, and TS media were tested in both solid and liquid forms to analyse compound production. The selected growth temperature was always 30 °C and the length of incubation ranged from 5 – 7 days. For example, in this study, surugamides were most commonly extracted from solid or liquid ISP2 medium after five days of growth. Antimycins, however, were extracted from liquid SFM medium after seven days of culture of *Streptomyces albus* S4.

2.6.2 Isolation of Antimycins by Solid Phase Extraction (SPE)

Production of antimycins occurs readily by *Streptomyces albus* S4 after culturing in liquid or solid SFM medium. For this study, 50 mL liquid SFM cultures were utilised for the production and analysis of antimycins. After growth of the bacteria for seven days at 30 °C, bacterial cells were removed by centrifugation (4000 x g, 15 minutes), yielding ~ 40 mL of supernatant after considering evaporative loss during the seven days of culture. In total 20 mL of the culture was extracted by solid phase extraction using SPE columns (Phenomenex Strata-XL C18 (100 μm, 100 mg, 3 ml)). SPE columns were first washed with one column volume of 100 % v/v methanol, after which culture supernatant was applied to SPE columns in 3 mL fractions and passed through by the function of a vacuum manifold containing a waste solvent trough. Columns were then washed with one column volume of water to remove polar and water-soluble compounds, and subsequently washed with two column volumes of 30 % v/v methanol. Remaining column-bound compounds were eluted by application of 300 μL of 100 % v/v methanol, which was collected in a microcentrifuge tube. Samples were centrifuged (15900 x g, 10 minutes) to remove and pellet any solid components before the supernatant was transferred to a sample vial for subsequent analysis.

2.6.3 Solvent Extraction of Solid Media

In the laboratory, *Streptomyces* spp. often produce the target compound only when cultured on agar plates (SFM, ISP2, TSA) for the requisite amount of time (5 – 7 days). After complete growth of the bacteria, spores were either collected and stocked to act as inoculum for future growth or remained on the plates for solvent extraction. Initially, agar plates were cut into small squares using a clean and dry scalpel or spatula, and transferred to a conical flask. Solvent was then added to the conical flask containing the agar slices at a volumetric ratio of 2 (solvent volume): 1 (agar volume, 1 small agar plate = 25 mL medium) and the mixture was left at room

temperature overnight, with gentle shaking on a rotary platform. Commonly used solvents in this study included ethyl acetate and methanol, with ethyl acetate used for extraction of surugamides from *Streptomyces albus* S4. After overnight extraction of compounds, the solvent was decanted into a round-bottomed flask and evaporated to dryness *in vacuo* using a rotary evaporator. The solid residue was resuspended in 300 – 500 µL of methanol and subsequently centrifuged (15900 x *g*, 10 minutes) to remove insoluble material, yielding samples ready for downstream analysis.

2.6.4 Solvent Extraction of Liquid Media

The production of specialised metabolites from various *Streptomyces* species' in liquid media has been well characterised and offers a scalable method for production of natural product type compounds. Media was prepared to the required volume, predominantly 50 mL for small-scale growths or 500 mL for large-scale cultures, and inoculated with concentrated spores or pre-cultured mycelia of the required strain of *Streptomyces*. After growth for 5 – 7 days as required, cultures were centrifuged (4000 x *g*, 15 minutes) to pellet bacterial cells. The total supernatant was then extracted with two volumes of a suitable solvent, followed by complete drying of the solvent layer *in vacuo*. For example, for the extraction of surugamides from large-scale liquid ISP2 cultures grown for five days at 30 °C, the supernatant (450 mL) was extracted with 900 mL ethyl acetate in a separating funnel, followed by elution and collection of the solvent layer. The solvent was then evaporated to dryness *in vacuo*, followed by resuspension of the solid residue in a minimal volume of methanol (300 – 500 µL) for centrifugation (15900 x *g*, 10 minutes) to remove any remaining solids and subsequent analysis.

2.6.5 HPLC Analysis of Chemical Extracts

Analysis of chemical extracts prepared from the cells and culture media of *Streptomyces* spp. was carried out using numerous methodologies. Initial analysis of these extracts was carried out using high performance liquid chromatography (HPLC) to give an indication of concentration, complexity, and composition of the respective extracts. Compounds were separated on reverse phase HPLC columns, with attached guard cartridges containing the same column chemistry (most commonly used columns were Phenomenex Luna C18 (150 x 4.6 mm, 100 Å pore, 5 µm particle) or Phenomenex Kinetex C18 (50 x 4.6 mm, 100 Å pore, 5 µm particle)). The solvents used for running HPLC analyses were: solvent A – water with 0.1 % trifluoroacetic acid (TFA) (v/v); solvent B – acetonitrile with 0.1 % TFA (v/v), and the flow

rate was maintained at 1.0 mL/min. A typical solvent gradient as used for the HPLC analysis was as follows: 0.0-1.0 min, 5% B; 1.0-10.0 min, 5-95% B; 10.0-17.0 min, 95% B; 17.0-20.0 min, 95-5% B; 20.0-30.0 min, 5% B. Post-run analysis was performed in the Clarity analytical software.

2.6.6 LC-HRMS(MS) Analysis of Chemical Extracts

To obtain a more detailed analysis of the composition of chemical extracts generated, samples were analysed by liquid chromatography high resolution mass spectrometry (LC-HRMS) or alternatively fragmentation of compounds occurred during the MS run to give analytical LC-HRMSMS data. LC-HRMS(MS) was carried out on a Bruker MaXis Impact II time of flight mass spectrometer coupled with a Dionex UltiMate 3000 HPLC equipped with a Waters Acquity Peptide CSH C18 column (100 x 2.1 mm, 1.7 μm). The system was calibrated after each run by the injection of sodium formate and mass spectra were generally acquired using electrospray ionisation in positive ion mode. In a typical run the following gradient was used: (solvent A, 0.1% formic acid in water (v/v); solvent B, 0.1% formic acid in acetonitrile (v/v); flow rate, 0.7 mL/min) 0.0–1.5 min, 1% B; 1.5–3.5 min, 75–95% B; 3.5–4.5 min, 95% B; 4.5–5.0 min, 95–1% B; 5.0–5.5 min, 1% B. Post run analysis was performed using the Bruker Data Analysis software (version 4.1).

2.6.7 Semi-Preparative HPLC Purification of Compounds

Compounds from chemical extracts, or more commonly from chemical reaction mixtures, were often required in a purer state for further analytical techniques and methodologies. For the purification of compounds with signature UV absorbances (due to the characteristics of the detector), semi-preparative HPLC was carried out to separate required compounds from complex mixtures using a Phenomenex Kinetex C18 Evo column (150 x 10.0 mm, 100 Å pore, 5 μm particle) equipped with C18 guard cartridge, and employing a flow rate of 4.7 mL/min. The solvents used for running semi-preparative HPLC purifications were: solvent A – water with 0.1 % trifluoroacetic acid (TFA) (v/v); solvent B – acetonitrile with 0.1 % TFA (v/v). A typical solvent gradient as used for semi-preparative HPLC was as follows: 0.0-1.0 min, 5% B; 1.0-10.0 min, 5-95% B; 10.0-17.0 min, 95% B; 17.0-20.0 min, 95-5% B; 20.0-30.0 min, 5% B. Fractions were collected manually after detection and visualisation of the required UV-vis peak.

2.6.8 Preparative HPLC Purification of Compounds

Compounds from generated chemical extracts, or from chemical reactions, often did not contain signature UV absorbances, abrogating the use of UV detectors in a semi-preparative HPLC set up. In these scenarios, as well as in cases where large volumes of sample required purification, preparative HPLC was utilised. All preparative HPLC runs were carried out on an Agilent 1260 Mass-Directed Preparative HPLC instrument equipped with a Phenomenex Kinetex EVO C18 column (100 × 21.2 mm, 5 μm). The solvent system was as follows: solvent A, 0.1% formic acid in water (v/v); solvent B, 0.1% formic acid in acetonitrile (v/v). The following gradient was used: 0-2 min, 5% B; 2-12 min, 5-95% B; 12- 14 min, 95% B; 14-15 min, 100-5% B. The flow rate was 20 ml/min. Fraction were automatically collected, after detection by the related mass signature peak, followed by pooling of the fractions and drying *in vacuo* to obtain the pure compound.

2.7 Biochemical Assays

2.7.1 Circular Dichroism (CD) Analysis of Proteins

Circular dichroism (CD) was carried out on purified proteins and allowed an initial determination of the secondary structural elements composing a protein of interest. Proteins were initially exchanged into required buffer systems (e.g. a HEPES based gel filtration buffer or phosphate buffered saline (PBS)). CD spectra (190 – 260 nm) were recorded using 200 μl protein solution (concentration approximately 0.2 mg/ml) in a 1 mm path length cuvette. CD experiments were conducted at 20 °C on an APP Chirascan CD spectropolarimeter.

2.7.2 Isothermal Titration Calorimetry (ITC)

ITC was carried out using purified proteins (the cell component) and a second purified protein or small molecule cofactor (the syringe component), and allowed the determination of numerous thermodynamic factors, such as the dissociation constant (K_D), the enthalpy of the interaction (ΔH), and the stoichiometry of the interaction. The cell component, at a concentration of 20 – 50 μM, was held in a thermostatted cell at 25 °C, with a reference power of 5 μcal s^{-1}. 40 μL of the syringe component, at a concentration 10x that used for the cell component, was taken up into the syringe. The system was primed and the syringe was inserted into the cell, with the syringe stirring at 750 rpm. Titration of the syringe component into the cell component occurred automatically according to the following timetable:

Table 14:

Injection number	Injection volume (μL)	Injection duration (seconds)	Time gap (seconds)
1	0.5	2	120
2 – 20	2	2	120

Heat changes as part of the raw thermograms were processed using NitPic, SEDFIT, and GUSSI (Schuck, 2000; Scheuermann and Brautigam, 2015; Brautigam et al., 2016) to yield sigmoidal isotherm plots for binding interactions and to estimate thermodynamic parameters of detected binding events.

2.7.3 Analytical Size Exclusion Chromatography

Analytical size exclusion chromatography (analytical-SEC) was carried out to determine and observe interactions between purified proteins, giving an indication as to the physiological nature and importance of such an interaction in *in vivo* settings. Analytical-SEC was carried out on a Superdex 200 10/300 GL column, equilibrated in a filtered and degassed 50 mM HEPES, 200 mM NaCl buffer, and elution of proteins and mixtures was subsequently carried out with the same buffer system. A set of protein markers was initially applied to the column and eluted, to give an indication of the size distribution of the column and to report elution volumes and times of marker protein of a known size. Test proteins were then run on the column individually to reveal the elution volume of individual proteins and to confirm proteins were eluting as expected, when compared to known marker proteins. The test proteins for which interaction complexes were putatively assigned or expected were mixed, usually in a 1: 1 molar concentration ratio, and incubated at room temperature for varying lengths of time. The mixed protein samples were then applied to the column and eluted using an identical buffer system. Fractions were collected to maintain protein stocks and elution volumes were checked to observe whether interaction complexes or individual proteins of expected sizes were eluted. Therefore, putative interactions could be assigned to occur between proteins of interest, and the data used to in conjunction with other methods to assess protein-protein interactions.

2.7.4 Spectrophotometric Analysis of Ketoreduction

Assays to test the ability of proteins to carry out the catalytic reduction of various ketone substrates were carried out at 30 °C in a BMG Labtech FLUOstar Omega microplate reader. Typical assay conditions were: 25 mM HEPES (pH 7.5), 1 mM NADPH, 0 – 5 μM protein,

and 0 – 50 mM substrate (dissolved in 25 mM HEPES with 3 % v/v DMSO). Subsequently, assays were also made with 250 mM and 500 mM final substrate concentrations. Reactions were monitored for their decrease in absorbance at $\lambda = 340$ nm, showing the usage of NADPH throughout the assay. The total assay volume was 150 µL and total assay time was 20 minutes, with readings taken every minute. Reactions were initiated by the addition of substrate after a preincubation period of enzyme with NADPH, lasting a total of 20 minutes. Assay plates were shaken for 3 seconds before each reading, and absorbance readings were corrected for the absorbance in the absence of enzyme. Raw readings were converted to rates of NADPH conversion to visualize the kinetic rates of each substrate and reaction condition.

2.7.5 β-galactosidase Based Analyses of Protein-Protein Interactions

Initial indication as to the formation of *in vivo* interaction complexes between specific proteins was obtained by carrying out β-galactosidase-based assays. Initially, the BAC-2-Hybrid (BAC2H) system was utilised to interrogate interactions between proteins (Karimova et al., 1998). The genes for proteins of interest were cloned in-frame, into BAC2H plasmids, pUT18 and pKT25, which each contained half of the split *E. coli* adenylate cyclase protein. The cloned plasmids were transformed into the either the non-revertant adenylate cyclase mutant BTH101 or DHM1 *E. coli* strains. The double transformed strains were plated onto LB agar containing IPTG and X-gal as an indicator. Interacting proteins lead to the breakdown of the chromogenic substrate and the formation of blue colonies, whilst strains producing non-interacting protein pairs remained white. A further β-galactosidase based assay utilised was an *in vitro o-*nitrophenyl β-D galactopyranoside (ONPG) assay with BAC2H cell lysates. Cells containing both cloned BAC2H plasmids were grown, either overnight or subcultured to a specific OD, and gently lysed. Insoluble components were pelleted by centrifugation and to the soluble proteins was added a solution of 2 mM ONPG. The cleavage reaction was allowed to occur for 30 minutes before the reactions were stopped by the addition of a final concentration 1.0 M sodium carbonate. The cleavage and release of the chromogenic substrate was measured in a plate reader at a wavelength of 420 nm, giving an indication of the amount of β-galactosidase present, and therefore, the nature and presence of interacting protein partners within the cell lysate.

2.7.6 Peptide-N-acetylcysteamine (Peptide-SNAC) Assays

The ability of proteins to utilise, and cleave or cyclise, SNAC-peptide substrates was assessed *in vitro* using the following reaction conditions in a final volume of 100 µL: 50 mM HEPES (pH 7.5), 0.1 mM SNAC-peptide, and 2 – 20 µM of purified protein. The enzymatic reaction or SNAC-peptide alone were incubated at 30 °C. After 1 – 5 hrs of incubation, samples were treated with 100 µL of 0.05 % v/v TFA and extracted with Phenomenex Strata-XL C18 (100 µm, 30 mg, 1 ml) solid-phase extraction columns and a vacuum manifold. Metabolites were eluted from the column with 120 µl of 100 % methanol. The methanolic extract was diluted 1:1 with 100 % v/v methanol prior to centrifugation for 10 mins at 15900 x *g* to remove any insoluble material before analysis by LC-HRMSMS on a Bruker MaXis Impact II TOF mass spectrometer equipped with a Dionex UltiMate 3000 high-performance liquid chromatography apparatus (using conditions described in section 2.6.6).

2.7.7 Crystal Trials and Structure Determination

Concentrated, ÄKTA-purified proteins were supplemented with cognate cofactors, usually to a final concentration of 1 mM, after which apo- and co-factor treated protein samples were subjected to crystallographic studies. Crystal trials were performed with the JCSG Core Suites I – IV (Qiagen) using proteins at final concentrations of 7 – 12 mg ml^{-1}. The trials were set up in 200 nL volumes (100 nL protein solution, 100 nL crystallisation condition), in 96 well crystallisation plates using a Formulatrix NT8 robot. Crystals were identified after multiple weeks of incubation at 25 °C using the Formulatrix storage hotel and imaging system. Plates containing putative crystals were obtained and protein crystals were picked and looped. The crystals used for data collection were formed in 0.1 M citrate phosphate buffer, pH 4.2, 0.2 M NaCl, and 10 % w/v PEG 3K (*apo*-AntM) and 0.1 M Tris, pH 8.5, 5 % w/v PEG 8K, 20 % v/v PEG 300, 10 % v/v glycerol (AntM with NADPH) and were cryoprotected by the addition of 50 % v/v glycerol before plunging into liquid nitrogen for storage. X-ray diffraction data was collected on beamline i04 at Diamond Light Source. All collected diffraction data was processed in CCP4, using the CCP4i2 graphical user interface. Data was initially scaled and reduced in AIMLESS and a suitable search model was identified and truncated in CHAINSAW. The structure of AntM in complex with NADPH was solved to 1.7 Å using the UrdMRed ketoreductase (PDB ID: 4osp) from the angucycline biosynthetic pathway as the search model in PHASER. The final structural model was generated by iterative rounds of manual model building in COOT and refinement in REFMAC5 (Emsley and Cowtan, 2004; McCoy et al., 2007; Stein, 2008; Murshudov et al., 2011; Winn et al., 2011, p.4; Evans and Murshudov, 2013).

2.8 Chemical Syntheses

2.8.1 Solid Phase Peptide Synthesis (SPPS) on 2-chlorotrityl chloride resins

SPPS for the preparation of peptides on solid supports was carried out using a step-by-step protocol, following established methodologies and developed for a simple and standard approach to many peptide syntheses. SPPS couplings were carried out in solvent-stable tubes with an attached tap, acting as a valve to release solvents and wash solutions. Peptide syntheses were predominantly carried out on 100 mg resin scales (resin loading 1.47 mmol g^{-1}), therefore requiring 77 mg of Oxyma (coupling additive, 3.5 eqs) and 35 μL of diisopropylcarbodiimide (DIC) (coupling reagent, 2 eqs). All amino acids utilised were N-terminal Fmoc-protected, apart from the terminal amino used in the synthesis (N-terminal of released peptide) which was Boc-protected. The protocol for the complete synthesis was as follows:

1. The resin was swelled in dichloromethane (DCM) for at least 30 minutes at room temperature, with stirring.

2. The first coupling reaction was carried out using 1.2 equivalents of protected amino acid and 4 equivalents of diisopropylethylamine (DIPEA) in DCM. The mixture was subsequently stirred for ~2 hours.

3. The unreacted handles on the resin were end-capped using sequential washes of 3x DCM/MeOH/DIPEA (17:2:1), 3x DCM, and 2x dimethylformamide (DMF). After the addition of each solvent or solvent mixture, the tube was shaken to ensure mixing of the resin and the valve opened to release any built-up pressure.

4. The first amino acid was deprotected using washes of 2x5min 20 % v/v piperidine in DMF and 5x1min DMF.

5. The next amino acid coupling solution was prepared by weighing out 5 equivalents of the subsequently required amino acid and oxyma in a vial. DMF was added to dissolve all components and the required volume of DIC was added to the solution. The solution was added to the resin and stirred for ~ 1 hour.

6. The coupling solution was eluted from the column and the resultant peptide deprotected using the series of washes; 3x2min DMF, 2x5min 20 % v/v piperidine, 5x1min DMF.

7. Steps 5 and 6 were repeated for all subsequently required amino acids.

8. For the removal of peptides from the resin with N-terminal Boc protection intact, the following washes were carried out: 3x1min DMF, 3x1min DCM, 3x1min MeOH. The resin was then dried over a vacuum or with a nitrogen line.

9. A solution of hexafluoroisopropanol (HFIP) in DCM, 1:4, was added to the dried resin and stirred for 1 hour, followed by elution of the solvent into an RB flask. A second cleavage and elution was carried out using 1:4 HFIP in DCM for 30 minutes to release all bound peptides and maximise yield. The solvent was then dried *in vacuo* to obtain the crude Boc-protected peptide.

If the terminal Boc protection was not required, trifluoroacetic acid (TFA) was used to cleave peptides of the resin, as well as affecting total deprotection of any protected side chains. The crude peptide, obtained after drying the peptide release solution, was purified using semi-preparative or preparative HPLC to obtain only the required peptide.

2.8.2 Synthesis of Peptide-SNACs

The crude protected peptide was dissolved in DCM (2 ml) to which was added DIC (63 mg, 0.5 mmol), HOBT (67 mg, 0.5 mmol) and N-acetylcysteamine (60 mg, 0.5 mmol). The solution was stirred at room temperature overnight followed by removal of solvent *in vacuo*. The crude thioester was globally deprotected by dissolving in 9:1 TFA: DCM and stirred at room temperature for 3 hours. The volatiles were removed in vacuo and the crude sample dissolved in a minimal volume of cold water. Ice cold diethyl ether was used to precipitate the peptide which was collected by centrifugation (4000 x *g*, 5 minutes, 4 °C) to afford the crude peptide-SNAC, which were subsequently purified by semi-preparative or preparative HPLC.

2.8.3 Solid Phase Peptide Synthesis (SPPS) on PEGA resins

SPPS on PEGA resins allowed the synthesis and attachment of peptides to resins through a synthesised linker, which more closely mimicked the 4'- phosphopantetheinyl linker used in the thiotemplating of NRPs when being synthesised on multimodular NRPSs. Initially the linker, composed of suberic acid, β-alanine, and ethanolamine, was synthesised on the resin following a previously established protocol (Kohli et al., 2002). Subsequent to complete linker formation, the desired peptide was synthesised on the resin in a manner analogous to the SPPS carried out on resins of other types and chemical compositions. PEGA resin syntheses were carried out on 100 mg scales (resin loading 0.47 mmol g^{-1}). A step by step protocol was established to allow standard synthesis of multiple PEGA resin-based peptide substrates:

1. PEGA resin was swelled in DCM for at least 30 minutes at room temperature, with stirring.

2. The resin was coupled with monomethyl suberic acid using 4.9 equivalents of benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate (PyBOP), 5 equivalents of hydroxybenzotriazole (HOBt), and 10 equivalents of DIPEA in DMF, with stirring for 2 hours.

3. The resin was washed with DMF and the coupling in step 2 was repeated.

4. The ester methyl group was deprotected to leave the free acid, using a base deprotection step of 30 minutes stirring in tetrahydrofuran (THF)/ MeOH/ 10N NaOH (6: 3: 1). This was followed by a wash with dilute acid.

5. 5 equivalents of β-alanine methyl ester were then coupled to the resin using the same procedure as in step 2.

6. The resin was washed with DMF and a second coupling of 5 equivalents of β-alanine methyl ester was carried out, followed by base deprotection (as in step 4) and dilute acid wash.

7. 20 equivalents of ethanolamine•HCl were coupled to the resin using the same method as in step 2, to result in PEGA resin with an appended chemical linker.

8. Amino acids required to form the requisite substrate peptide were coupled to the resin via the linker. This was carried out following same steps as performed for SPPS on 2-chlorotrityl chloride resins (Section 2.8.1, steps 5 and 6).

9. PEGA resin – linker – peptide substrates were used without further cleavage or purification in enzyme assays as required.

If cleavage of the peptide off the resin was required, for example to confirm the identity of the peptide, a solution of 1 M potassium hydroxide in methanol was added to the resin and stirred for two hours at room temperature. The resin was pelleted by centrifugation and soluble peptides were collected and analysed by HPLC or LC-HRESI-MSMS.

2.8.4 Synthesis of Amino Acid-Coenzyme A Conjugates

Amino acid – coenzyme A conjugates were synthesised for attempts to subsequently attach amino acids to PCP domains for downstream enzyme assays. The synthesis was performed according to previously described protocols (Stein et al., 2005). One equivalent of amino acid was mixed with 1.5 equivalents of coenzyme A, 1.5 equivlents of PyBOP, and 4 equivalents of potassium carbonate. All components were dissolved in a 1:1 mixture of THF and water.

The mixture was agitated for two hours at room temperature, after which the solvents were dried *in vacuo*. The amino acid amino group was then deprotected using a mixture of TFA, triisopropylsilane (TIPS), and water (95: 2.5: 2.5) for three hours at room temperature. The solvents were again evaporated *in vacuo* and the mixture was solubilised in a minimum volume of methanol. Amino acid – coenzyme A conjugates were purified from the mixture using semi-preparative or preparative HPLC (as detailed in sections 2.6.7 and 2.6.8).

2.9 Bioinformatic and Computational Analyses

2.9.1 General Bioinformatic Techniques

General and basic bioinformatic techniques were predominantly carried out using the corresponding internet servers as required. BLASTn and BLASTp (Altschul et al., 1990) analyses were used to compare nucleotide and amino acid sequences respectively, against large collated databases. Results gave indications as to the functionality of genes and proteins of unknown function, as well as corroborated functions of known proteins. Nucleotide sequences were translated to amino acid sequences by use of the ExPasy Translate tool and output protein sequences were analysed using the ExPasy ProtParam platform (Gasteiger et al., 2003), giving basic molecular parameters such as molecular weight and molar extinction coefficients. Amino acid sequence alignments were carried out using the Clustal Omega multiple sequence alignment tool or similar Clustal pairwise alignment platforms (Madeira et al., 2019). Genomes and genomic records were obtained from the NCBI Genbank database (Clark et al., 2016) and analysed for biosynthetic gene clusters by direct upload onto the AntiSMASH server (Medema et al., 2011; Blin et al., 2019). Results provided information on the BGC content of genomes and allowed initial visualisation of genes present within BGCs of interest. Records of known and annotated BGCs were accessed through the MiBIG database (Medema et al., 2015; Kautsar et al., 2020) and the BGC content of many *Streptomyces* genomes was available through the AntiSMASH database platform (Blin et al., 2017; Blin et al., 2021).

2.9.2 *In silico* Substrate Docking

Molecular docking of potential substrates into protein active site was carried out using AutoDock Vina 1.1.2 (Trott and Olson, 2010). AutoDock Tools 1.5.4 was used to prepare the ligands (potential substrates) and the protein (protein crystal structure) for docking. All nonpolar hydrogen atoms within the ligands were merged and all single bonds between atoms were defined as rotatable. A grid of a suitable size for protein binding site (e.g. 16 x 32 x 16 Å

for the AntM binding pocket) was defined within the protein, centred on the active site binding pocket and the location of any bound cofactors (e.g. NADPH). The output ligand-docked poses were manually assessed for parameters pertinent to substrate binding including, in the case of AntM substrate docking, but not limited to: active site binding; proximity to NADPH; and localization of the terminal sulfur atom. The most optimal poses were further analysed to aid comparisons between substrates.

Chapter 3 – CRISPR/Cas9 engineering of polyketide synthase substrate utilisation

Abstract

Streptomyces species are prodigious producers of natural products, often termed secondary or specialised metabolites, with an extremely wide variety of uses, both in the clinic and in research applications. The predominant natural use of these compounds in bacterial niches is as defence compounds against rival microorganisms, however siderophores, scavenging molecules, and signalling compounds are also commonly produced. Antimycins are hybrid non-ribosomal peptide/polyketide compounds with potent antifungal, antinematodal, and piscicidal activities. The molecules comprise a rare 3-formamidosalicylate starter unit appended to a 9-, 12-, 15-, or 18- membered ring, containing multiple lactones and a depsipeptide group. Although multiple different isoforms of antimycins are produced *in vivo*, a major area of research has been engineering novel antimycin derivatives, either through total chemical synthesis, or gene cluster and pathway engineering. Recently, the major method of engineering multimodular pathways and altering biosynthetic gene clusters has been the utilisation of CRISPR/Cas9 to modify sections of *Streptomyces* genomic DNA. The primary aim of this work was to establish the feasibility of using an *Escherichia coli*-based CRISPR/Cas9 system for precision engineering of natural product biosynthetic gene clusters, cloned onto large extrachromosomal pieces of DNA. The use of cloned clusters allows fast testing by heterologous expression and alleviates issues surrounding the alteration and mutation of genomic DNA, whilst the utilisation of *E. coli* as the host strain allows quick mutation and efficient selection of successfully altered constructs. A proof of principle study was carried out on the antimycin biosynthetic gene cluster, in which three active site variants of the acyltransferase domain of the sole polyketide synthase module, AntD, were produced. Surprisingly, two of the variants completely precluded the production of antimycins, whilst the final mutation facilitates the production of a novel antimycin variant. These results showed that fast *E. coli* based CRISPR/Cas9 alteration of cloned biosynthetic gene clusters is a viable technique for the precision engineering of large multimodular pathways, and the potential production of novel compounds with interesting activities.

3.1 Introduction

      3.1.1 Acyltransferase domains as a source of diversity in polyketide biosynthesis

Polyketide synthases are macromolecular machines, utilising a modular biosynthetic logic for the synthesis of complex molecules with extremely varied bioactivities. The core and accessory enzymatic domains involved in PKS biosynthesis (KS, AT, ACP, KR, DH, ER) all have defined roles, acting upon the intermediate tethered to the assembly line via a flexible Ppant moiety, permitting access to the active sites of these domains. In any enzymatic process, substrate selection plays a central role in ensuring product formation, particularly in reaction and catalytic cascades where subsequent enzymatic processes may be affected. AT domains are responsible for substrate selection and activation in the PKS biosynthetic process, thereby acting as the primary source of diversity observed in polyketide products (Khosla et al., 2014). Further diversification of polyketide intermediates is carried out by the remainder of the catalytic domains, particularly the β-keto processing accessory domains, with post-PKS enzymes also further modifying the scaffold. Most PKS extension modules, however, contain AT domains with the capability of activating malonyl- or methylmalonyl-CoA and therefore extending the intermediate by a two-carbon unit, with either no branching side group or a methyl chain, respectively. A number of AT domains, however, can utilise acyl-CoAs other than mCoA and mmCoA, leading to the biosynthesis of polyketide products with altered branching groups (Wang et al., 2015). Further, crotonyl-CoA reductase enzymes often provide AT domains with variable CoA-tethered substrates for activation and incorporation into growing intermediates (Wilson and Moore, 2012). The AT domains must therefore, be able to recognise specific, cognate substrates and subsequently function with the remainder of the module for tethering to ACP domains and downstream catalysis (Dutta et al., 2014; Mathews et al., 2017). The fact that AT domains play such key roles in PKS biosynthesis, particularly in substrate selection and activation, has led to immense interest in their modulation and alteration for the production of novel variants of metabolites and non-natural products.

3.1.2 Engineering PKSs and AT domains for altered functionality

As AT domains are the predominant enzyme responsible for the selection of substrates in PKS elongation modules, engineering their substrate specificity provides an obvious opportunity to alter the nature of the final compound produced (Dunn and Khosla, 2013). Since the discovery of PKSs, a large amount of energy has been focussed on the alteration and precision engineering of the complex pathways, particularly for the biosynthesis of novel compounds from type I, modular synthases. Therefore, a number of methodologies have been developed

with varying levels of success in returning functional systems and novel metabolites, including module swapping, domain swapping, and targeted site-directed mutation (Figure 3.1).
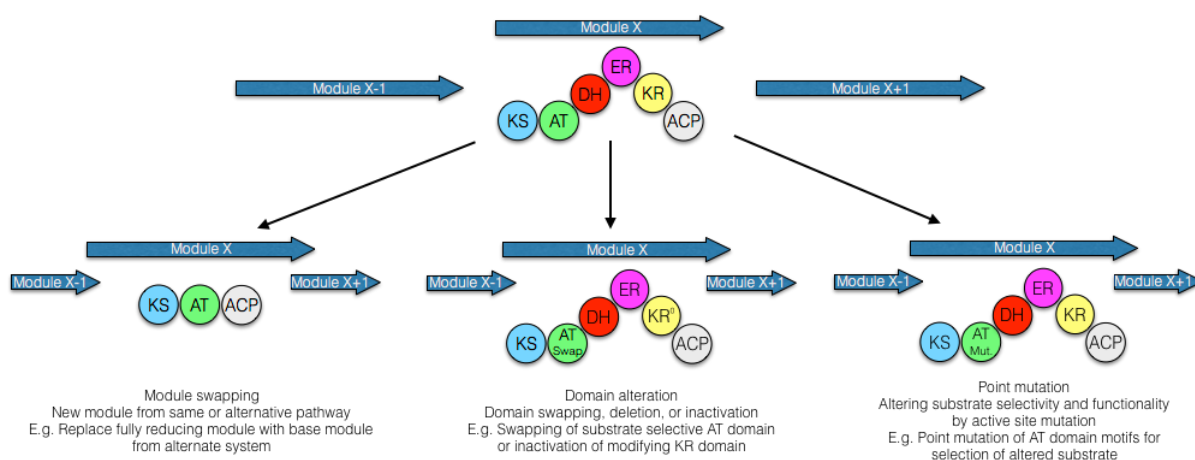
Figure 3.1: Overview of methodologies for the engineering of PKS systems. Module swapping, domain alteration, and point mutation strategies have all been utilised to alter PKS pathways, with varying levels of success. The underlying rationale of any engineering strategy is the alteration of substrate usage and the subsequent incorporation of the altered substrate into intermediates and final compounds formed.

Precision engineering of enzymes for altered catalytic functionality requires intimate and detailed knowledge about the structure, interactions, and dynamics of the protein, in order to rationally change the scope of protein function. In the context of PKSs, although much of the structural details are as yet unknown, several recently solved structures of domains, didomains, and modules have shed light on the structure of PKSs, and how the structure allows function and the biosynthesis of varied, complex compounds (discussed in section 1.5.3). PKSs share large amounts of functional similarities to fatty acid synthases, responsible for the production of long chain fatty acids by the use of a fully reducing modular biosynthetic logic (Revill et al., 1996; Smith and Tsai, 2007). The two systems were also found to share certain structural features; however, larger disparities were found between the types of proteinaceous pathways upon further analysis. Initial studies of PKS domains and modules were necessarily focussed on short constructs, particularly didomains, such as the structures from the erythromycin synthesising DEBS system (Khosla et al., 2007). More recent studies have focussed on complete, partially reducing modules (e.g. pikromycin module 5), which contradicted structures of isolated domains and didomains. Nonetheless, the structural analyses have provided insight into the complex functional states of PKS systems, with the dimeric enzymes

adopting complex arch-shaped folds and the interdomain linkers playing important roles in substrate shuttling and maintaining domain functionality (Gokhale and Khosla, 2000; Dutta et al., 2014; Whicher et al., 2014). The baseline knowledge of domain and module connectivity, as well as the in-depth knowledge of domain catalytic functionality, has allowed varied attempts at engineering systems and the alteration of function and compound production.

The modular nature of the vast majority of known, well understood, and bioactive compound-producing PKS systems, allows the simple theoretical creation of novel natural products through harnessing of the modularity of the pathways. For example, chimeric systems could be created by modular substitution or the fusing of two modules from distinct biosynthetic pathways to create a truncated system, but one with the potential for novel compound formation. However, module boundaries were found to extend beyond catalytic centres, with intermodular linkers playing important roles in the continuation of catalysis in a collinear manner, and also seemingly important in the dynamics of the pathway, essential for delivery of chemical intermediates to enzymatic centres (Weissman, 2016). Therefore, the retention of catalytic integrity and enzyme functionality was a major drawback to early efforts to engineer PKS pathways, with only a small number of chimeric or module swapped systems resulting in functionality and the subsequent formation of potentially novel products (Poust et al., 2014). Much of the early work on creating chimeric systems through module alterations and swapping were carried out on historically well studied, canonical PKS systems, particularly the DEBS pathway, which produces the erythromycin aglycone. The fusion of two or three modules in an *in vivo* testing platform led to the formation of functional systems in approximately 50 % of cases, highlighting the importance of cognate docking domains, or suitable replacements thereof (Staunton, 1998). Examples of successful engineering methodologies employing module swaps or additions have also been carried out in the DEBS system. Replacement of the cognate DEBS loading module with the promiscuous and wide-specificity loading module from the avermectin biosynthetic pathway led to the production of derivatives of erythromycin with extended side chains at the position corresponding to that added by the initial loading module (Marsden et al., 1998). Initial experimentation using the altered loading module and the first two extension modules of the cognate DEBS system in a *S. coelicolor* heterologous host resulted in production of the expected triketide lactones. Extension of the methodology to the full DEBS system in the native erythromycin producer yielded bioactive variants of the naturally produced heptaketide erythromycin, and showed the efficacy of such a system. Module insertion has also been shown to be an effective strategy in limited scenarios. Insertion

of an extension module from the rapamycin producing system into the DEBS system, preceding the second cognate extension module, produced the expected, larger products both *in vitro* and *in vivo* (Rowe et al., 2001). However, natural erythromycin aglycone products were also observed in fermentation broths, produced by skipping of the newly spliced module, indicating the inherent problems of engineering systems on a modular scale.

A more conservative option for engineering PKS systems than whole module swaps and insertions, is the methodology of altering or swapping single domains within pathways, usually AT domains, due to their control of modular substrate usage (Klaus and Grininger, 2018; Musiol-Kroll and Wohlleben, 2018). The alteration of other domains, however, such as the reductive, tailoring domains, has also been achieved, and is generally observed to be more successful due to the involvement of minimalistic changes to both the enzymatic domain and the subsequently synthesised compound (Weissman and Leadlay, 2005; Baerga-Ortiz et al., 2006). The replacement of AT domains has been predominantly carried out by two major methods; deletion and complementation, or *cis*-encoded domain substitutions. Both methods were also initially carried out on the well characterised erythromycin-synthesising DEBS system. Inactivation of certain mmCoA accepting AT domains and complementation of the activity *in trans* with AT domains of differing specificity, resulted in the production of compounds with altered groups at specific positions, such as those resulting from the usage of mCoA or ethylmalonyl-CoA (Stassi et al., 1998). A more controlled methodology is the direct swapping of AT domains from one pathway to another, as typified by the creation of a chimeric PKS including replacing an AT domain within the DEBS system with an AT domain from the rapamycin pathway (Petkovic et al., 2003). Subsequently a number of 6-dEB analogues were prepared by multiple AT domain swaps with domains of different specificities, from different pathways, leading to the incorporation of extender units derived from varied CoA attached precursors (McDaniel et al., 1999). Despite the success of a number of AT domain substitution strategies, many more attempts have been subject to failure or incur high production costs, making engineering efforts unviable or untenable for the rational production of novel compounds. Most engineered pathways suffer from very poor production titres, likely due to the disruption of important protein-protein interactions and/or the inability of downstream domains/modules to accommodate the engineered chemical diversity (Wu et al., 2001). Domain alterations have also highlighted the importance of precursor availability, with multiple engineered pathways requiring extra modifications, such as the addition of upstream precursor biosynthetic genes, for proper, altered functionality. These requirements also indicate

the extended nature of some assembly line-like systems, whose functionality extends beyond the megasynthase and requires multiple other proteins for efficient functionality and compound production.

A final, relatively commonly employed methodology for the production of novel metabolites from altered assembly lines is the minimal perturbation approach of site-directed mutagenesis (Drufva et al., 2020). Sequence alignments and the solving of structures of AT domains led to the identification of multiple residues, important in substrate selection and exhibiting divergence in AT domains of differing acyl-CoA specificities. A conserved AT domain motif resides approximately 100 residues towards the C-terminal of the active site, and has been shown to contribute to substrate control in various instances. The four amino acid motif is commonly composed of Tyr, Ala, Ser, and His (YASH) in mmCoA specific systems, and is generally His, Ala, Phe, and His (HAFH) in mCoA accepting domains, with the size of the third residue in the motif thought to be inversely proportional to the size of the accepted and activated substrate (Del Vecchio et al., 2003). The nature of this motif has led to its common targeting as part of precision engineering studies, such as the YASH motif containing AT domains in modules 1, 4, and 6 of the DEBS system. In these cases, motif switching to the mCoA specific HAFH residues led to promiscuous substrate selection and the incorporation of both acetyl and propionyl extender units (Sundermann et al., 2013). However, despite site directed mutagenesis being the most minimalistic approach to the alteration and engineering of assembly lines, *in vitro* kinetic analysis has shown that these perturbations severely restrict acyltransferase activity, thereby potentially causing promiscuity through attenuation of native activity rather than a complete switch in enzymatic substrate preference (Hans et al., 2003; Dunn et al., 2013). Engineering of PKS pathways, either through coarse or more site-specific methods, requires further analytical and structural knowledge, as well as an expanded toolbox of techniques, for the ultimate creation of efficient chimeric pathways and the subsequent production of novel secondary metabolites.

3.1.3 Heterologous expression of secondary metabolite gene clusters

The emergence of multiple antibiotic resistant strains of bacteria, coupled with the lack of novel antibiotic discoveries and reduced effectiveness of previously relied upon drugs, has led to renewed interest in the unearthing of novel natural products with interesting bioactivities. The renewed vigour in the search for antimicrobials has been accelerated by the analysis of many newly sequenced genomes of historically prolific producers of antibiotics, for their potential

capacity to produce a varied selection of specialised metabolites, many of which are as yet uncharacterised or unknown (Bachmann et al., 2014). Further study of many potentially interesting natural product producing pathways, however, is often precluded by the bacterial strain in question; either due to its poor growth characteristics or genetic intractability. Therefore, use of a known faster-growing and genetically tractable heterologous host bacterial strain can be a successful strategy, and is frequently adopted as the method of choice for analysing unknown gene clusters (Figure 3.2). These heterologous expression strategies require genetic clones of the pathways of interest, with multiple effective methodologies for the creation of such clones, such as library generation techniques, and specific DNA capture via recombination, commonly utilised (Lin et al., 2020).
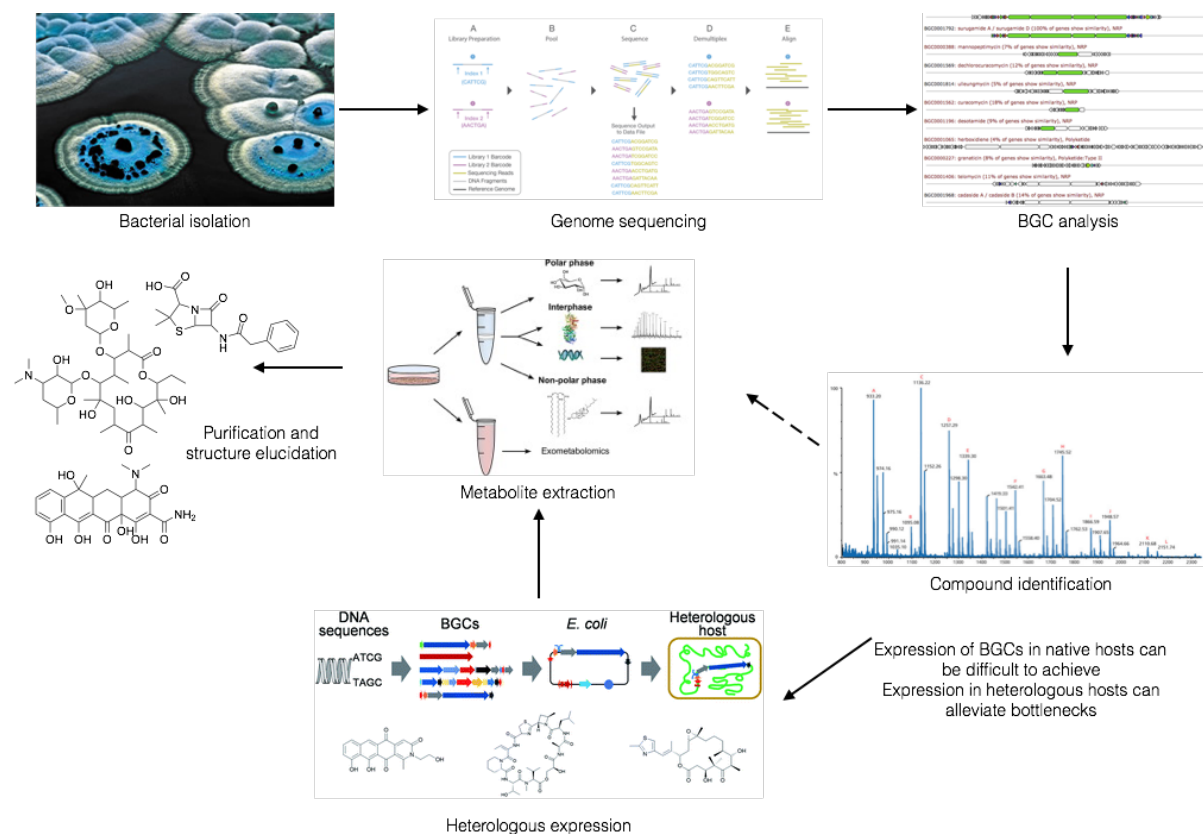
Figure 3.2: The general methodological pipeline utilised for the identification of novel natural products or specialised metabolites. Genome sequences can be obtained from bacteria from new niches or from public genome databases. After the analysis of the biosynthetic potential of the strain and identification of potential matching compounds, metabolite extraction and structural elucidation can yield potential novel bioactive compounds. Extraction of specialised metabolites can be difficult to achieve and heterologous expression of the relevant BGC in a disparate bacterial strain can alleviate bottlenecks.

Although many fast-growing bacterial strains are well studied and widely used as general hosts for various applications, the heterologous expression of natural product producing BGCs is predominantly carried out in genera such as *Streptomyces, Salinispora,* and *Saccharopolyspora,* bacteria generally closely related to the native BGC hosts (Wenzel and Müller, 2005). These strains are more suited to the use of BGC-encoded promoter systems, as well as being better equipped with required cofactors and precursor biosynthetic pathways. The creation of heterologous host strains, particularly in species of *Streptomyces*, as typified by the *S. coelicolor*, *S. albus*, and *S. venezuelae* heterologous hosts, usually combines selection of a theoretically ideal bacterial strain with an engineered and cleaner specialised metabolite profile, customarily carried out by the mutation of endogenous natural product producing BGCs (Baltz, 2010; Myronovskyi and Luzhetskyy, 2019). These engineered strains, more optimally engineered for metabolic flux towards incoming, heterologous BGCs, are then transformed with the requisite DNA and tested for the successful expression of the cluster, and production of the encoded specialised metabolite.

Numerous *Streptomyces* based heterologous hosts have been engineered for the expression of foreign BGCs and the subsequent formation of novel, or engineered variants of natural products. *S. coelicolor* is the model organism for the genus *Streptomyces*, with its genetic make-up, antibiotic production, and BGC content widely studied for numerous years, and increasing in pace with the sequencing of the whole genome of the strain in 2002 (Bentley et al., 2002). Multiple heterologous expression strains of *S. coelicolor* have subsequently been created, including the notable 'superhosts' *S. coelicolor* M1146 and *S. coelicolor* M1152 (Gomez-Escribano and Bibb, 2011). The M1146 strain contains complete pathway deletions of four biosynthetic gene clusters, including those responsible for the production of actinorhodin (*act*) and undecylprodigiosin (*red*), known for their distinctive pigmentations and colour profiles, as well as the *cpk* and *cda* clusters which, when functional, result in the formation of coelimycin and calcium-dependent antibiotic, respectively. The pathway deletions restrict the strain from production of usually high titre compounds, therefore allowing the flux of substrates, precursors, cofactors, and other required compounds to silent or non-cognate gene clusters, a methodology utilised in the creation of a variety of other heterologous hosts (for example those created in the *S. albus* J1074 and S4, *S. venezuelae*, and *S. lividans* backgrounds). The *S. coelicolor* M1152 strain, engineered from the M1146 strain, also contains mutations in *rpoB* and *rpsL*, a subunit of the RNA polymerase and a ribosomal protein,

respectively. These mutations have been shown to increase the production of antibiotics, thereby useful in the context of heterologous BGC expression, and also typify an alternate strategy in the creation of hosts through altered regulation of genes at both transcriptional and translational levels.

A wide variety of secondary metabolite BGCs have been successfully heterologously expressed in non-cognate host strains, including many metabolites whose production was not observed from the original host strain. Examples of natural products which have been produced from classic heterologous expression systems include epothilone, desotamide, and kanamycin, as well as actinorhodin and undecylprodigiosin (Nah et al., 2017). The BGCs responsible for the production of these compounds were isolated in a number of different methods (cosmid, BAC, PAC, TAR) and expressed in a variety of *Streptomyces* strains, often yielding the complete bioactive metabolite. The numerous examples highlight the functionality of heterologous expression as a method for accessing novel BGCs and compounds, however, heterologous expression strategies often require excess trial and error before a functional system can be found (Zhang et al., 2011). The unknown functional nature of the components of many BGCs, especially those with unknown metabolite products, coupled with the unpredictable nature of the behaviour of heterologous host strains, means selecting a suitable host for the expression of BGCs and the subsequent production of natural products is non-trivial and often a case of testing numerous available strains for pathway expression (Zhang et al., 2019). Therefore, the availability of a BGC clone and a suitable heterologous host strain are advantageous, but not a guarantee of successful compound production and novel structure elucidation. The production and availability of heterologous host strains for the expression of non-cognate BGCs, however, has been highly useful for the elucidation of novel natural products structures, and continues to be an important methodology for the discovery of bioactive compounds from a wide variety of cultured and uncultured microorganismal sources.

3.1.4 The utilisation of CRISPR/Cas9 in engineering specialised metabolite clusters

The discovery and development of CRISPR/Cas9 (Figure 3.3) as a tool for genome editing has revolutionised numerous fields, including the *in vivo* engineering of natural product biosynthetic gene clusters. Previous to the use of CRISPR/Cas9 the engineering or mutation of BGCs was carried out either on cloned clusters or BGCs held within the genome using Rec/ET recombineering or through the use of specialised vectors, such as suicide vectors for the deletion of often large stretches of genomic DNA (Miller and Mekalanos, 1988; Penfold and

Pemberton, 1992). CRISPR/Cas9 allows the targeting of specific areas of genomic DNA for the insertion, deletion, or site-directed mutagenesis of stretches of DNA, therefore useful in various BGC engineering projects, with wide-ranging goals or targets (Ebrahimi and Hashemi, 2020).
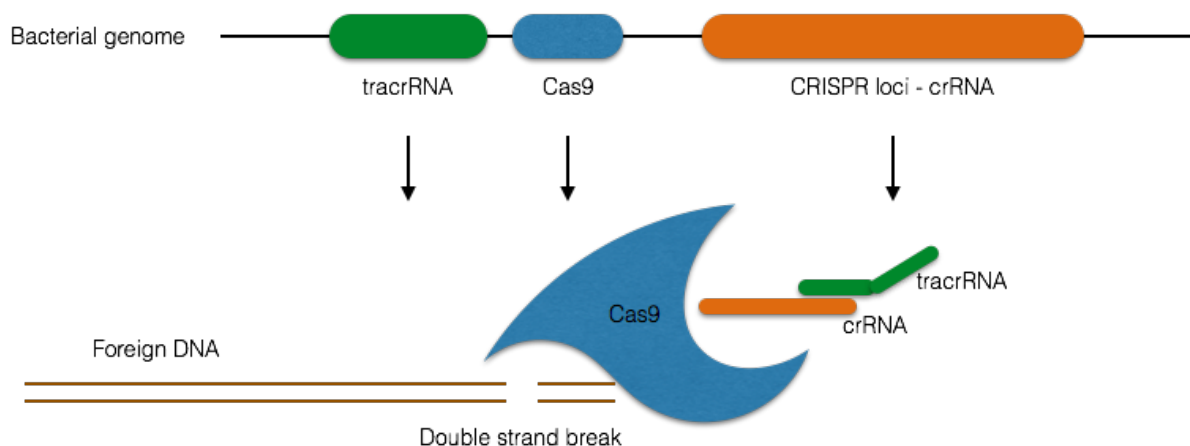
Figure 3.3: Simplified overview of the natural, bacterial CRISPR/Cas9 system used as a defence mechanism against foreign or invading organisms. Encountered viral DNA is stored in a CRISPR loci, subsequently used to make crRNA. The crRNA and tracrRNA guide the Cas9 cutting enzyme to the foreign DNA, resulting in double strand cleavage of the DNA.

A number of useful CRISPR/Cas9 based tools have been created for the manipulation and modulation of natural product BGCs, within common secondary metabolite producing strains, such as species of *Streptomyces*. These include systems for creating deletions or mutations, promoter insertions, or for cloning of whole clusters, such as the pCRISPR-Cas9 and pCRISPR-dCas9 systems, as well as the CATCH cloning system (Huang et al., 2015; Jiang et al., 2015, p.9; M.M. Zhang et al., 2017). A prominent example of a CRISPR/Cas9 system tailored for use in natural product producing *Streptomyces* strains is the pCRISPomyces system (Figure 3.4) (Cobb et al., 2015). The genome editing system is a single plasmid able to carry out multiplexed genome editing by the use of multiple sgRNAs and multiple editing templates, as well as a *Streptomyces* codon optimised variant of the *Streptococcus pyogenes cas9* gene. The plasmid, quickly assembled via Golden Gate assembly, mediating introduction of the sgRNA sequences prior to the introduction of the required editing templates, is transferred to *Streptomyces* for *in vivo* genome editing. Genome editing occurs at a frequency of 70 – 100 %, demonstrating a general upward trend in the efficacy of *in vivo* CRISPR/Cas9 systems for the editing of natural product BGCs. Whilst the efficiency and inherent usability of CRISPR/Cas9

systems is ever increasing, the process of editing BGCs within genomes can yet be problematic. For example, within the pCRISPomyces system, the process of curing the temperature sensitive replicon is time consuming and the replicon does not cure easily in some *Streptomyces* species, often leading to problems within the workflow. Overall, however, *In vivo* BGC editing methodologies are generally widely used and allow the manipulation of clusters for their activation or alteration of produced metabolites, making the techniques extremely valuable in the pursuit of novel bioactive compounds.
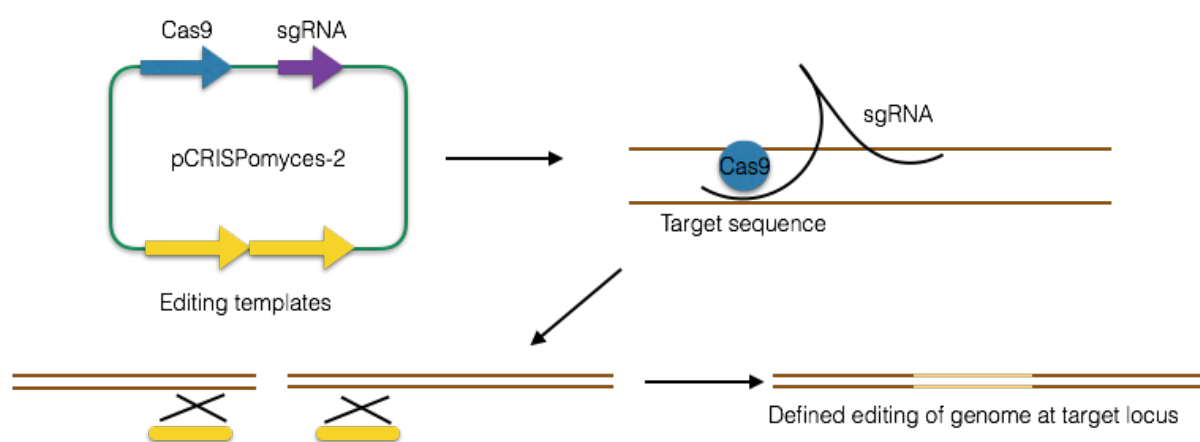
Figure 3.4: Overview of the pCRISPomyces system used to edit *Streptomyces* genomes *in vivo*. The replicative plasmid encodes a codon optimised *sSpcas9* gene, as well as the sgRNA and 1 kb editing templates used for inserting the required genetic alteration. The system can also be multiplexed for the alteration of multiple genetic loci in combination.

3.2 Aims and objectives

The aims of this project were to adapt, develop, and test a CRIPSR/Cas9 based system for the directed engineering of cloned specialised metabolite gene clusters. This involved a multi-faceted approach with the ultimate goal of simple, efficient, and fast mutation of large biosynthetic gene clusters, with high levels of specificity for only the required alteration. Biosynthetic gene clusters are usually very large sections of genomic DNA, and contain numerous multi-gene operons, encoding the proteins required for production of specific chemical compounds with useful activities in varying contexts. Due to the large sizes of BGCs, and the fact that many remain silent or cryptic under laboratory conditions, cloning and heterologous expression of BGCs has become an important and, often, successful strategy for the identification of novel metabolites. However, mutation and DNA alteration even in gene

clusters cloned into extrachromosomal sources of DNA, such as plasmids, cosmids, or BACs, can yet be challenging and an extremely long procedure. Whilst, CRISPR/Cas9 approaches are increasingly popular, and a CRISPR/Cas9 system for mutation of *Streptomyces* genomes has been developed, use of a CRISPR/Cas9 system on cloned BGCs in *E. coli* could expedite the process and potentially allow heterologous expression of mutated BGCs in strains that are more suited for compound production or more genetically tractable. Therefore, a plasmid based CRISPR/Cas9 system, predominantly utilised for chromosomal modifications in *E. coli*, was selected and adapted for targeting and modification of extrachromosomal DNA, held within the strain of *E. coli,* specifically the antimycin BGC held on a cosmid construct. This project, therefore, aimed to allow the targeting of cloned BGCs with CRISPR/Cas9, rather than chromosomal DNA, by specific sgRNA design, and subsequent transformation and PCR-based methods for mutant selection and isolation. In this scenario, the sole AT domain from the PKS module of the antimycin BGC was selected for specific mutation, in order to test the substrate specificity of the naturally promiscuous domain. The mutated, cloned antimycin BGC would then be tested for the production of antimycin-type depsipeptides by expression of the construct within *Streptomyces albus* S4, allowing extraction and visualisation of the antimycins produced.

3.3 Results

3.3.1 Antimycin production from *S. albus* Δ5 Cos213 strain

Members of the lab previously generated a chromatogram-simplified variant of *S. albus* S4 in which the entire antimycin BGC was deleted and four additional BGCs were mutagenized (candicidin, fredericamycin, surugamide, and albaflavenone) (Fazal et al., 2020). In theory, use of the Δ5 strain and a cosmid clone of the antimycin BGC would expedite the creation of mutant strains. Therefore, to initially establish the viability of this approach, Cos213 was moved into the Δ5 strain to determine if it could restore antimycin production. The strain was cultured under known antimycin production conditions, liquid SFM medium at 30 °C for seven days, after which antimycins and other produced secondary metabolites were extracted from the clarified supernatant by solid phase extraction (SPE) using small volume, reverse phase C-18 columns. Analysis of the extracts by LC-HRMS (Figure 3.5) showed the production of the regularly observed, wildtype suite of antimcyins with the molecular formulae: $C_{28}H_{40}N_2O_9$ (antimycin A1, 548.27 Da), $C_{27}H_{38}N_2O_9$ (antimycin A2, 534.26 Da), $C_{26}H_{36}N_2O_9$ (antimycin

A3, 520.24 Da), $C_{25}H_{34}N_2O_9$ (antimycin A4, 506.22 Da). These peaks, observed in high intensity, were indicative of the production of antimycins and presented with the characteristic double peak set. These 'early' and 'late' peaks were constantly present upon LCMS analysis of extracted antimycins, with the perceived reasoning being the production and presence of isomers of the antimycin compounds, with the same molecular formula and mass. The isomers potentially contained altered chemical groups at the R positions branching off the main nine-membered ring, with the total mass of the compound unchanged. The isomeric compounds could likely migrate differently on reverse phase chromatography columns, leading to the detection of two separate peaks, however this hypothesis was not confirmed. These results, however, showed that antimycins were produced by the Δ5 Cos213 strain under the utilised production conditions.
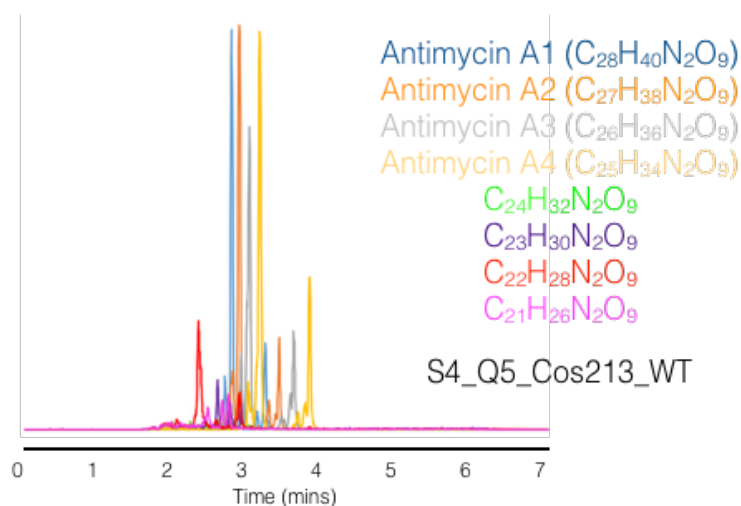
Figure 3.5: LC-HRESIMS analysis of extracts prepared from supernatants of the S4 Δ5 Cosmid 213 strain. Extracted ion chromatograms are shown for the four most commonly observed antimycins (antimycins A1 - A4) as well as four putative smaller antimycin species. Exact neutral m/z values used to determine EICs were: 548.2734, 534.2577, 520.2421, 506.2264, 492.2108, 478.1951, 464.1795, 450.1638. The y axis set to 6 x10^5.

3.3.2 Creation of strains with CRISPR/Cas9 mediated AT domain mutations

As antimycins were produced by the Δ5 Cos213 strain, subsequently required modifications to the antimycin BGC could be introduced on the cosmid clone, alleviating the reliance on homologous recombination strategies in *Streptomyces*. The overall workflow was to initially create all constructs required for the CRISPR/Cas9 mediated mutation of the antimycin BGC, followed by mutation creation in *E. coli* and confirmation by various techniques. Upon

successful mutation confirmation, the mutant constructs would be transferred into the *Streptomyces albus* S4 Δ5 strain for testing of antimycin production in an *in vivo* context.

The no-SCAR (Scarless Cas9 Assisted Recombineering) protocol (Reisch and Prather, 2015) was originally created as a two-plasmid method for engineering the *E. coli* genomic DNA in a single step. Previous methods for the generation of genome engineered variants of *E. coli* generally utilised antibiotic resistance cassettes to make the mutation(s), followed by excision of the cassette. This methodology, although relatively successful, required a two-step procedure and could result in scars or DNA remnants being present within the genome after excision of the resistance cassette. The no-SCAR protocol, however, uses CRISPR/Cas9 with a specifically designed sgRNA, targeting the region of genomic DNA of interest. After double strand break in *E. coli*, a linear repair template is introduced into the cells and stabilised by the use and expression of the λ-Red system. Homologous recombination between the genomic DNA containing the double strand break and the linear repair template creates a continuous, functional chromosome with no scars and containing the mutation required and encoded within the repair template (Figure 3.6). The first plasmid of the two-plasmid no-SCAR system contains the *cas9* gene under the control of the P_{TET} promoter, inducible by the addition of tetracycline or anhydrotetracyline to growing cells. The second plasmid contains the sgRNA, required for specific CRISPR/Cas9 functionality, also under the control of the P_{TET} promoter, as well as the three genes comprising the λ-Red system under the control of an arabinose inducible promoter. Development of CRISPR/Cas9 systems to date have focussed exclusively on editing genomic DNA, therefore alterations were made to the overall workflow in order to allow the targeted mutation of extrachromosomal constructs containing specialised metabolite BGCs.

Figure 3.6: The CRISPR/Cas9 system used to create point mutations in the antimycin BGC encoding cosmid. A single plasmid encodes the sgRNA construct, as well as the λ-Red system for stabilising incoming linear DNA. A second plasmid encodes the *cas9* gene under the control of the tetracycline-inducible promoter. Double strand break, followed by repair by homologous recombination yields cosmids containing the required site-specific alteration. The uppercase lettering used within DNA sequences indicates bases that were altered after repair of the double strand break. The majority of these changes (19/21) did not alter the amino acid encoded but allowed subsequent mutation screening by use of mismatched primers in a PCR.

Initially, a protospacer (GAGCACGTGCGAGTGGGCGG) targeting the cosmid region of interest, the AntD PKS AT domain motif III region, was cloned into the pKD plasmid backbone. The mutations required were all in this region and encoded changes in the amino acids comprising motif III of the AT domain, from the wildtype AAAH amino acid sequence to the three mutants with sequences: AAPH, HAFH, and YASH. The protospacer preceded a protospacer adjacent motif (PAM) of NGG on one of the strands of DNA and the sequence of the protospacer was checked to ensure that the *E. coli* genomic DNA would not be simultaneously targeted for double strand break. The targeting of the cosmid construct only was relatively easy to achieve due to the length of the protospacer, but primarily due to the

extremely high GC content of *Streptomyces*, and subsequently the antimycin BGC, compared to the *E. coli* genome. The specific protospacer was cloned into the pKD-sgRNA plasmid by circular polymerase extension cloning (CPEC), involving amplification of the plasmid in two parts, one of 3 kb and one fragment of 4 kb in length, followed by DpnI digest and gel purification and a self-priming PCR between the two linear DNA fragments to achieve circular plasmid formation. Once constructed, the protospacer containing vector and the pCas9-Cr4 vector was transformed into the *E. coli* strain, BW25113, already transformed with Cos213. The host strain was then cultured at 30 °C, due to the pKD-sgRNA plasmid containing a temperature sensitive origin of replication and to ensure no loss of plasmid. The *E. coli* BW25113 strain then contained the requisite DNA components for CRISPR/Cas9 targeting of Cos213 and mutant formation.

In order to create the mutant Cos213 constructs, the *E. coli* host strain with the three extrachromosomal constructs was grown to early log phase, before induction of the λ-Red system genes with arabinose and transformation with one of the linear double-stranded DNA repair templates containing the requisite mutation. Transformed cells were plated onto selective agar plates with anhydrotetracycline to induce Cas9 formation and subsequent double-strand break, specifically of the Cos213 DNA. Homologous recombination with the stabilised repair template yielded the mutant cosmid. The double-strand break carried out by the Cas9 enzyme, in theory provides counterselection by causing a break in the cosmid DNA and therefore not allowing growth on cosmid selective (apramycin containing) medium. However, constructs would still require checking by PCR in order to unambiguously confirm the genetic content and corroborate mutant formation. Due to the large genetic content of the host *E. coli* cells (genome, cosmid, two plasmids) and the fact that mutants were created in extrachromosomal cosmid constructs, DNA was purified from the cells by Miniprep, re-transformation, and growth on medium selective only for cosmids. Extrachromosomal DNA was then again isolated from numerous colonies and checked by PCR using mismatch primers, only capable of priming to mutated cosmids. This relied on base changes in the repair template, including those encoding the required amino acid mutation as well as single base changes in the template that maintained the amino acid sequence of the encoded protein but ensured the GC content of the repair template was manageable for *in vitro* techniques. These confirmation PCR assays showed a success rate for the CRISPR/Cas9 mediated mutation of ~ 10% (Figure 3.7) and resulted in recovery of cosmids containing the three requisite AT domain mutations within

Cos213 holding the antimycin BGC. After unambiguous confirmation of the sequence of the mutant constructs by DNA sequencing, the cosmids; Cos213, Cos213-AAPH, Cos213-HAFH, Cos213 YASH, were conjugated into the *Streptomyces albus* S4 Δ5 host strain. Genomic integration of the cosmids was confirmed by PCR and the strains were stored for subsequent expression and antimycin production experiments.
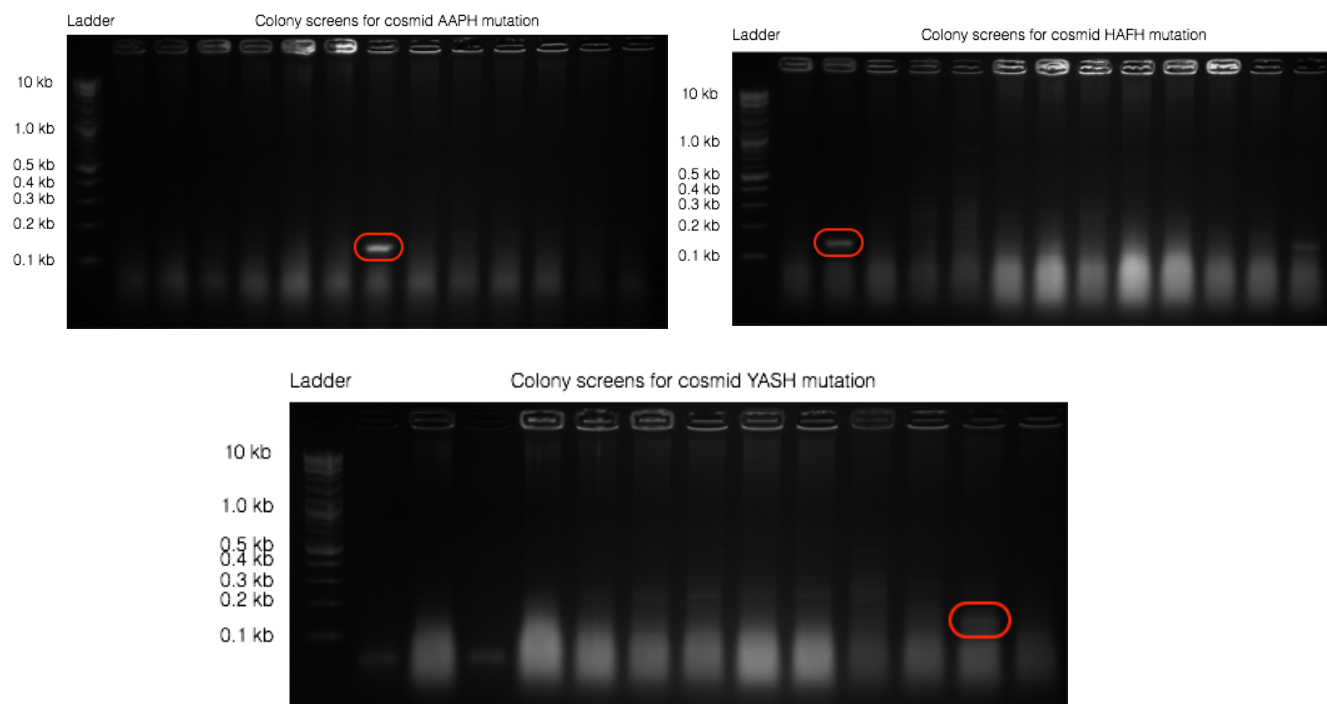
Figure 3.7: Screening gels indicating successful CRISPR/Cas9 mediated mutation of antimycin encoding cosmid constructs, with an expected product size of 143 bp. Cosmids were screened using mismatch primers, priming only to mutant constructs, with an observed success percentage of ~ 10 %.

3.3.3 Production of antimycins from AT-domain mutant strains

The wildtype AT domain present in the sole PKS module, AntD, from the antimycin pathway, is capable of selecting and catalysing the acyl-transfer of various malonyl-coenzyme A (mCoA) derivatives. This occurs via transfer of the CoA appended acyl groups onto the AT domain active site Ser residue, followed by transfer onto the conformationally dynamic thiol-terminating 4'-phosphopantetheinyl group, covalently linked to the ACP of the PKS module. The promiscuity of the AntD AT domain, which contains the residues AAAH at motif III, results in the formation of a battery of varied antimycin compounds, with differing chemical groups at the C7 position of the nine-membered ring forming the core of the antimycin-type compounds (Figure 3.8). These groups range commonly from two to six carbon units in a linear

alkyl chain. Alteration of one of the central motifs responsible for substrate selection was thought to potentially result in specific selection of a single mCoA derivative and a more substrate selective AT domain. This was due to the observed specificities of the cognate AT domains from which the three selected mutant motifs were obtained from; specifically, the HAFH and YASH motifs are commonly observed in several bacterial PKS AT domains and contribute to the selection of mCoA and mmCoA, respectively. The AAPH motif was derived from the sole AT domain involved in the biosynthesis of neoantimycin, an antimycin family compound formed from a central 15-membered ring (Skyrud et al., 2018). The PKS module, NatC, from the neoantimycin-synthesising system selects mCoA from the cellular pool and incorporates the moiety into the growing intermediate, with the module-encoded methyltransferase functioning twice post-selection of the mCoA precursor. The difference between the motif III sequences in $AntD^{AT}$ and $NatC^{AT}$ is a single amino acid change from Ala to Pro, therefore substitution of this residue within $AntD^{AT}$ could have some effect on substrate utilisation by the PKS module. Whilst small numbers of point mutations are not always sufficient for affecting or altering enzyme substrate specificity, single or multiple point mutations in judiciously selected protein locations have been shown to be effective in alterations of substrate specificity and kinetic utilisation for a wide variety of enzymatic systems. Therefore, mutation of motif III in the AntD AT domain was thought to be capable of resulting in changes in the substrate utilisation and subsequent observation of the antimycin product profile.

Cosmid integrants of the Δ5 strain were verified by PCR and subsequently tested for antimycin production under the previously utilised culture conditions. The strains were cultivated in liquid SFM at 30 degrees for seven days and the culture medium was subsequently extracted by solid phase extraction and analysed by LC-HRMS. Upon culture and extraction of antimycins, and other sufficiently hydrophobic compounds, from the culture broths of both the wildtype and mutant strains, methanolic extracts were subjected to LC-HRMS analysis, and any antimycin-type compounds were identified. The strain containing the wildtype copy of the antimycin BGC produced the expected and previously observed set of antimycins; namely antimycins A1, A2, A3, and A4, ranging from compounds containing 25 carbons to 28 carbon atoms in the complete molecule (Figure 3.8). No compounds corresponding to antimycins were observed at smaller m/z values, indicating that C_{25} antimycins are the smallest naturally produced under the conditions utilised. The most commonly observed chemical group at the

C7 position of the dilactone ring in the majority of the antimycins is a four-carbon alkyl chain, introduced by the selection and utilisation of butyl-malonyl CoA by AntD^{AT}.
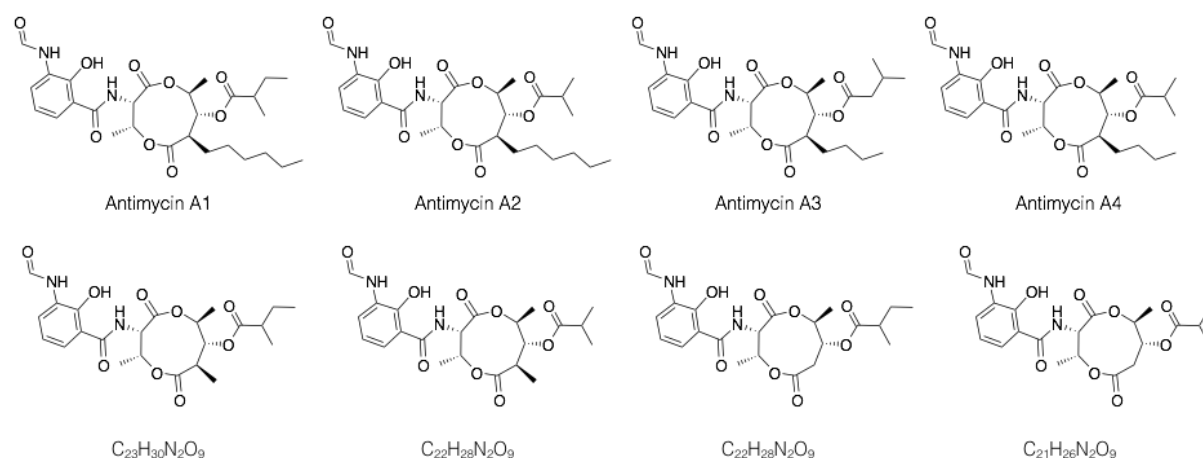
Figure 3.8: The structures of the wildtype suite of antimycin (A1 – A4) and the structures of putative smaller antimycin-type compounds that could be made from the utilisation of mCoA or mmCoA by mutant AT domains.

As the AT domain-mediated selection of mCoA and mmCoA, and the subsequent incorporation of these moieties at the C7 position of the antimycin dilactone ring, is very limited and seemingly unobserved among the naturally produced antimycin analogues, the alteration of AT domain selectivity motifs to those specific for the smallest acyl-CoA derivatives could be of interest in the area of pathway engineering and in the creation of novel antimycin orthologues. LC-HRMS analysis of extracts generated from the YASH and HAFH mutant strains displayed complete occlusion of the production of antimycins of any variety or size and seemed to wholly abolish the activity of the assembly line in the creation of antimycin-type secondary metabolites (Figure 3.9). Motif III has been shown to be important, though not all encompassing, in the selection of acyl-CoA derivatives for use by domains within PKS modules, and mutation of the residues comprising the motif to the commonly observed YASH and HAFH amino acid stretches has often been successful in the alteration of substrate selectivity and specificity. On the other hand, enzyme mutations for the modification of substrate utilisation rarely complies to universal rules, and functionality in the context of one system does not necessitate success in other, varied contexts. The failure of YASH and HAFH motif III mutations to effect changes in the substrate utilisation of the AntD PKS could be due to a variety of reasons, with the structural complexity of the enzyme, and the changes imposed

by the mutant residues, likely a key factor in the lack of AT domain-mediated activation of acyl-CoAs for subsequent antimycin production.
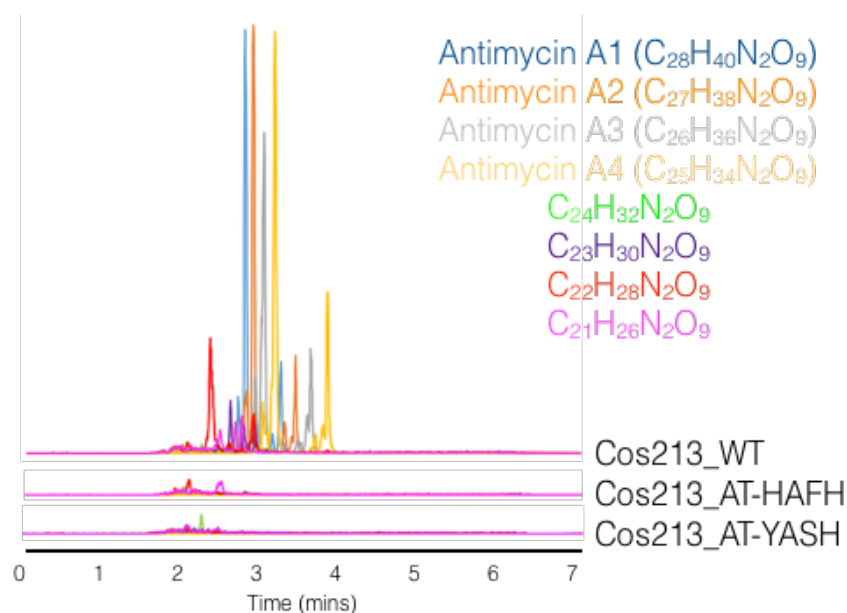
Figure 3.9: LC-HRESIMS analysis of extracts prepared from supernatants of the designated antimycin production strains. Extracted ion chromatograms are shown for the four most commonly observed antimycins (antimycins A1 - A4) as well as four putative smaller antimycin species. Exact neutral m/z values used to determine EICs were: 548.2734, 534.2577, 520.2421, 506.2264, 492.2108, 478.1951, 464.1795, 450.1638. The y axis was set to 6 x10^5 for each of the chromatograms.

The AAPH motif mutant, containing only a single point mutation from Ala to Pro at the third position of the amino acid stretch, was the most conservative motif mutant created. Culturing of the AAPH motif mutant strain and extraction of specialised metabolites from the culture medium yielded a complex mixture of antimycins, with a significantly deviated product profile when compared to the strain containing the wildtype antimycin BGC (Figure 3.10). The overall production of total antimycins was much lower than observed in the wildtype context, with much lower intensities seen in the mass spectra for extractions from the mutant strain. Strains expressing the wildtype antimycin BGC produced antimycins ranging from 25 carbons in the molecule, up to 28 carbons, with the 25 and 26 carbon-containing antimycins (antimycins A3 and A4) the most commonly produced. In the AAPH motif mutant strains, however, peaks corresponding to the 25 and 26 carbon antimycins were less abundant and, surprisingly, the larger antimycins were produced to a similar extent, or exceeding that produced by the wildtype

strain. Peaks were also observed in the mass spectral analysis which potentially corresponded to smaller antimycin-type compounds, such as those with the chemical formulae: $C_{21}H_{26}N_2O_9$, $C_{22}H_{28}N_2O_9$, and $C_{23}H_{30}N_2O_9$. The majority of these, however, did not resemble antimycins when MSMS fragmentation patterns were analysed, but the m/z of 464.1795 ([M+H]$^+$ of $C_{22}H_{28}N_2O_9$ antimycin) in the extract from the AAPH mutant strain was indicative of the production of an antimycin-type depsipeptide (containing MSMS peaks at *m/z* 136.0399 and 247.0719, Appendix 2). The observation of a C22 antimycin was interesting as a true antimycin of this size has not been previously observed, and this mass suggested the potential usage of a small acyl-CoA derivative, such as mCoA or mmCoA, in the final compound. However, as chemical groups of varying sizes and masses can be appended to both the C7 and C8 positions of the antimycin core ring, unambiguous confirmation of the nature of the group at the C7 position, and the composition of the acyl-CoA utilised by the modified AT domain, was not obtained. The standalone AT domain, AntB, is capable of catalysing the attachment of various chemical groups to the alcohol at the C8 position, resulting in the formation of a functionalised ester. Deletion of the *antB* gene, responsible for encoding the standalone AT domain, would potentially reduce the complexity at the C8 position in a predictable manner, leading to confirmation of the chemical side chain present at the C7 position, and the back-calculation of the acyl-CoA derivative utilised by the mutant AT domain of the PKS, AntD.
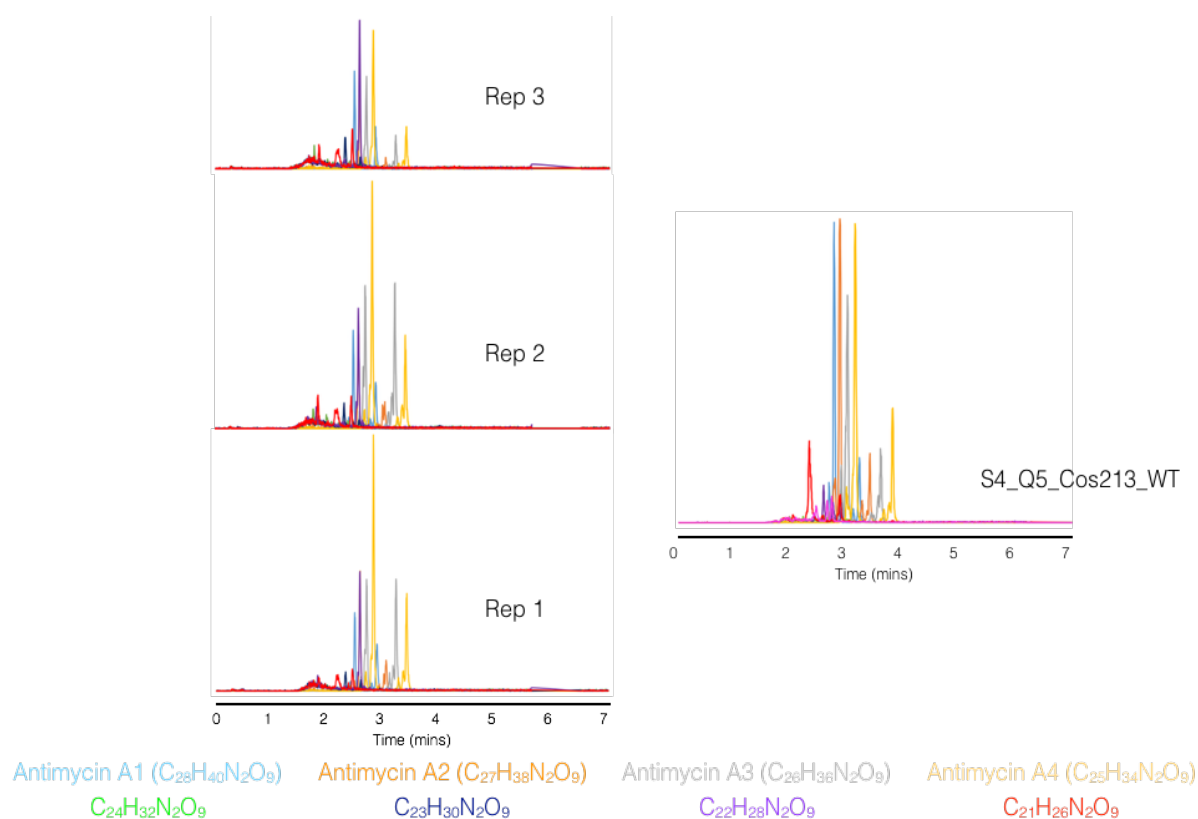
Figure 3.10: LC-HRESIMS analysis of extracts prepared from supernatants of three replicate growths of the AAPH AT domain motif mutant antimycin production strain. Extracted ion chromatograms are shown for the four most commonly observed antimycins (antimycins A1 - A4) as well as four putative smaller antimycin species (designated by the coloured key below the chromatograms). Analysis of extracts prepared from the WT strain are shown for comparison (right). Exact neutral m/z values used to determine EICs were: 548.2734, 534.2577, 520.2421, 506.2264, 492.2108, 478.1951, 464.1795, 450.1638. The y axis was set to 1.6×10^5 for each of the replicate chromatograms and at 6×10^5 for the WT chromatogram.

3.3.4 Cosmid deletion of *antB* and creation of mutant strains

AntB is a standalone acyltransferase responsible for installing the acyl group at the C8 position of the antimycin dilactone ring scaffold, via a post-NRPS tailoring, esterification reaction. The most commonly added chemical groups are four-carbon butyl esters or five-carbon, branched isopentanyl esters, however, chemical groups as small as two carbons and as large as seven carbon scaffolds have been observed at the C8 position. Whilst AntB is capable of installing a large variety of chemical groups at the corresponding position of the central antimycin nine-membered ring, and the variation observed at this position is likely helpful to the producing strain in natural environments as it could improve binding to its cellular target or reduce the

overall rate of resistance mutations arising, the variation at the C8 position leads to the production of multiple antimycin-type compounds, and makes chromatographic separation and identification more difficult. Therefore, deletion of the gene encoding production of the discrete acyltransferase would reduce the number of potential sources of variation observed within the antimycins, leaving the acyl group selected and activated by the AntD AT domain as the sole source of variety (Figure 3.11). The *antB* gene was deleted in the wildtype Cos213 and Cos213-AAPH via the ReDirect recombineering protocol (Gust et al., 2003). The mutated cosmids, with in-frame deletions, were conjugated into the Δ5 strain and analysed for antimycin production.
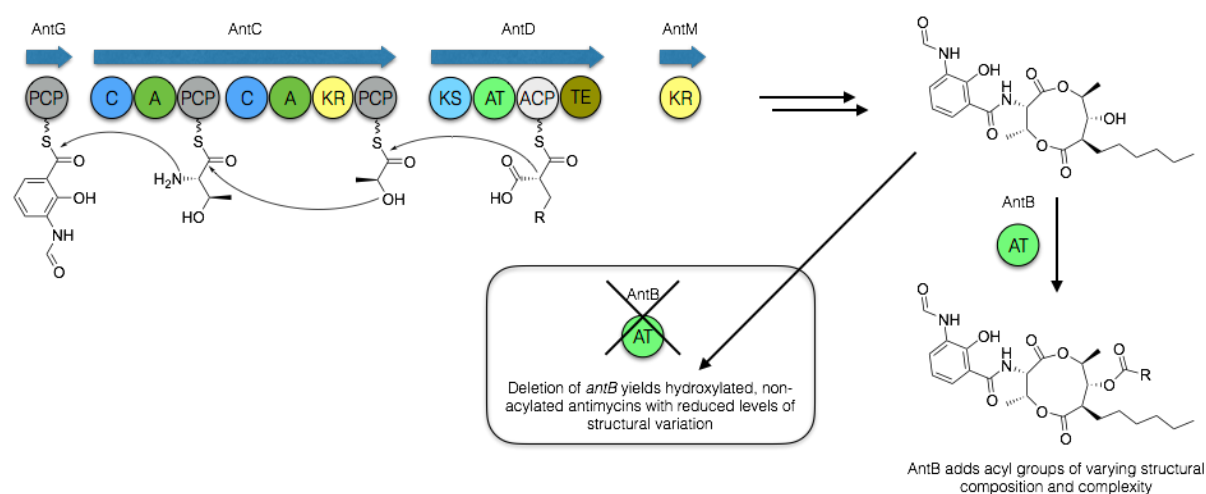
Figure 3.11: Overview of the antimycin biosynthetic pathway and the effect of the deletion of *antB*. Antimycins contain two regions of chemical variation, at the C7 and C8 positions. Deletion of AntB removes the structural complexity at the C8 position, reducing the total number of antimycins produced and reducing the chromatographic complexity.

3.3.5 Production of antimycins from *antB* deletion mutant strains

The *antB* deletion mutant *Streptomyces* strains were cultured under the same conditions used previously for the production of antimycins, and metabolites were extracted using the same SPE-based methodologies. The standalone acyltransferase, AntB, and its effects on the antimycins produced by the system have been previously characterised, generally resulting in de-esterified and hydroxylated products. The potential expected products from the Cos213Δ*antB* and Cos213ΔantB-AAPH strains are shown in Figure 3.12. Removal of the AntB installed acyl moiety at the C8 position of the dilactone scaffold generally results in the reduction in mass of antimycins produced corresponding to the formula of C_4H_7O or C_5H_9O.

$C_{23}H_{32}N_2O_8$ $C_{21}H_{28}N_2O_8$ $C_{18}H_{22}N_2O_8$ $C_{17}H_{20}N_2O_8$
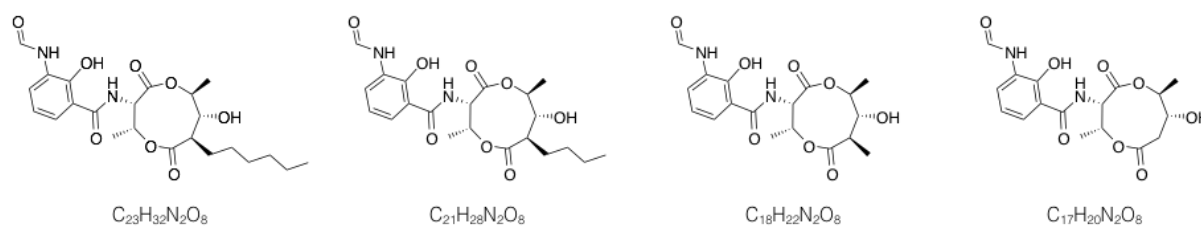
Figure 3.12: Structures of the expected antimycin products from the biosynthetic system containing a deletion in *antB* and therefore, no variant group on the external hydroxyl group of the antimycin core scaffold.

The deletion of *antB* in the wildtype background led to the expected production of a suite of antimycins with reduced masses compared to that produced by the strain containing the complete pathway (Figure 3.13). The masses observed corresponded to the singly charged $[M+H]^+$ of compounds with the formulae: $C_{23}H_{32}N_2O_8$ and $C_{21}H_{28}N_2O_8$, in turn representing commonly observed antimycins with four or six carbon alkyl chains at C7, and a hydroxy at C8 due to the lack of AntB producing the esterified product. These antimycin products, however, were seemingly not uniformly represented across different extracts prepared from bacteria cultured at the same time and in the same way, with the variability observed ascribed to originate from differences in the precursor pool and the substrates provided by the pathway crotonoyl-CoA reductase to the AntD AT domain, leading to alkyl chains of varying sizes present at the C7 position of the antimycin dilactone ring (six carbon alkyl chain for C23, and four carbon alkyl chain for the C21 antimycin). The overall observed trend showed production was highest for the C7 hexyl-containing compound, although overall intensities of spectral peaks was generally lower, likely due to a reduction in efficiency of the producing system *in vivo* from the absence of a biosynthetic enzyme. Production was lower, yet still detectable for the butyl-containing antimycin product ($C_{21}H_{28}N_2O_8$) at a later elution time. No peaks were observed for any smaller antimycin-type compounds that may have been biosynthesised from utilisation of small acyl-CoAs (mCoA, mmCoA) by the AntD AT domain. Overall, production of antimycins from these strains correlated with the expected production of deacylated compounds, with the hexyl and butyl groups at the C7 position of the antimycin core scaffold originating from acceptance and utilisation of varied acyl-CoA substrates in a manner analogous to that of completely wildtype systems.
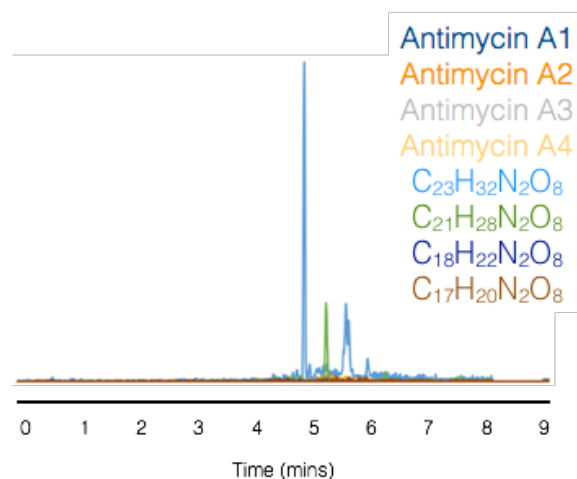
Figure 3.13: LC-HRESIMS analysis of extracts prepared from supernatants of the strain lacking *antB* and containing the wildtype motif in the AT domain of AntD. Extracted ion chromatograms are shown for the four most commonly observed antimycins (antimycins A1 - A4) as well as four putative smaller antimycin species, all containing a hydroxyl group at the C8 position of the antimycin core scaffold (designated by the coloured key). Exact neutral m/z values used to determine EICs were: 548.2734, 534.2577, 520.2421, 506.2264, 464.2159, 436.1846, 394.1376, 380.1220. The y axis was set to 7.0 x10^4 for each of the chromatograms.

The production of antimycins from the *antB* deletion and AT domain motif mutant strain showed a similar array of antimycins being produced and extracted, as well as a general reduction in the intensities of the mass spectral peaks observed, again showing a reduction in efficiency of the complete assembly line in the production of natural products (Figure 3.14). The composition of the antimycin product profile was also varied, when compared to those produced from the wildtype AT domain. Peaks were observed for one of the larger antimycins produced with a C8 hydroxyl group and C7 hexyl chain ($C_{23}H_{32}N_2O_8$). Peaks were observed at multiple elution times within the EIC for this species, potentially suggesting unusual interactions or behaviour on the chromatography column, although all detected peaks contained the same mass spectral signature. No peak was observed for the C7 butyl containing antimycin, however, suggesting that this antimycin-type compound is not produced under these conditions. Interestingly, the motif mutant and *antB* deletion strain also seemed to produce antimycins corresponding to the formula $C_{18}H_{22}N_2O_8$ (394.1376 Da), as shown by LC-HRMS analysis of extracts produced from the strain. The peak corresponding to a potential C18 antimycin, likely containing a methyl group at the C7 position of the dilactone ring, was consistently observed across all extracts produced from mutant strains. MSMS analysis, however, produced only low intensity fragment ions so could not confirm the nature of the

product. Compound breakdown during initial MS analysis did result in several peaks at m/z 136.0399, 164.0348, 247.0719 which are consistent with fragments produced from breakdown of the dilactone ring, or commonly, the 3-formamidosalicylate starter unit (Appendix 2). Although further identification of compounds would be required to unambiguously confirm their identities, these results show that mutation of motif III of PKS AT domains, whilst resulting in loss of overall efficiency of the assembly line, could result in the production of novel compounds, with potentially interesting biological activities.
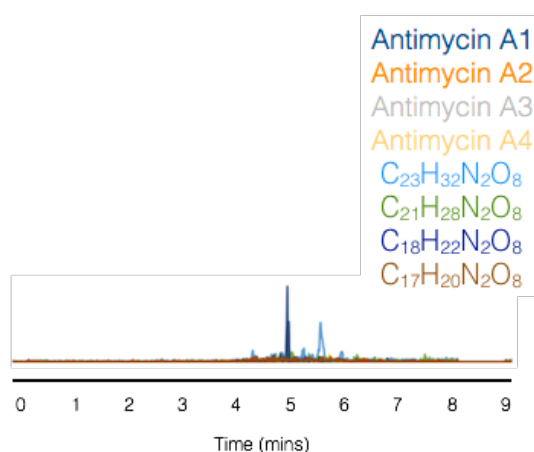
Figure 3.14: LC-HRESIMS analysis of extracts prepared from supernatants of the strain lacking *antB* and containing the variant AAPH motif in the AT domain of AntD. Extracted ion chromatograms are shown for the four most commonly observed antimycins (antimycins A1 - A4) as well as four putative smaller antimycin species, all containing a hydroxyl group at the C8 position of the antimycin core scaffold (designated by the coloured key). Exact neutral m/z values used to determine EICs were: 548.2734, 534.2577, 520.2421, 506.2264, 464.2159, 436.1846, 394.1376, 380.1220. The y axis was set to 7.0×10^4 for each of the chromatograms.

3.3.6 Structural analysis of the AntD AT domain binding pocket

Rational engineering of the enzymatic activities of proteins has extremely varied success rates in obtaining the desired outcomes, usually of altered substrate specificity or improved reaction kinetics. Generally, however, more specific information about the protein or system itself correlates greatly with improved rates of success, therefore obtaining structural information about enzymes at molecular levels of details allows rationalisation of the important motifs and interactions modulating protein functionality, and allows logical and reasoned explanations as to which residues, motifs, or secondary structural regions should be altered for maximum potential effect and benefit. The molecular basis of the promiscuity of the AntD AT domain

has been primarily ascribed to the nature of the binding pocket and the residues, such as the motif III amino acids, which make up the active site, allowing the binding and activation of various of mCoA derivatives of varying sizes. Whilst this work was in progress, a crystal structure of the AntD AT domain was solved (termed SpnD-AT, as the protein was isolated from the splenocin system which is capable of accepting benzylmalonyl-CoA but is otherwise identical to the antimycin system) and confirmed that the binding site residues are important in substrate selection, with the motif III residues (AAAH) playing key roles in substrate acceptance, or exclusion in the case of incorrect binding partners (Figure 3.15) (Li et al., 2018).
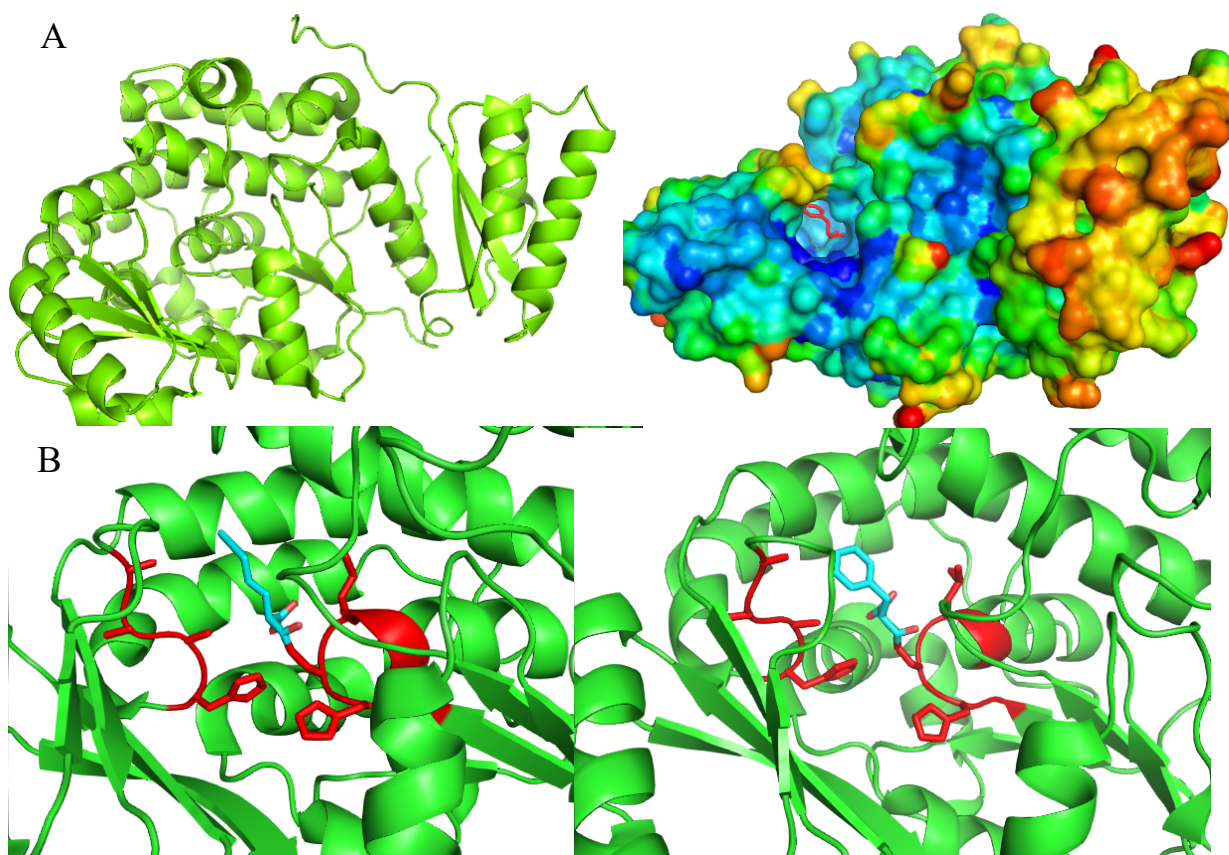
Figure 3.15: A) Overall structures of the monomer of the SpnD-AT domain (PDB ID: 5YDA). A benzyl-malonyl SNAC is bound in the active site of the surface structure. The AT domain resembles other structurally characterised PKS AT domains. B) Zoomed views of the SpnD-AT domain active site with bound pentynyl-malonyl SNAC (left) or benzyl-malonyl SNAC (right). The substrates are covalently bound to the active site Ser residue and held against the AAAH, loop motif, comprising the back of the binding site. Red colouring denotes the GHSQG active site motif, containing the Ser residue to which the selected acyl-CoA is bound, and the AAAH motif III region.

Analysis of the crystal structure of the AT domain showed that the enzyme contained a relatively wide binding cleft, with 14 amino acids forming the sides and the rear of the active site itself. The motif III residues form one of the sides of the pocket at a flexible hinge region, allowing access by substrates at the front of the pocket. The AAAH motif is also able to form repulsive interactions with incorrect substrates or substituents in incorrect configurations, causing steric clashes and occlusion of substrate binding. The residues forming the active site cleft (Figure 3.15) are predominantly small, hydrophobic amino acids, which do not impinge on the volume or space available, in which the substrate could potentially bind. The major residues involved in affecting substrate binding and subsequent catalysis of acyl transfer are the aforementioned AAAH motif, and the amino acids situated directly across from this motif, forming the opposite side of the active site pocket. Three Gly residues (^{91}GGG^{93}) form the majority of the cross-wall of the active site cleft, the border of the binding pocket across from the motif III amino acids. Ser173 and Leu174 form the remainder of the side of the pocket, and the Ser hydroxyl side chain, residing in close proximity to the motif III His, also acts as the catalytic moiety by the formation of a covalent bond with the malonyl extender unit by concomitant release of CoA. The AAAH motif, in combination with the spatially opposite tri-Gly motif, leaves a large cleft for the binding of mCoA derivatives with larger R groups, as observed in the biosynthesis of antimycins with relatively lengthy alkyl chains at the C7 position of the dilactone ring.
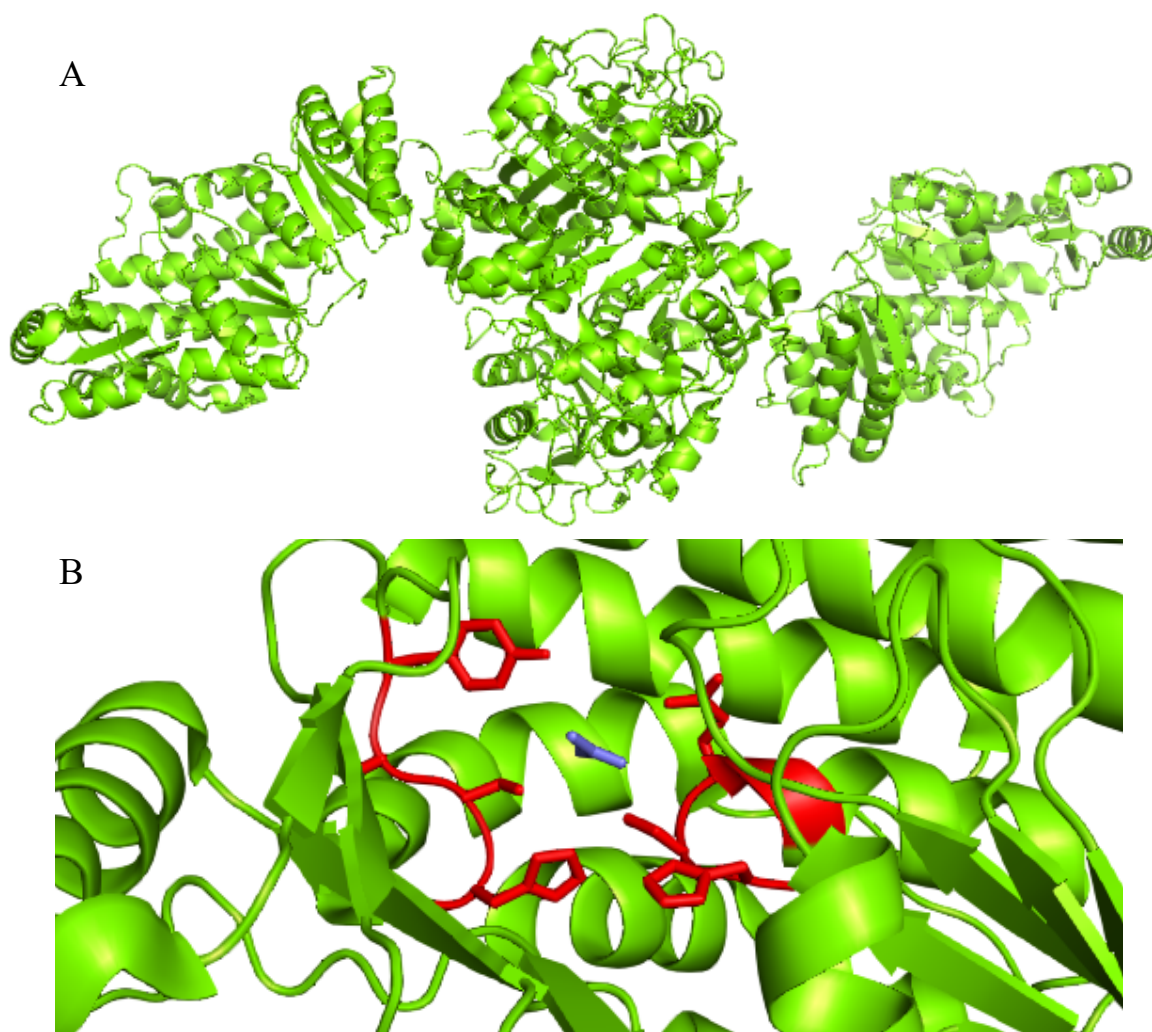
Figure 3.16: A) The dimeric structure of the third AT domain from the DEBS PKS pathway (PDB ID: 2QO3) (Tang et al., 2007). The protein dimerises through the N-terminal domain, and the C-terminal domain houses the catalytic motif. B) Zoomed view of the DEBS AT-3 domain active site, highlighting the YASH loop motif and the GHSQG catalytic motif, indicated in red colouring. An acetyl moiety is bound in the active site.

The 6-deoxyerythronolide B biosynthetic PKS system (DEBS), forming the erythromycin aglycone, has been well studied and a number of AT domain structures from the pathway have been solved (Figure 3.16) (Tang et al., 2006; Tang et al., 2007). All of the modules accept and catalyse the covalent ACP linking of mmCoA and all exhibit high levels of amino acid sequence identity, particularly in regions important for catalysis such as the YASH motif III sequence and the conserved location of the catalytic Ser in a GHSQG motif. The YASH motif forms one of the walls of the active site cleft, with the cross-wall formed primarily of a Gly, the catalytic residue at the bottom of the pocket, and three bulky Gln residues. One of the Gln

side chains protrudes directly into the centre of the binding site, forcing bound substrates closer to the YASH motif and thereby restricting the size of potential substrates and causing specificity of the AT domain for mmCoA. Mutation of the bulky residues lining the pocket to small amino acids has been shown to increase the promiscuity of the AT domain and allow activation of larger substrates. These structural analyses show that the residues forming the perimeters of the active site cleft are important in substrate binding and mediation of catalysis, and mutation of these residues is a viable strategy for the alteration of substrate usage of the enzymatic domain.
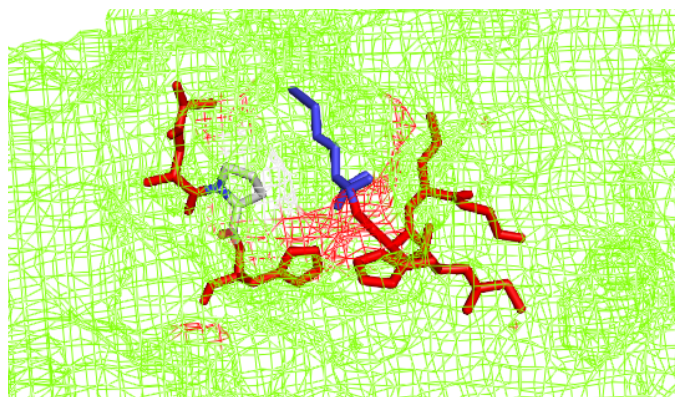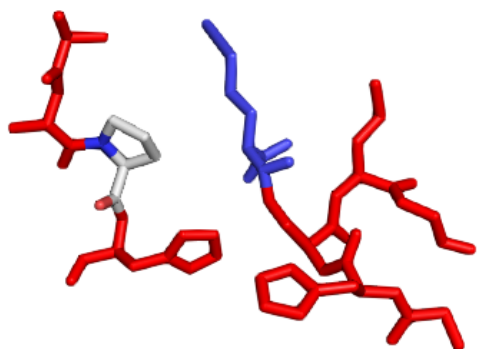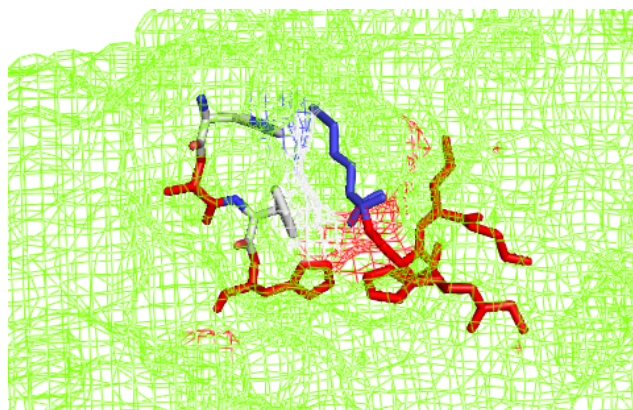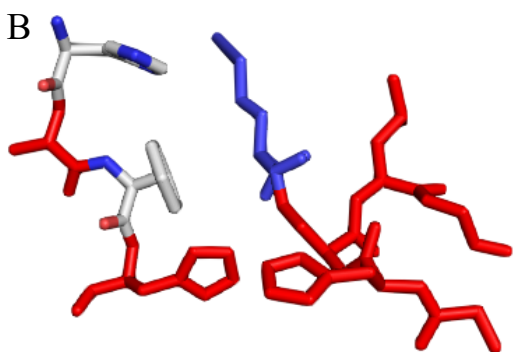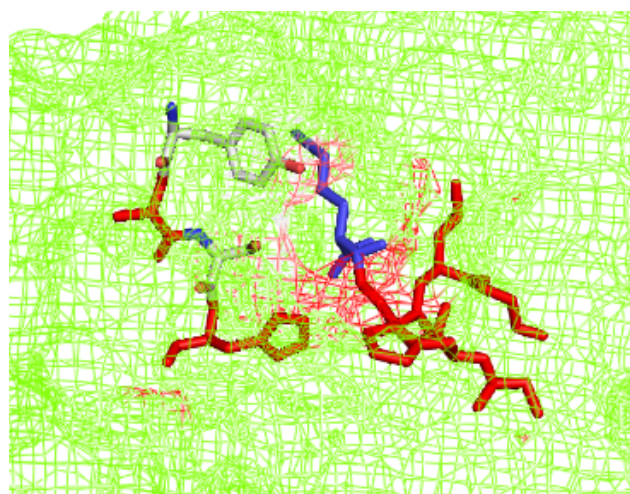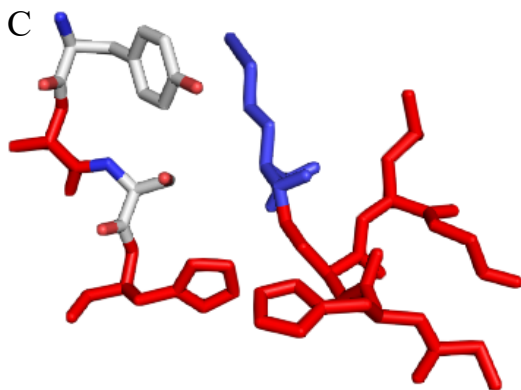
Figure 3.17: A) The AAPH motif (Ala to Pro mutation) modelled into the SpnD-AT domain structure, highlighted on the left with a zoomed view of the catalytic region and the motif III region. The motif III region and the conserved catalytic motif are shown in red in the mesh structure (right) with the bound pentynyl-malonyl SNAC coloured blue. The gap in the mesh indicates the binding site, with a pentynyl malonate group bound to the active site serine. Red or white colouring within the mesh indicates occlusion of the open binding site. The single Ala to Pro mutation slightly occludes the binding site but the substrate is well accommodated. B) The HAFH motif (Ala to His, Ala to Phe mutations) modelled into the SpnD-AT domain structure. Zoomed motif and binding site mesh views are shown. The mutations cause a more significant occlusion to the binding pocket. C) The YASH motif (Ala to Tyr, Ala to Phe mutations) modelled into the SpnD-AT domain structure. Zoomed motif and binding site mesh views are shown. The aromatic amino acids cause significant occlusion and repulsions in the active site pocket and restrict substrate access.

To understand in more details the structural basis of the lack of activity displayed by the YASH and HAFH motif III variants of the AntD AT domain, these motif mutations were modelled into the solved crystal structure of the protein in order to observe the effect on the global architecture of the protein and, primarily, the nature of the active site cleft. The mutants were separately modelled and, in each case, the most suitable rotamer was selected (Figure 3.17). The structure of the AT domain, with the YASH motif modelled, showed substantial intrusion into the binding pocket by the Ser, and more significantly, the Tyr residue, effectively reducing the space available for larger substrates (Figure 3.17C). This was particularly evident in the structure with a bound benzylmalonyl-SNAC substrate mimic which showed that the phenolic side chain of Tyr directly impinges on the binding site of the large R group of mCoA derivatives, and therefore, likely occludes binding of the majority of substrates for the AntD system leading to no antimycins being produced. The HAFH mutant seemed to similarly influence the space available within the binding pocket by the presence of large amino acid side chains, which were oriented towards the centre of the cleft (Figure 3.17B). The effect was increasingly pronounced with this motif due to the presence of two bulky, aromatic residues, with the His residue directed at the R group and the benzene ring of the Phe residue pointing towards the region occupied by the carbonyls forming the malonyl group. This dual occlusion severely restricts the space available within the pocket and consequently leads to lack of AT domain activity *in vivo*, and no antimycin production. Taken together, these structural analyses show that the space available within the binding pocket for the AntD AT domain is of prime

importance for the acceptance of the requisite substrate and for the formation of mature antimycin-type depsipeptides. The introduction of a single Ala to Pro point mutation to form an AAPH motif does not broadly reduce the overall space available in the active site pocket and seems to physically allow the binding of larger substrates, therefore resulting in the formation of certain antimycins (Figure 3.17A). The Pro residue, however, likely rigidifies a usually flexible hinge motif and restricts the motion of the neighbouring His residue, important in acid-base catalysis and the activation of substrates for covalent attachment into the active site and subsequently onto the ACP domain. The lack of movement in the hinge region may also lead to specification of the substrates that can potentially bind in the pocket, as there would be a reduced level of induced fit with potentially sub-optimal substrates, therefore leading to usage of only certain substrates and alteration of the antimycin product profile. Overall, the presence of the Pro residue allows the production of antimycins by acceptance of certain mCoA derivatives, whilst reduction of overall production likely occurred due to sub-optimal placement of catalytic residues in relation to the substrate. Molecular and structural rationalisation of alterations made to enzymes and proteins in directed engineering efforts can be extremely advantageous, and when combined with simple molecular techniques, such as CRISPR/Cas9, can be especially beneficial for the creation of improved systems for the potential synthesis of novel bioactive compounds.

3.4 Discussions and conclusions

Natural products are produced by a wide variety of bacterial species, with actinomycetes, and particularly *Streptomyces*, being prodigious producers of such secondary metabolites. The production of these useful bioactive compounds is usually carried out by specific proteolytic machinery, encoded by colocalised genes in a lengthy biosynthetic gene cluster. Many of these BGCs are, however, silent under normal laboratory conditions, and even if active, linking a produced metabolite or compound to its cognate BGC can be difficult due to the complexities of extracting and characterising such compounds (Huo et al., 2019). Although methods for activating silent BGCs in native hosts have become increasingly successful, the uncultivability or genetic recalcitrance of some organisms make these approaches difficult or impossible to carry out (Li et al., 2021). Cloning BGCs from such organisms followed by their heterologous expression is a common strategy in accessing exciting new chemistry. Although talented heterologous hosts exist, heterologous expression projects necessitate the cloning of large stretches of DNA, ideally corresponding to the whole of the BGC required for compound

production. Due to the length, and high GC genetic content, of the majority of BGCs, creating specific, markerless mutations and specifically engineering assembly line proteins can be a difficult chore. CRISPR/Cas9 has emerged as a genetically selective and rapidly programmable tool for the creation of specific mutations, when used in combination with established homologous recombination technologies. Various CRISPR/Cas9 systems have been used in the targeting and mutation of genomic DNA, including in species of *Streptomyces*, however targeting of extrachromosomal DNA constructs, specifically those containing cloned BGCs, has not been reliably carried out. The system adapted and developed in this work allows the fast, *E. coli*-based, targeting of natural product BGCs, for the creation of specifically engineered variants which can then be easily tested in multiple heterologous expression platforms, therefore allowing rapid and efficient characterisation of engineered BGCs.

The No-SCAR protocol was developed as a completely plasmid-contained CRISPR/Cas9 and λ-Red based homologous recombination system, for the creation of markerless *E. coli* genome mutations (Reisch and Prather, 2015). The adapted system utilised a regular *E. coli* cloning host and a specifically designed protospacer sequence, targeting the BGC, held on a multi-copy cosmid, or another extrachromosomal DNA construct. Use of linear, double-stranded DNA templates, stabilised by the λ-Red system, as templates for recombination and for the introduction of mutations was followed by the selection of positively mutated clones. The speed of the selection of successfully mutated clones was mediated primarily by the extrachromosomal nature of the constructs, and allowed a method of simply obtaining total extrachromosomal DNA from the *E. coli* strains, followed by re-transformation to obtain pure clones. Due to the nature of the homologous recombination template, an effective PCR screen was utilised to identify positively mutated clones, based on the introduction of extra, silent mutations in the template, to obtain a DNA template with a lower GC content. Positively identified clones, achieved at a rate of ~ 10%, could be subsequently used for expression tests of the cluster.

The antimycin BGC is composed a number of genes, whose product proteins are responsible for the production of the antimycin-type depsipeptides produced, as well as the regulation of the complete cluster (Seipke and Hutchings, 2013). The compound scaffold is formed predominantly by the NRPS and PKS modules, encoded by *antC* and *antD* respectively. Antimycins have been observed to have a myriad of uses and bioactivity against numerous

fungal species, as well as against fish, where it is used as a commercial piscicide. Due to these interesting uses and activities, the biosynthesis of antimycins has garnered attention as a potential method for creation of novel compound analogues with improved or an extended range of bioactivity. One of the most intriguing aspects of the biosynthesis is the presence of a largely promiscuous AT domain in the AntD PKS, capable of accepting and activating multiple mCoA variants provided by the pathway encoded crotonoyl CoA reductase enzyme (Liu et al., 2016). This promiscuity leads to the production of a variety of antimycins with alkyl groups of varying lengths at the C7 position of the dilactone scaffold. The enzymatic motifs enabling selection by AT domains have generally been identified, with motif III being readily conserved between AT domains that select mCoA (HAFH) or mmCoA (YASH). Therefore, mutation of motif III of the promiscuous AT domain from the AntD PKS was thought to potentially cause the biosynthesis of specific, novel antimycin-type depsipeptides.

The mutations, created by use of the adapted CRISPR/Cas9 system for mutation of large, extrachromosomally-contained BGCs, gave varied results. Mutation to the motifs mostly observed and conserved within other PKS systems (HAFH and YASH) resulted in complete loss of AT domain activity, and the subsequent loss of total antimycin production. However, a conservative point mutation to the motif observed only within the AT domain of neoantimycin biosynthesis (AAPH) resulted in production of antimycins, although a lower titre of compound was observed. Deletion of the gene encoding the standalone acyltransferase, AntB, responsible for the diversification of the hydroxyl group at the C8 position of the dilactone scaffold, resulted in a simpler chromatogram for the analysis of produced antimycins from the altered AT domain system. The commonly observed antimycins with four or six carbon linear alkyl chains at the C7 position were produced by the $\Delta antB$ variants and interestingly a peak for a potential antimycin with a C7 methyl group was observed. This antimycin could only have resulted from incorporation of mmCoA by the AntD AT domain, which differs from the natural substrate for the AAPH motif within the context of the neoantimycin system. The expected acceptance of the AAPH motif was mCoA, primarily due to investigations into the neoantimycin PKS module, showing dual functionality of the module-encoded methyltransferase domain at the same position of the antimycin intermediate. However, although only detected at low levels, single methylation events were also observed, indicating potential usage of mmCoA by NatC[AT] and showing consistency with the observed product from the instruction of the AAPH motif into AntD[AT] (Skyrud et al., 2018). Due to the relatively

low total levels of production from the mutant strain, fragmentation or MS2 patterns could not be obtained and the compound was not able to be purified. Although complete structural identification was restricted due to production titres, antimycin-type compounds readily break down under commonly utilised mass spectral conditions, resulting in breakdown peaks, or fragments, being produced during a regular MS run. Assignment of a number of these smaller, breakdown peaks allowed tentative assignment of the peak as resulting from the *in vivo* production of an antimycin with a methyl group at the C7 position, a novel antimycin-type compound not previously observed from *in vivo* extracts.

Engineering of enzymes involved in the biosynthesis of natural products, or those encoded by natural product BGCs, is of great interest and could contribute hugely towards the creation of improved variants of specialised metabolites, or in the biosynthesis of completely novel bioactive compounds. Structural data is of great use in this context as it enables a targeted approach to protein engineering, as well as the rationalisation of effects of mutations made to biosynthetic proteins (Wilkinson and Micklefield, 2007). Therefore, a mutational and bioengineering platform, based on high quality structural data, yields much more effective results, as well as outcomes that are increasingly applicable to similar proteins from alternative systems. A structure of the AT domain from the biosynthetic pathway of the splenocins, antimycins with aromatic groups at the C7 position, was published during the course of the study (Li et al., 2018). Analysis of the structure, specifically the active site of the protein and the motifs responsible for substrate selection, allowed development of a rationale for the outcomes observed for the mutations made in the AntD AT domain. Introduction of YASH and HAFH motifs, replacing the small amino acids present in the natural AAAH motif, abrogated production of any antimycins *in vivo*, seemingly due to occlusion of the active site space. The cognate AAAH motif III sequence is opposed in the natural binding pocket by a tri-Gly motif, importantly leaving copious amounts of space for the binding of a variety of large mCoA derivatives. Point mutation of the AntD AT domain motif III sequence to AAPH minorly alters the shape of the side of the pocket, but importantly does not protrude far into the active site itself, allowing substrates to bind relatively freely. The potential acceptance and formation of an mmCoA containing antimycin seems to show that the natural AntD AT domain is capable of accepting small mCoA derivatives, but due to the voluminous binding site, prefers to activate the larger compounds. Slight restriction of the binding site and alteration of the side wall shape allows activation of mmCoA by the AT domain, and therefore results in the formation of the antimycin with a C7 methyl group. Analysis of the AT domain structures from

the DEBS PKS system also showed that the amino acid residues forming the remainder of the binding site are important for substrate selection; the AT domains contain Gln residues directed into the active site, making the domains selective for mmCoA. Therefore, in mutation and engineering of enzyme active sites for the alteration of substrate scope, it is important to consider the totality of the pocket as it is likely that most residues play important roles in substrate binding or catalysis. Combining new molecular editing technologies with structural analysis towards the engineering of proteins encoded by natural product BGCs has the potential to yield novel metabolites with interesting bioactivities, and could, therefore, rapidly increase the availability of compounds targeting drug resistant, or other problematic microorganisms.

Chapter 4 – The structure and function of a standalone β-ketoreductase that acts concomitantly with biosynthesis of the antimycin core scaffold

Abstract

Non-ribosomal peptide synthetase (NRPS), polyketide synthase (PKS), and hybrid NRPS-PKSs are responsible for the production of a large number of structurally diverse and therapeutically important compounds. These long modular assembly lines frequently contain enzymatic domains serving to increase the diversity and chemical space of the product. Antimycin-type depsipeptides are formed by the action of such a hybrid NRPS-PKS biosynthetic system. The pathway contains a standalone β-ketoreductase enzyme, AntM, which is surprisingly required for the production of the antimycin core scaffold. In this chapter, AntM was characterised *in vivo*, and showed *in vitro* as an NADPH-dependent ketoreductase. X-ray crystal structures of the protein showed that AntM contains a long channel parallel to, and incorporating the NADPH binding motif. Substrate modelling and mutational analysis of the binding channel suggests that the enzyme preferentially binds and catalyses reduction of the C8 ketone within the final linear intermediate of antimycin biosynthesis, not the thioesterase-released cyclised product. An increased understanding of this class of enzymes could inform the detection, localisation, and characterisation of similar enzymes playing roles of expanded significance within biosynthetic pathways. Standalone enzymes are also an attractive target for biosynthetic pathway engineering and therefore, insights gained here could aid in the production of non-natural "natural" products.

4.1 Introduction

4.1.1 Reconstitution of antimycin biosynthesis *in vitro*

Antimycins were first isolated over 70 years ago, with the BGC and the hybrid NRPS/PKS pathway directing the synthesis discovered in 2011 (Dunshee et al., 1949; Yan et al., 2012; Sandy et al., 2012). The many compounds comprising the family of antimycins all contain the nine-membered dilactone scaffold and the rare 3-formamidosalicylate starter unit, giving the compounds potent antifungal and insecticidal activity, specifically targeting the cytochrome c reductase complex as part of the mitochondrial electron transport chain (Tappel, 1960). Antimycins are also important, commercially utilised piscicidal compounds, with recent reports also suggesting specific variants of the antimycins display anticancer activity by

targeting the Bcl-2/Bcl-x_L complexes, upregulated in certain forms of cancer (Tzung et al., 2001). The biosynthesis of antimycins is discussed in section 1.7.

The biosynthesis of antimycins in *in vivo* systems involves the action of 15, 16, or 17 genes, dependent upon the nature of the gene cluster harboured by the specific bacterial species (Seipke and Hutchings, 2013). The functionality of the hybrid pathway has been dissected by a combination of *in vivo* and *in vitro* methodologies, leading to a detailed characterisation of most aspects of antimycin biosynthesis. The nine-membered dilactone core of the antimycins is formed predominantly by the actions of the NRPS megaenzyme, AntC, as well as the PKS, AntD, with other enzymes acting as standalone factors for requisite tailoring reactions. *In vivo* biosynthesis of the antimycins is discussed in detail in section 1.7.3. In a study carried out by Sandy et al. (Sandy et al., 2012) the *in vitro* biosynthesis of the antimycin dilactone scaffold was shown to be carried out by a minimal set of six proteins from the BGC (Figure 4.1). The set of enzymes comprised the NRPS and PKS megasynthases, AntC and AntD respectively, as well as AntE, AntF, AntG, and AntM. AntF is an acyl-CoA ligase homologue and AntG is a standalone carrier protein, that act as priming and activation enzymes for the continuation of biosynthesis by the megaenzyme synthases. AntF adenylates the starter unit, the commercially available compound anthranilate, used *in lieu* of the structurally similar 3-FSA observed *in vivo*, which obviates the need for AntHIJKLNO usually functional in the construction of the starter unit. AntF subsequently loads the activated compound onto the Ppant extension of AntG, which can then provide the initiation unit for the dimodular NRPS AntC. L-Thr and pyruvate are condensed onto the starter unit by the NRPS, followed by the addition of a two-carbon PKS extender unit, provided to AntD by the crotonyl-CoA reductase, AntE. AntM is required for the reduction of the β-keto group and the action of the PKS-encoded TE domain results in cyclisation and offloading of the antimycin-type product.

The action of the six semi-purified proteins, incubated with the required cofactors, substrates, and precursors, led to the production of an antimycin-type compound, containing the central dilactone scaffold which typifies compounds from the family (Figure 4.1). The *in vitro* system was therefore, able to dissect the functionality and enzymology of the proteins required for antimycin biosynthesis and establish a potential platform for the production of antimycin derivatives or reprogramming of the biosynthesising assembly line. A number of interesting aspects of the biosynthesis were also noted, including the relaxed substrate specificity of the

acyltransferase domain of the PKS, AntD, which is capable of utilising a variety of acyl-CoA substrates. Interestingly, in *in vitro* reactions omitting the standalone ketoreductase domain, AntM, no unreduced, carbonyl-containing antimycin-type compounds were detected. Indeed, no compounds were produced when only AntC, AntD, AntE, AntF, and AntG were incubated with required substrates, showing that AntM functionality is required for the production of antimycins *in vitro* (Sandy et al., 2012).
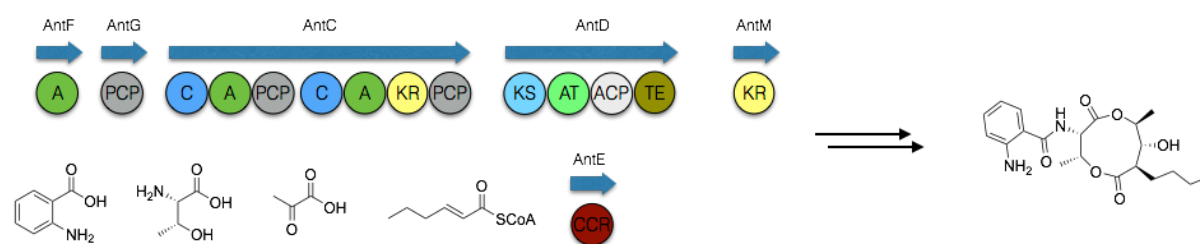
Figure 4.1: The system utilised to reconstitute the formation of the dilactone scaffold of antimycins *in vitro* (Sandy et al., 2012). A six-protein system (AntCDEFFM) along with the required substrates and cofactors resulted in the formation of an antimycin-type compound.

4.1.2 AntM as an essential β-ketoreductase

Upon discovery of the antimycin BGC, AntM was predicted to be a *trans*-acting KR domain, likely functional at the C8 position of the antimycin dilactone ring, predicted by analysis of the final compound and the presence of a diversified secondary alcohol at that position of the ring. Interestingly, *in vitro* reconstitution of the antimycin biosynthetic pathway, specifically utilising the enzymes involved in dilactone ring formation, showed that AntM was essential for the production of antimcyins from the *in vitro* system, as no C8 ketone-containing compounds were formed in reactions omitting the KR domain (Figure 4.2). AntM was seemingly, therefore, essential for the production of antimycins of any variety, an unusual case for an enzyme with β-ketoreductase activity.
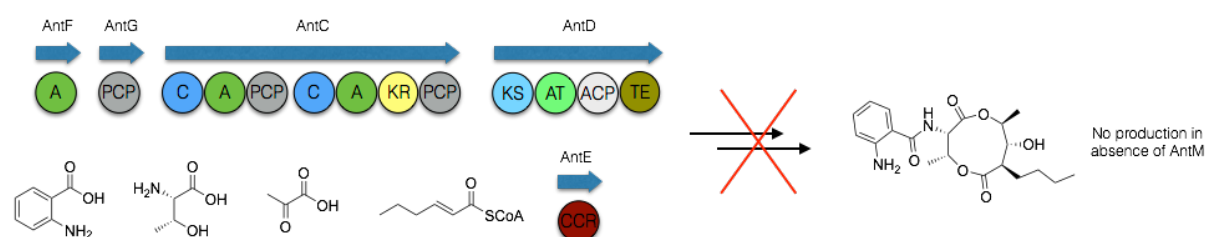
Figure 4.2: The requirement of AntM by the antimycin pathway was shown in the *in vitro* reconstitution system where no products were produced when AntM was omitted from the reaction mixture.

A number of enzymes and domains involved in compound biosynthesis have been shown to be essential for the assembly line-mediated production of natural products. These can include a variety of enzymes, involved at different stages of the biosynthetic pathway, including before and during production of the main compound scaffold (Finking and Marahiel, 2004). However, the essentiality of putative modifying domains is extremely rare and not commonly observed. The indication, therefore, that AntM functionality was an absolute requirement for the production of any antimycins from the *in vitro* reconstituted system was mechanistically unexpected and requires further investigation.

4.1.3 Standalone KRs in the biosynthesis of microbial natural products

Post-assembly line tailoring ketoreductase enzymes are often observed in PKS pathways, where enzymatic activity differs from *cis*-encoded KR domains, whose functionality occurs solely on β-keto groups of ACP-appended intermediates (Reid et al., 2003). Post-PKS KR domains can be functional on a variety of ketones in a range of scenarios, primarily as substrates are no longer tethered to assembly line domains (Olano et al., 2010). The wide range of ketones upon which post-PKS tailoring enzymes were active led to interest into the potential biocatalytic applications of the domains. However, no standalone KRs from type I PKS, or hybrid NRPS-PKS systems, have been characterised to date, and only the functionality of standalone KRs from type II PKS systems have been described in detail (Javidpour, Korman, et al., 2011).

Standalone KRs from type I PKS systems generally function on either ACP-bound intermediates, or assembly line-offloaded substrates. FdmC, the standalone KR from the type II PKS system leading to fredericamycin biosynthesis, is functional on an ACP-tethered intermediate, where it reduces the β-keto group as part of the chain initiation process (Das et al., 2010). Following ketoreduction, dehydration leads to alkene formation, yielding the hexadienyl intermediate utilised downstream for fredericamycin formation. The polyketide A-74528 similarly originates from an ACP-bound hexadienyl intermediate, and the SanC KR plays a similar role in the reduction of the β-keto group of the ACP-bound intermediate

(Fitzgerald et al., 2013). Despite these findings, standalone KR enzymes functional on ACP-tethered intermediates have not been characterised in detail and require further investigation.
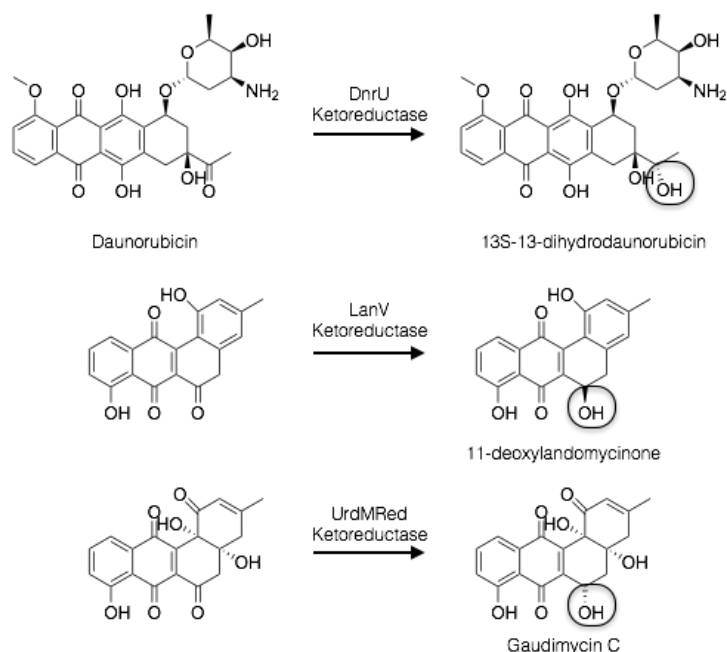
Figure 4.3: Overview of the catalysis carried out by a selection of post-assembly line KR domains, from the doxorubicin pathway (top), and from varying pathways to angucycline-type products (lower two panels). Varying substrate structures, stereo-, and regioselectivites are observed in the action of standalone KR domains. The resultant hydroxyl group is circled in the products of each catalytic reaction.

PKS pathways, including those biosynthesising the antitumor antibiotic doxorubicin and the anticancer angucyclines, often contain extensive post assembly line modifying enzymes, with KRs featuring prominently and playing important roles in the diversification of chemical scaffolds (Lomovskaya et al., 1999). Angucyclines, for example, are heavily modified after biosynthesis of the tetracyclic core in a thiotemplated manner, where hydroxylation and, particularly, reduction reactions are prominent (Figure 4.3) (Kharel et al., 2012). Tailoring KR enzymes are similar to their PKS module encoded counterparts, and belong to the short chain dehydrogenase (SDR) family of enzymes, employing the same Ser, Tyr, Lys catalytic triad. Enzymatic functionality of these KRs, however, occurs after compound release from the terminal ACP domain by the dedicated assembly line release factor, and often on ketone substrates in highly varied chemical loci (Mayer et al., 2005; Xiao et al., 2020). Due to the substrate variation of post-PKS tailoring KR enzymes, the structures of the reaction catalysing

enzymes are also seemingly adjusted to accommodate the varied substrates. Early studies into the structures of tailoring KRs showed that the active sites were less enclosed and more open to bulk solvent, potentially allowing the binding of larger and more structurally varied substrates, when compared to *cis*-encoded β-ketoreductases (Paananen et al., 2013; Patrikainen et al., 2014). The overall tertiary structures of the *cis*- and post-PKS enzymes are similar, with both possessing Rossmann fold architecture and binding the reduction cofactor NADPH, highlighting the subtle alterations to protein structure within enzyme classes that allow catalysis and functionality on a wide range of substrate compounds.

Despite structural variations that allow enzymatic functionality on a range of substrates, catalysis by post-PKS tailoring KR enzymes occurs through the same mechanistic route, beginning with cofactor and substrate binding in the open, extended binding site (Javidpour, Das, et al., 2011; Javidpour et al., 2013). The catalytic Tyr subsequently transfers a proton to the keto substrate and directional transfer of a hydride ion from the NADPH cofactor results in reduction of the substrate, and formation of the secondary alcohol product. The shift in substrate specificity between *cis*-encoded and post-PKS KR domains has allowed vast diversification of specialised metabolites, often yielding novel pharmacophores for varied biological activities.

4.2 Aims and objectives

Previous data suggested that AntM reduces the β-keto group installed by the final PKS module of AntD, leading to the formation of a secondary alcohol, which is further derivatised by a standalone acyltransferase enzyme, AntB. *In vitro* analysis also suggested that AntM may act upon a linear ACP-bound intermediate and its action may be required for compound release by the *cis*-encoded TE domain. Thus, the ultimate aim of this project was to probe this possibility by characterising AntM *in vivo*, *in vitro*, and structurally.

4.3 Results

4.3.1 Creation of a Δ*antM Streptomyces albus* S4 strain

The observation that AntM was obligately required for product release and antimycin production in *in vitro* studies was interesting for a number of different reasons. Upon first inspection of the antimycin biosynthetic gene cluster, *antM* is located at a genetic locus 7-8

genes downstream of the core genes involved in synthesis and production of the antimycin dilactone scaffold, potentially indicating differences in regulation and temporal gene expression. KR domains are also usually *cis*-encoded within PKS biosynthetic systems, where they catalyse β-ketoreduction of the carbonyl added by the preceding module, resulting in a secondary alcohol in the absence of other tailoring domains. The fact that AntM is a standalone ennzyme, and not encoded within the AntD PKS, also initially indicates a post-assembly line catalytic activity for the enzyme. A further point of interest was the apparent requirement of AntM β-ketoreductase activity, before TE-mediated macrocyclisation. The assembly line release reaction, however, does not directly involve the chemical position at which AntM is active, rather occurring between the Thr side-chain alcohol and the terminal carbonyl, participating in the pre-release thioester bond with the AntD ACP. All of these factors suggested interesting functionalities of AntM warranting detailed investigation.

The *in vitro* requirement of AntM in the production of antimycin scaffolds was an indirect indicator as to the *in vivo* essentiality of the enzyme. To determine if *antM* was an essential gene for the production of any antimycins *in vivo*, the ReDirect recombineering protocol (Gust et al., 2003) was used to generate a Δ*antM Streptomyces* strain, which was subsequently tested for production of specialised metabolites. ReDirect recombineering resulted in markerless, in-frame deletion of the *antM* gene on the cosmid holding the entire antimycin BGC (Cos213ΦC31), which was consequently conjugated into the *S. albus* S4 Δ5 strain, used to generate chemical extracts with a cleaner background for LC-HRMS analysis (Figure 4.4) (Fazal et al., 2020). The Δ5 strain contained a complete pathway deletion of the antimycin BGC, which would facilitate easier analysis of generated extracts for the presence of antimycin-type compounds.

A

antM, 777 bp

Primer 1

Hyg cassette, 1584 bp

Primer 2

In-frame scar sequence, 81 bp

Primer 1   Primer 2
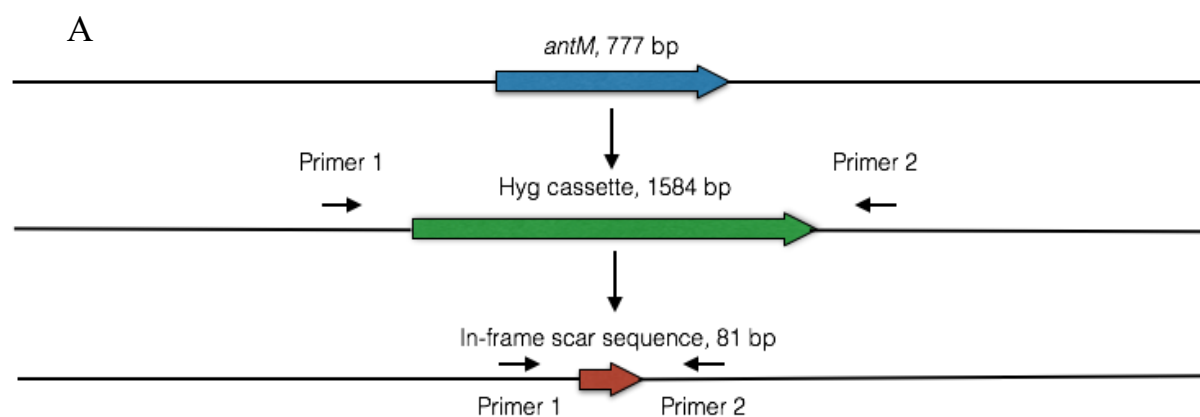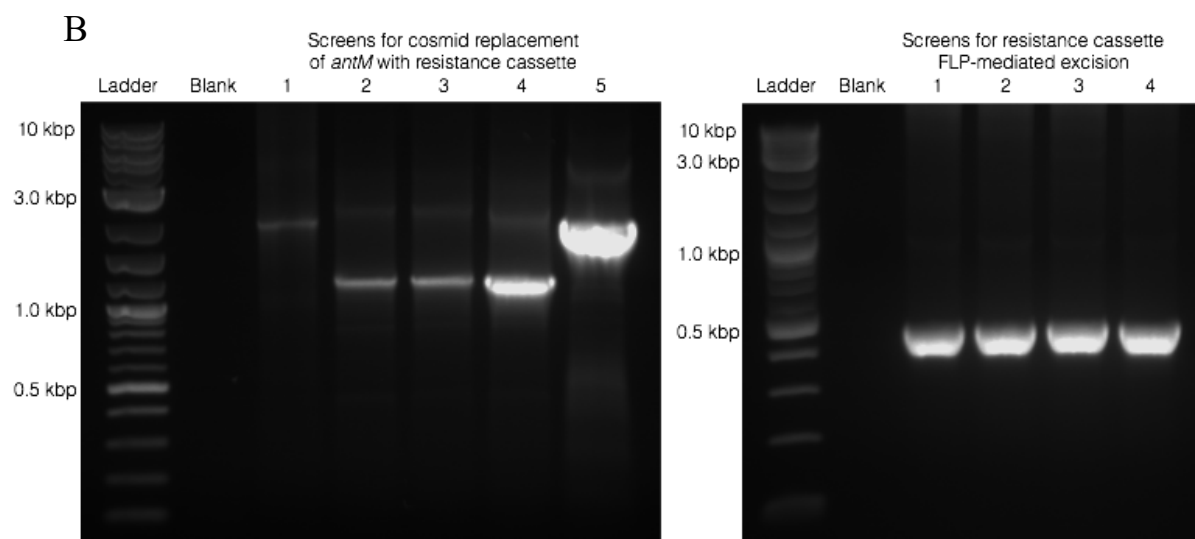
Figure 4.4: A) Schematic showing overview of deletion of *antM* gene and subsequent result of FLP-mediated excision, indicating relative sizes of genes. B) Agarose gels screening for success of *antM* cosmid deletion and replacement with resistance cassette (left), and subsequent FLP-mediated excision of the resistance cassette (right). The expected 1941 bp band was observed in lanes 1 and 5 of the gel screening for *antM* replacement with the hygromycin selective resistance cassette, indicating successful recombineering. The expected 438 bp band was observed in all lanes in the gel screening for excision of the resistance cassette and the in-frame scar sequence that remained. The success of the protocol afforded a markerless deletion of the *antM* gene.

4.3.2 Production of antimycins from the Δ*antM* strain

Chemical extracts were prepared by solid phase extraction (SPE) from liquid cultures of the *S. albus* S4 Δ5 *attB* ΦC31::Cos213 and *S. albus* S4 Δ5 *attB* ΦC31::Cos213 Δ*antM* strains, and analysed by LC-HRMS for the production of antimycins, or antimycin-type compounds. The strain containing genomic integration of the wildtype (WT) cosmid showed clear production of the WT suite of antimycins, namely antimyicns A1, A2, A3, and A4 (Figure 4.5). When the Δ*antM* strain was cultured for the production of specialised metabolites, no production of the initial suite of antimycins A1-4 was observed in SPE purified chemical extracts. Interestingly, antimycins harbouring a C8 carbonyl moiety, unreduced antimycins, were also not observed in extracts, corroborating previous *in vitro* data showing the lack of production in the absence of AntM (Figure 4.5). In addition, the corresponding unreduced linear antimycin intermediates, which could feasibly be produced by premature cleavage of the intermediate from the PKS assembly line, were also not present within extracts produced from the Δ*antM* strain.

A

B
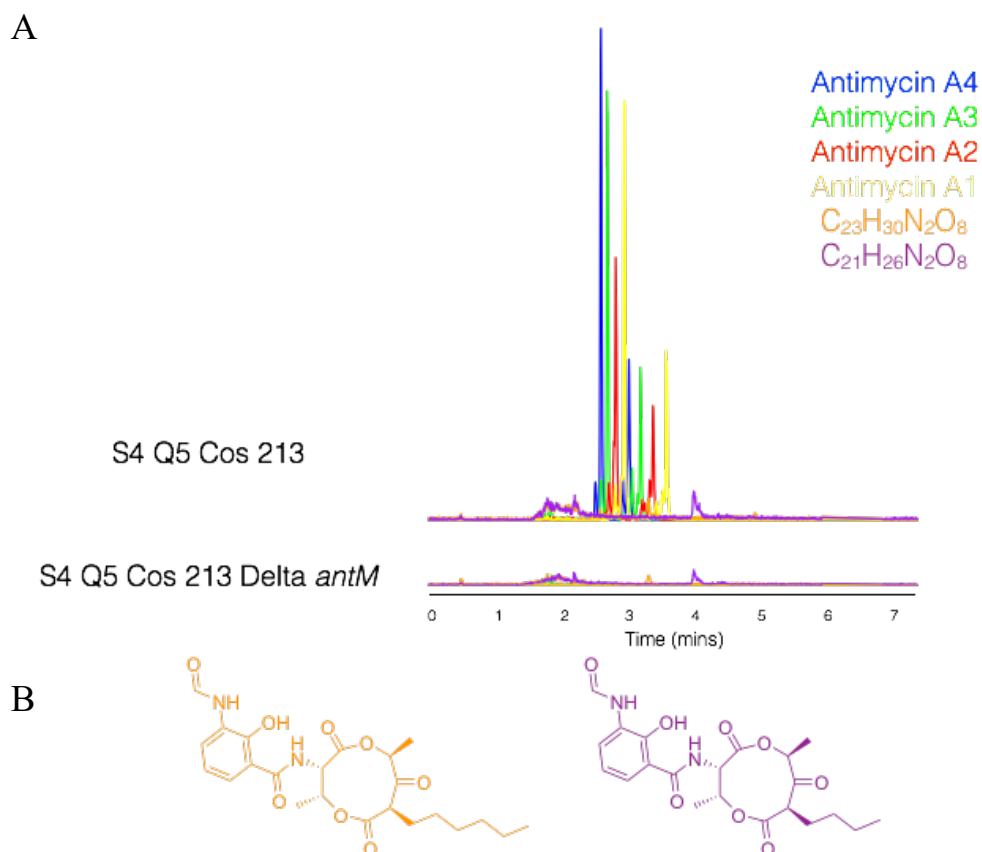
Figure 4.5: A) LC-HRESIMS analysis of chemical extracts prepared from the denoted antimycin production strain. Extracted ion chromatograms ($[M+H]^+$, $[M+Na]^+$) corresponding to each of the four commonly produced and observed antimycin variants (A4: $C_{25}H_{34}N_2O_9$, A3: $C_{26}H_{36}N_2O_9$, A2: $C_{27}H_{38}N_2O_9$, A1: $C_{28}H_{40}N_2O_9$), as well as two ketone-containing (or unreduced) antimycins ($C_{23}H_{30}N_2O_8$, $C_{21}H_{26}N_2O_8$) are shown. Exact neutral m/z values used to determine EICs were: 548.2734, 534.2577, 520.2421, 506.2264, 462.2002, 434.1689. The y-axis intensity was set to 1 x 10^5 in each case. B) The chemical structures of the two putative ketone-containing antimycins that may be produced *in vivo*, in the absence of AntM.

In order to confirm the phenotypic lack of antimycin production was due only to deletion of the *antM* gene and not any other random mutational events, the complementation strain *S. albus* S4 Δ5 *attB* ΦC31::Cos213 Δ*antM attB* ΦBT1::pIJ10257-*antM* was created. This strain harbours the wild-type *antM* gene, whose expression is driven by the *ermE** promoter. Methanolic extracts were once again prepared by SPE from the complementation strain, and the extracts analysed by LC-HRMS. Upon complementation of the *antM* deletion mutant with *antM,* the production titre of antimycins A1-4 were restored to near WT levels (Figure 4.6). Although antimycin production was restored, the product portfolio displayed a significant deviation from what was originally observed from the wildtype antimycin producing strain.

The antimycins produced displayed only the peak at the earlier elution time (between 2 and 3 minutes), rather than the characteristic early and late elution peaks. This could have been due to alteration in spatiotemporal expression of *antM*, as the gene was under the control of a constitutively active promoter in the complementation strain. This could have led to increased activity of the AntM KR domain, subsequently leading to a shifted pool of antimycin intermediates present and potentially a changed antimycin product portfolio. This possibility however, requires further investigation. These results, taken together, corroborate previous *in vitro* experiments indicating the absolute requirement of AntM for antimycin production. The results also indirectly indicate that β-ketoreduction of the C8 carbonyl by AntM likely occurs on the final linear biosynthetic intermediate while it is still covalently attached to AntD^{ACP}, making the AntM β-ketoreductase the only identified standalone enzyme of its type thus far.
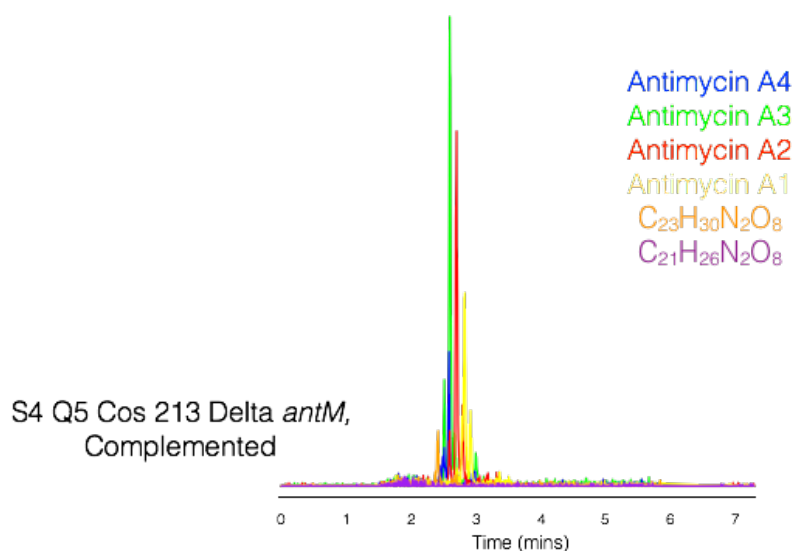
Figure 4.6: LC-HRESIMS analysis of chemical extracts prepared from the *antM* deletion strain, complemented with a functional, and constitutively expressed, copy of *antM*, *in trans*. Extracted ion chromatograms ([M+H]$^+$, [M+Na]$^+$) corresponding to each of the four commonly produced and observed antimycin variants (A4: $C_{25}H_{34}N_2O_9$, A3: $C_{26}H_{36}N_2O_9$, A2: $C_{27}H_{38}N_2O_9$, A1: $C_{28}H_{40}N_2O_9$), as well as two ketone-containing (or unreduced) antimycins ($C_{23}H_{30}N_2O_8$, $C_{21}H_{26}N_2O_8$) are shown. Exact neutral m/z values used to determine EICs were: 548.2734, 534.2577, 520.2421, 506.2264, 462.2002, 434.1689. The y-axis intensity was set to a maximum of 1 x 10^5.

4.3.3 Overproduction and purification of AntM

The hypothesis that the AntM β-ketoreductase may act on an ACP-bound linear intermediate was intriguing and led to studying this enzyme in greater detail *in vitro*. Therefore, the enzyme was overproduced and purified for use in *in vitro* assays. The coding sequence of the *antM* gene was cloned into the protein expression vector pET28a, a plasmid designed to express A es from the IPTG-inducible T7 promoter and produce N-terminal hexahistidine fusion proteins ((His)$_6$-AntM in this case). The AntM protein was purified from lysates by immobilised metal affinity chromatography (IMAC) with Ni-NTA resin, followed by a further purification step of gel filtration chromatography, yielding a pure AntM protein sample (Figure 4.7). Interestingly, (His)$_6$-AntM eluted off the gel filtration column as an apparent homo-tetrameric species, potentially composed of four AntM monomers. The identity of the purified protein was confirmed by SDS-PAGE analysis, showing that (His)$_6$-AntM had the anticipated monomeric molecular weight of 29988.70 Da. This was further corroborated by HRMS analysis of the protein sample, showing the molecular weight as expected for the recombinant protein.
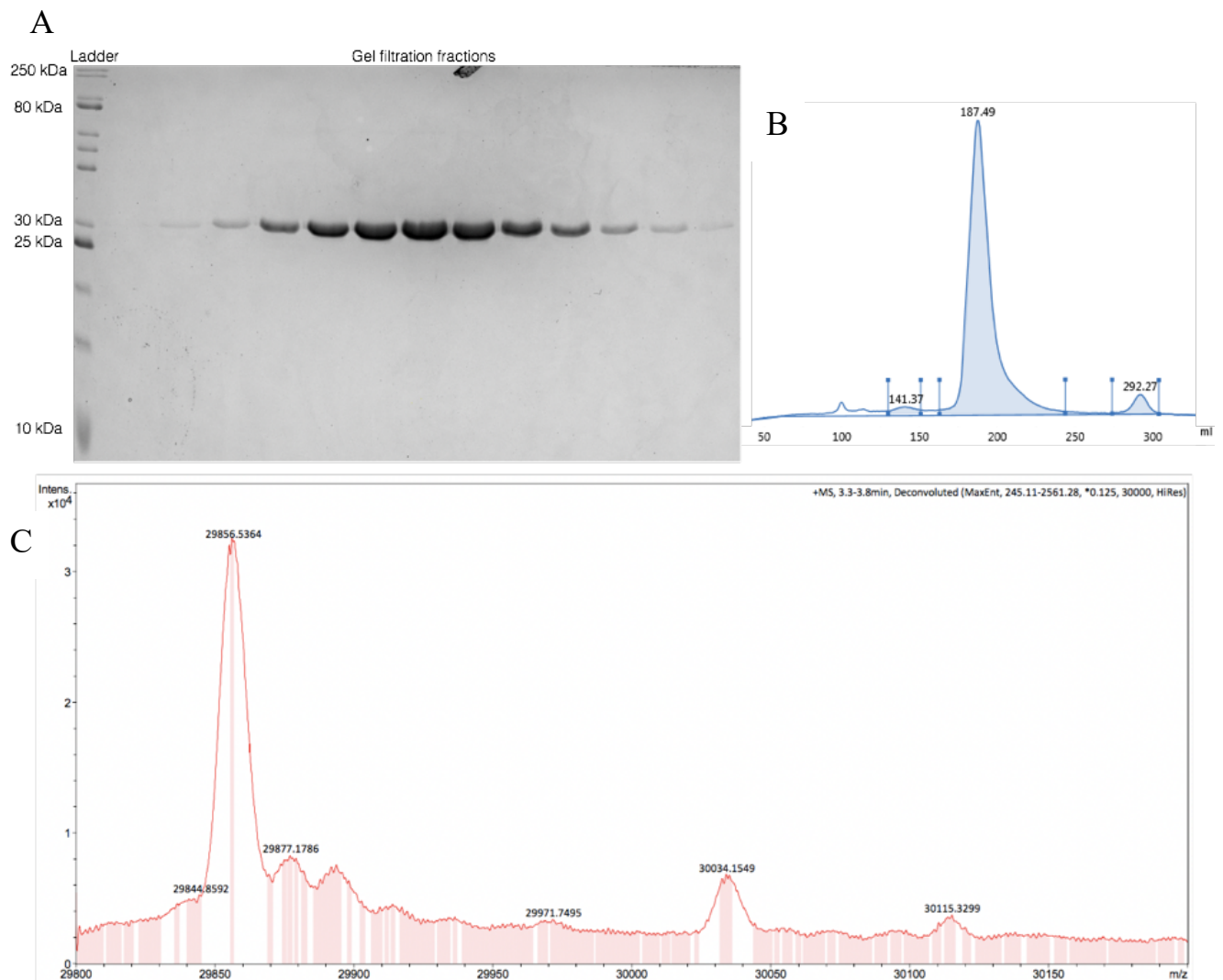
Figure 4.7: A) SDS-PAGE analysis of purified His$_6$-AntM, alongside molecular weight ladder. B) Absorption at 280 nm of elution buffer during gel filtration purification of His$_6$-AntM. The protein eluted at an elution volume of 187.49 mL indicative of a species of approximate mass 120,384 Da C) Deconvoluted single charge mass spectra of (His)$_6$-AntM. The major species has a mass of 29856.5364 Da, indicative of the His$_6$-AntM protein missing the N-terminal methionine residue.

4.3.4 ITC assays of AntM with reduction cofactors

In order to gain an initial understanding of AntM functionality, we sought to confirm the identity of its cognate nicotinamide cofactor by developing an isothermal titration calorimetry (ITC) assay. The previous *in vitro* antimycin reconstitution study utilised NADPH as the reduction cofactor, likely due to the fact that most β-ketoreductases involved in and associated with NRPS and PKS assembly lines bind and oxidise NADPH to achieve reduction of cognate substrates (Sandy et al., 2012). However, examples of β-ketoreductase enzymes using the non-phosphorylated nicotinamide cofactor NADH exist, as well as promiscuous enzymes, non-specific for the selection of reduction cofactors (Di Luccio et al., 2006).

The ITC assay was carried out using (His)$_6$-AntM, with a 10-fold molar excess of either NADPH, NADP$^+$, or NADH, in order to obtain thermodynamic parameters of any interactions that may occur. A number of different protein and cofactor concentration pairs were tested and titrations utilising 50 μM (His)$_6$-AntM with 500 μM nicotinamide cofactor gave the best heat change signal and highest quality raw thermograms. The titration of (His)$_6$-AntM with a 10-fold molar excess of NADH yielded no binding response, indicating that (His)$_6$-AntM does not utilise NADH *in vitro* and that the enzyme prefers phosphorylated nicotinamide cofactors (Figure 4.8). The ITC assay of recombinantly produced AntM with NADPH yielded a sigmoidal isotherm plot, with a 1:1 stoichiometry of binding and a dissociation constant (K_D) of 6.4 μM. Similarly, titrations performed with (His)$_6$-AntM and NADP$^+$ gave a sigmoidal isotherm plot with a 1:1 stoichiometry of binding and a K_D of 19.5 μM (Figure 4.8). AntM displayed a ~ 3-fold preference for binding NADPH over NADP$^+$, as expected for the use of NADPH as the cofactor for catalysis of the reduction reaction. The stoichiometry of binding showed each AntM monomer was capable of binding one cofactor molecule, and each

potentially capable of catalysing reduction of one antimycin intermediate, leading to subsequent release and product formation.
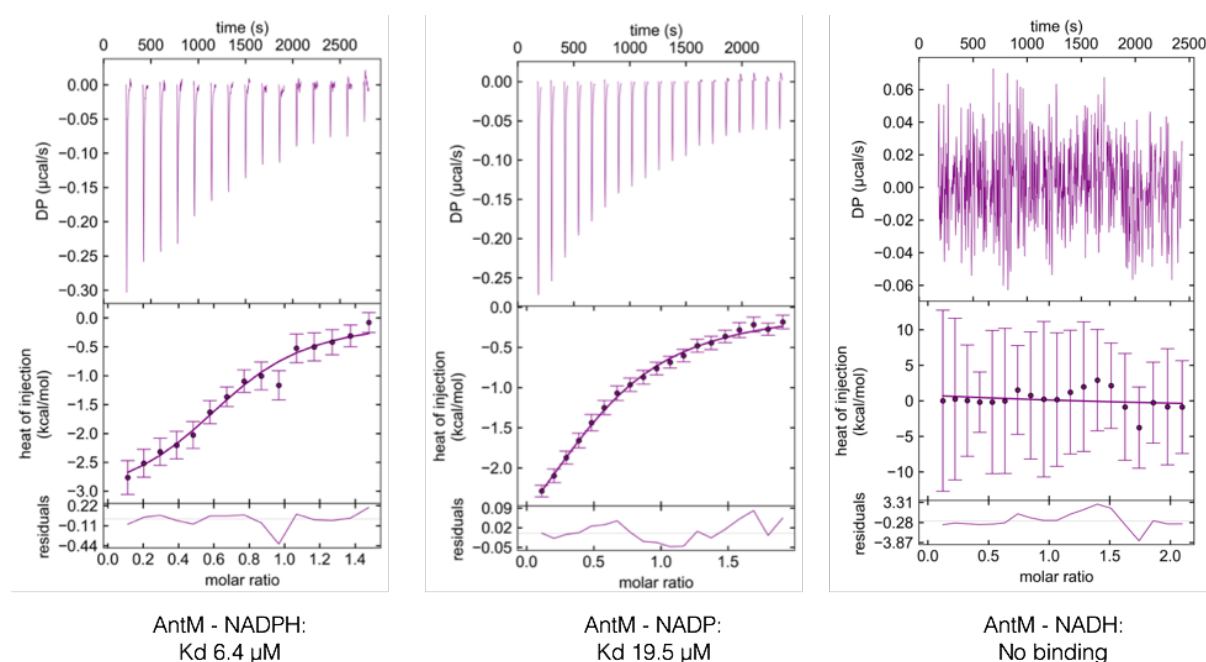
Figure 4.8: ITC traces for binding interaction analyses carried out with (His)$_6$-AntM and the designated nicotinamide cofactor. AntM showed binding with NADPH and NADP$^+$, but not with NADH.

4.3.5 Reduction assays of AntM with commercially available model substrates

The expected, natural substrate for AntM was a linear, peptide-ketide hybrid, incorporating a rare 3-FSA unit, making synthesis of the substrate synthetically challenging. PKS-related KR domains have historically been studied in isolation for their ability to reduce commercially available, and often commercially important, ketone substrates to the corresponding alcohols. The majority of the KR domains studied in this way are *cis*-encoded within their cognate PKS assembly lines (Bali et al., 2006). These KRs usually exhibit a convenient promiscuity with regards to their reduction of multiple ketone substrates, which allows their thorough kinetic characterisation. Due to the inaccessibility of the natural substrate, the use of alternative, simple ketone substrates could allow characterisation of AntM activity, useful for the probing of enzyme function and gaining an initial understanding of AntM substrate scope. Thus, AntM ketoreductase activity with a series of commercially available, simple ketone substrates was tested.

In total, eight substrates were selected for use in a continuous NADPH-coupled spectrophotometric assay. Ketone substrates were selected based on their structural simplicity, resemblance to antimycin biosynthetic intermediates or successful use with KRs from other PKS systems. Assays were initiated by the addition of the respective substrate, after an incubation period of the enzyme with the determined reduction cofactor, NADPH. Following initiation, the assays were monitored continuously for the loss of absorbance at $\lambda = 340$ nm, demonstrating the usage of NADPH throughout the assay, and thereby allowing the determination of the velocity of reduction and the kinetics of the catalytic activity. (His)$_6$-AntM was not active upon any of the substrates tested when using reasonable reaction conditions (5 μM (His)$_6$-AntM, 1 mM NADPH, 0-50 mM substrate) (Table 15).

Name	Structure	Assay
2-heptanone		
2-octanone		
3-octanone		
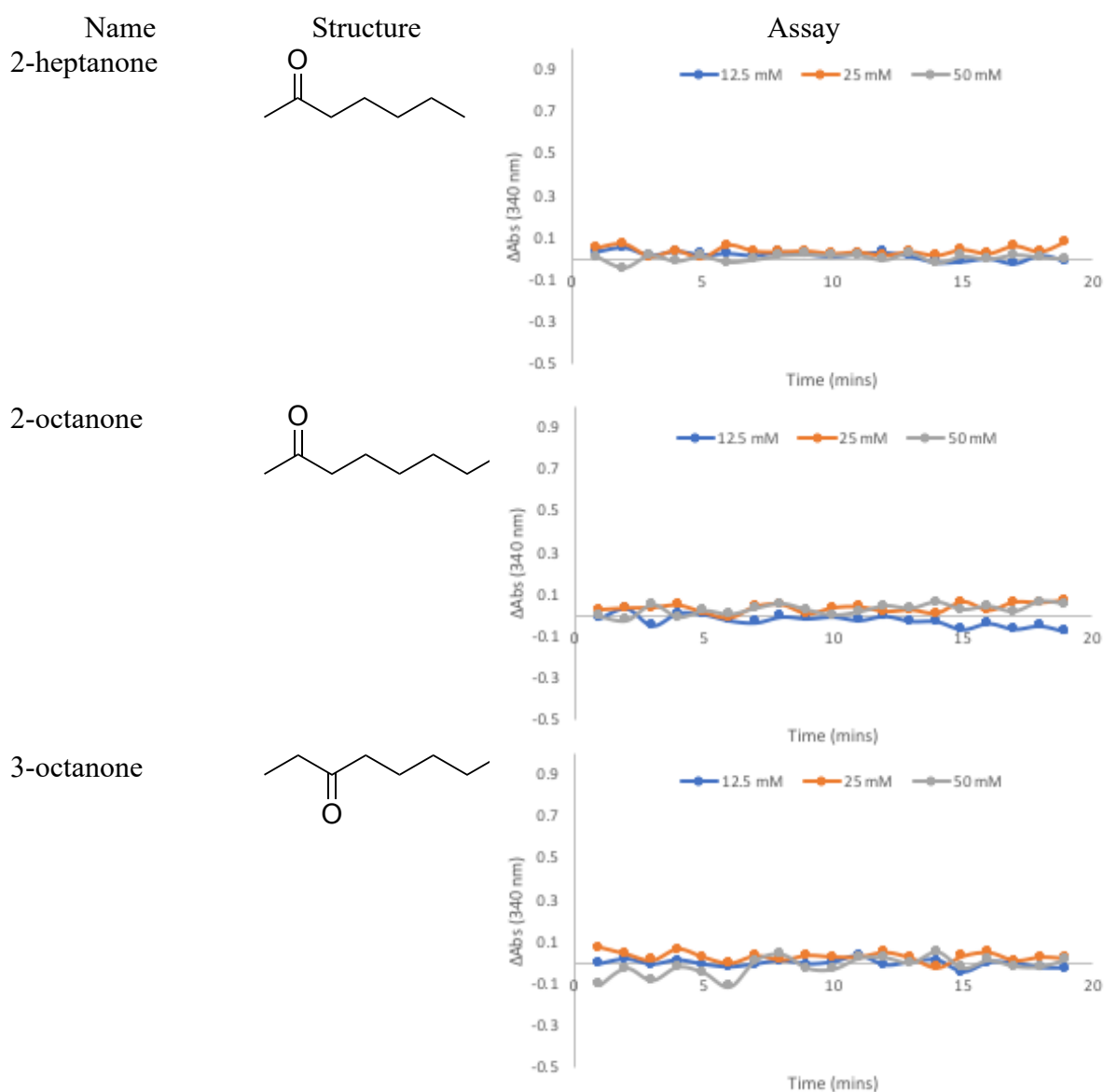
Cyclohexanone
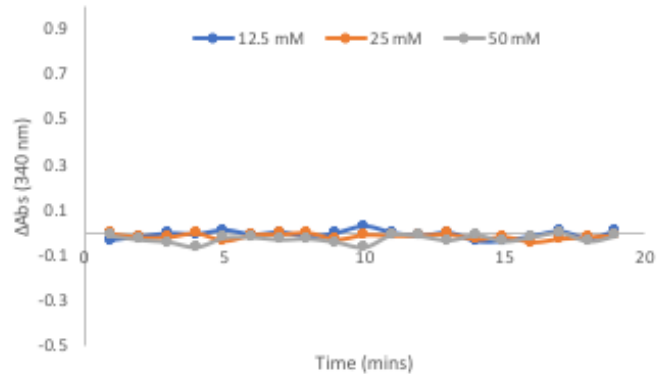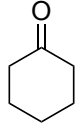
Acetophenone
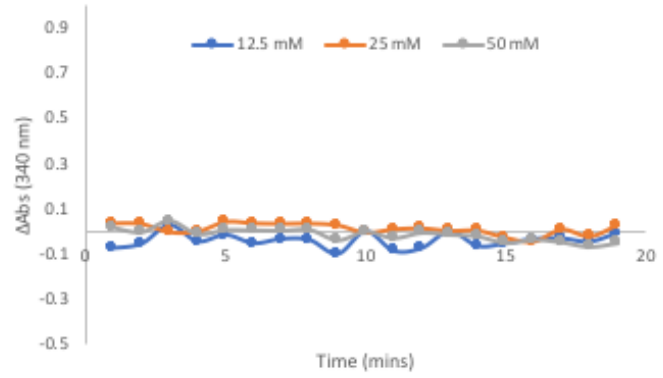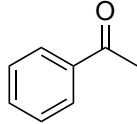
Propiophenone
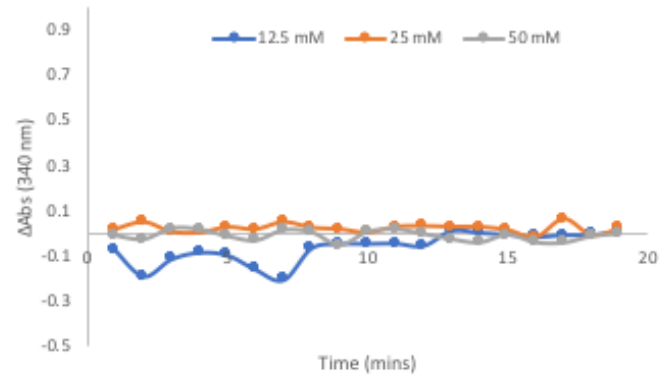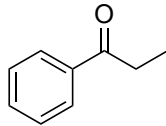
3-oxopentanoic
acid methyl ester
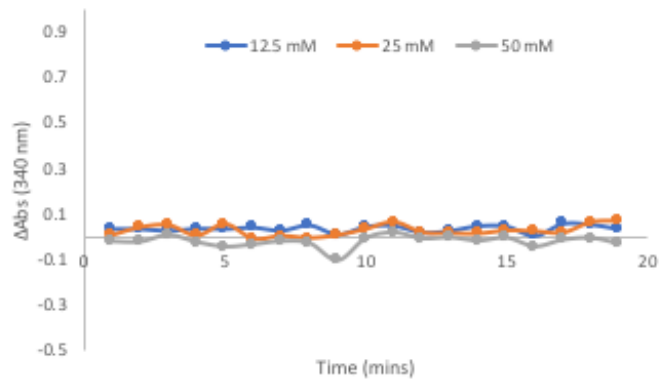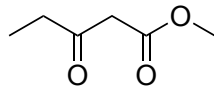
2,4-dioxopentanoic acid ethyl ester
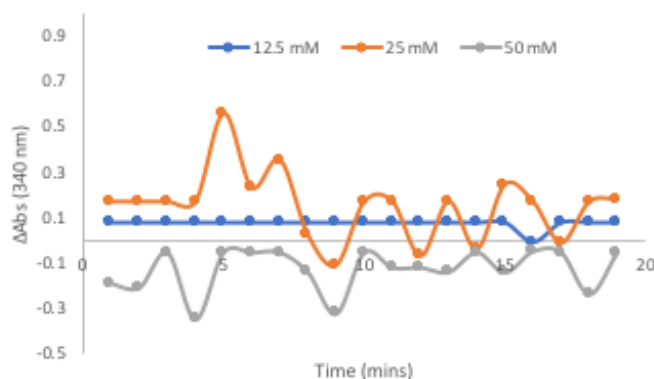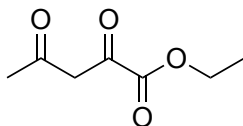
Table 15: NADPH-linked assays of (His)$_6$-AntM with the indicated, varying ketone substrates. Assays were performed by incubation of the required cofactor, substrate, and protein, and monitoring the loss of absorbance at $\lambda = 340$ nm. The assay data shown displays the corrected absorbance (corrected for background absorbance and absorbance in control wells) for three different substrate concentrations. No reduction, or NADPH loss, was seen for any of the substrates above background levels.

As a large proportion of the substrates tested were selected for structural simplicity and commercial availability, and contained little resemblance to those intermediates or products involved in the biosynthesis of antimycin, it was reasonable to expect and observe minimal AntM activity towards these substrates. However, a handful of the substrates tested (3-oxopentanoic acid methyl ester and 2,4-dioxopentanoic acid ethyl ester), contained structural motifs similar to those observed in late-stage antimycin intermediates, including a β-keto group where reduction would occur, as well as a neighbouring carbonyl group. The inability of AntM to utilise these compounds suggests that it may be more selective for its substrate than *cis*-encoded KR domains characterised using this approach.

4.3.6 Testing of antimycin intermediate mimic compounds for AntM binding

The total synthesis of cyclised antimycins has been achieved (Janetzko and Batey, 2014), and a pre-cyclised linear form of antimycin would likely proceed through a similar synthetic route. Due to the length and number of steps involved in the synthesis, a simpler mimic compound was designed, involving three reagents joined by peptide coupling reactions. The three constituent parts of the mimic were gamma-aminobutyric acid (GABA), L-Thr, and benzoic acid, with either a free carboxylic acid at the C-terminus or an N-acetylcysteamine (SNAC) group at the C-terminus. The identity of the synthesised substrates was confirmed by LC-HRMS, followed by testing in the previously developed ITC assay with the purified, recombinant AntM protein. (His)$_6$-AntM, preincubated with a molar equivalent or excess of

NADPH, was assayed separately with a 10-fold molar excess of the two antimycin mimic substrates. Neither of the substrates, however, displayed a binding response with AntM, and no corresponding signals were detected (Figure 4.9). A heat signal corresponding to a simple heat of dilution was observed, indicating no interaction between the protein and the small molecule. AntM did not bind either of these substrates, which could be due to the lack of a C8 ketone equivalent in the substrates or simply a further indication of its strict substrate specificity. An additional explanation could be AntM requires contact with an ACP domain to bind its cognate substrate.
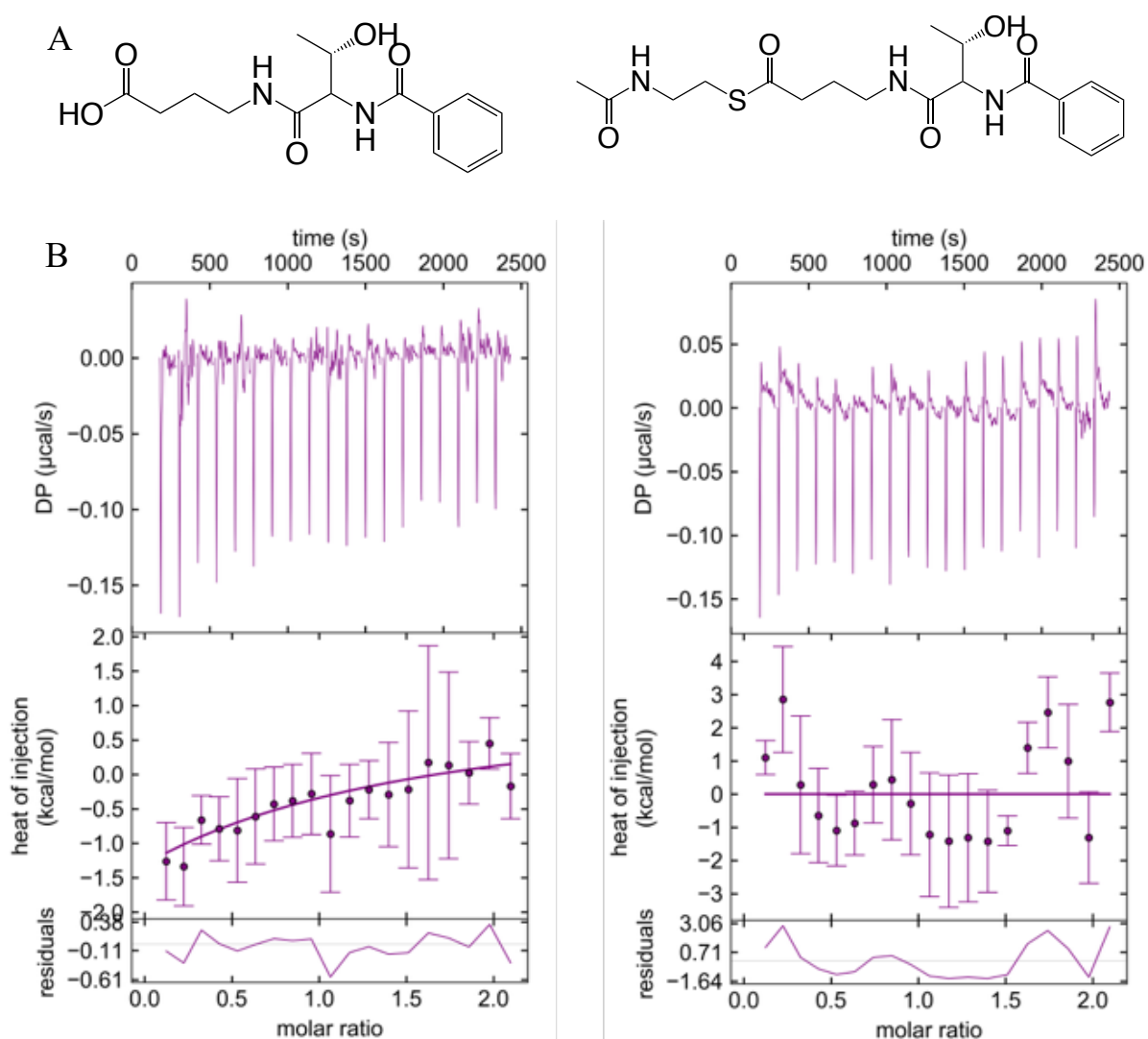
Figure 4.9: A) Chemical structures of synthesised antimycin mimic compounds. The basic mimic compound (left) and the SNAC-appended mimic (right) were both tested in an ITC assay with AntM. B) ITC assays of AntM with the antimycin mimic compounds. No reliable binding interaction was observed in each case.

4.3.7 Crystal trials of AntM

All previous *in vivo* and *in vitro* assays suggested that AntM would most likely only be functional on the terminal linear antimycin intermediate, whilst appended to the ACP of the PKS AntD. In order to provide further molecular detail and insight into this intriguing possibility, crystal trials were set up with the purified (His)$_6$-AntM protein, either in its *apo* form or complexed with reduction cofactors NADPH or NADP$^+$. Crystal trays, incubated at room temperature for multiple weeks, were imaged periodically and the images checked for protein crystal formation. Sizeable protein crystals were formed under numerous crystallisation conditions (Figure 4.10), as shown by clear and characteristic absorbance at $\lambda = 280$ nm, initially indicating the propensity with which the protein formed crystals, both in the *apo* form and when complexed with cofactors. Representative crystals, and those instinctively more likely to yield X-ray diffraction patterns, were selected and picked for storage and subsequent transport to the Diamond Light Source, for analysis in the X-ray beamline.
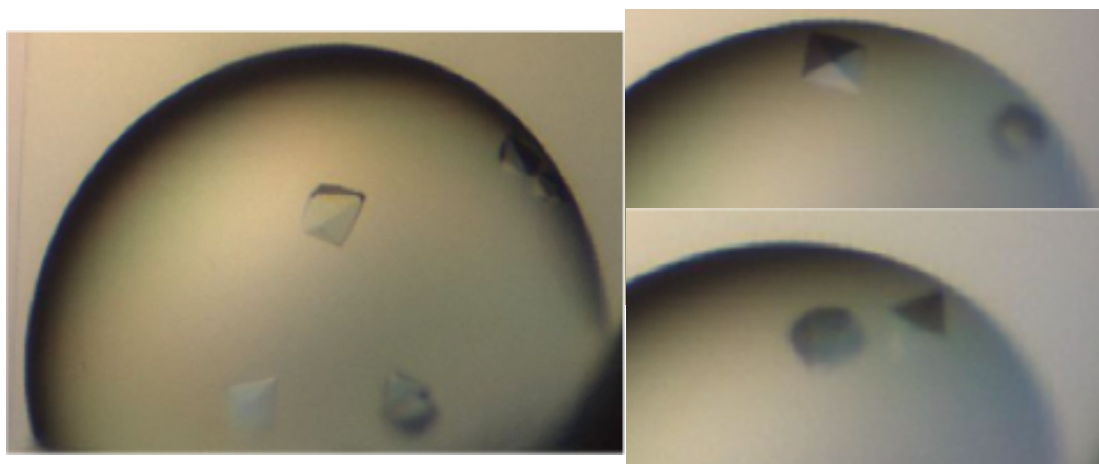
Figure 4.10: Protein crystals of (His)$_6$-AntM, formed in the presence of NADPH (left), and of the *apo*-protein (right). The crystal conditions were: 0.1 M citrate phosphate buffer, pH 4.2, 0.2 M NaCl, and 10% w/v PEG 3K (right); and 0.1 M Tris, pH 8.5, 5% w/v PEG 8K, 20% v/v PEG 300, 10% v/v glycerol (left).

4.3.8 The 1.7 Å crystal structure of AntM with NADPH

Multiple crystals were obtained and a representative sample of crystals were shot to obtain diffraction patterns, and subsequent protein structure data. A crystal obtained from the screen of AntM incubated with NADPH diffracted to 1.7 Å, and a full data set, consisting of 33683 reflections, was subsequently acquired. The structure of AntM in complex with NADPH was solved by molecular replacement using the UrdMred ketoreductase (PDB ID: 4osp) (Patrikainen et al., 2014) from the urdamycin biosynthetic pathway as the search model and

subsequent refinement, by utilisation of CCP4 and the CCP4i2 graphical user interface (specifically the AIMLESS, CHAINSAW, PHASER, COOT, and REFMAC5 programmes). Data collection and refinement statistics are shown in Table 16.

The AntM protein crystallised with one monomer per unit cell, with each monomer binding one molecule of the NADPH cofactor in the putative protein binding site. The data did not categorically confirm the existence of the homotetrameric species, as visualised by gel filtration chromatography in the initial purification of the His-tagged protein. Further gel filtration studies were therefore carried out with varying concentrations and amounts of protein utilised each time. In each size exclusion chromatography run, the protein eluted much earlier than expected for a monomeric protein with a mass of 29988.70 Da, rather being collected at an elution volume much more rationalisable with the mass of a homotetrameric species (elution at 189.19 mL suggesting ~ 120 kDa complex). Comparison of AntM with other biosynthetic pathway β-ketoreductase domains, including the UrdMred protein used for molecular replacement in the initial determination of the protein structure, showed that the majority of proteins with high sequence and structural similarity were homotetrameric proteins, as shown by various biochemical and biophysical assays, and further indicating that AntM likely exists in a functional tetrameric state.

A

B
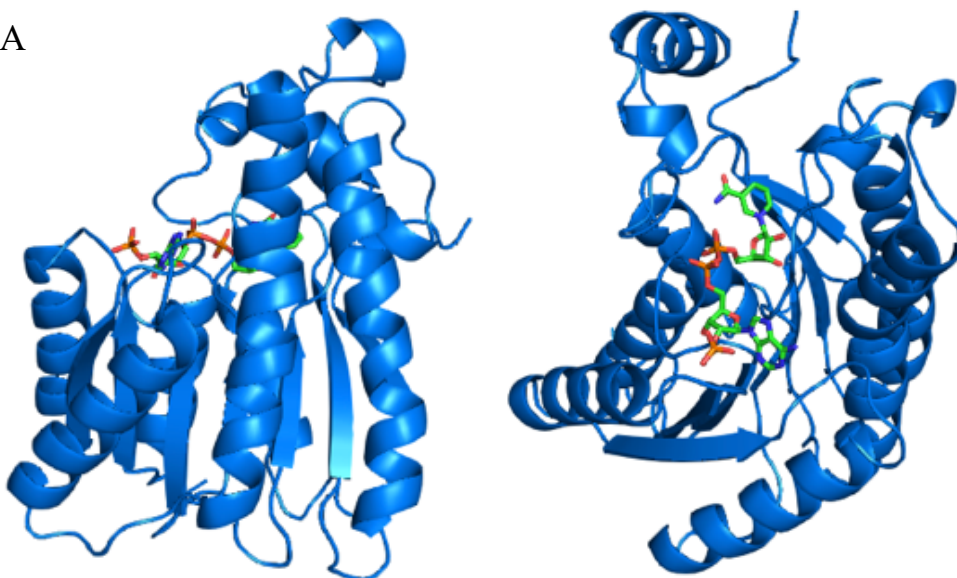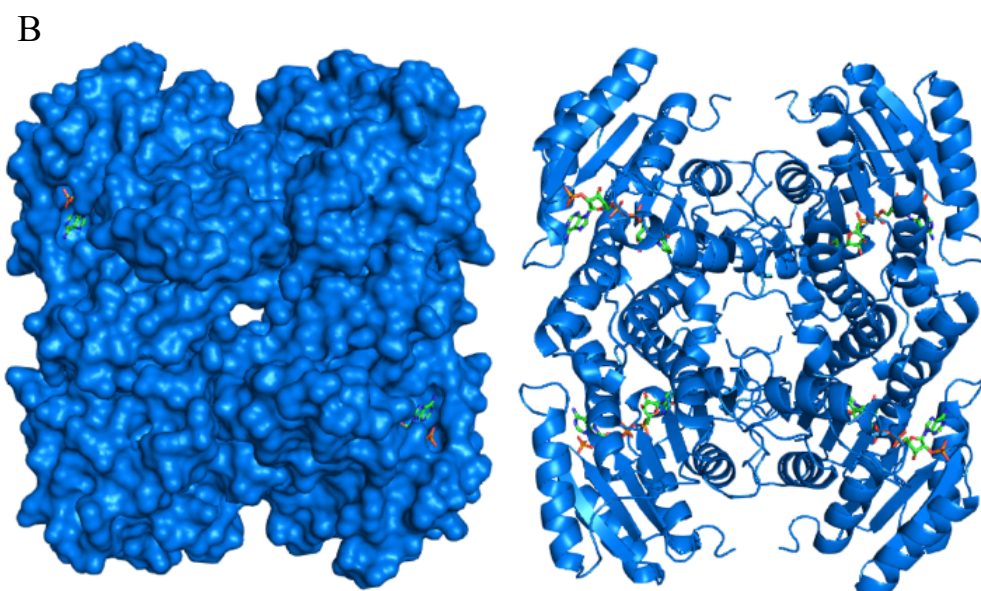
Figure 4.11: A) The overall structure of AntM, complexed with NADPH (shown as sticks), in the monomeric form. A side view (left) and top view (right) are shown. B) AntM, complexed with NADPH, in the tetrameric form. Surface (left) and ribbon (right) structures are shown. The monomeric form of the structure was directly solved (one monomer of AntM per unit cell, solved by molecular replacement) and the theoretical tetrameric structure was formed from analysis and tetramer construction around the protein symmetry axis.

	AntM-Apo	*AntM-NADPH*
Data collection		
Space group	P6₂22	P6₂22
Cell dimensions		
a, b, c (Å)	87.9, 87.9, 116.2	92.2, 92.2, 113.8
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Resolution (Å)	63.69 – 2.10 (2.16 – 2.10) *	65.36-1.70 (1.73-1.70) *
R_{merge}	0.155 (1.07)	0.063 (1.77)
R_{pim}	0.050 (0.35)	0.014 (0.44)
CC(1/2)	1 (0.75)	1.00 (0.96)
$I / \sigma I$	7.8 (1.5)	24.7 (1.7)
Completeness (%)	99.9 (98.7)	100.0 (100.0)
Multiplicity	9.0 (9.1)	21.2 (17.2)
Refinement		
Resolution (Å)	93.77-2.10	65.45-1.70
No. reflections	16855	33683
R_{work} / R_{free}	0.213/0.273	0.199/0.224

B-factors (Å^2)		
Protein	47.73	42.41
NADPH	n/a	42.64
Water	48.19	44.37
R.m.s. deviations		
Bond lengths (Å)	0.0094	0.0101
Bond angles (°)	1.649	1.656
PDB ID	7NM7	7NM8

*Values in parentheses are for highest-resolution shell.

Table 16: Data collection and refinement parameters for solved AntM structures.

The overall structure of AntM is typical of many SDR family proteins (IPR002347), with Rossmann fold architecture and a clear binding site for the dinucleotide reduction cofactor, NADPH. The monomer of AntM is composed of multiple alternating α-helices and β-strands, hydrogen bonded to form a β-sheet (Figure 4.11A). As in classical Rossmann fold containing enzymes, AntM comprises six β-strands, each alternating with an α-helix secondary structural motif. The core of the protein is therefore composed primarily of three longitudinal layers, with the central layer formed solely of β-strands and the external layers of the protein formed from helical motifs. External to the core Rossmann fold architecture and towards the C-terminus of the protein, a further β-strand and multiple short helices, as well as sections of amino acid loops, form the remainder of the protein. A minority of the loops connecting these external domains contained areas of poorly defined electron density, either due to the distal nature of the loops, artefacts of crystal packing and formation, or the possibility that the loops may be intrinsically disordered in the absence of substrate or other required binding partners. The structure of the tetramer of AntM is formed from four monomers of the protein, all of which are symmetry related (Figure 4.11B). Oligomer formation in the transverse direction seemingly occurs through interaction of two of the helical motifs involved in Rossmann fold formation on one monomer, with the same helices on another. The second interaction interface is formed through interaction of external loops as well as the seventh β-strand, not thought to participate in the formation of the Rossmann fold tertiary structure of the protein. Every monomer involved in tetramer formation is capable of binding a molecule of the cofactor NADPH, and therefore likely to be capable of catalysing β-ketoreduction during the biosynthesis of antimycins.

The position of NADPH is clearly visible within the crystal structure where it is coordinated parallel to a long channel by the amino acid stretch TGGSRGIG, a quintessential Gly-rich NADPH binding motif (Figure 4.12) (Hanukoglu and Gutfinger, 1989). The cofactor is also within the proximity of a catalytic triad; S178, Y191, K195, which is a conserved set of amino acids capable of carrying out β-ketoreduction in SDR family enzymes (Filling et al., 2002). Within the global structure of the protein, the NADPH cofactor molecule sits in the middle layer, above the set of parallel β-strands and between the two sets of external α-helices, amongst a series of loops connecting the alternating secondary structural motifs composing the nucleotide binding Rossmann fold. The cofactor coordinating residues, TGGSRGIG, are part of a loop connecting the first β-strand with α-helix 1, where the adenine, ribose, and phosphates are accommodated through a combination of hydrogen bonds and hydrophobic interactions, with the numerous Gly residues providing a complete lack of steric hindrance and chain flexibility to allow binding of the large cofactor. The residues forming the catalytic triad sit on the opposite side of the binding pocket to the NADPH coordinating residues and in close proximity to the reducing face of NADPH and the hydrogen that would be abstracted to catalyse reduction. The three amino acids, S178, Y191, K195, all reside within an average of 5 – 6 Å from the site of proton abstraction and within the range of catalytic activity.
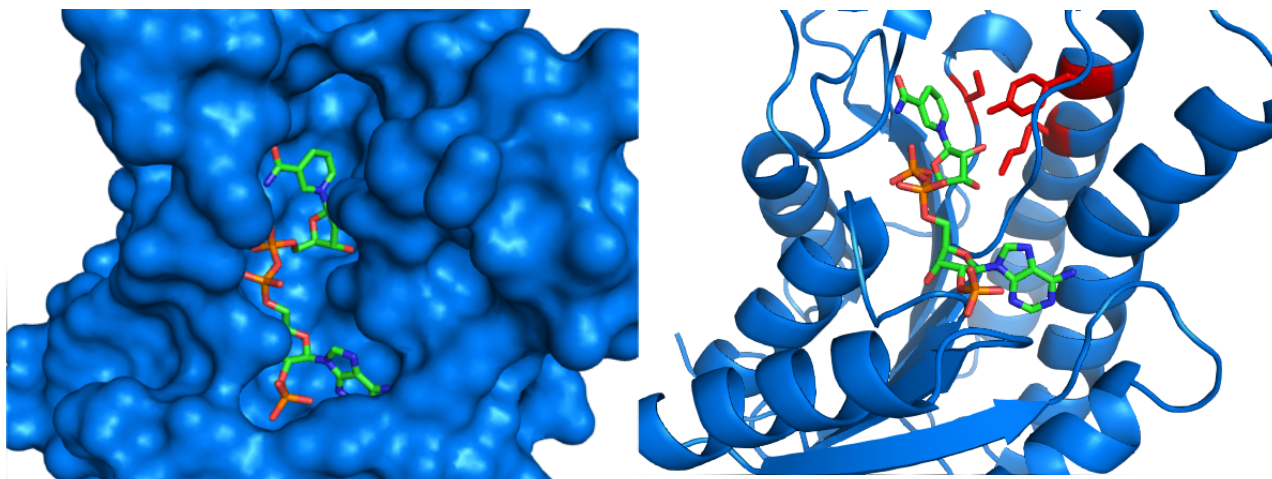
Figure 4.12: Structure of the extended binding pocket of AntM (left) and the accommodation of the NADPH cofactor, shown as sticks in both the surface and ribbon structures and coloured by atom. The location of the three residues (Ser, Tyr, Lys) required for AntM catalytic activity in relation to the cofactor (right, coloured in red).

The binding pocket itself is extended, primarily to accommodate the large reduction cofactor, but also extends slightly deeper into the protein when compared to other members of the same

SDR family proteins. The substrate usually binds above the NADPH molecule, in close proximity to the reducing face of the cofactor, in an open part of the extended binding groove. In the NADPH-complexed AntM crystal structure, the region above NADPH is mostly occluded by two bulky, aromatic residues at the entrance to the site (Figure 4.13). These gatekeeper residues (Y185, F223) would seemingly abrogate binding of any substrate in the usually occupied binding site, above the reducing face of the NADPH cofactor. The nature of the active site binding cleft, with it forming a deeper pocket in the protein could accommodate a larger and extended substrate, binding in a manner parallel to the cofactor. Without a structure of the tertiary complex, with the protein in a bound state with both cofactor and substrate, the nature and position of the substrate would be more difficult to categorically ascertain. The crystal structure of AntM complexed with NADPH, however, has given valuable insight into the structural variations of AntM compared with other β-ketoreductases, and raises intriguing possibilities regarding the essentiality of AntM in the biosynthesis of antimycins.
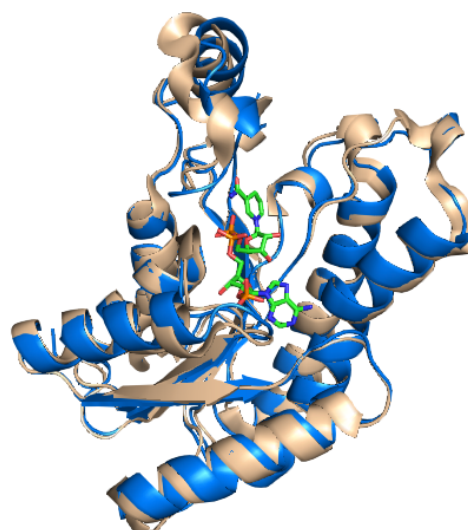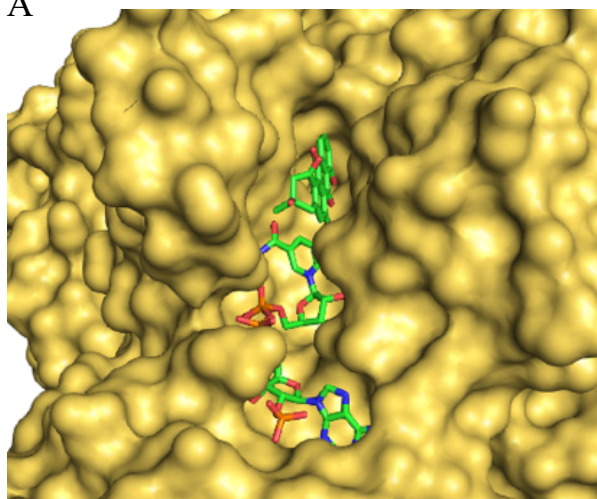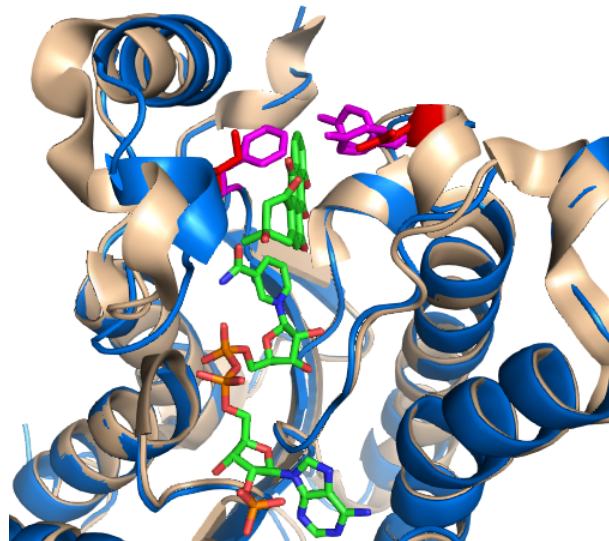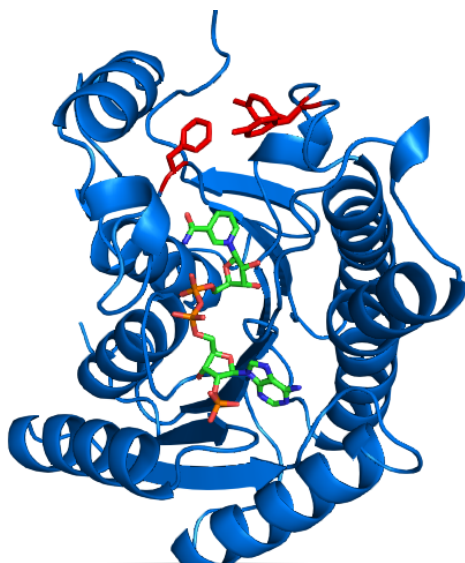
Figure 4.13: A) The surface structure of the UrdMred active site, with bound $NADP^+$ and the substrate analogue, rabelomycin (both shown as sticks and coloured by atom) (left); and an alignment of the crystal structures of UrdMred (gold) and AntM (blue) (right), indicating the structural similarities. NADPH is shown in the active site as a stick structure and colured by atom. B) The position of the two bulky gatekeeper residues (Phe and Tyr, coloured in red) in the overall structure of AntM, and in relation to the reduction cofactor (stick structure coloured by atom) (left). Tyr185 may adopt two distinct structural conformations, as judged by the presence of two areas of electron density within the structural maps. Alignment of UrdMred (gold) and AntM (blue) (right), highlighting the large AntM gatekeeper residues (pink), and the corresponding residues in UrdMred (red). NADPH and the substrate analogue, rabelomycin, are shown in the active site as stick structures and coloured by atom.

4.3.9 The 2.1 Å crystal structure of *apo*-AntM

A crystal obtained from the screen of AntM alone was shot and diffracted to 2.1 Å. A full data set, consisting of 16085 reflections, was acquired and the structure of *apo*-AntM was solved by molecular replacement. The previously solved structure of AntM in complex with NADPH was utilised as the search model, and subsequent refinement was carried out in CCP4 and the CCP4i2 graphical user interface with the AIMLESS, PHASER, COOT, and REFMAC5 programmes.
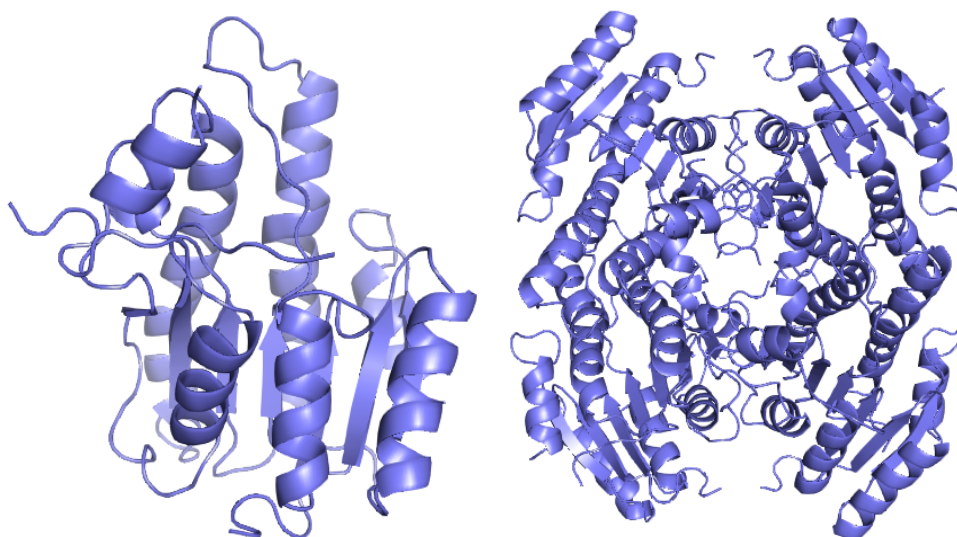
Figure 4.14: The structure of the AntM *apo* enzyme in its monomeric (left) and tetrameric (right) forms. The structures remain similar to the NADPH complexed forms; NADPH binding has little effect on the structure or overall topology of the enzyme.

The overall structure of AntM in its *apo* form (Figure 4.14) closely resembled its cofactor complexed form, with an immediately observable Rossmann fold tertiary structure of a central parallel β-sheet, sandwiched between external sets of α-helices. A clearly visible, solvent exposed, deep binding groove is also present at the surface of the protein, and encompasses the characterised NADPH binding motif as well as the established and conserved catalytic triad. Alignment of the *apo-* and NADPH complexed forms of AntM showed very little variation in the positioning of catalytically important residues and in overall secondary structural motifs. The only observable deviation between the two structures occurs in the final central β-strand, not directly a composite part of the nicotinamide binding super secondary structural motif. The remainder of the protein aligned extremely well with the previously solved complexed structure, potentially indicating that AntM, in its *apo* form, exists in a state capable of binding NADPH with minimal perturbations in amino acid residues and secondary structural motifs. Other residues potentially involved in catalysis or substrate binding, such as the gatekeeper residues above the reducing face of NADPH, were also unmoved between the two structures. The near-perfect alignment of the *apo-* and NADPH complexed forms of AntM, whilst not completely unexpected, could have occurred as an artefact of using the initially solved structure of AntM with NADPH as the search model to solve the phases of the *apo-* form of the protein. This could have forced the protein to adopt a similar structure, however, the binding of substrate rather than cofactor would be more likely to cause larger conformational changes in the protein and lead to variations in the overall protein morphology.

4.3.10 Computational docking of potential biosynthetic intermediates

The nature and dimensions of the NADPH and substrate binding channel as well as the orientation of the reducing face of NADPH, the catalytic triad, and other important residues within the binding pocket, imply binding of a potential linear molecule as a substrate. The hypothesis that AntM is functional upon the final, linear biosynthetic intermediate in antimycin biosynthesis, arising from the *in vitro* and *in vivo* essentiality of the ketoreductase for antimycin production, also suggests that AntM would bind a linear intermediate, likely when that intermediate is held on the terminal ACP of the assembly line. This linear intermediate was inaccessible, so docking studies were used to computationally assess this possibility. We modelled the terminal antimycin biosynthetic intermediate, containing a terminal SNAC group mimicking the phosphopantetheinyl extension required in ACP binding, as well as the unreduced, cyclized form of antimycin, into the putative AntM binding pocket with Autodock

Vina (Trott and Olson, 2010). A grid of 16 x 32 x 16 Å was defined, centred on the binding pocket and the position of the bound cofactor, and substrates were defined followed by computational assessment of the suitability of each provided substrate for the protein and the defined region. Output docking poses were manually assessed for suitability of binding and the potential for AntM catalytic activity on the selected substrate.

The initial modelling of the terminal linear biosynthetic intermediate and the cyclised and assembly line released form of antimycin both showed that the molecules could feasibly fit within the AntM binding pocket, alongside the NADPH cofactor. This was mostly due to the extended and deeper nature of the binding pocket itself, which gave ample space for binding compounds above and parallel to the cofactor. Average values for the dimensions of the pocket, estimated from averaging multiple measurements taken from different points within the pocket, were a binding cleft length of ~ 24.4 Å, a width of ~ 13.7 Å, and a depth of ~ 13.1 Å (approximate surface area of 102.4 Å^2 and volume of 4379 Å^3, assuming a cuboidal shape for the binding site). The optimal binding pose for the linear biosynthetic intermediate, however, demonstrated improved and more optimal potential binding parameters when compared to the cyclised antimycin product.
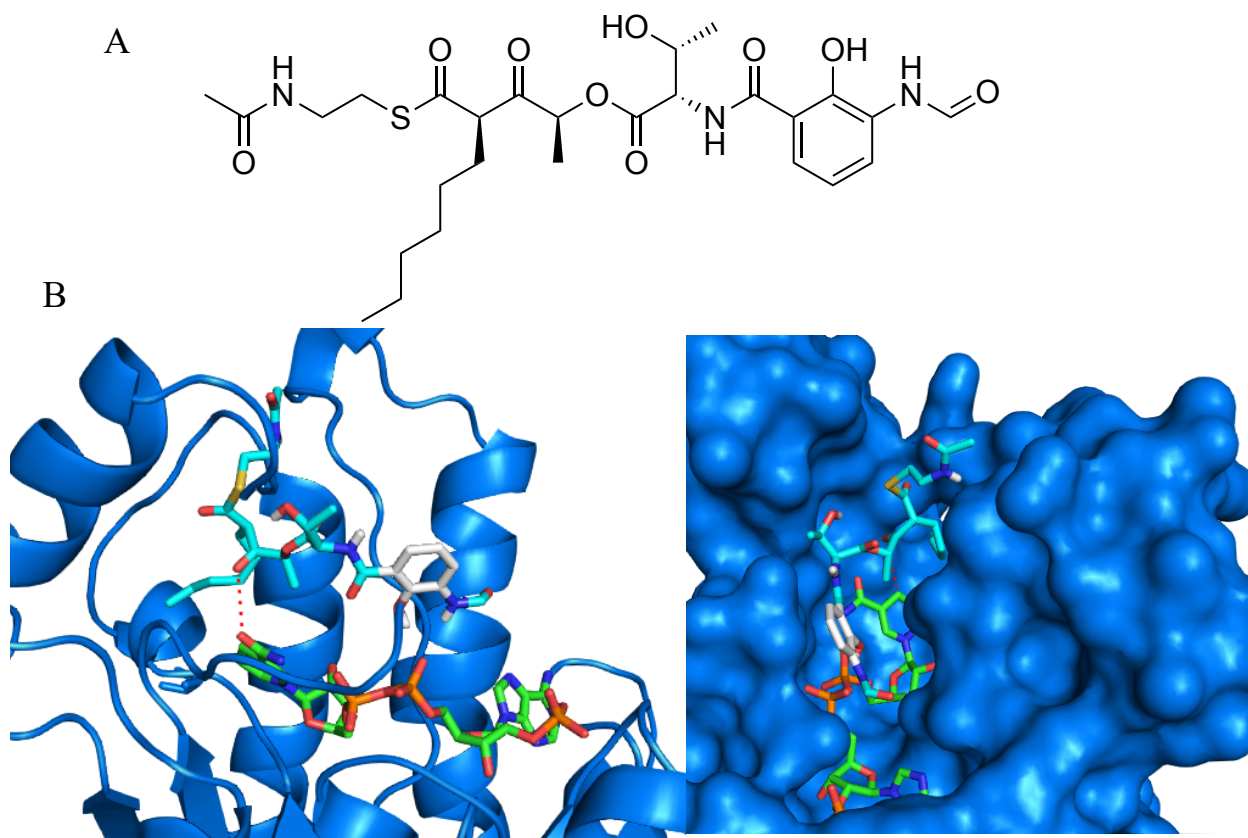
Figure 4.15: A) The chemical structure of the docked linear antimycin intermediate. B) The structure of AntM complexed with NADPH, and a linear antimycin intermediate docked into the active site in ribbon (left) and top surface (right) views. The dotted red line indicates the position of the reducing face of NADPH and the ketone that is the site of reductive activity.

The terminal linear biosynthetic intermediate from the antimycin pathway placed the C8 carbonyl, the location of AntM ketoreduction, 3.4 Å away from the reducing face of NADPH within the AntM binding pocket (Figure 4.15). The unreduced, cyclised product, however, was a distance of 5.9 Å from carbonyl to the hydride of the reduction cofactor (Figure 4.16). The approximated distance measurement of the linear intermediate from NADPH is generally consistent with that observed from other NADPH-utilising enzymes and systems (Warkentin et al., 2001; Le et al., 2017). The sulfur atom, modelled as part of the terminus of the linear intermediate and which would also be the atom at which thioesterification would occur to append the intermediate to AntD[ACP] by a ~ 20 Å phosphopantetheinyl group, was also outwardly positioned towards an external groove on the protein, which could feasibly accommodate the prosthetic group and potentially mediate interaction with the larger assembly line. The Gibbs free energy of each potential interaction, calculated by Autodock Vina as part of the docking process, yielded an increased negative value for the linear intermediate when compared to the unreduced, cyclised product, showing that AntM favoured binding of the biosynthetic intermediate (Linear: -8.3 kcal/mol, cyclised -7.5 kcal/mol). All of these parameters, assessed both manually and computationally, indicated that AntM likely binds and reduces the terminal intermediate of antimycin biosynthesis, rather than the thioesterease released product.
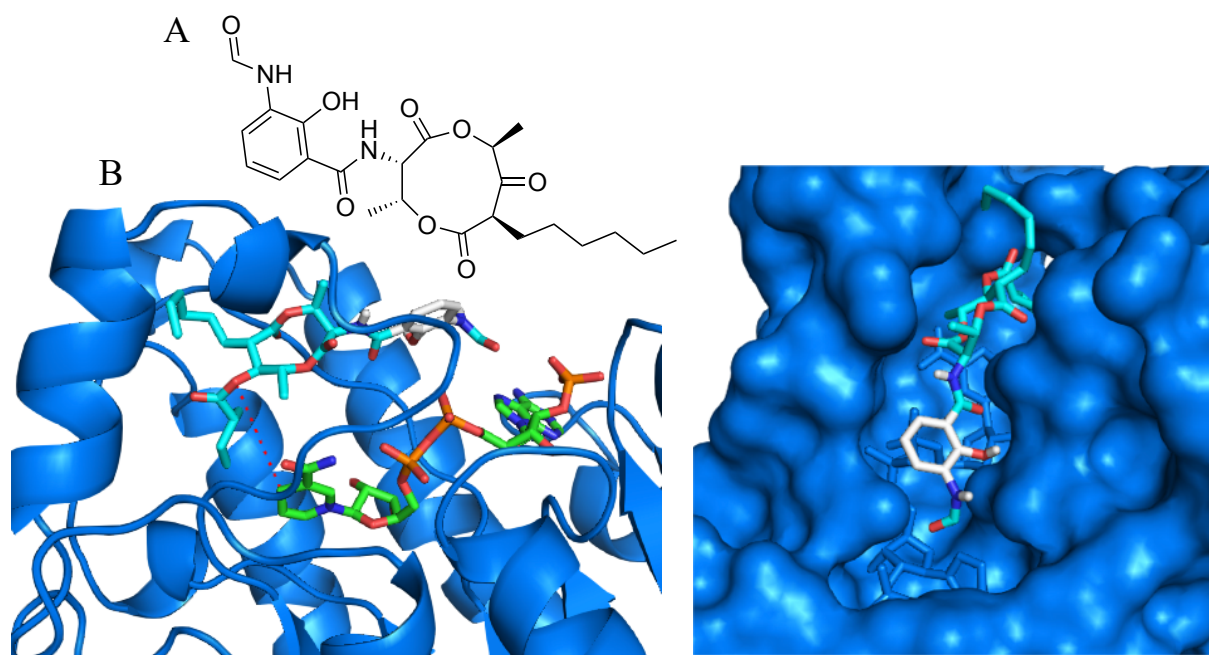
Figure 4.16: A) The chemical structure of the cyclised antimycin product used in docking studies. B) Docking of the cyclised, unreduced antimycin product into the AntM active site in ribbon (left) and surface (right) views. The dotted red line indicates the position of the reducing face of the NADPH cofactor and the site of ketoreduction.

4.3.11 Mutational analysis of the AntM binding pocket

The lack of AntM activity towards any simple ketone substrates *in vitro*, as well as the inability to detect a binding response with compound mimics, suggested that AntM activity towards its cognate substrate from the antimycin biosynthetic pathway would have to be probed in a more indirect manner. Structural analysis of AntM revealed the closest structural matches were KRs associated with type II PKS systems; specifically, UrdMred from *S. fradiae* (PDB entry 4OSP) and LanV from *S. cyanogenus* (PDB entry 4OSO), which are urdamycin and landomycin C6-ketoreductases, respectively (Paananen et al., 2013; Patrikainen et al., 2014). UrdMred shares 52.3% sequence identity with AntM, whilst LanV is 49% identical, however, the protomers of these enzymes superpose onto AntM with RMSDs of 0.91 and 0.93 Å, respectively. The structures of both UrdMred and LanV were also determined in the presence of a substrate mimic, rabelomycin, a small planar molecule significantly different to the expected substrate of AntM. This allowed comparison of the respective substrate binding sites from the three enzymes, which revealed interesting differences in the amino acid composition of the active site pocket. Two of the small side chains which form the rabelomycin binding site in UrdMred (Asn154 and Ile192) and LanV (Ser154 and Ser192) have been replaced by significantly

bulkier side chains of Tyr185 and Phe223 in AntM, with the Tyr residue occupying a dual conformation showing some flexibility in the region. These structural differences, however, suggested that AntM may therefore bind its substrate significantly differently compared to these previously studied enzymes.

To investigate the importance of the two residues in controlling access of potential AntM substrates to the binding pocket of the ketoreductase, a battery of pairwise mutant enzymes was created. Each of the residues was mutated to either Ala or Trp, or kept as the WT Tyr or Phe, giving a total of eight mutant proteins ($W^{185}W^{223}$, WA, WF, AW, AA, AF, YW, YA), with the expectation that Trp substitutions would occlude entry of the linear substrate into the binding pocket abrogating activity, whilst the Ala substitutions may open up the pocket without severely affecting activity. The mutant *antM* variants were used to genetically complement the Δ*antM* gene deletion to provide *in vivo* data on the importance of these positions. Mass spectral analysis revealed that mutation of any of the gatekeeper residues to a Trp residue resulted in a severe reduction, or complete attenuation, of antimycin production, as no products were observed in extracts made from cultures of the strains complemented with the Trp-containing *antM* genes (WW, WA, WF, AW, YW) (Figure 4.17). Production of antimycins was observed, however, in the remaining complementation strains, specifically the *antM* deletion strain complemented with $AntM^{Y185A}$, $AntM^{F223A}$, and $AntM^{Y185A-F223A}$ (Figure 4.17). The observation that a Trp residue at either of the gatekeeper positions occludes antimycin production, and therefore also ketoreduction, suggests that these residues play an important role in substrate access or binding within the AntM active site.

To ensure that the observed phenotypes were not the result of effects on cofactor binding, a representative subset of the proteins were overproduced, purified and used in ITC to confirm that NADPH binding had not been disrupted. ITC analysis of $(His)_6$-$AntM^{Y185W}$, $(His)_6$-$AntM^{F223W}$ and $(His)_6$-$AntM^{Y185W-F223W}$ mutant protein binding with NADPH showed that the mutant variants all bound to NADPH equally well as the wild-type protein, with K_D values in the micromolar range (Figure 4.17). These analyses indicated that the purified, mutant proteins were likely folded and active, and importantly uncompromised in NADPH cofactor binding. These data therefore lend strong support to our structural analysis which suggests that these residues are important for the correct positioning of the substrate in the active site to allow antimycin maturation, and that this most likely occurs on the linear substrate whilst it is still bound to the $AntD^{ACP}$.

A

Antimycin A1
Antimycin A2
Antimycin A3
Antimycin A4
Unreduced A1
Unreduced A3

S4_Q5_Cos213_DeltaAntM_pIJ-AntM
S4_Q5_Cos213_DeltaAntM
S4_Q5

YA
YW
AF
AA
AW
WF
WA
WW

B

YW   1.89 µM
WF   14.23 µM
WW   3.43 µM
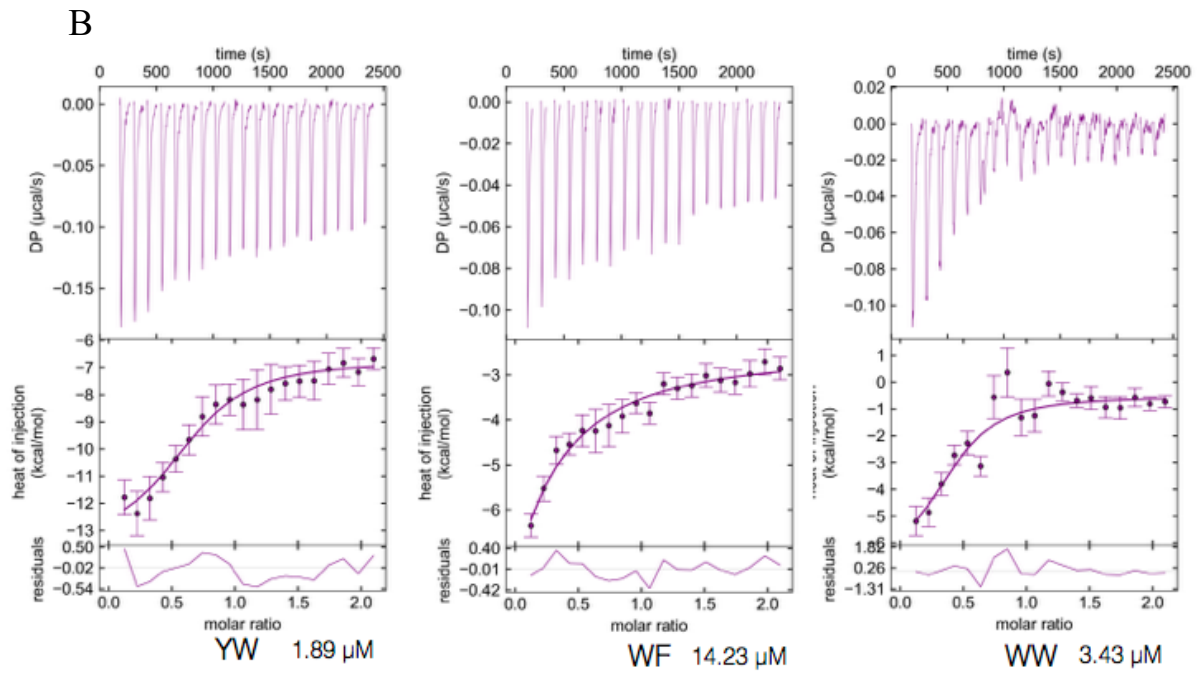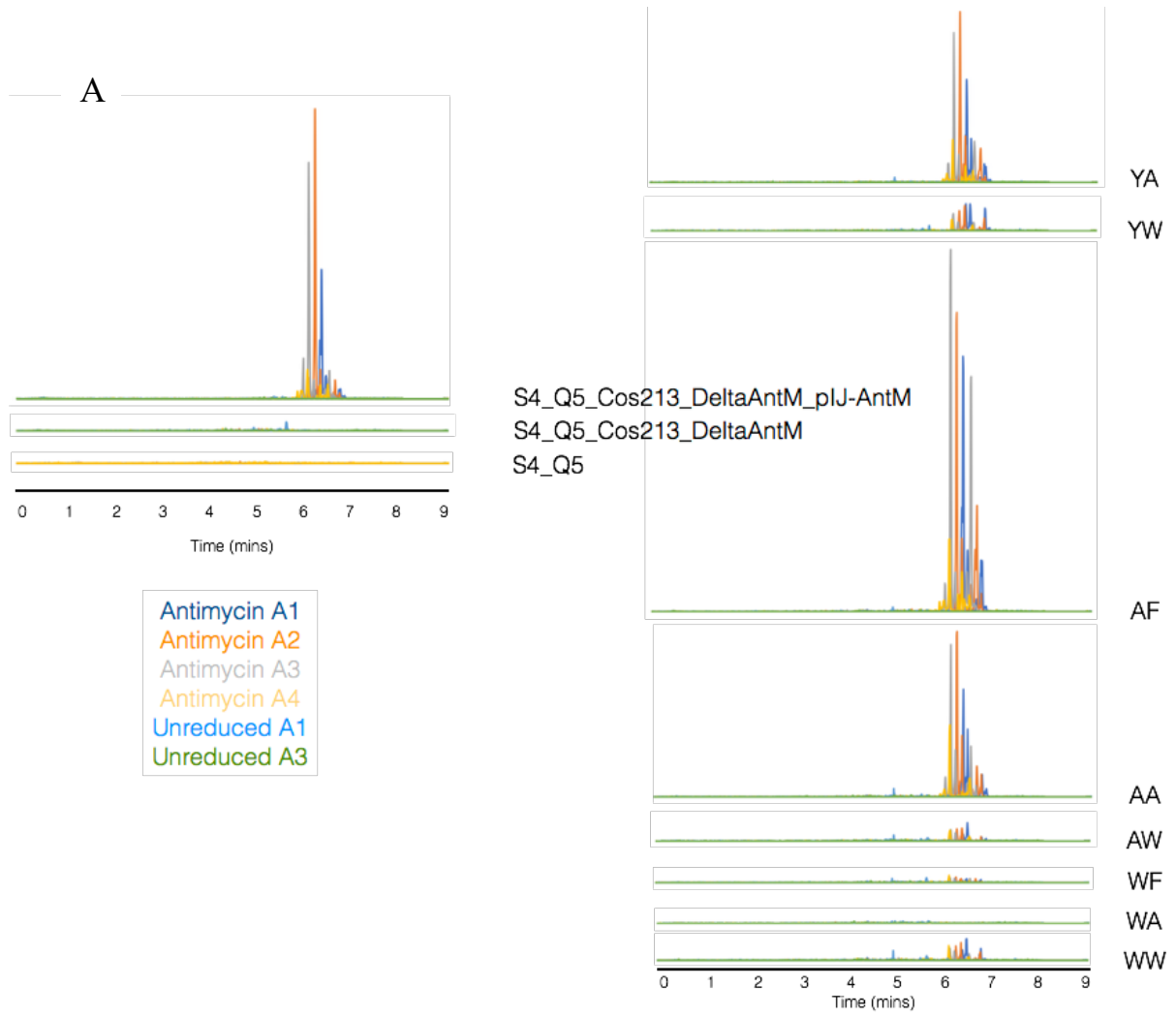
Figure 4.17: A) LC-HRESIMS analysis of chemical extracts prepared from designated antimycin production strains, containing either an *antM* deletion, the WT form of the protein, or one of eight *antM* gatekeeper mutants. Extracted ion chromatograms ([M+H]$^+$, [M+Na]$^+$) corresponding to each of the four commonly produced and observed antimycin variants (A4: $C_{25}H_{34}N_2O_9$, A3: $C_{26}H_{36}N_2O_9$, A2: $C_{27}H_{38}N_2O_9$, A1: $C_{28}H_{40}N_2O_9$), as well as two ketone-containing (or unreduced) antimycins ($C_{23}H_{30}N_2O_8$, $C_{21}H_{26}N_2O_8$) are shown in each case. Exact neutral m/z values used to determine EICs were: 548.2734, 534.2577, 520.2421, 506.2264, 462.2002, 434.1689. The y-axis was set to 2.5 x10^5 for all chromatograms. B) ITC assays of each of the designated AntM mutants with the reduction cofactor, NADPH. Calculated K_D values are noted below each of the isotherm plots.

4.3.12 Bioinformatic analysis of AntM- family ketoreductases

In vivo and *in vitro* data obtained for the catalytic activity of AntM had suggested that the enzyme was most likely to be functional on a linear biosynthetic intermediate, whilst tethered to final ACP of the PKS AntD. Whilst ketoreductase function on ACP tethered intermediates has been observed for type II PKS systems, it was yet to be observed for functionality of a standalone enzyme on an NRPS or type I PKS assembly line. In order to evaluate how widespread KRs like AntM were and to understand their relationship to cis-encoded KR domains, a phylogenetic analysis was performed. Initial bioinformatic analyses were carried out using the Minimum Information about a Biosynthetic Gene (MIBiG) Cluster sequence repository version 2.0 (Kautsar et al., 2020). MIBiG hits for AntM were identified by carrying out a Blastp search against the database, whilst also retrieving a set of *cis*-encoded KRs from a wide variety of systems with varied biosynthetic rationale. The representative set of ketoreductase domains involved in secondary metabolism of specialised metabolites were aligned using default parameters in MUSCLE (Edgar, 2004). The resulting alignment was used to infer an approximate maximum likelihood phylogeny and a phylogenetic tree. The majority of the bioinformatic analyses were carried out by Dr Ryan Seipke.
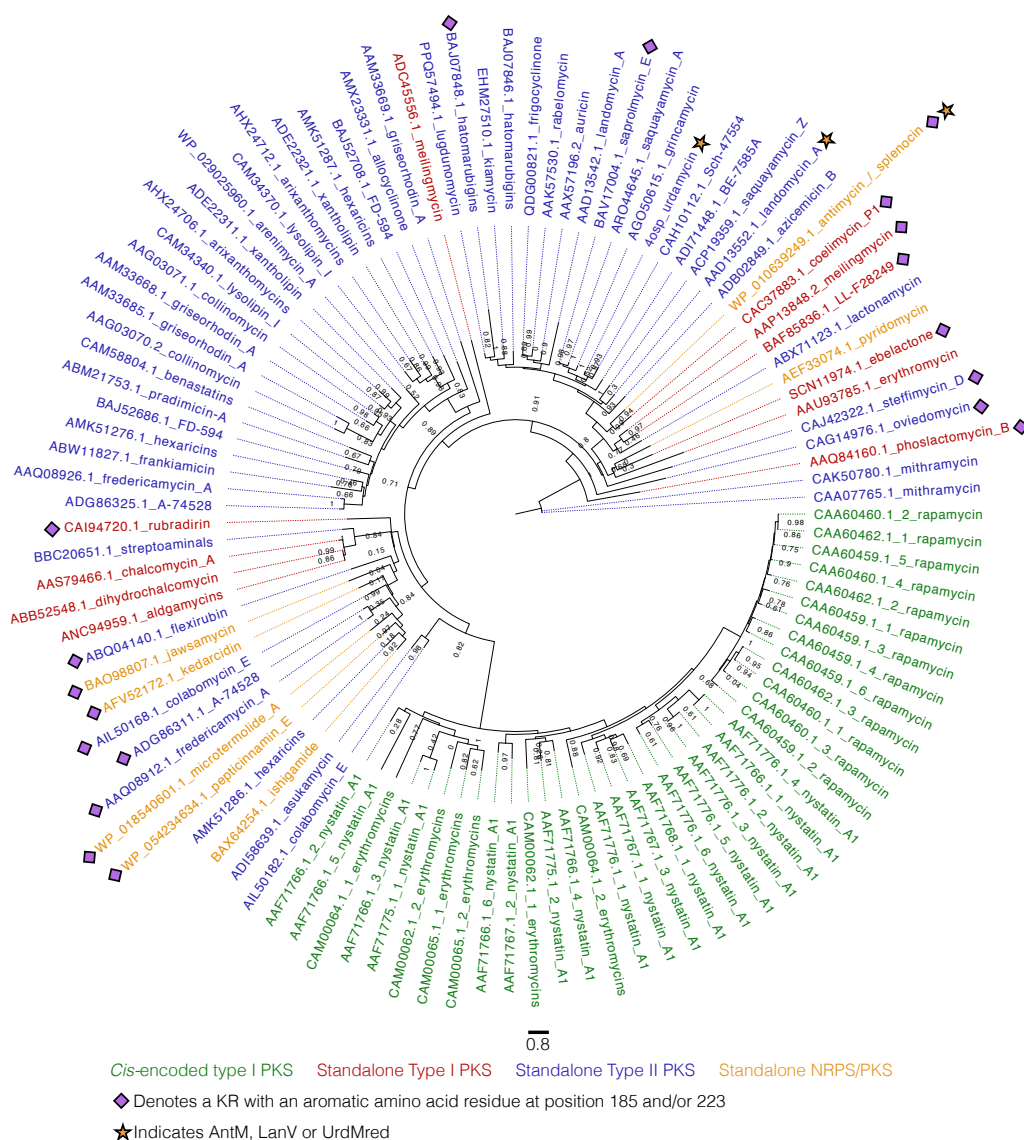
Figure 4.18: Phylogenetic tree showing the relatedness of AntM to a variety of KR domains in the MIBiG database. Two predominant clades are observable; those KRs that are *cis*-encoded in type I PKS assembly lines; and those that are part of other PKS assembly lines, with varied biosynthetic logic. The position of AntM within the tree is indicated, and the enzyme most closely resembles standalone KRs from type II PKS systems, despite having functionality reminiscent of *cis*-encoded type I PKS KR.

The resulting phylogenetic tree surprisingly indicated that AntM was only distantly related to *cis*-encoded KRs from assembly line systems and was generally more closely aligned with KRs from type II PKS systems (Figure 4.18). AntM also forms a small, distinct clade with KRs from different pathways, which also belong to different biosynthetic classes: SgvK (griseovirdin pathway, NRPS/*trans*-AT PKS), Mcl38 (merochlorin pathway, terpene/type II

PKS), and CpkI (cryptic polyketide/coelimycin pathway, type I PKS) (Kaysser et al., 2012; Gomez-Escribano et al., 2012; Xie et al., 2012). There is no proposed role for these KRs within their respective biosynthetic pathways, however similarity to AntM and biosynthetic origins could suggest that these KRs also act concomitantly with biosynthesis in the formation of their cognate natural products. These results suggested that AntM was atypical in terms of functionality in the context of NRPS or type I PKS biosynthetic systems and more closely resembled KRs from distinct type II PKS pathways, providing more evidence for AntM catalytic activity occurring concomitantly with biosynthesis. KRs from other natural product biosynthetic systems also showed levels of phylogenetic similarity with AntM, suggesting that other systems have developed intermediate modification enzymes to function during assembly line-based biosynthesis, potentially to maintain a level of control or fidelity over the production of likely toxic compounds. The concomitant functionality strategy could also ensure certain chemical handles are introduced into natural products before release of the compound or to provide a level of late-stage biosynthetic regulation, in order to guarantee production of the correct specialised metabolite.

4.4 Discussions and conclusions

A variety of standalone enzymes exist within non-ribosomal peptide synthetase and polyketide synthase systems, with the vast majority having defined spatiotemporal specificities regarding the catalytic activity of the protein (Walsh et al., 2001). A majority of these standalone enzymes are responsible for the catalytic diversification of intermediate chemical scaffolds, leading to the introduction of varied functional groups and potentially the targeting of novel or useful intracellular targets. Therefore, the use of standalone enzymes can add an extra dimension or level of complexity to NRPS or PKS systems and allow the catalysis of varied chemistries, not normally observed within these assembly line pathways. NRPS, PKS, or hybrid systems generally contain multiple standalone enzymes, catalysing various different reactions and responsible for varied roles within the system, such as the formation and supply of precursors, chemical diversification of substrates or intermediates, as well as other required and important activities, such as the offloading or release of mature compounds from terminal modules of assembly lines (Staunton and Weissman, 2001; Süssmuth and Mainz, 2017). The S-form antimycin BGC contains four genes encoding standalone proteins, functional during antimycin dilactone assembly in substrate supply, intermediate diversification, or post-assembly line tailoring. In addition, the BGC contains eight genes responsible for the production and

tethering of the 3-formamidosalicylate starter unit, which is then supplied to the two subsequent NRPS modules and terminating PKS module. One of the four enzymes involved in dilactone assembly, AntM, has been shown to catalyse ketoreduction at the C8 position of the antimycin dilactone scaffold, which is subsequently further diversified by a discrete acyltransferase catalysing an esterification reaction (Sandy et al., 2012). Interestingly, AntM was seemingly essential to the formation of antimycins and its exclusion precluded not only the formation of mature compounds but also unreduced variants, an unusual observation for standalone, tailoring enzymes.

Here, the essentiality of AntM that was observed *in vitro* was confirmed *in vivo* by gene deletion analysis and LC-HRMS. The fact that no antimycin-type compounds are released from the cognate assembly lines in the absence of AntM suggests that the standalone KR acts upon the terminal ACP-bound biosynthetic intermediate and its reduction is essential for offloading/cyclisation by the AntD TE domain. Analysis of AntM *in vitro* led to further insights into the structure and enzymology of the KR. Much like a variety of other assembly line resident KR domains, NADPH was shown to be the specifically required reduction cofactor, whilst the use of simple ketone substrates in a reduction assay, whilst effective for many *cis*-encoded KR domains in PKS pathways, yielded no assayable activity. The usage of synthesised substrate mimics, mimicking the terminal linear antimycin intermediate, also yielded no response in an ITC assay, showing no enzyme-substrate binding occurred. The lack of activity towards *in vitro* substrates again showed that AntM was specific with regards to the substrate and timing of catalytic activity, requiring a specific substrate, potentially tethered to the cognate ACP domain from antimycin biosynthesis.

The structures solved of AntM in this study gave valuable insights into the catalytic nature of the standalone KR enzyme. The overall structural motifs present in the enzyme were typical of such proteins, whilst a large solvent exposed binding pocket was clearly and immediately visible within the structures. The NADPH reduction cofactor was located parallel to a defined Gly rich region, a structural motif conserved amongst enzymes utilising the phosphorylated nicotinamide cofactor. The catalytic triad (S178, Y191, K195) was also prominently located within the active site and all amino acid side chains were located close to the presumed site of catalytic activity, in close proximity to the reducing face of NADPH. The post-PKS tailoring KR, UrdMred, from the urdamycin biosynthetic pathway, used as the search model for solving the structure of AntM, can also provide insight into the nature and functionality of AntM. The

two enzymes share 52.3% sequence identity, including the catalytic triad residues, and only contain a single amino acid difference between the cognate NADPH binding motifs. The structural similarities between the proteins are abundantly visible from alignment of the structures; the two enzymes have similar secondary and tertiary structural motifs, largely designated by the capability of the protein in binding NADPH, as well as the catalytic capacity to carry out ketoreduction in a relatively large, solvent-exposed binding pocket. The urdamycin KR domain was solved as a homotetramer formed from two symmetry related dimers; an overall quarternary structure similar to the observed homoteteameric form of AntM. The major difference between the two protein structures, which could potentially impact or alter functionality of the enzymes, lie in the nature of the active site of the protein. The active site of AntM was similarly open and extended, with the reduction cofactor held deep in the binding site, however the space directly above the nicotinamide ring of the NADPH cofactor was occupied by two large, aromatic gatekeeper amino acids; Tyr and Phe. These gatekeepers would seemingly occlude the binding of compounds in the position occupied by the substrate in the tertiary structure of the UrdMred ketoreductase. The same residues in the UrdMred structure were Asn and Ile, relatively smaller residues providing lower levels of steric hindrance to the upper part of the pocket, in close proximity to the nicotinamide ring of NADPH. The proposed abrogation of substrate binding in the same part of the pocket as used in other similar KR domains, therefore would necessitate an altered binding conformation for the substrate of AntM, as visualised by *in silico* substrate docking.

Multiple *in vitro* and *in vivo* observations suggested binding of substrate compounds in the AntM active site occurred in an altered conformation compared to other similar, structurally characterised KRs, with AntM acting on a linear biosynthetic intermediate being the most likely scenario. *In silico* docking studies suggested that the terminal linear intermediate of antimycin biosynthesis was indeed the most likely substrate for AntM, as judged by the conformation of the docked compound compared to the alternative substrate, the cyclised, released, and C8 unreduced form of antimycin. The overall bound conformation of the linear substrate; the distance of the C8 carbonyl from the reducing face of the NADPH cofactor; and the calculated Gibbs free energy of binding of the linear substrate, all indicated that the linear, ACP-tethered intermediate is the preferred substrate for binding in the AntM active site and for subsequent catalysis. Mutation of the two larger gatekeeper residues (Tyr185 and Phe223) also indicated the importance of the two residues in substrate binding and catalysis during antimycin biosynthesis. When mutated to larger residues, the protein retained the ability to

bind NADPH, as indicated by ITC analysis *in vitro*, however; antimycin production was severely affected *in vivo*. Production of antimycins was only observed in pathways containing AntM mutants in which the large, aromatic residues were converted to smaller Ala residues. These results indicated that the two gatekeeper residues play important roles in enzyme activity, presumably by controlling access to the substrate binding channel, an eventuality much more suited to, and indicative of, AntM binding of a linear intermediate, rather than post-PKS ketoreduction of a cyclised product.

Bioinformatics can be a useful tool in the analysis of proteins and genes, and especially in the comparison of newly characterised proteins with those of unknown or cryptic functionality (Jenke-Kodama and Dittmann, 2009). Phylogenetic analysis of AntM showed that the protein was most similar to KRs from type II PKS systems, minimally composed of KS and ACP domains, with the remaining required domains predominantly acting *in trans* to carry out diversification of the chemical scaffold. Therefore, KRs from type II PKS systems are capable of acting on ACP-tethered intermediates, and concomitantly with biosynthesis of the core scaffold of the secondary metabolite. The clading of AntM predominantly with type II PKS KR domains was initially surprising due to the large variation between the assembly lines within which the KRs are functional; whilst the chemical reactions catalysed are extremely similar, large variations exist in the biosynthetic method between type II PKS systems and hybrid NRPS/type I PKS systems. Interestingly, AntM also formed a small clade with KRs from other varied pathways, including the griseoviridin, merochlorin, and coelimycin pathways, all of which synthesise specialised metabolites by following different biosynthetic methodologies. The specific functions of these KRs are yet to be determined. However, this indicates that the proposed methodology of AntM functionality; specifically, catalysis on an assembly line-tethered intermediate, could also be utilised by other KR domains from a variety of pathways, leading to the formation of a variety of bioactive compounds. In addition, the standalone KR, NatE, involved in the formation of neoantimycins, ring expanded members of the antimycin family, seems to operate in a very different manner (Skyrud et al., 2018; Zhou et al., 2018). Notably, unreduced neoantimcyins can be formed in the absence of the KR domain, showing that it is likely a post-PKS tailoring enzyme. Additionally, NatE functionality occurs on a chemical component added by the penultimate biosynthetic module, thereby less likely to effect TE domain mediated offloading, and the resultant hydroxyl from NatE function is not subsequently enzymatically derivatised, as is the case in antimycin biosynthesis. The neoantimycin KR also exhibits little sequence similarity to AntM and it is these exceptional

differences that show that these KRs are likely evolutionarily divergent and, therefore, perform significantly altered roles within their cognate hybrid pathways.

Efficiently engineering natural product biosynthetic pathways is an important and longstanding goal within the fields of secondary metabolite biosynthesis and the finding and development of novel compounds with interesting bioactivities. The use of standalone enzymes, and extending their use to alternative pathways can potentially be a powerful method, and one that is more amenable to augmentation. Understanding, in detail, the functionality of these enzymes would allow more targeted efforts at extending their activities to other systems of interest. The enzymatic activity of AntM, the standalone KR from the biosynthetic pathway of antimycins, has been characterised *in vitro*, *in vivo*, and structurally, and showed that the enzyme reduces the C8 carbonyl of the terminal linear intermediate while it remains tethered to the ACP domain of the PKS module AntD. Unearthing the specific mechanisms modulating function of the enzyme, such as potential interactions with assembly line proteins, would allow the development of the enzyme as a useful tool for future engineering endeavours. Importantly, this work demonstrated that AntM is the first standalone KR that acts concomitantly with an assembly-line like system to produce the core chemical scaffold and bioinformatics analysis suggests that there are other standalone KRs from assembly-line like systems that may function in a similar way.

Chapter 5 – A standalone cyclase offloading strategy for assembly line release of surugamides

Abstract

Non-ribosomal peptide synthetases (NRPSs) are responsible for the production of a large number of therapeutically relevant compounds. NRPSs are large assembly line-like machines whose terminal step in biosynthesis is the hydrolytic release, and frequently, macrocyclisation of the aminoacyl-S-thioester by an embedded thioesterase (TE). The surugamide biosynthetic pathway is composed of two NRPS assembly lines. One produces surugamide A, which is a cyclic octapeptide, and the other produces surugamide F, a linear decapeptide. The terminal module of each system lacks an embedded TE, which begs the question of how the peptides are released from the assembly line (and cyclised in the case of surugamide A). An alpha/beta hydrolase and β-lactamase were characterised *in vivo* and the former was found to be a type II TE for surugamide A, but not surugamide F, and the latter was a *trans*-acting release factor for both compounds. *In vitro* substrate utilisation assays unambiguously established that the β-lactamase, SurE, can produce mature surugamides A and F from *N*-acetylcysteamine (SNAC) thioester mimics of the cognate terminal biosynthetic intermediates, and further substrate assays developed the scope and promiscuity of the enzyme. Using bioinformatics, it was estimated that ~12% of filamentous Actinobacteria harbour an NRPS system lacking an embedded TE and instead use a standalone cyclase release strategy. This expands the paradigmatic understanding of how non-ribosomal peptides are released from the terminal NRPS module and adds a new dimension to the synthetic biology toolkit, potentially useful in the search for novel antibiotics.

5.1 Introduction

5.1.1 The desotamide family of antibiotics

The desotamide family of antibiotics are a group of predominantly cyclic, peptide antibiotics, synthesised by NRPS systems with the known members ranging in size between six and ten amino acids in length (Fazal et al., 2020). They are typified by the presence of at least one Trp or Phe residue and a C-terminal Gly or D-amino acid, which is a prerequisite for cyclisation of the peptide. Due to the presence of the terminal D-amino acid, the final NRPS modules

synthesising the compounds invariantly contain an E domain, catalysing the observed switch in stereochemical configuration, whilst all the NRPS systems also lack a canonical biosynthesis terminating TE domain (discussed in detail in the following Results section). The compounds comprising the family also commonly contain modified or bespoke precursors, extending the variety of amino acids observed in the antibiotics. The founding member of the family is desotamide A, originally discovered in 1997 from the fermentation broth of *Streptomyces* sp. NRRL 21611 (Miao et al., 1997). Later, a suite of structurally similar cyclic octapeptides named surugamides A-E were discovered from a marine microbe named *Streptomyces* sp. JAMM992, followed by four additional desotamide analogues (desotamides B-D) produced by *S. scopuliridis* SCSIO ZJ46 (Takada et al., 2013; Song et al., 2014). A further two desotamide analogues (E and F) were subsequently discovered, as well as C-terminal D-Orn-containing wollamides A and B, all produced by the same organism; *Streptomyces* nov. sp. MST-115088. Soon after came the discovery of surugamide F, a linear decapeptide previously unobserved during the initial discovery of surugamides A-E (Ninomiya et al., 2016). Additional chlorinated hexapeptide members of the desotamide family were recently identified, including the ulleungmycins produced by *Streptomyces* sp. KCB13F003 and noursamycins/curacomycins produced by *S. noursei* ATCC 11455 and *S. curacoi* NBRC 12761, respectively (Kaweewan et al., 2017; Son et al., 2017; Mudalungu et al., 2019).
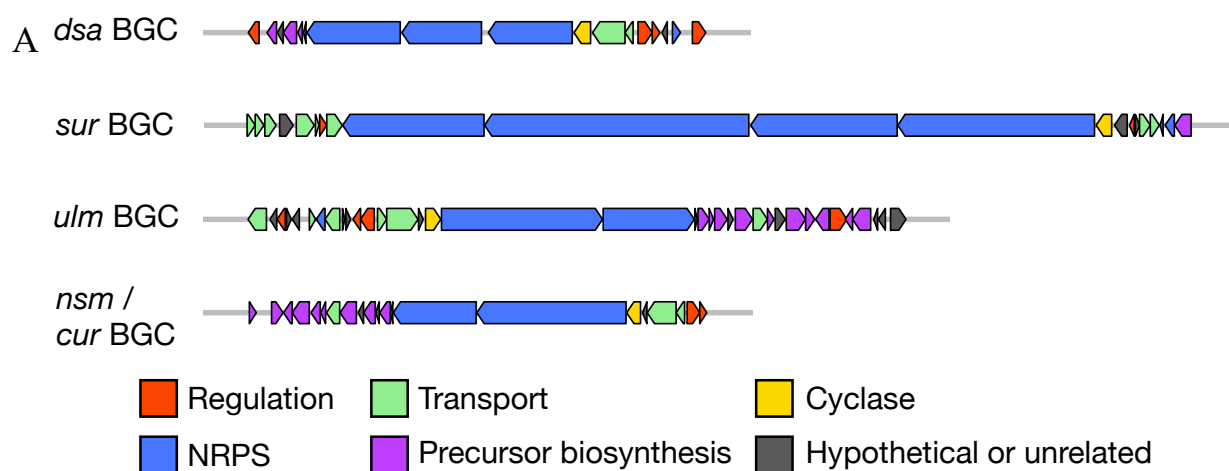
B
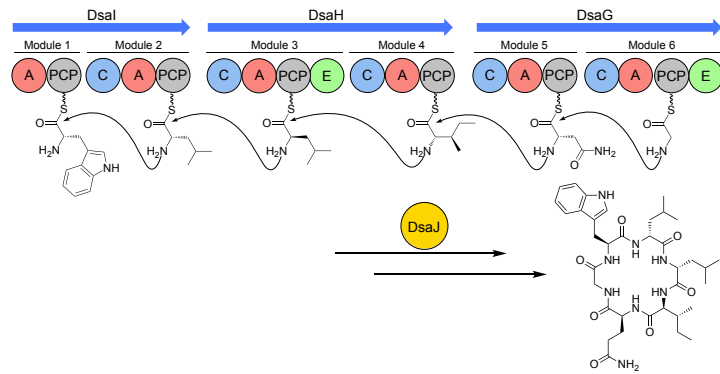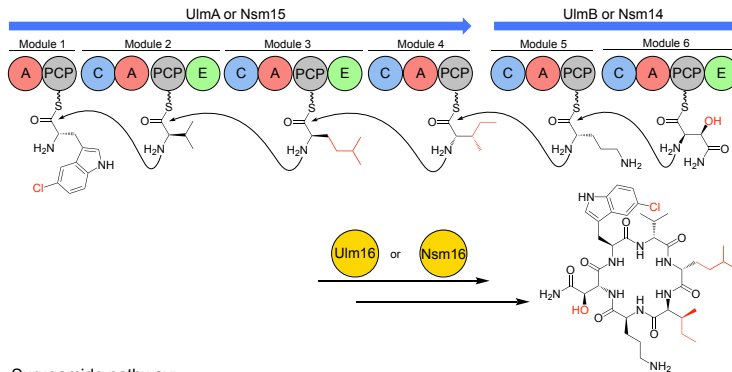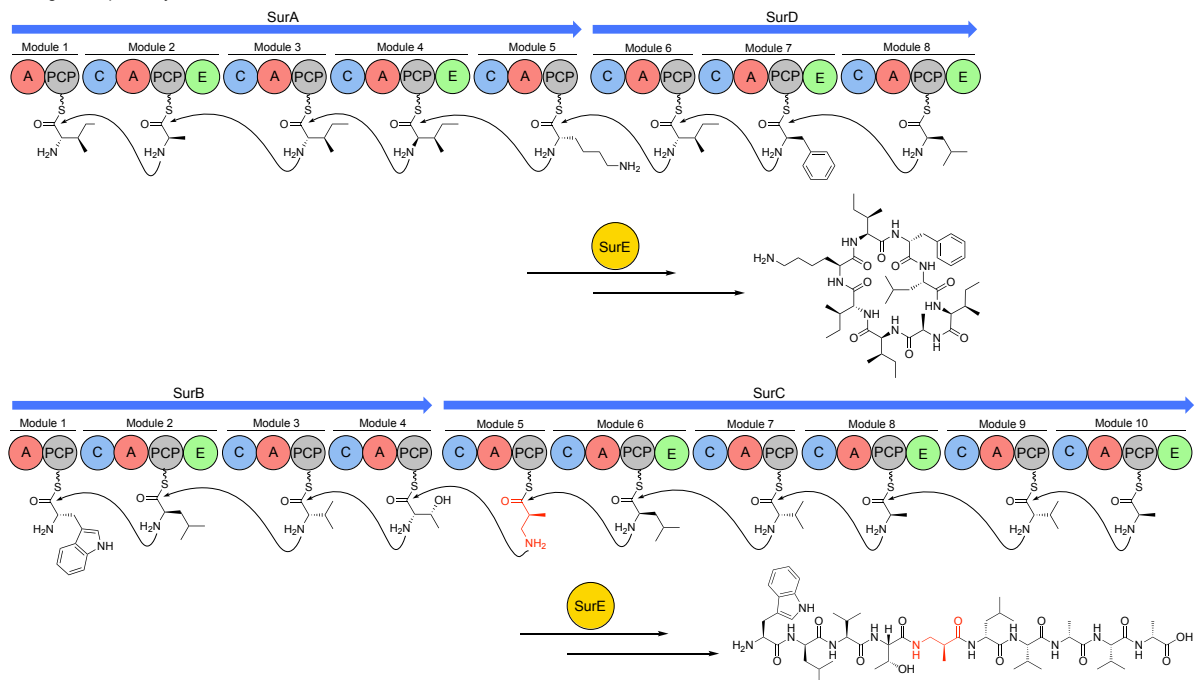
Figure 5.1: Overview of the desotamide family of antibiotics. A) The BGC organisation of all members of the family; genes are colour-coded dependent upon function of the gene product. B) Overall biosynthetic schemes of the pathways in the desotamide family of antibiotics. The NRPS systems are shown, with the A domain specificity indicated by the amino acid bound to the downstream PCP. Standalone cyclase enzymes are indicated below assembly lines, next to pathway products. Structures indicated in red are non-canonical or altered amino acids.

Biosynthetic gene clusters for all members of the desotamide family of antibiotics have been discovered and reported (Figure 5.1). The BGCs share multiple similarities in the organisation of the large NRPS genes and the remainder of the biosynthetic, regulatory, and transport related genes present within the cluster. The NRPS genes generally contain the encoding potential for the canonical set of core peptide synthesising domains, and as such, obey the collinearity principle with the number and nature of the domains and modules all rationalisable with the structure of the final compound. The surugamide BGC differs from the traditional NRPS logic by containing the biosynthetic potential for the synthesis of two structurally distinct compounds, a cyclic octapeptide and a linear decapeptide (Ninomiya et al., 2016). This was observed to be due to the abnormal organisation of the NRPS genes and their subsequent organisation into complete NRPS systems at the protein level: *surA* and *surD* form the NRPS system leading to the production of the surugamides A-E and *surB* and *surC* lead to the formation of the linear surugamide F (Figure 5.1). Upon formation of the NRPSs, however, the systems function as do the remainder of the desotamide family pathways.

The antimicrobial target or targets for members of the desotamide family remain unknown, but all compounds possess a minimum inhibitory concentration against Gram-positive indicator organisms that is in the micromolar range, and wollamides are active against *Mycobacterium tuberculosis* (Khalil et al., 2014). A structure-activity relationship study was recently performed with synthetic derivatives of wollamide B, which revealed that the Trp and Leu residues in the first and second positions of the macrocyclic ring, respectively, are essential for bioactivity and that it could be enhanced by altering the C-terminal D-Orn residue to D-Arg or D-Lys, but not to their L-stereoisomers (Khalil et al., 2019). The antibacterial effects of the remainder of the compounds comprising the family remain to be studied.

5.1.2 Precursor biosynthesis in the desotamide family of antibiotics

The desotamide family of antibiotics contains compounds composed of a variety of amino acids, many of which are bespoke, altered, or non-natural. The multitude of D-amino acids observed as integral and C-terminal residues are installed via A domain mediated selection and activation of the cognate L-amino acid, followed by epimerisation at the α-carbon by the module-encoded E domain. However, the desotamide family is composed of a variety of other amino acids, which contain chemically diverse groups. Examples of these include hydroxylated

and halogenated amino acids, formed by the action of dedicated enzymes on PCP-bound precursors; the biosynthesis of these precursors is discussed in section 1.6.1.

A well characterised example of the origin of a bespoke precursor is the biosynthesis of L-*allo*-Ile, a β-carbon epimerised form of Ile. E domains, *cis*-encoded within NRPS modules are capable of epimerising the α position of chiral amino acids, though the origins of β-epimerised Ile residues were unknown. The biosynthetic origin of the non-proteinogenic L-*allo*-Ile precursor was originally identified from the desotamide and marfomycin pathways (Figure 5.2) (Li et al., 2016). The two-enzyme system consists of an aminotransferase and an isomerase, whose collective action results in isomerisation at the β-carbon of Ile and occurs initially through covalent linking of the α-amino group of Ile to PLP, itself covalently linked to a Lys residue within the active site of the aminotransferase. This is followed by two deprotonations at the α- and β-carbons, catalysed by the aminotransferase and isomerase, respectively. Re-protonation of the β-carbon from the opposite side to initial abstraction of the proton results in the formation of the diastereoisomer, L-*allo*-Ile, upon release of the amino acid from the enzyme active site. This amino acid can then be directly adenylated by A domains within NRPS modules and incorporated into growing NRPs.
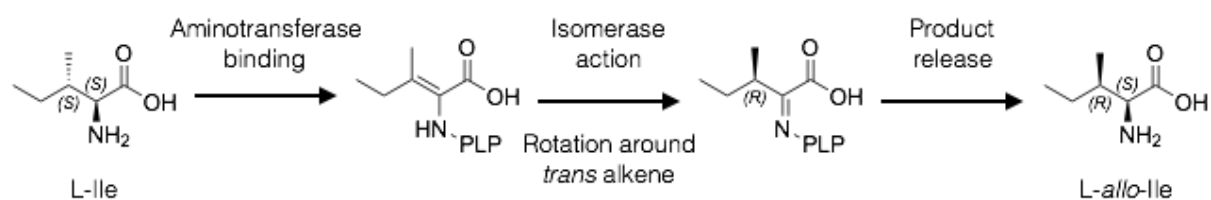
Figure 5.2: Simplified overview of the two-enzyme system that results in the formation of L-*allo*-Ile from L-Ile, by isomerisation at the β-carbon. The amino acid is covalently bound to PLP in the active site of an aminotransferase, followed by deprotonation, rotation around the formed *trans* olefin, and reprotonation from the opposite side resulting in an isomerised product.

Other bespoke precursors used in NRP biosynthesis are available from primary metabolic pathways, but the enzymes required for their creation are also frequently encoded within NRPS BGCs. This is likely due to insufficient availability of the primary metabolite pool and may also be due to spatiotemporal factors with respect to the expression and cellular location of the assembly line. Examples of such amino acid precursors are Orn and Kyn, produced in the

primary metabolic pathways of Arg formation and Trp decomposition, respectively (Cunin et al., 1986; Kurnasov et al., 2003). Enzymes such as amidinotransferases and Trp-2,3-dioxygenases are relatively commonly observed in natural product BGCs (including the ulleungmycin and noursamycin BGCs from the desotamide family) where they catalyse important reactions in the process of precursor supply. The BGC encoded amidinotransferase enzymes resemble commonly observed arginine:glycine amidinotransferases, which canonically produce Orn as a byproduct during the formation of creatine (Chen et al., 2013). In the context of secondary metabolite biosynthesis, the enzymes likely produce Orn from Arg as required by the NRPS assembly line. Trp-2,3-dioxygenases generally oxygenate the indole ring of Trp to produce *N*-formyl- L-Kyn, which can be enzymatically transformed, or spontaneously hydrolysed, to form Kyn. This amino acid can then be directly used and incorporated into the final compound, or utilised in the downstream synthesis of further precursor compounds (Miao et al., 2005). Much like the amidinotransferases, Trp-2,3-dioxygenases are usually encoded as part of the gene clusters that require them, however, in the case of the desotamide BGC, no gene candidate was identified leading to the suggestion that Kyn is obtained from primary metabolic processes.

Surugamide F, the largest member of the desotamide family of antibiotics, contains an unusual β-amino acid, 3-amino-2-methylpropionic acid (AMPA), at the fifth position of the compound. AMPA has been shown to be derived from two different sources: pyrimidine degradation and decarboxylation of 3-methylaspartate (Figure 5.3) (Beck et al., 2007; Webb et al., 2007). Catabolism of the pyrimidine base thymine begins with NAD(P)H-dependent reduction of the alkene within the heterocyclic six-membered ring, catalysed by dihydropyrimidine dehydrogenase. This is followed by amide bond hydrolysis and ring opening by a dihydropyrimidinase and the formation of AMPA is achieved through release of ammonia and carbon dioxide by β-ureidopropionase. The biosynthesis of AMPA can also occur through a divergent mechanistic route. The enzyme CrpG from the pathway biosynthesising the cryptophycins, was shown to have shared amino acid identity with pyruvoyl-dependent Asp decarboxylases, such as the characterised *E. coli* enzyme PanD, whose active, catalytic form is generated through internal proteolytic cleavage of the initially expressed proenzyme. CrpG was observed to effectively catalyse decarboxylation at the α-carboxylic acid of 3-methylaspartate, leading to the formation of AMPA. The enzyme activity was diastereoselective regarding the substrate as (2S,3R)-3-methylaspartate was preferred by 3-4 orders of magnitude over other

diastereoisomers, as well as L- and D-Asp. The formation of AMPA within the surugamide pathway is yet unresolved, however AMPA biosynthetic genes can be putatively identified within the *sur* BGC and, upon further investigation, could yield the methodology utilised by the surugamide pathway in the biosynthesis of the unusual β-amino acid precursor.

A

B
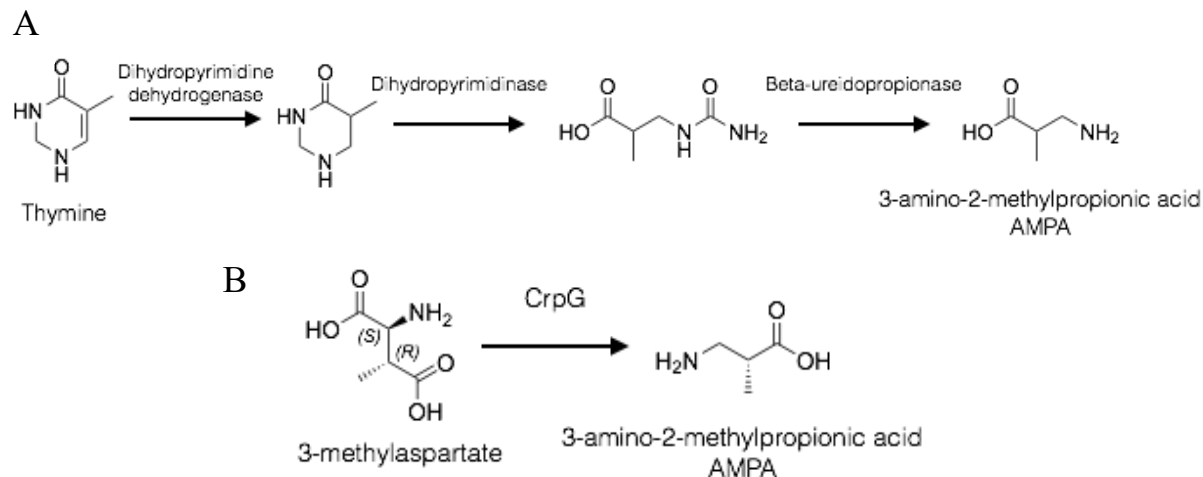
Figure 5.3: Two distinct biosynthetic pathways towards the unusual β-amino acid, AMPA, contained within the mature surugamide F peptide. A) AMPA formation through catabolism of the pyrimidine thymine. Three enzymes carry out reduction, ring opening, and the ultimate formation of the amino acid. B) A divergent mechanism for the formation of AMPA observed within the cryptophycin biosynthesising system, where the pyruvoyl-dependent enzyme, CrpG, catalyses the decarboxylation of 3-methylaspartate, resulting in formation of the β-amino acid.

5.1.3 Type I and type II thioesterase domains

Thioesterase domains, and proteins with structural folds similar to those of TE domains, play critical roles in the context of NRPS biosynthetic systems, as well as in pathways synthesising other classes of specialised metabolite. The two major types of TE domain observed as part of NRPS systems are the type I and type II TE domains. Type I TE domains are predominantly observed at the C-terminus of NRPS termination modules and catalyse the offloading of mature peptide scaffolds, usually by hydrolysis or macrocyclisation reactions. Type I TE domains are also prevalent in PKS systems where their disruption, similar to the NRPS counterparts, results in complete loss of compound production. The role and catalysis of type I TEs is discussed in section 1.4.5.

Type II TEs also play important roles during biosynthesis, however are less frequently observed within NRPS systems compared to terminating type I TEs. Type II TEs are *trans-*

acting enzymes and act as discrete hydrolytic enzymes with a variety of roles in peptide formation (Schwarzer et al., 2002). During NRPS functionality type II TEs are mainly catalytically active in the removal of any misprimed compounds from the Ppant extension of PCP domains, which can occur through the incorrect action of PPTase enzymes (Figure 5.4). These enzymes generally load a conserved Ser residue on PCP domains with the Ppant extension by usage of CoA, but can occasionally utilise acetyl-CoA, predominantly due to the relatively high cellular concentration of the compound, as well as its similarity to the required coenzyme. Type II TEs can also catalytically remove aberrantly loaded substrates and intermediates from PCP domains, usually caused by potential A domain mediated acceptance of incorrect amino acids (Figure 5.4) (Yeh et al., 2004). The proofreading type II TE enzymes, which hydrolyse any incorrectly loaded compounds yielding the terminal thiol of the Ppant extension, therefore must be capable of the recognition of multiple, variable substrates, tethered to a range of PCP domains within the NRPS system. Type II TEs have also been observed to possess other varied roles during specialised metabolite biosynthesis. In certain fatty acid synthase (FAS) systems, the thioesterases have a terminating and chain release role, similar to the type I TEs seen in PKS and NRPS pathways, whilst type II TE enzymes also take part in starter unit biosynthesis and selection in certain, rarer scenarios (Nakamura et al., 2015).
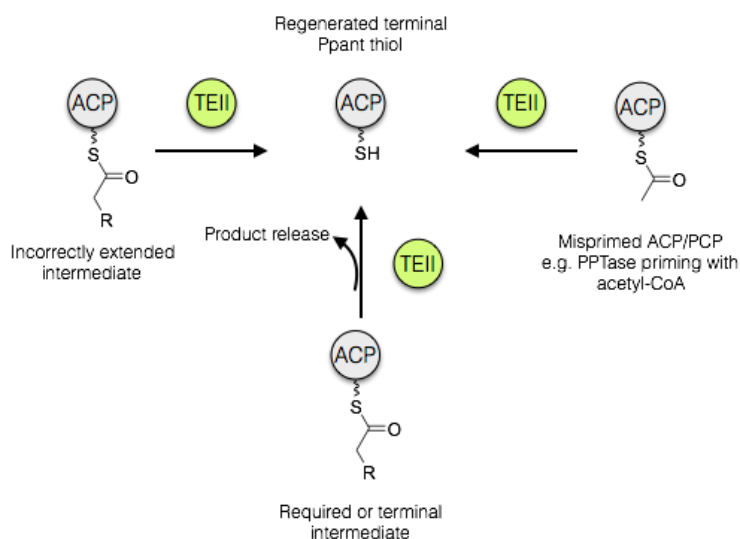
Figure 5.4: The predominant roles of type II TE domains in NRPS and PKS systems. The domains primarily function to regenerate the terminal Ppant thiol of carrier protein domains for further compound formation and catalysis. Functional roles include removal of aberrant intermediates, regeneration of misprimed carrier proteins, and limited product release (clockwise from top left).

Type II TEs have similar proofreading functions in both NRPS and PKS systems and much of the study surrounding the functionality of the enzymes has been carried out on the proteins as part of canonical type I PKS pathways. Studies of the type II TEs from the tylosin, coelimycin, and pikromycin PKS pathways showed the enzymes to be adept at the cleavage of acylated ACPs, as well as CoA and N-acetylcysteamine (NAC) bound substrate mimics (Kim et al., 2002; Kotowska et al., 2014). As ACP domains are acylated as part of the PKS biosynthetic process, type II TEs have the potential to carry out the redundant and unwanted cleavage of required substrates and intermediates, leading to a potential loss of yield and a less efficient rate of compound formation. Kinetic analysis of the TE enzymes, however, has shown that the cleavage rate for biosynthetic intermediates is much lower than observed for small, unwanted compounds, which likely allows the KS domain to utilise the bound substrate or intermediate prior to a potential cleavage reaction (Heathcote et al., 2001). Analysis of the NRPS type II thioesterases from the surfactin and bacitracin systems showed that the enzymes possessed a broad substrate specificity towards acetyl-PCP and PCPs bound with aminoacyl or peptidyl groups (Linne et al., 2004; Koglin et al., 2008). The role of the TEs in the cleavage of amino acids or peptides from PCP domains was thought to be a by-product of the promiscuity of the enzyme, with little relevance to the physiological role of the domain in NRPS functionality.

Type II thioesterases generally show typical α/β hydrolase folds, comparable to the overall topology of terminating type I TEs, which also possess the characteristic seven stranded β-sheet with eight surrounding helices (Figure 5.5) (Claxton et al., 2009). The catalytic triad is also characteristically composed of Ser, Asp, and His, and mutation of any of these residues to those with aliphatic side chains leads to an enzyme with abolished catalytic activity. Further analysis of conserved residues between numerous type II TE enzymes revealed a structurally important Asp residue, which when mutated yielded an unstable protein, although the amino acid was observed to have no role in catalysis (Linne et al., 2004). α/β hydrolase enzymes contain external lid regions, generally occupying central positions in relation to the remainder of the protein structure. The type II TEs, however, possess a shorter and shifted lid motif, leading to only slight occlusion of the active site residues, and subsequently a more accessible binding pocket in comparison to type I TEs. The accessibility of the binding pocket allows the proofreading TE to dock onto and bind a variety of PCPs within its cognate system, a required aspect for enzyme function. The type II TE enzyme, however, also has a shallower active site

cavity than other α/β hydrolase enzymes, therefore only permitting binding and catalysis on small compounds bound to the thiol group terminating extensions of loaded PCP domains (Koglin et al., 2008; Claxton et al., 2009). The structure of the enzyme, particularly the nature of the shallow, accessible binding pocket, correlates with the function of the type II TE as a proofreading enzyme, functional in the removal of small molecules from misprimed PCPs, which if left unchanged would cause the NRPS system to be non-functional and thus a waste of cellular resources.
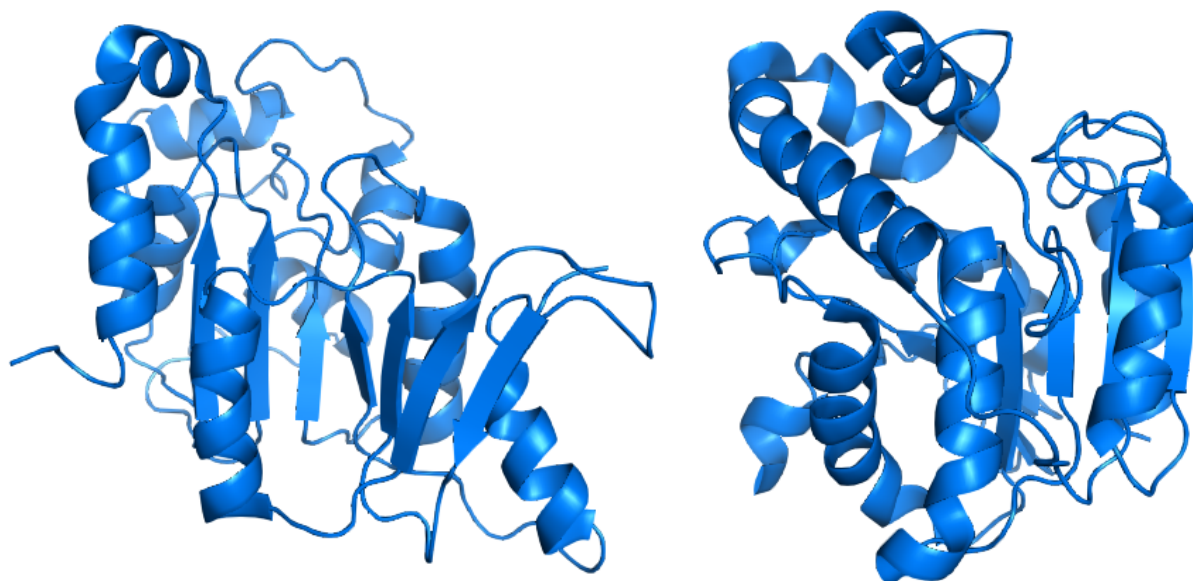
Figure 5.5: The structures of the type II TE domains from the surfactin biosynthetic system (2RON, left) and RifR from the rifamycin system (3FLA, right). The similar structures contain the core α/β-hydrolase fold motif, with β-sheets surrounded by α-helices. The flexible loop regions which partially occlude the active sites of the domains are at the top of the structures in each case.

5.1.4 *In vitro* chemoenzymatic cyclisation

Cyclisation of peptides is an important strategy and imparts a number of advantages and benefits on a peptide, particularly antibiotic and other therapeutic peptides. Cyclic peptides exhibit better biological activities, primarily due to conformational rigidity allowing better binding of target molecules. The entropy cost of an active peptide binding a protein target is reduced when binding cyclic peptides. Cyclic peptides, particularly those that are cyclised at the N- and C-termini, are much less susceptible to hydrolysis by peptidases, which can readily cleave linear, more flexible peptides in a cellular environment. A number of cyclic peptides have also been found to be able to cross membranes better than similar linear peptides (Joo,

2012; Gang et al., 2018). A number of NRPS-synthesised peptides are cyclised upon termination of biosynthesis and cleavage from the assembly line, including peptides with antibiotic and immunosuppressive activities. Vancomycin and gramicidin are cogent examples of cyclic NRPs with potent antibacterial effects (Bischoff et al., 2001).

Natural products have inspired a large proportion of the drugs in clinical use and subsequently a large effort has been committed to the chemical synthesis of mimics of specialised metabolites, including cyclic peptides. The *in vitro* cyclisation of synthetic peptides, however, can be extremely difficult to achieve, leading to backlogs in synthetic procedures and the exploration of alternative methods of cyclisation. Chemical methods for cyclisation of peptide scaffolds necessitate the use of protecting groups on amino acid side chains to avoid off target cyclisation, and methodologies often also utilise harsh conditions and reagents (Lambert et al., 2001). The use of enzymes as catalysts for the cyclisation reaction, however, is a potential method for the synthesis of cyclic peptides without the requirement for severe reaction conditions, and with a much simpler synthetic pipeline of peptide synthesis, removal of protecting groups, followed by chemoenzymatic cyclisation using a suitable enzyme (Nuijens et al., 2019).

The use of enzymatic proteins, such as NRPS TE enzymes, as biosynthetic tools for peptide cyclisation, whilst potentially extremely effective on both small- and large-scale applications, requires the use of specific substrates, both in terms of peptide composition due to enzyme specificity, as well as in the peptide terminating group. TE domains, in addition to other terminating enzymes, are naturally functional on substrates tethered to PCP (or ACP) domains, and held as thioesters at the terminus of the Ppant extension. Activity of these domains on basic peptides, terminating in amino and carboxy groups, is therefore limited, however, simple, synthetic substrate mimics are commonly utilised. N-acetylcysteamine (SNAC) thioester mimics were first utilised to study the cyclising TE domain from the erythromycin synthesising PKS system, and act as structural mimics for the terminating portion of the Ppant extension (Figure 5.6) (Jacobsen et al., 1997; Gokhale et al., 1999). TycC-TE, the thioesterase from the tyrocidine A synthetase, was the first NRPS system TE domain to be characterised *in vitro*, by the use of SNAC-thioester mimics of tyrocidine (Trauger et al., 2000). Incubation of the SNAC-thioester mimic of the tyrocidine A peptide with the cognate TE domain led to formation of the cyclic peptide, as observed from the *in vivo* system, with only a small amount of linear product formed from hydrolysis of the substrate mimic. The use of SNAC-thioester

mimics of tyrocidine allowed detailed characterisation of the TycC-TE domain as an enzyme which is only selective for the residues at the N- and C-terminus of the peptide, thereby allowing the screening of compound libraries with internally altered residues for improvements in bioactivity (Figure 5.6). This study demonstrated the utility of SNAC-thioester mimics as substrates, along with excised TE domains, for the biocatalytic preparation of cyclic peptides with interesting bioactivities. A number of further studies have demonstrated the effectiveness of SNAC-thioesters as substrate mimics for a wide variety of NRPS domains, including alternative termination enzymes, as well as *trans*-acting and post assembly line modification enzymes (Ehmann et al., 2000; Kohli et al., 2001).
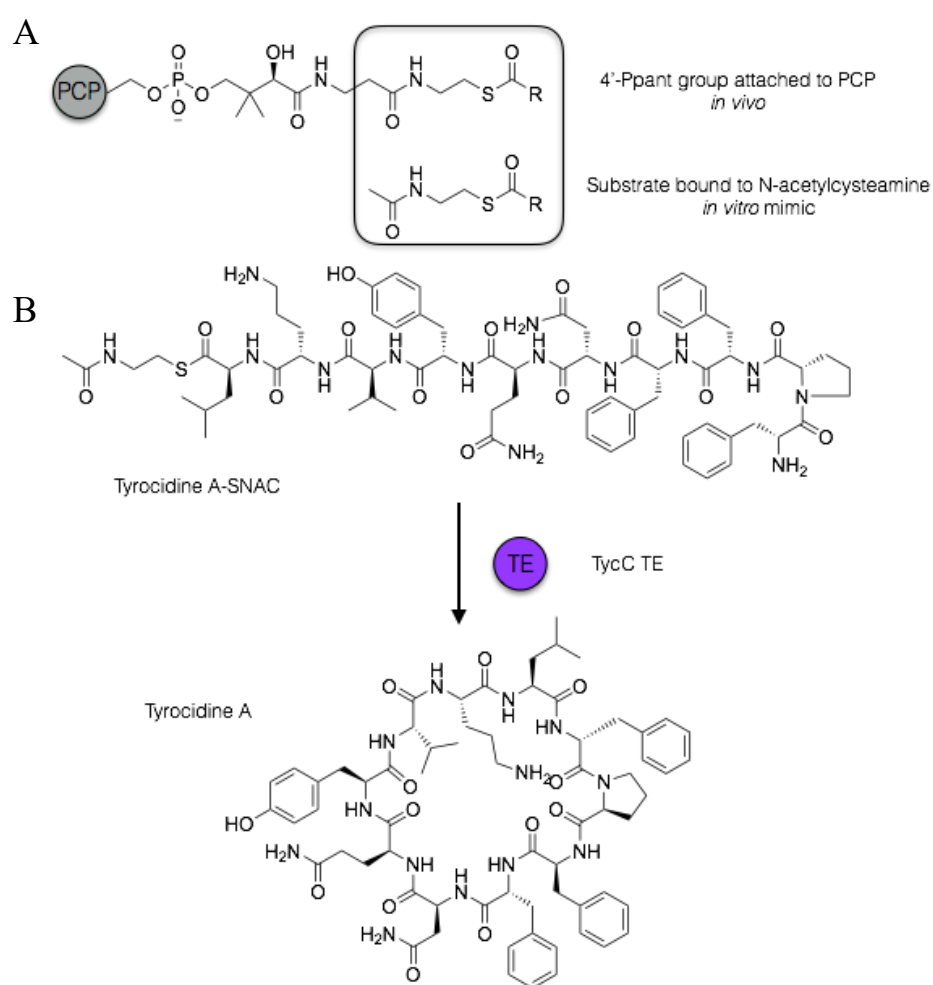
Figure 5.6: A) Overview of N-acetylcysteamine linked subtrates, outlining the similarity of the mimic substrate to the naturally observed 4'-Ppant group. B) Example of the usage of SNAC-substrates showing the TycC TE mediated cyclisation and offloading of tyrocidine A and formation of the mature compound.

Excised TE domains offer an attractive methodology for the synthesis of NRPs with altered amino acid compositions, with the potential for altered or improved bioactivity profiles. However, in multiple *in vitro* scenarios, the yield of the product, cyclic peptide has been observed to be relatively low, problematic when the purified enzymes are used as biocatalytic tools for the production of active peptides. A number of other substrate mimics have been developed in order to achieve improved yields or for ease of synthesis, therefore providing a set of substrate mimics for various applications. One such substrate mimic makes use of acrylamide polyethylene glycol (PEGA) solid phase peptide synthesis (SPPS) resins, upon which peptides are synthesised via a biomimetic linker (Figure 5.7) (Kohli et al., 2002). The linker region acts as a structural mimic of the Ppant extension used in *in vivo* systems and is capable of acting as the recognisable docking point for excised TE domains, which can then carry out their catalytic function on the tethered peptide, again demonstrated with the TycC-TE domain. The use of PEGA-linker-peptide type substrates alleviates the need of lengthy synthetic and purification steps prior to enzyme assaying, thereby providing a simple, scalable method for the characterisation of excised NRPS domains, as well as the potential for increased production of active compounds. Nonetheless, *in vitro* chemoenzymatic methods have proved invaluable in the characterisation of the enzymology and functionality of numerous NRPS domains, insights which play important roles toward the desired outcome of improved biocatalytic enzyme functionality and the production of novel compounds with altered or enhanced bioactivities.
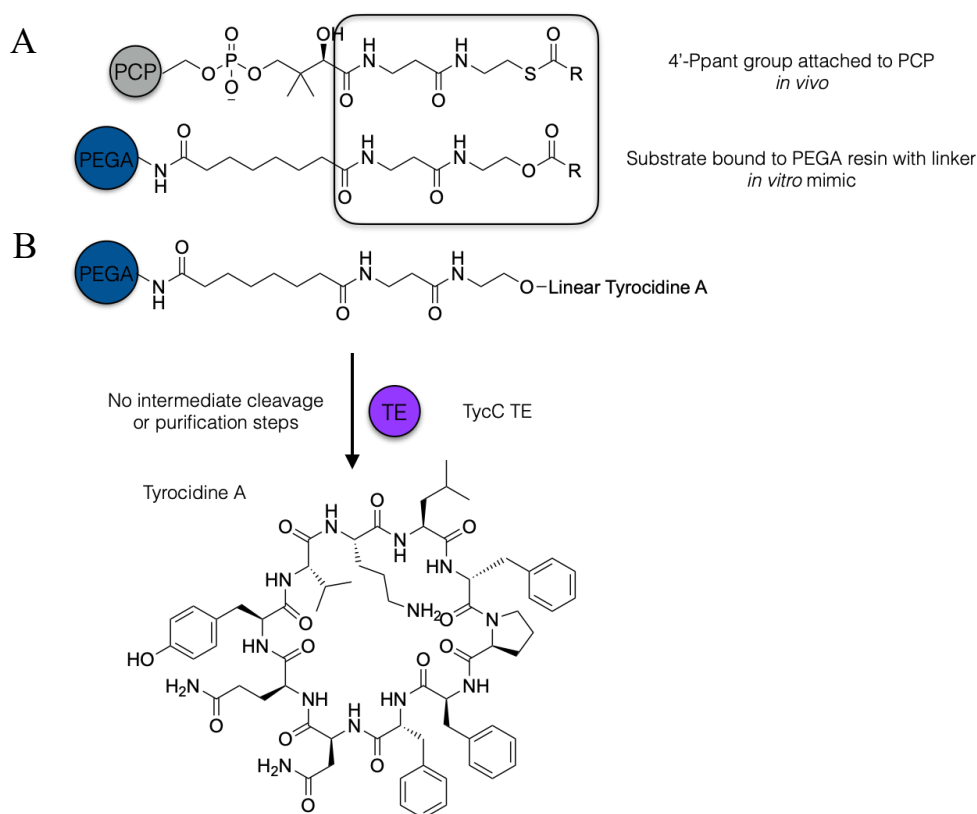
Figure 5.7: A) Overview of the similarity between PEGA-linker tethered peptides and those observed naturally as part of *in vivo* NRPS systems. The terminal part of the linker identically mimics the 4'-Ppant moiety, with the exception of an oxo-ester linkage at the terminus. B) The usage of PEGA-linker-peptide substrates as initially exemplified by the TycC TE domain from the tyrocidine biosynthesising pathway. The TE domain release the mature, cyclised form of the peptide from the resin without the need for lengthy preceding steps.

5.2 Aims and objectives

The aims of this project were to illuminate the method of non-ribosomal peptide offloading for the surugamide biosynthetic system. Surugamides, encompassing numerous isoforms of surugamide A and surugamide F, a cyclic octapeptide and linear decapeptide respectively, had been shown to be produced by various species of *Streptomyces*, including the commonly used heterologous expression strain *Streptomyces albus* J1074. Sequence and biosynthetic gene cluster analysis (BGC) had shown that the BGC contained four NRPS genes (*surABCD*) which, when expressed, directed the formation of both surugamide A (from genes *surAD*) and surugamide F (from genes *surBC*). Each of the cognate NRPS assembly lines, however, lacked canonical terminating domains and the final domain within each pathway was an epimerase. An alternative offloading mechanism must therefore be responsible for the terminal step in

surugamide biosynthesis, likely encoded within the surugamide BGC itself. This project aimed to identify and characterise the off-loading factor in detail.

Much of this work was published in the following publications: (Thankachan et al., 2019; Fazal et al., 2020).

5.3 Results

5.3.1 Production of surugamides *in vivo*

The surugamides are a group of NRPs, originally isolated from the marine *Streptomyces* sp. JAMM992. Alongside isolation of the peptides, the BGC directing the formation of the surugamides was identified, allowing the analysis of other *Actinobacterial* genome sequences for the potential of harbouring the same set of NRPS genes. Surugamides were also recently detected in fermentation broths generated from *Streptomyces albus* J1074. In addition, a strain genetically similar to J1074 and one with which we work with routinely in the lab, *Streptomyces albus* S4, harboured a putative surugamide BGC. In order to verify that this surugamide BGC could produce surugamides, the wildtype strain of *S. albus* S4 was cultured in liquid ISP2 medium for seven days at 30 °C, after which mycelia and solid particulate matter were collected by centrifugation. The supernatant was extracted with two volumes of ethyl acetate in a liquid-liquid extraction system, before evaporation of the solvent and analysis of the resultant extract by LC-HRMS. The mass spectral analysis showed that the major surugamide peptides were produced, with sizeable quantities of the cyclic octapeptide surugamide A observed, but a much lower intensity was observed for the linear decapeptide surugamide F (Figure 5.8). The fact that the *surBC* genes, that produce the NRPS assembly lines that synthesise surugamide F, are predicted to be co-transcribed with the *surAD* NRPS genes producing surugamide A, would have initially suggested a similar level of production for the two compounds. However, without further analysis into the transcriptional and translational control, as well as the assembly line synthesis of the longer decapeptidic compound, the much lower rate of production is yet unknown, with the most viable explanation likely being the nature of the unusual amino acid AMPA and the low expected cellular quantities of the substrate. However, these conditions showed that the production of the surugamides was easily achievable from *S. albus* S4 by utilisation of basic culture and extraction conditions, whilst other laboratory conditions were also found to be viable (such as plate-based growth and methanolic extraction, for example).
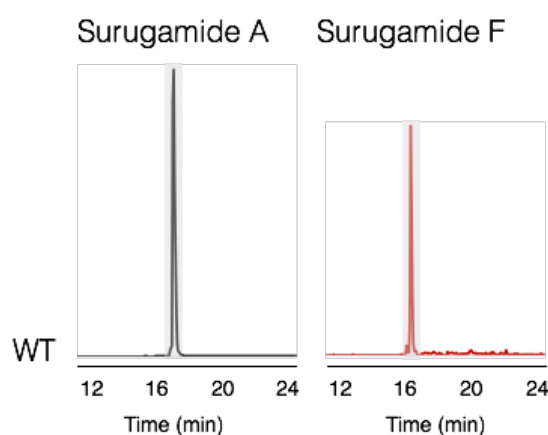
Figure 5.8: LC-HRESIMS analysis of chemical extracts prepared from the *S. albus* S4 WT strain. Extracted ion chromatograms ([M+H]$^+$, [M+Na]$^+$) corresponding to surugamide A ($C_{48}H_{81}N_9O_8$) and surugamide F ($C_{52}H_{85}N_{11}O_{12}$) are shown (Thankachan et al., 2019). Exact neutral m/z values used to determine EICs were: 911.6208 and 1055.6379. The y-axes are set to 1×10^6 for surugamide A spectra and 1×10^5 for surugamide F spectra.

5.3.2 Assembly line release of surugamides *in vivo*

Having verified *S. albus* S4 does indeed produce surugamides, a peculiarity of the *sur* BGC became obvious; all assembly lines involved in the production of the compounds lack a *cis*-encoded TE domain. Initially it was assumed that the terminating domain for the assembly line that produces surugamide A (*surAD*) would be a macrocyclising TE domain, whereas the terminating enzyme for surugamide F production (*surBC*) would likely be a TE domain employing a simple hydrolytic strategy. However, upon inspection of the NRPS assembly lines composing the surugamide pathway, no such termination motifs were observed in *cis* at the C-terminus of both pathways and no alternative termination domains, such as reductase or specialised condensation domains, were present. Instead both pathways terminated in epimerase domains, which would provide the D- amino acid observed at the final position of all surugamide peptides. It was therefore assumed that a novel termination strategy may be employed to offload both surugamide A and surugamide F form their cognate assembly lines, which would involve an enzymatic domain encoded elsewhere within the surugamide BGC.

In order to investigate the possibility of other genes within the surugamide BGC being responsible for the offloading of the mature peptides, the remaining genes from the antiSMASH assigned gene cluster were evaluated for their bioinformatic potential to catalyse

the requisite reaction (Figure 5.9). The majority of the genes encoded within the pathway were predicted to encode transporters or pathway regulators, activities inconsistent with the cyclisation of surugamide A and the hydrolytic offloading of surugamide F. Interestingly, a large number of genes near the predicted border of the pathway formed part of the folate biosynthetic pathway, potentially indicating a surugamide pathway requirement for one carbon methyl transfer reactions, a possibility requiring further study. A number of putative or hypothetical proteins also formed part of the pathway, however, two genes, *surE* and *surF,* were of particular interest due to their Pfam assignments. The *surE* gene encodes for a protein belonging to the β-lactamase superfamily. Members of this family are typically hydrolytic enzymes; for example, β-lactamases cleave the lactam ring of β-lactam antibiotics. Whilst SurE could potentially catalyse the cleavage of chemical bonds in the context of surugamide biosynthesis, the protein did not resemble canonical TE domains from other NRPS pathways, catalysing either macrocyclisation or hydrolytic offloading. The only previous observation of a β-lactamase superfamily protein potentially playing a key role in termination of synthesis, was as part of the mannopeptimycin biosynthetic pathway. The lipoglycopeptide pathway employs NRPS biosynthesis logic but similarly lacks a C-terminal offloading domain, leading to the suggestion that MppK, a β-lactamase family protein, could be responsible for peptide offloading, although this hypothesis was not tested in any detail. Overall, the hydrolytic capabilities of this family of enzymes suggested a potential role in offloading of the linear peptide, surugamide F.

The *surF* gene encodes for a protein belonging to the α/β hydrolase family. Type I and II thioesterase domains typically contain α/β hydrolase folds, potentially suggesting that SurF could be a standalone TE, offloading mature products from the surugamide synthesising assembly line. However, genes encoding standalone TEs are typically type II proofreading enzymes. Type II TEs, whilst not essential for production of the secondary metabolite, dramatically increase the rate of production and fidelity of the pathway by removing any aberrant substrates and intermediates which may have formed. In order to test which of the two candidate enzymes, SurE and SurF, were responsible for peptide offloading from the surugamide assembly lines, an *in vivo* strategy was employed to test the effect of deletions in the production of mature surugamides.
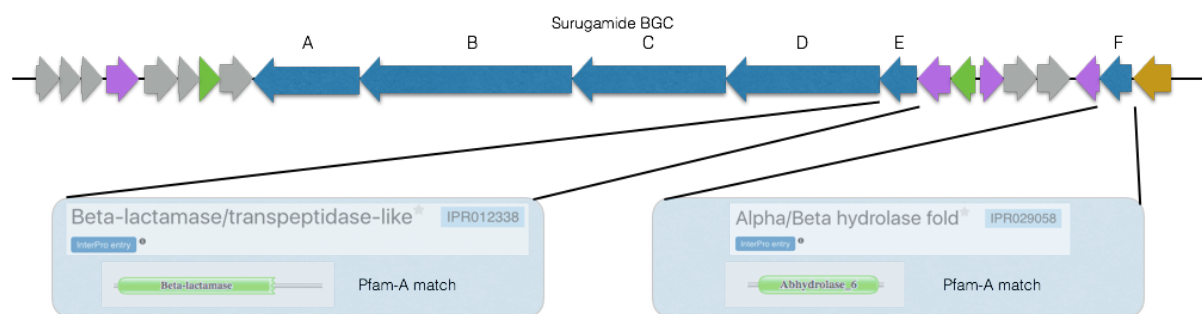
Figure 5.9: The surugamide BGC. Analysis of the genes within the BGC and the expected functionalities of the gene products led to the identification of SurE and SurF as potential surugamide offloading factors. InterPro and Pfam bioinformatic analyses of gene products are shown beneath the relevant genes.

The Redirect recombineering protocol was used to generate deletion mutants of *surE* and *surF* in the *S. albus* S4 background, giving the Δ*surE* and Δ*surF* strains. Complementation strains were also created by supplying the deleted gene (*surE* or *surF*) under the control of a constitutive promoter into the deletion mutants. All cloning steps for creation of deletion mutants and complementation strains was carried out by Dr Divya Thankachan (Seipke Lab, University of Leeds). Chemical extracts were prepared from the two deletion and two complementation strains and analysed by LC-HRMS (Figure 5.10). Deletion of *surF* did not abolish production of the surugamides, but did lead to severe reduction in titres of both major compounds. This initially suggested that SurF acts as a type II TE for the surugamide pathway, which was corroborated by the mass spectral analysis of the complementation strain showing that titres returned to near wildtype levels in the presence of an overexpressed form of the *surF* gene. LC-HRMS of extracts prepared from the Δ*surE* strain showed that production of the surugamides was completely abolished, and was restored to near wildtype levels under conditions of complementation of the deleted *surE* gene. These analyses showed that SurE is essential for the synthesis of surugamides A and F *in vivo*, and likely acts as the dedicated release factor for both compounds.
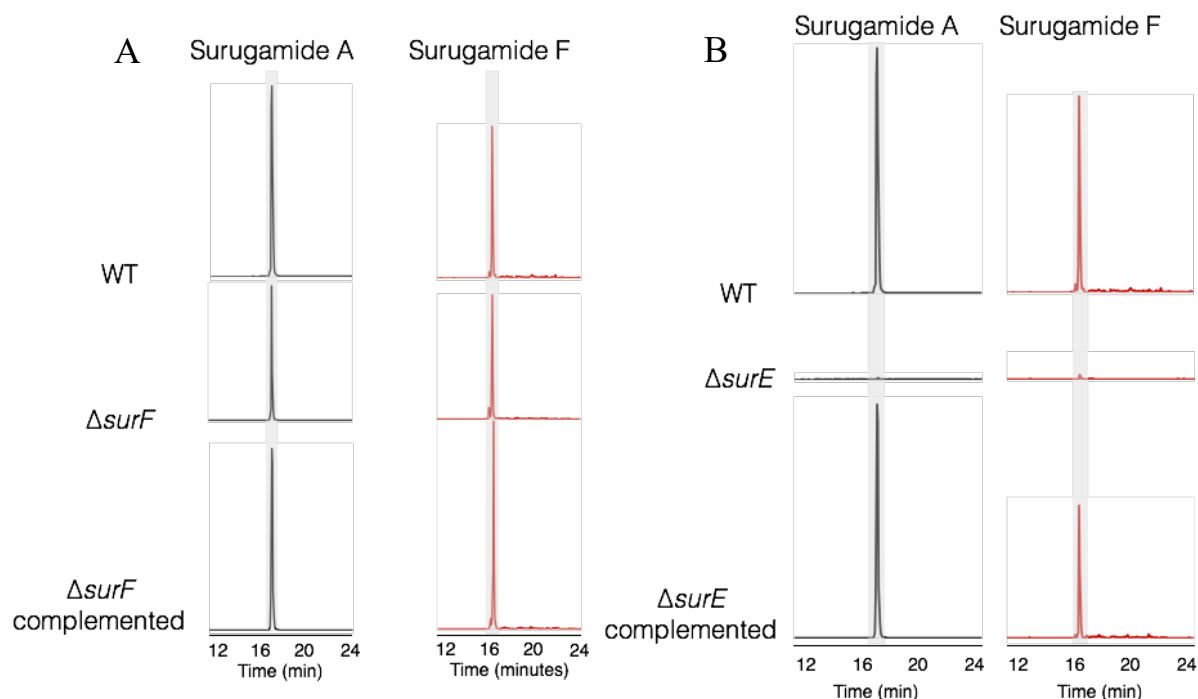
Figure 5.10: A) LC-HRESIMS analysis of chemical extracts prepared from the denoted surugamide production and *surF* gene deletion strains. Extracted ion chromatograms ([M+H]$^+$, [M+Na]$^+$) corresponding to surugamide A ($C_{48}H_{81}N_9O_8$) and surugamide F ($C_{52}H_{85}N_{11}O_{12}$) are shown. Exact neutral m/z values used to determine EICs were: 911.6208 and 1055.6379. The y-axes are set to 1 x 10^6 for surugamide A spectra and 1 x 10^5 for surugamide F spectra. Deletion of *surF* leads to reduced levels of surugamide production, and complementation returned surugamide formation to near WT levels. B) LC-HRESIMS analysis of chemical extracts prepared from the denoted surugamide production and *surE* gene deletion strains. Extracted ion chromatograms ([M+H]$^+$, [M+Na]$^+$) corresponding to surugamide A ($C_{48}H_{81}N_9O_8$) and surugamide F ($C_{52}H_{85}N_{11}O_{12}$) are shown. Exact neutral m/z values used to determine EICs were: 911.6208 and 1055.6379. The y-axes are set to 1 x 10^6 for surugamide A spectra and 1 x 10^5 for surugamide F spectra. Deletion of *surE* abolishes surugamide production, whilst complementation of the gene in *trans* restores production. Adapted from (Thankachan et al., 2019).

5.3.3 Production and purification of SurE

In vivo analyses had shown that SurE was the likely the release factor for all surugamides observed and produced within organisms containing the cognate BGC. In order to unambiguously show that SurE was a standalone enzyme acting as the offloading domain for all surugamides, the functionality of the protein was studied *in vitro*. The *surE* coding sequence was therefore cloned into the expression vector pET28a, by use of the NdeI and HindIII restriction enzymes. The pET28a-*surE* plasmid, which would produce an N-terminally His-tagged protein when expressed from the IPTG inducible T7 promoter system, was transformed into *E. coli* BL21-Gold(DE3) cells. The cells were cultured in auto-induction medium for 60 hours, after which time the cells were collected by centrifugation and lysed. (His)$_6$-SurE was purified from lysates by IMAC with Ni-NTA resin followed by gel filtration chromatography (Figure 5.11). This yielded a pure sample of the recombinant protein as judged by SDS-PAGE analysis, which also showed that the purified protein had the expected molecular weight of 49,764.00 Da. Elution of the protein from the gel filtration column at the expected elution volume for a monomeric species confirmed the molecular weight of the protein.
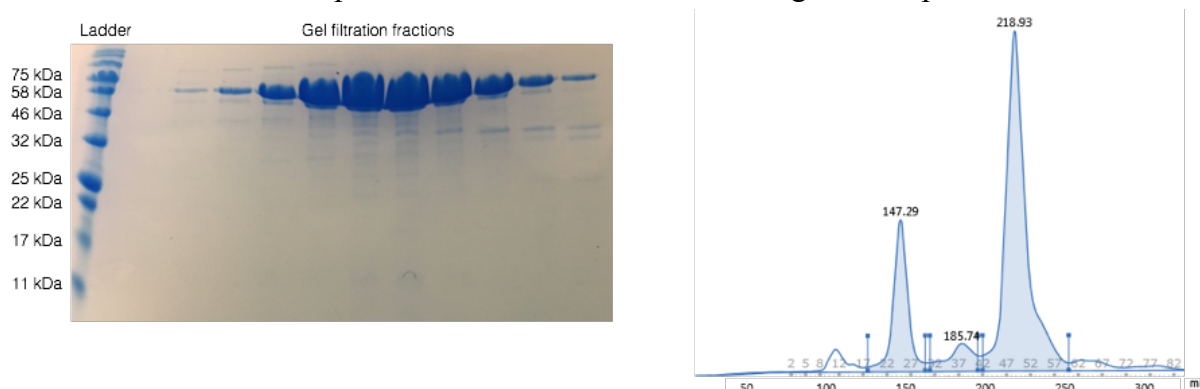
Figure 5.11: SDS-PAGE analysis of purified (His)$_6$-SurE, alongside molecular weight ladder (left). Absorption at 280 nm of elution buffer during gel filtration purification of (His)$_6$-SurE (right). The protein eluted at an elution volume of 2108.93mL indicative of a monomeric species.

5.3.4 Synthesis of SNAC-surugamide substrates

In order to characterise SurE *in vitro*, the terminal biosynthetic intermediate for surugamides A and F were synthesised by solid phase peptide synthesis (SPPS) and converted to N-acetylcysteamine (SNAC) thioester mimics (Figure 5.12). The peptides were synthesised as SNAC-thioesters to mimic the termini of the flexible 4'-phosphopantetheinyl linker that joins the growing peptide chain to the proteinaceous NRPS machinery, and allows reactions to occur at the enzymatic domains of the modules. Peptides were purified by preparative or semi-preparative HPLC and the identity of each of the SNAC-peptide mimics was confirmed by LC-

HRMS. Two truncated surugamide-type peptides were also synthesised to test the substrate scope of SurE (Figure 5.12). A truncated surugamide A-type hexapeptide retaining six out of the eight amino acids, but with the exclusion of the internal neighbouring L-Lys and L-Ile residues, was synthesised and converted into a SNAC-thioester. Interestingly, the four C-terminal amino acids of surugamide F are a repeated dipeptide motif (Val- D-Ala- Val- D-Ala). To test the substrate scope of SurE on altered octapeptidic substrates, a truncated version of surugamide F, missing the terminal Val- D-Ala dipeptide, was synthesised and converted to a SNAC-thioester. The identity of each of the peptides was confirmed by LC-HRMS, after purification by preparative or semi-preparative HPLC. A thorough investigation of the substrate tolerance and capabilities of SurE is of great of interest for chemoenzymatic and synthetic biology applications, and work towards this endeavour is continuing within the lab. A full list of synthesised substrates is shown in Appendix 1.
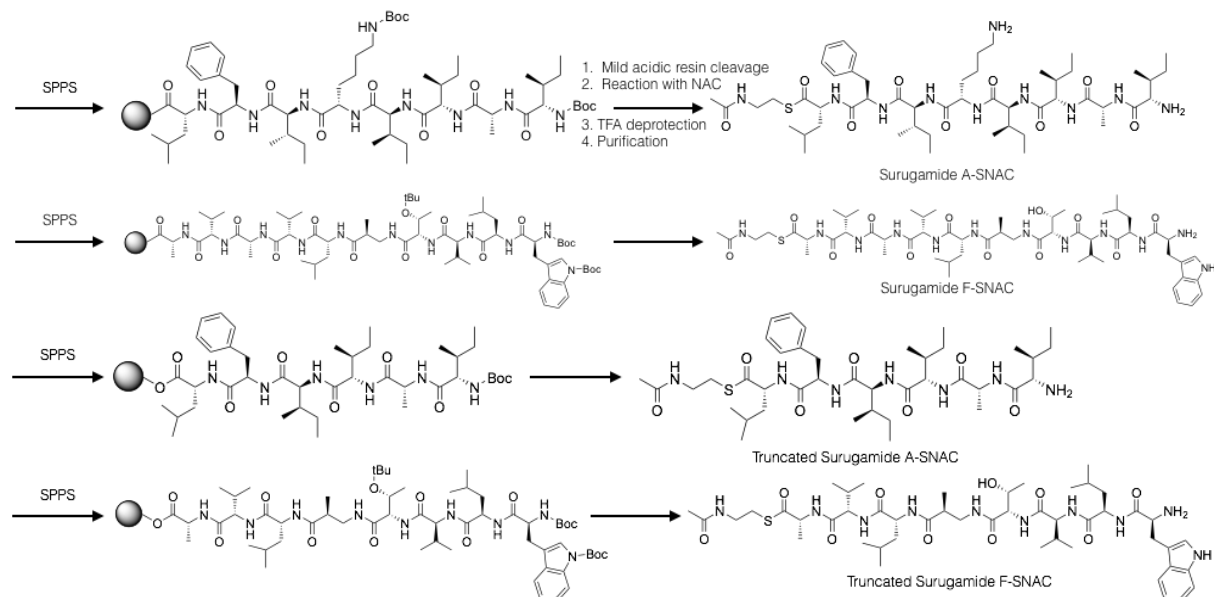
Figure 5.12: Schematic overview of the synthesis of SNAC-substrates prepared for initial testing with SurE. Protected peptides were synthesised by SPPS on 2-chlorotrityl chloride resin, before the same four steps of mild acidic cleavage, N-acetylcysteamine reaction, deprotection with TFA, and purification were undertaken, yielding the required SNAC-surugamide substrates.

5.3.5 SurE cyclisation assays with SNAC substrates

Substrate assays of the recombinantly produced (His)$_6$-SurE enzyme with the synthesised and purified SNAC-surugamide substrates allowed unambiguous confirmation of the activity of the offloading domain, originally observed *in vivo*. *In vitro* substrate assays involving the

recombinant SurE protein and the synthesised SNAC-surugamide A substrate, showed accumulation of a dehydrated surugamide A species by LC-HRMS. To confirm the nature of the produced compound, the products of the *in vitro* assay were co-injected with the extracted natural product and analysed by LC-HRMS (Figure 5.13). These data revealed that the *in vitro* produced surugamide A co-eluted with the natural product, indicating that the compound was indeed cyclic and not a dehydrated linear form of surugamide A. In the absence of the enzyme or the substrate, no changes were observed in the reaction mixture, indicating a lack of activity in the absence of either SNAC-substrate or cyclisation enzyme. In addition, no cyclisation and surugamide A formation was seen in assays utilising a mutant form of SurE, where the catalytically essential Ser had been replaced with Ala (Figure 5.13B). Taken together, and in combination with the *in vivo* observations, these results indicated that SurE was the essential cyclisation enzyme and release factor for surugamide A.
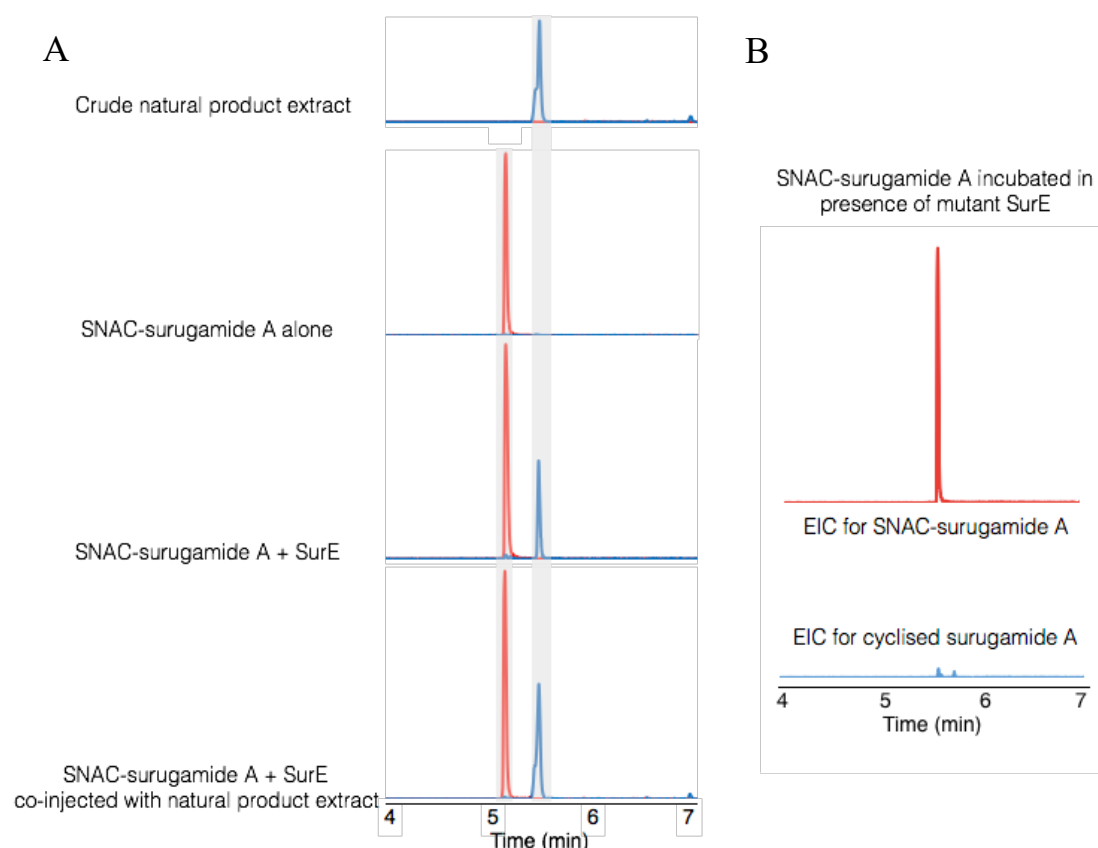
Figure 5.13: A) LC-HRESIMS analysis of *in vitro* reactions of SurE with SNAC-surugamide A. EICs ([M+H]$^+$, [M+Na]$^+$, [M+H]$^{2+}$) are shown for SNAC-surugamide A (red, $C_{52}H_{90}N_{10}O_9S$) and surugamide A (blue, $C_{48}H_{81}N_9O_8$). Exact neutral m/z values used to determine EICs were: 911.6208 and 1030.6613. The intensity scale for all EICs is 1 x 10^6. B)

LC-HRESIMS analysis of *in vitro* reaction of the mutant form of SurE with SNAC-surugamide A. No production of cyclic surugamide A was observed.

SurE was also incubated with SNAC-surugamide F (Figure 5.14). As expected, no changes were observed in the absence of the enzyme or the substrate. Accumulation of a linear, hydrated surugamide F species was observed under assay conditions, showing that SurE was also likely responsible for the simple hydrolytic offloading of surugamide F from the assembly line, as observed *in vivo*. Interestingly, a dehydrated, cyclised form of surugamide F was also observed. Initially, this result was discounted as low-level non-catalytic or spontaneous cyclisation. However, two different groups published on the same subject at a similar time as this work had been completed, and both detected the presence of cyclosurugamide F in SurE assays using the same substrate (Matsuda et al., 2019; Zhou et al., 2019). However, one study showed accumulation of cyclosurugamide F as the major product, with the other showing the cyclic form as a minor, side product. In all scenarios, however, SurE was found to be essential for the release of surugamide F from the NRPS assembly line, with the significance of the cyclic form of surugamide F yet to be determined.
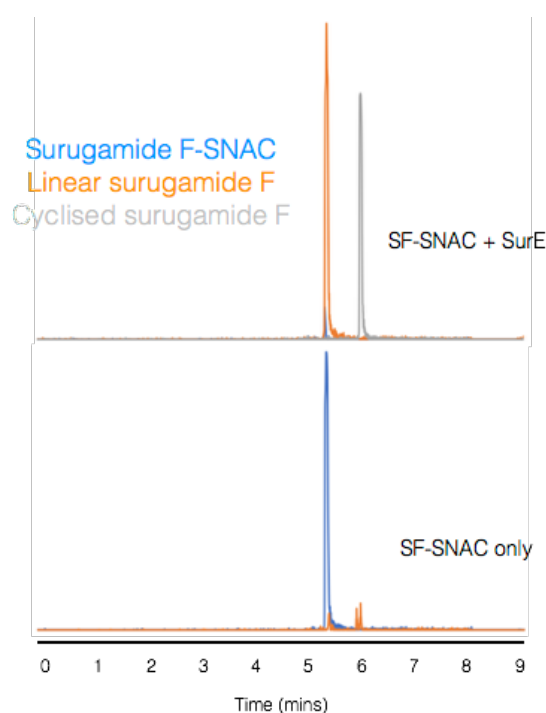
Figure 5.14: LC-HRESIMS analysis of *in vitro* reactions of SurE with a SNAC-surugamide F mimic. EICs ($[M+H]^+$, $[M+Na]^+$) are shown for SNAC-surugamide F (blue, $C_{55}H_{90}N_{12}O_{12}S$), cyclised surugamide F (grey, $C_{51}H_{81}N_{11}O_{11}$), and surugamide F (orange, $C_{51}H_{83}N_{11}O_{12}$). Exact

neutral m/z values used to determine EICs were: 1145.6522, 1023.6117, 1041.6223. The y-axis was set to 7×10^5 for each of the chromatograms.

Finally, SurE was incubated with truncated forms of surugamide A and F (Figure 5.15). Although, SurE utilises substrates eight and ten amino acids in length, linear or cyclised product could not be detected when the enzyme was incubated with our truncated surugamide A hexapeptide substrate. This could indicate a lack of activity on peptides of a shorter length than eight amino acids or it could be due to the specific composition of the peptide used in the assay. Interestingly, when SurE was incubated with the truncated surugamide F mimic (last two amino acids excluded), only a linear product was detected, suggesting SurE exclusively catalysed the hydrolytic release of the peptide substrate. It is also possible that the peptide sequence itself lends itself to hydrolytic release and formation of a linear product, rather than a macrocyclised product. Overall, it is these data that suggest that SurE is indeed capable of offloading peptides from assembly lines using both macrocyclisation and hydrolysis as release strategies.
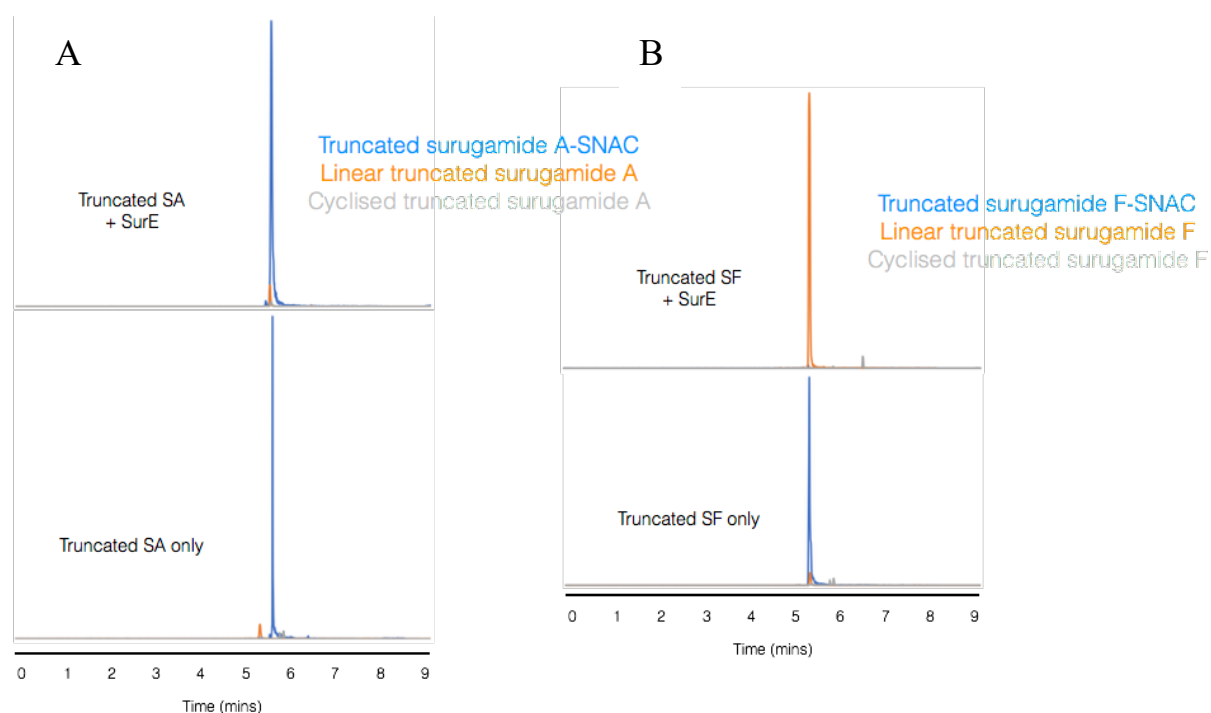
Figure 5.15: LC-HRESIMS analysis of *in vitro* reactions of SurE with truncated SNAC-surugamide mimics. A) Analysis of assays of SurE with a truncated surugamide A-SNAC peptide. EICs ([M+H]$^+$, [M+Na]$^+$) are shown for SNAC-truncated surugamide A (blue, $C_{40}H_{67}N_7O_7S$), linear truncated surugamide A (orange, $C_{36}H_{60}N_6O_7$), and cyclised truncated

surugamide A (grey, $C_{36}H_{58}N_6O_6$). Exact neutral m/z values used to determine EICs were: 789.4823, 688.4523, 670.4418. The y-axis was set to 3.5 x10^5 for each of the chromatograms. B) Analysis of assays of SurE with a truncated surugamide F-SNAC peptide. EICs ([M+H]$^+$, [M+Na]$^+$) are shown for SNAC-truncated surugamide F (blue, $C_{47}H_{76}N_{10}O_{10}S$), linear truncated surugamide F (orange, $C_{43}H_{69}N_9O_{10}$), and cyclised truncated surugamide F (grey, $C_{43}H_{67}N_9O_9$). Exact neutral m/z values used to determine EICs were: 972.5467, 871.5167, 853.5062. The y-axis was set to 3.5 x10^5 for each of the chromatograms.

5.3.6 Synthesis of PEGA-surugamide substrates

The use of acrylamide poly-ethylene glycol (PEGA) resins allows the fast determination of the substrate tolerance of certain enzymes by use of specifically selected linker and substrate portions. The PEGA resin itself is stable to the strong acids and bases utilised in SPPS methodologies, so allows formation of substrates with minimal loss or effects on the resin. The resin also has low to medium level peptide loading capacity, ideal for presentation of substrates to enzymes in downstream assays without hindering access of proteins to the tethered peptides. Initially, a linker region was synthesised on the resin, primarily following the previously reported synthesis in the usage of such resins for enzymatic assays (Kohli et al., 2002). The linker consisted of suberic acid, β-alanine, and ethanolamine, giving the complete linker a very similar chemical structure to the 4'-phosphopantetheinyl group utilised *in vivo* to tether substrates and intermediates to cognate NRPS assembly lines (Figure 5.16). The only major difference between the synthesised linker and the 4'-phosphopantetheinyl group was the terminating alcohol, which would lead to ester linked substrates, rather than the sulfur containing thiol functionality. The presence of the alcohol allowed a much simpler downstream substrate synthetic route, due to the relative ease of forming ester linked peptides, whereas thioester linked substrates would have provided issues with the addition of further amino acids to the substrate using regular SPPS conditions. SurE would be tested on ester linked substrates, initially on the wildtype surugamide A peptide, to confirm offloading and cyclisation was a viable strategy for ester linked substrates, as well as for release of PEGA-resin tethered substrates.
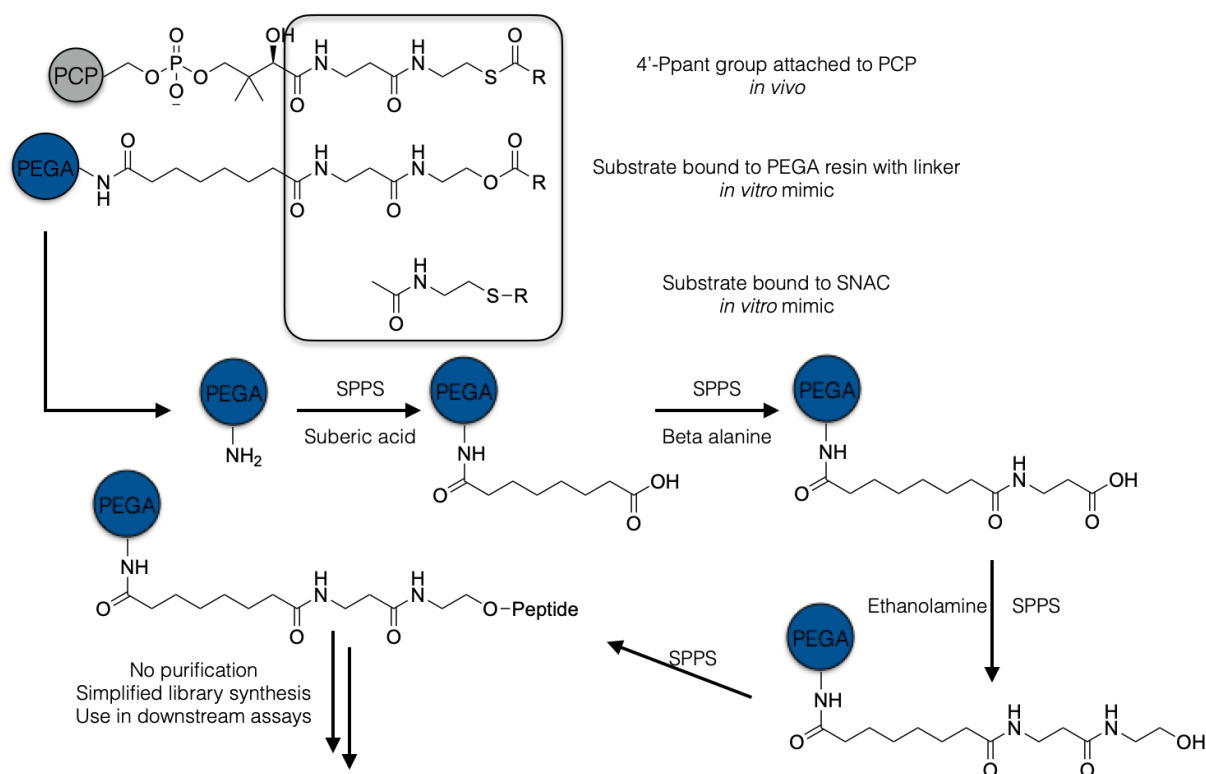
Figure 5.16: Overview of the synthetic route towards PEGA-linked substrates and the similarity of the linker region to the 4'-Ppant group utilised in *in vivo* biosynthetic systems. PEGA-linker-substrates potentially provide faster routes to the testing of biosynthetic enzymes due to the *in vivo* similarity, ease of synthesis, and the lack of obligate requirement for the utilisation of extensive purification methodologies.

PEGA-resin linked substrates would theoretically allow a much quicker validation of the substrate scope of SurE, primarily due to the ease of synthesis and minimal purification required of the complete compound. However, the methodology was, as yet, only verified for the usage of the *cis*-encoded TE domain from the tyrocidine biosynthetic pathway, and therefore, required verification and validation of functionality on the novel standalone release factor from surugamide biosynthesis. The wildtype surugamide A octapeptide was synthesised onto the linker domain, tethered to the PEGA-resin. These substrates were used without further modification or purification, and were employed directly in enzyme assays to determine the viability of the methodology, which, if successful, would allow the rapid synthesis and testing of modified peptide substrates for a complete determination of the scope of SurE substrate tolerance and acceptance.

5.3.7 Assays of SurE with PEGA based substrates

PEGA-based substrates were used in substrate assays with the recombinantly produced $(His)_6$-SurE enzyme in a manner similar to that used for SNAC-substrates. After incubation of the enzyme with the PEGA-linked surugamide A peptide for an extended length of time to attempt to allow maximum conversion of substrate into product (5 µM enzyme, 10 mg loaded resin, ~ 5h), showed the accumulation of an ion with an *m/z* consistent with that of the cyclised form of surugamide A (Figure 5.17). This showed that PEGA-linked peptides could indeed be used as viable substrates for offloading and cyclisation assays with SurE. Shorter incubation times were also tested, but resulted in much lower levels of product formation and peptide offloading. Interestingly, the appearance of a product with a mass corresponding to the linear, hydrated form of surugamide A was also observed at the end-point of the assay, potentially indicating that the reaction conditions were not completely suitable for cyclisation, therefore leading to hydrolytic offloading of surugamide A peptides. In order to confirm that the products observed from the assay were not results of random offloading by the enzyme, control peptides are required upon which SurE would not be expected to be active. For example, a PEGA-linked peptide with a C-terminal L-Leu residue should show no production of cyclic peptides in the presence of SurE, as has been observed for the SNAC-conjugated peptide. Overall, the presence of the cyclised surugamide A peptide at the end of the assay was a promising indicator for the viability of this simplified PEGA-resin based assay for probing offloading and cyclisation by the SurE release factor. However, the generally low levels of product formed indicates that whilst PEGA-substrates are viable, they may not be the ideal substrates for these assays, and require further optimisation.
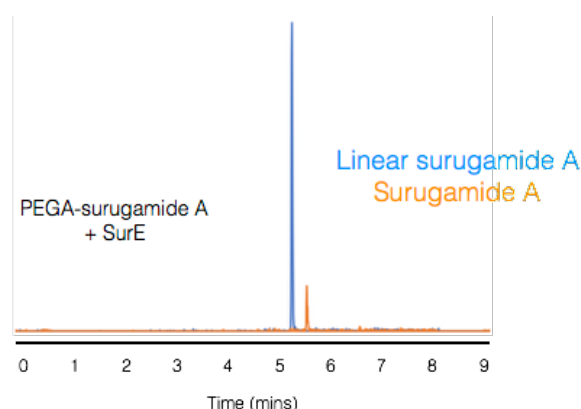
Figure 5.17: LC-HRESIMS analysis of metabolites produced after incubation of SurE with PEGA-surugamide A. EICs ($[M+H]^+$, $[M+Na]^+$) are shown for linear surugamide A (blue, $C_{48}H_{83}N_9O_9$) and cyclised surugamide A (orange, $C_{48}H_{81}N_9O_8$). Exact neutral m/z values used

to determine EICs were: 929.6314, 911.6208. The y-axis was set to 5 x10^4 for each of the chromatograms.

5.3.8 Attempted structural determination of SurE

As SurE was a novel enzyme functioning to offload and cyclise peptides from the surugamide NRPS assembly line, structural information about the protein itself, as well as the protein in complex with substrates would provide immense levels of detail with regards to the activity and catalytic cycle of the enzyme. Detailed structural information would allow confirmation and identification of important structural motifs and catalytic residues within the protein, as well as give more specific information on the nature of cyclic peptide formation within the enzyme. The protein structure would also make future efforts to engineer the enzyme's substrate more rationalisable, as well as make easier any efforts to interrogate potential interactions made with the assembly line during offloading. To this end, multiple crystal trials were set up using various concentrations of the recombinantly produced SurE protein, ranging from 7 – 12 mg ml^{-1}, with the Qiagen JCSG Core I – IV crystal screens. A number of crystal-like species were formed and observed, however, the vast majority of these returned no typical protein UV$_{280}$ signal, indicating that these crystals were likely formed of crystal screen or buffer salts. A small proportion of the crystals returned vague UV$_{280}$ signals, and whilst the crystals were small in size, were sent to the Diamond Synchrotron for X-ray shooting to determine if protein was present and if crystallisation conditions required optimisation. These crystals, when shot with X-ray radiation however, did not give a diffraction pattern representative of the presence of a proteinaceous sample. Nonetheless, efforts to find a suitable crystal condition and subsequently solve the structure of the protein were yet ongoing.

Whilst work was ongoing to attempt to solve the structure of the protein, a study was published containing a 2.2 Å resolution structure of SurE (Figure 5.18) (Matsuda et al., 2020). The crystal structure showed that SurE was formed of two distinct domains, an α/β hydrolase fold domain, typical of type I and II TE domains, and a C-terminal lipocalin domain, linked by an extended unfolded linker region. SurE was found to contain a wider and shallower binding site, compared to other structurally similar β-lactamase enzymes, whilst the positioning of the catalytic tetrad (Ser-Lys-Tyr-His) residues are mostly conserved. The C-terminal lipocalin domain and unfolded linker regions were hypothesised to play important roles in substrate binding and catalysis, and likely underwent conformational changes in response to the binding

of peptides within the active site cleft. Computational modelling of a simplified peptide substrate (N-formyl- D-Leu) mimicking the C-terminus of linear surugamide A showed that the D-amino acid was accommodated well within the active site and made preferential interactions close to the catalytic residues and with residues within the binding site, such as a conserved Arg residue. Modelling of a L-Leu containing compound, however, showed that L-configured amino acids are likely excluded from the binding pocket due to steric clashes, therefore abrogating any potential SurE catalytic activity. To confirm the modelling observations, peptide substrates upon which SurE was not functional were soaked into crystals of the protein, and the crystals structure was solved. Electron density was observed for the terminal D-Leu residue, yet only faint density was observed for the remainder of the peptide. This structure showed the D-Leu residue in very close proximity to the catalytic Ser residue of SurE, with altered conformations of the lipocalin and linker regions in close proximity with the substrate, showing these regions likely had an important role in catalysis of peptide cyclisation.
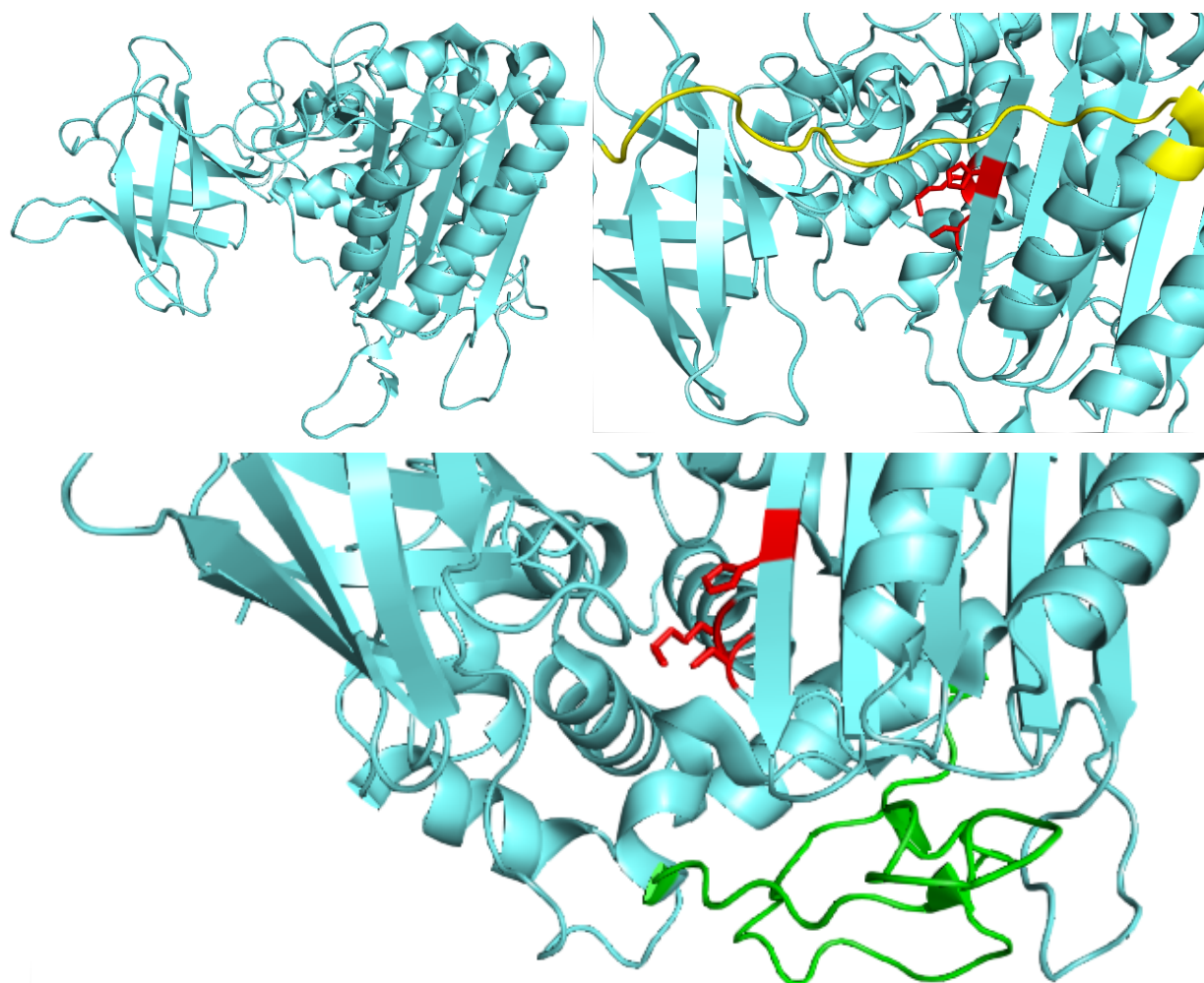
Figure 5.18: Structural analysis of SurE (PDB ID: 6KSU, (Matsuda et al., 2020)). Upper left: Overview of the monomeric structure of SurE. An N-terminal PBP-type domain and C-terminal lipocalin domain constitute the protein structure. Upper right: Zoomed view of SurE,

indicating the Ser-Lys-His catalytic residues (red). The loop region linking the two domains of the protein is shown in yellow. Bottom: The active site of SurE and the flexible loop region (green) important in the catalysis of cyclisation by SurE. The loop region is capable of folding into the active site and occupying hydrophobic pockets, thereby mediating substrate binding and catalysis. Catalytic residues are shown in red.

5.3.9 Production and purification of domains from the SurC and SurD NRPS assembly lines

SurE was shown to catalyse the cyclisation and offloading of multiple peptides from the surugamide assembly lines. Two theoretical methods of peptide offloading could be employed by the offloading cyclase (illustrated in Figure 5.19): a methodology that involves interaction of SurE with the protein assembly line, which then facilitates release of mature surugamides; or a methodology that is substrate driven, where SurE forms no specific interactions with the NRPS and only recognises its chemical substrates for peptide cyclisation. In order to maintain a high level of specificity and fidelity of the offloading reaction, it follows that SurE would form specific protein-protein interactions (PPIs) with the NRPS assembly line, whilst recognising a specific substrate before catalysing peptide cyclisation. However, SurE functionality has been observed on SNAC-peptide mimics of substrates tethered via 4'-phosphopantetheinyl linkers and no specific helper proteins or PPIs were required for activity. In order to test the potential PPIs occurring between SurE and the surugamide pathway NRPS modules, the production of domain and multi-domain proteins from the SurC and SurD NRPS assembly lines was carried out in *E. coli* to attempt subsequent *in vitro* interaction assays. The terminal modules of both SurC and SurD have the domain organisation: C-A-PCP-E, allowing activation of the final substrate amino acid by the A domain, switching of the stereochemistry to the D- configuration by the E domain, and, addition to the upstream polypeptide by the C domain. In total, 10 domain and multi-domain constructs were cloned into *E. coli* expression vectors and trialled for soluble protein expression in the BL21-Gold expression system. The constructs made were as follows: C-A-PCP-E, A-PCP-E, PCP-E, A-PCP, PCP for the terminal modules of each surugamide synthesising system, SurC and SurD. Only the PCP and PCP-E proteins from the SurC terminal module (surugamide F synthesising) were produced as soluble protein under the simple protein production conditions trialled, whilst the remainder of the SurC and all of the SurD constructs produced the required protein solely in the insoluble fraction. The identity of the soluble SurC PCP (11345.63 Da) and PCP-E (59534.42 Da) proteins was confirmed by SDS-PAGE after purification by IMAC using Ni-NTA resin and

subsequent gel filtration chromatography (Figure 5.19). Upon confirmation of protein production, the proteins were buffer exchanged into a suitable, minimal buffer for use in downstream assays.
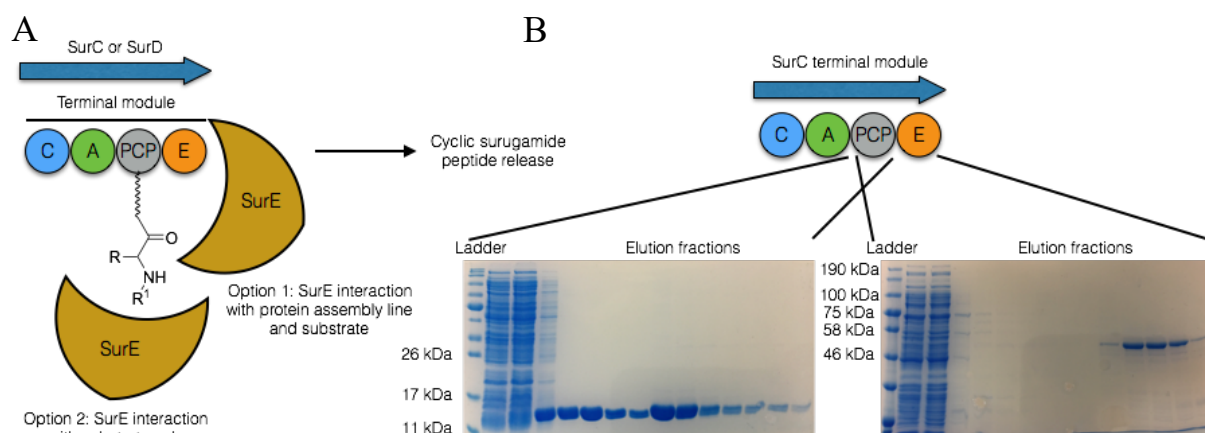
Figure 5.19: A) Overview of potential interactions that could mediate offloading and cyclisation of terminal surugamide intermediates by SurE. Option 1 illustrates a mode of SurE action dependent on protein-protein interactions with the main NRPS assembly line, whilst option 2 depicts SurE offloading mediated by recognition of its chemical substrate only. B) In order to investigate option 1 *in vitro,* a number of proteins were cloned and production was attempted. Soluble forms of the terminal SurC-PCP domain, as well as the PCP-E didomain were successfully produced and purified.

5.3.10 Preliminary analysis of interactions between SurE and the SurC/SurD assembly lines

To initially examine the potential occurrence of protein-protein interactions between the offloading cyclase, SurE, and the terminal module of the NRPS assembly line during peptide cyclisation, an initial *in vitro* methodology was utilised. The SurE, SurC PCP, and SurC PCP-E proteins were produced in *E. coli* and purified to homogeneity for *in vitro* assays. The proteins were separately buffer exchanged into identical buffer solutions and initially used in an analytical gel filtration assay, at a 1:1 concentration ratio. The proteins were mixed (SurE with SurC PCP or SurE with SurC PCP-E) and incubated for one hour prior to injection onto the gel filtration column, and proteins or protein complexes were eluted from the column using one column volume of gel filtration buffer. Analysis of the resultant elution profile showed the appearance and elution of two separate peaks in both gel filtration runs involving SurE with either SurC PCP or SurC PCP-E. This was eminently visible in the SurE/SurC PCP experiment where the two proteins are of very different masses and sizes (SurE: 49764.00 Da, SurC PCP

11345.63 Da) but was yet discernible in the assay of the two similarly sizes, larger proteins SurE and SurC PCP-E. This initially showed there was no interaction between the terminal portions of the SurC assembly line and the peptide offloading cyclase, SurE. Furthermore, ITC assays were carried out with SurE and the two assembly line proteins, utilising the same buffer exchanged stocks of purified proteins. SurE, at a concentration of 50 μM, was held in the cell, whilst the assembly line protein (SurC PCP or SurC PCP-E), in a ten times excess at a concentration of 500 μM, was automatically titrated into the cell. Processing and analysis of the resulting raw thermogram showed no trace representative of the occurrence of an interaction and no formation of a protein-protein complex (Figure 5.20). Instead, peaks indicative of a heat of dilution were observed in the processed thermograms, as the small volume injections of the assembly line proteins were diluted into the buffer of the cell component with no additional heat released from complexation or interaction of the injected protein with the cell contained, SurE protein.
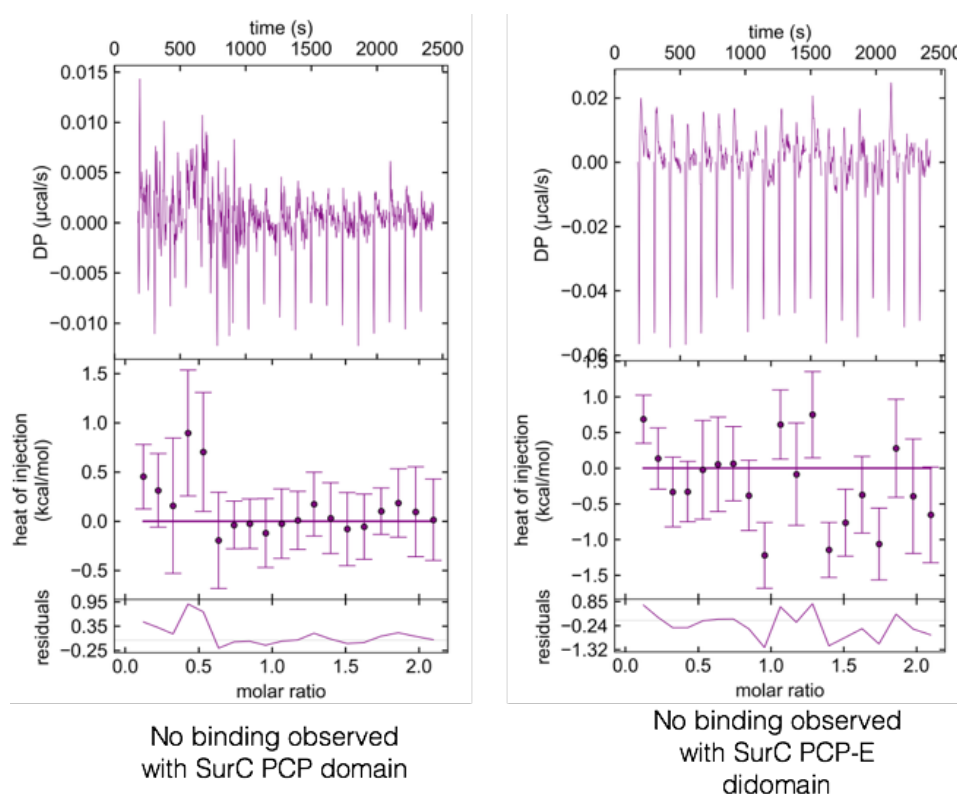
Figure 5.20: ITC assays between SurE and the PCP or PCP-E domains. No binding interaction between the two proteins was observed in each assay.

These *in vitro* assays showed that no interaction complexes formed between SurE and SurC PCP or SurC PCP-E domain proteins, under the conditions utilised. This initially suggests that

SurE does not interact with assembly line proteins when catalysing the release and cyclisation of surugamide proteins from the SurC and SurD terminal NRPS modules. This would then imply that SurE only recognises its substrate and catalyses the subsequent cyclisation of assembly line tethered peptides. The finding that peptides of both eight and ten amino acids in length are offloaded by SurE suggests that the cyclase has a level of inherent promiscuity in terms of substrate acceptance, accompanied by the fact that numerous isoforms of surugamide A are produced *in vivo*. This added further credence to the idea that specific PPIs would ensure the correct functionality of SurE in terminating the synthesis of surugamide peptides only, and potentially act as a restriction on the natural promiscuity of the cyclase. However, *in vitro* functionality on SNAC-peptide mimics and the lack of complex formation in assays utilising purified proteins, shows that specific substrate recognition could be the sole or primary determinant in SurE function. The lack of interaction observed between SurE and SurC NRPS proteins could also have been due to the nature of the assembly line proteins. The proteins were used as purified, with mass spectral analysis showing that the proteins were in the *apo* form, and lacking the 4'-phosphopantetheinyl group of *holo* proteins or the subsequently tethered amino acid of loaded proteins. Usage of either *holo* or loaded forms of the PCP and PCP-E proteins may have allowed interaction of the cyclase with the NRPS proteins to occur, whilst utilisation of PCP domains with amino acids or peptides tethered via 4'-phosphopantetheinyl groups for substrate utilisation, protein-protein interaction, and functional analysis assays would be the most ideal and representative biomimetic system available, as SurE functionality *in vivo* very likely occurs in this manner.

5.3.11 Bioinformatics of SurE-type cyclases

The mannopeptimycin BGC, responsible for the synthesis of the titular highly derivatised lipoglycopeptide, encodes a protein with sequence similarity to SurE and potentially with the requisite activity to offload and cyclise the terminal, linear biosynthetic intermediate (Magarvey et al., 2006). The hypothesis, although not rigorously tested or evaluated, posited that MppK was the cyclase and release factor, and indispensable for the production of mature mannopeptimycins. Simple BLAST searches of the SurE amino acid sequence led to the identification of numerous homologues of the protein, including MppK, whilst putative orthologues have been identified within the desotamide and ulleungmycin BGCs (Fazal et al., 2020). As numerous characterised BGCs contained a homologous gene to *surE* and the predicted function of each of the putative proteins was confidently predicted as a standalone, terminating cyclase, it was realised that the offloading mechanism employed by the surugamide

system could be widespread throughout numerous other NRPS biosynthetic pathways. Presently however, the only other experimentally verified β-lactamase type protein acting as an NRP offloading cyclase is from the desotamide pathway, specifically *dsaJ* from the desotamide BGC (Ding et al., 2020). DsaJ shares 43% amino acid sequence identity to SurE and also contains the characteristic Ser-Lys-Tyr-His catalytic tetrad that defines the enzymatic activity of the group of proteins (Figure 5.21). Surprisingly, DsaG, the protein harbouring the final two NRPS modules required for the biosynthesis of the desotamides, was initially observed to terminate in a canonical TE domain, thought to be responsible for mature peptide release and concomitant cyclisation. However, upon re-analysis of the desotamide BGC by multiple groups and specifically the domains present within the final NRPS module, DsaG was found to terminate in an E domain, rather than the offloading TE domain (Li et al., 2015). This aligned with the apparent requirements of the presence of an offloading cyclase, and DsaJ was subsequently found to be required for the *in vivo* formation of desotamides of any variety or composition, validating its functional similarity to SurE.

Figure 5.21: A) The last stage of desotamide A biosynthesis and the pathway requirement for the standalone cyclase; DsaJ. No products are observed in the absence of the cyclase enzyme. B) Pairwise sequence alignement of SurE (top) with DsaJ (bottom), the only other experimentally characterised standalone cyclase enzyme. The proteins exhibit 51.4 % sequence

similarity. The catalytic tetrad (Ser, Lys, Tyr, His) are conserved and their respective locations are boxed within the alignments.

The identification and confirmation of SurE as the standalone offloading cyclase for the surugamides, laid the foundation for the *in silico* observation of proteins from other NRPS pathways, catalysing the same chemical reaction and having the same enzymatic activity. The observation of homologues of SurE in other known and characterised NRPS pathways led to the question of how widespread this offloading and cyclisation strategy could be. A bioinformatic pipeline was constructed and carried out to address this question, with the work of database curation, pathway analysis, and similarity network creation carried out by Dr Ryan Seipke. From a manually curated database of 1421 filamentous actinobacterial genomes available from GenBank (Clark et al., 2016), organisms harbouring orthologues of the SurABCD NRPSs and SurE were identified. Candidate genomes were then analysed with antiSMASH to identify NRPS biosynthetic systems without a cis-encoded TE domain, but that possessed an adjacent SurE homologue (Blin et al., 2019). A total of 166 organisms contained at least one NRPS system employing a standalone cyclase offloading strategy. These BGCs were subsequently analysed with BiG-SCAPE (Navarro-Muñoz et al., 2020) to generate a BGC similarity network comprising 15 related subnetworks and 12 singletons (Figure 5.22) (Thankachan et al., 2019). Upon inspection of the network it became apparent that the number of biosynthetic modules in an NRPS system (4-10 modules) was a major factor influencing formation of subnetworks, showing that SurE-type offloading cyclases may be able to cyclise peptides ranging in length from four to ten amino acids, with wide varieties of amino acid compositions, useful for downstream chemical biology and engineering applications. Interestingly, the desotamide BGC was not resolved within the network, suggesting a distant relatedness between this BGC and other pathways leading to the production of peptides of a similar size and composition. SurE and its homologues, however, were present amongst a large proportion of BGCs and genomes, showing that the offloading cyclase based termination strategy is a widespread mechanism where the majority of the compounds produced by BGCs harbouring the mechanism are unknown. Discovery of these pathways and compounds, either prior or subsequent to the identification of the pathway encoded offloading cyclase, would expand the repertoire of available cyclase enzymes for interesting uses in *in vivo* and, particularly, *in vitro* chemical biology applications.
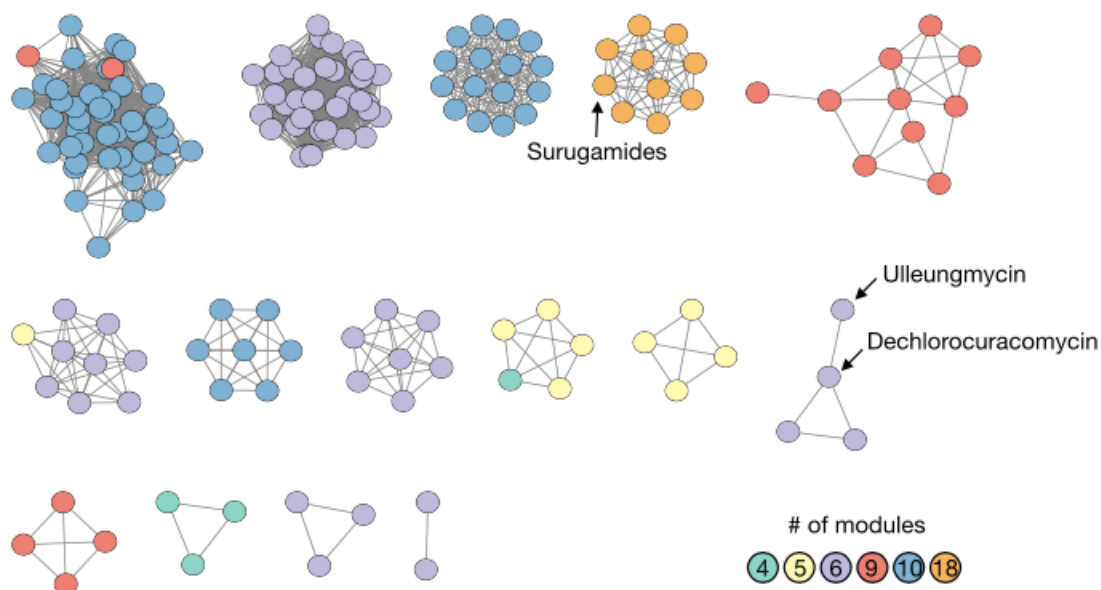
Figure 5.22: Gene cluster similarity network for NRPS BGCs without an embedded, *cis*-encoded TE domain, but encoding an orthologue of the *surE* gene. Each node represents a disparate BGC and is colour-coded based on the number of putative biosynthetic modules encoded by the BGC.

5.4 Discussion and conclusions

The canonical paradigm of non-ribosomal peptide synthesis by NRPS megasynthases has been modular synthesis by individual enzymatic domains catalysing consecutive peptide elongations, followed by mature compound release by a physically connected, *cis*-acting thioesterase domain. The manner of TE domain-mediated peptide release is relatively well conserved across most NRPS systems, however a number of exceptions have been discovered among systems containing alternative enzymatic strategies, such as modified condensation domains or reductase domains for peptide offloading (Du and Lou, 2010). Spontaneous, non-enzymatic offloading of mature compounds from terminal PCP domains has also been observed in NRP biosynthesis, however this strategy is rare and not seemingly evolutionarily preferred. Bioinformatically, the NRPS systems directing the formation of the surugamides lack canonical C-terminal release factors, instead terminating in epimerase domains, responsible for the switching of stereochemistry of the final amino acid. The lack of a dedicated, *cis*-encoded domain responsible for the termination of peptide synthesis led to curiosity about the method of peptide release employed by the system.

Bioinformatic analysis of the complete surugamide pathway led to the identification of two possible trans-acting enzymes, potentially with the requisite activity to catalyse chain release and mature surugamide formation. The *in vivo* results presented showed that SurE, and not SurF, was the dedicated domain for the formation and release of cyclic peptides. SurF was observed to be a type II TE by the reduction in extracted compound titres of both the cyclic surugamide A and the linear peptide, surugamide F. Type II TEs generally catalyse the removal of aberrant amino acids and peptide intermediates from assembly line PCP domains, which would otherwise result in stalling or termination of synthesis and the subsequent reduction in efficiency and rate of product formation. SurF, however, was initially inconsistent with this functionality as it contained an altered hydrolase signature motif (GTSLG) when compared to established, well characterised proof-reading thioesterases (GHSMG) and also belonged to a different InterPro family, IPR029058 rather than IPR1223. The reduction in produced compound titres, as well as the return to wildtype production levels upon *in trans* supplementation of the enzyme, confirmed that SurF likely acts as the type II TE for the surugamide system. The observed bioinformatic variables may indicate an altered method of enzymatic action or protein structure, and will be investigated further. SurE was identified as another potential release factor for the surugamide peptides, and *in vivo* data confirmed this was the case by complete abolition of production of both the cyclic and linear surugamides upon deletion of the *surE* gene, and restoration of production upon *in trans* complementation. This represented a novel method of enzymatic NRP release from assembly lines and allowed the relatively facile determination and discovery of the desotamide family of antibiotics; a collection of related peptide antibiotics, whose NRPS systems employ the same standalone cyclase offloading mechanism (Fazal et al., 2020).

Upon identification of SurE as the dedicated standalone cyclase for the surugamide system, and as a completely novel enzymatic methodology for the release of cyclic NRPs from their cognate assembly lines, attention was turned to ways of better understanding the structure and functionality of the protein and the way these parameters pertain to cyclic peptide formation. The protein was purified in sufficiently large quantities and to a good level purity for attempts at forming protein crystals, however efforts to obtain a crystal structure were unsuccessful. After the conclusion of these initial crystallisation attempts, a high-resolution crystal structure of SurE was published and allowed detailed insights into the molecular controls underpinning efficient catalysis (Matsuda et al., 2020). The requirement for a C-terminal D-configured amino

acid, as observed *in vivo* for both surugamides and confirmed by the presence of an E domain in the final module, was rationalised structurally as part of the requirement of heterochirality at the termini of peptidic substrates. The peptide C-terminal residue is held in a hydrophobic pocket of SurE, with an Arg residue forming hydrogen bonds to the backbone carbonyl of the substrate. The Arg residue is also important in excluding peptides terminating in L-amino acids due to the formation of large and disruptive steric clashes, which culminate in the orientation of the terminal amino acid away from catalytic residues and out of the pocket. The structural rationale also paves the way for additional efforts into engineering of the substrate scope of the protein, or directed engineering of the protein itself. Whilst the enzyme is not completely indiscriminate towards amino acids at the centre of the substrate peptide, preliminary analysis of the substrate scope has shown that alterations are tolerated, making SurE an attractive biocatalyst for the production of cyclic peptides ranging from eight to ten amino acids in length.

N-acetylcysteamine thioesters have been used over the course of a number of years as simplified mimics of the natural 4'-phosphopantetheinyl moiety, used as the ~ 20 Å flexible linker between the PCP domain and the synthesised peptide (Trauger et al., 2000). As NRPS pathways generally use this thiotemplating methodology for the synthesis of bioactive compounds, enzymes involved in NRPS systems, particularly standalone enzymes, seem to have evolved to recognise parts of the thioester linker region or only display functionality when peptide substrates are bound to PCP domains. Therefore, providing a short thioester-based handle, added onto peptides via a peptide coupling reaction, allows assaying of enzymatic activity without the need for extensive chemical synthetic processes. A number of SNAC-peptide substrates have been trialled with SurE to attempt to characterise the substrate scope of the enzyme. Alterations at internal positions of the substrate peptide were seemingly well tolerated, albeit with a titre reduction in cyclic product formation. Changing of the C-terminal amino acid to D-Phe led to efficient product formation, however, alterations to D-Ala, Gly, or an L-configured amino acid led to no or aberrant product formation (Matsuda, Kobayashi, et al., 2019; Zhou et al., 2019; Matsuda et al., 2020). In this work, the parameter of substrate length was tested and SurE showed no cyclic product formation from a SNAC-thioester mimic of a truncated, hexapeptidic form of surugamide A, indicating that SurE is at least partly selective for substrates eight or ten amino acids in length. The wildtype and truncated, octapeptidic forms of surugamide F were also tested with the SurE cyclase. Both assays showed the accumulation of a major linear product, showing that SurE could potentially catalyse hydrolytic release of certain peptide substrates. Further study, however, into the surugamide F

offloading mechanism, and biosynthesis in general, is required in order to completely characterise the system. PEGA-based substrates, peptide substrates synthesised on solid-phase peptide synthesis beads with a linker mimicking the naturally observed 4'-phosphopantetheinyl moiety, were also tested for functionality with SurE. The standalone cyclase enzyme was capable of forming cyclic, mature surugamide A from the tethered peptides, indicating that PEGA-tethered peptides can be utilised as substrates for cyclisation assays, a methodology that would greatly reduce the time and effort required in the synthesis and purification of substrates for SurE cyclisation assays and would potentially allow the testing of multiple different substrates quickly and efficiently.

SurE functionality resulting in the efficient cyclisation of linear surugamide peptides also depends on the ability of the enzyme to effectively recognise its assembly line-tethered cognate substrate in an *in vivo* environment. Therefore, efficient formation of an enzyme-substrate complex or the formation of beneficial protein-protein interactions between the cyclase and the terminal NRPS module holding the final intermediate, are a prerequisite for the productive functionality of SurE. Interactions between different proteins involved in non-ribosomal peptide and polyketide biosynthetic systems is vitally important in the functionality of the individual proteins, as well as the ultimate synthesis of the required bioactive compound (Weissman and Müller, 2008; Izoré and J. Cryle, 2018). Much of the research in this area has focussed on elucidating the interactions occurring between NRPS modules held on different polypeptides, which require the dynamic capability to form PPIs allowing the transfer of intermediates between modules. The identification of communication mediating (COMM) domains, as well as specific interaction motifs in PKS and hybrid systems, has shown, in part, the mechanisms utilised for modular interactions (Hahn and Stachelhaus, 2004). However, the identification of dedicated standalone domains catalysing important chemical reactions and their functionality in the context of NRPS assembly lines, has added an extra level of complexity to the process of NRP synthesis, and the details regarding the formation of interaction complexes is mostly unknown. SurE showed no interaction *in vitro* with purified proteins from the terminal module of the SurC assembly line. The PCPs used in these assays were in the *apo* form, and contained no tethered substrate or 4'-phosphopantetheinyl linker, which could be prerequisites for the formation of any interaction complex.

Although much remains to be discovered regarding the specific enzymatic functionality of SurE, the overall mechanism of cyclic NRP release by a standalone cyclase domain is relatively

widespread. Bioinformatic analysis of a large collection of filamentous actinobacterial genomes showed that ~ 11% of organisms potentially contained an NRPS system with a standalone cyclase offloading strategy for the synthesis of cyclic, bioactive compounds. This showed the widespread nature of the release strategy, although the majority of the pathways identified in the analysis were undiscovered or unknown. The bioinformatic analysis did, however, allow the determination and definition of the desotamide family of antibiotics; known cyclic compounds formed by NRPS pathways with a standalone cyclase offloading strategy. The compounds forming the desotamide family (desotamides, wollamides, surugamides, ulleungmycins, noursamycins, curacomycin) range in length from six to ten amino acids in length and are comprised with a variety of unnatural amino acids, leading to highly modified and varied cyclic scaffolds, with unknown or interesting bioactivities.

The pharmaceutical potential displayed by many NRPs, alongside the challenges of antimicrobial resistance and shortages of drugs to treat a variety of other diseases, has led to an increase in the discovery of novel NRPS pathways as well as a heightened interest in the engineering of existing synthetic platforms for the creation of bioactive products. To achieve the efficient engineering of natural product biosynthetic pathways, a varied enzymatic toolkit is required in order to provide the varied chemistries for the creation of new scaffolds. This work has contributed to the discovery and further characterisation of the standalone cyclase enzyme from the surugamide BGC, SurE, and established it as an enzyme capable of cyclising and releasing peptides from NRPS pathways or from synthetic substrate mimics. SurE has the potential to be a versatile biocatalyst, which naturally cyclises peptides eight and ten amino acids in length, and can also tolerate residue alterations at various positions of the peptidic substrate. Further characterisation of the enzyme, utilising the *in vitro* assay tools developed in this work, would allow the development of the protein as a valuable tool in the synthesis of varied cyclic peptides, products that can be difficult to synthesis using purely chemical synthetic methods. SurF was also identified as the type II thioesterase for the surugamide system, whilst bioinformatic analysis showed that both SurE and SurF were interesting enzymes, with various differences from characterised enzymes carrying out the same function is disparate pathways. Most interestingly, homologues of SurE were found to exist in a large number of undiscovered pathways. With access to the genomes of talented producing organisms becoming increasingly available, as well as the increased usability and versatility of genome mining tools, discovery of novel pathways synthesising interesting, bioactive

compounds, alongside the discovery of new enzymology in the context of specialised metabolite biosynthesis, could be hugely accelerated or, potentially, be expected.

Chapter 6 – General discussions and perspectives

The sequencing of the first *Streptomyces* genome, that of *Streptomyces coelicolor* A3(2) in 2002, and the observation that the biosynthetic potential of this microorganism was much greater than initially hypothesised, kick-started the genomics era of antibiotic discovery (Bentley et al., 2002). Genome mining, now a near-ubiquitous term in the natural product field, involves the analysis of large volumes of genome sequence data for the presence of BGCs, which could subsequently encode the requisite protein machinery for the production of bioactive compounds, with structural complexity not yet rivalled by chemical synthetic methods (Wilkinson and Micklefield, 2007). Analysis of genomic data and genome mining studies have indicated the vast and unexplored nature of the compounds naturally produced. As the majority of clinically used antibiotics are derived from natural products, the exploration of the untapped remainder of this versatile source of bioactive compounds could yield extremely medically important compounds (Newman and Cragg, 2016).

Natural product BGCs usually contain the requisite genetic potential to encode the enzymes required for secondary metabolite production. Whilst the functionality of many of the most important biosynthetic proteins (such as the assembly line-like NRPS and PKS proteins) is relatively well-studied, many enzymes involved in the biosynthesis of secondary metabolites are functionally cryptic. Characterisation of such enzymes could potentially lead to the discovery of novel or interesting enzymology, or proteins with substrate scopes extended beyond that which has been previously characterised. Augmentation or alteration of the enzymes involved in natural product biosynthesis is an interesting strategy that can yield novel secondary metabolites, with altered or exchanged functional groups, potentially useful methodologies in the search for new bioactive chemical scaffolds (Winn et al., 2016).

The work in Chapter 3 describes attempts to rationally engineer the substrate scope of the naturally promiscuous AT domain, involved in the biosynthesis of antimycin. The AT domain is naturally capable of accepting various acyl-CoAs, produced by the pathway specific CCR enzyme. Mutation of motif III, a short amino acid stretch shown to be involved in substrate control, to the previously characterised YASH and HAFH motifs caused the complete attenuation of antimycin production. However, a more conservative mutation of the natural AAAH motif involving an Ala to Pro alteration at position three of the motif, led to antimycin

production *in vivo*, including the production of a C7 methylated species, not previously observed naturally. This indicated that a minimal approach to pathway engineering could potentially lead to the production of novel natural products.

The biosynthesis of the antimycins is relatively well characterised, particularly with regards to the biosynthetic roles played by the NRPS and PKS assembly lines, AntC and AntD, respectively (Sandy et al., 2012; Seipke and Hutchings, 2013b). This knowledge has permitted pathway engineering efforts, such as that undertaken with AntD[AT]. One uncharacterised aspect of antimycin biosynthesis, however, is the functionality of the standalone β-ketoreductase, AntM. The enzyme was proposed to catalyse C8 ketoreduction of the antimycin core scaffold, likely as a post-PKS tailoring enzyme. Structural and functional characterisation of the enzyme, both *in vivo* and *in vitro*, suggested that AntM functions as the antimycin C8 ketoreductase, concomitantly with biosynthesis, and therefore, on a linear, ACP-tethered intermediate. Bioinformatically, AntM was shown to not be unique in this regard, but represents the first standalone β-ketoreductase that acts during core scaffold assembly by a type I PKS system. Structural analysis shed light on the functionality of the protein and yielded interesting mechanistic insights, which could be useful in the reprogramming of this, and other standalone enzymes, for combinatorial biosynthesis. Rational pathway engineering requires high levels of structural and mechanistic detail with regards to the enzymes and pathways targeted for alteration. For many natural product biosynthetic pathways, this level of detail remains unattained, leading to the variability observed in efforts to engineer pathways and the failure of many combinatorial biosynthesis projects. High levels of specific detail regarding the functional aspects of secondary metabolite biosynthetic enzymes, would likely make any subsequent efforts to alter or augment their activity much more amenable to success. The data presented in this work, with regards to characterisation of the standalone β-ketoreductase AntM, and the AntD[AT] domain, could be of use in any future attempts to engineer the antimycin biosynthetic system.

Functionally, AntM represents an interesting class of ketoreductase enzymes, whose activity on a linear, ACP-tethered intermediate in the context of a type I PKS system, has not been previously observed. The data presented in Chapter 4, illustrate that AntM is essential for antimycin production *in vivo*, likely due to its functionality on an assembly line-tethered intermediate. The inaccessibility of the proposed linear substrate led to the usage of alternative

substrates in an assay, as well as molecular docking, which suggested the preferred substrate of AntM was the linear ACP-tethered intermediate. Obtaining a suitable linear substrate compound and unambiguously confirming the nature of AntM catalytic activity, would be an overarching goal of any future investigations. Another interesting avenue of investigation would be probing the reason for essentiality of AntM action, with regards to the formation of antimycins *in vivo*. It is an interesting hypothesis to suggest that the TE domain, responsible for cyclisation and offloading of the terminal biosynthetic intermediate, requires the action of AntM, and therefore a hydroxyl group at the C8 position, as a prerequisite for catalytic action. The fact that the hydroxyl group created by AntM catalysis is not involved in the cyclisation of antimycin-type compounds, may suggest a sort of sensing carried out by the TE domain, ensuring that no release of compounds containing a C8 carbonyl occurs, however this requires further analysis and experimental detail. Additionally, an overarching theme regarding the function of standalone enzymes in the context of natural product biosynthesis, is the potential requirement of protein-protein interactions with the main assembly line for the action of such enzymes. The use of specific substrate mimics (such as SNACs) has often abrogated the need for the purification of certain proteins (such as ACPs or PCPs) in the testing of the functionality of standalone enzymes. However, a variety of enzymes have been found to require carrier protein tethered substrates, likely meaning specific PPIs are required for catalytic activity. Probing any potential interaction formed by AntM with the sole PKS module of antimycin biosynthesis could provide interesting insights into the functional mechanism by which standalone ketoreductases catalyse chemical reactions, also extremely useful in the downstream alteration of enzymes and systems for combinatorial biosynthesis applications.

Chapter 5 describes the *in vitro* characterisation of SurE, the founding member of a novel class of standalone peptide cyclases involved in the biosynthesis of non-ribosomal peptides. SurE was observed to be essential for the biosynthesis of the surugamides *in vivo,* where it catalysed the release of peptides of two distinct lengths and composition. The cyclase catalyses the release of octapeptides from the NRPS assembly line using a cyclisation release mechanism, and decapeptides using a simple, hydrolytic methodology. This was confirmed *in vitro* by the use of SNAC-substrate mimics, which confirmed the natural substrate scope of the enzyme and the dual peptide release methodologies utilised by SurE. The observation and analysis of SurE represented a novel peptide release strategy utilised by NRPS systems, and allowed the bioinformatic unearthing of a multitude of similar enzymes, from disparate NRPS pathways, showing the widespread nature of the offloading cyclase release strategy.

The natural promiscuity of SurE immediately raises the prospect of functionality, or functionality of an engineered enzyme, on peptidic substrates of differing lengths. This was tested in part by using truncated forms of the natural surugamide peptides, which showed that SurE was active on truncated surugamide F (octapeptide), but not a truncated, hexapeptide form of surugamide A. This showed that the cyclase was selective for peptide length, but alterations to amino acid composition were tolerated, within the boundaries of incorporating a C-terminal D-amino acid and heterochirality at the peptide termini. A thorough investigation of the substrate scope of SurE would be useful in further characterisation of the enzyme, a study which would likely be aided by the use of PEGA-linked substrates, as shown in this study, making the synthesis and assaying of peptides much faster and much more facile.

Structural detail of SurE was obtained during the course of this study and illuminated some of the structural and mechanistic detail of the enzyme's functionality (Matsuda et al., 2020). Further structural detail, such as structures with bound ligands, could provide increased insight into the requirements for SurE catalysis, and potentially allow subsequent studies aimed at a long-term goal of completely switching the stereochemical control of the enzyme, as well as ultimately altering the substrate scope. SurE, much like AntM, is a standalone enzyme, whose substrate is a linear compound, tethered to assembly line-resident carrier proteins. Logically, SurE would likely interact with the assembly line protein, before or during peptide cyclisation. Probing any interactions SurE makes with the relevant assembly lines synthesising the varying surugamides, whilst experimentally difficult to achieve, would yield important insights into the mode of action of standalone enzymes. These potential insights, along with increased structural and mechanistic detail, would hugely benefit any attempts at refactoring these enzymes for the creation of novel natural products.

7 - References

Abraham, E.P., Chain, E., Fletcher, C.M., Gardner, A.D., Heatley, N.G., Jennings, M.A. and Florey, H.W. 1941. Further observations on penicillin. *The Lancet*. **238**(6155), pp.177–189.

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

Alekseyev, V.Y., Liu, C.W., Cane, D.E., Puglisi, J.D. and Khosla, C. 2007. Solution structure and proposed domain–domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Science*. **16**(10), pp.2093–2107.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*. **215**(3), pp.403–410.

Ames, B.D., Liu, X. and Walsh, C.T. 2010. Enzymatic processing of fumiquinazoline F: a tandem oxidative-acylation strategy for the generation of multicyclic scaffolds in fungal indole alkaloid biosynthesis. *Biochemistry*. **49**(39), pp.8564–8576.

Ames, B.D., Nguyen, C., Bruegger, J., Smith, P., Xu, W., Ma, S., Wong, E., Wong, S., Xie, X., Li, J.W.-H., Vederas, J.C., Tang, Y. and Tsai, S.-C. 2012. Crystal structure and biochemical studies of the trans-acting polyketide enoyl reductase LovC from lovastatin biosynthesis. *Proceedings of the National Academy of Sciences*. **109**(28), pp.11144–11149.

Aminov, R.I. 2010. A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the Future. *Frontiers in Microbiology*. **1**, pp. 1-7.

Austin, M.B. and Noel, J.P. 2003. The chalcone synthase superfamily of type III polyketide synthases. *Natural Product Reports*. **20**(1), pp.79–110.

Bachmann, B.O., Van Lanen, S.G. and Baltz, R.H. 2014. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *Journal of Industrial Microbiology and Biotechnology*. **41**(2), pp.175–184.

Baerga-Ortiz, A., Popovic, B., Siskos, A.P., O'Hare, H.M., Spiteller, D., Williams, M.G., Campillo, N., Spencer, J.B. and Leadlay, P.F. 2006. Directed Mutagenesis Alters the Stereochemistry of Catalysis by Isolated Ketoreductase Domains from the Erythromycin Polyketide Synthase. *Chemistry & Biology*. **13**(3), pp.277–285.

Bali, S., O'Hare, H.M. and Weissman, K.J. 2006. Broad Substrate Specificity of Ketoreductases Derived from Modular Polyketide Synthases. *ChemBioChem*. **7**(3), pp.478–484.

Baltz, R.H. 2008. Renaissance in antibacterial discovery from actinomycetes. *Current Opinion in Pharmacology*. **8**(5), pp.557–563.

Baltz, R.H. 2010. Streptomyces and Saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. *Journal of Industrial Microbiology & Biotechnology*. **37**(8), pp.759–772.

Barajas, J.F., Phelan, R.M., Schaub, A.J., Kliewer, J.T., Kelly, P.J., Jackson, D.R., Luo, R., Keasling, J.D. and Tsai, S.-C. 2015. Comprehensive Structural and Biochemical Analysis of the Terminal Myxalamid Reductase Domain for the Engineered Production of Primary Alcohols. *Chemistry & Biology*. **22**(8), pp.1018–1029.

Barka, E.A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Meier-Kolthoff, J.P., Klenk, H.-P., Clément, C., Ouhdouch, Y. and van Wezel, G.P. 2016. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiology and molecular biology reviews: MMBR*. **80**(1), pp.1–43.

Barke, J., Seipke, R.F., Grüschow, S., Heavens, D., Drou, N., Bibb, M.J., Goss, R.J., Yu, D.W. and Hutchings, M.I. 2010. A mixed community of actinomycetes produce multiple antibiotics for the fungus farming ant Acromyrmex octospinosus. *BMC Biology*. **8**(1), pp.109-116.

Barkei, J.J., Kevany, B.M., Felnagle, E.A. and Thomas, M.G. 2009. Investigations into Viomycin Biosynthesis Using Heterologous Production in Streptomyces lividans. *Chembiochem : a European journal of chemical biology*. **10**(2), pp.366–376.

Beck, Z.Q., Burr, D.A. and Sherman, D.H. 2007. Characterization of the β-Methylaspartate-α-decarboxylase (CrpG) from the Cryptophycin Biosynthetic Pathway. *ChemBioChem*. **8**(12), pp.1373–1375.

Beld, J., Sonnenschein, E.C., Vickery, C.R., Noel, J.P. and Burkart, M.D. 2013. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Natural Product Reports*. **31**(1), pp.61–108.

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., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.-H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.-A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J. and Hopwood, D.A. 2002. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature*. **417**(6885), pp.141–147.

Bibb, M.J. 2005. Regulation of secondary metabolism in streptomycetes. *Current Opinion in Microbiology*. **8**(2), pp.208–215.

Bilyk, B., Kim, S., Fazal, A., Baker, T.A. and Seipke, R.F. 2020. Regulation of Antimycin Biosynthesis Is Controlled by the ClpXP Protease. *mSphere*. **5**(2).

Bischoff, D., Pelzer, S., Bister, B., Nicholson, G.J., Stockert, S., Schirle, M., Wohlleben, W., Jung, G. and Süssmuth, R.D. 2001. The Biosynthesis of Vancomycin-Type Glycopeptide Antibiotics—The Order of the Cyclization Steps. *Angewandte Chemie International Edition*. **40**(24), pp.4688–4691.

Blin, K., Medema, M.H., Kottmann, R., Lee, S.Y. and Weber, T. 2017. The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Research*. **45**(Database issue), pp.D555–D559.

Blin, K., Shaw, S., Kautsar, S.A., Medema, M.H. and Weber, T. 2021. The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes. *Nucleic Acids Research*. **49**(D1), pp.D639–D643.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H. and Weber, T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*. **47**(W1), pp.W81–W87.

Bloudoff, K., Rodionov, D. and Schmeing, T.M. 2013. Crystal Structures of the First Condensation Domain of CDA Synthetase Suggest Conformational Changes during the Synthetic Cycle of Nonribosomal Peptide Synthetases. *Journal of Molecular Biology*. **425**(17), pp.3137–3150.

Bloudoff, K. and Schmeing, T.M. 2017. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. **1865**(11, Part B), pp.1587–1604.

Bobek, J., Šmídová, K. and Čihák, M. 2017. A Waking Review: Old and Novel Insights into the Spore Germination in Streptomyces. *Frontiers in Microbiology*. **8**(1).

Bosch, F. and Rosich, L. 2008. The Contributions of Paul Ehrlich to Pharmacology: A Tribute on the Occasion of the Centenary of His Nobel Prize. *Pharmacology*. **82**(3), pp.171–179.

Brautigam, C.A., Zhao, H., Vargas, C., Keller, S. and Schuck, P. 2016. Integration and global analysis of isothermal titration calorimetry data for studying macromolecular interactions. *Nature Protocols*. **11**(5), pp.882–894.

Bruner, S.D., Weber, T., Kohli, R.M., Schwarzer, D., Marahiel, M.A., Walsh, C.T. and Stubbs, M.T. 2002. Structural Basis for the Cyclization of the Lipopeptide Antibiotic Surfactin by the Thioesterase Domain SrfTE. *Structure*. **10**(3), pp.301–310.

Bush, K. 2018. Past and Present Perspectives on β-Lactamases. *Antimicrobial Agents and Chemotherapy*. **62**(10).

Caboche, S., Leclère, V., Pupin, M., Kucherov, G. and Jacques, P. 2010. Diversity of Monomers in Nonribosomal Peptides: towards the Prediction of Origin and Biological Activity. *Journal of Bacteriology*. **192**(19), pp.5143–5150.

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

Caffrey, P., Bevitt, D.J., Staunton, J. and Leadlay, P.F. 1992. Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of the erythromycin-producing polyketide synthase from Saccharopolyspora erythraea. *FEBS Letters*. **304**(2), pp.225–228.

Callahan, B., Thattai, M. and Shraiman, B.I. 2009. Emergent gene order in a model of modular polyketide synthases. *Proceedings of the National Academy of Sciences*. **106**(46), pp.19410–19415.

Challis, G.L., Ravel, J. and Townsend, C.A. 2000. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemistry & Biology*. **7**(3), pp.211–224.

Charkoudian, L.K., Liu, C.W., Capone, S., Kapur, S., Cane, D.E., Togni, A., Seebach, D. and Khosla, C. 2011. Probing the interactions of an acyl carrier protein domain from the 6-deoxyerythronolide B synthase. *Protein Science*. **20**(7), pp.1244–1255.

Chater, K.F. 2016. Recent advances in understanding Streptomyces. *F1000Research*. **5**.

Chen, A.Y., Cane, D.E. and Khosla, C. 2007. Structure-Based Dissociation of a Type I Polyketide Synthase Module. *Chemistry & Biology*. **14**(7), pp.784–792.

Chen, Y., Unger, M., Ntai, I., McClure, R.A., Albright, J.C., Thomson, R.J. and Kelleher, N.L. 2013. Gobichelin A and B: Mixed-Ligand Siderophores Discovered Using Proteomics. *MedChemComm*. **4**(1), pp.233–238.

Clardy, J., Fischbach, M.A. and Walsh, C.T. 2006. New antibiotics from bacterial natural products. *Nature Biotechnology*. **24**(12), pp.1541–1550.

Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Sayers, E.W. 2016. GenBank. *Nucleic Acids Research*. **44**(Database issue), pp.D67–D72.

Claxton, H.B., Akey, D.L., Silver, M.K., Admiraal, S.J. and Smith, J.L. 2009. Structure and Functional Analysis of RifR, the Type II Thioesterase from the Rifamycin Biosynthetic Pathway. *Journal of Biological Chemistry*. **284**(8), pp.5021–5029.

Cobb, R.E., Wang, Y. and Zhao, H. 2015. High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. *ACS Synthetic Biology*. **4**(6), pp.723–728.

Conti, E., Stachelhaus, T., Marahiel, M.A. and Brick, P. 1997. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *The EMBO Journal*. **16**(14), pp.4174–4183.

Cunin, R., Glansdorff, N., Piérard, A. and Stalon, V. 1986. Biosynthesis and metabolism of arginine in bacteria. *Microbiological Reviews*. **50**(3), pp.314–352.

Cupp-Vickery, J.R. and Poulos, T.L. 1995. Structure of cytochrome P450eryF involved in erythromycin biosynthesis. *Nature Structural Biology*. **2**(2), pp.144–153.

Das, A., Szu, P.-H., Fitzgerald, J.T. and Khosla, C. 2010. Mechanism and engineering of polyketide chain initiation in fredericamycin biosynthesis. *Journal of the American Chemical Society*. **132**(26), pp.8831–8833.

Datsenko, K.A. and Wanner, B.L. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*. **97**(12), pp.6640–6645.

Davies, J. 2006. Where have All the Antibiotics Gone? *The Canadian Journal of Infectious Diseases & Medical Microbiology*. **17**(5), pp.287–290.

Del Vecchio, F., Petkovic, H., Kendrew, S.G., Low, L., Wilkinson, B., Lill, R., Cortés, J., Rudd, B.A.M., Staunton, J. and Leadlay, P.F. 2003. Active-site residue, domain and module swaps in modular polyketide synthases. *Journal of Industrial Microbiology and Biotechnology*. **30**(8), pp.489–494.

Di Luccio, E., Elling, R.A. and Wilson, D.K. 2006. Identification of a novel NADH-specific aldo-keto reductase using sequence and structural homologies. *Biochemical Journal*. **400**(1), pp.105–114.

Ding, W., Dong, Y., Ju, J. and Li, Q. 2020. The roles of genes associated with regulation, transportation, and macrocyclization in desotamide biosynthesis in Streptomyces scopuliridis SCSIO ZJ46. *Applied Microbiology and Biotechnology*. **104**(6), pp.2603–2610.

Dobson, C.M. 2004. Chemical space and biology. *Nature*. **432**(7019), pp.824–828.

Dong, C., Flecks, S., Unversucht, S., Haupt, C., van Pée, K.-H. and Naismith, J.H. 2005. Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science*. **309**(5744), pp.2216–2219.

Drufva, E.E., Hix, E.G. and Bailey, C.B. 2020. Site directed mutagenesis as a precision tool to enable synthetic biology with engineered modular polyketide synthases. *Synthetic and Systems Biotechnology*. **5**(2), pp.62–80.

Du, L. and Lou, L. 2010. PKS and NRPS release mechanisms. *Natural Product Reports*. **27**(2), pp.255–278.

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

Dunn, B.J. and Khosla, C. 2013. Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. *Journal of The Royal Society Interface*. **10**(85), p.20130297.

Dunshee, B.R., Leben, Curt., Keitt, G.W. and Strong, F.M. 1949. The Isolation and Properties of Antimycin A. *Journal of the American Chemical Society*. **71**(7), pp.2436–2437.

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. and Skiniotis, G. 2014. Structure of a modular polyketide synthase. *Nature*. **510**(7506), pp.512–517.

Ebrahimi, V. and Hashemi, A. 2020. Challenges of in vitro genome editing with CRISPR/Cas9 and possible solutions: A review. *Gene*. **753**, p.144813.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. **32**(5), pp.1792–1797.

Ehmann, D.E., Gehring, A.M. and Walsh, C.T. 1999. Lysine biosynthesis in Saccharomyces cerevisiae: mechanism of alpha-aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5. *Biochemistry*. **38**(19), pp.6171–6177.

Ehmann, D.E., Trauger, J.W., Stachelhaus, T. and Walsh, C.T. 2000. Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chemistry & Biology*. **7**(10), pp.765–772.

Emsley, P. and Cowtan, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallographica. Section D, Biological Crystallography*. **60**(Pt 12 Pt 1), pp.2126–2132.

Erb, T.J., Berg, I.A., Brecht, V., Müller, M., Fuchs, G. and Alber, B.E. 2007. Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proceedings of the National Academy of Sciences of the United States of America*. **104**(25), pp.10631–10636.

Erb, T.J., Brecht, V., Fuchs, G., Müller, M. and Alber, B.E. 2009. Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase. *Proceedings of the National Academy of Sciences of the United States of America*. **106**(22), pp.8871–8876.

Evans, P.R. and Murshudov, G.N. 2013. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography*. **69**(Pt 7), pp.1204–1214.

Fazal, A., Thankachan, D., Harris, E. and Seipke, R.F. 2020. A chromatogram-simplified Streptomyces albus host for heterologous production of natural products. *Antonie Van Leeuwenhoek*. **113**(4), pp.511–520.

Fazal, A., Webb, M.E. and Seipke, R.F. 2020. The Desotamide Family of Antibiotics. *Antibiotics*. **9**(8), p.452.

Felnagle, E.A., Barkei, J.J., Park, H., Podevels, A.M., McMahon, M.D., Drott, D.W. and Thomas, M.G. 2010. MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry*. **49**(41), pp.8815–8817.

Felnagle, E.A., Jackson, E.E., Chan, Y.A., Podevels, A.M., Berti, A.D., McMahon, M.D. and Thomas, M.G. 2008. Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products. *Molecular pharmaceutics*. **5**(2), pp.191–211.

Filling, C., Berndt, K.D., Benach, J., Knapp, S., Prozorovski, T., Nordling, E., Ladenstein, R., Jörnvall, H. and Oppermann, U. 2002. Critical Residues for Structure and Catalysis in Short-chain Dehydrogenases/Reductases. *Journal of Biological Chemistry*. **277**(28), pp.25677–25684.

Finking, R. and Marahiel, M.A. 2004. Biosynthesis of Nonribosomal Peptides. *Annual Review of Microbiology*. **58**(1), pp.453–488.

Fischbach, M.A. and Walsh, C.T. 2006. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chemical Reviews*. **106**(8), pp.3468–3496.

Fitzgerald, J.T., Charkoudian, L.K., Watts, K.R. and Khosla, C. 2013. Analysis and refactoring of the A-74528 biosynthetic pathway. *Journal of the American Chemical Society*. **135**(10), pp.3752–3755.

Flärdh, K. and Buttner, M.J. 2009. Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium. *Nature Reviews Microbiology*. **7**(1), pp.36–49.

Fleming, A. 1929. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. *British journal of experimental pathology*. **10**(3), pp.226–236.

Frueh, D.P., Arthanari, H., Koglin, A., Vosburg, D.A., Bennett, A.E., Walsh, C.T. and Wagner, G. 2008. Dynamic thiolation–thioesterase structure of a non-ribosomal peptide synthetase. *Nature*. **454**(7206), pp.903–906.

Fu, J., Bian, X., Hu, S., Wang, H., Huang, F., Seibert, P.M., Plaza, A., Xia, L., Müller, R., Stewart, A.F. and Zhang, Y. 2012. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nature Biotechnology*. **30**(5), pp.440–446.

Gang, D., Kim, D.W. and Park, H.-S. 2018. Cyclic Peptides: Promising Scaffolds for Biopharmaceuticals. *Genes*. **9**(11).

Gao, X., Haynes, S.W., Ames, B.D., Wang, P., Vien, L.P., Walsh, C.T. and Tang, Y. 2012. Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain. *Nature Chemical Biology*. **8**(10), pp.823–830.

Garg, A., Xie, X., Keatinge-Clay, A., Khosla, C. and Cane, D.E. 2014. Elucidation of the cryptic epimerase activity of redox-inactive ketoreductase domains from modular polyketide synthases by tandem equilibrium isotope exchange. *Journal of the American Chemical Society*. **136**(29), pp.10190–10193.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*. **31**(13), pp.3784–3788.

Gay, D.C., Gay, G., Axelrod, A.J., Jenner, M., Kohlhaas, C., Kampa, A., Oldham, N.J., Piel, J. and Keatinge-Clay, A.T. 2014. A Close Look at a Ketosynthase from a Trans-Acyltransferase Modular Polyketide Synthase. *Structure*. **22**(3), pp.444–451.

Gehring, A.M., Bradley, K.A. and Walsh, C.T. 1997. Enterobactin Biosynthesis in Escherichia coli: Isochorismate Lyase (EntB) Is a Bifunctional Enzyme That Is Phosphopantetheinylated by EntD and Then Acylated by EntE Using ATP and 2,3-Dihydroxybenzoate. *Biochemistry*. **36**(28), pp.8495–8503.

Gelpi, A., Gilbertson, A. and Tucker, J.D. 2015. Magic bullet: Paul Ehrlich, Salvarsan and the birth of venereology. *Sexually Transmitted Infections*. **91**(1), pp.68–69.

Gerber, R., Lou, L. and Du, L. 2009. A PLP-dependent polyketide chain releasing mechanism in the biosynthesis of mycotoxin fumonisins in Fusarium verticillioides. *Journal of the American Chemical Society*. **131**(9), pp.3148–3149.

Gokhale, R.S., Hunziker, D., Cane, D.E. and Khosla, C. 1999. Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chemistry & Biology*. **6**(2), pp.117–125.

Gokhale, R.S. and Khosla, C. 2000. Role of linkers in communication between protein modules. *Current Opinion in Chemical Biology*. **4**(1), pp.22–27.

Gomez-Escribano, J.P. and Bibb, M.J. 2011. Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters. *Microbial Biotechnology*. **4**(2), pp.207–215.

Gomez-Escribano, J.P., Song, L., Fox, D.J., Yeo, V., Bibb, M.J. and Challis, G.L. 2012. Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the cpk gene cluster of Streptomyces coelicolor M145. *Chemical Science*. **3**(9), pp.2716–2720.

Gould, K. 2016. Antibiotics: from prehistory to the present day. *Journal of Antimicrobial Chemotherapy*. **71**(3), pp.572–575.

Gust, B., Challis, G.L., Fowler, K., Kieser, T. and Chater, K.F. 2003. PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proceedings of the National Academy of Sciences*. **100**(4), pp.1541–1546.

Hahn, M. and Stachelhaus, T. 2004. Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proceedings of the National Academy of Sciences of the United States of America*. **101**(44), pp.15585–15590.

Haltli, B., Tan, Y., Magarvey, N.A., Wagenaar, M., Yin, X., Greenstein, M., Hucul, J.A. and Zabriskie, T.M. 2005. Investigating β-Hydroxyenduracididine Formation in the Biosynthesis of the Mannopeptimycins. *Chemistry & Biology*. **12**(11), pp.1163–1168.

Hamamoto, K., Kida, Y., Zhang, Y., Shimizu, T. and Kuwano, K. 2002. Antimicrobial Activity and Stability to Proteolysis of Small Linear Cationic Peptides with D-Amino Acid Substitutions. *Microbiology and Immunology*. **46**(11), pp.741–749.

Hans, M., Hornung, A., Dziarnowski, A., Cane, D.E. and Khosla, C. 2003. Mechanistic analysis of acyl transferase domain exchange in polyketide synthase modules. *Journal of the American Chemical Society*. **125**(18), pp.5366–5374.

Hanukoglu, I. and Gutfinger, T. 1989. cDNA sequence of adrenodoxin reductase. *European Journal of Biochemistry*. **180**(2), pp.479–484.

Haslinger, K., Redfield, C. and Cryle, M.J. 2015. Structure of the terminal PCP domain of the non-ribosomal peptide synthetase in teicoplanin biosynthesis. *Proteins*. **83**(4), pp.711–721.

Heathcote, M.L., Staunton, J. and Leadlay, P.F. 2001. Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chemistry & Biology*. **8**(2), pp.207–220.

Hertweck, C. 2009. The biosynthetic logic of polyketide diversity. *Angewandte Chemie International Edition*. **48**(26), pp.4688–4716.

Hertweck, C., Luzhetskyy, A., Rebets, Y. and Bechthold, A. 2007. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Natural Product Reports*. **24**(1), pp.162–190.

Holzbaur, I.E., Harris, R.C., Bycroft, M., Cortes, J., Bisang, C., Staunton, J., Rudd, B.A. and Leadlay, P.F. 1999. Molecular basis of Celmer's rules: the role of two ketoreductase domains in the control of chirality by the erythromycin modular polyketide synthase. *Chemistry & Biology*. **6**(4), pp.189–195.

Hong, H.-J., Hutchings, M.I., Hill, L.M. and Buttner, M.J. 2005. The role of the novel Fem protein VanK in vancomycin resistance in Streptomyces coelicolor. *The Journal of Biological Chemistry*. **280**(13), pp.13055–13061.

Hopwood, D.A. 1967. Genetic analysis and genome structure in Streptomyces coelicolor. *Bacteriological Reviews*. **31**(4), pp.373–403.

Hoyer, K.M., Mahlert, C. and Marahiel, M.A. 2007. The Iterative Gramicidin S Thioesterase Catalyzes Peptide Ligation and Cyclization. *Chemistry & Biology*. **14**(1), pp.13–22.

Huang, H., Zheng, G., Jiang, W., Hu, H. and Lu, Y. 2015. One-step high-efficiency CRISPR/Cas9-mediated genome editing in Streptomyces. *Acta Biochimica et Biophysica Sinica*. **47**(4), pp.231–243.

Huo, L., Hug, J.J., Fu, C., Bian, X., Zhang, Y. and Müller, R. 2019. Heterologous expression of bacterial natural product biosynthetic pathways. *Natural Product Reports*. **36**(10), pp.1412–1436.

Hutchings, M.I., Truman, A.W. and Wilkinson, B. 2019. Antibiotics: past, present and future. *Current Opinion in Microbiology*. **51**, pp.72–80.

Izoré, T. and J. Cryle, M. 2018. The many faces and important roles of protein–protein interactions during non-ribosomal peptide synthesis. *Natural Product Reports*. **35**(11), pp.1120–1139.

Jacobsen, J.R., Hutchinson, C.R., Cane, D.E. and Khosla, C. 1997. Precursor-Directed Biosynthesis of Erythromycin Analogs by an Engineered Polyketide Synthase. *Science*. **277**(5324), pp.367–369.

Janetzko, J. and Batey, R.A. 2014. Organoboron-Based Allylation Approach to the Total Synthesis of the Medium-Ring Dilactone (+)-Antimycin A1b. *The Journal of Organic Chemistry*. **79**(16), pp.7415–7424.

Jaremko, M.J., Davis, T.D., Corpuz, J.C. and Burkart, M.D. 2020. Type II non-ribosomal peptide synthetase proteins: structure, mechanism, and protein–protein interactions. *Natural Product Reports*. **37**(3), pp.355–379.

Javidpour, P., Bruegger, J., Srithahan, S., Korman, T.P., Crump, M.P., Crosby, J., Burkart, M.D. and Tsai, S.-C. 2013. The Determinants of Activity and Specificity in Actinorhodin Type II Polyketide Ketoreductase. *Chemistry & Biology*. **20**(10), pp.1225–1234.

Javidpour, P., Das, A., Khosla, C. and Tsai, S.-C. 2011. Structural and biochemical studies of the hedamycin type II polyketide ketoreductase (HedKR): molecular basis of stereo- and regiospecificities. *Biochemistry*. **50**(34), pp.7426–7439.

Javidpour, P., Korman, T.P., Shakya, G. and Tsai, S.-C. 2011. Structural and Biochemical Analyses of Regio- and Stereospecificities Observed in a Type II Polyketide Ketoreductase. *Biochemistry*. **50**(21), pp.4638–4649.

Jenke-Kodama, H. and Dittmann, E. 2009. Bioinformatic perspectives on NRPS/PKS megasynthases: Advances and challenges. *Natural Product Reports*. **26**(7), pp.874–883.

Jiang, W., Zhao, X., Gabrieli, T., Lou, C., Ebenstein, Y. and Zhu, T.F. 2015. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nature Communications*. **6**(1), p.8101.

Jones, S.E. and Elliot, M.A. 2018. 'Exploring' the regulation of Streptomyces growth and development. *Current Opinion in Microbiology*. **42**, pp.25–30.

Joo, S.H. 2012. Cyclic Peptides as Therapeutic Agents and Biochemical Tools. *Biomolecules & Therapeutics*. **20**(1), pp.19–26.

Kapoor, G., Saigal, S. and Elongavan, A. 2017. Action and resistance mechanisms of antibiotics: A guide for clinicians. *Journal of Anaesthesiology, Clinical Pharmacology*. **33**(3), pp.300–305.

Karimova, G., Pidoux, J., Ullmann, A. and Ladant, D. 1998. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proceedings of the National Academy of Sciences*. **95**(10), pp.5752–5756.

Katz, L. and Baltz, R.H. 2016. Natural product discovery: past, present, and future. *Journal of Industrial Microbiology & Biotechnology*. **43**(2–3), pp.155–176.

Kautsar, S.A., Blin, K., Shaw, S., Navarro-Muñoz, J.C., Terlouw, B.R., van der Hooft, J.J.J., van Santen, J.A., Tracanna, V., Suarez Duran, H.G., Pascal Andreu, V., Selem-Mojica, N., Alanjary, M., Robinson, S.L., Lund, G., Epstein, S.C., Sisto, A.C., Charkoudian, L.K., Collemare, J., Linington, R.G., Weber, T. and Medema, M.H. 2020. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Research*. **48**(D1), pp.D454–D458.

Kaweewan, I., Komaki, H., Hemmi, H. and Kodani, S. 2017. Isolation and Structure Determination of New Antibacterial Peptide Curacomycin Based on Genome Mining. *Asian Journal of Organic Chemistry*. **6**(12), pp.1838–1844.

Kaysser, L., Bernhardt, P., Nam, S.-J., Loesgen, S., Ruby, J.G., Skewes-Cox, P., Jensen, P.R., Fenical, W. and Moore, B.S. 2012. Merochlorins A–D, Cyclic Meroterpenoid Antibiotics Biosynthesized in Divergent Pathways with Vanadium-Dependent

Chloroperoxidases. *Journal of the American Chemical Society*. **134**(29), pp.11988–11991.

Keating, T.A., Marshall, C.G. and Walsh, C.T. 2000. Vibriobactin Biosynthesis in Vibrio cholerae: VibH Is an Amide Synthase Homologous to Nonribosomal Peptide Synthetase Condensation Domains. *Biochemistry*. **39**(50), pp.15513–15521.

Keating, T.A., Marshall, C.G., Walsh, C.T. and Keating, A.E. 2002. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nature Structural Biology*. **9**(7), pp.522–526.

Keatinge-Clay, A. 2008. Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *Journal of Molecular Biology*. **384**(4), pp.941–953.

Keatinge-Clay, A.T. 2016. Stereocontrol within polyketide assembly lines. *Natural Product Reports*. **33**(2), pp.141–149.

Keatinge-Clay, A.T. 2012. The structures of type I polyketide synthases. *Natural Product Reports*. **29**(10), pp.1050–1073.

Keatinge-Clay, A.T., Shelat, A.A., Savage, D.F., Tsai, S.-C., Miercke, L.J.W., O'Connell, J.D., Khosla, C. and Stroud, R.M. 2003. Catalysis, Specificity, and ACP Docking Site of Streptomyces coelicolor Malonyl-CoA:ACP Transacylase. *Structure*. **11**(2), pp.147–154.

Kelemen, G.H., Viollier, P.H., Tenor, J., Marri, L., Buttner, M.J. and Thompson, C.J. 2001. A connection between stress and development in the multicellular prokaryote Streptomyces coelicolor A3(2). *Molecular Microbiology*. **40**(4), pp.804–814.

Khalil, Z.G., Hill, T.A., De Leon Rodriguez, L.M., Lohman, R.-J., Hoang, H.N., Reiling, N., Hillemann, D., Brimble, M.A., Fairlie, D.P., Blumenthal, A. and Capon, R.J. 2019. Structure-Activity Relationships of Wollamide Cyclic Hexapeptides with Activity against Drug-Resistant and Intracellular Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy*. **63**(3), pp.1984–13.

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

Kharel, M.K., Pahari, P., Shepherd, M.D., Tibrewal, N., Nybo, S.E., Shaaban, K.A. and Rohr, J. 2012. Angucyclines: Biosynthesis, mode-of-action, new natural products, and synthesis. *Natural Product Reports*. **29**(2), pp.264–325.

Khosla, C., Gokhale, R.S., Jacobsen, J.R. and Cane, D.E. 1999. Tolerance and Specificity of Polyketide Synthases. *Annual Review of Biochemistry*. **68**(1), pp.219–253.

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

Khosla, C., Tang, Y., Chen, A.Y., Schnarr, N.A. and Cane, D.E. 2007. Structure and Mechanism of the 6-Deoxyerythronolide B Synthase. *Annual Review of Biochemistry*. **76**(1), pp.195–221.

Kieser, T.B., Buttner, M.J., Chater, K.F., Bibb, M.J. and Hopwood, D.A. 2000. *Practical Streptomyces genetics*. Norwich, United Kingdom: The John Innes Foundation.

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

Klaus, M. and Grininger, M. 2018. Engineering strategies for rational polyketide synthase design. *Natural Product Reports*. **35**(10), pp.1070–1081.

Kluge, B., Vater, J., Salnikow, J. and Eckart, K. 1988. Studies on the biosynthesis of surfactin, a lipopeptide antibiotic from Bacillus subtilis ATCC 21332. *FEBS Letters*. **231**(1), pp.107–110.

Koglin, A., Löhr, F., Bernhard, F., Rogov, V.V., Frueh, D.P., Strieter, E.R., Mofid, M.R., Güntert, P., Wagner, G., Walsh, C.T., Marahiel, M.A. and Dötsch, V. 2008. Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature*. **454**(7206), pp.907–911.

Kohli, R.M., Trauger, J.W., Schwarzer, D., Marahiel, M.A. and Walsh, C.T. 2001. Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. *Biochemistry*. **40**(24), pp.7099–7108.

Kohli, R.M., Walsh, C.T. and Burkart, M.D. 2002. Biomimetic synthesis and optimization of cyclic peptide antibiotics. *Nature*. **418**(6898), pp.658–661.

Kong, K.-F., Schneper, L. and Mathee, K. 2010. Beta-lactam Antibiotics: From Antibiosis to Resistance and Bacteriology. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica*. **118**(1), pp.1–36.

Korman, T.P., Crawford, J.M., Labonte, J.W., Newman, A.G., Wong, J., Townsend, C.A. and Tsai, S.-C. 2010. Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. *Proceedings of the National Academy of Sciences*. **107**(14), pp.6246–6251.

Kotowska, M., Ciekot, J. and Pawlik, K. 2014. Type II thioesterase ScoT is required for coelimycin production by the modular polyketide synthase Cpk of Streptomyces coelicolor A3(2). *Acta Biochimica Polonica*. **61**(1).

Kraas, F.I., Helmetag, V., Wittmann, M., Strieker, M. and Marahiel, M.A. 2010. Functional Dissection of Surfactin Synthetase Initiation Module Reveals Insights into the Mechanism of Lipoinitiation. *Chemistry & Biology*. **17**(8), pp.872–880.

Kröger, A., Klingenberg, M. and Schweidler, S. 1973. Further Evidence for the Pool Function of Ubiquinone as Derived from the Inhibition of the Electron Transport by Antimycin. *European Journal of Biochemistry*. **39**(2), pp.313–323.

Kuranaga, T., Fukuba, A., Ninomiya, A., Takada, K., Matsunaga, S. and Wakimoto, T. 2018. Diastereoselective Total Synthesis and Structural Confirmation of Surugamide F. *Chemical and Pharmaceutical Bulletin*. **66**(6), pp.637–641.

Kuranaga, T., Matsuda, K., Sano, A., Kobayashi, M., Ninomiya, A., Takada, K., Matsunaga, S. and Wakimoto, T. 2018. Total Synthesis of the Nonribosomal Peptide Surugamide B and Identification of a New Offloading Cyclase Family. *Angewandte Chemie International Edition*. **57**(30), pp.9447–9451.

Kurnasov, O., Jablonski, L., Polanuyer, B., Dorrestein, P., Begley, T. and Osterman, A. 2003. Aerobic tryptophan degradation pathway in bacteria: novel kynurenine formamidase. *FEMS microbiology letters*. **227**(2), pp.219–227.

Kwan, D.H. and Leadlay, P.F. 2010. Mutagenesis of a modular polyketide synthase enoylreductase domain reveals insights into catalysis and stereospecificity. *ACS chemical biology*. **5**(9), pp.829–838.

Lachance, H., Wetzel, S., Kumar, K. and Waldmann, H. 2012. Charting, Navigating, and Populating Natural Product Chemical Space for Drug Discovery. *Journal of Medicinal Chemistry*. **55**(13), pp.5989–6001.

Lambert, J.N., Mitchell, J.P. and Roberts, K.D. 2001. The synthesis of cyclic peptides. *Journal of the Chemical Society, Perkin Transactions 1*. (5), pp.471–484.

Law, J.W.-F., Pusparajah, P., Mutalib, N.-S.A., Wong, S.H., Goh, B.-H. and Lee, L.-H. 2019. A Review on Mangrove Actinobacterial Diversity: The Roles of Streptomyces and Novel Species Discovery. *Progress In Microbes & Molecular Biology*. **2**(1).

Le, C.Q., Oyugi, M., Joseph, E., Nguyen, T., Ullah, M.H., Aubert, J., Phan, T., Tran, J. and Johnson-Winters, K. 2017. Effects of isoleucine 135 side chain length on the cofactor donor-acceptor distance within F420H2:NADP+ oxidoreductase: A kinetic analysis. *Biochemistry and Biophysics Reports*. **9**, pp.114–120.

Lee, T.V., Johnson, L.J., Johnson, R.D., Koulman, A., Lane, G.A., Lott, J.S. and Arcus, V.L. 2010. Structure of a Eukaryotic Nonribosomal Peptide Synthetase Adenylation Domain That Activates a Large Hydroxamate Amino Acid in Siderophore Biosynthesis. *Journal of Biological Chemistry*. **285**(4), pp.2415–2427.

Li, L., MacIntyre, L.W. and Brady, S.F. 2021. Refactoring biosynthetic gene clusters for heterologous production of microbial natural products. *Current Opinion in Biotechnology*. **69**, pp.145–152.

Li, Q., Khosla, C., Puglisi, J.D. and Liu, C.W. 2003. Solution Structure and Backbone Dynamics of the Holo Form of the Frenolicin Acyl Carrier Protein,. *Biochemistry*. **42**(16), pp.4648–4657.

Li, Q., Qin, X., Liu, J., Gui, C., Wang, B., Li, J. and Ju, J. 2016. Deciphering the Biosynthetic Origin of l-allo-Isoleucine. *Journal of the American Chemical Society*. **138**(1), pp.408–415.

Li, Q., Song, Y., Qin, X., Zhang, X., Sun, A. and Ju, J. 2015. Identification of the Biosynthetic Gene Cluster for the Anti-infective Desotamides and Production of a

New Analogue in a Heterologous Host. *Journal of natural products*. **78**(4), pp.944–948.

Li, T.-L., Huang, F., Haydock, S.F., Mironenko, T., Leadlay, P.F. and Spencer, J.B. 2004. Biosynthetic Gene Cluster of the Glycopeptide Antibiotic Teicoplanin: Characterization of Two Glycosyltransferases and the Key Acyltransferase. *Chemistry & Biology*. **11**(1), pp.107–119.

Li, Y., Zhang, W., Zhang, H., Tian, W., Wu, L., Wang, S., Zheng, M., Zhang, J., Sun, C., Deng, Z., Sun, Y., Qu, X. and Zhou, J. 2018. Structural Basis of a Broadly Selective Acyltransferase from the Polyketide Synthase of Splenocin. *Angewandte Chemie International Edition*. **57**(20), pp.5823–5827.

Lin, S., Lanen, S.G.V. and Shen, B. 2009. A free-standing condensation enzyme catalyzing ester bond formation in C-1027 biosynthesis. *Proceedings of the National Academy of Sciences*. **106**(11), pp.4183–4188.

Lin, Z., Nielsen, J. and Liu, Z. 2020. Bioprospecting Through Cloning of Whole Natural Product Biosynthetic Gene Clusters. *Frontiers in Bioengineering and Biotechnology*. **8**.

Linne, U., Doekel, S. and Marahiel, M.A. 2001. Portability of Epimerization Domain and Role of Peptidyl Carrier Protein on Epimerization Activity in Nonribosomal Peptide Synthetases. *Biochemistry*. **40**(51), pp.15824–15834.

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

Linne, U., Schwarzer, D., Schroeder, G.N. and Marahiel, M.A. 2004. Mutational analysis of a type II thioesterase associated with nonribosomal peptide synthesis. *European Journal of Biochemistry*. **271**(8), pp.1536–1545.

Liu, H. and Reynolds, K.A. 2001. Precursor Supply for Polyketide Biosynthesis: The Role of Crotonyl-CoA Reductase. *Metabolic Engineering*. **3**(1), pp.40–48.

Liu, J., Zhu, X., Kim, S.J. and Zhang, W. 2016. Antimycin-type depsipeptides: discovery, biosynthesis, chemical synthesis, and bioactivities. *Natural Product Reports*. **33**(10), pp.1146–1165.

Liu, J., Zhu, X., Seipke, R.F. and Zhang, W. 2015. Biosynthesis of Antimycins with a Reconstituted 3-Formamidosalicylate Pharmacophore in Escherichia coli. *ACS Synthetic Biology*. **4**(5), pp.559–565.

Lohman, J.R., Ma, M., Cuff, M.E., Bigelow, L., Bearden, J., Babnigg, G., Joachimiak, A., Phillips, G.N. and Shen, B. 2014. The crystal structure of BlmI as a model for nonribosomal peptide synthetase peptidyl carrier proteins. *Proteins*. **82**(7), pp.1210–1218.

Lomovskaya, N., Otten, S.L., Doi-Katayama, Y., Fonstein, L., Liu, X.-C., Takatsu, T., Inventi-Solari, A., Filippini, S., Torti, F., Colombo, A.L. and Hutchinson, C.R. 1999. Doxorubicin Overproduction in Streptomyces peucetius: Cloning and

Characterization of the dnrU Ketoreductase anddnrV Genes and the doxA Cytochrome P-450 Hydroxylase Gene. *Journal of Bacteriology*. **181**(1), pp.305–318.

Losey, H.C., Peczuh, M.W., Chen, Z., Eggert, U.S., Dong, S.D., Pelczer, I., Kahne, D. and Walsh, C.T. 2001. Tandem action of glycosyltransferases in the maturation of vancomycin and teicoplanin aglycones: novel glycopeptides. *Biochemistry*. **40**(15), pp.4745–4755.

Mach, B., Reich, E. and Tatum, E.L. 1963. Separation of the biosynthesis of the antibiotic polypeptide tyrocidine from protein biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*. **50**(1), pp.175–181.

MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H. and MacNeil, T. 1992. Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene*. **111**(1), pp.61–68.

Madeira, F., Park, Y. mi, Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D. and Lopez, R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*. **47**(W1), pp.W636–W641.

Magarvey, N.A., Ehling-Schulz, M. and Walsh, C.T. 2006. Characterization of the Cereulide NRPS α-Hydroxy Acid Specifying Modules: Activation of α-Keto Acids and Chiral Reduction on the Assembly Line. *Journal of the American Chemical Society*. **128**(33), pp.10698–10699.

Magarvey, N.A., Haltli, B., He, M., Greenstein, M. and Hucul, J.A. 2006. Biosynthetic Pathway for Mannopeptimycins, Lipoglycopeptide Antibiotics Active against Drug-Resistant Gram-Positive Pathogens. *Antimicrobial Agents and Chemotherapy*. **50**(6), pp.2167–2177.

Manivasagan, P., Venkatesan, J., Sivakumar, K. and Kim, S.-K. 2014. Pharmaceutically active secondary metabolites of marine actinobacteria. *Microbiological Research*. **169**(4), pp.262–278.

Marahiel, M.A. 2016. A structural model for multimodular NRPS assembly lines. *Natural Product Reports*. **33**(2), pp.136–140.

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

Masschelein, J., Sydor, P.K., Hobson, C., Howe, R., Jones, C., Roberts, D.M., Ling Yap, Z., Parkhill, J., Mahenthiralingam, E. and Challis, G.L. 2019. A dual transacylation mechanism for polyketide synthase chain release in enacyloxin antibiotic biosynthesis. *Nature Chemistry*. **11**(10), pp.906–912.

Mathews, I.I., Allison, K., Robbins, T., Lyubimov, A.Y., Uervirojnangkoorn, M., Brunger, A.T., Khosla, C., DeMirci, H., McPhillips, S.E., Hollenbeck, M., Soltis, M. and Cohen, A.E. 2017. The Conformational Flexibility of the Acyltransferase from the Disorazole Polyketide Synthase Is Revealed by an X-ray Free-Electron Laser Using a

Room-Temperature Sample Delivery Method for Serial Crystallography. *Biochemistry*. **56**(36), pp.4751–4756.

Matsuda, K., Kobayashi, M., Kuranaga, T., Takada, K., Ikeda, H., Matsunaga, S. and Wakimoto, T. 2019. SurE is a trans-acting thioesterase cyclizing two distinct non-ribosomal peptides. *Organic & Biomolecular Chemistry*. **17**(5), pp.1058–1061.

Matsuda, K., Kuranaga, T., Sano, A., Ninomiya, A., Takada, K. and Wakimoto, T. 2019. The Revised Structure of the Cyclic Octapeptide Surugamide A. *Chemical and Pharmaceutical Bulletin*. **67**(5), pp.476–480.

Matsuda, K., Zhai, R., Mori, T., Kobayashi, M., Sano, A., Abe, I. and Wakimoto, T. 2020. Heterochiral coupling in non-ribosomal peptide macrolactamization. *Nature Catalysis*. **3**(6), pp.507–515.

Mayer, A., Taguchi, T., Linnenbrink, A., Hofmann, C., Luzhetskyy, A. and Bechthold, A. 2005. LanV, a Bifunctional Enzyme: Aromatase and Ketoreductase during Landomycin A Biosynthesis. *ChemBioChem*. **6**(12), pp.2312–2315.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. 2007. Phaser crystallographic software. *Journal of Applied Crystallography*. **40**(4), pp.658–674.

McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, Melanie, Betlach, Mary and Ashley, G. 1999. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proceedings of the National Academy of Sciences*. **96**(5), pp.1846–1851.

McLean, T.C., Hoskisson, P.A. and Seipke, R.F. 2016. Coordinate Regulation of Antimycin and Candicidin Biosynthesis. *mSphere*. **1**(6).

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E. and Breitling, R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research*. **39**(Web Server issue), pp.W339–W346.

Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., de Bruijn, I., Chooi, Y.H., Claesen, J., Coates, R.C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D.J., Fewer, D.P., Garg, N., Geiger, C., Gomez-Escribano, J.P., Greule, A., Hadjithomas, M., Haines, A.S., Helfrich, E.J.N., Hillwig, M.L., Ishida, K., Jones, A.C., Jones, C.S., Jungmann, K., Kegler, C., Kim, H.U., Kötter, P., Krug, D., Masschelein, J., Melnik, A.V., Mantovani, S.M., Monroe, E.A., Moore, M., Moss, N., Nützmann, H.-W., Pan, G., Pati, A., Petras, D., Reen, F.J., Rosconi, F., Rui, Z., Tian, Z., Tobias, N.J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X., Yim, G., Yu, F., Xie, Y., Aigle, B., Apel, A.K., Balibar, C.J., Balskus, E.P., Barona-Gómez, F., Bechthold, A., Bode, H.B., Borriss, R., Brady, S.F., Brakhage, A.A., Caffrey, P., Cheng, Y.-Q., Clardy, J., Cox, R.J., De Mot, R., Donadio, S., Donia, M.S., van der Donk, W.A., Dorrestein, P.C., Doyle, S., Driessen, A.J.M., Ehling-Schulz, M., Entian, K.-D., Fischbach, M.A., Gerwick, L., Gerwick, W.H., Gross, H., Gust, B., Hertweck, C., Höfte, M., Jensen, S.E., Ju, J., Katz, L., Kaysser, L., Klassen, J.L., Keller, N.P.,

Kormanec, J., Kuipers, O.P., Kuzuyama, T., Kyrpides, N.C., Kwon, H.-J., Lautru, S., Lavigne, R., Lee, C.Y., Linquan, B., Liu, X., Liu, W., Luzhetskyy, A., Mahmud, T., Mast, Y., Méndez, C., Metsä-Ketelä, M., Micklefield, J., Mitchell, D.A., Moore, B.S., Moreira, L.M., Müller, R., Neilan, B.A., Nett, M., Nielsen, J., O'Gara, F., Oikawa, H., Osbourn, A., Osburne, M.S., Ostash, B., Payne, S.M., Pernodet, J.-L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J.M., Salas, J.A., Schmitt, E.K., Scott, B., Seipke, R.F., Shen, B., Sherman, D.H., Sivonen, K., Smanski, M.J., Sosio, M., Stegmann, E., Süssmuth, R.D., Tahlan, K., Thomas, C.M., Tang, Y., Truman, A.W., Viaud, M., Walton, J.D., Walsh, C.T., Weber, T., van Wezel, G.P., Wilkinson, B., Willey, J.M., Wohlleben, W., Wright, G.D., Ziemert, N., Zhang, C., Zotchev, S.B., Breitling, R., Takano, E. and Glöckner, F.O. 2015. Minimum Information about a Biosynthetic Gene cluster. *Nature chemical biology*. **11**(9), pp.625–631.

van der Meij, A., Worsley, S.F., Hutchings, M.I. and van Wezel, G.P. 2017. Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiology Reviews*. **41**(3), pp.392–416.

Miao, S., Anstee, M.R., LaMarco, K., Matthew, J., Huang, L.H.T. and Brasseur, M.M. 1997. Inhibition of Bacterial RNA Polymerases. Peptide Metabolites from the Cultures of Streptomyces sp. *Journal of Natural Products*. **60**(8), pp.858–861.

Miao, V., Coëffet-LeGal, M.-F., Brian, P., Brost, R., Penn, J., Whiting, A., Martin, S., Ford, R., Parr, I., Bouchard, M., Silva, C.J., Wrigley, S.K. and Baltz, R.H. 2005. Daptomycin biosynthesis in Streptomyces roseosporus: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology (Reading, England)*. **151**(Pt 5), pp.1507–1523.

Miller, V.L. and Mekalanos, J.J. 1988. A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in Vibrio cholerae requires toxR. *Journal of Bacteriology*. **170**(6), pp.2575–2583.

Mudalungu, C.M., von Törne, W.J., Voigt, K., Rückert, C., Schmitz, S., Sekurova, O.N., Zotchev, S.B. and Süssmuth, R.D. 2019. Noursamycins, Chlorinated Cyclohexapeptides Identified from Molecular Networking of Streptomyces noursei NTR-SR4. *Journal of Natural Products*. **82**(6), pp.1478–1486.

Mulichak, A.M., Lu, W., Losey, H.C., Walsh, C.T. and Garavito, R.M. 2004. Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: interactions with acceptor and nucleotide ligands. *Biochemistry*. **43**(18), pp.5170–5180.

Munita, J.M. and Arias, C.A. 2016. Mechanisms of Antibiotic Resistance. *Microbiology spectrum*. **4**(2).

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. and Vagin, A.A. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*. **67**(4), pp.355–367.

Musiol-Kroll, E.M. and Wohlleben, W. 2018. Acyltransferases as Tools for Polyketide Synthase Engineering. *Antibiotics*. **7**(3), p.62.

Myronovskyi, M. and Luzhetskyy, A. 2019. Heterologous production of small molecules in the optimized Streptomyces hosts. *Natural Product Reports*. **36**(9), pp.1281–1294.

Nah, H.-J., Pyeon, H.-R., Kang, S.-H., Choi, S.-S. and Kim, E.-S. 2017. Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in Streptomyces Species. *Frontiers in Microbiology*. **8**(1).

Nakamura, H., X. Wang, J. and P. Balskus, E. 2015. Assembly line termination in cylindrocyclophane biosynthesis: discovery of an editing type II thioesterase domain in a type I polyketide synthase. *Chemical Science*. **6**(7), pp.3816–3822.

Navarro-Muñoz, J.C., Selem-Mojica, N., Mullowney, M.W., Kautsar, S.A., Tryon, J.H., Parkinson, E.I., De Los Santos, E.L.C., Yeong, M., Cruz-Morales, P., Abubucker, S., Roeters, A., Lokhorst, W., Fernandez-Guerra, A., Cappelini, L.T.D., Goering, A.W., Thomson, R.J., Metcalf, W.W., Kelleher, N.L., Barona-Gomez, F. and Medema, M.H. 2020. A computational framework to explore large-scale biosynthetic diversity. *Nature Chemical Biology*. **16**(1), pp.60–68.

Newman, D.J. and Cragg, G.M. 2016. Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products*. **79**(3), pp.629–661.

Ninomiya, A., Katsuyama, Y., Kuranaga, T., Miyazaki, M., Nogi, Y., Okada, S., Wakimoto, T., Ohnishi, Y., Matsunaga, S. and Takada, K. 2016. Biosynthetic Gene Cluster for Surugamide A Encompasses an Unrelated Decapeptide, Surugamide F. *ChemBioChem*. **17**(18), pp.1709–1712.

Nuijens, T., Toplak, A., Schmidt, M., Ricci, A. and Cabri, W. 2019. Natural Occurring and Engineered Enzymes for Peptide Ligation and Cyclization. *Frontiers in Chemistry*. **7**(1).

O' Neill, J. 2016. Report on Antimicrobial Resistance.

Oefner, C., Schulz, H., D'Arcy, A. and Dale, G.E. 2006. Mapping the active site of Escherichia coli malonyl-CoA–acyl carrier protein transacylase (FabD) by protein crystallography. *Acta Crystallographica Section D: Biological Crystallography*. **62**(6), pp.613–618.

Olano, C., Méndez, C. and Salas, J.A. 2010. Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Natural Product Reports*. **27**(4), pp.571–616.

Paananen, P., Patrikainen, P., Kallio, P., Mäntsälä, P., Niemi, J., Niiranen, L. and Metsä-Ketelä, M. 2013. Structural and functional analysis of angucycline C-6 ketoreductase LanV involved in landomycin biosynthesis. *Biochemistry*. **52**(31), pp.5304–5314.

Patrikainen, P., Niiranen, L., Thapa, K., Paananen, P., Tähtinen, P., Mäntsälä, P., Niemi, J. and Metsä-Ketelä, M. 2014. Structure-Based Engineering of Angucyclinone 6-Ketoreductases. *Chemistry & Biology*. **21**(10), pp.1381–1391.

van Pée, K.-H. and Patallo, E.P. 2006. Flavin-dependent halogenases involved in secondary metabolism in bacteria. *Applied Microbiology and Biotechnology*. **70**(6), pp.631–641.

Penfold, R.J. and Pemberton, J.M. 1992. An improved suicide vector for construction of chromosomal insertion mutations in bacteria. *Gene*. **118**(1), pp.145–146.

Petkovic, H., Lill, R.E., Sheridan, R.M., Wilkinson, B., McCormick, E.L., McAthrur, H. A. I., Staunton, J., Leadlay, P.F. and Kendrew, S.G. 2003. A Novel Erythromycin, 6-Desmethyl Erythromycin D, Made by Substituting an Acyltransferase Domain of the Erythromycin Polyketide Synthase. *The Journal of Antibiotics*. **56**(6), pp.543–551.

Peypoux, F., Bonmatin, J.M. and Wallach, J. 1999. Recent trends in the biochemistry of surfactin. *Applied Microbiology and Biotechnology*. **51**(5), pp.553–563.

Poust, S., Hagen, A., Katz, L. and Keasling, J.D. 2014. Narrowing the gap between the promise and reality of polyketide synthases as a synthetic biology platform. *Current Opinion in Biotechnology*. **30**, pp.32–39.

Prescott, J.F. 2014. The resistance tsunami, antimicrobial stewardship, and the golden age of microbiology. *Veterinary Microbiology*. **171**(3), pp.273–278.

Quadri, L.E., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P. and Walsh, C.T. 1998. Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry*. **37**(6), pp.1585–1595.

Ray, L., Valentic, T.R., Miyazawa, T., Withall, D.M., Song, L., Milligan, J.C., Osada, H., Takahashi, S., Tsai, S.-C. and Challis, G.L. 2016. A crotonyl-CoA reductase-carboxylase independent pathway for assembly of unusual alkylmalonyl-CoA polyketide synthase extender units. *Nature Communications*. **7**(1), p.13609.

Reeves, C.D., Murli, S., Ashley, G.W., Piagentini, M., Hutchinson, C.R. and McDaniel, R. 2001. Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. *Biochemistry*. **40**(51), pp.15464–15470.

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

Reimer, J.M., Eivaskhani, M., Harb, I., Guarné, A., Weigt, M. and Schmeing, T.M. 2019. Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science*. **366**(6466).

Reimer, J.M., Haque, A.S., Tarry, M.J. and Schmeing, T.M. 2018. Piecing together nonribosomal peptide synthesis. *Current Opinion in Structural Biology*. **49**, pp.104–113.

Reisch, C.R. and Prather, K.L.J. 2015. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in Escherichia coli. *Scientific Reports*. **5**, p.15096.

Revill, W.P., Bibb, M.J. and Hopwood, D.A. 1996. Relationships between fatty acid and polyketide synthases from Streptomyces coelicolor A3(2): characterization of the fatty acid synthase acyl carrier protein. *Journal of Bacteriology*. **178**(19), pp.5660–5667.

Ribeiro da Cunha, B., Fonseca, L.P. and Calado, C.R.C. 2019. Antibiotic Discovery: Where Have We Come from, Where Do We Go? *Antibiotics*. **8**(2).

Rix, U., Fischer, C., Remsing, L.L. and Rohr, J. 2002. Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Natural Product Reports*. **19**(5), pp.542–580.

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

Rowe, C.J., Böhm, I.U., Thomas, I.P., Wilkinson, B., Rudd, B.A.M., Foster, G., Blackaby, A.P., Sidebottom, P.J., Roddis, Y., Buss, A.D., Staunton, J. and Leadlay, P.F. 2001. Engineering a polyketide with a longer chain by insertion of an extra module into the erythromycin-producing polyketide synthase. *Chemistry & Biology*. **8**(5), pp.475–485.

Ruan, X., Pereda, A., Stassi, D.L., Zeidner, D., Summers, R.G., Jackson, M., Shivakumar, A., Kakavas, S., Staver, M.J., Donadio, S. and Katz, L. 1997. Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives. *Journal of Bacteriology*. **179**(20), pp.6416–6425.

Samel, S.A., Czodrowski, P. and Essen, L.-O. 2014. Structure of the epimerization domain of tyrocidine synthetase A. *Acta Crystallographica Section D: Biological Crystallography*. **70**(5), pp.1442–1452.

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

Sandy, M., Rui, Z., Gallagher, J. and Zhang, W. 2012. Enzymatic Synthesis of Dilactone Scaffold of Antimycins. *ACS Chemical Biology*. **7**(12), pp.1956–1961.

Sandy, M., Zhu, X., Rui, Z. and Zhang, W. 2013. Characterization of AntB, a Promiscuous Acyltransferase Involved in Antimycin Biosynthesis. *Organic Letters*. **15**(13), pp.3396–3399.

Scheuermann, T.H. and Brautigam, C.A. 2015. High-precision, automated integration of multiple isothermal titration calorimetric thermograms: new features of NITPIC. *Methods (San Diego, Calif.)*. **76**, pp.87–98.

Schoenian, I., Paetz, C., Dickschat, J.S., Aigle, B., Leblond, P. and Spiteller, D. 2012. An Unprecedented 1,2-Shift in the Biosynthesis of the 3-Aminosalicylate Moiety of Antimycins. *ChemBioChem*. **13**(6), pp.769–773.

Schoppet, M., Peschke, M., Kirchberg, A., Wiebach, V., D. Süssmuth, R., Stegmann, E. and J. Cryle, M. 2019. The biosynthetic implications of late-stage condensation domain

selectivity during glycopeptide antibiotic biosynthesis. *Chemical Science*. **10**(1), pp.118–133.

Schracke, N., Linne, U., Mahlert, C. and Marahiel, M.A. 2005. Synthesis of Linear Gramicidin Requires the Cooperation of Two Independent Reductases. *Biochemistry*. **44**(23), pp.8507–8513.

Schuck, P. 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophysical Journal*. **78**(3), pp.1606–1619.

Schwarzer, D., Finking, R. and Marahiel, M.A. 2003. Nonribosomal peptides: from genes to products. *Natural Product Reports*. **20**(3), pp.275–287.

Schwarzer, D., Mootz, H.D., Linne, U. and Marahiel, M.A. 2002. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proceedings of the National Academy of Sciences*. **99**(22), pp.14083–14088.

Scott, T.A. and Piel, J. 2019. The hidden enzymology of bacterial natural product biosynthesis. *Nature Reviews Chemistry*. **3**(7), pp.404–425.

Seidel, J., Miao, Y., Porterfield, W., Cai, W., Zhu, X., Kim, S.-J., Hu, F., Bhattarai-Kline, S., Min, W. and Zhang, W. 2019. Structure–activity–distribution relationship study of anti-cancer antimycin-type depsipeptides. *Chemical Communications*. **55**(63), pp.9379–9382.

Seipke, R.F., Barke, J., Brearley, C., Hill, L., Yu, D.W., Goss, R.J.M. and Hutchings, M.I. 2011. A single Streptomyces symbiont makes multiple antifungals to support the fungus farming ant Acromyrmex octospinosus. *PloS One*. **6**(8), p.e22028.

Seipke, R.F. and Hutchings, M.I. 2013a. The regulation and biosynthesis of antimycins. *Beilstein Journal of Organic Chemistry*. **9**(1), pp.2556–2563.

Seipke, R.F. and Hutchings, M.I. 2013b. The regulation and biosynthesis of antimycins. *Beilstein Journal of Organic Chemistry*. **9**, pp.2556–2563.

Seipke, R.F., Patrick, E. and Hutchings, M.I. 2014. Regulation of antimycin biosynthesis by the orphan ECF RNA polymerase sigma factor σ (AntA.). *PeerJ*. **2**, p.e253.

Shaw-Reid, C.A., Kelleher, N.L., Losey, H.C., Gehring, A.M., Berg, C. and Walsh, C.T. 1999. Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E. coli EntF catalyzes both elongation and cyclolactonization. *Chemistry & Biology*. **6**(6), pp.385–400.

Shen, B. 2003. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology*. **7**(2), pp.285–295.

Sieber, S.A. and Marahiel, M.A. 2003. Learning from Nature's Drug Factories: Nonribosomal Synthesisof MacrocyclicPeptides. *Journal of Bacteriology*. **185**(24), pp.7036–7043.

Silver, L.L. 2011. Challenges of Antibacterial Discovery. *Clinical Microbiology Reviews*. **24**(1), pp.71–109.

Siskos, A.P., Baerga-Ortiz, A., Bali, S., Stein, V., Mamdani, H., Spiteller, D., Popovic, B., Spencer, J.B., Staunton, J., Weissman, K.J. and Leadlay, P.F. 2005. Molecular basis of Celmer's rules: stereochemistry of catalysis by isolated ketoreductase domains from modular polyketide synthases. *Chemistry & Biology*. **12**(10), pp.1145–1153.

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

Skyrud, W., Liu, J., Thankachan, D., Cabrera, M., Seipke, R.F. and Zhang, W. 2018. Biosynthesis of the 15-Membered Ring Depsipeptide Neoantimycin. *ACS chemical biology*. **13**(5), pp.1398–1406.

Smith, S. and Tsai, S.-C. 2007. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Natural Product Reports*. **24**(5), pp.1041–1072.

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., Kim, B.Y., Jang, J.-H. and Ahn, J.S. 2017. Genomics-Driven Discovery of Chlorinated Cyclic Hexapeptides Ulleungmycins A and B from a StreptomycesSpecies. *Journal of natural products*. **80**(11), pp.3025–3031.

Song, Y., Li, Q., Liu, X., Chen, Y., Zhang, Y., Sun, A., Zhang, W., Zhang, J. and Ju, J. 2014. Cyclic Hexapeptides from the Deep South China Sea-Derived Streptomyces scopuliridis SCSIO ZJ46 Active Against Pathogenic Gram-Positive Bacteria. *Journal of natural products*. **77**(8), pp.1937–1941.

Stachelhaus, T., Mootz, H.D., Bergendahl, V. and Marahiel, M.A. 1998. Peptide Bond Formation in Nonribosomal Peptide Biosynthesis: Catalytic Role Of The Condensation Domain. *Journal of Biological Chemistry*. **273**(35), pp.22773–22781.

Stachelhaus, T., Mootz, H.D. and Marahiel, M.A. 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & Biology*. **6**(8), pp.493–505.

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

Stanišić, A. and Kries, H. 2019. Adenylation Domains in Nonribosomal Peptide Engineering. *ChemBioChem*. **20**(11), pp.1347–1356.

Stassi, D.L., Kakavas, S.J., Reynolds, K.A., Gunawardana, G., Swanson, S., Zeidner, D., Jackson, M., Liu, H., Buko, A. and Katz, L. 1998. Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering. *Proceedings of the National Academy of Sciences*. **95**(13), pp.7305–7309.

Staunton, J. 1998. Combinatorial biosynthesis of erythromycin and complex polyketides. *Current Opinion in Chemical Biology*. **2**(3), pp.339–345.

Staunton, J. and Weissman, K.J. 2001. Polyketide biosynthesis: a millennium review. *Natural Product Reports*. **18**(4), pp.380–416.

Stein, D.B., Linne, U. and Marahiel, M.A. 2005. Utility of epimerization domains for the redesign of nonribosomal peptide synthetases. *The FEBS Journal*. **272**(17), pp.4506–4520.

Stein, N. 2008. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography*. **41**(3), pp.641–643.

Steller, S., Sokoll, A., Wilde, C., Bernhard, F., Franke, P. and Vater, J. 2004. Initiation of Surfactin Biosynthesis and the Role of the SrfD-Thioesterase Protein. *Biochemistry*. **43**(35), pp.11331–11343.

Strieker, M., Nolan, E.M., Walsh, C.T. and Marahiel, M.A. 2009. Stereospecific Synthesis of threo- and erythro-β-Hydroxyglutamic Acid During Kutzneride Biosynthesis. *Journal of the American Chemical Society*. **131**(37), pp.13523–13530.

Strieker, M., Tanović, A. and Marahiel, M.A. 2010. Nonribosomal peptide synthetases: structures and dynamics. *Current Opinion in Structural Biology*. **20**(2), pp.234–240.

Sundermann, U., Bravo-Rodriguez, K., Klopries, S., Kushnir, S., Gomez, H., Sanchez-Garcia, E. and Schulz, F. 2013. Enzyme-directed mutasynthesis: a combined experimental and theoretical approach to substrate recognition of a polyketide synthase. *ACS chemical biology*. **8**(2), pp.443–450.

Süssmuth, R.D. and Mainz, A. 2017. Nonribosomal Peptide Synthesis—Principles and Prospects. *Angewandte Chemie International Edition*. **56**(14), pp.3770–3821.

Takada, K., Ninomiya, A., Naruse, M., Sun, Y., Miyazaki, M., Nogi, Y., Okada, S. and Matsunaga, S. 2013. Surugamides A–E, Cyclic Octapeptides with Four d-Amino Acid Residues, from a Marine Streptomyces sp.: LC–MS-Aided Inspection of Partial Hydrolysates for the Distinction of d- and l-Amino Acid Residues in the Sequence. *The Journal of Organic Chemistry*. **78**(13), pp.6746–6750.

Tang, L. and McDaniel, R. 2001. Construction of desosamine containing polyketide libraries using a glycosyltransferase with broad substrate specificity. *Chemistry & Biology*. **8**(6), pp.547–555.

Tang, Y., Chen, A.Y., Kim, C.-Y., Cane, D.E. and Khosla, C. 2007. Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chemistry & Biology*. **14**(8), pp.931–943.

Tang, Y., Kim, C.-Y., Mathews, I.I., Cane, D.E. and Khosla, C. 2006. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proceedings of the National Academy of Sciences of the United States of America*. **103**(30), pp.11124–11129.

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

Tappel, A.L. 1960. Inhibition of electron transport by antimycin A, alkyl hydroxy naphthoquinones and metal coordination compounds. *Biochemical Pharmacology*. **3**, pp.289–296.

Tarry, M.J., Haque, A.S., Bui, K.H. and Schmeing, T.M. 2017. X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure*. **25**(5), pp.783-793.e4.

Thackray, S.J., Mowat, C.G. and Chapman, S.K. 2008. Exploring the mechanism of tryptophan 2,3-dioxygenase. *Biochemical Society Transactions*. **36**(Pt 6), pp.1120–1123.

Thankachan, D., Fazal, A., Francis, D., Song, L., Webb, M.E. and Seipke, R.F. 2019. A trans-Acting Cyclase Offloading Strategy for Nonribosomal Peptide Synthetases. *ACS chemical biology*. **14**(5), pp.845–849.

Trauger, J.W., Kohli, R.M., Mootz, H.D., Marahiel, M.A. and Walsh, C.T. 2000. Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature*. **407**(6801), pp.215–218.

Trott, O. and Olson, A.J. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of computational chemistry*. **31**(2), pp.455–461.

Tsai, S.-C., Lu, H., Cane, D.E., Khosla, C. and Stroud, R.M. 2002. Insights into Channel Architecture and Substrate Specificity from Crystal Structures of Two Macrocycle-Forming Thioesterases of Modular Polyketide Synthases,. *Biochemistry*. **41**(42), pp.12598–12606.

Tsai, S.-C., Miercke, L.J.W., Krucinski, J., Gokhale, R., Chen, J.C.-H., Foster, P.G., Cane, D.E., Khosla, C. and Stroud, R.M. 2001. Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: Versatility from a unique substrate channel. *Proceedings of the National Academy of Sciences*. **98**(26), pp.14808–14813.

Tzung, S.-P., Kim, K.M., Basañez, G., Giedt, C.D., Simon, J., Zimmerberg, J., Zhang, K.Y.J. and Hockenbery, D.M. 2001. Antimycin A mimics a cell-death-inducing Bcl-2 homology domain 3. *Nature Cell Biology*. **3**(2), pp.183–191.

Ventola, C.L. 2015. The Antibiotic Resistance Crisis. *Pharmacy and Therapeutics*. **40**(4), pp.277–283.

Waksman, S.A., Schatz, A. and Reynolds, D.M. 2010. Production of antibiotic substances by actinomycetes. *Annals of the New York Academy of Sciences*. **1213**, pp.112–124.

Wallace, K.K., Bao, Z.-Y., Dai, H., Digate, R., Schuler, G., Speedie, M.K. and Reynolds, K.A. 1995. Purification of Crotonyl-CoA Reductase from Streptomyces collinus and Cloning, Sequencing and Expression of the Corresponding Gene in Escherichia coli. *European Journal of Biochemistry*. **233**(3), pp.954–962.

Walsh, C.T. 2016. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Natural Product Reports*. **33**(2), pp.127–135.

Walsh, C.T., Chen, H., Keating, T.A., Hubbard, B.K., Losey, H.C., Luo, L., Marshall, C.G., Miller, D.A. and Patel, H.M. 2001. Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Current Opinion in Chemical Biology*. **5**(5), pp.525–534.

Wang, F., Wang, Y., Ji, J., Zhou, Z., Yu, J., Zhu, H., Su, Z., Zhang, L. and Zheng, J. 2015. Structural and Functional Analysis of the Loading Acyltransferase from Avermectin Modular Polyketide Synthase. *ACS Chemical Biology*. **10**(4), pp.1017–1025.

Warkentin, E., Mamat, B., Sordel-Klippert, M., Wicke, M., Thauer, R.K., Iwata, M., Iwata, S., Ermler, U. and Shima, S. 2001. Structures of F420H2:NADP+ oxidoreductase with and without its substrates bound. *The EMBO Journal*. **20**(23), pp.6561–6569.

Webb, M.E., Marquet, A., Mendel, R.R., Rébeillé, F. and Smith, A.G. 2007. Elucidating biosynthetic pathways for vitamins and cofactors. *Natural Product Reports*. **24**(5), pp.988–1008.

Weissman, K.J. 2016. Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. *Natural Product Reports*. **33**(2), pp.203–230.

Weissman, K.J. and Leadlay, P.F. 2005. Combinatorial biosynthesis of reduced polyketides. *Nature Reviews Microbiology*. **3**(12), pp.925–936.

Weissman, K.J. and Müller, R. 2008. Protein–Protein Interactions in Multienzyme Megasynthetases. *ChemBioChem*. **9**(6), pp.826–848.

Wenzel, S.C. and Müller, R. 2005. Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Current Opinion in Biotechnology*. **16**(6), pp.594–606.

Whicher, J.R., Dutta, S., Hansen, D.A., Hale, W.A., Chemler, J.A., Dosey, A.M., Narayan, A.R.H., Håkansson, K., Sherman, D.H., Smith, J.L. and Skiniotis, G. 2014. Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature*. **510**(7506), pp.560–564.

Wilkinson, B. and Micklefield, J. 2009. Chapter 14. Biosynthesis of nonribosomal peptide precursors *In*: *Methods in Enzymology*., pp.353–378.

Wilkinson, B. and Micklefield, J. 2007. Mining and engineering natural-product biosynthetic pathways. *Nature Chemical Biology*. **3**(7), pp.379–386.

Wilson, M.C. and Moore, B.S. 2012. Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Natural Product Reports*. **29**(1), pp.72–86.

Winn, M., K. Fyans, J., Zhuo, Y. and Micklefield, J. 2016. Recent advances in engineering nonribosomal peptide assembly lines. *Natural Product Reports*. **33**(2), pp.317–347.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A. and Wilson,

K.S. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*. **67**(Pt 4), pp.235–242.

Wu, N., Tsuji, S.Y., Cane, D.E. and Khosla, C. 2001. Assessing the Balance between Protein−Protein Interactions and Enzyme−Substrate Interactions in the Channeling of Intermediates between Polyketide Synthase Modules. *Journal of the American Chemical Society*. **123**(27), pp.6465–6474.

Xiao, X., Elsayed, S.S., Wu, C., van der Heul, H.U., Metsä-Ketelä, M., Du, C., Prota, A.E., Chen, C.-C., Liu, W., Guo, R.-T., Abrahams, J.P. and van Wezel, G.P. 2020. Functional and Structural Insights into a Novel Promiscuous Ketoreductase of the Lugdunomycin Biosynthetic Pathway. *ACS Chemical Biology*. **15**(9), pp.2529–2538.

Xie, X. and Cane, D.E. 2018. pH-Rate profiles establish that polyketide synthase dehydratase domains utilize a single-base mechanism. *Organic & Biomolecular Chemistry*. **16**(47), pp.9165–9170.

Xie, X., Garg, A., Keatinge-Clay, A.T., Khosla, C. and Cane, D.E. 2016. The Epimerase and Reductase Activities of Polyketide Synthase Ketoreductase Domains Utilize the Same Conserved Tyrosine and Serine Residues. *Biochemistry*. **55**(8), pp.1179–1186.

Xie, X., Meehan, M.J., Xu, W., Dorrestein, P.C. and Tang, Y. 2009. Acyltransferase mediated polyketide release from a fungal megasynthase. *Journal of the American Chemical Society*. **131**(24), pp.8388–8389.

Xie, Y., Wang, B., Liu, J., Zhou, J., Ma, J., Huang, H. and Ju, J. 2012. Identification of the biosynthetic gene cluster and regulatory cascade for the synergistic antibacterial antibiotics griseoviridin and viridogrisein in Streptomyces griseoviridis. *Chembiochem: A European Journal of Chemical Biology*. **13**(18), pp.2745–2757.

Xu, F., Nazari, B., Moon, K., Bushin, L.B. and Seyedsayamdost, M.R. 2017. Discovery of a Cryptic Antifungal Compound from Streptomyces albusJ1074 Using High-Throughput Elicitor Screens. *Journal of the American Chemical Society*. **139**(27), pp.9203–9212.

Yan, Y., Liu, N. and Tang, Y. 2020. Recent developments in self-resistance gene directed natural product discovery. *Natural Product Reports*. **37**(7), pp.879–892.

Yan, Y., Zhang, L., Ito, T., Qu, X., Asakawa, Y., Awakawa, T., Abe, I. and Liu, W. 2012. Biosynthetic Pathway for High Structural Diversity of a Common Dilactone Core in Antimycin Production. *Organic Letters*. **14**(16), pp.4142–4145.

Yeh, E., Blasiak, L.C., Koglin, A., Drennan, C.L. and Walsh, C.T. 2007. Chlorination by a Long-Lived Intermediate in the Mechanism of Flavin-Dependent Halogenases. *Biochemistry*. **46**(5), pp.1284–1292.

Yeh, E., Kohli, R.M., Bruner, S.D. and Walsh, C.T. 2004. Type II Thioesterase Restores Activity of a NRPS Module Stalled with an Aminoacyl-S-enzyme that Cannot Be Elongated. *ChemBioChem*. **5**(9), pp.1290–1293.
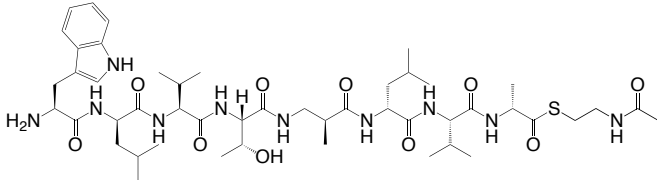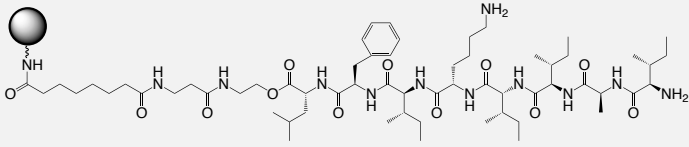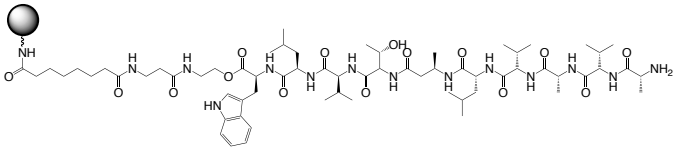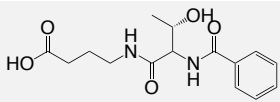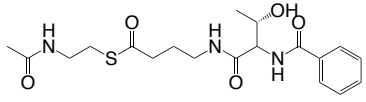
Yin, X. and Zabriskie, T.M. 2004. VioC is a Non-Heme Iron, α-Ketoglutarate-Dependent Oxygenase that Catalyzes the Formation of 3S-Hydroxy-L-Arginine during Viomycin Biosynthesis. *ChemBioChem*. **5**(9), pp.1274–1277.

Zaman, S.B., Hussain, M.A., Nye, R., Mehta, V., Mamun, K.T. and Hossain, N. 2017. A Review on Antibiotic Resistance: Alarm Bells are Ringing. *Cureus*. **9**(6).

Zehner, S., Kotzsch, A., Bister, B., Süssmuth, R.D., Méndez, C., Salas, J.A. and van Pée, K.-H. 2005. A Regioselective Tryptophan 5-Halogenase Is Involved in Pyrroindomycin Biosynthesis in Streptomyces rugosporus LL-42D005. *Chemistry & Biology*. **12**(4), pp.445–452.

Zhang, H., A. Boghigian, B., Armando, J. and A. Pfeifer, B. 2011. Methods and options for the heterologous production of complex natural products. *Natural Product Reports*. **28**(1), pp.125–151.

Zhang, J.J., Tang, X. and Moore, B.S. 2019. Genetic platforms for heterologous expression of microbial natural products. *Natural Product Reports*. **36**(9), pp.1313–1332.

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

Zhang, L., Mori, T., Zheng, Q., Awakawa, T., Yan, Y., Liu, W. and Abe, I. 2015. Rational Control of Polyketide Extender Units by Structure-Based Engineering of a Crotonyl-CoA Carboxylase/Reductase in Antimycin Biosynthesis. *Angewandte Chemie International Edition*. **54**(45), pp.13462–13465.

Zhang, M.M., Wong, F.T., Wang, Y., Luo, S., Lim, Y.H., Heng, E., Yeo, W.L., Cobb, R.E., Enghiad, B., Ang, E.L. and Zhao, H. 2017. CRISPR–Cas9 strategy for activation of silent Streptomyces biosynthetic gene clusters. *Nature Chemical Biology*. **13**(6), pp.607–609.

Zhang, Z., Pan, H.-X. and Tang, G.-L. 2017. New insights into bacterial type II polyketide biosynthesis. *F1000Research*. **6**.

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

Zhou, Y., Lin, X., Williams, S.R., Liu, L., Shen, Y., Wang, S.-P., Sun, F., Xu, S., Deng, H., Leadlay, P.F. and Lin, H.-W. 2018. Directed Accumulation of Anticancer Depsipeptides by Characterization of Neoantimycins Biosynthetic Pathway and an NADPH-Dependent Reductase. *ACS chemical biology*. **13**(8), pp.2153–2160.

Zhou, Y., Lin, X., Xu, C., Shen, Y., Wang, S.-P., Liao, H., Li, L., Deng, H. and Lin, H.-W. 2019. Investigation of Penicillin Binding Protein (PBP)-like Peptide Cyclase and Hydrolase in Surugamide Non-ribosomal Peptide Biosynthesis. *Cell Chemical Biology*. **26**(5), pp.737-744.e4.

8 – Appendices

8.1 Appendix 1

Table 17: List and structures of chemically synthesised substrates.

Substrate number	Name	Structure
1	Linear Surugamide A	
2	Linear Surugamide A – SNAC	
3	Surugamide F	
4	Surugamide F – SNAC	
5	Truncated Surugamide A	
6	Truncated Surugamide A – SNAC	
7	Truncated Surugamide F	

8	Truncated Surugamide F – SNAC	
9	PEGA – linker – Surugamide A	
10	PEGA – linker – Surugamide F	
11	Antimycin mimic	
12	Antimycin mimic – SNAC	

8.2 Appendix 2

MS2 analysis of antimycin-type depsipeptides:
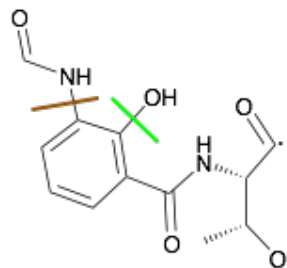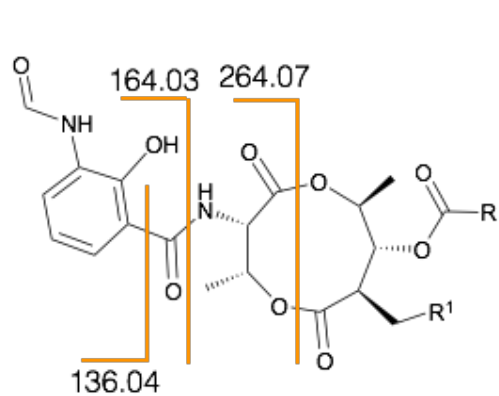
Mass of initial breakdown product = 264.07

Loss of —— = 220.06

Loss of —— = 247.07

(Note: Breakdown of antimycins also occurs readily under normal conditions of ESI-MS)