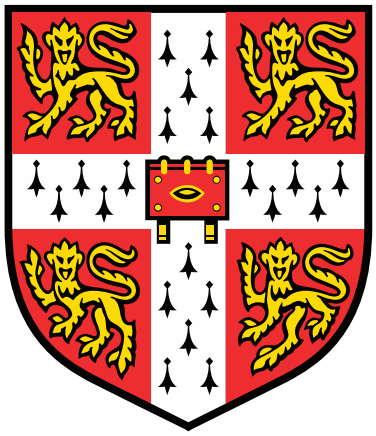
Mining and Manipulation of Antibiotic Biosynthesis in *Streptomyces*



Jacob A. Pollock

Christ's College

Department of Biochemistry

University of Cambridge

September 2018

This dissertation is submitted for the degree of Doctor of Philosophy

Mining and manipulation of antibiotic biosynthesis in *Streptomyces*

Jacob A. Pollock

New small-molecule drugs are needed, both to address existing disease and to combat the rise of antibiotic resistance in pathogenic microbes. A major source of antibiotics and other valuable therapeutic agents remains the natural products produced by *Streptomyces* and allied bacteria. The advent of genome-level sequencing has changed how such bioactive products are identified in two ways: first, by enabling new approaches to engineering of known clusters to obtain new analogues; and secondly by enabling mining of their large (8-12 Mbp) genomes to make rapid links between valuable compounds and the gene clusters responsible for their biosynthesis; as well as a complete inventory of orphan clusters for decryption. In this work, both of these aspects were explored.

Accelerated Evolution (AE) is a new method of induced recombination in *Streptomyces* that produces libraries of analogues of a parent compound in a single experiment. It has been previously applied to a typical assembly-line biosynthetic pathway involving a multimodular polyketide synthase, that for the immunosuppressant rapamycin. Here, the utility of AE was initially explored with several polyketide antibiotics, before focussing on the polyene filipin. Unfortunately, this system proved refractory to the AE method, since only recombinants which had either lost filipin production, or which maintained some level of filipin production, were recovered. These results suggested that deeper understanding of the fundamental processes of recombination in *Streptomyces* are now needed to allow the AE approach to flourish.

A genome mining approach was taken to identify and analyse the gene cluster in *Streptomyces albus* DMS40763 for pseudouridimycin (PUM), a rare C-nucleoside antibiotic and a newly-recognised selective inhibitor of bacterial RNA polymerase. The fully-sequenced genome (8.01 Mbp) of *S. albus* was analysed and 27 biosynthetic gene clusters were found, including those for 11 novel assembly-line systems. A strong candidate for the PUM gene cluster was identified through BLAST searches with a gene probe (*truD*) found in the gene cluster for a different C-nucleoside in other *Streptomyces* spp., hinting at a common mechanism for formation of the C-nucleoside moiety. Identification of the enzymes predicted to be encoded by the gene cluster allowed a detailed biosynthetic pathway to be proposed. Several genes from the PUM cluster were expressed and purified as recombinant proteins in *Escherichia coli*, and the proposed enzymatic roles of two of them were verified. One of these, a pseudouridine oxidase, also showed activity against uridine. Consistent with this finding, feeding of uracil to a mutant *S. albus* lacking the pseudouridine synthase-like gene *truD* gave production of the novel N-linked analogue of PUM, suggesting that the biosynthetic enzymes may be sufficiently flexible to incorporate other, non-natural nucleosides into PUM analogues for testing.

Preface

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text

This thesis does not exceed the prescribed word limit.

Acknowledgements

I owe a great debt to the many people who have supported me during my time at Cambridge. I wish to thank all my friends and colleagues for their time, advice and good humour that has helped me through my PhD.

First and foremost, I wish to thank my supervisor Peter Leadlay for inviting me to join his group and for his endless enthusiasm and patience. He has been an excellent source of inspiration and after every meeting I have left his office with fresh ideas to pursue and the motivation to chase them.

Secondly, I want to thank the Biotechnology and Biological Sciences Research Council for the opportunity they have given me to attend Cambridge and work towards my PhD in the Biochemistry Department.

I wish to thank the many members of the Leadlay lab who have helped me over the four years I have been here. They have all been great company and have always been willing to listen and help me with any problems I ran into. I especially want to thank Anya, who's patience and incredible knowledge and skill was invaluable to me during the start of my time in the Leadlay lab, and Hui, who has given me so much assistance with the later parts of my PhD.

I would also like to thank Bill Broadhurst for his very kind help with the NMR experiments and figures in this work.

I want to thank my sister who has been very supportive, and very kindly helped me to edit this thesis. I want to thank my parents for everything they have done for me. I would not be where I am without their generosity.

Finally, I want to thank Liz whose love has made me stronger.

Contents

Abbreviations

6-dEB	6-deoxyerythronolide B
A570	absorbance at 570 nm
A600	absorbance at 600 nm
ACP	acyl carrier protein
AE	accelerated evolution
Amp	ampicillin
AntiSMASH	antibiotics & Secondary Metabolite Analysis SHell
Apr	apramycin
APU	5'-aminopseudouridine
AT	acyltransferase
BLAST	basic local alignment search tool
bp	base pair
Cam	chloramphenicol
CoA	coenzyme A
DEBS	6-deoxyerythronolide B synthase
DH	dehydratase
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
ER	enoylreductase
ESI	electrospray ionisation
FA	formic acid
FAS	fatty acid synthase
gDNA	genomic DNA
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high performance liquid chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
Kan	kanamycin
kbp	kilo base pairs
kDa	kilo Dalton
KR	ketoreductase
KS	ketosynthase
LCMS	liquid chromatography-mass spectrometry
m/v	mass to volume ratio
m/z	mass to charge ratio
Mbp	mega base pairs
MDR	multidrug-resistant
MS	mass spectrometry
MT	methyltransferase
N-PUM	*N*-nucleoside isomer of pseudouridimycin
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction

PDB	protein data bank
PEG	polyethylene glycol
PK	polyketide
PKS	polyketide synthase
PPTase	4'-phosphopantetheinyl transferase
PUM	pseudouridimycin
R5P	ribose-5'-phosphate
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
ssDNA	single-stranded DNA
TE	thioesterase
TFA	trifluoroacetic acid
Thio	thiostrepton
tRNA	transfer ribonucleic acid
TSB	tryptic soya broth
U	uridine
UV	ultraviolet light
v/v	volume to volume ratio
Vis	visible light
WT	wild type
Ψ	pseudouridine

Chapter 1: Introduction

1.1: Actinomycete Natural Products and their Targets

1.1.1: The Role of Natural Products in Medicine

Natural products, and the field of research that attempts to understand and eventually control them, lie at the interface between chemistry and biology. Their complex, diverse structures have captured the imagination of chemical scientists and led to a flowering of modern methods of stereoselective total synthesis [1–3]. Meanwhile, the exquisite specificity of action of many natural products has been invaluable in the mapping of metabolic pathways and regulatory cascades. Last but not least, natural products have been a continuing source of leads and inspiration in drug discovery [4–6]

The simplest definition of a natural product is a small molecule that is produced by a biological source [7]. The term natural product generally refers to secondary metabolites. These are non-essential molecules; they are not required for the growth, development and reproduction of the producing organism. The function of secondary metabolites has been widely discussed and there have been a number of views as to why they exist at all; Bu'Lock, who coined the term 'secondary metabolite' stated "it is the process of secondary biosynthesis which is seen as advantageous, and not, in the general case, its products, and in my opinion this is still the general type of explanation" [8]; Zähner *et al* held the view that we should "rid ourselves of the simplistic idea that antibiotics are formed as a defence mechanism, and recognise instead that antibiotics are nothing more than secondary metabolites which possess, more or less incidentally, an antibiotic effect" [9]; while Williams *et al* argue that "all natural products have evolved under the pressure of natural selection to bind to specific receptors" [10]. Perhaps the most widely-accepted view now is that of Firn and Jones [11], who identify the ability to produce such bioactive specialised metabolites as the valuable selectable trait, rather than the ability to produce a particular molecule and that the secondary metabolism will have evolved such that traits that optimise the production and retention of chemical diversity at a minimum cost will have been selected for.

The earliest records of natural products have been found on clay tablets from Mesopotamia, dated at 2600 BC. These documented oils from *Cupressus sempervirens* (Cypress) and *Commiphora* (Myrrh) species, which are still used to treat colds, coughs and inflammation [12]. Studies of the ancient Egyptian Ebers Papyrus identified about 800 complex prescriptions and over 700 natural agents such as *Aloe vera*, *Boswellia carteri* (frankincense) and the oil of *Ricinus communis* (castor oil) [13]. The first

treatise written on the concepts and practice of Indian Ayurveda was written almost 3000 years ago and contains hundreds of plant derived medicines [14]. Some of the most important drugs in use today have ancient roots; the anti-inflammatory agent, acetylsalicyclic acid (aspirin) is derived from salicin, isolated from the bark of *Salix alba* (willow tree) [15] and the alkaloid morphine is derived from the plant *Papaver somniferum* (opium poppy) [15] (Figure 1.1).

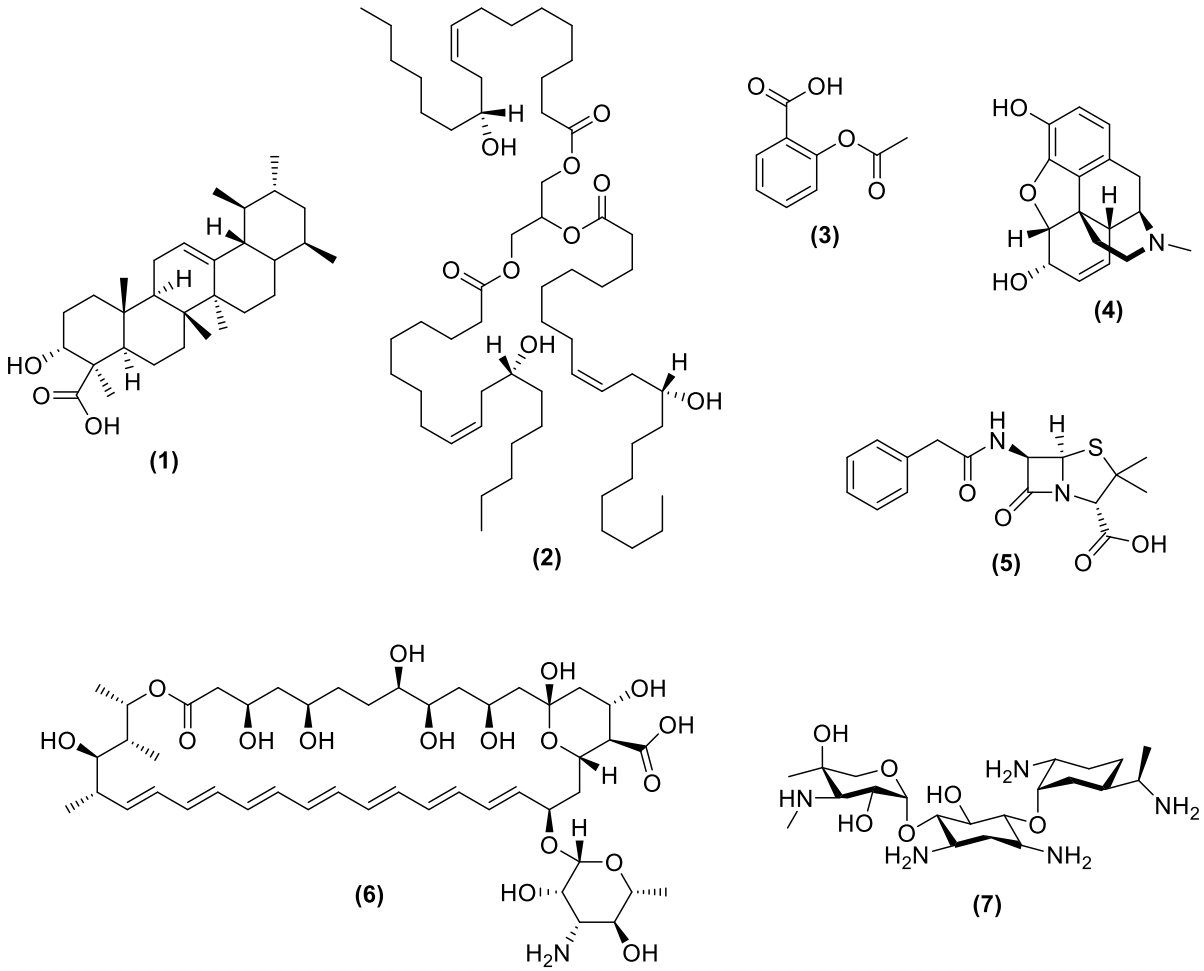


Figure 1.1: Historical and modern examples of therapeutic natural products. (1) β-boswellic acid, one of the active ingredients in frankincense **(2)** A trimester of glycerol and ricinoleic acid, the major component of castor oil **(3)** acetylsalicyclic acid **(4)** morphine **(5)** benzylpenicillin **(6)** amphotericin B **(7)** gentamicin.

The development of modern medicine has also been shaped in part by the impact of natural products. In 1929 Alexander Fleming famously discovered penicillin, derived from the fungus *Penicillium notatum*. The clinical studies and commercialisation of synthetic penicillins that followed revolutionised drug discovery and the practice of medicine, allowing operations to be performed at

much lower risk, and the treatment of previously life-threatening diseases. Anti-infectives remain an important class of natural products, as in 2016 three of the top ten global causes of death were infections, including tuberculosis and respiratory infections. Many of these compounds are found in the World Health Organisation (WHO) Model List of Essential Medicines, including amphotericin B, rifampicin, vancomycin, and gentamicin. Natural products possess a vast chemical diversity and have had a profound impact on the development of antitumour, immunosuppressnt and cholesterol-lowering compounds [16,17], however the focus of the work described here is on anti-infective natural products, their producers, and the genes and enzymes responsible for their biosynthesis.

Sixteen years after Alexander Fleming's discovery of penicillin, he shared the 1945 Nobel Prize in Physiology or Medicine for the discovery of penicillin. Then, in 1952, Selman A. Waksman received the Nobel Prize for the discovery of streptomycin [18], an antibiotic with activity against penicillin-resistant bacteria, and importantly exhibiting activity against *Mycobacterium tuberculosis*, the causative agent of tuberculosis. This was the first time natural products were systematically sought through screening, and it heralded the age of antibiotics and what is known as 'the Golden Age' of natural product drug discovery. Half of the drugs commonly used today and many of the recognised classes `of antibiotics were discovered during this period [19] (Figure 1.1).

1.1.2: Producers of Natural Products

Antibiotic-producing organisms are primarily bacterial, but it is important to note that plants and fungi also have specialised secondary metabolic pathways that can generate huge chemical diversity. Indeed, penicillin, one of the best known and most influential antibiotics, is derived from a fungus, *Penicillium notatum*. There are approximately 100,000 fungal species known, although the total expected number is far in excess of one million. The variety of species and their diversity means fungi are a rich source of new natural products. An extension of screening from traditionally used saprophytic terrestrial strains to marine and endophytic strains, accompanied by new screening strategies, has led to the discovery of many novel fungal metabolites [20]. Hypocreolide, a nonenolide derived from the terrestrial ascomycete *Hypocrea lactea*, shows antibacterial, antifungal, and cytotoxic activities[21]. Benzophomopsin A, a new benzoxepin isolated from an endophytic *Phomopsis* strain derived from a Japanese cherry tree, potently inhibits Ca^{2+} signalling [22]. The novel compounds; JBIR-37 and JBIR-38, isolated from the marine sponge-derived fungus *Acremonium* sp., do not exhibit any bioactivities, but they are the first examples of glycosyl benzenes from fungi [23], showing that advances in chemical screening techniques continue to identify interesting chemotypes.

Plants have historically been a primary source of medicines. In a pre-synthetic era it is estimated that 80% of medicines were obtained from roots, bark and leaves [24]. With an estimated 200,000 natural products being generated by plants[25], including many valuable alkaloids and terpenoids, such as artemisinin, plants are an important source of novel chemical diversity. Plant natural product biosynthesis tends to focus on the generation of key intermediate scaffolds which are then tailored, or enzymatically decorated, by a small collection of enzymatic transformations, to form different classes of natural products. The most widely used drug for the treatment of breast cancer is paclitaxel (Taxol), isolated from the bark of *Taxus brevifolia* (Pacific Yew). Initial screenings of the compound required the bark from three 100 year old trees to produce 1 gram of paclitaxel [26]. A total synthetic route has been developed to help meet the current annual demand for 100 to 200 kg of the compound [15], but it is expensive. Fortunately, baccatin III was found to be present in the needles of *T. brevifolia* in much higher quantities and can be used as a starting point for paclitaxel semi-synthesis. Unfortunately, only a handful of plant secondary metabolic pathways are fully defined [25], and this lack of knowledge of the genes and enzymes involved in bioynthesis has hindered progress. However there have been success stories, such as artemisinin. This potent antimalarial was first isolated from *Artemisia annua*, which was originally used in traditional Chinese medicine. The supply of plant-derived artemisinin is unstable, resulting in complications for drug manufacturers, but through huge effort, the pathway for a precursor to artemisinin, artemisinic acid, has been successfully engineered into *Saccharomyces cerevisiae* [27], and an elegant chemical conversion devised to turn this into the bioactive natural product.

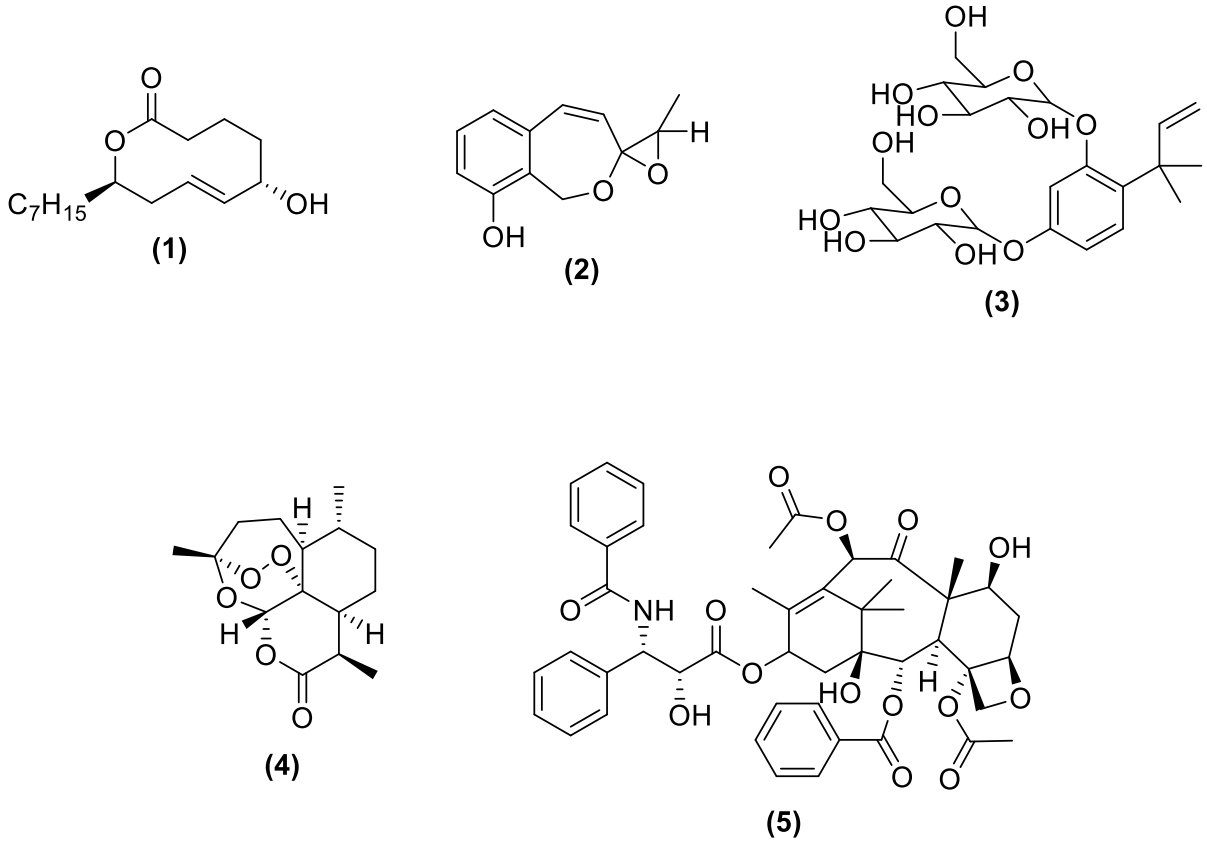


Figure 1.2: Natural products derived from fungi and plants. (1) hypocreolide A **(2)** benzophomopsin A **(3)** JBIR-37 **(4)** artemisinin **(5)** paclitaxel.

Actinomycetes are prolific producers of natural products. More than 5000 antibiotics have been identified from this order of organisms, with over 90% of those coming from the genus *Streptomyces* [28]. The discoveries of streptothricin and streptomycin in the 1940s led to large scale screening of candidates, but it has been estimated that only a small percentage of *Streptomyces* antibiotics have been discovered [29]. A number of other genera are also important contributors to natural product diversity. The genus *Bacillus* produces, for example, lanthionine-containing antibiotics, bacteriocin, and gramicidin [30,31]. The order Myxococcales are also recognised as prolific producers of natural products. They contain some of the largest genomes of all bacteria [32] which together encode over 500 natural products, mostly the products of polyketide synthase (PKS), non-ribosomal peptide synthase (NRPS), or hybrid PKS-NRPS pathways [33,34]. Pathogenic species of the genus *Burkholderia* are also a rich source of new natural products, producing for example enacyloxins, which are active against Gram-positive and Gram-negative pathogens [35]; rhizoxin, an antimitotic agent [36]; and the glidobactins, antitumor antibiotics [37].

After the 'Golden Age' of natural product drug discovery, it was assumed that the diversity of natural products had been fully evaluated and since the 1990s pharmaceutical companies have largely abandoned natural product screening in favour of combinatorial chemistry, structure-assisted drug design, and antibody-based therapies. However, new technologies and screening strategies can uncover many more promising drug candidates from natural products. For example, new screening methods have identified many new β-lactams, which as of 2003 held approximately 65% of the global antibiotics market. β-lactams remain at the forefront of the fight against antibiotic resistance with the development of the current 5th generation β-lactam derived cephalosporins, ceftaroline [38], and ceftobiprole [39] which are both highly active against MRSA and other Gram-positive bacteria.

Considering that an estimated 99% of all bacteria species are uncultured [40], bacteria still represent a largely untapped reservoir of novel chemical diversity.

1.1.3: The Life Cycle of *Streptomyces* and related actinomycetes

Streptomyces are filamentous actinomycetes, Gram-positive and of characteristically high G+C% content. They are abundant in soil and are believed to have had a last common ancestor around 400 million years ago [41], about the same time at which the land was being colonised by plants. Their major role in wider ecosystems was probably the solubilisation of cell walls, or surface components of plants, fungi and insects [42], composting early organic material and helping to form primeval soil. This ecological function can be seen still at the genome level. *Streptomyces reticuli* has an estimated 456 genes encoding proteins involved in degradation and utilisation of cellulose and other carbohydrates [43]. This is a common situation among *Streptomyces* and suggests that they may have evolved to live in mixed communities [44].

The developmental cycle of *Streptomyces* on a surface begins with formation of substrate mycelium, a vegetative tip growth phase that produces branching filaments that only septate occasionally so that each cell contains multiple chromosomes. In response to nutrient depletion, typical *Streptomyces* alter their growth pattern, producing specialised aerial hyphae, or aerial mycelium, which undergoes septation to generate a series of compartments each containing a single copy of the (linear) chromosome. Finally, the compartments differentiate to create a chain of heat-labile but desiccation-resistant spores for dispersion [45] (Figure 1.3). A complex signalling cascade controls the transition from vegetative growth to reproduction [46], and might provide checkpoints for the coordination of secondary metabolism [47].

The developmental cycle in liquid cultures is different to growth on a surface and has not been as extensively studied. This is mainly because most *Streptomyces* do not sporulate in liquid medium. Most industrial processes involving *Streptomyces* are performed in liquid cultures, so the majority of the work in this area has focused on mycelial morphology, medium composition, and bioreactor design [48–50]. Mycelial morphology, for example, is correlated with secondary metabolite production in at least some cases [51,52], but not in others [53]. In practice, each strain requires individual optimisation, which normally accompanies successive rounds of mutation and selection for improved production.

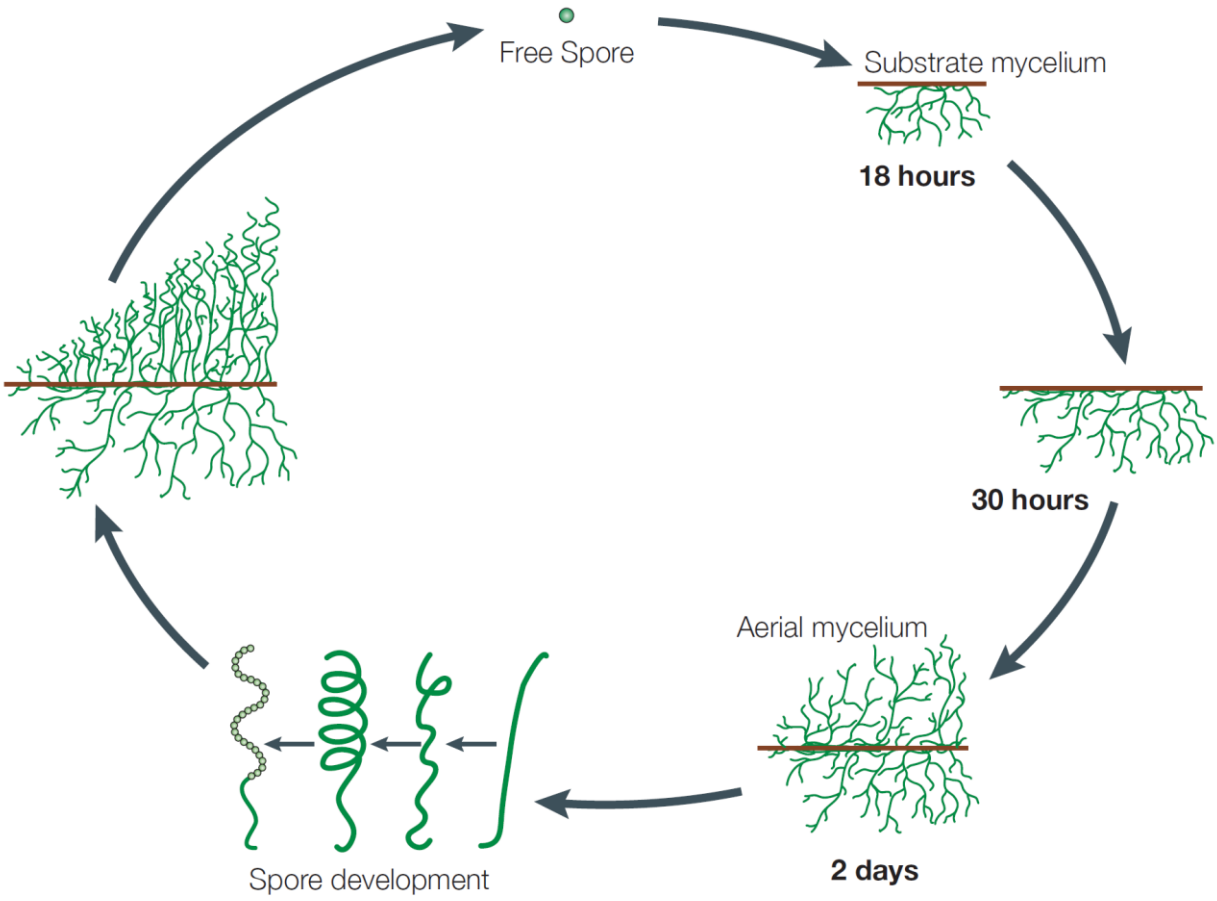


Figure 1.3: The life cycle of *Streptomyces* spp. Image adapted from Angert *et al*. [47].

1.1.4: Antibiotic modes of action and bacterial mechanisms of resistance

The chemical structure of an antibiotic determines its class, which in turn determine the bacterial physiology or biochemistry of the drug's target. The primary modes of action for antibiotics are illustrated in Figure 1.4. Common examples of antibiotic classes and their biological targets are; β-lactams which target cell walls, fluoroquinolones which target DNA and RNA synthesis, sulphonamides which target folate synthesis, lipopeptides which target the cell membrane, and aminoglycosides which target protein synthesis.

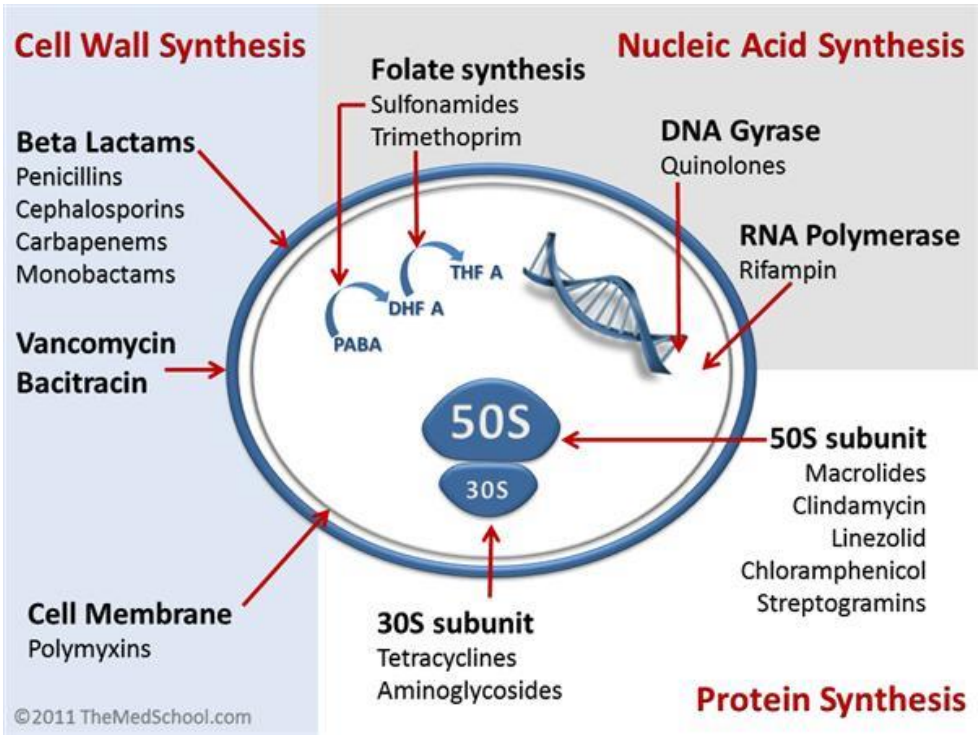


Figure 1.4: Mechanisms of action of antibiotics.

Self-resistance is an evidently important feature for antibiotic-producing organisms. Self-resistance is closely connected to acquired resistance, the ability of target organisms to modify themselves in some way to overcome the effects of an antibiotic. The development and/or recruitment of genes that confer resistance is of clear value for survival. Many genes responsible for resistance against several classes of antibiotics have been identified, most systematically by high-throughput screens of high-density mutant libraries created by targeted insertion or random transposon mutagenesis [54,55]. An example of intrinsic resistance (Figure 1.5) is given by daptomycin, a clinically-used lipopeptide which is effective against Gram-positive, but not Gram-negative bacteria. Gram-negative bacteria have fewer phospholipids in the cytoplasmic membrane significantly reducing

daptomycin uptake [56]. Other mechanisms by which bacteria protect themselves include: minimising intracellular concentration of the antibiotic by dedicated active efflux systems; modification of the intracellular target by genetic mutation or post-translational modification; or inactivation of the antibiotic by hydrolysis or covalent modification.

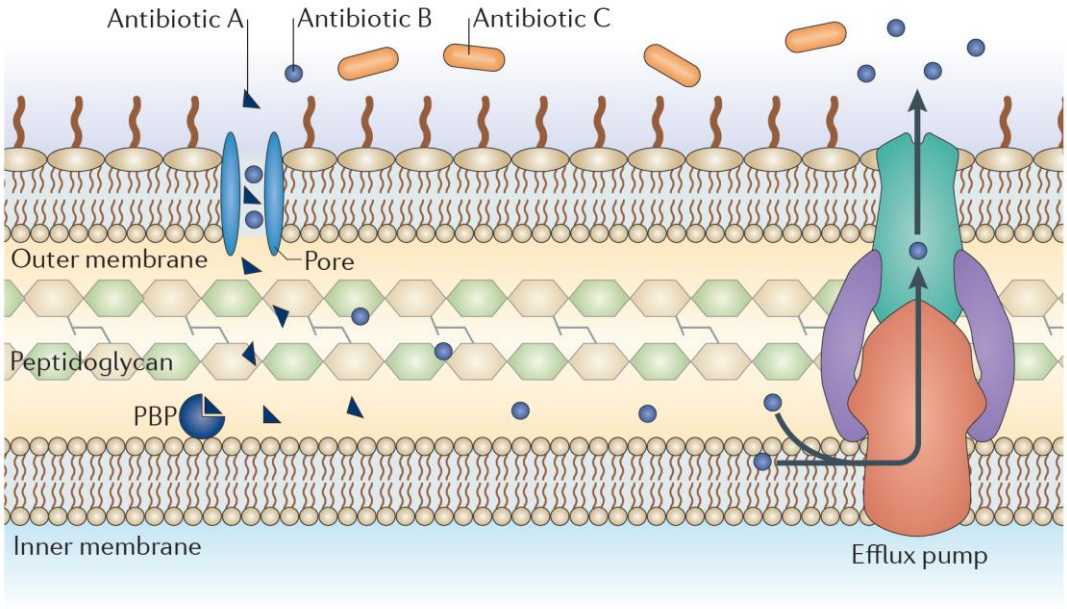


Figure 1.5: Intrinsic mechanisms of resistance. Antibiotic A can enter the cell via a cell membrane spanning pore and inhibit peptidoglycan synthesis. Antibiotic B can also enter the cell but is removed by efflux. Antibiotic C cannot cross the membrane. Adapted from Blair *et al*. [56].

Bacterial efflux pumps actively transport molecules out of cells and are major factors to the intrinsic resistance to antibiotics for many Gram-negative bacteria. Some pumps are highly substrate specific, such as the Tet family efflux pumps, which are specific for tetracycline antibiotics [57]. Others have a wider substrate specificity, such as AcrB in *E. coli*. AcrB forms a protein complex with two adapter proteins, and resides in the inner cell membrane. Structural studies have shown a large, flexible binding pocket which can accommodate molecules of different sizes and structures, allowing the transport of a wide range of antibiotics from the cell [58].

Antibiotic modifying enzymes are another important factor in resistance. Penicillins and cephalosporins are hydrolysed by a range of β-lactamase enzymes [56,59]. These enzymes were first detected in 1979 after analysis of bacterial isolates resistant to cefamandole, a cephalosporin [60]. Newer examples have been found to follow, with increasing rapidity, upon the introduction into

clinical use of improved β-lactams [56]. This type of resistance mechanism is particularly dangerous. The resistance can be provided by a single gene carried on a plasmid, meaning the risk of horizontal transfer is especially high, which can have serious implications for the treatment of severe infections.

Resistance genes have been present in the environment long before the modern development of antibiotics [61] but antibiotic resistance is now one of the biggest threats to global health, food security and development. A particularly striking example of the dangers of antibiotic resistance was the isolation of a strain of methicillin-resistant *Staphylococcus aureus* (MRSA) with resistance to vancomycin, normally used as a drug of last resort [62]. The rapid spread of resistance genes across bacterial populations has been driven by the widespread and poorly managed use of antibiotics in human health, and increasingly in agriculture and aquaculture, leading to increased concentrations of antibiotics in the environment and creating perfect conditions for the development of resistance. To combat microbial infections in the era of multi-drug resistance bacteria the traditional approaches to drug discovery need to change. Economic incentives designed to encourage pharmaceutical investment into new antibiotics, and a focus on accurate, personalised diagnosis and narrow spectrum antibiotics are steps in the right direction to treat serious infections in the resistance era [63].

1.2: Biosynthetic Genes and Enzymes of Actinomycetes

1.2.1: Biosynthetic Gene Cluster Organisation and Regulation.

One of the most important discoveries in the early days of actinobacterial gene sequencing was that the genes required for the biosynthesis of secondary metabolites are clustered together on the bacterial chromosome, although that clustering is highly variable across and within organisms [64]. Most gene clusters are organised into several operons, each under the control of a single upstream promoter, ensuring tight coupling of transcription and translation [65]. Within a biosynthetic gene cluster, the core biosynthetic genes encode enzymes that generate the scaffold of the metabolite. Its structure will depend on the class of metabolite, for example polyketide, non-ribosomal peptide, or terpene. In addition to the core biosynthetic genes, there are genes that encode tailoring enzymes that modify the scaffold and generate diversity and bioactivity, such as oxidoreductases, methyltransferase, acyltransferases, and glycosyltransferases. Many gene clusters will also contain cluster-specific regulatory (CSR) genes, as well as genes conferring self-resistance, and genes encoding mechanisms to transport the product out of the cell. As discussed above, the mechanisms for self-resistance are very important for an organism. Erythromycin A is a potent inhibitor of protein synthesis and causes this inhibition in susceptible bacteria by binding to their ribosome. The erythromycin

biosynthetic gene cluster contains a gene, *ermE*, which confers self-resistance to erythromycin by the dimethylation of a specific adenine base in its ribosomal RNA [66]. One of the first actinobacterial antibiotic biosynthetic gene clusters to be identified was that for the coloured isochromanequinone actinorhodin [67]. The gene cluster has been characterised in detail and is illustrated in Figure 1.6, showing that the core minimal PKS genes form only a small component of a complex system.

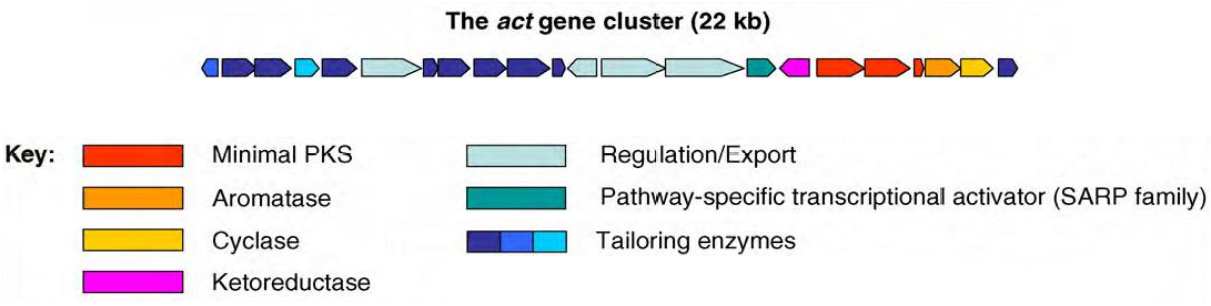


Figure 1.6: The actinorhodin gene cluster. The functions of genes are indicated by colour. Adapted from Osbourn [64].

In *Streptomyces*, and in other actinobacteria, the expression of secondary metabolic pathways is typically controlled by complex regulatory cascades that send signals to pathway-specific regulators and are often coordinated with morphological changes in the organism, or environmental stimuli [68]. While the pathway-specific regulators, such as the SARP [69] or LAL [70] families of regulators, are responsible for transcriptional control of individual gene clusters, there are also global, or pleiotropic regulators that control multiple gene clusters. The highly phosphorylated guanosine nucleotide, ppGpp, has received attention for its role in antibiotic production in *Streptomyces*. It has been shown that ppGpp is required for activating the transcription of genes required for antibiotic biosynthesis under nitrogen limited conditions, although the exact mechanism is elusive [71]. γ-butyrolactones are produced by many *Streptomyces* [72], and have also been implicated in the onset of secondary metabolism [68]. A-factor, a typical γ-butyrolactone from *S. griseus*, has been implicated in both the activation of secondary metabolism and morphological differentiation [73].

1.2.2: Evolution and Function of Biosynthetic Gene Clusters

The clustered organisation of biosynthetic genes suggests that at some point in their evolution they were selected for as a group, rather than as individual genes. If the production of the natural product were favoured for the producing organism then it would be beneficial for them to be close together,

increasing the probability of them being transferred to subsequent generations of the organism [74]. This does not explain the origins of the core biosynthetic genes however, which are often considered the founders of the gene cluster. Three types of evolutionary pathway are possible: linear, divergent and convergent [74]. A linear pathway is the easiest to imagine. The first catalytic enzyme of the primordial gene cluster could have arisen from random mutation of an enzyme involved in the primary metabolism of the starting material. If the product of that enzyme could be secreted from the cell, and provided some advantage to the organism, for instance by interacting with a receptor from another organism, then the producing organism will be selected for over others. Subsequent duplication, or mutation of that first enzyme, or the occurrence of a new enzyme may result in the transformation of the product, which may provide additional advantage over the original mutant organism. It can be imagined that repeated steps of this process could generate a multi-step biosynthetic pathway where each step provides further advantage over the previous steps. A divergent evolutionary pathway can be described by a slight modification of the linear one. If, at the point of divergence, two mutations of an enzyme are generated that can perform different reactions on the same substrate and providing they both provide greater advantage than the original organism, then both will be selected for. At that point one can imagine two parallel linear evolutionary pathways that diverge as they improve upon the original biosynthesises. Convergent pathways are more difficult to imagine occurring randomly, they require a mutation resulting in an enzyme that can bring together two different biosynthetic pathways, with the condensed product of the new pathway conferring greater advantage to the organism than the previous two products. An example of this kind of evolutionary pathway may be dynemicin, an antibiotic with both enediyne and anthracycline structural features [75].

1.2.3: Polyketides and Polyketide Synthases

Polyketides are a large and diverse class of biologically active natural product polymers, which can be broadly divided into aromatic compounds, in which the carbon skeleton that is built up is then stabilised by dehydration and aromatic ring formation; and complex reduced polyketides, in which the more chemically stable reduced chain adopts a wide range of structures including linear products, macrocycles, and polyethers, and may also be found covalently linked to amino acid units within hybrid polyketide-peptide compounds [17] Polyketides have been harvested from a number of organisms besides bacteria, including most importantly fungi. In the case of insects, marine sponges, and plants the actual producing organism is very often a commensal microorganism [17]. Between them

polyketides have provided clinically useful compounds with antibacterial, antifungal, antiviral, and antitumour activity, as well as immunosuppressant and cholesterol-lowering compounds. This family of natural products also includes toxins such as the potent fungal carcinogen aflatoxin from *Aspergillus* spp. [76]; macrolide mycolactone produced by *Mycobacterium ulcerans* [77], the causative agent of the necrotizing skin disease Buruli ulcer; and the potent respiratory toxin bongkrekic acid produced by *Burkholderia gladioli*, a notorious contaminant of the traditional Indonesian dish, tempe bongkrek [78].

Polyketides are an exciting source of drug candidates as they cover a region of chemical space that appears to interact with biological targets very effectively [79,80]. Over 7000 known polyketide structures have led to more than 20 commercial drugs, a hit rate of 0.3%, far greater than the hit rate of <0.001% typically found for synthetic compound libraries [81]. These drugs have also proved very successful commercially, with annual peak sales of the top six totalling $15 billion. Over the last thirty years, fundamental research on the genes and enzymes of polyketide biosynthesis has transformed our understanding of the pathways and mechanisms involved, and this in turn has presented new opportunities to exploit natural products as leads in drug discovery.

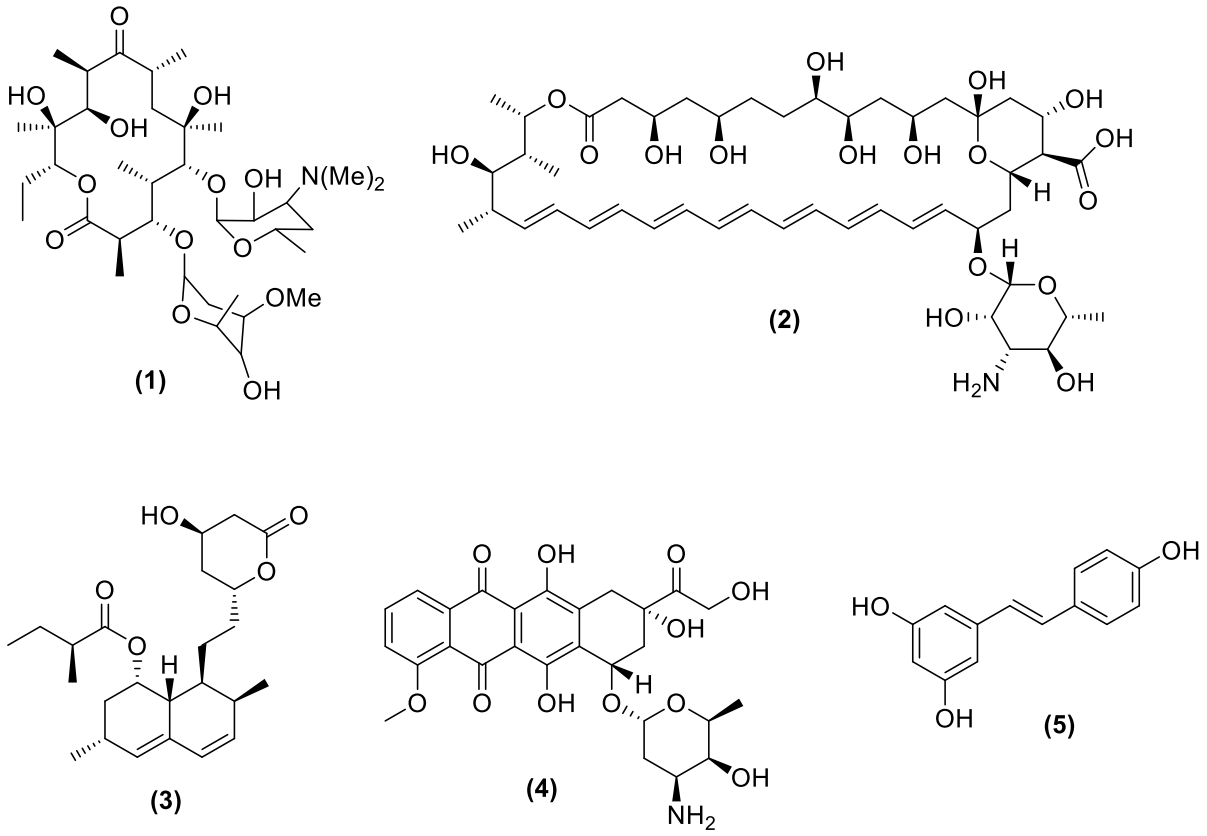


Figure 1.7: Examples of polyketide natural products. (1) modular type I PKS product erythromycin A; **(2)** modular type I PKS product amphotericin B; **(3)** iterative type I PKS product lovastatin; **(4)** type II PKS product doxorubicin; **(5)** type III PKS product resveratrol.

Polyketide synthases (PKSs) are the key enzymes catalysing the biosynthesis of polyketide polymers. They use a mechanism for chain assembly that closely resembles that of fatty acid synthases (FASs), in which simple carboxylic acid extender units provided by primary metabolism are linked together by iterative Claisen-like condensation reactions [17] (see Figure 1.8). Assembly on a typical fatty acid synthase starts by transfer of an acetyl unit from acetyl-CoA onto the active site of a ketosynthase (KS) enzyme as a thioester. The transfer is catalysed by an acyltransferase (AT) which in FAS systems also transfers a malonyl unit from malonyl-CoA to the 4'-phosphopantetheinyl prosthetic group of an acyl carrier protein (ACP). The KS catalyses Claisen-like condensation to give the 3-ketoacyl-ACP product, after which reductive processing takes place catalysed by, successively, a β-ketacyl-ACP reductase (KR); a dehydratase (DH); and an enoylreductase to give a fully reduced butyryl thioester, which the KS transfers back to its active site, freeing the ACP for attachment of a further malonate extender unit. This iterative process continues until the chain length is such that chain termination catalysed by a thioesterase (TE) or transferase competes effectively with back-transfer of the chain to the KS. In *E. coli* and most other bacteria, the individual enzymatic activities of the FAS are housed on individual proteins (a Type II system), while in yeast and animal FAS the enzymes are domains within a covalently-linked multifunctional enzyme (a Type I system) [17].

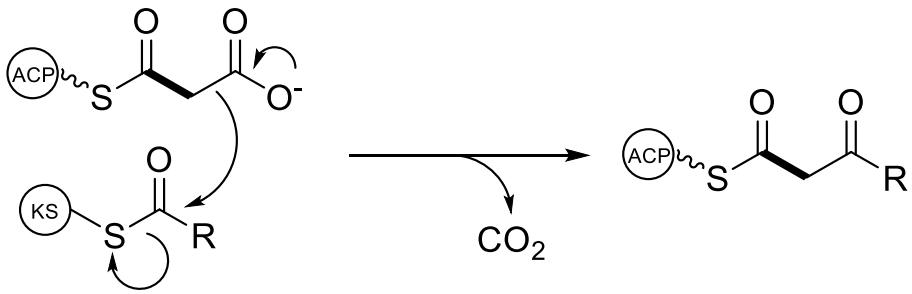


Figure 1.8: Claisen-like condensation of an ACP-bound malonyl extender unit with the growing polyketide backbone. An activated malonyl extender unit, bound to an ACP domain by a phosphopantetheine prosthetic group reacts with the polyketide intermediate held by the KS domain, expelling carbon dioxide and forming a new polyketide intermediate which is passed to the upstream KS domain for the next round of elongation.

The major difference between FAS and PKS is that in the latter type of synthase reductive processing is not complete. In Type II dissociated PKSs of bacteria, and in the iterative Type I PKS multienzymes of fungi, reduction may take place rarely or not at all, giving a full-length, highly-reactive poly-β-oxo-thioester intermediate which is stabilised by formation of aromatic rings. In bacterial modular PKSs reduction is variable and usually incomplete, as discussed below. Table 1.1 summarises these different types of synthase and their characteristic products.

Table 1.1: Summary of polyketide synthase types (adapted from Weissman [82]**).**

Type of PKS	**Mode of operation**	**Mode of substrate activation**	**Typical product**	**Typical organisms**
I	Modular	ACP	Reduced	Bacteria
I	Iterative	ACP	Aromatic and reduced	Bacteria and Fungi
II	Iterative	ACP	Aromatic	Bacteria
III	Iterative	CoA	Aromatic	Bacteria, Fungi and Plants

Type III PKSs are relatively simple homodimeric enzymes, with a single KS-like multifunctional active site which interacts with CoA-bound substrates (no ACP is involved) to catalyse the iterative condensations of acetate units to the CoA-bound starter units. They are found in bacteria, fungi and plants [83], typically yielding mono- and bi-cyclic aromatic products [84] formed through intramolecular condensations and aromatisation of the linear intermediate within the same PKS active site cavity. The type III PKSs were originally discovered in plants [83], where they include chalcone synthases and stilbene synthases. They generate a vast array of small polyphenolic molecules important to plant survival and fitness. The phenolic compounds were likely early adaptations in plants, involved in the transition of plants from aquatic to terrestrial environments [85]. These compounds are involved in UV protection, flower colour, pollen development, root nodulation, plant architecture and chemical defence [86]. These systems provide diversity through the choice of CoA-bound starter molecules, the number of elongation steps and the mechanisms of cyclisation. Downstream tailoring enzymes can then further increase diversity, transforming the initial scaffolds into a huge range of different compounds [87]. The relatively simple architecture of the type III enzymes has allowed for a mechanistic framework to be developed, offering insights into the factors involved in starter unit selection, chain

extension and control of ring formation [82]. These insights can be helpful when considering similar factors in the more complex type II and type I systems.

Type II PKSs consist of several different enzymes, each expressed from a distinct gene. Marine- and soil-dwelling actinomycetes are the only known organism to employ the type II system. They are a rich source of bioactive polyphenols such as tetracyclines and produce doxorubicin, an anthracycline, one of the most effective classes of anticancer drugs ever developed [88]. Chain initiation, most commonly starting with acetate, and elongation, using malonyl-CoA extension units, are achieved by a "minimal PKS" which consists of a complex of two ketosynthase-like condensing enzymes (KS_α and KS_β), and an ACP to which the polyketide intermediate is tethered. The folding of the growing polyketide intermediate can be modified by additional enzymes such as a KR, cyclases and aromatases, to generate an aromatic product. Once released these polyphenols are often extensively modified by hydroxylases, glycosylases and other tailoring enzymes [89]. A malonyl-CoA:ACP transferase (MCAT) activity to recruit new malonyl extension units is thought a necessary component of the minimal PKS, however genes encoding proteins with this activity are absent in most type II PKS gene clusters and it was proposed that an endogenous MCAT is recruited from fatty acid biosynthesis [90]. An alternative model for the recruitment of malonyl extender units was proposed by Simpson *et al.* that ACP self-malonylates. Although they demonstrated this activity *in vitro*, it unfortunately required elevated concentrations of malonyl-CoA unlikely to be found *in vivo* [91]. The multienzyme complexes of type II PKS have been notoriously hard to study and only recently, through structural studies, has there been direct evidence for the KS_α, KS_β heterodimer [92].

Type I PKSs have proven to be the most complicated and most versatile class of PKS. Iterative Type I PKSs are commonly found in fungi where they use a single set or module of enzymes acting iteratively to generate the diverse polyketide structures found in fungi natural products. Fungal polyketides can broadly be grouped into three classes, based on the extent to which the enzymes make use of reducing domains. These are illustrated in Figure 1.9. The classes of non-reduced compounds such as orsellinic acid, and partially reduced compounds such as 6-MSA were characterised first from the limited set of known fungal PKS genes. Degenerate primers, designed to amplify the conserved β-ketoacyl synthase domains identified many new fungal PKS genes, which also identified a new class of highly reduced compounds, such as lovastatin [93].

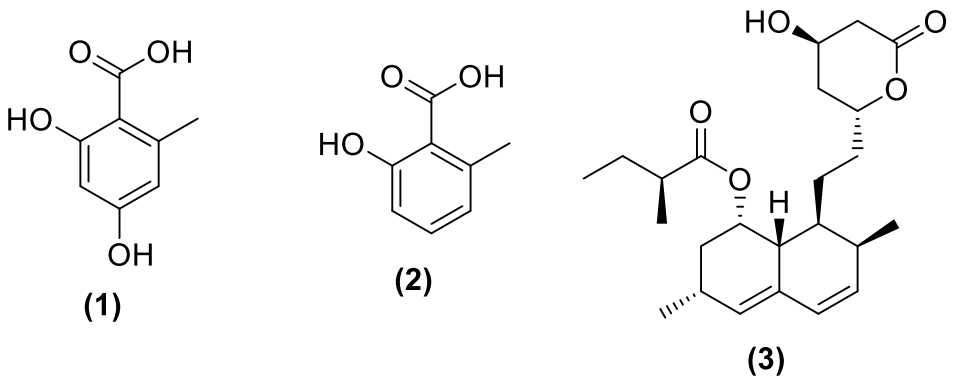


Figure 1.9: Examples of iterative Type I PKS products. (1) orsellinic acid; **(2)** 6-MSA; **(3)** lovastatin.

1.2.4: Modular Type I Polyketide Synthases and the Assembly-line Paradigm

The modular PKS for the 14-membered macrolide erythromycin A

Macrolides are glycosylated polyketides with potent antibacterial activity (via inhibition of protein biosynthesis) and a ring size between 12 and 18, including the 16-membered spiramycin produced by *Streptomyces ambofaciens* [94] and the 12-membered methymycin produced by *Streptomyces venezuelae* [95]. Erythromycin A (Figure 1.10) [96] is a prominent member of this class, for together with its semi-synthetic derivatives it has been in worldwide clinical use since the 1960s. Erythromycin biosynthesis starts with the synthesis of the macrocyclic aglycone precursor 6-deoxyerythronolide B (6-dEB) assembled from the successive condensation of the propionyl starter unit with six methylmalonyl extender units catalysed by six extension modules housed within three multi-modular PKS enzymes [97]. Tailoring enzymes then catalyse post-PKS modifications to generate erythromycin A.

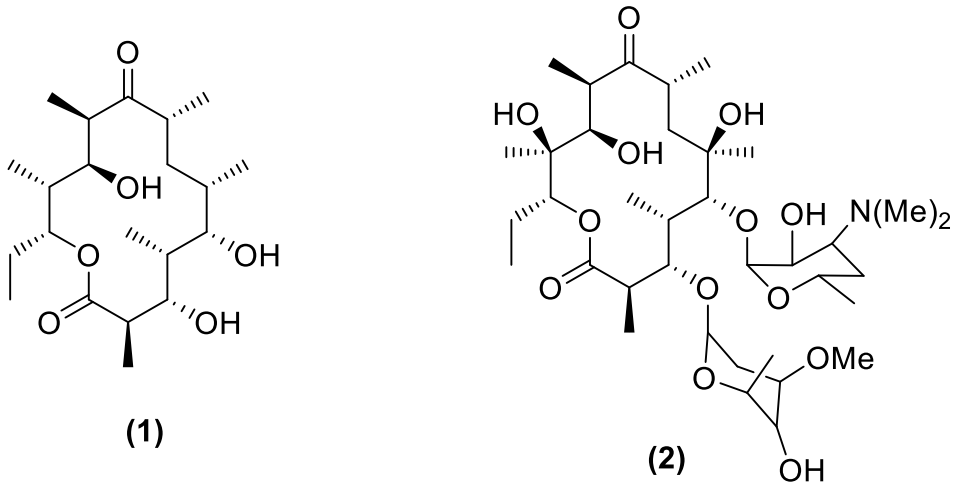


Figure 1.10: Erythromycin biosynthesis. (1) the biosynthetic intermediate aglycone 6-deoxyerythronolide (6-dEB) B; (2) erythromycin A.

In the early 1990s the genes encoding deoxyerythronolide B synthase (DEBS), the PKS responsible for the precursor to erythromycin, from the producing organism *Saccharopolyspora erythraea*, were sequenced and characterised by two independent groups [98–101]. These were the first PKS genes to be sequenced in their entirety and they revealed for the first time the modular architecture that became the paradigm for type I PKSs and rationalised the biosynthetic relationship between many known reduced polyketides and the PKS multienzymes that produced them. The modular architecture of these biological machines has resulted in some of the largest proteins in nature. A single subunit of MLSA1, the enzyme responsible for generating nine of the eighteen acyl links in the mycolactone polyketide chain, has a molar mass of 1.8MDa [102].

A PKS module can be defined as a set of activities (domains) responsible for incorporating a particular precursor at a designated point in the nascent acyl chain and for determining its structure in the final polyketide produced [103]. Figure 1.11 shows the modular organisation of DEBS along with the growing polyketide chain of 6-dEB.

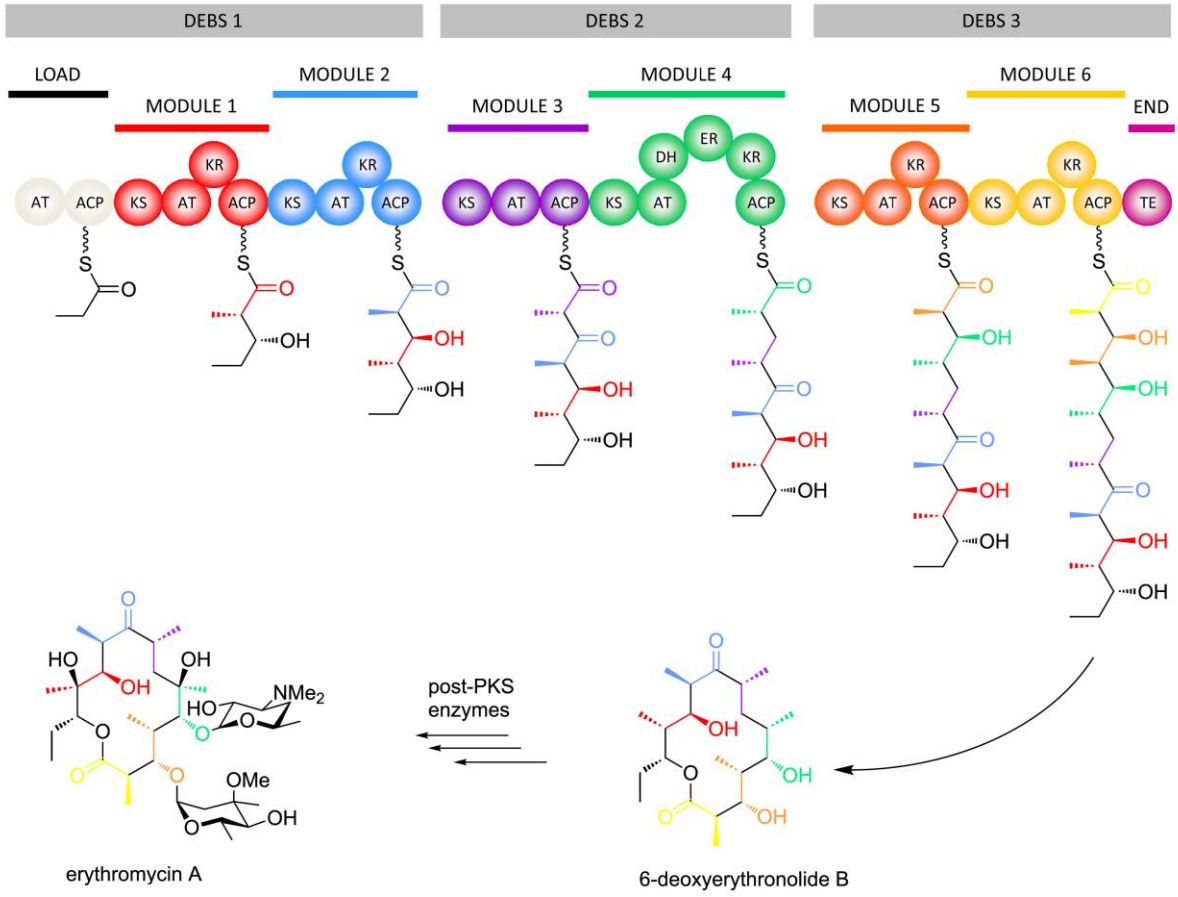


Figure 1.11: Schematic of the erythromycin PKS assembly line. Image adapted from Weissman [104]. The DEBS genes are not to scale, but indicate which modules are encoded by which genes. DEBS 1 = 10.5 kb, DEBS 2 = 10.5 kb, DEBS 3 = 9.5 kb.

DEBS is made up of three components, DEBS 1-3, which are encoded respectively by three large genes, *ery*AI-III [100]. Each polypeptide contains two modules of FAS-related activities, with DEBS 1 also containing a loading module comprising a propionyl-CoA specific AT and an ACP, to recruit the propionyl starter unit. The remaining six modules contain (2*S*)-methylmalonyl specific AT domains [105] and DEBS 3 contains a C-terminal TE domain which is responsible for cleaving and cyclising the completed heptaketide to generate 6-dEB [103]. Strikingly, each module contains the complement of reductive domains needed to achieve the appropriate oxidation state after each cycle of chain extension. DEBS proved to be the prototype of a very large number of such multi-modular systems, composed of several large enzymes with multiple modules in each enzyme, each module is responsible for a given catalytic cycle in the polyketide biosynthesis[82]. All intermediates remain bound to the assembly line throughout chain assembly.

The general reaction scheme is similar for all modular type I PKSs. First, the KS is acylated with a polyketide intermediate from the upstream module, and the extender unit is selected by the AT domain which then charges the ACP with it. The modular arrangement allows different extender units to be utilised at different cycles of chain growth, and natural ATs have been identified that are specific respectively for malonyl-CoA, methylmalonyl-CoA, ethylmalonyl-CoA, hydroxymalonyl-CoA, and methoxymalonyl-CoA [106,107]. This larger pool of extender units helps explain the great molecular diversity of complex polyketides obtained from modular systems. Secondly, the (alkyl)malonyl-ACP docks to the acylated KS which then catalyses the condensation reaction, extending the polyketide chain; the resulting β-ketoacyl-ACP intermediate then migrates to the active sites in turn of the processing enzymes of the so-called 'reductive loop' (KR, DH and ER) [108]; finally the processed polyketide intermediate is ferried either to the adjacent downstream module, or to a TE domain which catalyses chain-release [109]. The stereospecific action of the reductive loop domains (KR, DH and ER) provides exquisite control over the shape of the final product [82].

Extension of the modular paradigm to other major classes of complex polyketide

Polyether ionophores are an abundant class of polyketides with broad-spectrum activity against bacteria and parasites [110]. More recently, certain polyethers such as salinomycin (Figure 1.9) have been shown to be selective killers of cancer stem cells [111] and to be active in an animal model of latent TB [112]. Unfortunately, their clinical use is limited by their significant toxicity [113], however targeted modification of these compounds may provide promising drug leads. They are produced exclusively by actinobacteria, with the vast majority produced by the genera Streptomyces and Actinomadura [114]. Polyethers are characterised by chains of tetrahydrofuran and tetrahydropyran rings connected via C-C bonds or spiro linkages (Figure 1.9). The high oxygen content of polyether ionophores allow them to chelate monovalent metal ions such as Na^+ and K^+, and in a minority of cases divalent ions such as Mg^{2+} and Ca^{2+} [115]. The resulting complex has a hydrophobic exterior surface, allowing transport of the chelated ions through lipid membranes, resulting in a collapse of trans-membrane cell potentials and cell death [116].

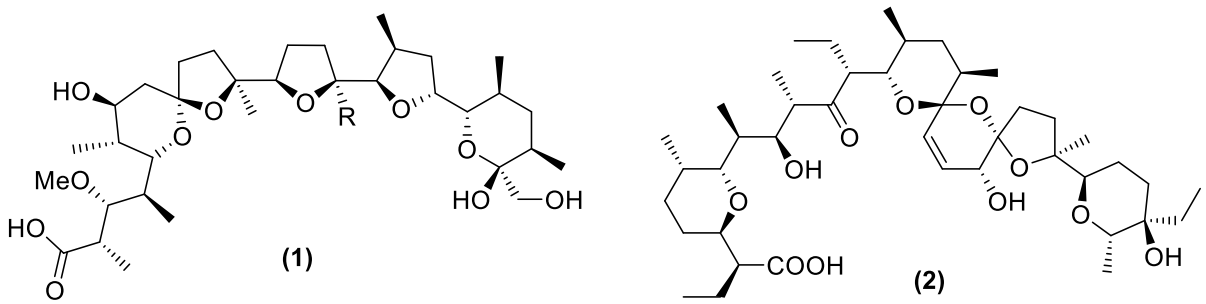
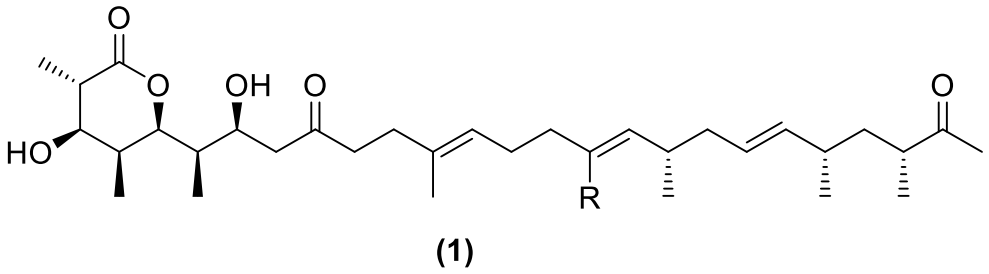


Figure 1.12: Examples of polyether antibiotics. (1) monensin A, R=C_2H_5, monensin B, R=CH_3 **(2)** salinomycin.

Monensin (Figure 1.12) is an iconic polyether ionophore [117] which has found a widespread application in animal husbandry as a coccidiostat and growth promoter [118]. Monensin A differs from B by the presence of an ethyl substituent rather than methyl at position C16. Both compounds selectively bind Na^+ ions [118]. The monensin gene cluster from *S. cinnamonensis* was identified and sequenced by Leadlay *et al.* in 2001 [118]. It consists of 8 open reading frames (*monAI* to *monAVIII*) encoding the 12 extension modules of the monensin PKS, flanked by tailoring genes that encode enzymes required for post-PKS processing of the monensin carbon backbone, as well as efflux and regulatory proteins [82].

The difference between monensins A and B arises in module 5, where the acyltransferase (AT) domain allows the incorporation of either methyl or ethylmalonyl-CoA. *MonCI* is a unique epoxidase which epoxidises three double bonds in the linear precursor, and *monBI* and *monBII* are novel ring-opening epoxide hydrolases/cyclases responsible for forming the ether rings in a cascade reaction [119,120]. The monensin gene cluster also has additional ketosynthase (KS)-like and ACP-like ORFs. It is proposed that the KS-like enzyme transfers the full-length triene intermediate onto the ACP-like protein where the oxidative cyclisations occur, followed by further tailoring and final release of monensin from the discrete ACP catalysed by thioesterase MonCII [121]. Deletion of the *monCI* epoxidase gene results in the accumulation of the shunt product premonensin (Figure 1.13) in which the double bonds are intact.



(1)

Figure 1.13: Premonensin, shunt product of monensin biosynthesis. (1) premonensin A, R=C_2H_5, premonensin B, R=CH_3.

Macrocyclic polyene polyketides

Amphotericin B and nystatin are valuable macrocyclic polyenes, amphotericin being used in front-line treatment of life-threatening systemic fungal infections. Such polyenes are composed of a large macrolactone ring with distinct polyol and polyene regions and are often glycosylated with a mycosamine sugar residue (see Figure 1.8). Their advantages over other antifungal compounds such as azoles is their potent fungicidal action, and the extremely low incidence of resistance [122]. Azoles disrupt ergosterol biosynthesis, a crucial component of the fungal cell membrane, but are fungistatic and resistance can arise more readily [123]. Macrocyclic polyenes instead interact directly with ergosterol, forming pore-like structures in the cell membrane that lead to cell death [124]. Unfortunately macrocyclic polyenes also interact with cholesterol in mammalian cell membranes, albeit with much lower affinity, and because of this relative toxicity amphotericin is always administered as an expensive liposomal formulation (Ambisome) [125].

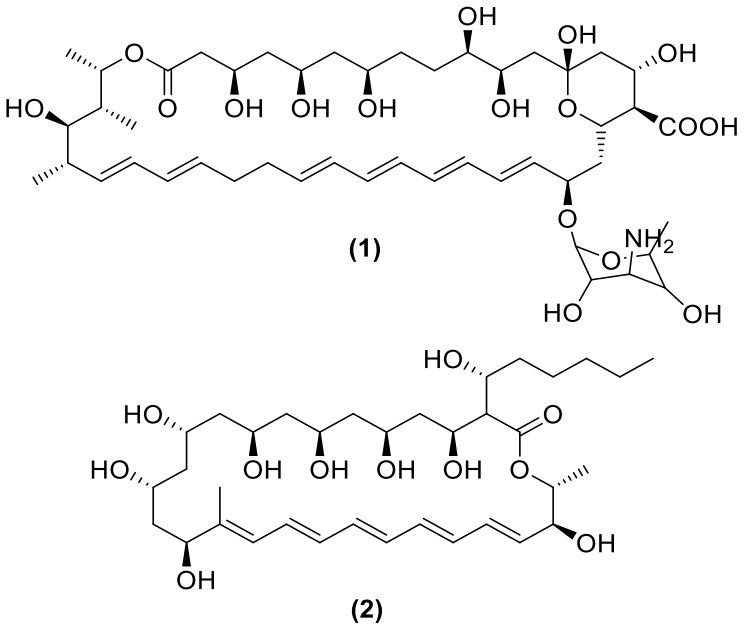


Figure 1.14: Examples of macrocyclic polyenes. (1) nystatin A1 **(2)** filipin III. Filipin biosynthesis incorporates an unusual hexylmalonyl extender unit in the final chain extension.

Nystatin, produced by *Streptomyces noursei*, was the first useful macrocyclic polyene to be identified. It is widely used for the topical treatment of oral, gastro-intestinal and genital candidosis. Nystatin biosynthesis involves the successive condensation of 19 malonyl and methylmalonyl extender units catalysed by separate modules housed within the 6 multimodular enzymes of the nystatin type 1 PKS. The polyene region of amphotericin contains a conjugated heptaene, but in nystatin this region of the molecule is interrupted into diene and tetraene regions because of the presence of an enoylreductase (ER) domain in module 5 (Fig. 1.14). Once the polyketide chain is complete, the biosynthesis is completed by oxidation of a methyl group to an exocyclic carboxyl group, hydroxylation at C-10 in the polyol region and glycosylation with a mycosamine sugar residue [124].

The macrocyclic antifungal polyene filipin was isolated from *Streptomyces filipinensis* in 1955 [126]. Filipin is synthesised by a type I PKS with five multi-modular enzymes consisting of 14 modules condensing both malonyl and methylmalonyl extender units. The final PKS extension module introduces a hexylmalonyl unit to generate the unusual extended alkyl chain. Post-assembly steps involve specific hydroxylations catalysed by cytochrome P450 monooxygenases [127]. Filipin lacks the mycosamine moiety found in many other macrocyclic polyenes, but maintains its affinity to membrane sterols and can form pore-like structures [128]. Filipin is too toxic for clinical use as it has similar affinity

to both ergosterol and cholesterol. This property does however make it a valuable probe for the detection of cholesterol in cellular membranes [129].

Rapamycin, a macrocyclic hybrid polyketide-nonribosomal peptide

Rapamycin (Figure 1.15) is a classic example of a hybrid polyketide-nonribosomal peptide and historically has played an important role in PKS engineering. It is a potent antifungal and immunosuppressive and possesses a structural motif which is responsible for the specific binding to FK506-binding proteins. This binding triggers subsequent binding of the binary complex to the mammalian target of rapamycin (mTOR) which mediates its biological activities [130]. In 2001, Rowe and colleagues spliced an entire PKS extension module from rapamycin between the first two extension modules of DEBS 1 in one of the first attempts at combinatorial biosynthesis [131]. When the extended DEBS 1 was co-expressed in *Sa. erythraea* with DEBS 2 and DEBS 3 production of a novel octaketide was observed. Due to rapamycin's (and its analogues') properties as the only known inhibitor of mTOR, and as the most selective inhibitor of kinase activity known to date [132], it has been the target of extensive work to build a library of related compounds with desirable therapeutic properties. Some of the more promising work has focussed on the exotic start unit of rapamycin or rather, its removal. Several rapamycin analogues (rapalogs) that underwent clinical trials contained semisynthetic alterations at positions derived from the starter unit [132] and in 2005 Gregory and colleagues showed that novel starter units can be efficiently incorporated into rapamycin by feeding them to a *rapK*-deleted mutant (*rapK* is involved in the synthesis of the 4,5-dihydroxycyclohex-1-enecarboxylic acid starter unit) [132]. The rapamycin scaffold has undergone further modification to generate rapalogs lacking the O- and C-linked methyl groups at position 16 and 17, respectively (Figure 1.15) [133]. These modifications were based on knowledge gained from X-ray structures of rapamycin bound to mTOR and aimed at reducing steric constraints and increase positive interactions. The group's efforts were rewarded by the production of rapalogs that displayed enhanced inhibition of cancer cell lines [133]. Rapamycin and its analogues exemplify the potential modular biosynthetic enzymes have for bioengineering and guided knowledge-based modifications.

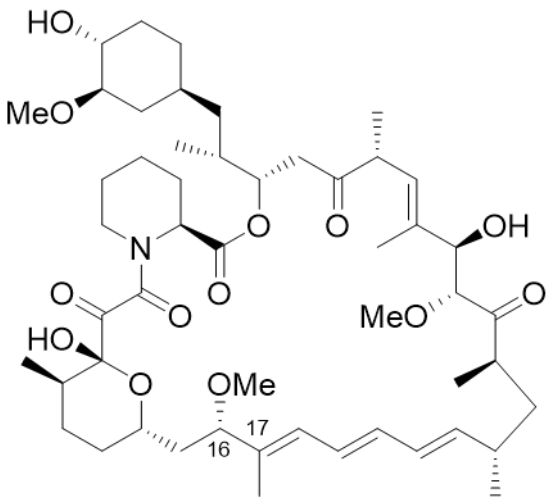


Figure 1.15: Structure of the macrocyclic hybrid polyketide-nonribosomal peptide rapamycin.

The active site specificity of PKS domains

As the number of characterised PKS systems has increased, detailed comparisons have revealed active site sequence motifs that correlate with observed chemical and stereochemical outcomes within extension modules. This was confirmed with the sequencing of the rapamycin PKS gene cluster in 1995 [134]. In many cases these have been validated by biochemical experiments and confirmed by atomic resolution X-ray crystal structures of individual domains [104,135]. In the erythromycin PKS (DEBS) the choice of methylmalonyl-CoA could be identified by the motif of YASH 100 residues downstream of the active site serine, while malonyl-CoA specific AT domains would have a HAFH motif [135]. The presence of structural data, also explained the stereospecificity for the (2S)-isomer of methylmalonyl-CoA in AT domains with the YASH motif, as there would be a steric clash between the (2R)-methyl group and residues in the active site.

Similar analyses of KR, DH and ER domains have provided explanations for substrate specificity. All reductase active KR domain active sites contain a tetrad of Tyr, Ser, Lys, and Asn residues [136] and bind the NADPH cofactor in the same orientation. Therefore the alternative stereochemical outcomes of ketoreduction are thought to arise from opposite modes of binding the substrate into the active site [104]. Several motifs were identified by comparative sequence analysis that predict the directions, referred to as A- and B-type KRs [137]. An LDD motif in the region between amino acid positions 88 and 103 is a strong indication of a B-type and is absent in the A-type KRs. The role of these residues in influencing the direction of substrate binding is still unknown, possibly because as yet no

KR has been co-crystallised in a complex with both the polyketide intermediate and the NADPH cofactor [104]. However there have been two alternative mechanisms to account for substrate positions proposed by Keatinge-Clay and Bonnett *et al.* [138,139].

DH domains typically form *trans* double bonds, although *cis* double bonds are also present [140]. By studying the stereochemistry of the evolutionary related DH domains from fatty acid synthesis and comparing that with PKS DH domains it appears that control of the stereochemistry of the dehydration step is ultimately determined by the KR domain and should always yield a *trans* double bond because the catalytic machinery is not in the correct position to yield a *cis* double bond [141]. The origin of *cis* double bonds is less clear. There is little evidence to show they are generated directly, or if they form a *trans* bond before undergoing isomerisation through post-PKS tailoring reactions [142]. The only clear evidence for *cis* bond formation by a specific module comes from the phoslactomycin PKS where it was shown that the modules of the PKS could discriminate between *cis* and *trans* intermediates and that they do not contain domains capable of catalysing double bond isomerisation [143].

The final domain with reducing activity is the ER domain which acts on *trans* double bonds, reducing them to generate fully-saturated alkyl groups. This reduction has stereochemical consequences when a side chain, such as a C-2 methyl substituent, is present. Reduction can produce either (2*R*)- or (2*S*)-configuration depending on the side from which the double bond is protonated [104]. Through similar comparisons of sequences that determined key residues in the previously discussed domains, a correlation was identified between ER sequence motifs and the stereochemical outcome of their reaction [144]. A conserved Y residue predicts a (2*S*)-configuration, while a conserved V residue predicts a (2*R*)-configuration. However, subsequent mutagenesis experiments revealed that additional factors affect stereocontrol of the ER domain. When the conserved Y was replaced with a V the stereospecificity swapped from a (2*S*)- to a (2*R*)-configuration, as was predicted, however in the reciprocal experiment where the conserved V was replaced with a Y there was no observed change in stereospecificity. A combination of the solved structure of a KR-ER didomain from the second module of the spinosyn PKS [145] and further mutagenesis experiments [146] helped to explain these observations. In the active site pocket there is a conserved K residue positioned opposite the conserved Y residue. When the Y is present it acts as the proton donor in the double bond protonation reaction, but in its absence the K residue acts as a proton donor. However, introducing a Y at the conserved position into a (2R)-producing ER domain was not enough to overcome the K proton donor, and more work will be required to identify the remaining stereochemical determinants. This result highlights the prevailing situation across PKS structural studies, that despite the great progress that has been made since their discovery, there is still a lot that remains unknown regarding the architecture of PKSs. However, these

bioinformatic, structural and biochemical insights provided the essential knowledge for rational engineering of PKSs to produce novel chemical diversity and predict the possibility of even finer control of these molecular assembly lines in the future.

The erythromycin PKS has been a model system for many investigations into PKS engineering, leading to the discovery of several interesting properties of PKS systems, namely stuttering and skipping. Stuttering of a PKS module refers to its ability to catalyse additional extension cycles before the growing polyketide chain is transferred to the next module [147]. Skipping of a module has the opposite effect, it allows a module to be bypassed, passing the polyketide chain onwards without catalysing an extension cycle.[131]

Both the stuttering and the skipping properties of the erythromycin producing system were discovered as a result of modification of the PKS, where they highlight an important limitation in current polyketide engineering techniques. The desired products of a modified system are often found as the minor product whereas the major product will be the product of the parent system, despite the modifications. Meanwhile, a number of natural modular PKS systems have been described in which non-colinearity is an essential feature [148]. For example, formation of the antibiotic borrelidin requires that module 5 of the *bor* PKS acts three times (identically) before the chain is handed on [149]. In the case of the marginolide azalomycin, the iteration of module 1 actually achieves different outcomes, as an essential part of producing the natural product [150]. Such programmed iteration offers the possibility of generating "aberrant" products from a single production line and offers a possible clue to a process that might have been important in the natural evolution of these assembly-line systems.

In addition to the functional domains of each DEBS module, they contain additional domains comprised of short α-helical regions at the N- and C- termini. These have been associated with docking to neighbouring modules thanks to solved crystal and NMR structures of these domains [151,152]. The heptaketide product of DEBS cannot be produced by DEBS alone, it requires additional enzymatic activities to generate 6-dEB. When the DEBS was purified and acylated *in vitro*, it was shown that all six AT domains retained their specificity for (2*S*)-methylmalonyl, however the configuration of the methyl branches in the product was different from that produced *in vivo*. Stereoselection during the biosynthesis must also involve epimerisation activities that DEBS is unable to perform.

Following the sequencing of DEBS, Leonard Katz and colleagues performed pioneering work on the relationship between the arrangement of domains in PKS systems and the structure of the resulting polyketide using erythromycin as a model system. First the group disabled the KR domain of module 5 by removing a part of the gene; this resulted in a mutant strain which produced two erythromycin analogues in which a hydroxyl group was replaced by a keto group [98]. The group then

disabled the ER domain in module 4 through mutations in the putative NADPH binding site. The resulting mutant strain produced an erythromycin analogue with a double bond at the predicted position [153]. The structure of these analogues had been predicted by knowledge of the function and position of the inactivated domains. This was the first signs of what became known as co-linearity, the relationship between the arrangement of the modules at the DNA level and their order of use in the biosynthesis of the corresponding polyketide.

The sequences for other macrolide PKSs, such as tylosin and spiromycin, were among the first to appear after the DEBS sequence [154,155]. What was remarkable at the time was the similarity of their organisation to DEBS, strengthening the evidence for co-linearity, and making it a general rule for type I PKS systems [156]. The importance of co-linearity is that it allows for predictions of the polyketide structure, given the PKS DNA sequence is known. The prediction became more confident as more PKS sequence data became available, as discussed above, making the prospect of engineering novel polyketides so tempting, since modifications of the organisation of domains within the PKS lead to predictable outcomes in the polyketide metabolites. The architecture of modular Type I PKSs has encouraged efforts at combinatorial biosynthesis, in which different natural PKS modules are "mixed and matched" to create hybrid PKS systems [157]. Combinatorial biosynthesis has resulted in several hundred new polyketide structures over the last 20 years and plays an important part in current polyketide drug discovery [82].

As the number of PKSs discovered to follow the DEBS paradigm increased, an increasing number of PKSs were identified that didn't [103]. Two major classes of modular PKSs are now distinguished: *cis*-AT PKSs, including DEBS and others, and *trans*-AT. The defining feature of a *trans*-AT is the absence of an AT domain from all extension modules. Instead the acyltransferase activity is provided on a single discrete protein that provides each module with the same extender unit [158], as illustrated by the virginiamycin PKS (Figure 1.16). *Trans*-AT PKSs have other characteristic features including; irregular domain arrangements, duplicated and inactive domains, atypical enzymatic functions, such as integrated methyltransferase domains, and modules distributed between two polypeptides [104].

Stereochemical control operates similarly in the two systems, as both contain the full suite of reducing domains, which act in the same way. There are differences however in the specificity for the AT domains. In *cis*-AT PKSs, the AT domains have a greater range of specificity, some ATs even specifying long-chain linear or branched extender units. In contrast, the majority of discrete AT domains from *trans*-AT PKSs only provide malonyl-CoA, although there are exceptions, such as KirCII, the *trans*-AT from kirromycin biosynthesis which is capable of loading ethylmalonyl-CoA onto a

specific ACP [159]. Methyl branches in the products of *trans*-AT PKSs are introduced by methyltransferase domains during chain elongation [160].

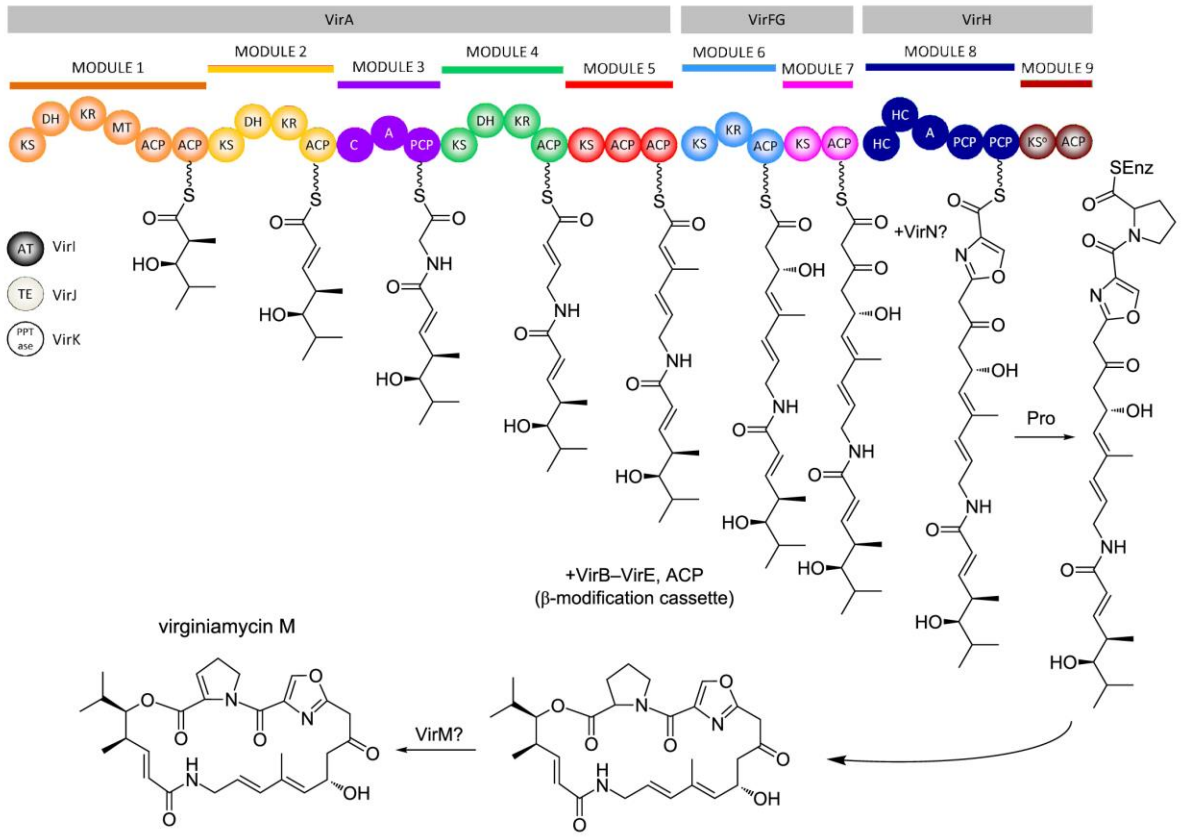


Figure 1.16: Schematic of the virginiamycin *trans*-AT PKS assembly line. Image adapted from Weissman [104]. The virginiamycin genes are not to scale, but indicate which modules are encoded by which genes.

The differences in architecture suggests independent, or divergent evolutionary pathways for the two classes of modular PKS systems. Nguyen *et al.* argue that while *cis*-AT PKSs evolved through gene duplication, the *trans*-AT class evolved through horizontal recruitment and assembly of substrate-specific KS domains [161]. However, recent studies have suggested that at least some *trans*-AT PKSs share a common ancestor with *cis*-AT PKSs [162]. Solved crystal structures have revealed that remnants of the linker regions found between KS-AT didomains in *cis*-AT PKSs can be recognised within *trans*-AT PKSs.

1.3: Genome-level Analysis; Sequencing, Assembly, and Annotation

1.3.1: Evolution of Sequencing and Assembly Technologies

The initial characterisation of biosynthetic gene clusters from *Streptomyces* and allied bacteria, leading to the uncovering of the assembly line paradigm for both PKS and nonribosomal peptide synthetase (NRPS) systems, was made using conventional Sanger sequencing, later automated using fluorescent chain terminators. Even several actinomycete genomes were sequenced in this way. The first was *Mycobacterium tuberculosis* in 1998 [163], followed several years later by the model organism *Streptomyces coelicolor* [164] and by *Streptomyces avermitilis* [165] The *S. coelicolor* sequence was generated by sequencing an ordered library of cosmid and BAC clones. In contrast the S. avermitilis sequence was the first to attempt a whole-genome shotgun approach. The final assembly had to rely on additional sequence data from over 10,000 cosmids, additional cosmid shotgun sequencing, hundreds of PCR products and many manually curated sequences [165]. The assembly of the *S. avermitilis* genome highlighted issues that any attempt at sequencing actinomycete genomes would encounter; high GC% content, and large sections of highly conserved repeats. In 2005 and 2006 the so-called Next Generation Sequencing (NGS) technologies were introduced: 454 and Solexa/Illumina sequencing. These took advantage of increased processing power to allow for massively parallel sequencing of shotgun fragments [166,167]. Several other complete actinomycete genomes were reported at this stage, including those of *Sa. erythraea* [168] and *S. griseus* [169]. Rapid improvements in the Illumina platform have since then made it the method of choice for genome-scale sequencing of bacterial and small fungal genomes. Single Molecule Real Time (SMRT) sequencing [170], usually referred to as PacBio, and Nanopore sequencing [171] allow sequencing of single DNA molecules, without amplification and in real time, to generate read lengths of up to several kilobases [172]. As yet, the high costs of PacBio and the high error rate of Nanopore sequencing means that they have yet to supplant Illumina as the method of choice outside major sequencing centres.

A search for actinobacteria in the NCBI database reveals 15,600 genome assemblies, more than double the number of entries since 2015, at 7057 [173]. The most important discovery to come from all these sequence data is that actinobacteria contain many more putative biosynthetic gene clusters than are expected from compounds isolated under laboratory conditions, in some cases the underrepresentation of metabolites is up to 10-fold, with some species dedicating up to 11% of their genome to secondary metabolism [174]. This realisation has led to much focus on these orphaned gene clusters and the biosynthetic potential that they represent.

1.3.2: Prediction of Biosynthetic Gene Clusters

To help identify putative orphaned gene clusters, automated BLAST searches [175] can be used to identify similarities with proteins in the public databases. The improvement of recognition of characteristic protein sequence motifs and prediction of substrate specificity for biosynthetic genes has been the main focus for this field, particularly for PKS and NRPS annotation [176]. The most commonly used software is antiSMASH, which aims to define biosynthetic loci covering the entire range of known secondary metabolite classes [177]. Its current iteration, antiSMASH 4.0 [178], is an extremely useful tool for the identification of biosynthetic gene clusters in actinomycetes, but manual curation is often required to obtain an accurate assessment of the biosynthetic potential of a gene cluster, as will be discussed in following chapters.

1.3.3: Genome and Metagenome Mining

The combination of advances in sequencing technologies and the development of robust biosynthetic gene cluster prediction software has led to a new paradigm for natural product discovery. The new approach uses sequence data and annotation to identify biosynthetic gene clusters of interest, and then study the product, or if it is an orphan gene cluster, attempt to 'wake up' the cluster and generate the product. There are three strategies used to obtain and characterise unknown products of a biosynthetic gene cluster (Figure 1.17): activation of expression in the host strain; heterologous cloning and expression; and mutation of key genes followed by metabolic profiling. An example of the first approach was the identification of the stambomycins, a family of 51-membered glycosylated macrolides, synthesised by an PKS gene cluster spanning almost 150kb found in *Streptomyces ambofaciens* [179]. Fermentation extracts showed no evidence for the product of this gene cluster, but after constitutive expression of a LuxR activator gene present in the cluster, stambomycins were detected and could be characterised. The second strategy was used in the identification and characterisation of grisemycin, a peptide natural product produced by *S. griseus* [180]. The gene cluster had previously been identified by comparing homologous genes to the characterised cypemycin gene cluster and a mass signal could be seen that matched the predicted mass of the unknown compound. To verify that the gene cluster was involved in grisemycin biosynthesis the entire gene cluster was cloned into a plasmid and transformed into *S. coelicolor* A3(2) and heterologous production was confirmed by mass spectrometry. The third strategy is useful when the gene cluster is expressed, but the product is unknown, for example in the identification and characterisation of the cuevaenes,

ansamycins from *Streptomyces* sp. LZ35 [181]. By identification and subsequent knocking out of key biosynthetic genes, such as Cuv10, a putative 3-hydroxybenzoate (3-HBA) synthase, the authors could study the metabolic profiles of several mutants and identify the product of the gene cluster. By feeding 3-HBA to the ΔCuv10 mutant, the authors could also confirm the function of the putative gene.

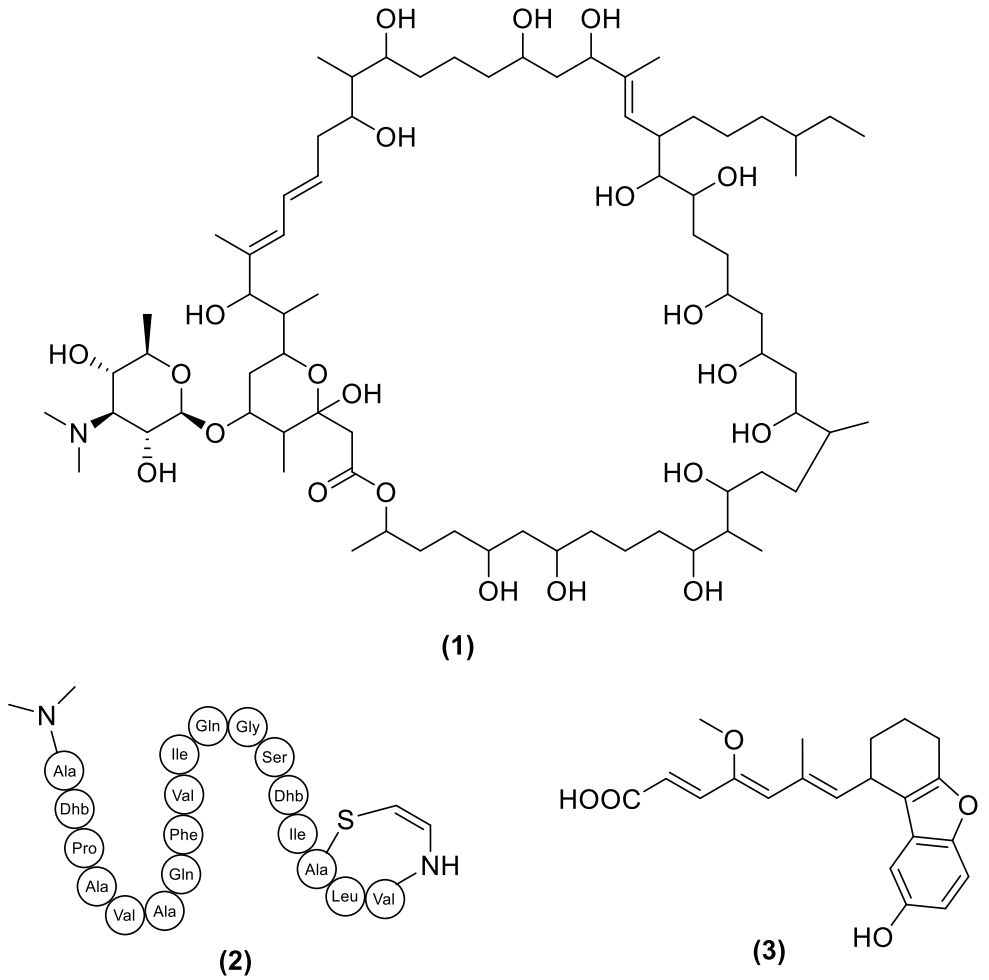


Figure 1.17: Natural products identified through genome mining. (1) stambomycin A **(2)** grisemycin **(3)** cuevaene A.

Metagenomic mining is an important tool to estimate the total biosynthetic potential of organisms in certain complex environments, such as soil, and to identify if there are any gene clusters of note and worth pursuing. It has been estimated that over 99% of microorganisms from the environment resist cultivation in a laboratory setting with unculturable organisms in every prokaryotic group, and even entire divisions with no cultured representative [182]. This is an extremely deep reservoir of potential novel natural products, and achievements such as the identification of novel

pentacyclic type II polyketides [183], and a novel indolotryptoline with antiproliferative activity [184] encourage further study. The characterisation of the potent cell wall synthesis inhibitor teixobactin [185] and isolation of other organisms previously thought unculturable [182] have shown the importance of non-traditional culturing techniques in this endeavour.

1.4: Engineering of Complex Polyketide Biosynthesis

1.4.1 Engineering in the native host

Two different experimental approaches are available for the engineering of the biosynthetic pathway to complex polyketides, which are encoded by gene clusters typically between 70-150 kbp in size. The first of these is the use of homologous recombination between identical sequences to delete, alter, or swap segments of genomic DNA in the natural host organism is the first strategy used to attempt engineering of PKS systems. This approach ensures the presence of the required precursors and any tailoring enzymes. The architecture of PKSs is modular, and the boundaries and active sites of each domain have been identified and are increasingly well understood. The hypothesis behind combinatorial biosynthesis was that if the component enzymes of a PKS system were sufficiently autonomous to carry out their natural functions in alternative contexts and with new substrates then it should be possible to create novel PKSs that produce novel polyketide products by rearranging the domains.

Through the directed manipulation of domains, hybrid PKSs have indeed been produced with novel properties including: modified levels of β-keto reduction in the avermectin PKS [186]; increased and decreased polyketide chain lengths in the erythromycin [131] and spinosyn PKS [187]; respectively, and changed selection for starter [188] and extension [189] units in model PKS systems based on DEBS. However, these experiments have also shown that modular PKSs are not entirely modular, and some are markedly less modular than others. Crucially, we do not yet understand well enough the spatial architecture of native PKS modules required for non-native domains to function together correctly [157], which inevitably leads to slow, inefficient and low yielding enzymatic reactions.

Despite these problems, over the last few years several successful examples have also been published of a knowledge-based approach to the systematic engineering of the biosynthetic pathways to complex polyketides. These have included for example the generation of macbecin analogues which combine extreme potency against their Hsp90 target with greater safety thanks to elimination of the undesirable quinone moiety in this and other antitumour ansamycins such as geldanamycin [190]; and

the generation of over 130 different rapalogs from *S. rapamycinicus* [132,133]. It is important to note that an integrated approach to engineering the pathway is essential: additional manipulation of precursor genes and/or genes for late-stage tailoring maximises the number of new compounds that are obtained from PKS engineering. While this approach to engineering organisms does result in new insights and in the production of valuable chemicals it has the drawback of requiring detailed knowledge of the PKS system and of its target, in order to design the alterations to the molecule accordingly. The approach is also slow and labour-intensive, and it requires careful initial exploration of the strain to confirm that it is genetically tractable and amenable to engineering. Homologous recombination can be very difficult to achieve in organisms such as *Streptomyces* spp. Also, vectors suitable for *Streptomyces* may not function in rare actinomycetes such as *Saccharopolyspora*.

1.4.2 Engineering in a heterologous host strain

An alternative engineering strategy is the transfer to, and expression of, the target PKS gene cluster in a heterologous host that is more tractable to genetic engineering. Heterologous expression has already been discussed as a valuable tool in (meta)genome mining, both of known clusters and of orphan clusters (section 1.3.3). *S. coelicolor*, *S. lividans, S. avermitilis* and *S. albus* are often used. *S. coelicolor* can naturally produce high yields of secondary metabolites, but by 'cleaning' the strain by removing endogenous plasmids, and deleting or inactivating biosynthetic gene clusters, the yields of heterologously expressed natural products have been increased [191,192]. A derivative of *S. avermitilis* from which endogenous gene clusters have been systematically removed has also been used successfully for expression of heterologous polyketides [193]. Cleaned strains have the benefit of a simpler metabolic profile, easing analysis of any gene cluster introduced to the organism.

The general approach to heterologous expression is to amplify the PKS genes from the less tractable strain, assemble them in *E. coli*, and transfer them by conjugation into more tractable strains. Once in the new host, the genes can then be integrated into the chromosome by homologous recombination or through a suitable attachment site. An example is the successful heterologous expression of the gene cluster for FK506, a clinically relevant immunosuppressant. The entire 83.5kb FK506 gene cluster was assembled into a 130kb PAC clone and used to transform different *S. coelicolor* derivatives, giving yields approximately five-fold higher than in the wild-type strain, *S. tsukubaensis* [194]. Progress in this area has focussed on ways to improve the promiscuity of the heterologous hosts, and to improve the expression of introduced gene cluster [157]. Identifying promoters that function well with heterologous systems is an important factor, however it appears that PKS gene clusters prefer

their own PKS-associated promoter [195], so creating a single system to support a range of PKS gene clusters is unlikely. An important pitfall is the propensity of the tandemly repetitive PKS gene sequences to undergo unwanted recombination events during transfer to *E. col*i and thence to the *Streptomyces* host.

E. coli itself has been developed as a host for heterologous expression of PKS genes. However, extensive engineering was required to allow production of polyketides. *E. coli* naturally produces acetyl-CoA and malonyl-CoA, but lacks any significant quantity of other precursors, such as propionyl-CoA, or methylmalonyl-CoA [196]. Despite this, a series of modifications has resulted in an *E. coli* strain that can convert exogenous propionate into 6-dEB, the precursor scaffold to erythromycin, with a yield comparable to an overproducing mutant strain [196]. More recent advances have developed upon the 6-dEB producing *E. coli* strain, introducing 17 new heterologous genes to provide the enzymes necessary to glycosylate 6-dEB converting it to erythromycin C [197]. This work establishes a basis for production of other glycosylated polyketides in *E. coli*.

More recently, the potential for *Saccharomyces cerevisiae* as a heterologous host has been shown. A strain of *S. cerevisiae* was modified to contain two additional plasmids, carrying the genes encoding 6-methylsalicyclic acid (6-MSA) synthase from *Penicillium patulum* and a PPTase from *Aspergillus nidulans*, respectively. Yields of 6-MSA were increased by replacing the promotor sequence of ACC1, the gene encoding acetyl-CoA carboxylase, which catalyses the conversion of acetyl-CoA into malonyl-CoA with a strong, constitutive promotor [198]. *S. cerevisiae* is a popular host for cloning thanks to the simplicity and high efficiency of homologous recombination in yeast, which requires only short overlapping sequences of 29 nucleotides to be able to assembly multiple DNA fragments in a single step [199]. Developments in yeast heterologous expression have included additional tools to aid with cloning in this organism. Using an any-gene-any-plasmid approach [200], a yeast cloning cassette can be introduced into any vector, along with multiple DNA fragments, at efficiencies approaching 100%, far higher than commonly used ligase-based techniques. This approach could be of great use to the study of PKSs, allowing rapid, efficient assembly of PCR products into a yeast or *E. coli* compatible plasmid. Despite this exciting progress, heterologous expression it still remains to be seen whether heterologous expression will deliver libraries of productive mutant PKS assembly lines.

1.4.3: Exploratory approaches to engineering of *Streptomyces* spp.

Recently, a chance discovery by a group at Isomerase Therapeutics developed into a new method for rapidly recombining PKS gene clusters to generate novel polyketide analogues [201]. The method, which has been termed accelerated evolution (AE), takes advantage of the many highly homologous regions within PKS gene clusters to replace, add, or remove modules to generate diversity, in a way that mimics a plausible mechanism for the natural evolution of modular PKSs [201]. What is particularly exciting about AE is that the diversity appears to come from genetic rearrangements induced by the relaxation step of the homologous recombination, meaning a single experiment could generate a library of analogous molecules. Furthermore, while some knowledge of the target PKS is required, this approach is exploratory in that it has the potential to create a range of unpredicted products and by studying the crossover points, could reveals new insights into PKS evolution and architecture.

Another interesting approach to exploring the potential for combinatorial biosynthesis of modular systems is the reconstitution *in vitro* of (parts of) a biosynthetic pathway using recombinantly expressed and purified enzymes. Biochemical experiments were vital to the early development of our understanding of assembly-line polyketide synthases. They were used to show, surprisingly, that modular PKSs multienzymes are homodimeric [17]; to show that the DEBS AT is stereospecific, accepting only the (2S) isomer of methylmalonyl-CoA as substrate [105]; and to show that individual modules of DEBS are autonomously active [202]. Now, *in vitro* reconstitution offers not only critical validation of proposed biosynthetic mechanisms but also insight into potential bottlenecks in biosynthesis, and into the substrate specificity of individual enzymes, information which can be used to guide *in vivo* engineering. There have been many examples of the application of *in vitro* reconstitution and the insights these attempts can bring [203]. One of the first attempts to produce an aromatic polyketide *in vitro* was by Carreras and colleagues when they reconstituted the actinorhodin PKS from its purified components [204]. Using ^{14}C-labeled malonyl-CoA they observed shunt product intermediates of the actinorhodin biosynthetic pathway. Through this work they also observed that the malonyl-CoA:ACP transacylase (MAT) is shared by the fatty acid synthase in the natural host, and notably, was the discovery that the KS domain and chain length factor (CLF) of type II PKSs formed a heterodimer. The insights gained by this *in vitro* reconstitution paved the way for further mechanistic analyses of the how the CLF controls polyketide chain length, and enabled the production of more advanced intermediates through *in vitro* reconstitution [205,206]. A more recent example is found in the elucidation of the biosynthesis of antifungal C-nucleoside malayamycin which identified the key precursor as 5'-pseudouridine-monophosphate (5'-Ψ-MP) and provided the opportunity for engineering through mutasynthetic or semisynthetic approaches (Hong *et al*., (2018) *ms*. in the press).

Aims of the project

This project aimed to evaluate two powerful and complementary approaches to the discovery and engineering of biosynthetic gene clusters in actinomycete bacteria, and their potential to deliver novel chemical diversity of potential therapeutic value.

First, the recently introduced (and still poorly-understood) technique termed accelerated evolution (AE), for which the published work on rapamycin analogues (rapalogs) provides a worked example, would be applied to three other antibiotic producing strains, *Saccharopolyspora erythraea*, *Streptomyces filipinensis*, and *Streptomyces cinamonensis* and evaluated for its potential to generate not only antibiotic analogues, but potentially also novel insights into PKS evolution.

Secondly, a genome-level analysis of *Streptomyces albus* DSM40763, producer of the novel antibacterial compound pseudouridimycin (PUM), would be undertaken. This would allow a detailed survey of the biosynthetic potential of the strain. In particular, it would provide access to the biosynthetic gene cluster for PUM, a remarkable C-nucleoside antibiotic which is now known to target a novel site on bacterial RNA polymerase and would allow *in vitro* and *in vivo* studies to trace the biosynthetic pathway and provide a platform for generation of novel analogues.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Chemicals

All chemicals were purchased from Sigma-Aldrich and were of analytical grade. All solvents were purchased from Fischer Scientific and were of HPLC grade. Unless stated all water was purified with the MilliQ Integral Water Purification system (Merck Millipore). All antibiotics were purchased from Sigma-Aldrich and were filtered before being diluted to the stock concentration listed in Table 2.1 and were stored at -20°C.

Table 2.1. Antibiotics used in this work

Antibiotic	Solvent	Stock Concentration (mg/ml)	Working Concentration (µg/ml)
Ampicillin	H_2O	100	100
Apramycin	H_2O	50	50
Kanamycin	H_2O	50	50
Chloramphenicol	EtOH	25	25
Nalidixic acid	0.15M NaOH	25	25
Thiostrepton	DMSO	50	50

2.1.2 Enzymes

Table 2.2 contains a list of all purchased enzymes used in this study. Enzymes were stored at -20°C and were used with the provided buffers.

Table 2.2: Enzymes used in this work

Enzyme	Seller
Fast Digest restriction endonucleases	ThermoFisher Scientific
Phusion High-Fidelity PCR Master Mix	New England Biolabs
Q5 High-Fidelity DNA Polymerase	New England Biolabs
Gibson Assembly Master Mix	New England Biolabs
NEBuilder HiFi DNA Assembly Master Mix	New England Biolabs
BioMix Red	Bioline
Proteinase K	Sigma-Aldrich
Lysozyme	Sigma-Aldrich

2.1.3 Culture Media

The media found in Table 2.2 were used in this study. Solid phase media were created by adding agar to the liquid media at a ratio of 20g/l.

Table 2.3: Components of the media used in this work

Medium	Component	Mass per 1L of medium (g)
TSB	Tryptic soya broth (Difco)	30
	MilliQ water	Up to 1L
TSBY	TSB	30
	Sucrose	103
	Yeast Extract	5
	MilliQ water	Up to 1L
LB	Tryptone	10
	Yeast Extract	5
	NaCl	10
	MilliQ water	Up to 1L, pH 7.0
2TY	Tryptone	16
	Yeast Extract	10
	NaCl	5
	MilliQ	Up to 1L, pH 7.0

SFM*	Soya flour	20
	D-Mannitol	20
	Tap water	Up to 1L
	$MgCl_2$	10 ml/L of 1M stock

*Tap water was added to soya flour and autoclaved. D-Mannitol was added to the mixture which was autoclaved a second time. 1M $MgCl_2$ solution was added just before use.

2.1.4 Buffers

Table 2.4: Buffer used in this work

Chemically Competent Cell Buffers	
Buffer A	50 mM $CaCl_2$, 10 mM CH_3COOK, pH 6.2
Buffer B	50 mM $CaCl_2$, 10 mM CH_3COOK, 20% (v/v) glycerol, pH 6.2
Genomic DNA Isolation Buffers	
SET Buffer	75 mM NaCl, 25 mM EDTA, 20 mM Tris-HCl, pH 7.5
Protein Purification Buffers	
Binding Buffer	20 mM Tris-HCl, 500 mM NaCl, pH 7.9
Wash Buffer	20 mM Tris-HCl, 500 mM NaCl, 1 M Imidazole, pH 7.9
Strip Buffer	20 mM Tris-HCl, 500 mM NaCl, 100 mM Imidazole, 100 mM EDTA, pH 7.9
SDS-PAGE Protein Buffer	10% (w/v) SDS, 10 mM DTT, 20% (v/v) glycerol, 200 mM Tris-HCl, 0.05% (w/v) Bromophenol blue, pH 6.8

2.1.5 Oligonucleotides

A complete list of oligonucleotides used in this work is given in the appendix.

2.1.6 Vectors

Table 2.5: Vectors created and used in this work

Plasmid	Features	Reference
pKC1139	oriT, ori (psG5), aac(3)IV, LacZ α	[207]
pET28a	ori, f1 ori, T7, T7 tag, His tag, T7 term, *lacI*, *kanR*	Invitrogen
pYH7	Derived from pHZ1358, *oriT*, *ori* (pIJ101), *ori* (pBR322), rep, *sti*, *tsr*, *bla*, *acc*(3)IV	[208]
pKC1139-Fil1	Derived from pKC1139, contains a fragment of module 11 from the filipin PKS gene cluster in A936	This work
pKC1139-Fil2	Derived from pKC1139, contains a fragment of module 11 from the filipin PKS gene cluster in A936	This work
pKC1139-Fil3	Derived from pKC1139, contains a fragment of module 11 from the filipin PKS gene cluster in A936	This work
pKC1139-Fil4	Derived from pKC1139, contains a fragment of module 11 from the filipin PKS gene cluster in A936	This work
pKC1139-Fil5	Derived from pKC1139, contains a fragment of module 2 from the filipin PKS gene cluster in A936	This work
pKC1139-Ery1	Derived from pKC1139, contains a fragment of module 6 from the erythromycin PKS gene cluster in A009	This work
pKC1139-Ery2	Derived from pKC1139, contains a fragment of module 6 from the erythromycin PKS gene cluster in A009	This work
pKC1139-Ery3	Derived from pKC1139, contains a fragment of module 6 from the erythromycin PKS gene cluster in A009	This work
pKC1139-Ery4	Derived from pKC1139, contains a fragment of module 6 from the erythromycin PKS gene cluster in A009	This work
pKC1139-Pre1	Derived from pKC1139, contains a fragment of module 2 from the monensin PKS gene cluster in A568	This work
pKC1139-Pre2	Derived from pKC1139, contains a fragment of module 2 from the monensin PKS gene cluster in A568	This work
pKC1139-Pre3	Derived from pKC1139, contains a fragment of module 2 from the monensin PKS gene cluster in A568	This work

pKC1139-Pre4	Derived from pKC1139, contains a fragment of module 2 from the monensin PKS gene cluster in A568	This work
pKC1139-Pre5	Derived from pKC1139, contains a fragment of module 2 from the monensin PKS gene cluster in A568	This work
pKC1139-Pre6	Derived from pKC1139, contains a fragment of module 2 from the monensin PKS gene cluster in A568	This work
pET28a-psmH	Derived from pET28a, contains the *psmH* gene from the pseudouridimycin gene cluster in A980	This work
pET28a-psmI	Derived from pET28a, contains the *psmI* gene from the pseudouridimycin gene cluster in A980	This work
pET28a-psmK	Derived from pET28a, contains the *psmK* gene from the pseudouridimycin gene cluster in A980	This work
pET28a-psmM	Derived from pET28a, contains the *psmM* gene from the pseudouridimycin gene cluster in A980	This work
pET28a-psmO	Derived from pET28a, contains the *psmO* gene from the pseudouridimycin gene cluster in A980	This work
pET28a-yeiN	Derived from pET28a, contains the *yeiN* gene from *E. coli* O81	This work
pYH7-ΔpsmN	Derived from pYH7, contains DNA designed for knocking out the *psmN* gene in A980	This work
pYH7-ΔpsmU	Derived from pYH7, contains DNA designed for knocking out the *psmU* gene in A980	This work
pYH7-ΔpsmB	Derived from pYH7, contains DNA designed for knocking out the *psmB* gene in A980	This work

2.1.7 Bacterial Strains

Table 2.6: Bacterial strains used in this work

Escherichia coli strains	**Genotype**	**Reference**
DH10B	F$^-$ *endA1 deoR*$^+$ *recA1 galE15 galK16 nupG rpsL* Δ*(lac)X74* ϕ80*lacZΔM15 araD139* Δ*(ara,leu)7697 mcrA* Δ*(mrr-hsdRMS-mcrBC)* Str^R λ$^-$	Invitrogen
BL21	*E. coli* B F$^-$ *ompT gal dcm lon hsdS*$_B$($r_B^- m_B^-$) [*malB*$^+$]$_{K-12}$(λ^S)	NEB
ET12567/pUZ8002	F$^-$ dam-13::Tn9 dcm-6 hsdM hsdR zjj-202::Tn10 recF143 galK2 galT22 ara-14 lacY1 xyl-5 leuB6 thi-1 tonA31 rpsL136 hisG4 tsx-78 mtl-1 glnV44	(MacNeil *et al*. 1992) [209]

Streptomyces and Saccharopolyspora strains	**Genotype**	**Reference**
S. filipinensis DSM40112 (A936)	Filipin-producing wild type strain	DSMZ
S. erythraea NRRL 2338 (A009)	Erythromycin-producing wild type strain	NRRL
S. cinnamonensis (A568)	ΔmonCI, ΔmonBI, ΔmonBII, premonensin-producing overproducer strain	PFL
S. albus DSM40763 (A980)	Pseudouridimycin-producing wild type strain	DSMZ
A980:ΔpsmN	A980 with psmN knocked out	This work
A980:ΔpsmU	A980 with psmU knocked out	This work
A980:ΔpsmB	A980 with psmB knocked out	This work

2.2 Microbiological Methods

2.2.1 Growth and Maintenance of *E. coli*

E. coli strains were grown and maintained in solid or liquid LB or 2TY supplemented with the appropriate antibiotics. Unless stated all *E. coli* cultures were grown at 37°C overnight in Falcon tubes or on plates. Liquid cultures were shaken at 220 rpm.

2.2.2 Growth and Maintenance of *Streptomyces*

Streptomyces cultures were grown and maintained in solid or liquid TSB, TSBY or SFM. All *Streptomyces* cultures were grown at 30°C in conical flasks with metal springs to provide aeration, or on plates. Liquid cultures were shaken at 220 rpm for 72 hrs. Solid media cultures were incubated for 3-14 days. For spore isolation cultures were grown on solid SFM media for 14 days. Spores were collected by the addition of 1-2 ml MilliQ water and agitation by sterilised cotton buds. The spore suspension was cleaned by filtration through cotton wool.

2.2.3 Growth of *Streptomyces* for metabolite production

For production of filipin, seed cultures of *S. filipinensis* were grown in 5 ml SFM for 48 hours at 30°C, 200 rpm in 50 ml flasks with metal springs. 10% (v/v) of the seed medium was used to inoculate a larger culture of 50 ml SFM in a 250 ml flask with metal springs. The larger culture was incubated for up to 7 days at 30°C, 200 rpm.

For production tests of a large number of *S. filipinensis* mutants, a 96 well plate with 2 ml wells was used. Each well was filled with 1 ml of SFM and a small glass bead was added to prevent clumping. The wells were inoculated with a small piece of sporulating mycelium grown on SFM agar for 7 days. The 96 well plates were incubated for 5-7 days at 30°C, 200 rpm.

For production of pseudouridimycin, seed cultures of A980 were grown in 5 ml TSBY for 48 hours at 30°C, 200 rpm in 50 ml flasks with metal springs. 10% (v/v) of the seed medium was used to inoculate a larger culture of 50 ml SFM in a 250 ml flask with metal springs. The larger culture was incubated for up to 7 days at 30°C, 200 rpm.

2.2.4 Preparation of chemically competent cells

For the preparation of chemically competent cells an overnight culture of *E. coli* ET12567/pUZ8002 or DH10B was used as an inoculum (0.5 ml into 50 ml) of 2TY with the appropriate antibiotics. The culture was incubated at 37°C with shaking at 220 rpm until an optical density of A_{600} = 0.3-0.5 was achieved. The cells were pelleted by centrifugation and resuspended in 30 ml ice cold buffer A (Table 2.3). After 1 hr incubation with occasional agitation the cells were pelleted by centrifugation and resuspended in 1-3 ml ice cold buffer B. The culture was aliquoted (50-100 µl) in the cold room. All tips, cryotubes, and racks were pre-cooled in the cold room. Aliquots were frozen in dry ice and stored at -80°C.

2.2.5 Transformation of *E. coli*

An aliquot of *E. coli* ET12567/pUZ8002 or DH10B competent cells was thawed on ice for 20 mins. 1-5 µl (2-200 ng) of DNA was added and the mixture was left of ice for 30 mins. Cells were submitted to heat shock for 30 s at 42°C and then placed on ice for 2 mins. 250-750 µl was added and the mixture was incubated at 37°C for 1 hr with shaking at 220 rpm. 50-200 µl was spread on an LB or 2TY plate with appropriate antibiotics and incubated at 37°C overnight. Successfully transformed colonies were identified through colony PCR (cPCR).

2.2.6 Transformation of *Streptomyces* by conjugation

Transformations of *Streptomyces* with integrative plasmids were carried out through intergenic conjugation with *E. coli* ET12567/pUZ8002 following protocols based on those in the laboratory manual “Practical Streptomyces Genetics”.

Competent cells of *E. coli* ET12567/pUZ8002 were prepared and transformed with the integrative plasmid as described above. Single colonies were picked and used to inoculate 10 ml LB with chloramphenicol, kanamycin, and the desired selective antibiotic and were grown overnight at 37°C, 220 rpm. Fresh LB (10 ml) containing half concentration of antibiotics was inoculated with 100 µl of the overnight culture and grown at 37°C, 220 rpm until an A_{600} of 0.4-0.6 was reached. Meanwhile a culture of *Streptomyces* was grown in 50 ml TSBY for 48 to 72 hrs. *E. coli* and *Streptomyces* cells were harvested by centrifugation and were washed three times in LB. The *Streptomyces* pellet was

resuspended in 500 µl LB and the *E. coli* pellet was resuspended in 100 µl LB. 100 µl of the *E. coli* and the *Streptomyces* for each conjugation plate were mixed together before being spread on an SFM agar plate and incubated at 30°C for 12-20 hrs. The plates were overlaid with 1 ml of MilliQ water containing nalidixic acid and the selective antibiotic. The plates were then incubated at 30°C until colonies were visible.

Conjugations using *Streptomyces* spores instead of a liquid culture were performed as above, except that before mixing with resuspended *E. coli* the spores were heat shocked at 50°C for 10 mins before incubation on ice for 10 mins.

2.3 Molecular Biology Methods

2.3.1 Isolation of Plasmid DNA from *E. coli*

Plasmid DNA was isolated from 5-10 ml cultures grown overnight using the Plasmid Mini Kit I (Omega BioTek) following the manufacturer's protocol.

2.3.2 Isolation of Genomic DNA from Streptomyces

Streptomyces genomic DNA was isolated using the salting out procedure outlined in the laboratory manual "Practical Streptomyces Genetics".

A culture of *Streptomyces* was grown in 10 ml TSBY at 30°C, 220 rpm for 48 hrs. The cells were then harvested by centrifugation. The pellet was resuspended in 5 ml SET buffer, 100 µl of lysozyme solution (50 mg/ml) was added and the mixture was incubated at 37°C for 1 hr. 140 µl proteinase K solution (20 mg/ml) was added, the mixture was mixed, then 600 µl SDS solution (10%), pre-warmed to 55°C, was added. The mixture was incubated at 55°C for 2 hrs with occasional inversion. 2 ml of 5M NaCl was added and the mixture was allowed to cool to 37°C. 5 ml chloroform was added, and the mixture was mixed for 30 mins at room temperature before being centrifuged for 15 mins at 5000 x *g*. The upper layer was transferred to a fresh tube, 0.6 volumes of isopropanol was added, and the tube was inverted until the DNA became visible. The DNA was spooled onto a pipette tip, transferred to a fresh tube and washed twice with 5 ml ethanol (70%). The DNA was then air dried and redissolved in 0.5-2 ml MilliQ water.

2.3.3 Polymerase Chain Reaction

All polymerase chain reaction (PCR) reactions were carried out on a GeneAmp PCR system 9700 (PE Applied Biosystems). For cloning, a 2x Phusion Master Mix with GC buffer or a 2x Q5 High-fidelity Master Mix was used (Table 2.7, condition 1). For colony PCR a 2x Bioline BioMix with Taq polymerase (Table 2.7, condition 2) was used. Typical reactions conditions are shown in Table 2.7 and typical PCR temperature cycling conditions are shown in Table 2.8.

Table 2.7: Components required for a single polymerase chain reaction

Components	Reaction conditions 1 (µl)	Reactions conditions 2 (µl)
Polymerase Mix	25	10
DMSO	1.5	1
Primer 1 (10 pM)	2.5	1
Primer 2 (10 pM)	2.5	1
Template DNA	X	X
MilliQ water	Up to 50 µl	Up to 20 µl

Table 2.8: Thermal cycling conditions for PCR. The annealing temperature is dependent on the melting temperature of the primers

PCR cycle stage (30 cycles)	Temperature (°C)	Time (sec)
Initial denaturation	95	120
Denaturation	95	10
Annealing	X	20
Extension	72 (68 for Taq)	30-60 per kb
Final extension	72 (68 for Taq)	300
Rest	4	∞

2.3.4 Colony PCR for *E. coli*

Routine screening of transformed *E. coli* was performed by colony PCR (cPCR) of single colonies. Isolated colonies were picked with a toothpick and agitated in a 250 µl PCR tube containing 20 µl of

cPCR master mix. The toothpick was then streaked on an LB or 2TY plate containing appropriate antibiotics. The plates were incubated at 37°C overnight. The PCR tubes were incubated according to the conditions described above.

2.3.5 Colony PCR for *Streptomyces*

Screening of large numbers of transformed *Streptomyces* colonies was performed by a modified cPCR protocol. Small amounts of mycelium were scraped off each colony using a toothpick and agitated in a 250 µl PCR tube containing 20 µl of 1:1 DMSO:MilliQ water. The tubes were microwaved at a high setting for 30 s on, 30 s off for a total microwave times of 2 min. 1 µl of the mixture was used as the template DNA in the PCR protocol described above.

2.3.6 Agarose Gel Electrophoresis

DNA separation and visualisation was performed by agarose gel electrophoresis using a horizontal flow tank. Agarose was dissolved in 1x TAE buffer (4.85 g Tris, 1.14 ml glacial acetic acid, 2 ml 0.5M EDTA Ph 8.0) at 0.7% (w/v) and SYBR Safe gel stain (3 µl to 100ml buffer) was added for DNA visualisation. DNA samples were mixed with a 6x loading buffer (Fermentas) and GeneRuler 1kb Plus DNA ladder (Fermentas) was used as a molecular DNA length marker. DNA was visualised with a UV transilluminator (254 nm).

2.3.7 Recovery of DNA from Agarose Gel

For preparative purposed DNA fragments bands were cut from the gel and purified using a QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's protocol.

2.3.8 Restriction Enzyme Digestion of DNA

Analytical and preparative DNA digestion was performed using the following restriction endonuclease digestion protocol. The components and volumes are shown in Table 2.9. Digestion mixtures were incubated at 37°C for 0.5-1 hrs, then separated and visualised by gel electrophoresis.

Table 2.9: Components required for DNA restriction digest

Components	Analytical digestion mix (µl)	Preparative digestion mix (µl)
DNA to be digested	X (100 ng of DNA)	X (1 µg of DNA)
10x digestion buffer	1	5
Restriction enzyme	0.2	1
MilliQ water	Up to 10 µl	Up to 50 µl

2.3.9 Ligation of DNA Fragments

Ligations of DNA fragments were carried out using the protocol laid out in Table 2.10. The ligation reaction mixture was incubated at 16°C overnight, after which the entire reaction volume was used for transformation of *E. coli*.

Table 2.10: Components required for DNA ligation

Components	Volume (µl)
T4 DNA ligase	1
10x ligation buffer	1
Linearised vector DNA	X (20-50 ng of DNA)
Insert DNA	Y (5x molar excess of insert DNA)
MilliQ water	Up to 10 µl

2.3.10 Isothermal Assembly of DNA Fragments

Isothermal assemblies of DNA fragments were performed using the NEBuilder HiFi DNA Assembly Master Mix (NEB) following the protocol laid out in Table 2.11. The reaction mixture was incubated at 50°C for 1 hr, after which 10 µl of the reaction volume was used for transformation of *E. coli*.

Table 2.11: Components required for isothermal assembly of DNA fragments

Components	**Volume (µl)**
Assembly Master Mix	10
Linearised vector DNA	X (50-100 ng of DNA)
Insert DNA	Y (3x molar excess of insert DNA)
MilliQ water	Up to 20 µl

2.3.11 DNA Sequencing

All cloning was confirmed by sequencing. All routine sequencing reactions were carried out by the DNA Sequencing Facility in the Department of Biochemistry, University of Cambridge.

Whole genome sequencing reactions were carried out by the DNA Sequencing Facility and assembly and automated annotation was performed by Dr Markiyan Samborskyy of the Leadlay Lab.

2.3.12 Protein Expression in E. coli

Protein expression was carried out by taking advantage of the T7 promoter in pET28a to express the target gene in the presence of Isopropyl-β-D-1-thiogalactopyranoside (IPTG).

pET28a vectors containing the target gene downstream of the T7 promoter were constructed following the protocols described above and used to transform *E. coli* BL21(DE3) cells. A single colony was inoculated into LB (10 ml) with appropriate antibiotics and incubated overnight. This seed culture was then used to inoculate LB (1 L) with antibiotics and was incubated until A_{600} = 0.5-1.0 was reached. IPTG was added (final concentration 0.1-0.3 mM) and the culture was incubated at 16°C overnight.

2.3.13 Affinity Chromatography

The pET28a expression system generates His tagged proteins, and so the target protein can be purified from the cell culture by Nickel-NTA affinity chromatography (nitrilotriacetic acid (NTA) is a chelating ligand coupled to a cross-linked 6% agarose resin that when charged with nickel can purify 6xHis tagged recombinant proteins).

The overnight culture, as described above, was pelleted by centrifugation, the supernatant discarded, and the pellet resuspended in binding buffer (30 ml), see table 2.4. The resuspended culture was sonicated on ice and the lysate was centrifuged. This supernatant was then used for chromatography, the steps of which are summarised in Table 2.12.

Table 2.12: Wash steps used in affinity chromatography

Solutions added to column in order	**Volume (ml)**
MilliQ water	10
$NiCl_2$ (50 mM)	4
MilliQ water	4
Binding buffer	4
Lysate	All supernatant
Wash buffer (0 mM Imidazole)	15
Wash buffer (10 mM Imidazole)	15
Wash buffer (40 mM Imidazole)	5
Wash buffer (80 mM Imidazole)	2
Wash buffer (100 mM Imidazole)	2
Wash buffer (150 mM Imidazole)	2
Wash buffer (200 mM Imidazole)	2
Wash buffer (300 mM Imidazole)	2
Strip buffer	5
MilliQ water	20
MilliQ water	20
Ethanol (20%)	5

The lysate fraction, all wash fractions and a sample of the pellet were analysed by SDS-PAGE gel electrophoresis. A sample of each fraction (10 µl) was added to SDS-PAGE protein buffer (10 µl), see table 2.3. The mixture was boiled for 10 mins before being loaded onto the SDS-PAGE gel.

The SDS-PAGE gels were visualised by the addition of InstantBlue, and incubation at room temperature until the bands were visible.

2.4 Chemical and Analytical Methods

2.4.1 Analysis of Polyketide Products

For analysis of filipin, 1 ml was sampled from *S. filipinensis* production cultures and were extracted with 1-2 ml of ethyl acetate. The mixture was shaken vigorously for 5 mins before being separated by centrifugation. The upper layer was taken, and the solvent was removed by drying in air. The residue was redissolved in 500 µl of methanol. Samples were centrifuged before HPLC analysis.

For analysis of pseudouridimycin and related molecules, 1 ml of A980 production cultures was sampled and the cells pelleted by centrifugation. The supernatant was collected and was centrifuged before HPLC analysis.

2.4.2 Liquid Chromatography and Mass Spectrometry

HPLC-MS analysis was performed using an HPLC (Hewlett Packard, Agilent Technologies 1200 series) coupled to a Finnigan MAT LTQ mass spectrometer (Thermo Finnigan) fitted with an electrospray ionisation (ESI) source. The mass spectrometer was run in positive ionisation mode, scanning from *m*/*z* 150 to 1800 with fragmentation at 12 % normalised collision energy.

2.4.3 HPLC analysis

The HPLC was fitted with a Prodigy C18 (250 mm × 4.6 mm, 5µm, Phenomenex) column. Filipin samples were eluted using a solvent system of solvent A (H_2O + 0.1% FA) and solvent B (MeOH), with a linear gradient starting at 50% A for 3 mins, 50%-10% A over 9 mins, 10% A for 8 mins, 10%-0% A over 1 min, 0% A for 2 mins, and 0%-50% A over 4 mins at a flow rate of 0.8 ml/min. Pseudouridimycin and related

standards and intermediates were eluted using a solvent system of solvent A (H_2O + 0.1% TFA) and solvent B (CH_3CN), with a linear gradient starting at 100% A for 5 mins, 100%-95% A over 10 mins, 95%-0% A over 10 mins, 0% A for 4 mins, and 0%-100% A over 1 min at a flow rate of 0.4 ml/min.

2.4.4 NMR analysis

All samples for NMR spectroscopy were prepared with substrates as concentrations of 1 mM, and enzymes at a concentration of 10 µM, in 50 mM phosphate or HEPES buffer, pH 7.5, 100 mM NaCl, supplemented with 10% D2O (Sigma) and 0.0025% 3,3,3-trimethylsilylpropionate (Sigma) in 5 mm Ultra-Imperial grade NMR tubes (Wilmad). All spectra were recorded at 298 K on a Bruker Avance 800 MHz spectrometer equipped with a z-shielded gradient triple resonance probe. One dimensional 1H spectra were acquired by averaging 32 scans using the first row of a NOESY experiment with Watergate solvent suppression.

2.4.5 Measurement of DNA and Protein Concentration

For the measurement of concentrations of DNA and protein in samples, a Nanodrop spectrophotometer (ND-100 v3.2.1) was used.

2.4.6 Bioinformatics and software

Genome sequences were analysed and visualised with Artemis [ref], and genomes were annotated by AntiSMASH 4.0 [178]. BLAST [175] was used to search for DNA and protein homologues in public databases, and Clustal Omega [210] was used to align multiple DNA and protein sequences. SnapGene (GSL Biotech, Chicago, IL) was used to design vectors, and Primer3 [211] was used to design primers. ChemDraw (PerkinElmer, CambridgeSoft) was used to draw chemical structures and calculate exact masses. Xcalibur v2.1 (Thermo Finnigan) was used to process and visualise HPLC-MS data.

Chapter 3: Homologous Recombination Based Engineering of Modular Polyketide Synthases

3.1: Introduction

Recently an attempt at using homologous recombination to carry out a conventional domain swap in the modular rapamycin PKS (Figure 3.1) led to the serendipitous discovery by Dr Steve Kendrew (then at Biotica Technology) of an efficient method to increase the structural diversity of a polyketide template following a single integration event [201]. Instead of the expected product, the various progeny generated by the experiment produced a variety of unexpected rapalogs at comparable titres to the wild type production of rapamycin. Each of the progeny strains produced a single rapalog, and the various progeny, between them producing an array of different rapalogs, were generated from a single crossover event. Follow up studies isolated and characterised the rapalog compounds. It became clear that each rapalog was directly related to rapamycin but lacked between 2 or 12 carbon atoms from the rapamycin backbone, and in one case had 2 additional carbon atoms. The structures matched the predicted polyketide structures that would be produced by truncated or expanded rapamycin PKS multienzymes. This result was repeatable, and further investigation showed that the diversity was generated during the second crossover event. The second recombination event can occur between several homologous regions of DNA in the PKS gene cluster, leading to a mixture of recombinant strains with varying number of modules in the target PKS, as illustrated schematically in Figure 3.2. This integration leads to a high ratio of productive strains to non-productive strains, where the productive strains can be isolated and separated giving multiple individual strains each producing a specific novel polyketide (see Figure 3.3). The productivity of the recombinant strains was surprisingly not much reduced from the parent strain. This is unusual, as manipulations of PKS gene clusters, especially those creating hybrid PKSs, tend to lead to strains with low productivity [212]. Analysis of the genome sequences for the rapalog producing strains revealed that all genomic changes were confined to the PKS region of the rapamycin BGC. For each rapalog producer the new PKS systems matched the predicted sequence of domains based on the rapalog structures. The architecture of the mutant PKSs was consistent with homologous recombination having occurred between modules at regions of high sequence similarity. Additionally, the apparent deletion of one or more modules was often more complex than might be expected. For example, sequence analysis of the mutant producing the 1-module expanded rapalog revealed that the PKS had lost almost two modules (modules 2 and 3) and that three extra modules (modules 11, 12 and 13 from *rapC*) had been inserted into *rapA*. In the cases

where PKS architecture was similar, it was observed that the contractions had occurred at different junction points. Taking the data together, it appeared that the junction points tended to occur either in the coding sequences for the KS or AT domains, or in a region just upstream of the ACP.

It is possible that this approach mimics how modular PKSs might have evolved, utilising processes involving gene duplication, module duplications, homologous recombination (both inter and intra-BGC) and horizontal transfer of genes or gene clusters [213]. Due to the similarities between this process and natural evolution, this approach to PKS engineering has been called Accelerate Evolution (AE).

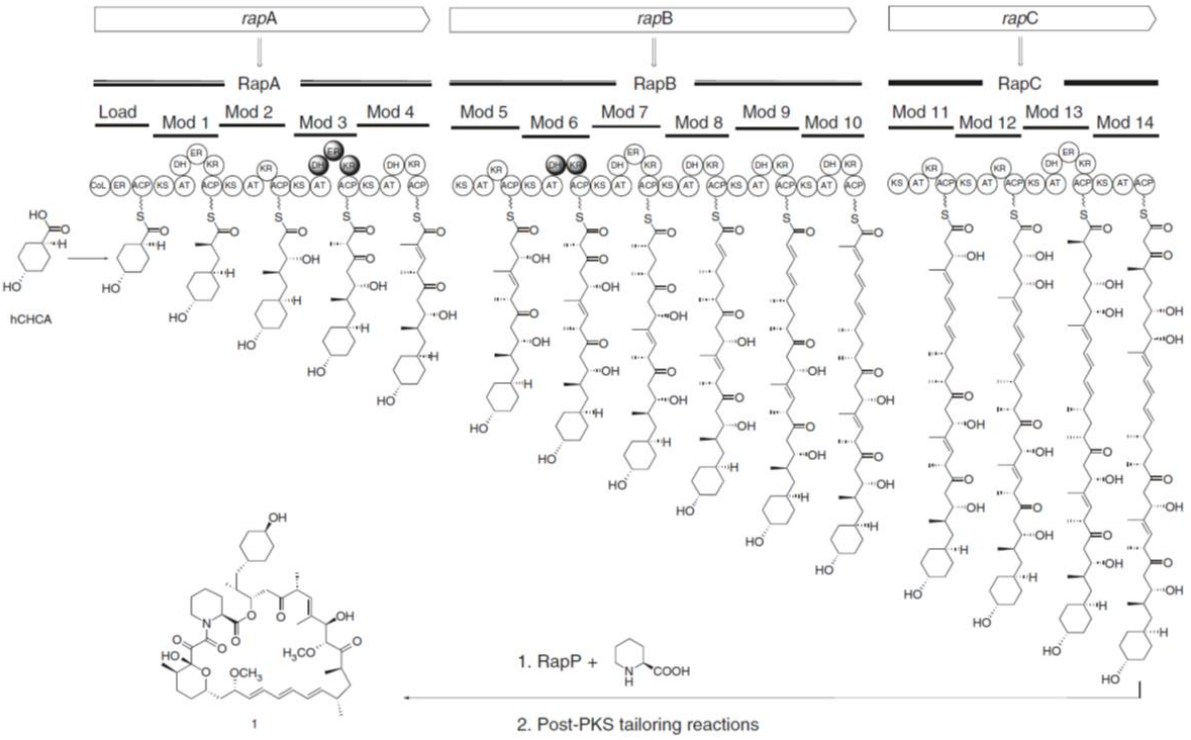


Figure 3.1: Schematic diagram of the rapamycin PKS multienzyme. Adapted from Wlodek et al. [201]. CoL = CoA ligase like, hCHCA = *trans*-4-hydroxycyclohexanecarboxylic acid which is fed to the strain to initiate biosynthesis of rapamycin. Shaded circles indicate inactive domains.

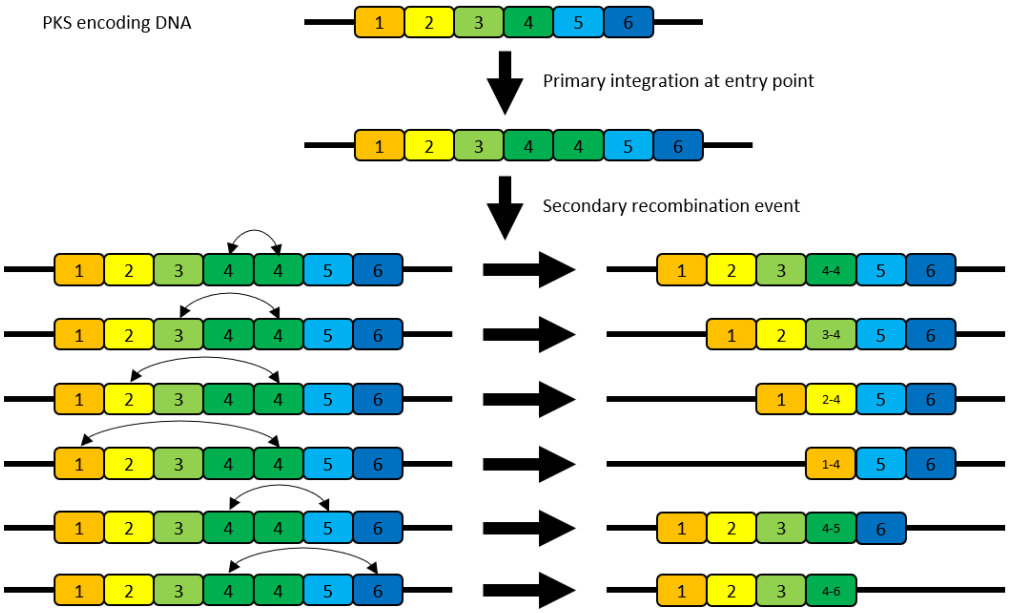


Figure 3.2: Schematic representation of the recombineering process. Primary integration uses entry point DNA encoding a region of module 4 from the PKS, leading to a series of different potential secondary recombination outputs.

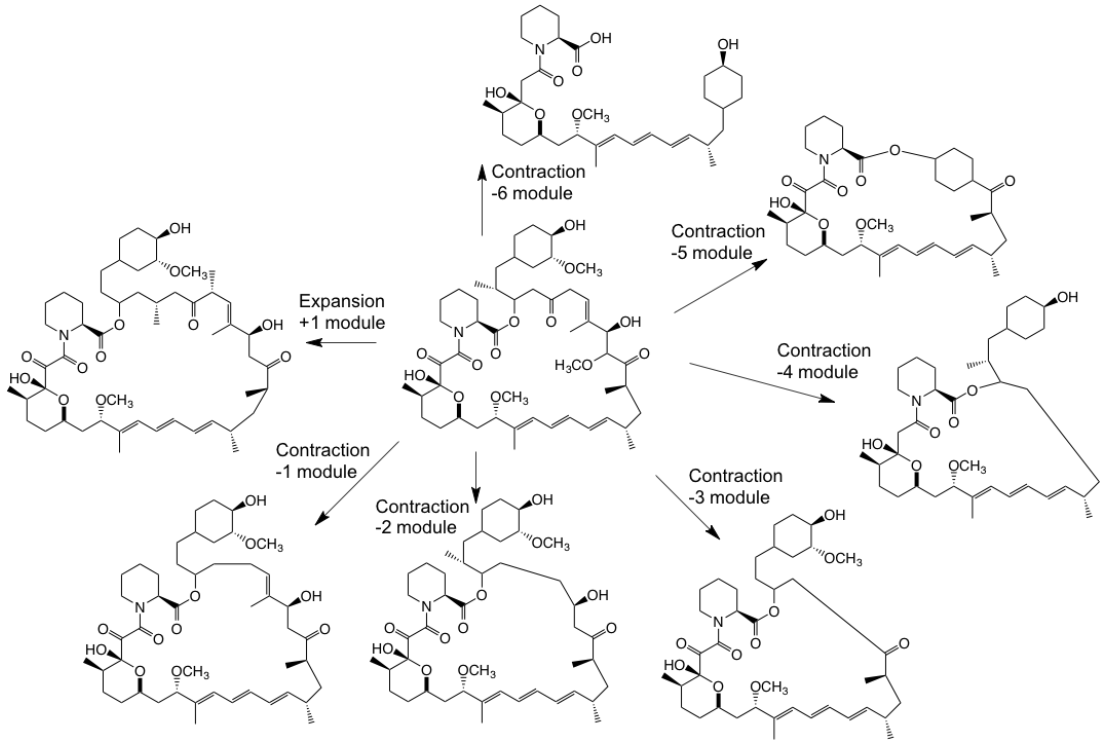


Figure 3.3: Representation of the outputs from recombineering on the rapamycin PKS. In this example, initial recombination was into DNA encoding module 3 of the rapamycin PKS. Adapted from Wlodek et al. [201].

AE differs from traditional engineering approaches by its apparent lack of target. Typically, a conventional attempt at engineering PKS would first decide upon a target molecule then, utilising the co-linearity principle, design a PKS system that could generate that polyketide. AE instead picks a system which could generate useful analogues, then introduces homologous DNA at a target site within the PKS which results in a library of analogous compounds. This principle is illustrated in Figure 3.4.

Wlodek and colleagues at Isomerase Therapeutics began optimising this process to further study it after the rapalogs were identified. To simplify the work, they used short regions of homologous DNA rather than the pairs of flanking DNA sequences that are used to remove or introduce DNA. When designing this entry point DNA, a desired initial integration site can be selected for. This marks a central point from where the second recombination event can occur between homologous regions in other modules, leading to the expanded or contracted macrocyclic ring or linear chain structures of the modified polyketide. The original work suggests a preferential target for the entry point DNA would be a sequence of >100 bases of homologous sequence which is >80% identical (but quickly a standard length of approximately 2 kb was adopted). This high sequence homology is common between comparable domains within PKS multi-enzymes and especially between domains of the same PKS. If the desired entry point lacks sufficient sequence identity, it is possible that the PKS could be prepared for the application of AE by increasing the homology of comparable domains using silent mutations of the sequence. Due to the clinical importance of many polyketide compounds, the ability to create libraries of analogues could lead to novel compounds with therapeutic activities. Replicating the success of AE with the rapamycin gene cluster in other systems would provide important evidence for the general use of this technique, even for the most well-known polyketides such as erythromycin and monensin. AE is still in its infancy, and so will require a thorough investigation of its mode of action, and optimisation of the protocol for each actinomycete strain may well be required.

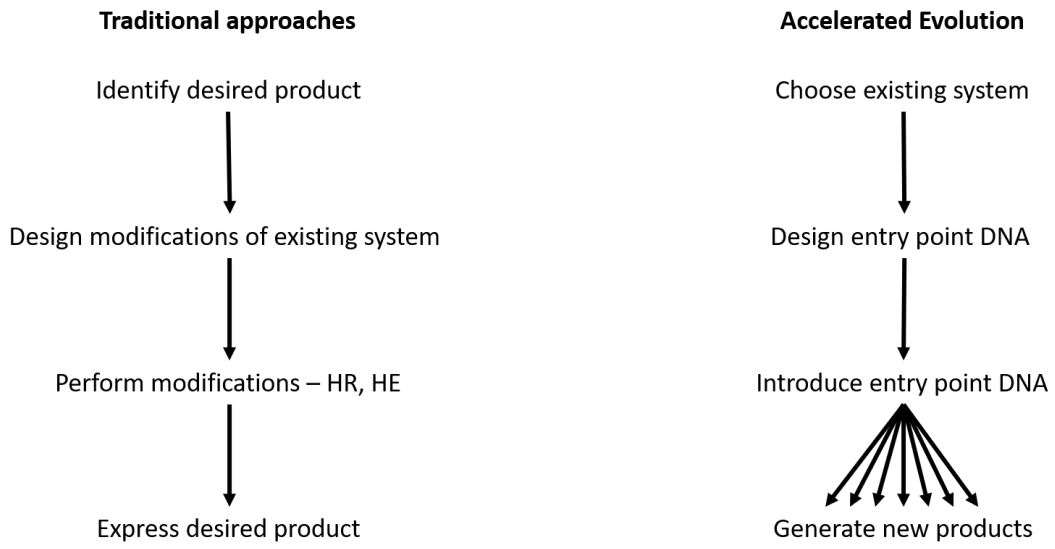


Figure 3.4: Comparison between the design logic of a traditional engineering-based approach and an accelerated evolution-based approach.

There are a number of factors to consider when designing an AE experiment. These are; the suitability of the PKS system for AE, the entry point DNA (the sequence of DNA homologous to the desired integration site), and the vector chosen to carry the entry point DNA into the cell and initiate the secondary recombination event. There are several characteristics which through an examination of the proposed mechanisms, experimental methods, and analytical techniques should determine whether a PKS system is suitable for engineering by AE:

1. High production levels in wild type
2. UV spectrum
3. Low number of background peaks
4. No tailoring enzymes
5. High-quality whole genome sequence
6. Co-linear arrangement of PKS genes
7. Exotic unit required (conditional synthesis)
8. Strain genetically tractable

Most of these characteristics are chosen to assist in characterisation of the fermentation products of any AE-generated strains. A high production level is helpful for any engineering method, and a characteristic UV spectrum allows for rapid analysis by HPLC and UV-vis spectrometry. Due to the nature of AE it will generate products that may have unpredictable molecular weights. Choosing systems that have a low number of background peaks and no post-PKS tailoring enzymes makes

identifying and characterising new mass signals easier. A high-quality whole genome sequence makes genetic analysis easier, and a co-linear arrangement of PKS genes refers to a linear arrangement of PKS genes. Using the rapamycin PKS as an example, a co-linear arrangement would involve the *rapA* gene followed by *rapB* and *rapC*. This is a common feature among PKS gene clusters, but not universal. The co-linear arrangement allows for the maximum possible range of contractions, while minimising the creation of any progeny that have inactive an PKS, through loss of a TE or starting KS domain. An exotic unit involved in the biosynthesis is not mandatory, but will make analysis easier, as it creates the possibility of conditional synthesis. The pathway required to generate the exotic unit can be knocked out, and then the exotic unit fed to the fermentation media to initiate production. The final characteristic, essential for any engineering attempt, is that the strain must be genetically tractable.

Once a system has been selected on these criteria, the choice of the entry point DNA is also an important consideration. It must be a sequence that has a high sequence similarity to multiple comparable domains within the PKS. The entry point DNA also determines the module where primary integration occurs. This is thought to mark the centre from which recombination can occur, although the above examples show that the recombination can also occur at locations some distance from the insertion module. Work by Wlodek and colleagues highlighted certain 'hot-spots' where the most common junction points are found. These were in the KS and AT domains, and in the interdomain region following the ACP domain.

A controllable origin of replication (for example temperature controlled) is known to be critical to the success of the second recombination event, but it must also contain selection markers for both the parent host (for example *E. coli*) and the PKS-producing hosts. There are a number of vectors with temperature sensitive origins of replication, such as pSG5 [214] or pKC1139 [207]. pSG5 is found naturally as a temperature-sensitive plasmid in *Streptomyces ghanaensis* DSM 2932 which replicates only at temperatures below 34°C. The nucleotide sequence of its minimal replicon is the template for the pGM vector family. The naturally temperature-sensitive replication of pSG5 and its derivatives makes it ideal for this technique because the replication can be controlled to induce recombination events. The vector pKC1139 in particular contains the temperature sensitive replicon from pSG5, as well as an apramycin resistance gene, multiple cloning sites and an *oriT* fragment which allows for transfer of the plasmid to the PKS-producing host via conjugation [215]. Assembly of the entry point DNA into the vector is typically achieved using Gibson isothermal ligation assembly [216] in *E. coli* DH10B.

The work by Wlodek *et al*. is not the first recorded evidence of an AE-like modification of a PKS multienzyme, but it is the first to recognise its significant and to repeat and characterise the effect. In 2002 Sergey Zotchev and colleagues from the Norwegian Institute of Science and Technology

reported that an attempt to manipulate the *nysC* gene of the nystatin PKS yielded a recombinant strain that produced a hexaene nystatin derivative [217]. Further investigation revealed that the hexaene compound was a 36-membered macrocycle (compared to the 38-membered nystatin backbone) and was fully decorated by the post-PKS enzymes. Sequence analysis of the recombinant strain revealed an in-frame deletion in the *nysC* gene that had resulted in the loss of an entire module. The recombination occurred between the KR domains of modules 4 and 5, which share 99.9% sequence similarity, and have a G+C% of 64%, much lower than the 74% found in the rest of the gene cluster, suggesting that those sequences were acquired from another organism through horizontal gene transfer. The Norwegian group came to similar conclusions that the modification was generated by homologous recombination between regions of highly similar DNA through a process that mimicked the evolution of the nystatin PKS cluster, however they did not repeat the experiment to investigate whether this was a random and deleterious occurrence, or a process that could be replicated and made useful.

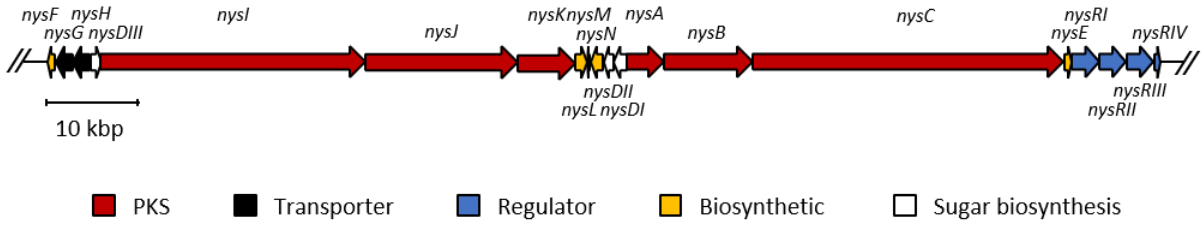
The significance of AE technology is that from a single exconjugant (primary), representing a single integration event, a wealth of diversity can be generated via the random truncations. The technique appears to overcome the limitations of conventional PKS engineering, namely the interference of domain boundaries resulting in poor productivity, by removing the need to artificially determine domain boundaries and instead allowing recombination processes to occur and the selecting for the productive strains. Furthermore, as a single strain can undergo a series of recombination events between different sites in the PKS gene sequence one can potentially create multiple stable strains that have had one or more module inserted or deleted during a single experiment. The analysis of AE-generated hybrid PKS modules could provide valuable insights to guide the rational engineering of modular PKS. Additionally, AE has the potential to become a standard approach to the engineering of PKS in its own right. Compared to rational PKS engineering, where the design of a single DNA construct is aimed at the production of a single strain that produces a single desired compound, AE allows for the rapid parallel creation of a number of strains each producing a member of a family of novel compounds. This work was done with kind advice from scientists at Isomerase Therapeutics, particularly Dr Aleksandra Wlodek.

3.2: Results and Discussion

3.2.1: Nystatin Gene Cluster Rearrangement

The first recorded instance of an AE-like deletion of a PKS module was in the nystatin PKS. Therefore, it made sense to evaluate the nystatin producer *S. noursei* as a possible testbed for extension of AE, since the machinery for homologous recombination could be assumed present in this strain. Other integration sites would be tested in the nystatin PKS, with the aim of eventually generating a library of compounds. The nystatin gene cluster spans 267 kbp. There are six core PKS genes; *nysA,B,C,I,J,K* which between them encode 19 modules (see Figure 3.5).

(A)



(B)

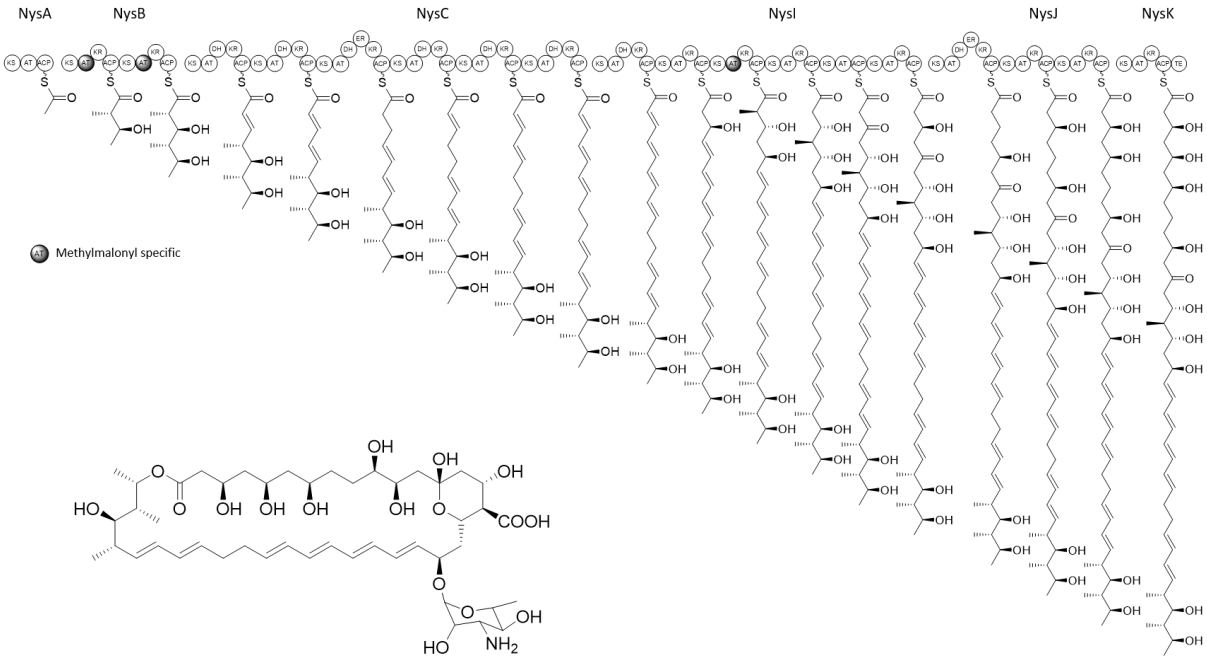


Figure 3.5: Nystatin PKS biosynthesis. (A) Nystatin gene cluster: *Nys*A/B/C/I/J/K: PKS genes, *nys*L: P450 monooxygenase, *nys*M: ferredoxin, *nys*M: P450 monooxygenase, *nys*DII: amino transferase, *nys*DI: glycosyltransferase. **(B)** Nystatin PKS biosynthetic pathway and completed nystatin structure.

As Figure 3.5 illustrates, the nystatin PKS genes are not co-linear with the arrangement of the genes in the cluster, and while there are large regions where homologous recombination can occur harmlessly, the location of *nysA* and *nysK*, which are essential to production of nystatin, presents the risk of generating recombinant nystatin BGCs that are inactive. To maximise the chances of generating novel nystatin derivatives an experiment based on CRISPR/Cas9 engineering was designed to rearrange the nystatin gene cluster.

To obtain the nystatin gene cluster rearrangement, a vector was constructed that contained the cassette necessary to make two accurate cuts to cleave *nysA*, *B*, and *C* (*nysCBA*) from the genome, and to re-join the digested ends. The protocol for designing and creating the vector was based on the protocol outlined in the supplementary materials of Cobb *et al*. [218]. The *nysCBA* PKS genes were chosen for mobilisation, as this region is flanked by larger regions of non-coding DNA than the *nysIJK* genes. Once the vector had been designed, it would be transferred into the host through intergeneric conjugation and the desired cleavage performed. The next step would be to design a second vector that contained a cassette with *nysCBA*, and the required homologous arms to induce homologous recombination and insert *nysCBA* into a co-linear position within the nystatin BGC. This vector would be inserted, and progeny would be screened for the desired insertion. Figure 3.6 illustrates the targeted rearrangement of the nystatin BGC.

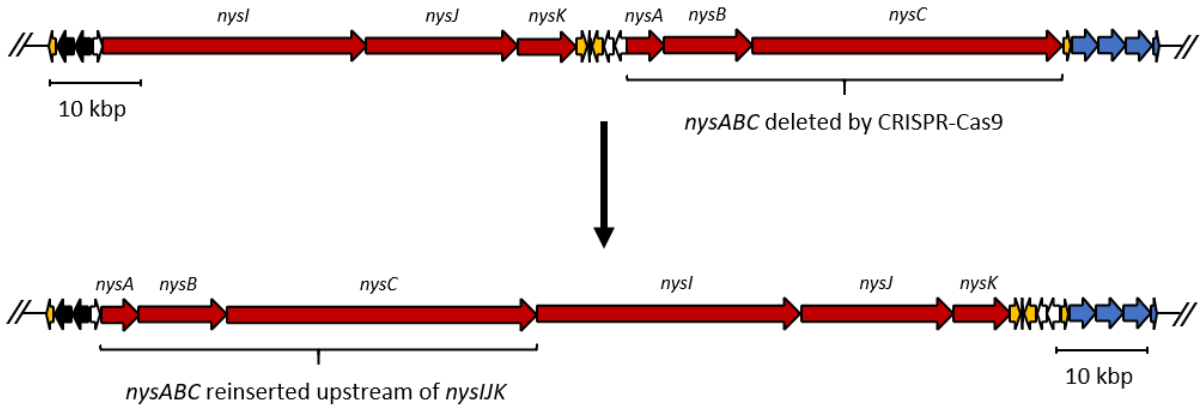


Figure 3.6: The scheme for the rearrangement the nystatin biosynthetic gene cluster. *NysABC* is deleted by two targeted double stranded breaks mediated by the CRISPR-Cas9 complex. The *nysABC* cassette is then constructed and inserted back into the cluster upstream of *nysIJK* to complete the rearrangement.

The *S. noursei* genome was sequenced and assembled by the Department of Biochemistry DNA Sequencing Facility, and the sequence of the nystatin gene cluster was fully checked and confirmed (Figure 3.5). The genes *nys*CBA were to be deleted and reinserted in sequential order adjacent to *nys*KJI. The template sequence was created by amplifying 1kbp of genomic DNA on either side of *nys*CBA to provide flanking arms. The flanking arms were purified by gel electrophoresis and were ligated together with pCRISPomyces2, previously linearised by digestion with the restriction enzyme *Xba*I, by using Gibson isothermal assembly (Figure 3.7) [216].

Candidate crRNA sites were identified using an online tool [219] and target sites were chosen based on the 21bp target sequence and the position relative to *nys*CBA. These crRNA sites were included in the sgRNA sequence containing appropriate promoters and terminators (Figure 3.8). The designed sgRNA was bought as a gene fragment from Eurofins Genomics and ligated into pCRISPomyces2 via Golden Gate assembly (Figure 3.9). The completed pC2-nysCBA plasmid was verified by PCR amplification of the region containing the sgRNA and template inserts, and by sequencing.

Transformation of pC2-nysCBA into *S. noursei* was attempted through conjugation with chemically competent ET12567/pUZ8002 cells which had previously been transformed with pC2-nysCBA. Conjugations were performed on several different media; SFM, MAM, ABB13 and R6, varying concentration of $MgCl_2$, using both mycelium and spores of *S. noursei* and varying incubation times before overlaying the conjugation plate with antibiotics (summarized in Table 3.1).

CRISPR/Cas9 gene editing has been used successfully to edit *Streptomyces spp.* genomes[218,220] and so a deletion of approximately 40 kb in the nystatin gene cluster was considered achievable. This deletion would be the first step in a large-scale rearrangement of the nystatin cluster, an achievement in its own right, and one which would allow a more efficient and extensive application of AE technology. Despite many attempts at optimising the conjugation of ET12567/pUZ8002 with *S. noursei* exconjugants were never observed, so this ambitious attempt at large-scale rearrangement was abandoned.

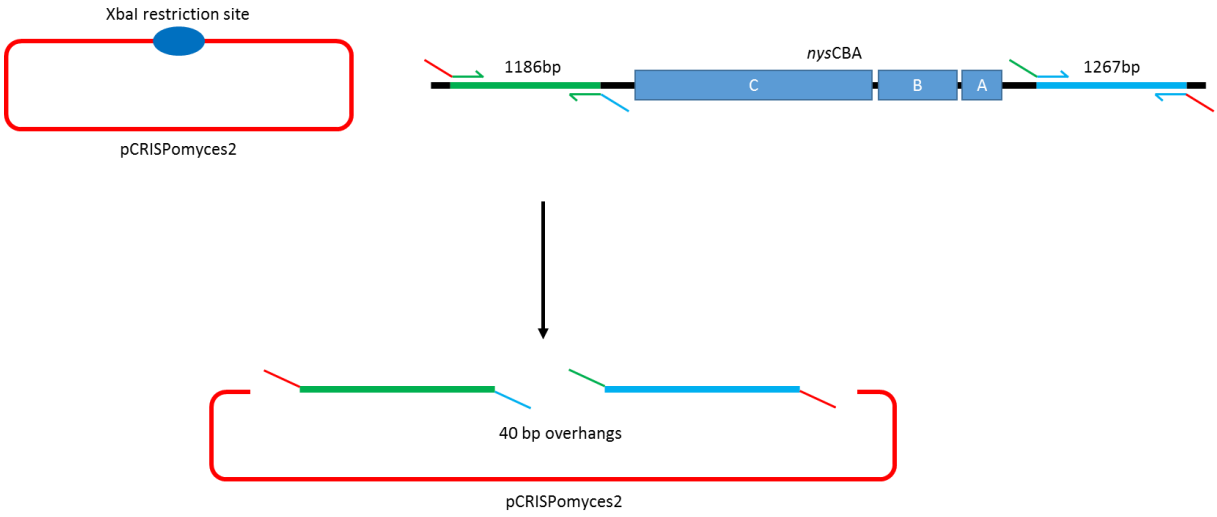


Figure 3.7: Scheme for three-fragment assembly of the *nys*CBA flanking arms into pCRISPomyces2 to form the template sequence. pCRISPomyces2 was linearised by digestion with the restriction enzyme *Xba*I. PCR primers had 40bp overhangs, forming complementary regions to allow isothermal assembly.

5′acgctccctttctttagaagaggcgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaa
cttgaaaaagtggcaccgagtcggtgctttttttagcataaccccttggggcctctaaacgggtcttgaggggttt
tttggctgctccttcggtcggacgtgcgtctacgggcaccttaccgcagccgtcggctgtgcgacacggacggat
cgggcgaactggccgatgctgggagaagcgcgctgctgtacggcgcgcaccgggtgcggagcccctcggcgagcg
gtgtgaaacttctgtgaatggcctgttcggttgctttttttatacggctgccagataaggcttgcagcatctggg
cggctaccgctatgatcggggcgttcctgcaattcttagtgcgagtatctgaaaggggatacgcactagtgattc
cccggtccggttt-3′

BbsI-Spacer 1-tracrRNA tail-T7 Terminator-gapdhp(EL)-Spacer 2-BbSI

Figure 3.8: Design of sgRNA. BbsI indicates the sticky ends, complementary to the sticky ends left by *Bbs*I digestion of pCRISPomyces2. Spacers 1 and 2 are the crRNAs that will guide Cas9 to the target sequences. The tracrRNA tail is the second part of the sgRNA and forms a scaffold to bind to Cas9. T7 is a terminator and gapdhp(EL) is a promoter. The missing promoter for Spacer 1 and the tracrRNA tail and terminator for Spacer 2 are found on the pCRISPomyces2 plasmid.

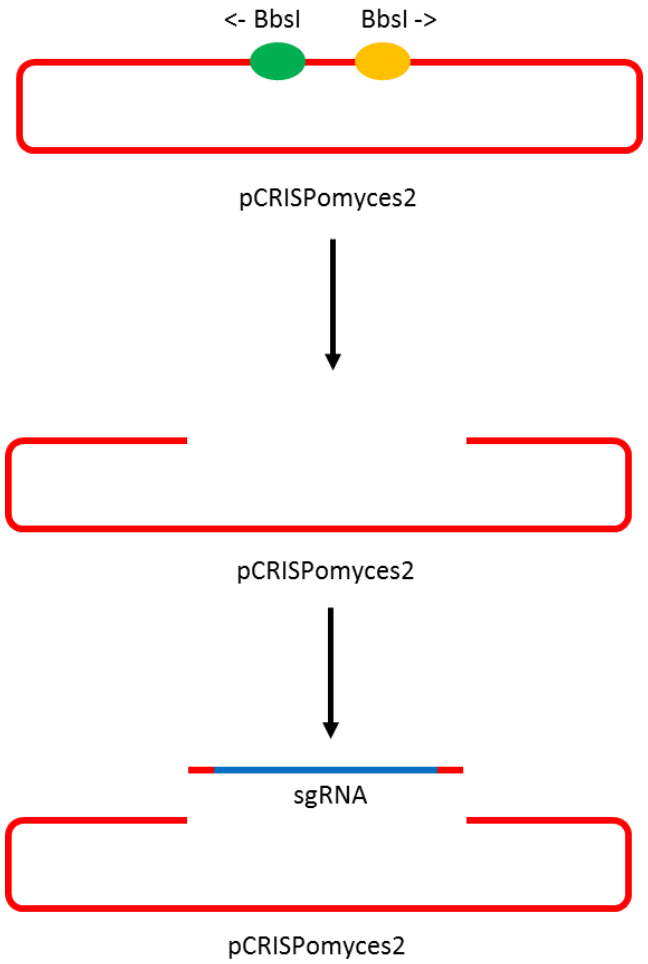


Figure 3.9: Golden gate assembly of sgRNA into pCRISPomyces2 linearised by digestion with the restriction enzyme *Bbs*I. *Bbs*I cuts outside its recognition sequence and so the arrows above indicate the direction in which *Bbs*I cuts, allowing for a seamless insertion of the sgRNA fragment.

Table 3.1: Conditions attempted for conjugation between ET12567/pUZ8002 and *S. noursei*.

Medium	$MgCl_2$ concentration	*S. noursei* cell type	Incubation time	Antibiotic concentrations
SFM	10mM	Mycelium	20hrs (Nal + Apr)	Nal: 25ng/µl
MAM	20mM	Spores	14hrs (Nal + Apr)	Apr: 50ng/µl
ABB13			3hrs (Nal),	
R6			14hrs (Apr)	

3.2.2: Nystatin Accelerated Evolution

Accelerated Evolution is a new technology. It is known that entry point DNA requires high homology to comparable sequences in the PKS gene cluster, and that in the case of rapamycin most new junctions were found within the KS and AT domains, and C-terminal of the ACP domain. Besides that, there are few guidelines for choosing a suitable site for the entry point DNA. In this case, the entry point DNA was chosen by aligning all modules against one another and choosing the region of highest mutual percentage identity to counterparts in all other modules. (see Figure 3.10). It was believed that this should maximise the chance of recombination with sites in neighbouring modules after initial integration at the entry point.

% Identity	NysA1	NysB1	NysB2	NysC1	NysC2	NysC3	NysC4	NysC5	NysC6	NysI1	NysI2	NysI3	NysI4	NysI5	NysI6	NysJ1	NysJ2	NysJ3	NysK1
NysA1	100	75.18	73.18	76.2	71.24	70.76	72.18	62.88	71.32	74.66	62.96	71.72	73.74	76.54	76.54	75.1	76.89	75.29	76.9
NysB1	75.18	100	74.39	76.01	73.24	72.76	74.03	64.13	72.92	75.66	64.13	78.1	74.41	76.59	76.59	77.04	77.17	75.41	77.74
NysB2	73.18	74.39	100	76.44	76.06	74.26	76.13	67.21	74.96	75.06	67.53	79.29	84	76.94	76.94	77.07	76.1	75.82	75.29
NysC1	76.2	76.01	76.44	100	75.53	74.11	75.93	64.21	73.71	76.71	64.85	74.52	74.57	78.89	78.89	78.99	76.62	78.33	78.53
NysC2	71.24	73.24	76.06	75.53	100	88.97	86.31	75.82	87.48	73.62	76.21	74.12	73.88	74.68	74.68	74.69	75.75	74.76	74.39
NysC3	70.76	72.76	74.26	74.11	88.97	100	86.31	77.78	95.93	73.14	78.01	73.02	72.94	74.52	74.52	73.43	74.96	74.92	73.6
NysC4	72.18	74.03	76.13	75.93	86.31	86.31	100	74.26	87.72	76.1	74.41	74.9	74.75	75.71	75.71	75.55	77.8	75	74.31
NysC5	62.88	64.13	67.21	64.21	75.82	77.78	74.26	100	77.31	65.15	96.87	66.59	67.53	64.84	64.84	65.49	66.59	67.37	66.98
NysC6	71.32	72.92	74.96	73.71	87.48	95.93	87.72	77.31	100	74.82	77.23	73.49	73.41	74.52	74.52	74.45	75.91	74.29	73.92
NysI1	74.66	75.66	75.06	76.71	73.62	73.14	76.1	65.15	74.82	100	65.71	74.88	76.16	76.54	76.54	76.87	76.18	76.58	76.3
NysI2	62.96	64.13	67.53	64.85	76.21	78.01	74.41	96.87	77.23	65.71	100	67.29	67.61	65.48	65.48	65.96	67.14	67.85	67.45
NysI3	71.72	78.1	79.29	74.52	74.12	73.02	74.9	66.59	73.49	74.88	67.29	100	79.53	75.5	75.5	75.97	77.83	74.23	75.83
NysI4	73.74	74.41	84	74.57	73.88	72.94	74.75	67.53	73.41	76.16	67.61	79.53	100	74.58	74.58	77.33	75.33	74.82	75.86
NysI5	76.54	76.59	76.94	78.89	74.68	74.52	75.71	64.84	74.52	76.54	65.48	75.5	74.58	100	100	79.68	77.41	78.28	78.16
NysI6	76.54	76.59	76.94	78.89	74.68	74.52	75.71	64.84	74.52	76.54	65.48	75.5	74.58	100	100	79.68	77.41	78.28	78.16
NysJ1	75.1	77.04	77.07	78.99	74.69	73.43	75.55	65.49	74.45	76.87	65.96	75.97	77.33	79.68	79.68	100	78.52	76.73	78.96
NysJ2	76.89	77.17	76.1	76.62	75.75	74.96	77.8	66.59	75.91	76.18	67.14	77.83	75.33	77.41	77.41	78.52	100	82.47	78.11
NysJ3	75.29	75.41	75.82	78.33	74.76	74.92	75	67.37	74.29	76.58	67.85	74.23	74.82	78.28	78.28	76.73	82.47	100	82.82
NysK1	76.9	77.74	75.29	78.53	74.39	73.6	74.31	66.98	73.92	76.3	67.45	75.83	75.86	78.16	78.16	78.96	78.11	82.82	100
Average	72.96	74.19444	75.37056	75.16889	76.74611	76.88556	77.06167	69.76944	77.10611	74.48222	70.12056	74.57278	74.72389	76.60333	76.60333	75.63944	76.01056	75.73611	75.73944

Figure 3.10: Sequence homology between KS domains in the nystatin PKS gene cluster. The highlighted column shows the sequence similarity between the KS in module 9 and all other KS domains. The KS domain in module 9 was chosen as the entry point for AE as it has the highest sequence similarity compared to all other KS domains.

The region with the highest identity contained most of the KS domain of module 6, therefore the whole KS6 domain (1383 bp) was chosen as the entry point DNA. The KS6 domain was amplified by PCR and purified using gel electrophoresis. This fragment was then inserted into pKC1139 which had been linearised by the restriction enzymes *Xba*I and *Eco*RI using Gibson isothermal assembly, to create the plasmid pKC1139-nysKS6. Figure 3.11 shows a general scheme for the creation of all pKC1139 derived vectors used in this work and Figure 3.12 shows the PCR amplification of the entry

point DNA and bands confirming the presence of the entry point DNA insert in pKC1139 generated by cPCR.

Transformation of pKC1139-nysKS6 into *S. noursei* was attempted through conjugation at 37°C with chemically competent ET12567/pUZ8002 cells which had previously been transformed with pKC1139-nysKS6. The same conditions as described in Table 3.1 were used, but once again failed to produce exconjugants despite attempts at optimisation using literature precedents as a guide [221]. The difficulty of conjugating into *S. noursei* could have been overcome through further attempts at optimisation, or though other methods of transformation such as electroporation or transduction. However, due to the difficulties presented, it was decided that rather than continue working with nystatin, focus would be shifted to other strains which are known to be easily conjugatable, and where the natural arrangement of the PKS genes favours AE.

Accordingly, three other modular PKS systems were evaluated for AE: filipin, erythromycin and monensin. Each of these systems have their PKS genes organized sequentially in the cluster and in the case of erythromycin and monensin they had been extensively characterised and effective conjugation protocols were well known.

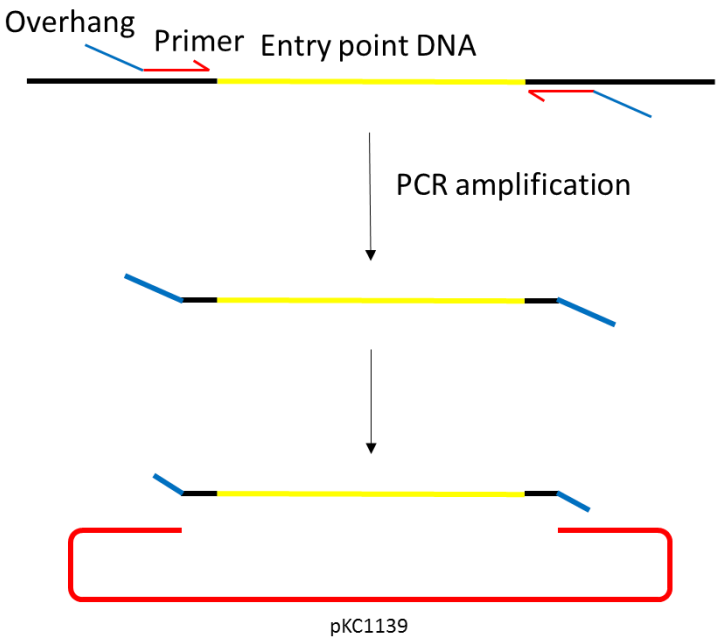


Figure 3.11: General scheme of amplification and ligation of entry point DNA into pKC1139. The primers include 20bp overhangs that are complementary to the ends of the linearised pKC1139. pKC1139 is linearised by *XbaI* and *EcoRI* digestion.

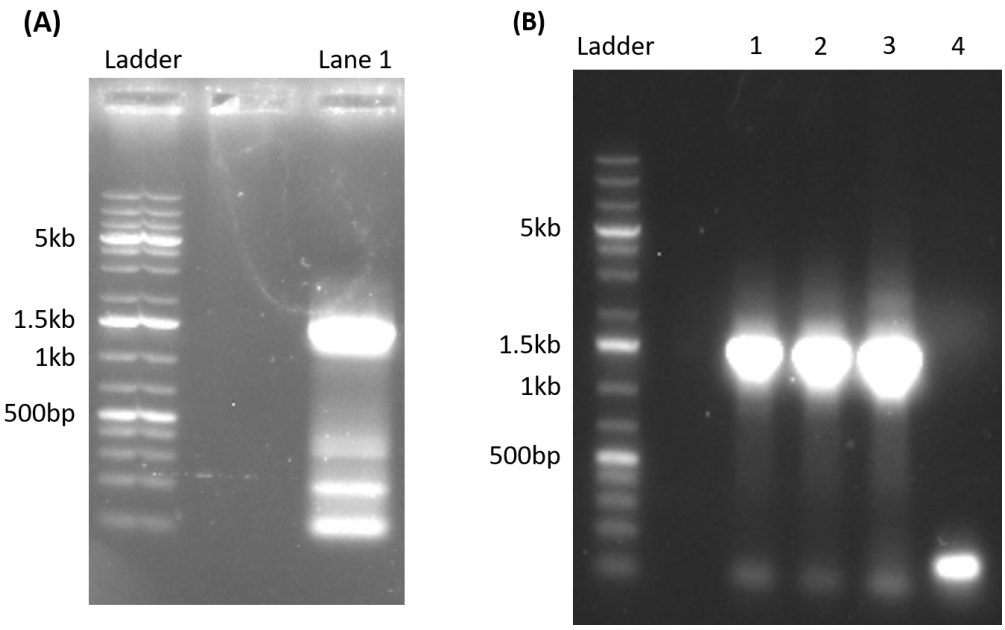


Figure 3.12: Gel images showing amplification of entry point DNA and screening of the pCK1139-nysKS6 vector. (A) Amplification of entry point DNA; **(B)** Screening of pCK1139-nysKS vector by cPCR. The 'Ladder' lane contains the DNA ladder (1 Kb Plus DNA Ladder). The numbers to the side of each ladder refer to the size of each relevant band in the ladder. Each numbered lane refers to a colony of E. coli that was screened for the correct vector.

3.2.3: Filipin, Erythromycin and Monensin Accelerated Evolution.

Erythromycin and monensin are well known polyketide compounds with useful therapeutic activity, and their biosynthesis has been well characterised over many years. Filipin is not useful as a therapeutic agent, but its biosynthesis, along with other macrolide polyenes, such as amphotericin and nystatin, is well characterised. They were chosen for several reasons, the first criterion being the ease with which the strains can be manipulated, but each gene cluster also presents characteristics that make it a plausible candidate system for AE. All three antibiotics are produced at easily detectable production levels, accompanied by a low number of background peaks. Filipin biosynthesis involves few tailoring enzymes, and for monensin a mutant strain of *S. cinnamonensis* was available in which the gene *monCI* was knocked out, preventing oxidative cyclisation of the linear polyketide backbone and instead producing the simple triene shunt product, premonensin [222]. . High quality genome sequences were already available for the erythromycin and monensin producers, and the filipin-producing strain, *S. filipinensis* had been acquired and was submitted for whole genome sequencing and assembly. Finally, in all three gene clusters the PKS genes were aligned in the same direction in the order in which their encoded enzymes act, maximising the scope for productive recombination.

Filipin had the additional advantage of a distinctive UV spectrum and an exotic hexylmalonyl-CoA extender unit, the synthesis of which could potentially be knocked out and used as a basis for conditional synthesis of filipin analogs. Figure 3.13 shows the polyketide biosynthetic pathway for the three chosen PKS systems and highlights the chosen entry points in the PKS genes and structures.

The initial entry point DNA for erythromycin and monensin AE was designed using the same method as described above for nystatin, and once again KS domains (KS6 for erythromycin and KS2 for monensin) were chosen as they held the highest sequence similarity between modules. Meanwhile the *S. filipinensis* genome was sequenced, assembled and annotated by the Department of Biochemistry DNA Sequencing Facility staff. Once the assembly was completed the entry point DNA, also a KS domain (KS11), was chosen for filipin. KS11 was chosen on the basis of its high sequence similarity with the KS domain of other modules, and because module 11 is involved in biosynthesis of part of the polyol region of filipin. The reasoning was that if contractions were generated, they would be more likely to leave the characteristic UV signature of filipin intact, which would help detection of novel filipin analogues. The entry point DNA for all three clusters was amplified by PCR and purified using gel electrophoresis. The fragments were then inserted into pKC1139, which had been linearised by the restriction enzymes *Xba*I and *Eco*RI, through Gibson isothermal assembly to create the plasmids pKC1139-eryKS6, pKC1139-preKS2 and pKC1139-filKS11 for the erythromycin cluster, monensin and filipin gene clusters, respectively (Figure 3.11). Successful assemblies were screened for using cPCR. Constructs were used to transform the *S. erythraea*, *S. cinnamonensis*, and *S. filipinensis* hosts was attempted through intergeneric conjugation at 37°C with chemically competent ET12567/pUZ8002 cells which had previously been transformed with each construct. The initial conjugation attempts failed to produce exconjugants. Standard conjugation protocols were known to be effective in these strains [127,223,224] and so there must have been another issue, such as the vector, the entry point DNA, or the modifications to the conjugation protocol. The AE protocol requires conjugations to occur at 37°C. Exconjugants are restreaked to ensure no contamination is carried over, and once confirmed as clean they are analysed by PCR to confirm the presence of the vector in the host genome. Only at this point are the cells grown at 30°C, which was described as vital to the success of AE in the original experimental protocol. The importance of the initial higher temperature is due to the pSG5 origin of the pKC1129 vector. This origin of replication is non-permissive at temperatures higher than 34°C, therefore the plasmid cannot replicate and must integrate into the genome to survive. The pKC1139 vector is commonly used as a suicide vector for transformation of *Streptomyces* spp. [225]. Why the temperature control is so vital to success is currently unknown but if the pSG5 origin were responsible for this effect one might expect it to have been observed before now.

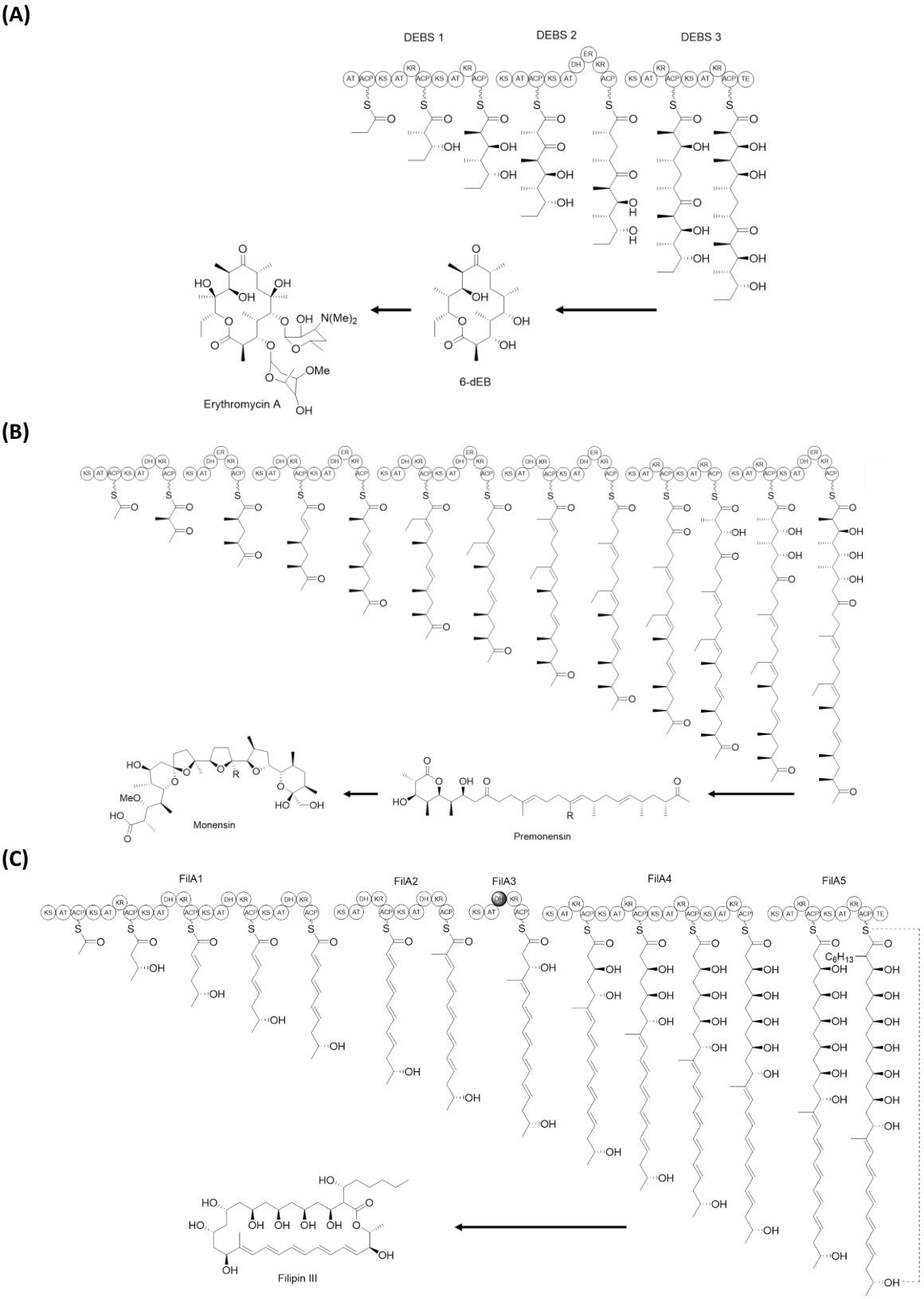


Figure 3.13: PKS biosynthetic pathways of erythromycin, monensin, and filipin. The PKS biosynthetic pathways, intermediates and completed structures for; **(A)** erythromycin **(B)** monensin (the simple shunt compound "premonensin" was the target for the AE work) **(C)** filipin.

One hypothesis is that lowering the temperature initiates replication from the pSG5 origin within the host genome. This would generate single stranded DNA, which is known to be highly recombinogenic [226] and could start strand invasion and recombination at neighbouring regions of homology, creating a recombination event between the entry point and the site of secondary integration, that truncates the PKS-encoding region between the two sites. An alternative hypothesis instead suggests might be that the temperature sensitive origin is just an efficient means to transform DNA into *Streptomyces* spp., and that the inter-module recombination is the secondary effect of the induced stress of repairing double stranded breaks, or other DNA damage caused by the integration of the pKC1139 vector.

The modified protocol appeared to interfere with the growth of exconjugants. A larger-scale conjugation was attempted to overcome the reduced probability of generating exconjugants. This work resulted in exconjugants which grew when patched onto selective media. Consultation with our collaborators at Isomerase Therapeutics at this point confirmed that this approach did reduce the likelihood of generating exconjugants, and more importantly the 'hot-spots' within rapamycin PKS modules that were discussed in the introduction of this chapter had been identified, and their importance in generating mutants was highlighted. This indicated the need to evaluate several constructs for each system, covering an entire module, and to carry out conjugations on a larger scale.

The AE experiments were therefore repeated for a new series of constructs. Entry point DNA for the new constructs was designed using the same module within each of the three gene clusters as in the initial experiments with the KS region only. For each chosen module, four constructs were designed to cover the entire module from the ACP domain in the previous module, to the KR domain in the chosen module. Figure 3.14 shows the sequence similarity across each module in the respective PKS gene clusters and Figure 3.15 shows the region each construct covers within the chosen module. The sequence similarity plots show that across the three PKS systems, the regions of highest sequence similarity consistently lie within the KS domains, although additional regions of high similarity can be found in the AT domains, and in interdomain regions, such as the sequence C-terminal of the ACP domain.

The entry point DNA for each construct was amplified by PCR and purified using gel electrophoresis. The fragments were then inserted into pKC1139, which had been pre-linearised by the restriction enzymes *Xba*I and *Eco*RI, by Gibson isothermal assembly, to create each construct (Table 2.2). Transformation of each construct was attempted via conjugation between ET12567/pUZ8002 and the host strains at 37°C. Figure 3.16 shows the PCR amplification of the entry point DNA and bands confirming the presence of each entry point DNA insert in pKC1139 vectors for the filipin entry point constructs.

Table 3.2: pKC1139-based plasmids for AE recombineering

Plasmid	Entry point DNA	Length (bp)
pKC1139-filKS11	Filipin domain KS11	1284
pKC1139-filKS-AT11	Filipin domains KS-AT11	2056
pKC1139-filAT-KR11	Filipin domains AT-KR11	2023
pKC1139-filACP-KS11	Filipin domains ACP10-KS11	1945
pKC1139-eryKS6	Erythromycin domain KS6	1278
pKC1139-eryKS-AT6	Erythromycin domains KS-AT6	2031
pKC1139-eryAT-KR6	Erythromycin domains AT-KS6	2062
pKC1139-pryACP-KS6	Erythromycin domains ACP5-KS6	1560
pKC1139-preKS2	Premonensin domain KS2	1278
pKC1139-preKS-AT2	Premonensin domains KS-AT2	1890
pKC1139-preACP-KS2	Premonensin domains ACP1-KS2	1727
pKC1139-preER-KR-ACP2	Premonensin domains ER-KR-ACP2	2475

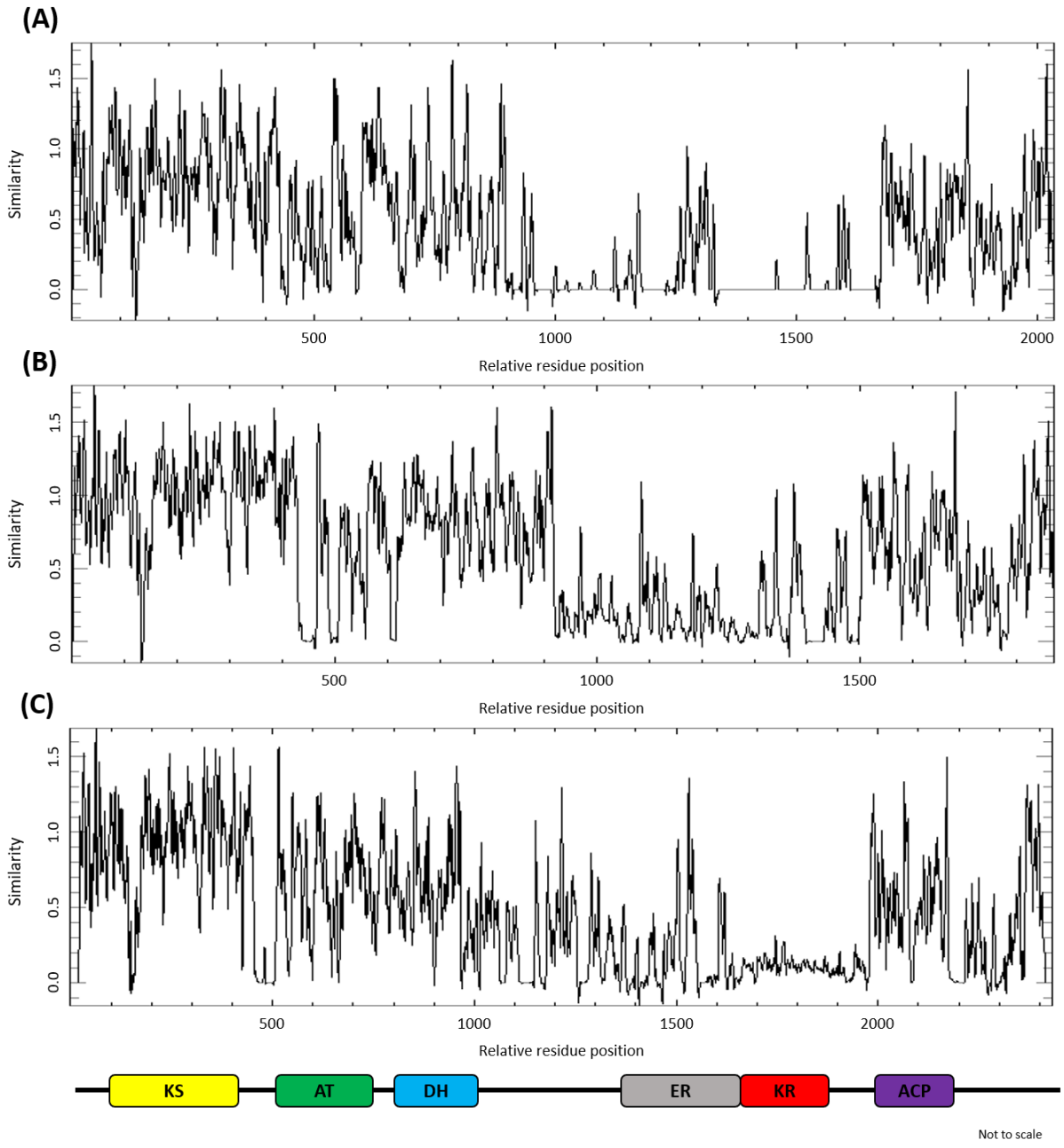


Figure 3.14: Sequence similarity between each module in the respective PKS gene clusters. (A) Erythromycin **(B)** Monensin **(C)** Filipin. The charts were created by aligning all module sequences from each PKS system and plotting sequence similarity using Plotcon (http://emboss.bioinformatics.nl/cgi-bin/emboss/plotcon). The schematic of a PKS module is not to scale but indicates regions of the module with the highest sequence similarity.

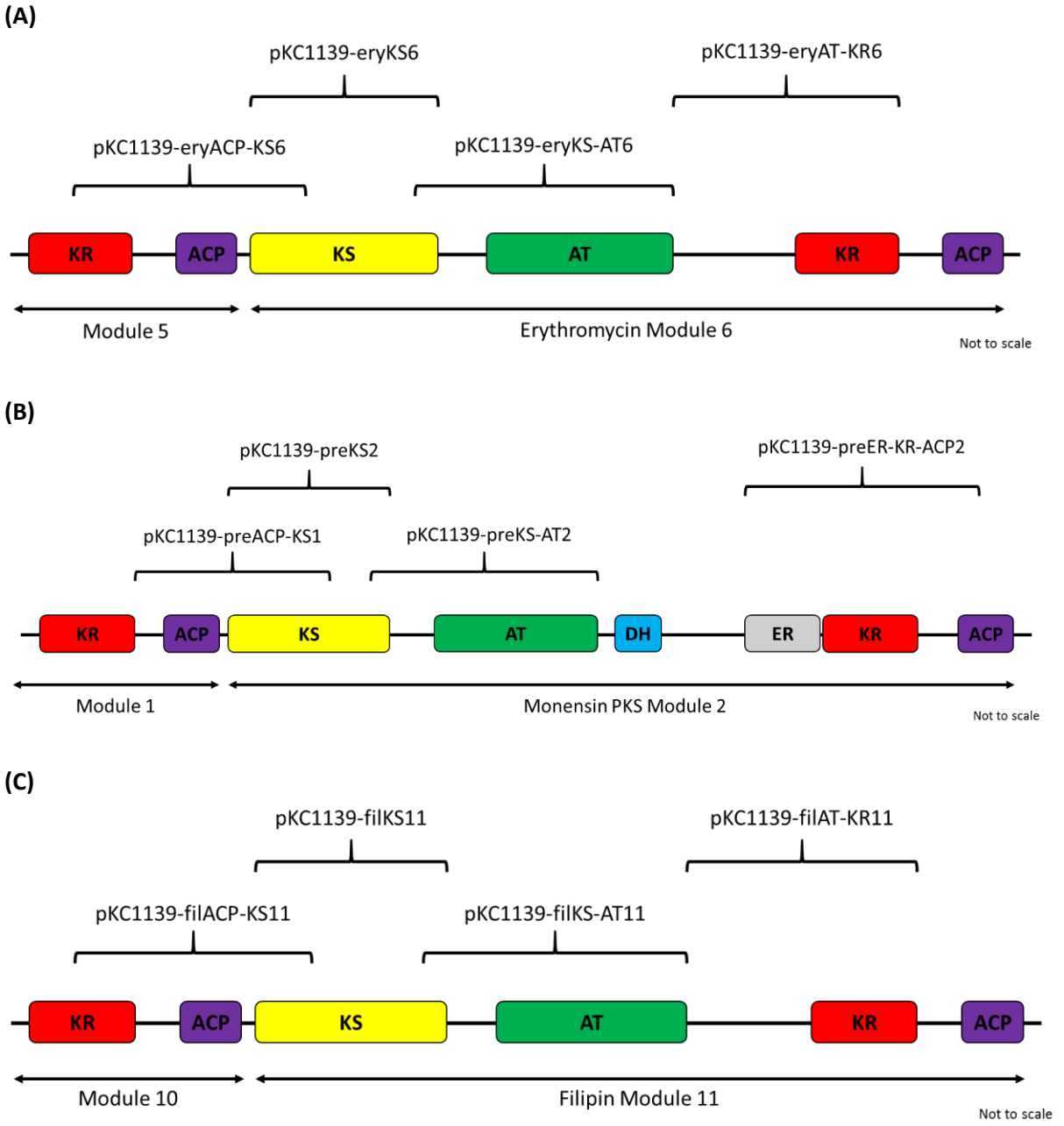


Figure 3.15: Entry point DNA constructs covering the length of the chosen modules in each gene cluster. (A) Erythromycin PKS module 6 **(B)** Monensin PKS module 2 **(C)** Filipin PKS module 11.

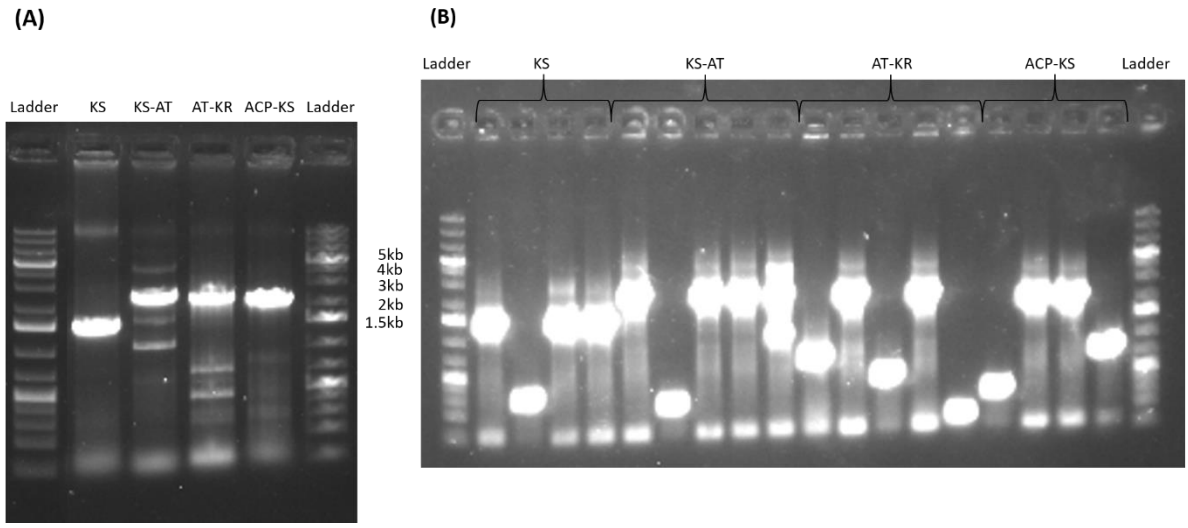


Figure 3.16: Gel images showing amplification of entry point DNA and screening of the pCK1139-derived vectors for the filipin constructs. (A) Amplification of entry point DNA; **(B)** Screening of pCK1139 filipin vectors. The 'Ladder' lane contains the DNA ladder (1 Kb Plus DNA Ladder). The numbers to the side of each ladder refer to the size of each relevant band in the ladder. Each set of 4 lanes in (B) are from a colony of E. coli that was screened for the correct expression vector.

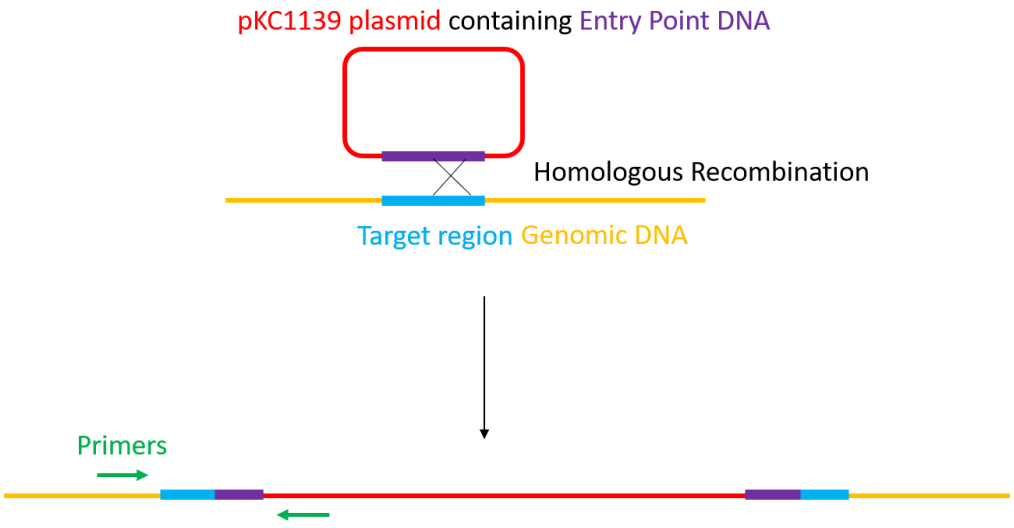


Figure 3.17: Scheme for cPCR screening of exconjugants for primaries. A successful primary will have the pKC1139 plasmid integrated within the target region of the PKS module. To screen for this pairs of primers were designed that would amplify a sequence of DNA across the junction of the integration point. As long as one primer is on the plasmid and one is on upstream of the target region, a DNA fragment should be amplified regardless of the exact point of integration.

(A)

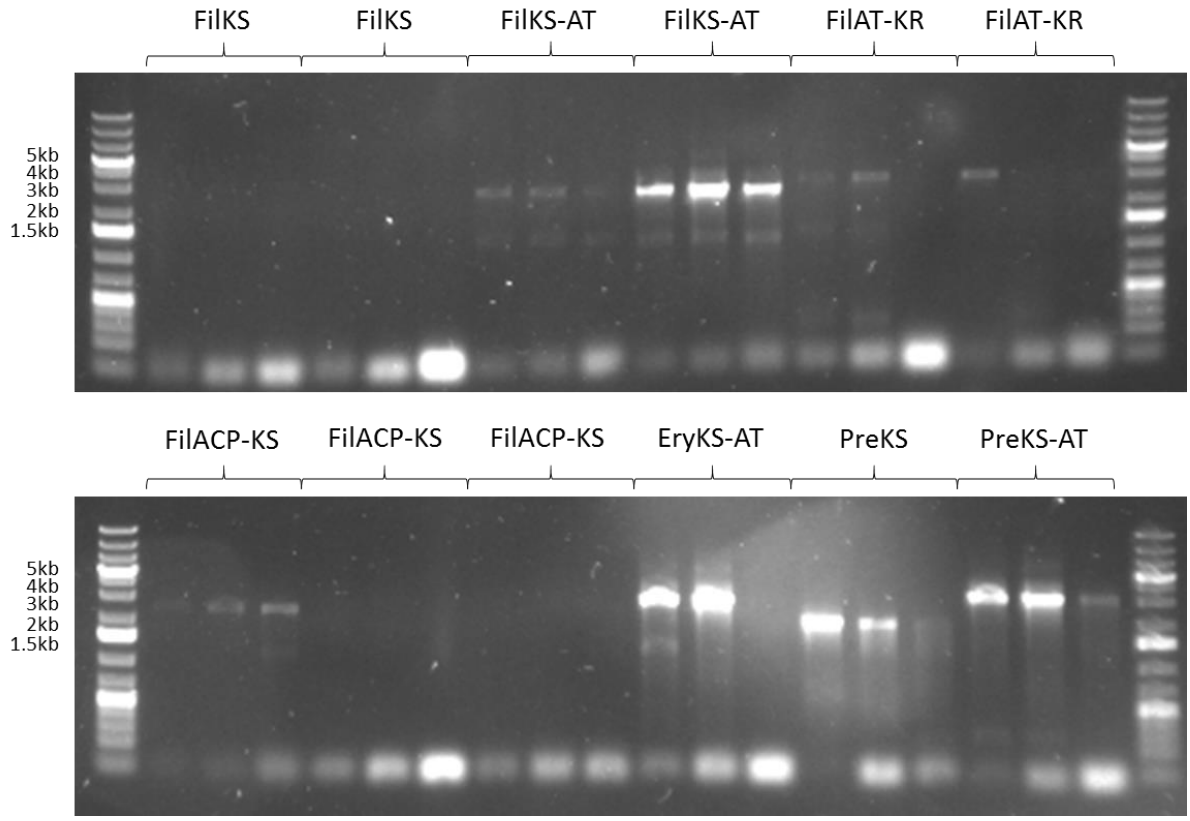


Figure 3.18: Confirmation of primary exconjugants by colony PCR. Each set of three bands show PCR amplification of the junction regions between inserted vectors and genomic DNA. Each band observed is of the expected size, however many lanes do not any bands. This is an intrinsic issue with *Streptomyces* colony PCR. It is unknown whether lanes are missing bands due to a failure of the PCR, or due to the vector not being present in the host genome.

This attempt produced several exconjugants for each strain. These were restreaked onto SFM plates containing nalidixic acid and apramycin. The exconjugant was cleared of contamination by several rounds of restreaking on fresh SFM plates containing antibiotics. The presence of the pKC1139 derived construct in the cells was confirmed using *Streptomyces* cPCR and sequencing. Figure 3.17 shows the scheme for screening of primary exconjugants and Figure 3.18 shows the results of the screening.

The effectiveness of colony PCR of sporulating *Streptomyces* colonies was investigated for this screening process. A small amount of mycelium was picked, the cells were lysed by boiling in a 50% DMSO solution, and the resulting mixture was used as the template for a PCR reaction. While initially promising, cPCR required several rounds of optimisation before it could produce products observable by gel electrophoresis in the majority of samples. As can be seen in Figure 3.18, the observable bands

are clear and of the expected size. However further optimisation is required before *Streptomyces* cPCR can reliably detect successful integration events.

The screening revealed only one erythromycin and two monensin primary exconjugants. In contrast, every one of the filipin conjugations produced colonies resistance to apramycin, and of those all but the FilKS-AT11 construct were confirmed as primaries. Once exconjugants were confirmed, the selective pressure of an antibiotic was removed to induce a secondary recombination event and removed the integrated plasmid. This is referred to a relaxing and was achieved by inoculating colonies into liquid TSBY medium without antibiotics and incubating the cultures at 30°C. Each culture was re-inoculated after 48 hrs incubation, three times, then finally the cultures were diluted and streaked out to single colonies on an SFM plate without antibiotics. The loss of antibiotic resistance was confirmed by patching single colonies onto pairs of SFM plates, one without apramycin and one with apramycin. The only colonies that lost their resistance to apramycin were derived from *S. filipinensis*. It is unclear whether the relative success of the filipin constructs compared to those of erythromycin and monensin is due to chance, or some property of *S. filipinensis* which lends itself to success with AE.

In addition to the loss of apramycin resistance, conferred by the integrated pKC1139 vector, putative filipin mutants were screened by PCR. The reasoning behind the PCR screen is outlined in Figure 3.19 and the results of the screen are presented in Figure 3.20. This screen is not ideal as it relies on the disappearance of a band to identify recombinant DNA. However, considering the number of potential recombinants to be screened, this method provided a rapid way to prioritise strains for fermentation.

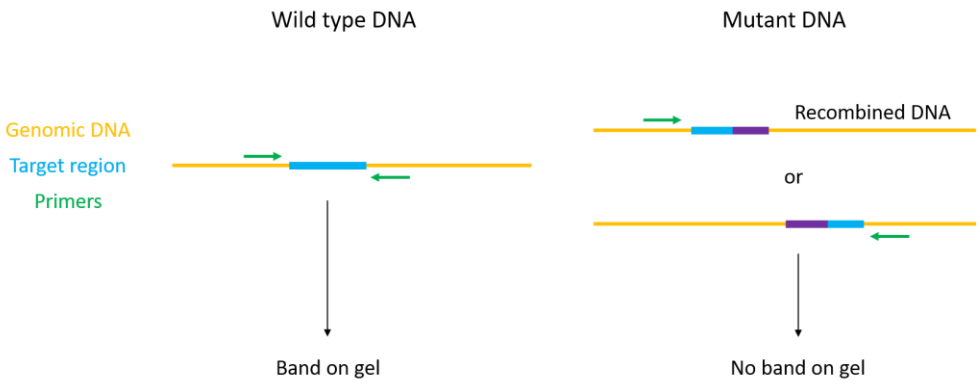


Figure 3.19: Scheme for relaxant screening of putative filipin recombinant strains. By using primer pairs that amplify across the target region, a wild type strain can be identified. A recombinant strain would not give a product. This method is obviously not a reliable way to confirm a recombinant strain but provided a starting point for further confirmation.

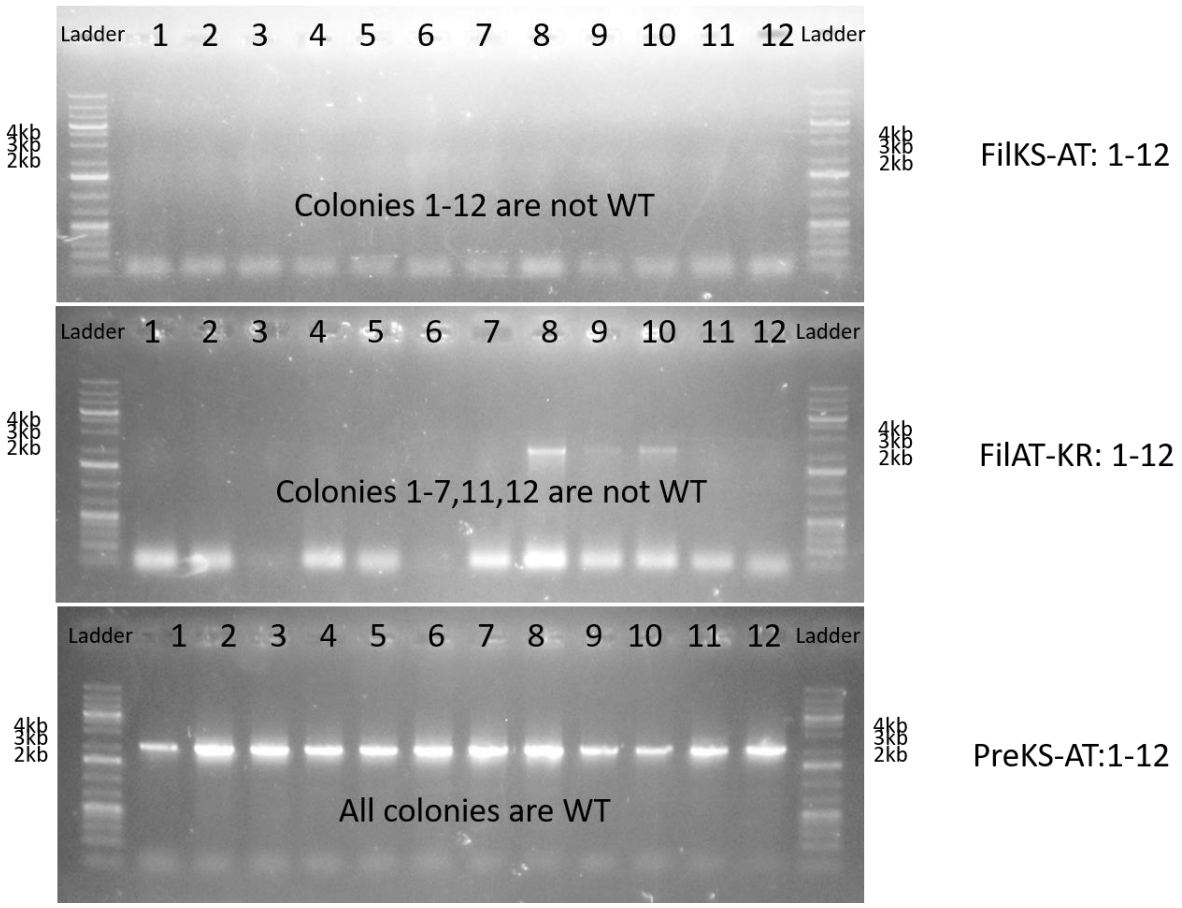


Figure 3.20: Gel images showing the screening for filipin recombinant strains. This screening approach identifies all 12 samples of the FilKS-AT relaxant strains and 9 of the FilAT-KR relaxant strains as being putative recombinants. In contrast the screening of the PreKS-AT relaxant strains identifies all 12 samples as being wild type.

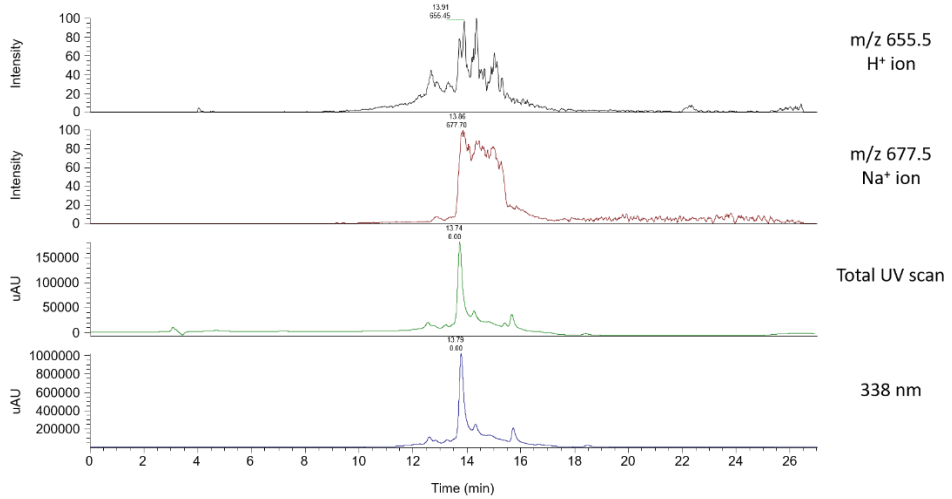
The first batch of experiments resulted in 53 putative recombinant strains. These were grown following the protocol outlined in Chapter 2. After 5 to 7 days of growth the fermentation culture was sampled by extraction of 1 ml of culture with 2 ml of ethyl acetate. The ethyl acetate was evaporated off, and the remaining residue was dissolved in methanol. The samples were then analysed by LCMS, following the procedure in Chapter 2. All but one sample appeared to produce wildtype filipin. The remaining sample (sample 2.8.2) gave an LCMS spectrum that showed a profile with the characteristics of a filipin-like UV signature, a non-filipin mass, and a shorter retention time than filipin. Figure 3.21 illustrates the differences between wildtype *S. filipinensis*, a characteristic sample of a putative recombinants, and sample 2.8.2. The shorter retention time of sample 2.8.2 is indicative of a filipin-related analogue, perhaps generated by a contraction in the number of modules, and the reduction in yield is commonly seen in recombinant strains. To test whether this might be a novel filipin

derivative I reinoculated the strain from the plate it had been stored on, to repeat the fermentation at a larger volume and confirm the result.

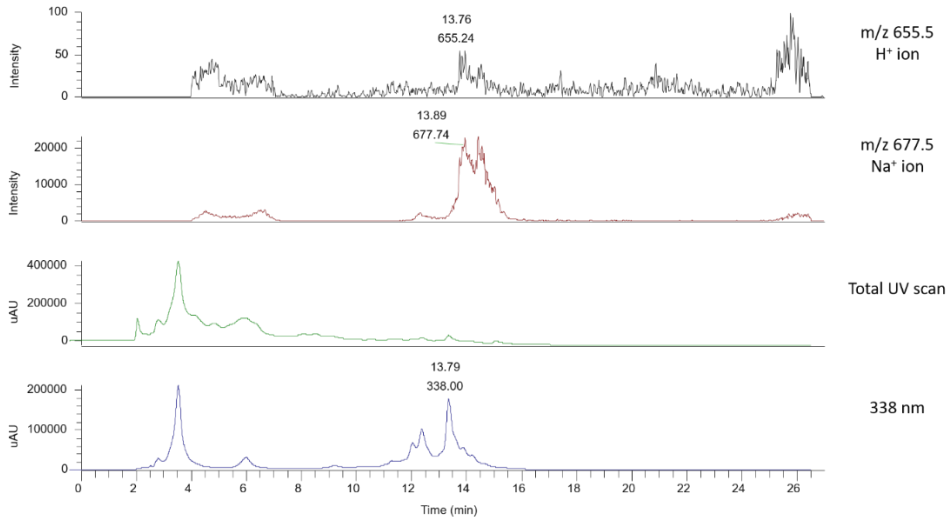
Unfortunately, despite multiple attempts using liquid media and on SFM agar the strain did not grow. The reasons for this are unknown and might be wholly unconnected with filipin biosynthesis. It did however raise the possibility that an altered filipin analogue might be toxic to the cells. The reduction in the yield I observed in sample 2.8.2 was also a serious concern. To address this a production screen was set up using 13 different media. All media were inoculated with 0.5 ml of the same seed culture and incubated at 30°C for 5 days. There was an approximately 200-fold difference in the production level of filipin between the lowest and highest producing media, with SFM being the highest yielding medium. Ethyl acetate extractions of SFM cultures had a strong yellow colour, indicative of a high concentration of filipin. The differences are shown in Figure 3.22. Once SFM had been identified as the best medium for producing filipin it was used for all further fermentations of *S. filipinensis*.

Further work focussed on repeating the attempts at AE of *S. filipinensis* and on maximising the output of potential recombinant strains and trying to shorten the time required for analysis. Approximately five to ten exconjugants were generated for each of the filipin constructs and these were relaxed to generate hundreds of putative recombinants. To maximise the number of putative recombinants that could be screened for loss of apramycin resistance, 96 deep-well plates (2ml in each well) were adopted for the relaxation stage. These plates contained 1 ml of SFM and a single glass bead to prevent clumping. The wells were inoculated from a seed culture from a confirmed primary exconjugant, and after several rounds of reinoculation into fresh non-selective media the cultures were diluted and spread out to obtain single colonies. These in turn were screened for loss of apramycin resistance. Primary exconjugants were screened for loss of production of filipin, resistance to apramycin and were also screened by cPCR. Loss of production of filipin is expected if the target region has been interrupted by the integrated vector. Figure 3.23 illustrates the loss of production in a primary exconjugant.

(A)



(B)



(C)

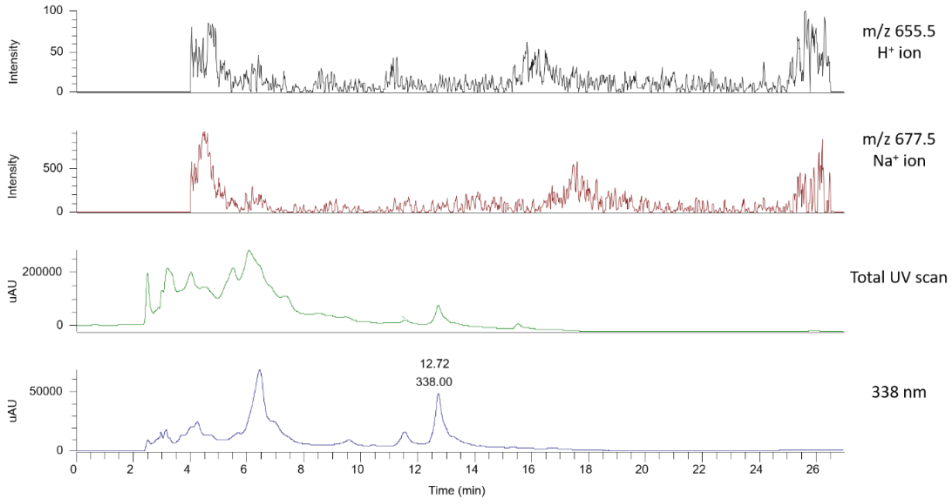


Figure 3.21: LCMS analysis of putative recombinant strains. (A) Wildtype *S. filipinensis* production **(B)** *S. filipinensis* sample 2.4.4 **(C)** *S. filipinensis* sample 2.8.2.

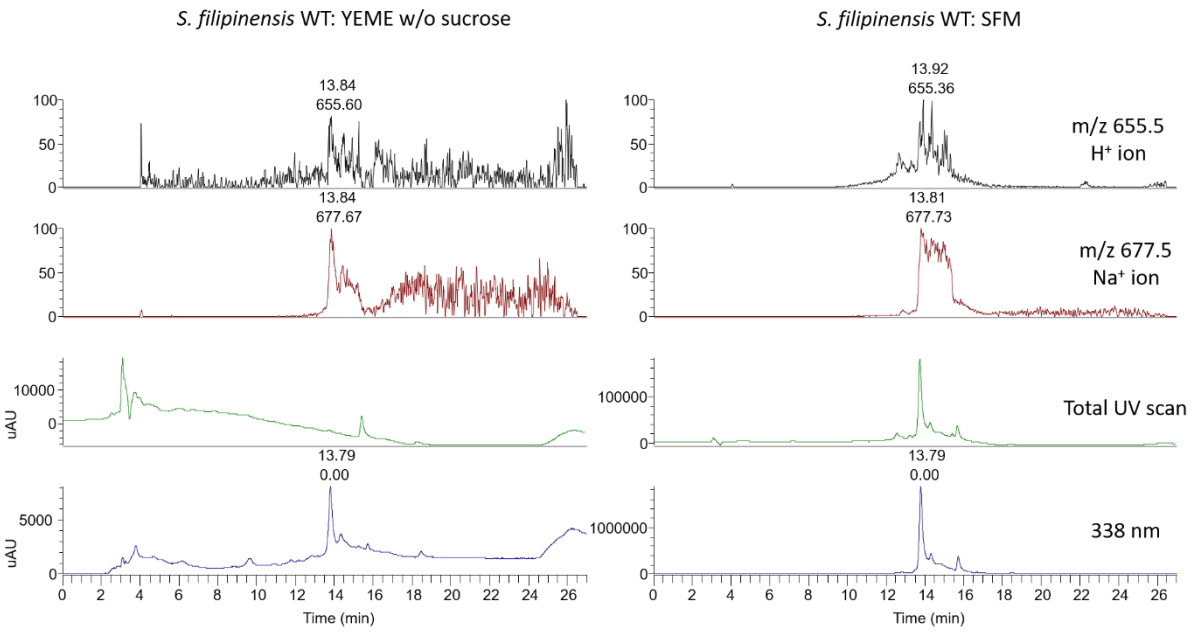


Figure 3.22: LCMS analysis of the filipin production screen. Based on the relative intensity of the UV signature of filipin SFM media allows an approximate 200-fold increase in yield of filipin compared to YEME without sucrose.

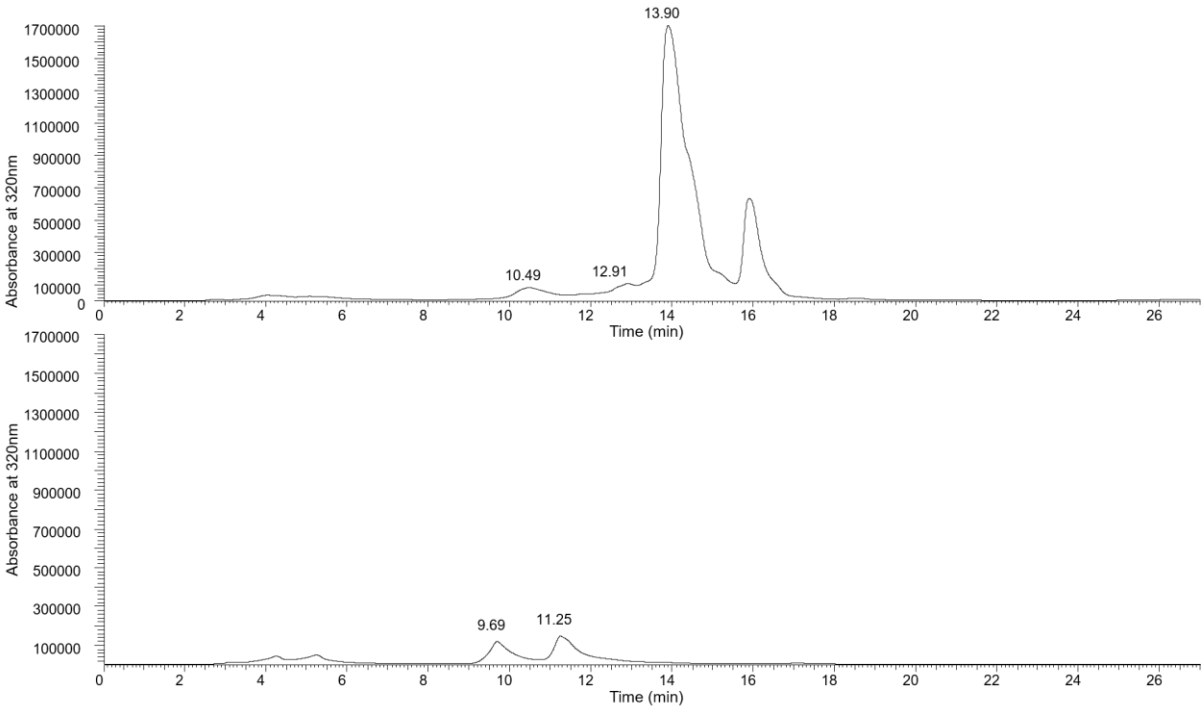


Figure 3.23: LCMS analysis of the loss of filipin production in a primary. The characteristic peak at 13.9 mins has been lost, indicating a loss of filipin production. The peak at 16 mins was always observed with filipin production and is most likely a filipin related compound.

Unfortunately, despite the new screening protocols, and the large number of potential recombinant strains being fermented and analysed by LCMS no other potential filipin derivatives were observed. Perhaps one reason for this was the lack of processing capacity in the lab, which was not really set up for such high throughput screening. The total number of potential recombinant strains that were generated from the last batch of primaries was approximately 900. Of those only 128 were analysed by LCMS. Except for 14 samples, all 128 appeared to produce filipin, matching the molecular weight and retention time of wildtype filipin. The remaining 14 samples appeared to have lost filipin production, without gaining the ability to make an altered product. This effect has also been observed with AE experiments on the antibiotic tylosin [201].

3.3: Concluding Remarks

The work described in this chapter was an attempt to widen the application of AE in *Streptomyces* spp. It began with the optimistic target of applying AE to the nystatin, erythromycin, premonensin, and filipin PKS systems, however the scale of that original goal, and the issues found with conjugation into the nystatin and to a lesser extent the erythromycin and premonensin producing strains led to a focus of the work on the filipin PKS system. Despite this focus, and the large number of putative recombinants generated and screened, no new filipin derivatives were characterised.

In future work, it would be highly advantageous to mutate *S. filipinensis* to produce filipin only when supplemented. One possible route to this might be through blocking the strain's ability to produce the hexylmalonyl-CoA building block required for the last chain-extension step on the PKS. *FilB* has been identified as a 2-octenoyl-CoA reductase that supplies the hexylmalonyl-CoA for filipin biosynthesis [227]. Knocking out this gene would provide the basis for conditional synthesis, through feeding a cell-permeable precursor of hexylmalonyl-CoA to the strain. Having an *S. filipinensis* strain with conditional filipin synthesis would address concerns about generating toxic metabolites and would greatly ease analysis of any recombinant strains.

A significant barrier to the wider application of AE is that the underlying mechanism behind its generation of derivatives is not understood. It is possible that there is a specific requirement for AE to function correctly that happens to be found in the rapamycin producing strain, and not in others. The fact that this effect has only been recorded once, other than the paper by Wlodek *et al*. suggests that AE is possible in other strains but is extremely rare. This would also explain why a possible filipin derivative was observed in sample 2.8.2 but was never repeated. With the benefit of hindsight, it was

perhaps premature to begin trying to generalise the technology. The success of AE with rapamycin is undeniable, but I believe in the absence of a better understanding of the mechanism, it relied too heavily on luck to achieve an effect that had only been observed in a very small number of polyketide producing strains. The study of homologous recombination in *Streptomyces* is decades behind research focussing on *E. coli* [228,229] which makes it difficult to identify the similarities or differences required to make a strain amenable to AE. One way to explore this initially would be to continue AE on *S. rapamycinicus*, aimed at another of the metabolites of that strain, such as the long chain linear polyene clethramycin [230]. It would be of great interest to know whether AE on a different metabolite was equally productive.

AE is a fascinating technology that if generalised has the potential to bring about a paradigm shift in the field of polyketide biosynthesis. The ability to rapidly generate a library of derivatives from a single experiment could lead to therapeutic agents with novel activities, improved efficacies, or agents that are able to overcome antibacterial resistance, especially if the generation of mutants could be linked to a high-throughput selection or screen.

Chapter 4: Genome-level analysis of the pseudouridimycin producing strain *Streptomyces albus* DSM40763

4.1 Introduction

As discussed in the introduction, advances in DNA sequencing and genome annotation have provided huge benefits to the study of bacterial natural products. *Streptomyces* and allied genera have exceptionally large genomes (8-12 Mbp) and their high G+C % content and frequent regions of repetitive DNA pose significant technical problems for sequence analysis. This is reflected in the fact that the vast majority of publicly deposited draft genomes are highly fragmented and poorly interpretable. However, high-quality, well-annotated, genome sequences are afforded by the in-house Department of Biochemistry Sequencing Facility and the customised assembly and automated annotation pipeline developed by Dr Markiyan Samborskyy. Together with the cluster-finding software AntiSMASH this provides an accurate estimate of the biosynthetic potential of a given strain and allows recovery of the full sequence of even the largest gene clusters. Their work has confirmed that on average a *Streptomyces* strain houses 20-25 clusters, although an exceptional strain such as *Streptomyces malaysiensis* DSM 4137 has 53 annotated clusters (Samborskyy *et al.*, *ms* in preparation). Typically, most of these clusters do not give rise to detectable products under standard laboratory growth conditions. This genomics-based approach is now the method of choice for helping identify species and assessing the biosynthetic potential of a strain. However, it is only the first step in correctly matching a gene cluster to its product.

Pseudouridimycin (PUM) (Figure 4.1) is a nucleoside antibiotic first isolated under the name strepturidin in 1993. It was identified through the antibacterial activity of an extract of *Streptomyces albus* culture against the Gram-negative nitrogen-fixing soil bacterium *Azotobacter chroococcum* [231], and tests proved positive for guanidine groups, sugars and amino sugars, and negative for aromatic amines, organic acids, amino acids and peptides. A proposal for its structure was made in 2014 [232], together with a tentative proposal for its biosynthetic pathway. The structure was initially characterised using HPLC-HR-ESI-(+)-Orbitrap-MS and IR spectroscopy. A molecular formula of $C_{17}H_{27}O_9N_8$ was identified by high-resolution mass spectroscopy, and IR spectroscopy revealed the presence of amine groups, hydroxyl groups, and amide carbonyls. UV spectroscopy showed characteristic absorptions for heterocycles. These insights were combined with the results of extensive ^{1}H and ^{1}H-^{13}C 2D-NMR spectroscopy, which identified four distinct spin systems, to complete

the structural elucidation. The first spin system was identified as N_{α}-hydroxylated glutamine, the second was a triol-containing carbon chain, supposedly derived from diamino aldopentose, the third spin system was identified as 2-aminoimidazole-5-one, and the final spin system was identified as uracil. Signals indicative of the connection between these spin systems were also observed, identifying the structure seen in Figure 4.1 as strepturidin.

PUM was also independently reported by Ebright, Donadio, and colleagues in 2017 [233] through screening a library of 3,000 actinobacterial and fungal culture extracts for inhibitory activity against *E. coli* RNA polymerase (RNAP) [234]. PUM was isolated from two culture extracts with antibacterial activity, and it was shown to selectively inhibit bacterial RNAP in both Gram-negative and Gram-positive bacterial strains, and against both drug-sensitive and drug-resistant strains. It also showed no cross-resistance with the potent, and broad-spectrum bacterial RNAP inhibitor rifampicin, exhibiting additive antibacterial activity when co-administered with rifampicin (a constituent of the standard DOTS treatment for TB). This suggests that PUM acts through a different mechanism to rifampicin, which makes it an exciting target for study, and a possible drug lead.

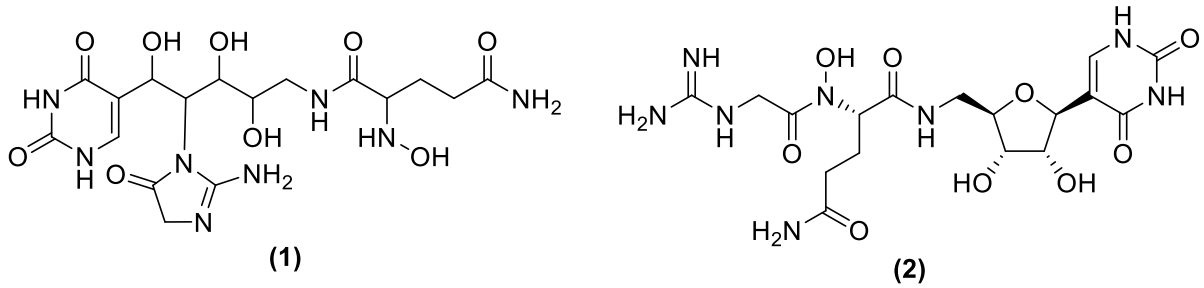


Figure 4.1: Structures of (1) strepturidin and (2) pseudouridimycin. Deoxy-pseudouridimycin refers to a loss of the *N*-hydroxyl group.

The PUM structure shown in Figure 4.1 was independently established [233] by high-field NMR but additional ^{15}N-HSQC spectroscopy indicated the uracil moiety was C-linked to C1' of 6'-amino-ribose, which HMBC spectra indicated was linked in turn to the glutaminyl moiety at the N6' of 6'-amino-ribose. Further data indicated the presence of glycine C-linked to the glutamine Nα, and of formamidine C-linked to the glycine Nα. The structure was completed when the absence of a glutaminyl Nα signal in the N-HSQC spectrum suggested the presence of a hydroxamic acid, which was confirmed by reduction of PUM with $TiCl_3$, yielding deoxy-PUM (Figure 4.1). The stereochemistry of PUM shown in Figure 4.1 was identified through hydrolysis and a comparison of the products with

commercial standards using GC-MS. The work to characterise PUM structure does not acknowledge the earlier report, but it seems likely that strepturidin is actually mischaracterised PUM.

The fully elucidated structure of PUM shows several interesting features. Most strikingly it is a C-nucleoside, meaning that the uracil base is connected to the sugar through a C-C rather than a N-C bond. Dr Hui Hong in this laboratory has recently analysed the biosynthetic pathway to the C-nucleoside malayamycin A and has shown that it is derived from pseudouridine 5'-monophosphate (Ψ-MP) (Hong et al. (2018) *ms*. in the press). The other interesting feature is its modular nature, apparently the result of a convergent pathway in which 6'-amino-pseudouridine, N-hydroxylated glutamine, and guanidinoacetate moieties are then linked together. The work in this Chapter describes a genome mining approach to uncovering the PUM gene cluster, with a view to examining the biochemistry of the pathway compared to that of malayamycin and exploring the potential for preparing analogues of the PUM molecule. At the same time, the genome mining approach would allow a detailed overview of the unexplored biosynthetic potential of this strain.

4.2 Results and Discussion

4.2.1 Whole-genome sequencing of *Streptomyces albus* DSM40763

Streptomyces albus DSM40763 was obtained from the DSMZ collection. Whole-genome sequence was obtained in-house by Shilo Dickens and colleagues at the NextGen DNA Sequencing Facility in the Department of Biochemistry using a combination of shotgun MiSeq and long-range mate-pair MiSeq data. Sequence data was assembled into scaffolds, using an in-house pipeline, by Dr. Markiyan Samborskyy. Open reading frames were identified and visualised by the genome browser Artemis [235].

Streptomyces albus DSM40763 is coded as A980 in the Leadlay laboratory strain collection. A980 has a genome size of 8.01 Mbp with 6774 putative protein-coding sequences (Figure 4.2). The final assembly has a single large scaffold, A980_AS6_SC1, containing 97% of the genome and housing all but one of the biosynthetic gene clusters. Seven small scaffolds of decreasing size, which could not be mapped to the core sequence due to a lack of overlapping regions, were also obtained. These are named A980_AS6_SC2, A980_AS6_SC3, A980_AS6_SC4, A980_AS6_SC5, A980_AS6_SC6, A980_AS6_SC7, and A980_AS6_SC8. The final biosynthetic gene is found in A980_AS6_SC2. The G+C percentage content of the genome is approximately 72% and the strain does not appear to house a plasmid.

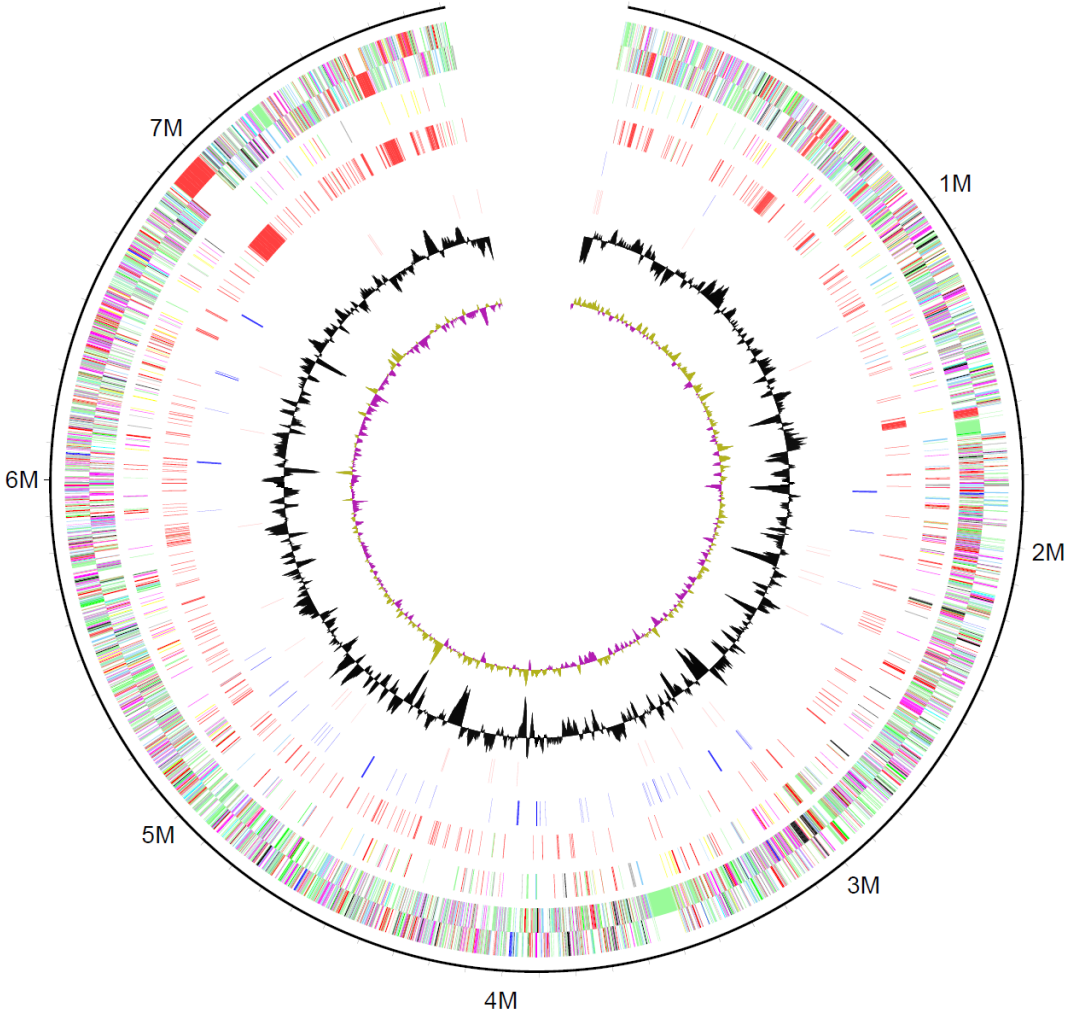


Figure 4.2: Map of the linear *Streptomyces* A980 chromosome scaffold 1, comprising 97.5% of the total genome. The scale is numbered in megabases. The outermost two rings show genes on the forwards and reverse strands, respectively. The coloured bands represent the class of gene; red, secondary metabolism, black, energy metabolism, pale blue, regulators, yellow, central metabolism, cyan, degradation of large molecules, magenta, degradation of small molecules, dark green, surface associated, orange, conserved hypothetical, pale green, unknown, grey, miscellaneous. The third ring shows the essential genes for cell division, DNA replication, transcription, translation, and amino acid biosynthesis. The fourth ring shows the genes of secondary metabolism. The large gene cluster around the 7 Mbp mark is discussed below. The fifth ring shows the mobile genetic elements (blue, transposases, red, prophages or integrated plasmids). The black sixth ring shows the GC content. The innermost ring shows the GC bias (khaki, >1, purple, <1).

4.2.2 Phylogenetic studies and whole genome comparisons

To explore the phylogenetic context of A980 a BLAST [175] comparison was made of its 16S rRNA sequence to that of other *Streptomyces* spp. in the NCBI database. This revealed that A980 is correctly assigned as *S. albus*, however the 16S rRNA nearest neighbours were too closely related to produce a dendrogram that gave any meaningful information on the phylogeny. Instead a multilocus sequence alignment (MLSA) was performed using the sequence of the five housekeeping genes, *atpD*, *gyrB*, *recA*, *rpoB* and *trpB* [236]. The phylogenetic tree in Figure 4.3 shows that A980 is closely related to *S. albus* type strains.

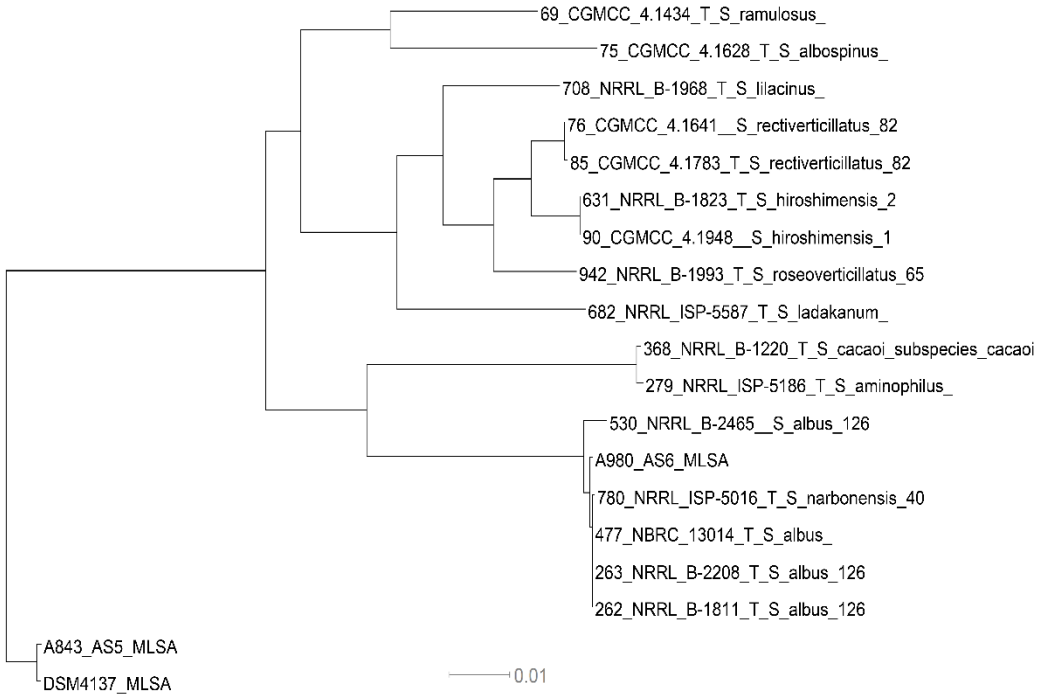


Figure 4.3: Phylogenetic tree based on MLSA of five housekeeping gene sequences of A980 and other Streptomyces spp. Evolutionary history was inferred using the Maximum Likelihood Method [237]. All positions containing gaps were eliminated and there were 2514 nucleotides in the final dataset. Evolutionary analysis was conducted by MEGA7 [238].

Comparisons of 16S rRNA and housekeeping genes offer useful information on the phylogeny of an organism but give little information on the differences across the rest of the genome. To address this MUMmer3, a bioinformatics software for the rapid alignment of whole or draft genome sequences [239], was used to compare the genome sequence of A980 against those of the type strains

S. albus NRRL B-1811 and *S. albus* NBRC 13014. The alignments were generated and visualised as dot plots by Dr. Markiyan Samborskyy (Figure 4.4). The genome comparisons illustrated by Figure 4.4 show that both the type strains NRRL B-1811 and NBRC13014 are highly homologous to A980, with no major DNA rearrangements. The only difference of any significance occurs at the 3' end of the genome. A980 is slightly longer at the 3' end, and there is a break in the homology between A980 and the type strains, where A980 has additional nucleotides not found in the other two strains. This is probably the large gene cluster highlighted in Figure 4.3. The genomes for both NRRL B-1811 and NBRC13014 were contained within over 100 scaffolds which had to be aligned against the A980 sequence before the comparison could be made.

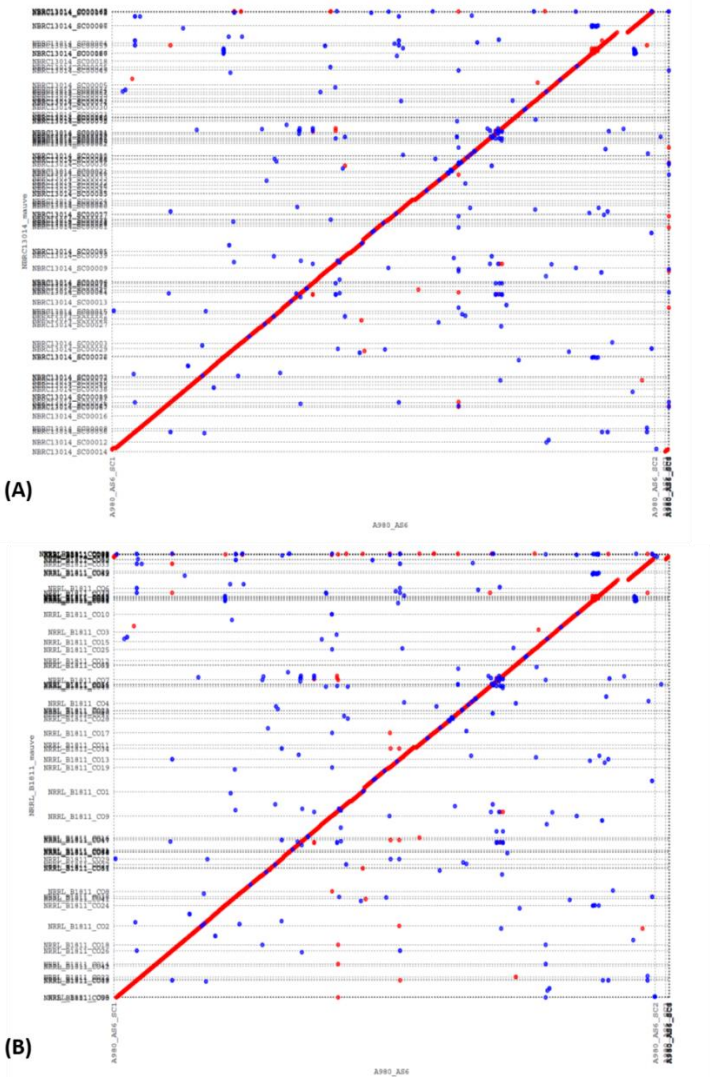


Figure 4.4: Whole-genome sequence alignments of A980. The analysis was conducted in MUMmer3 with a minimum value of 128 base pairs. Red, forwards strand match, blue, reverse strand match. (A) Comparison against NBRC13014. (B) Comparison against NRRL B-1881.

4.2.3 Annotation of the genome of *Streptomyces albus* DSM40763

After sequencing and assembly of the A980 genome an initial detection and annotation of gene clusters was carried out using AntiSMASH 4.0 [178]. This analysis revealed 27 gene clusters across two scaffolds. The biosynthetic gene clusters tend to be located at either end of the linear chromosome, a feature observed in other *Streptomyces* spp. [173]. The putative pseudouridimycin gene cluster was not directly identified by AntiSMASH. However, screening of the A980 genome sequence with a biosynthetic gene, TruD, from the pathway to the C-nucleoside malayamycin A, identified a likely candidate, as described later in this Chapter, which had been listed by AntiSMASH under "Other". The full list of annotated gene clusters is shown in Table 4.1.

The 27 gene clusters revealed by AntiSMASH analysis were manually investigated to further elucidate the biosynthetic potential of A980. Six gene clusters were annotated as terpene biosynthetic gene clusters. The putative products of five of the clusters, 2,3,8,14, and 19 were identified as phytoene, geranylgeranyl pyrophosphate, dehydrosqualene, geosmin, and selinadiene, respectively, by comparison against the PDB database. The final cluster annotated as a terpene, cluster 26, contained several putative biosynthetic genes including a putative cyclodipeptide synthase and a putative phytoene synthase. Homologous gene clusters with this combination of genes have previously been identified and proposed to generate a fused cyclodipeptide-terpene product [240].

Two putative siderophore gene clusters were identified; clusters 11 and 18. When analysing cluster 11 with BLAST, none of the three biosynthetic genes had a high similarity with any sequences in the PDB database. Cluster 18 however, contains four genes which when compared to the PDB database with BLAST showed high similarity to the four genes responsible for the biosynthesis of the iron chelator, desferrioxamine [241].

Cluster 9 is annotated as a putative bacteriocin biosynthetic gene cluster. However, the cluster does not contain any genes which might encode a core peptide, or any posttranslational modification enzymes expected in a true bacteriocin gene cluster. Further analysis using BAGEL3, an online platform for identification of genes encoding bacteriocins and bactericidal post-translationally modified peptides [242], found no evidence of bacteriocin biosynthetic genes in the region identified by AntiSMASH.

Upon comparison against the PDB database cluster 17 appears to contain the three genes necessary for ectoine biosynthesis; diaminobutyric acid acetyltransferase, diaminobutyric acid transaminase, and ectoine synthase [243].

AntiSMASH identified two putative lanthipeptide gene clusters; cluster 4 and 7. Cluster 4 is likely a class I lanthipeptide due to the presence of LanB dehydratase and LanC cyclase homologues responsible for the formation of a carbon-carbon double bond and the cyclisation of the precursor peptide [244]. Cluster 7 contains a gene, ORF01056, with a central kinase-like domain and a C-terminal LanC-like domain, indicative of a class III lanthipeptide.

Two putative lassopeptide gene clusters were identified; clusters 12 and 27. However a closer examination of the gene clusters indicated that neither would be capable of generating lassopaptides. The genetic architecture required for lassopeptide biosynthesis is well known, composed of three genes that are strictly necessary for the biosynthesis [245]. A precursor peptide encoded by the A gene is processed by the B and C maturation enzymes to generate the lasso fold. All lassopeptide gene clusters share homologues of these genes. Neither cluster 12 nor 27 held homologues of these genes, suggesting that they have been erroneously annotated.

AntiSMASH identified four nonribosomal peptide synthase (NRPS) gene clusters; clusters 1, 13, 21, and 25. Cluster 1 encodes an unknown tripeptide. Cluster 13 contains a single NRPS gene containing single condensation, adenylation, peptidyl carrier (PCP) and thioesterase domains, suggesting a putative iterative NRPS. A second gene is a putative methyltransferase. Cluster 21 appears to encode an unknown dipeptide. AntiSMASH predicts cluster 25 to generate a pentapeptide, however the cluster is not as well organised as expected for a functional NRPS gene cluster. The NRPS genes are separated by at least 10 kb of unrelated genes, there are two genes with adenylation domains and no PCP domains, and no clear initiation module. Biochemical and genetic investigation will be required to better elucidate this gene cluster.

Two polyketide synthase (PKS) gene clusters were identified by AntiSMASH; clusters 16 and 23. Cluster 16 is annotated as a Type II PKS. A BLAST analysis against the PDB database identifies two genes; ORF08410 and ORF08411 as homologues for the KS_{α} and KS_{β} required for actinorhodin biosynthesis [92,246]. Cluster 23 encodes an unknown type 1 PKS with one loading module and seven extension modules, the last including a thioesterase (TE) domain. Each extension module contains a ketoreductase (KR) domain, and modules 1 and 2 contain dehydratase (DH) domains. Assuming the colinearity principle of PKSs, this gene cluster encodes a polyene/ol.

AntiSMASH also identified four type 1 PKS hybrids. Cluster 10 is annotated as a PKS-NRPS hybrid, cluster 15 as a PKS-lanthipeptide, cluster 20 as a PKS-oligosaccharide, and cluster 24 is annotated as a PKS-terpene hybrid. Cluster 10 encodes two putative functional PKS genes, two putative functional NRPS genes, and two genes containing lone adenylation domains. The two PKS genes include one loading module and 3 extension modules, while the two NRPS genes include one

loading module and one extension module. The PKS and NRPS genes are held in separate genes, so there is unlikely to be direct functional hybridisation between the NRPS and PKS proteins. Instead a possible biosynthetic pathways could include the PKS and NRPS components being synthesised individually, then coupled together by a ligase, or in a cyclosporin-like pathway [247], the PKS could be converted to an amino acid before being incorporated into the natural product by the NRPS enzyme. Cluster 15 contains an interesting mixture of PKS genes and lanthipeptide biosynthetic genes. It is annotated as a lanthipeptide-PKS hybrid; however, this hybrid cluster is most likely several clusters, incorrectly annotated. It contains two genes that encode precursor peptides, only one of which has local LanB dehydratase and LanC cyclase homologues. There is also a gene encoding a PKS loading module, and several other genes that encode independent putative PKS biosynthetic enzymes.

Cluster 20 appears to encode the biosynthetic genes for a large polyketide. It contains seven putative PKS genes containing one loading module and 20 extension modules, the last containing a TE domain. There are also two putative glycosyltransferase enzymes, suggesting that this polyketide is glycosylated. The final hybrid gene cluster is cluster 24. The more interesting biosynthetic genes of this cluster are a putative PKS gene that appears to be a single extension module, and a squalene/phytoene-like terpene synthase. The lack of a loading module hints that this PKS gene could encode a single module iterative type I PKS such as the PKS responsible for the biosynthesis of phenolic moiety in avilamycin [248], or for the biosynthesis of the enediyne in calicheamicin [249]. The cluster also contains a putative oxidoreductase, aminotransferase and carboxylate-amine ligase, which could be responsible for the amination of either the terpene or the iterative PKS product and subsequent ligation with the other compound to give the final product.

Clusters 5, 6, and 22 were annotated as "Other". At first glance, none of the three appear to be functional biosynthetic gene clusters. However closer investigation of cluster 6 reveals several genes of interest, including a TruD homologue.

Table 4.1: Biosynthetic gene clusters found in A980. Clusters were analysed using AntiSMASH 4.0. Gene cluster limits were kept from AntiSMASH predictions. Annotated gene clusters were manually identified from key biosynthetic genes using BLAST [175].

Cluster	AntiSMASH 4.0 prediction	Nucleotide		Gene ID		Annotation
		From	To	From	To	
A980_AS6_SC1						
1	NRPS	75234	128829	114	190	Unknown
2	Terpene	148468	172932	219	262	Phytoene
3	Terpene	221673	243559	338	373	Geranylgeranyl Pyrophosphate
4	Lanthipeptide	345987	370632	535	567	Unknown
5	Other	408817	449653	616	671	-
6	Other	634198	675436	955	1016	PUM
7	Lanthipeptide	691897	725605	1042	1090	Unknown
8	Terpene	939670	966374	1423	1463	Dehydrosqualene
9	Bacteriocin	1434686	1445516	2150	2167	Unknown
10	Type I PKS-NRPS	1577987	1641100	2369	2440	Unknown
11	Siderophore	1782846	1798150	2602	2620	Unknown
12	Lassopeptide	2545416	2568398	3702	3731	Unknown
13	NRPS	2718705	2762826	3953	4012	Unknown
14	Terpene	5505171	5527402	7965	7999	Geosmin
15	Lanthipeptide-Type I PKS	5746467	5797885	8324	8399	Unknown
16	Type II PKS	5783012	5825590	8372	8441	Actinorhodin-like Type II PKS
17	Ectoine	5995802	6006206	8685	8697	Ectoine
18	Siderophore	6074265	6086076	8811	8819	Desferrioxamine
19	Terpene	6247497	6284125	9080	9144	Selinadiene
20	Type I PKS-Oligosaccharide	6883231	7074071	10057	10199	Unknown
21	NRPS	7394703	7441235	10671	10738	Unknown
22	Other	7455353	7498613	10763	10815	-
23	Type I PKS	7486455	7567810	10802	10885	Unknown
24	Type I PKS-Terpene	7586501	7642165	10916	11006	Unknown
25	NRPS	7632513	7711931	10992	11094	Unknown
26	Terpene	7728151	7760470	11113	11161	Unknown
A980_AS6_SC3						
27	Lassopeptide	9383	31822	10	38	Unknown

4.2.4 Bioinformatic analysis of uncharacterised gene clusters

Four uncharacterised biosynthetic gene clusters were identified from the AntiSMASH 4.0 analysis of the whole-genome sequence of A980. These clusters were chosen for their potential to generate novel compounds, or because their annotated gene cluster required further deconvolution to understand the biosynthetic potential. In the following section clusters 15 (a lanthipeptide-PKS hybrid), 16 (an actinorhodin-like type II PKS), 20 (a large type I PKS-oligosaccharide), and 26 (a terpene) are analysed bioinformatically to assess their potential to generate novel chemical diversity. The analysis and predictions were reasoned from comparative sequence analysis with homologous genes and gene clusters and where possible by study of enzyme active sites.

Cluster 15: a novel lanthipeptide-PKS hybrid natural product?

AntiSMASH identifies cluster 15 as a lanthipeptide-PKS hybrid gene cluster. This would certainly generate novel chemical diversity and have intriguing biosynthesis. Gene clusters containing lanthipeptide and PKS elements have been identified before, such as the GCF133 family of biosynthetic clusters where a LanKC homologue appears beside type II PKS genes [250]. Cluster 15 spans a region of 51.4 kbp and contains 47 contiguous open reading frames encoding two precursor peptides, one pair of LanB and LanC homologues, another smaller LanB homologue, and one PKS enzyme. There are also ORFs encoding an acyl-CoA synthetase, two acyl-CoA dehydrogenases, a FabH-like protein, an FkbH-like protein, an ACP, a TE, and a P450. There is also a 4'-phosphopantetheinyl transferase, an uncharacterised FAD-binding enzyme, one transport related enzyme, four regulatory enzymes and 21 ORFs annotated as 'other'. The gene organisation of cluster 15 is shown in figure 4.5.

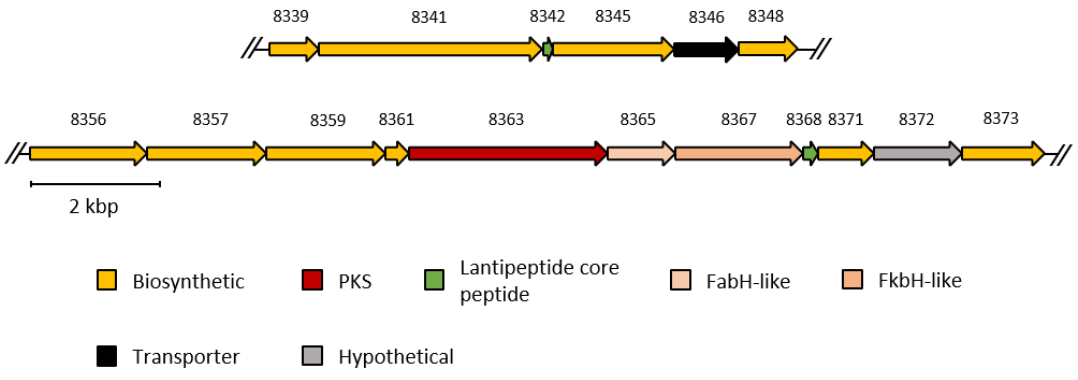


Figure 4.5: Organisation of the core biosynthetic gene cluster 15 in A980

It becomes clear that AntiSMASH has picked up two distinct biosynthetic gene clusters. Cluster 15a encodes a putative class I Lanthipeptide where the LanB and LanC homologues services two precursor peptides. One of those peptides, ORF08368, has been embedded in the middle of another cluster. It is unlikely that the products of the two clusters interact, but the location of ORF08368 raises questions regarding the origin of this genetic architecture. Cluster 15b appears to encode a natural product made up of products from polyketide and fatty acid synthases. The PKS gene encodes only one module containing one KS domain, one AT domain and an ACP domain. The substrate specificity of the AT domain is predicted to be malonyl-CoA by AntiSMASH, using the Minowa substrate specificity predictor [251]. The PKS modules is unlikely to be a type II iterative system, due to the lack of the two ketosynthase domains, KS_{α} and KS_{β}, typical of such systems [252]. The product generated by this gene cluster will probably only have a single malonyl moiety.

The FabH and FkbH homologues give further clues as to what class of compound this gene cluster might generate. FabH is involved in type II fatty acid synthesis. It is known to condense malonyl-ACP with acetyl-CoA, yielding acetoacetyl-CoA [253]. The protein which gives its name to the FkbH family of enzymes is involved in the biosynthesis of the immunosuppressant FK520 [254]. It catalyses the formation of glyceryl-CoA, as part of the methoxymalonyl-CoA biosynthesis, an exotic building block used in the biosynthesis. The FkbH family of enzymes contain a central domain that is highly homologous to phosphatases from the haloacid dehalogenase family [255]. When FkbH homologues are aligned a subfamily of enzymes with an additional N-terminal extension appear. Tmn16 is an FkbH family enzyme with this N-terminal extension. It catalyses the first step in the formation of the tetronate moiety in the tetronomycin biosynthesis [256]. When the N-terminal extension was removed, the truncated Tmn16 was unable to transfer the glyceryl moiety to the downstream ACP, halting synthesis [256]. Investigation of FkbH homologues with the N-terminal extension reveals that they are all involved in tetronate formation, providing a useful marker to identify the subfamily of FkbH family enzymes and the class of natural product they are used to generate. When the FkbH homologue in cluster 15b is aligned with several FkbH family enzymes of both types it clearly shows a N-terminal extension, indicating it is involved in tetronate formation (Figure 4.6). Tetronate ring formation requires several genes; an FabH-like protein, an FbkH-like protein and an ACP. All of these are present in cluster 15b, making it very likely that cluster 15b generates a compound with a tetronate moiety.

Figure 4.6: Sequence alignment of ORF08367 from A980 with FkbH homologues. The presence of an N-terminal extension indicates enzymes involved in tetronate formation. The N-terminal extension is highlighted in yellow. Sequences were aligned with Clustal Omega.

Cluster 16: an actinorhodin-like type II PKS?

Cluster 16 is an uncharacterised biosynthetic gene cluster which appears to encode an actinorhodin-like type II PKS. The cluster spans 42.6 kbp and contains 47 contiguous open reading frames. The core biosynthetic genes in this cluster (orf08400, orf08410, and orf08411) encode an ACP and two beta-ketoacyl synthases, which share high identity scores with the KS_{α} and KS_{β} domains from actinorhodin biosynthesis, respectively. The rest of the biosynthetic genes in cluster 16 consist of; a type II TE, three dehydrogenases, two cyclases, one of which is identical to the F2 cyclase in tetracenomycin biosynthesis [257], and two acetyl-CoA carboxylases. There are also two SARP family transcriptional regulators and a multidrug efflux enzyme, similar to the actinorhodin transporter form *S. coelicolor*. The gene cluster organisation is shown in Figure 4.7. A section of the gene cluster (orf08398 to orf08423) containing the core biosynthetic genes is conserved across several other *Streptomyces* spp as identified by a BLAST homology search. The homologues can be found in *Streptomyces* sp. HPH0547, *Streptomyces* sp. AA1529 and *Streptomyces flavogriseus*, the xantholipin producer [258] and show near identical gene arrangement, except for a set of five ORFs containing the ACP and cyclase, which is inverted. The conservation of this gene cluster across species encourages the view that the cluster may be functional and expressed under certain conditions.

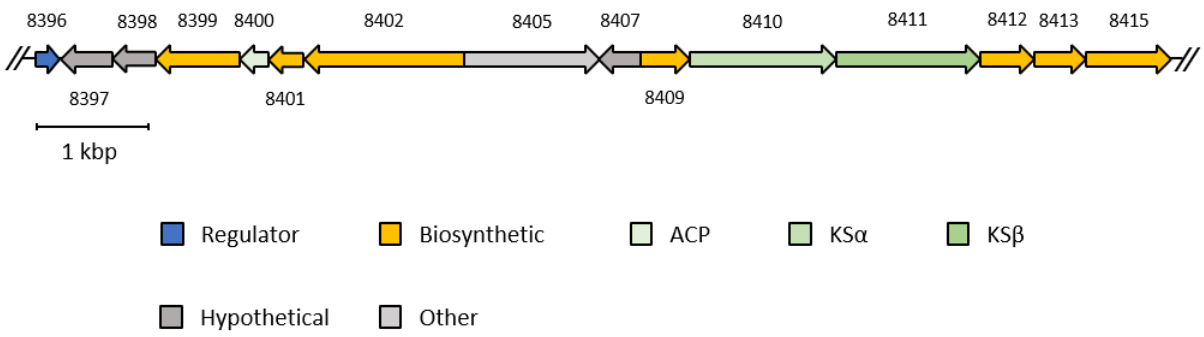


Figure 4.7: Organisation of the core biosynthetic gene cluster 16 in A980.

The genes contained by cluster 16 strongly implicate an aromatic type II polyketide product as the end product of the pathway. The extension unit used in the biosynthesis of this product is most likely malonyl-CoA, as indicated by the presence of acetyl-CoA carboxylases (though enzymes annotated as acetyl-CoA carboxylases often also accept propionyl-CoA). The number of extension units in the product is likely to closely resemble the actinorhodin biosynthetic pathway due to the similarity between the respective KS_{α} and KS_{β} proteins of each pathway. KS-CLF didomains are known to have a series of 'gates' within the tunnel that the growing polyketide extrudes into. These gates can be open or closed depending on the size of the amino acid residues at those locations [259]. The length of the growing chain is controlled by these gates, as the chain extension appears to halt once it reaches a closed gate. However, the true number is unsafe to predict, and biochemical evidence is needed.

Actinorhodin biosynthesis starts with the condensation of one starter unit and seven extender malonyl units into the octaketide PKS product. At the heptaketide stage the chain has reached the end of the tunnel, closed off by a gate. The final extension causes the chain to buckle, allowing for intramolecular cyclisation [92]. This is followed by series of reduction and spontaneous and catalysed cyclisations. After the final cyclisation the scaffold is decorated by hydroxylation at the C-6 and C-8 positions [260] followed by dimerization to form actinorhodin (figure 4.8).

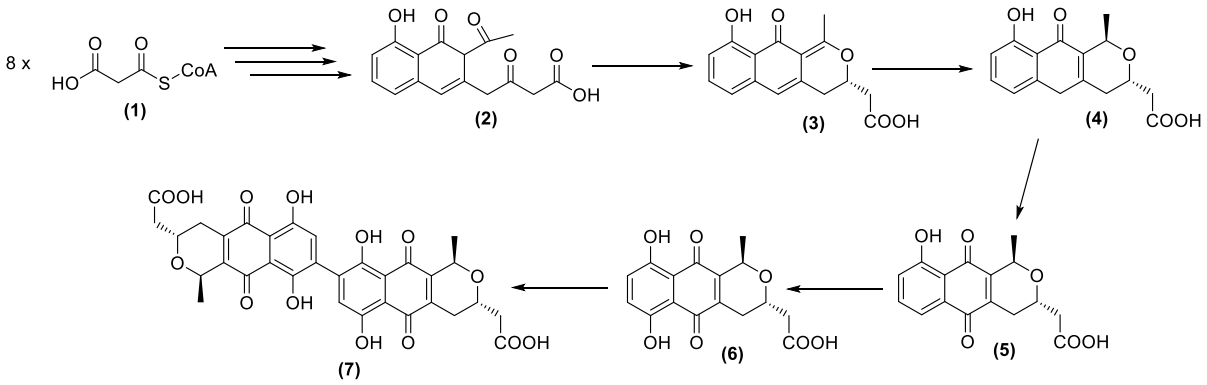


Figure 4.8: Actinorhodin biosynthetic pathway. (1) eight malonyl-CoA used to generate the polyketide backbone **(2)** bicyclic intermediate **(3)** (*S*)-DNPA **(4)** 6-deoxy-dihydrokalafungin **(5)** dihydrokalafungin **(6)** 8-oxy-dihydrokalafungin **(7)** actinorhodin.

The putative biosynthetic pathway encoded by cluster 16 appears to be an unremarkable type II PKS, however the presence of different enzymes with high sequence identity to genes found in different type II PKS gene clusters hints at an interesting evolution. Perhaps this is an example of a relatively new gene cluster formed by the combination of multiple other clusters and hasn't had time for the individual genes to diverge. Whatever the case, this gene cluster warrants further investigation.

Cluster 20: a novel PKS-oligosaccharide?

Cluster 20 is annotated by AntiSMASH as a type I PKS oligosaccharide. The cluster appears to encode a large polyol tailored with multiple sugar moieties, indicated by the seven glycosyltransferase genes found in the cluster. Cluster 20 is huge, spanning 190.8 kbp. Its core biosynthetic genes consist of seven type I PKS genes encoding a starter module and 20 extension modules with a terminal type I TE domain. The cluster also contains many post-PKS tailoring enzymes; three aminotransferases, three epimerases, three monooxygenases, two oxidoreductases, and seven glycosyltransferases in addition to drug resistance transporters and several LuxR transcriptional regulators. Figure 4.9 shows the organisation of the gene cluster.

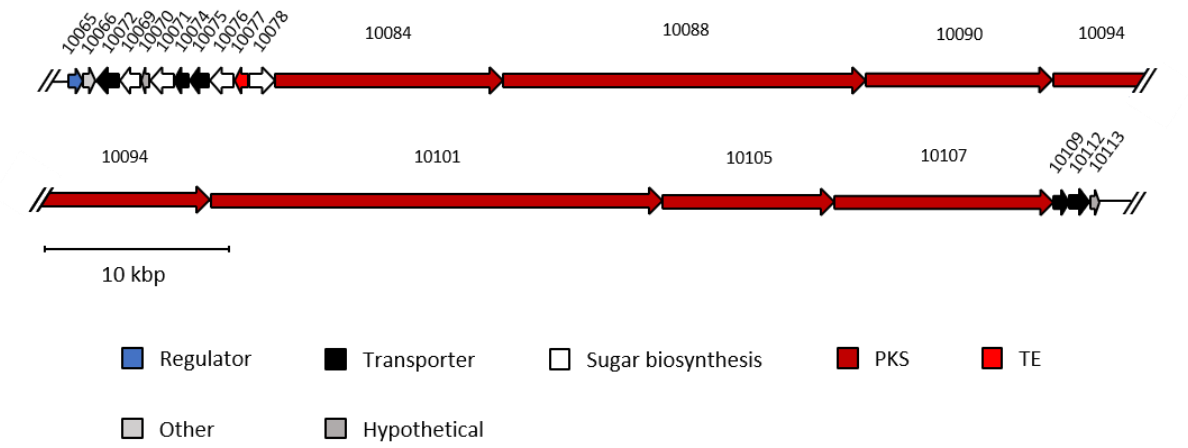


Figure 4.9: Organisation of the core biosynthetic gene cluster 20 in A980.

From an understanding of the relationship between the sequence of amino acid residues and the function of PKS domains, as discussed in the introduction, the specificity of each domain can be predicted. There is one loading AT domain, and 20 extension AT domains. Nine of the AT domains are specific for methylmalonyl-CoA, having the characteristic YASH motif, the remaining 12 are specific for malonyl-CoA, as indicated by a HAFH motif [104,135]. There are 19 KR domains, as module 14 lacks a KR, despite having a DH domain. Seven of the KR domains have the characteristic LDD motif, indicating they are B type KR domains [137]. The remaining KR domains are A type, and three of the them are A2 type, as indicated by the conserved WxxxxH motif in the catalytic site [137]. The six DH domains found in this PKS all have the characteristic motifs required for activity [152], however the DH from module 14 won't have any activity due to the lack of a KR domain in that module to generate the necessary hydroxyl group. Module 16 also has a ER domain that appears to be active. The KS in the loading module was confirmed as a starting KS by the presence of the characteristic QASS motif. Figure 4.10 summarises the sequence alignments and characteristic sequence motifs and Figure 4.11 shows the hypothetical full-length predicted polyketide after release from the PKS.

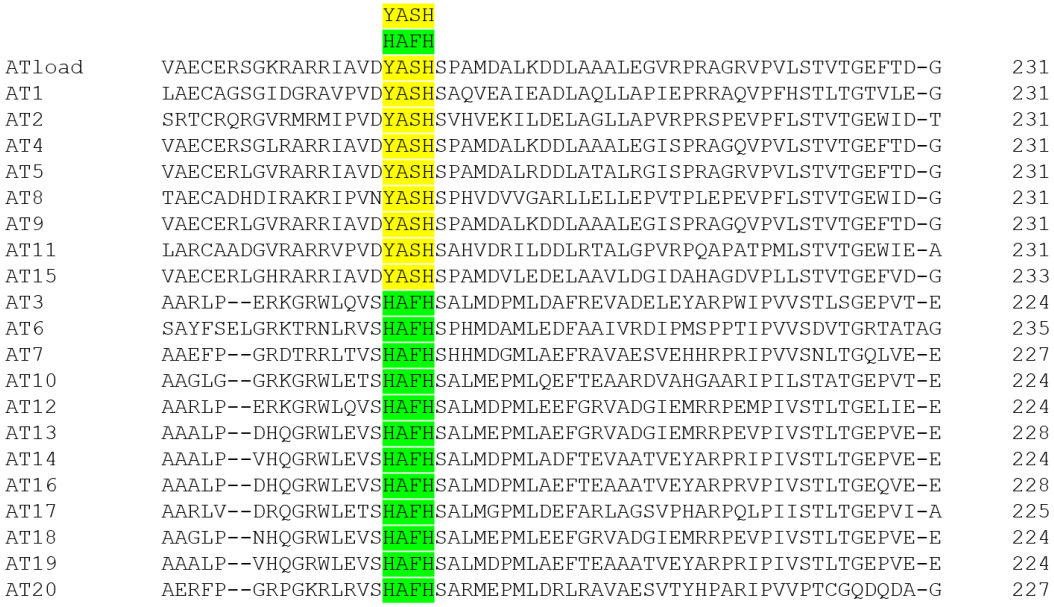


Figure 4.10: Sequence alignments of cluster 20 PKS domains and amino acid motifs conferring activity. The YASH motif indicates a methylmalonyl-CoA specific AT domain. The HAFH motif indicates a malonyl-CoA specific AT domain. The LDD motif indicates a B-type KR domain. The WxxxxH motif indicates an A2-type KR domain. Sequences were aligned with Clustal Omega.

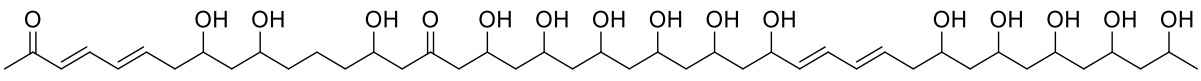


Figure 4.11: Predicted polyketide intermediate from cluster 20 PKS.

The presence of a pair of dienes in the polyketide backbone at first suggested cluster 20 may encode a novel bafilomycin [261], however the polyketide generated by cluster 20 is far larger than any bafilomycin class molecule, and the dienes are too far apart to form the characteristic bafilomycin structure. The PKS genes do not have any known homologues so this cluster likely encodes a novel polyketide scaffold. The length of the polyketide backbone is longer than usual, but not unexpectedly large [179]. The post-PKS tailoring enzymes however, are of interest. The cluster contains three aminotransferase genes that are labelled as eryC1 homologues (ORF10069, ORF10071 and ORF10132). EryC1 is involved in erythromycin biosynthesis either as a structural gene in the formation of the deoxyamino-sugar desosamine [66]. On closer inspection only ORF10069 is a close homologue to eryC1, but the other two are homologous to enzymes involved in sugar biosynthesis. The cluster also contains several glycosyltransferases. At least one appears to be a true biosynthetic glycosyltransferase, while the others appear to be involved in antibiotic self-resistance. These glycosyltransferase enzymes will glycosylate sugar moieties on the molecule and once the product is transported outside the cell the molecule is converted back to the antibiotic by an extracellular hydrolysis enzyme [262]. Cluster 20 also has an NDP hexose 2,3-dehydratase, and an NDP hexose 3-ketoreductase, which are both involved in 2,6-dideoxysugar biosynthesis [263]. There is also a putative epimerase, and a putative 3,4-ketoisomerase to further modify the sugar molecule, which may also be aminated by one of the aminotransferase enzymes. The putative sugar biosynthesis is summarised in Figure 4.12.

Finally, there are also two genes that are implicated in polyisoprenoid biosynthesis. ORF 10113 is a putative homologue of a polyisoprenoid binding protein from *Thermus thermophilus* and a member of the YceI family of enzymes that are involved in isoprenoid metabolism, transport, or storage [264]. ORF10137 is a putative menaquinone biosynthetic enzyme, also from *Thermus thermophilus*.

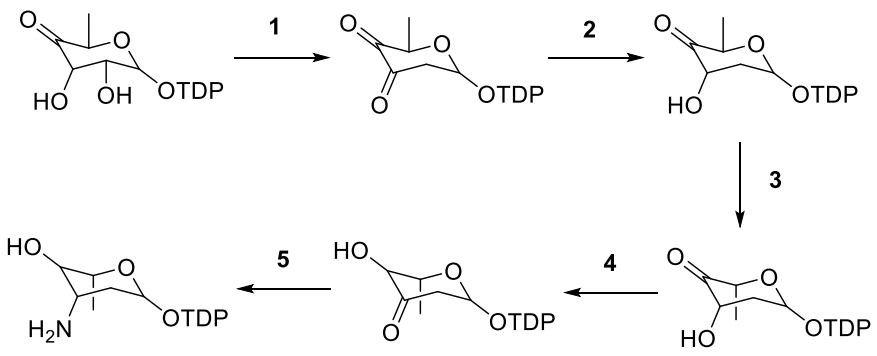


Figure 4.12: TDP-deoxysugar biosynthesis. (1) 2,3-dehydratase **(2)** 3-ketoreductase **(3)** epimerase **(4)** 3,4-ketoisomerase **(5)** aminotransferase.

While the exact product of this gene cluster is difficult to predict, the product is definitely novel, as there are no similar PKS gene clusters in the database. It appears to be a novel combination of type I PKS and isoprenoid, but further investigation is required to properly characterise the cluster. The first attempts at studying this cluster should begin by ensuring the cluster is active under normal laboratory conditions. Knocking out or interrupting key biosynthetic genes and then comparing the fermentation profiles of the mutant and wildtype can be an effective method for this. It is also possible to take advantage of recent advances in overexpression mutants. The cluster has three LuxR family transcriptional regulators, two FscRIII homologues and one FscRIV homologue (from the candicidin biosynthetic pathway [265]). These are positive regulatory genes, and can be overexpressed by introducing a strong promoter to boost expression of the gene cluster and increase the yield of the product [179].

Cluster 26: a hybrid cyclodipeptide-terpene?

Cluster 26 spans 14.0 kbp, however the core biosynthetic genes only cover 3.3 kbp. The core cluster contains a phytoene synthase, a cyclodipeptide synthase (a family of peptide bond-forming enzymes that utilise two amino acids activated as aminoacyl-tRNA for the synthesis of cyclodipeptides [266]), and two methyltransferases. The extended cluster also has a hydrolase and oxidoreductase. The presence of a LuxR transcriptional regulator encourages the view that this is a real biosynthetic cluster that may be active or has the potential to be activated. Figure 4.13 shows the organisation of cluster.

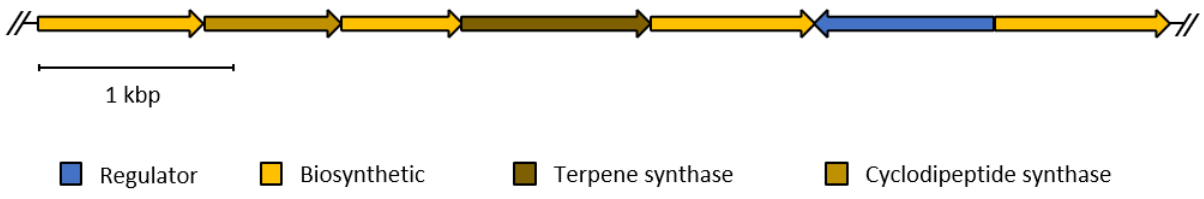


Figure 4.13: Organisation of the core biosynthetic gene cluster 26 in A980.

The presence of a cyclodipeptide synthase in close proximity to a phytoene synthase is unusual and hints at a novel hybrid structure. Cyclodipeptides are widespread in nature and exhibit a broad variety of biological activities [267]. They are typically under the control of LuxR regulators, and are commonly associated with a range of tailoring enzymes such as oxidoreductases, hydrolases, methyltransferases and peptide ligases [267], however an association with terpene biosynthetic genes is unusual. One study by Skinnider *et al*. aimed to group genetically encoded cyclodipeptides in an effort to create an online tool to facilitate genome mining of cyclodipeptides. They identified a family of cyclodipeptide gene clusters that contained terpene biosynthetic machinery. This class of cyclodipeptide synthases had a methyltransferase domain fused to the C-terminus [240]. When the cyclodipeptide synthase in cluster 26 is inspected by BLAST analysis, a methyltransferase domain is revealed in the C-terminus. This evidence strongly supports the view that the product of cluster 26 is a novel cyclodipeptide-terpene hybrid.

4.2.5 Identification of the putative pseudouridimycin biosynthetic gene cluster

Pseudouridine (Ψ) has been referred to as the fifth base, because it is the most abundant post-transcriptional alteration found in Nature [268]. Several pseuduridine synthases have been identified, usually named from their counterparts in *E. coli*. The reaction they catalyse uses as substrate a preformed tRNA or other ncRNA, and each enzyme isomerises one or more specific uridine residues to pseudouridine (Ψ) [269]. The Ψ moiety in the PUM structure could be generated through several ways, which will be discussed in detail in the next chapter. One of those could be through a dedicated pseudouridine synthase, such as the TruD-like [270] enzymes found by Dr Hui Hong in the recently-characterised gene clusters for the C-nucleoside malayamycin A (Hong *et al*., (2018) *ms*. in the press) from *Streptomyces malaysiensis* DSM14702 and *Streptomyces chromofuscus*. The sequence of TruD from *S. malaysiensis* was used as a probe in a BLAST search of the *S. albus* DSM40763 (A980) genome, yielding only one significant hit (43% identity, 56% similarity). The sequence of this putative A980 TruD homologue was aligned with the authentic TruD enzymes using Clustal Omega (Figure 4.14). It shares regions of conserved sequence and possesses the essential active site aspartic acid residue within one of these conserved regions.

```
TruD_EC         MIEFDNLTYLHGKPQGTGLLKANPEDFVVVEDLGFEP--DGEGEHILVRILKTGCNTRFV        58
TruD_SA         ------------MTTGNPVLKHRPEDFRVRENLVVPLTRAQSAPHRYLLLHKCGYTTMEA        48
TruD_SM         --------------VEYPVLKAHDTDFVVQESMVLPAQDDPDAPYQYLRLRKRGYTTFEA        46
TruD_SC_        --------------VEYPILKAHDTDFVVQESMVLPVQDDQDAPFQYLRLRKRGYTTFEA        46

                                        *
TruD_EC         ADALAKFLKIHAREVSFAGQKDKHAVTEQWLCARVPGK-----EMPDLSAFQLEGC---Q        110
TruD_SA         VRWVADRLGLPSSEVGYAGLKDEDGVTEQLLSVPVTARTAAGGPVEEEPALSEEPGRWLR        108
TruD_SM         LERIAALCGVSASDLSTAGLKDEDAVTEQHIACRGRLPRD---AITEFNARHASGARTMT        103
TruD_SC_        VERIAAVCGVPASDLSTAGLKDEDAVTEQHLACRGGLSPD---GIAEFNTRHGSGERTMR        103


TruD_EC             VLEYARHKRKLRLGALKGNAFTLVLREVSNRDDVEQRLID--ICVKGVPNYFGAQRFGIG    168
TruD_SA         LHHYGYGHEALTVGRLNGNGFRVVLRDLDEATAARLAERG--RRNLLFVNYYDTQRFGVP        166
TruD_SM         LTTHGYGHLNLRAGQLEGNAFRITVRGLSSGFATTLGALGERAENLFFVNYYDVQRFGVA        163
TruD_SC_        LSLHGYGHQGLRAGHLEGNGFRIVVRGLSAGFVRTVEALGERAGDLFFVNYYDTQRFGVA        163


TruD_EC         GSNLQGA-----------------------------QRWAQTNTP----VRDRNKRSFWL        195
TruD_SA         GGPKRTHLVGEALLKEDWGLARRELATLGAPESGTARRWTGPDRDLFRRGLDARTTAFYL        226
TruD_SM         GGPRTTHQIGGALLEEDYGTALALVRESQSPEADHARRFTGPAAQFFD-RIDPRVRAFYL        222
TruD_SC_        GGPRTTHQIGRALLEEDHGTALGLVRESKSPEAERARLFTGSAAAFFD-RIDPRVRAFYL        222


TruD_EC         SAARSALFNQIVAERLKKADVNQVVDGDALQLAGRGSWFVATTEELAELQRRVNDKELMI        255
TruD_SA         AAHSSADWNARVRDLIATSCPGDSVETS---VDGLPYRFPTSGADVAALQAAYH------        277
TruD_SM         CARASWVWNGQLAALVKKVSHHPLDETL---REGLPYVFTTHRDDVLALLRHAT------        273
TruD_SC_        CAHASYVWNAQLAALLRKVGRHPLDETV---REGIPYTFATHRDDVLALLQEAT------        273


TruD_EC         TAALPG-SGEWGTQREALAFEQAAVAAETELQALLVREKVEAARRAMLLYPQQLSWNWWD        314
TruD_SA         --ELPYTRYSWR---------DGPVERA-TVRPTVVQTVVTVERT-------GPDDAFPG        318
TruD_SM         --SLPYERYRWS---------NAAIRRTQGLRRTVVQTRIRVDRT-------SEDSDTPG        315
TruD_SC_        --TLPYERHRWD---------GGEMRRTQGLRPTVVQTRVRVDRT-------GEDTDTPG        315

TruD_EC         DVTVEIRFWLPAGSFATSVVRELINTTGDYAHIAE---      349
TruD_SA         RHAVEVSFLLPSGCYATAALRQLVLQC-----------      345
TruD_SM         RYACEIALFLPPGCYATNAVSQWVAWIARTDAPRNANA      353
TruD_SC_        RHACELTLFLPSGCYATNAVAQWVAWAGSER-------      346
```

Figure 4.14. Sequence alignment of TruD family of tRNA pseudouridine synthases. The sequences of TruDs [TruD_SM: TruD from *Streptomyces malaysiensis*; TruD_SC: TruD from *Streptomyces chromofuscus*; TruD_SA: TruD from *Streptomyces albus* A980; TruD_EC: TruD from *E. coli* (accession no. AQZ29382)] were aligned using Clustal Omega. The catalytic aspartate is denoted with an asterisk.

Further evidence that this gene was involved in the biosynthesis of PUM came when the genes flanking the TruD homologue were investigated. In close proximity to the TruD gene are two putative ligases and an amidinotransferase. Retrobiosynthesis applied to the structure of PUM shows that similar enzymes would be appropriate to produce PUM. The pseudouridine moiety would require modification before it would be suitable for ligation. This gene cluster also contains a putative Glucose-Methanol-Choline (GMC) oxidoreductase and an aminotransferase. This cluster is clearly a plausible candidate for PUM biosynthesis.

4.3 Concluding remarks

The whole-genome sequence of a putative pseudouridine producer, *Streptomyces albus* DSM40763 (A980), has been obtained through shotgun DNA sequencing using the Illumina platform and assembled into three scaffolds, with the largest scaffold containing 97% of the whole genome. Phylogenetic analysis has clarified that A980 is a true *S. albus* and shares high similarity with both NBRC13014 and NRRL B-1881 *S. albus* type strains.

Initial annotation was performed by AntiSMASH 4.0, and subsequent manual curation has revealed a number of novel biosynthetic gene clusters, including several clusters that appear to govern production of novel classes of natural products. The level to which this genome can be analysed for its potential to generate novel chemical diversity is due in large measure to the availability of a high-quality genome sequence. Large biosynthetic gene clusters such as cluster 20 cover huge regions of the genome. Cluster 20 itself spans over 190 kbp, which is larger than many of the genome scaffolds often found in online databases. Proper genome-level analysis is impossible with such fragmented genomes. The number of biosynthetic gene clusters identified in this chapter that deserve further investigation illustrates the potential of genome mining and highlights how vital high-quality genome sequences are to the field. Finally, the pseudouridimycin gene cluster was identified with high confidence. Detailed analysis and characterisation of the cluster is discussed in the next Chapter.

Chapter 5: The Enzymology of Pseudouridimycin Biosynthesis

5.1: Introduction

5.1.1: Structure and Function of Nucleoside Antibiotics

Nucleoside antibiotics are a large, diverse family of natural products derived from nucleosides and nucleotides, a number of these are illustrated in Figure 5.1. They are produced almost exclusively by *Streptomyces* and related actinomycetes. Nucleoside antibiotics exhibit a broad range of biological activities, such as antibacterial, antifungal, antiviral, immunosuppressive, and antitumor activities [271]. Such a broad spectrum of activities is due to the involvement of nucleosides and nucleotides in almost all cellular metabolism [272]. Nucleoside antibiotics can be classified into three major groups based on their biological functions; inhibitors of peptidoglycan biosynthesis, by competitively inhibiting MraY, the translocase that initiates the lipid cycle of peptidoglycan biosynthesis [273], inhibitors of fungal cell wall biosynthesis, by competitively inhibiting fungal chitin synthases, and inhibitors of protein biosynthesis, but competitively inhibiting peptidyl transferase [272]. Nucleoside antibiotics often contain unusual structural features, arising from unusual enzymatic reactions. The rapid expansion of available genomic data has allowed biosynthetic studies of nucleoside antibiotics to accelerate, providing the foundation for engineering of the biosynthesis of these molecules.

Nikkomycins are potent competitive inhibitors of fungal chitin synthases as they mimic the natural substrate for the synthase, UDP-*N*-acetylglucosamine (Figure 5.2). Nikkomycin is nontoxic to mammals, making it a potentially valuable antifungal therapeutic agent. Phase I clinical trials have investigated the dosing effects of nikkomycin Z for the treatment of coccidioidomycosis [274]. It is also an effective treatment for the ecologically significant fungal infection caused by *Batrachochytrium dendrobatidis* which is threatening several endangered species of frogs [275]. Nikkomycins are composed of a di- or tri-peptide backbone containing a non-proteinogenic amino acid residue, hydroxypyridyl-homothreonine (HPHT), and an aminohexuronic acid that is *N*-glycosidically linked to a uracil base (in nikkomycins Z and J) or to a 4-formyl-4-imidazolin-2-one base (in nikkomycins X and I) [272]. The isolation of the nikkomycin biosynthetic cluster from *S. tendae* and its heterologous expression in *S. lividans* enabled a thorough investigation into the biosynthesis of this family of compounds [273]. It has been postulated that the peptide and nucleoside fragments of nikkomycin are biosynthesised separately, then linked by a peptide bond [276]. Analysis of the gene cluster led to the prediction of eleven enzymatic steps required for the generation of the peptide core [277], and feeding experiments indicated that the

HPHT moiety is generated from L-lysine [278]. The exact biosynthetic pathway for the aminohexuronic acid remains to be determined, although uridine and pyruvate have been established as the metabolic precursors [279].

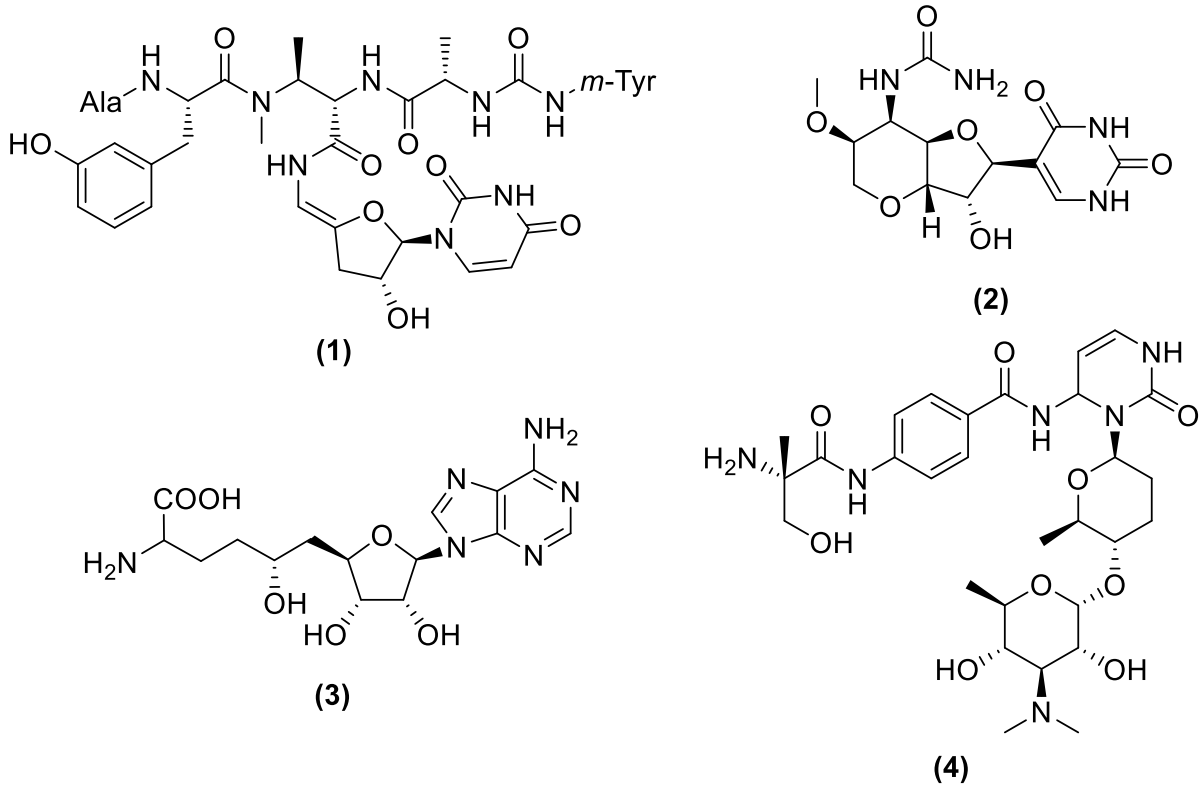


Figure 5.1: Examples of nucleoside antibiotics from *Streptomyces* spp. (1) pacidamycin 1 **(2)** malayamycin A **(3)** sinefungin **(4)** amicetin.

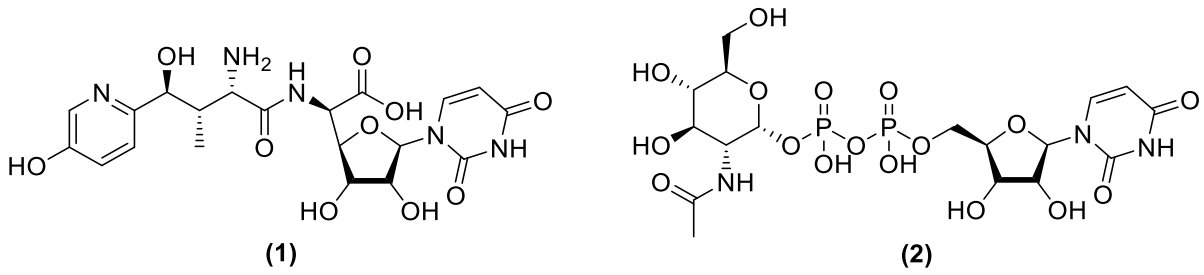


Figure 5.2: Nikkomycin Z mimics fungal chitin synthase substrate UDP-*N*-acetylglucosamine. (1) nikkomycin Z **(2)** UDP-*N*-acetylglycosamine.

Another class of nucleoside antibiotics targets bacterial RNA polymerase (RNAP) to inhibit RNA synthesis. RNAP is particularly suitable as a target for broad-spectrum antibacterial therapy as bacterial RNAP is an essential enzyme, bacterial RNAP subunit sequences are highly conserved, and the bacterial RNAP subunit sequences are not highly conserved in eukaryotic RNAP, which provides a basis for selectivity [280]. Rifampin is a member of this class of molecules and is part of the rifamycin family of antibiotics. It is a potent and broad-spectrum semi-synthetic antibiotic whose precursor is obtained from the fermentation broth of *S. mediterranei*. It is active against most Gram-positive bacteria, some Gram-negative bacteria, and mycobacteria [281]. The bacterial RNAP inhibitory activity was known to totally block formation of the second or third phosphodiester bond in the growing RNA chain [282], however it had no effect of RNAP which had already started synthesising a long transcript. These insights led to a hypothesis that rifampicin's activity is conferred by a steric block of the path the RNA chain took, which was confirmed later by a crystal structure of rifampicin in complex with *Taq* RNAP from *E. coli* [283] (figure 5.3). Rifampin was shown to bind to the β subunit of the RNAP, adjacent to the RNAP active site, which also matched the findings that mutations conferring rifampin resistance are almost exclusively found on the *rpoB* gene, encoding the β subunit [284].

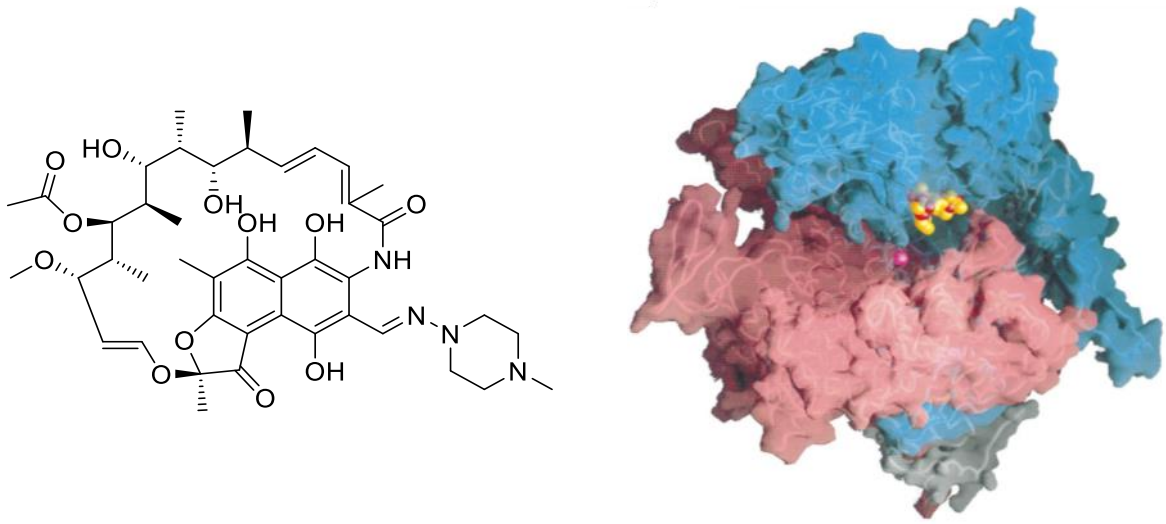


Figure 5.3: Structure of rifampin and the three-dimensional structure of *Taq* RNA polymerase in complex with rifampin. Adapted from Campbell *et al*. [283].

Unfortunately, these mutations occur with high frequency, leading to rifampin-resistant strains of bacteria. This has led the medical community to a voluntary restriction of its use for treatment of TB or in emergencies [283], and to a search for alternative front-line drugs. The only other antibiotics currently in use targeting bacterial RNAP are the macrocyclic polyketides known as lipiarmycins (fidaxomicins) which function by binding to the end of the RNAP 'clamp', trapping the RNAP in an open state and therefore allosterically inhibiting RNAP-DNA interaction [285]. Lipiarmycins

suffer from the same high frequency of mutations giving rise to bacterial resistance as rifampin. Therefore, there is an urgent need for new antibacterial drugs that inhibit this important target.

5.1.2: Structure and Function of Pseudouridimycin

As discussed in the previous chapter, one such molecule has been identified as a selective bacterial RNAP inhibitor. Pseudouridimycin (PUM) is a nucleoside antibiotic with the favourable characteristics of no cross-resistance with other bacterial RNAP inhibitors, additive antibacterial activity when co-administered, and spontaneous resistance rates an order-of-magnitude lower than those of rifampin [280]. Initial studies of the biochemical basis for the RNAP inhibition by PUM conclude that PUM functions as a nucleoside-analogue inhibitor that competes with UTP for occupancy of the RNAP NTP addition site [280]. Crystal studies rationalise the low spontaneous resistance rates to PUM and the selectivity for bacterial RNAP over human RNAP. PUM makes direct contact with ten functionally critical residues, the substitution of which would compromise RNAP activity, and all residues in contact with PUM are conserved in bacterial RNAP, while four important residues are not conserved in human RNAPs.

Additionally, PUM has an interesting structure (Figure 5.4), enabling structure-based design of analogues with increased potency and selectivity, and its C-nucleoside structure hints at an unusual biosynthetic pathway involving one of the family of prolific pseudouridine synthases.

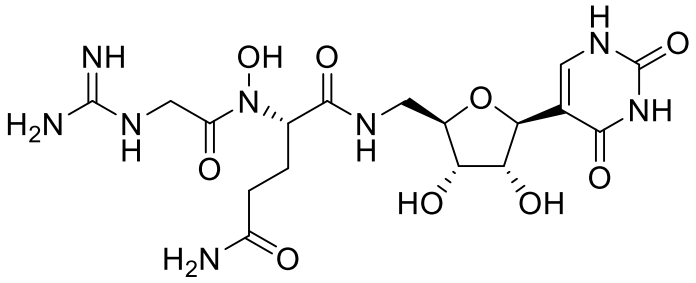


Figure 5.4: Structure of pseudouridimycin. The uracil moiety is linked to the rest of the nucleoside sugar by the unusual C-C linkage of pseudouridine.

5.1.3: Pseudouridine and Pseudouridine Synthases

The unusual C-C linked nucleotide found in pseudouridimycin is due to a pseudouridine moiety linked to the dipeptide that makes up the rest of pseudouridimycin. Pseudouridine is the most abundant posttranscriptional modification of RNA. In fact it is found in such abundance (4% of nucleotides in yeast tRNA) , that, as mentioned in Chapter 4, it was dubbed the fifth nucleotide when first discovered in 1957 [286]. The enzymatic reaction was characterised much later. Studies using RNase digestion and 2D TLC analysis of an *in vitro*-transcribed ^{14}C-uridine-labelled RNA found that *E. coli* cultural extracts would isomerise uridines that are incorporated into RNA chains, and that the modifications were only found on specific tRNA [287]. Subsequent studies demonstrated that pseudouridine synthases require no cofactor, and the uridine base is not released during the isomerisation reaction [288–290]. The reaction is proposed to proceed through a covalent intermediate formed between an aspartate residue and C6 of uridine, allowing a 180° rotation of the pyrimidine and subsequent re-ligation, resulting in the C-C glycosidic bond. A number of pseudouridine synthases have been identified, the most recent being TruD that modifies U13 in glutamic acid tRNA [270,291]. *TruD* is unique among the pseudouridine synthase family in that is shares no sequence homology with the rest of the family. In a database search, 58 homologues were found, none of which had a previously known RNA-binding motif, indicating this class of enzymes had a novel RNA-binding sequence. The only conserved aspartate in the aligned homologues was mutated and shown to be essential for activity [292]. Despite the lack of significant sequence homology, the catalytic domain of TruD is strikingly similar in its structure to the other members of pseudouridine synthases, with TruA being its closest structural homologue (Figure 5.5). The only structurally conserved residue in the active site is the catalytic aspartate, but there are other retained structural features involved in recognition and catalysis [291]. *TruD* has an additional domain, unlike other pseudouridine synthases. This insertion domain has no significant structural similarities to any PDB entry and has a strong positive surface charge on the side facing the catalytic domain. This domain is most likely involved in RNA binding, and accounts for the lack of any other RNA binding motif in TruD. A catalytic cleft is formed in between the insertion and catalytic domains with the catalytic Asp protruding from one face of the cleft where, presumably, the tRNA binds into the positively charged pocket placing U13 in the correct position to undergo isomerisation.

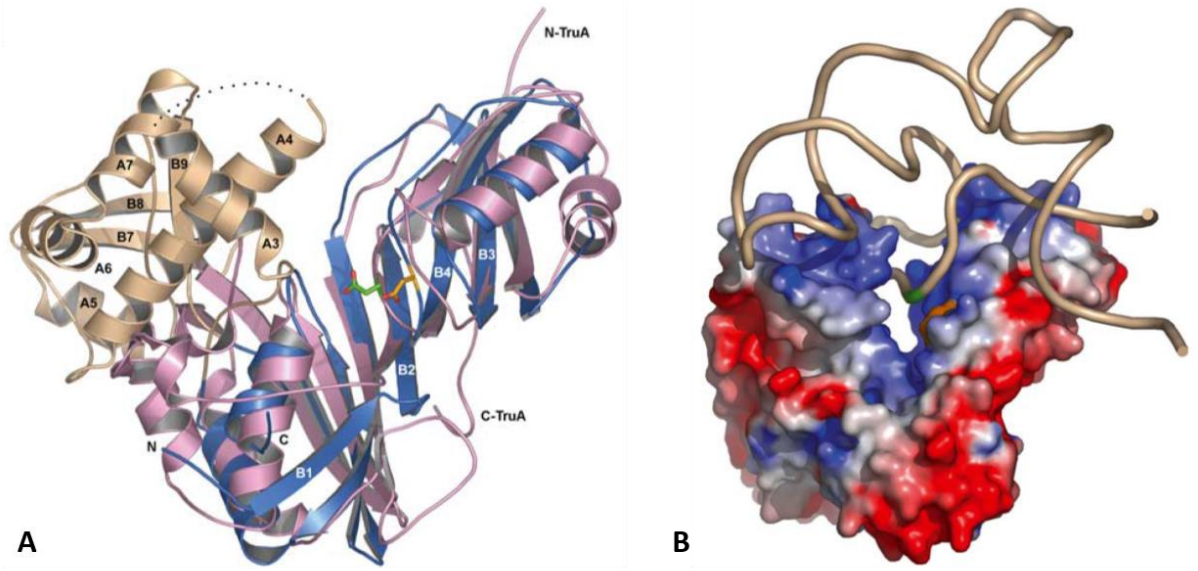


Figure 5.5: Structure and catalytic activity of *truD*. (A) Superposition of TruD and its closest structural homologue, TruA. TruD is coloured in blue (catalytic domain) and beige (insertion domain) and TruA is coloured in pink. **(B)** A rigid docking model of TruD with $tRNA^{Glu}$. The surface of the catalytic Asp80 is coloured in orange and the U13 to be modified is coloured in green. Adapted from Ericsson *et al*. [291].

The work in this Chapter describes an analysis of the biosynthetic pathway to PUM and in particular attempts to reconstitute parts of the pathway *in vitro*, to pave the way for future knowledge-based generation of analogues of this bioactive C-nucleoside.

5.2: Results and Discussion

5.2.1: Identification and Characterisation of the Pseudouridimycin Biosynthetic Gene Cluster

The structure of PUM had been characterised by Maffioli *et al.* [280] in 2017. The obvious choice to probe for the PUM cluster in the genome sequence of *S. albus* A980 was therefore a pseudouridine synthase. As noted in Chapter 4, probing the A980 genome sequence with a TruD-like sequence from malayamycin producers being studied in this laboratory produced a single definite hit. The sequence of TruD from yeast was also used to probe the entire Leadlay lab strain collection. This returned five hits; *Saccharopolyspora cavernae* DSMZ 45825, *Nocardia* sp. NBRC 14326 (a thiolactomycin producer), the two malayamycin producers *Streptomyces malaysiensis* DSM 14702 and *Streptomyces chromofuscus* ATCC 49982, and *Streptomyces albus* DSM 40763 (A980). Looking at the genes

surrounding the TruD homologue in A980 revealed many biosynthetic genes that could be involved in PUM biosynthesis. The putative biosynthetic gene cluster (BGC) is illustrated in Figure 5.6.

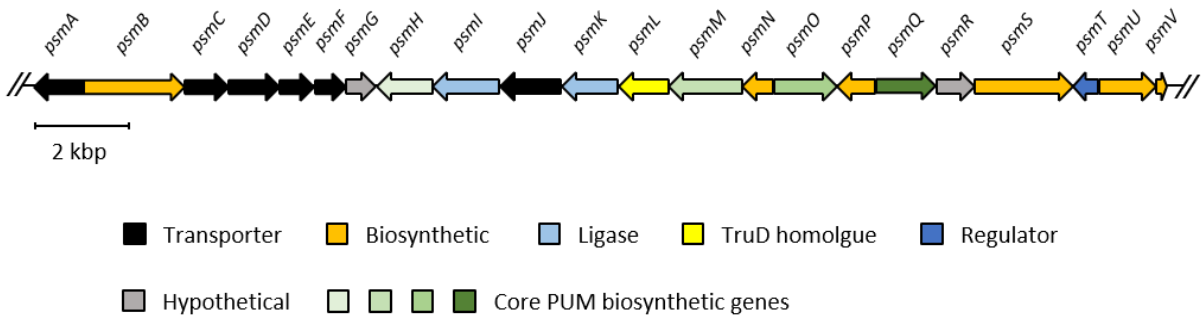


Figure 5.6: Organisation of the putative pseudouridimycin biosynthetic gene cluster.

Once the candidate bioynthetic gene cluster had been identified in A980, the encoded enzymes could be scrutinised for their fit with an initial proposal for the biosynthetic pathway. The structure of PUM revealed three units linked together by amide bonds (Figure 5.7); an N-terminal guanidinoacetate, N-hydroxylated glutamic acid, and C-terminal 5'-amino-pseudouridine (APU). If the N-hydroxylation occurs as the final step in the pathway, which is chemically reasonable, glutamic acid is recruited directly from primary metabolism. Guanidinoacetate is known to be the product of an amidinotransferase reaction between arginine and glycine [293] again directly from primary metabolism. The C-terminal component, APU, might be derived from pseudouridine (Ψ) via oxidation of Ψ at the 5' position to form an aldehyde, coupled to the action of an aminotransferase to generate APU. Analogous oxidation-aminotransfer steps are widespread in aminoglycoside biosynthetic pathways [294]. The origin of the Ψ substrate remains obscure. If a TruD homologue in A980 catalyses the conventional TruD-catalysed reaction, post-transcriptional modification of an ncRNA, there must be a mechanism by which Ψ can be subsequently be released from the ncRNA. This could be as part of normal RNA turnover, but nothing is known in detail about this in *Streptomyces*. Alternatively, the Ψ required for PUM biosynthesis is generated by a different mechanism. At any rate, a coupled oxidoreductase, aminotransferase reaction on Ψ remains a plausible way to generate APU. Alternatively, an Fe(II)-2-oxoglutarate-dependent oxygenase enzyme in capuramycin biosynthesis has been described that converts uridine 5'-monophosphate to uridine 5'-aldehyde in a single step [295]. Conceivably an analogous enzyme could convert 5'-ΨMP directly to the 5'-aldehyde. The three components of Figure 5.7 must be linked together, which likely requires two different two ligase

enzymes, and finally the glutamate amide nitrogen must be hydroxylated to complete the biosynthesis.

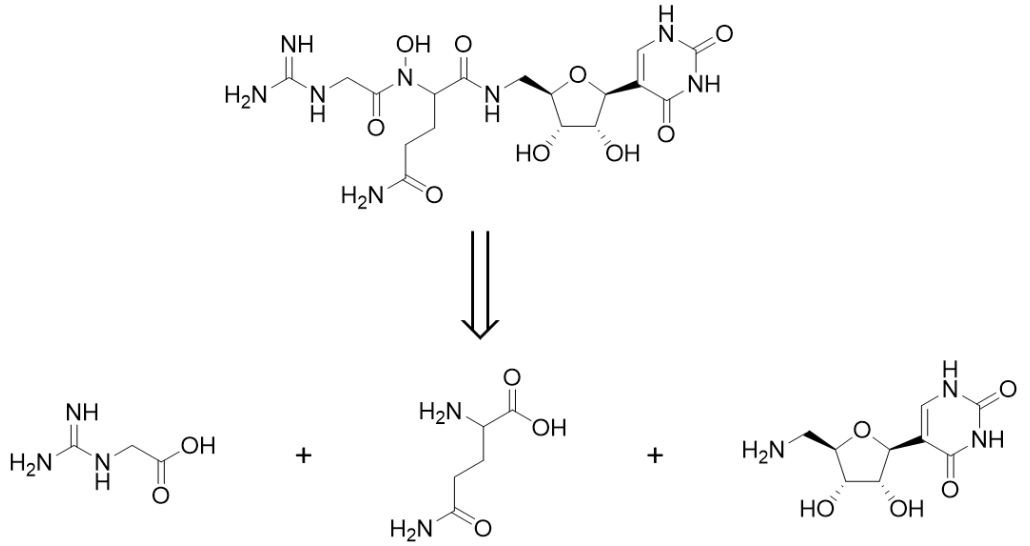


Figure 5.7: Retrobiosynthetic analysis of pseudouridimycin.

The putative PUM cluster was manually curated. Its boundaries were provisionally defined when genes were encountered that were unlikely to be involved in PUM biosynthesis. The putative cluster covers 26 kbp and contains 22 protein coding sequences. It is summarised in Table 5.1.

Table 5.1: Summary of the putative pseudouridimycin biosynthetic gene cluster. The function of each gene has been identified by comparing its sequence against the NCBI database using a BLAST search. Similarities to homologues of known functions have been included, as well as the predicted function of the gene in the PUM cluster. Pathway enzymatic activities are highlighted in bold.

Gene	No. AAs	Protein homologue	Accession number	Identity/similarity	Predicted function
psmA	351	ABC transporter substrate-binding protein, *Streptomyces sp*. NRRL F-5639	WP_030405507.1	99/100	export

psmB	701	hydrolase CocE/NonD family protein, *Streptomyces sp.* HPH0547	WP_016471205.1	100/100	biosynthetic
psmC	303	substrate-binding region of ABC-type glycine betaine transport system, *Conexibacter woesei* DSM 14684	ADB51680.1	47/62	export
psmD	355	glycine/betaine ABC transporter, *Streptomyces sp.* NRRL F-6602	WP_030405504.1	99/99	export
psmE	237	ABC transporter permease, *Streptomyces sp.*	WP_030886564.1	96/96	export
psmF	216	ABC transporter permease, *Streptomyces sp.* NRRL F-5053	WP_030895287.1	93/95	export
psmG	204	Hypothetical Protein	-	-	-
psmH	384	glycine amidinotransferase, *Streptomyces sp.* NRRL F-5639	WP_031027661.1	99/100	**Formation of guanidinoacetate, N-terminal component**
psmI	463	D-ala-D-ala ligase, *Enterorhabdus mucosicola*	WP_028026109.1	29/37	**Ligation of the C-terminal and central components**
psmJ	425	membrane transporter, *Streptomyces sp.* NRRL F-5639	WP_031027667.1	99/99	export
psmK	394	NikS-like protein, *Actinomyces sp.* Lu 9419	ACH85565.1	55/65	**Ligation of the central and C-terminal components**
psmL	345	TruD family tRNA pseudouridine synthase, *Streptomyces sp.* HPH0547	WP_016471195.1	100/100	**Generates Ψ residues embedded in tRNA?**
psmM	509	glucose-methanol-choline (GMC) oxidoreductase,	KPC68900.1	100/100	**5' oxidation of PU**

		Streptomyces sp. NRRL F-6602			
psmN	212	adenylate kinase, *Streptomyces mobaraensis*	WP_004951438.1	69/76	unknown
psmO	440	aminotransferase, *Streptomyces sp.* NRRL F-6602	KPC69014.1	99/100	**Generates APU, C-terminal component**
psmP	267	serine/threonine protein kinase, *Streptomyces griseoflavus*	WP_040908662.1	55/65	unknown
psmQ	408	acyl-CoA dehydrogenase, *Streptomyces sp.* NRRL F-5917	WP_030405496.1	99/99	**Glutamate N-hydroxylation candidate**
psmR	264	UPF0317 protein, Streptomyces sp. HPH0547	EPD92375.1	100/100	unknown
psmS	685	peptidase M4, *Streptomyces albus*	WP_037611685.1	100/100	unknown
psmT	178	TetR family transcriptional regulator, *Streptomyces sp.* NRRL F-5639	WP_031027676.1	99/100	regulator
psmU	396	cytochrome P450, *Streptomyces sp.* CNT360	WP_051352406.1	86/92	**Glutamate N-hydroxylation candidate**
psmV	72	cytochrome P450, *Streptomyces varsoviensis*	WP_030874200.1	81/91	unknown

As noted in Table 5.1, there is an excellent fit between a number of the genes in the candidate cluster and the enzymology expected for the otlined PUM biosynthetic pathway. No other locus in the *S. albus* A980 genome matchedf these reqirements. With the biosynthetic gene cluster for PUM confidently identified, its sequence was used to design *in vitro* experiments to obtain biochemical evidence for the proposed pathway. While this was carried out, the Ebright and Donadio groups followed up their Cell paper (2017) on the PUM target by independently reporting in 2018 the PUM cluster from an unrelated *Streptomyces* spp., and they generated several deletions to investigate the sequence of biosynthesis.[296] Their results will be discussed in full, in the context of my own results, later in this chapter. A comparison of the Sosio *et al.* PUM cluster [296] with that in A980 is summarised in Figure 5.8. It reveals an essentially identical, but inverted, core sequence of genes (*psmG* to *psmQ*), with some minor differences in the rest of the cluster.

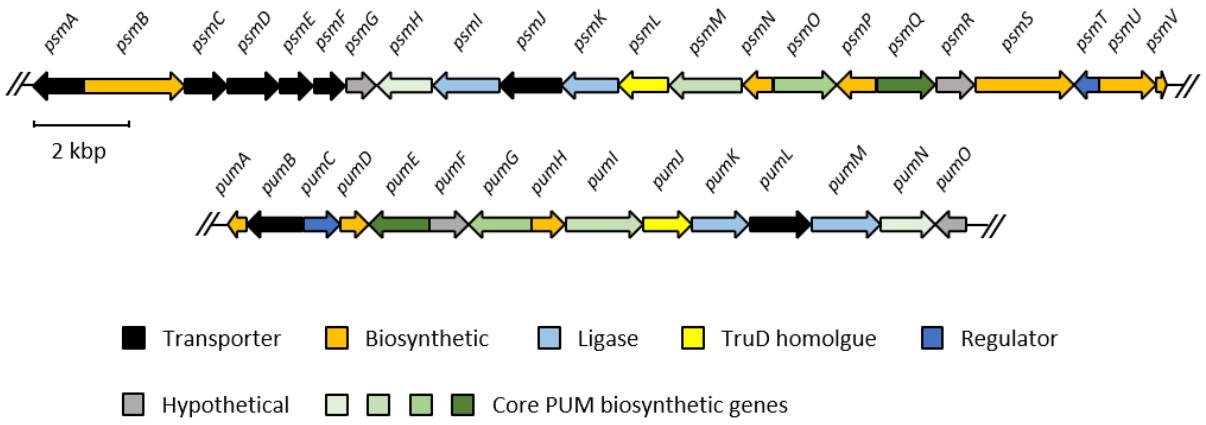


Figure 5.8: Comparative sequence analysis of the pseudouridimycin biosynthetic gene cluster in A980 with the biosynthetic gene cluster characterised by Sosio *et al* [296]. The Sosio cluster is smaller, and the A980 gene cluster contains several biosynthetic genes that are only putatively involved in PUM biosynthesis. Further work is required to define the boundaries of this cluster. The clusters are not identical, the 5' and 3' ends hold different genes but the core biosynthetic genes, transporter genes, and other hypothetical genes (*psmG* to *psmQ* and *pumE* to *pumO*) are identical but inverted with respect to each other.

5.2.2: *In vitro* Assays

The *in vivo* biosynthesis of PUM appears to follow a convergent path: the building blocks guanidinoacetate and APU are synthesised and then ligated to glutamic acid (Figure 5.7). If the building blocks could be obtained enzymatically, it was hoped then to use recombinant candidate ligases PsmL and PsmK to generate deoxy-PUM *in vitro* (Figure 5.10). *In vitro* assays were designed to recapitulate formation of the building blocks other than glutamate. This would provide substrates for the putative ligases. Also, this would open the way to testing alternative substrates, especially those with an altered nucleoside moiety.

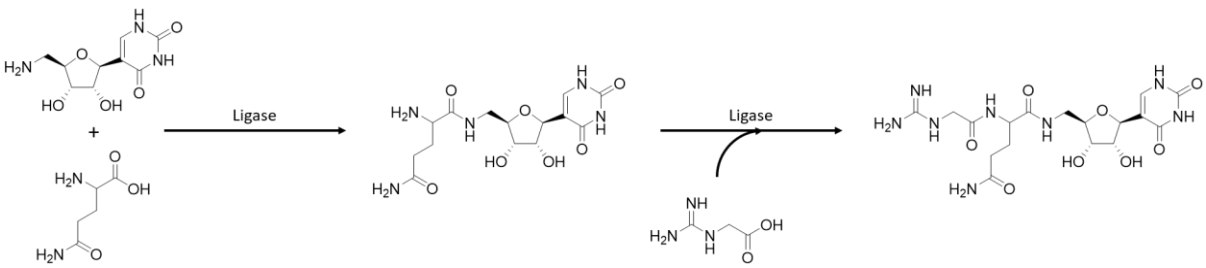


Figure 5.9: Reactions between guanidinoacetate, glutamic acid, and 5'-aminopseudouridine catalysed by the products of the *psmI* and *psmK* genes.

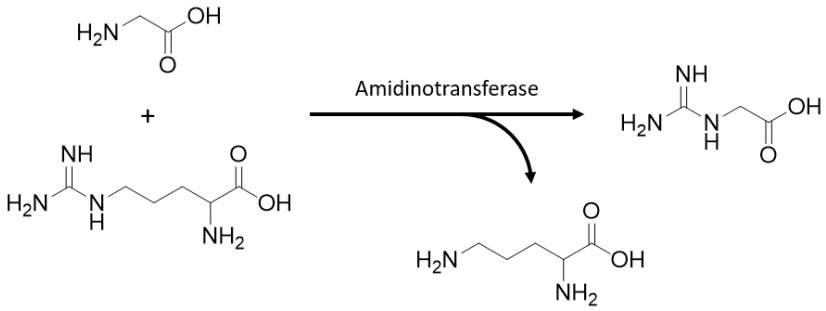


Figure 5.10: Reaction between arginine and glycine catalysed by PsmH.

The production of guanidinoacetate from arginine and glycine is a well-known reaction. [297] Testing whether PsmH catalyses this reaction (Figure 5.10) was expected to be straightforward. The recombinant PsmH required was obtained by cloning and expressing the gene in *E. coli*, and purification of the N-terminally His_6-tagged enzyme by affinity chromatography, as detailed later in this section. The assay conditions were based on work by Li et al. [297]. The conditions are summarised in Table 5.2, and the reaction was performed in a 0.05M phosphate buffer with 0.1M NaCl at pH 7.4. The reaction mixture was incubated at 30°C for 2 hours

Table 5.2: Reaction conditions for PsmH assay

Component	Volume (µl)	Final concentration
PsmH	100	15 µM
glycine	20	1 mM
arginine	20	1 mM
phosphate buffer	60	0.05 M

Since the substrates and products have no chromophore, the products of the reaction were initially tested for using Marfey's reagent. Marfey's reagent (1-fluoro-2-4-dinitrophenyl-5-L-alanine amide) reacts with primary amines (Figure 5.11) and is commonly used as a derivatisation agent for the UV detection of amino acids and for the resolution of chiral mixtures of amino acids [298]

The derivatisation worked well, and could be used to detect glycine, arginine, and one of the products, ornithine. However, after some trials using standard samples of the starting materials and products with TLC and HPLC (not shown) ^{1}H NMR was used instead to visualise the progress of the reaction, with the major advantage that the reaction could be monitored *in situ* as it progressed. The ^{1}H NMR analysis was performed as described in section 2.4.4. Spectra for the starting materials (arginine and glycine) and for the products (ornithine and guanidinoacetate) were recorded first.

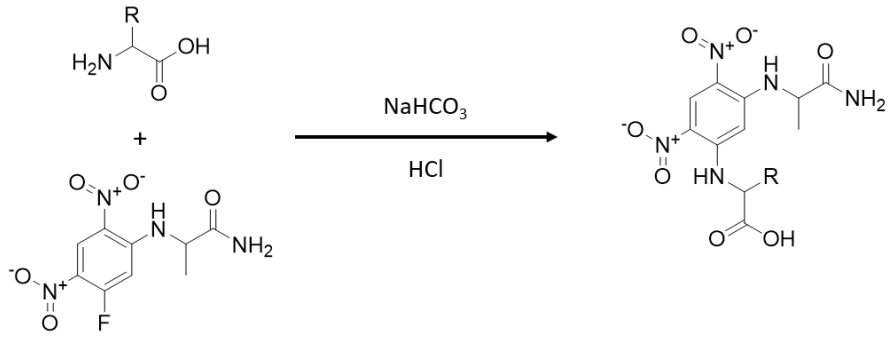


Figure 5.11: Reaction between Marfey's reagent and a primary amine.

These spectra are shown in Figure 5.12. The CH2 region of the spectrum (Figure 5.12A) shows signals from arginine and ornithine, but not from glycine or guanidinoacetate, as expected. The difference in the chemical shifts of the arginine and ornithine signals could be used to track the reaction, but guanidinoacetate gives a unique singlet at 3.8 ppm in the Hα region (Figure 5.22B) which was the most convenient way to monitor its appearance (Figure 5.13).

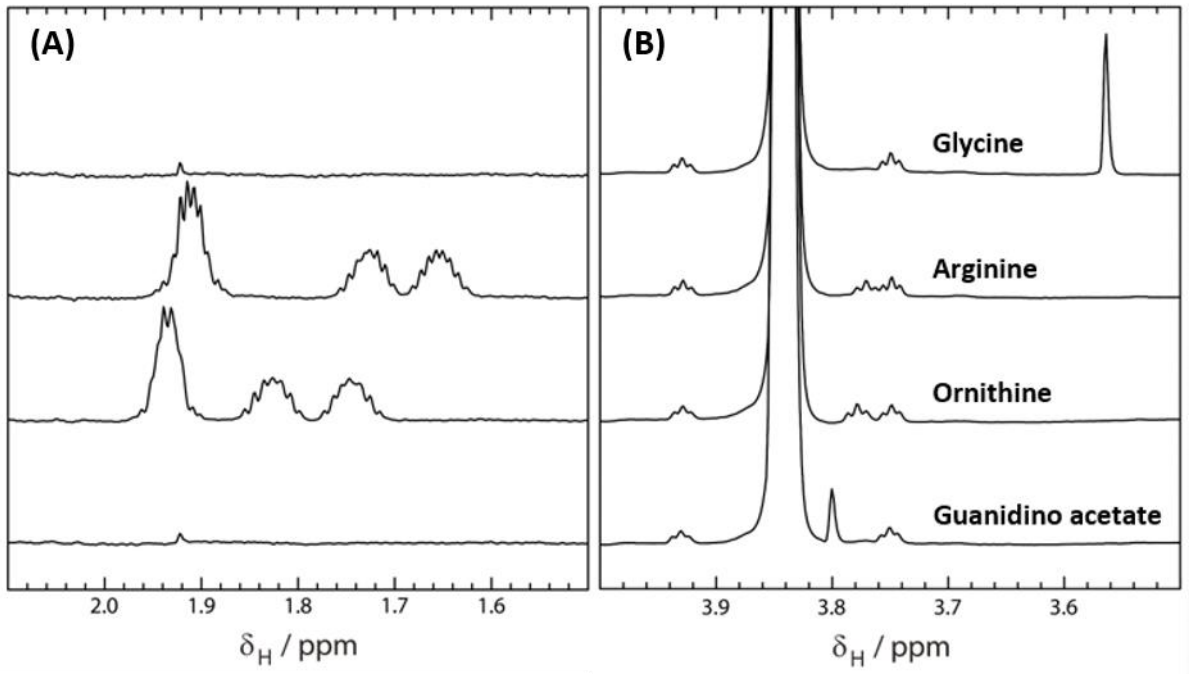


Figure 5.12: NMR spectra of starting materials and products in the arginine:glycine amidinotransferase-catalysed reaction. Portions of the ^{1}H NMR spectra of glycine, arginine, ornithine and guanidinoacetate are shown. **(A)** CH2 region (1.5 – 2.1 ppm); **(B)** Hα region (3.5 – 4.0 ppm).

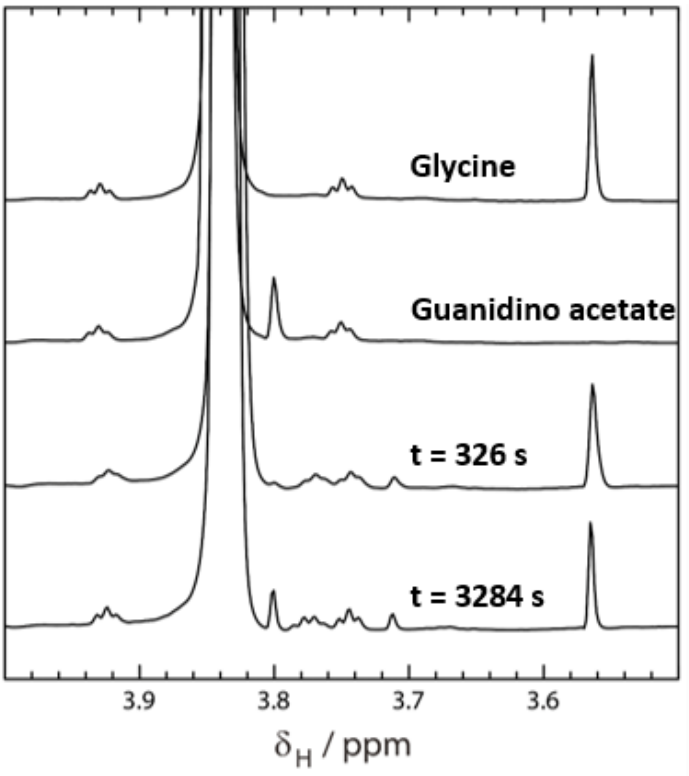


Figure 5.13: The *psmH*-catalysed reaction monitored by the appearance of a signal at 3.8 ppm. ^{1}H NMR spectra showing the substrate glycine, the product guanidinoacetate and the spectra of the assay mixture at the first and last time points.

The NMR data showed that guanidinoacetate is indeed generated by the reaction between glycine and arginine catalysed by PsmH and confirms the identity of this building block. As guanidinoacetate is readily commercially available, there is obviously no need to use the recombinant PsmH to generate it.

If pseudouridine (Ψ) is the correct starting material for biosynthesis of the APU building block, it can be converted to APU in a two step, coupled oxidation/transamination reaction. The oxidation of a hydroxy group followed by transamination is a well-known coupled reaction in *Streptomyces* biosynthesis (Figure 5.14). To produce the Ψ starting material *in vitro*, advantage would be taken of a pathway already found useful in this laboratory in the study of malayamycin biosynthesis. YeiN from *E. coli* is a pseudouridine-5'-phosphate glycosidase, normally involved in recycling of Ψ. Recombinant YeiN (the kind gift of Dr Hui Hong) also catalyses the reverse reaction between ribose-5'-phosphate (R5P) and uracil and generates pseudouridine-5'-phosphate (5'-ΨMP) [299] which may be subsequently dephosphorylated by a commercially-available phosphatase enzyme to give Ψ (Figure 5.15).

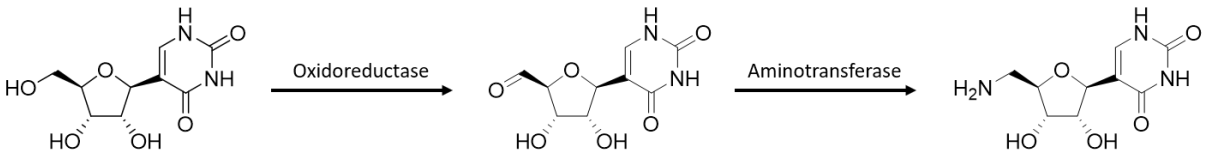


Figure 5.14: Conversion of pseudouridine to 5'-aminopseudouridine proposed to be catalysed by PsmM and PsmO.

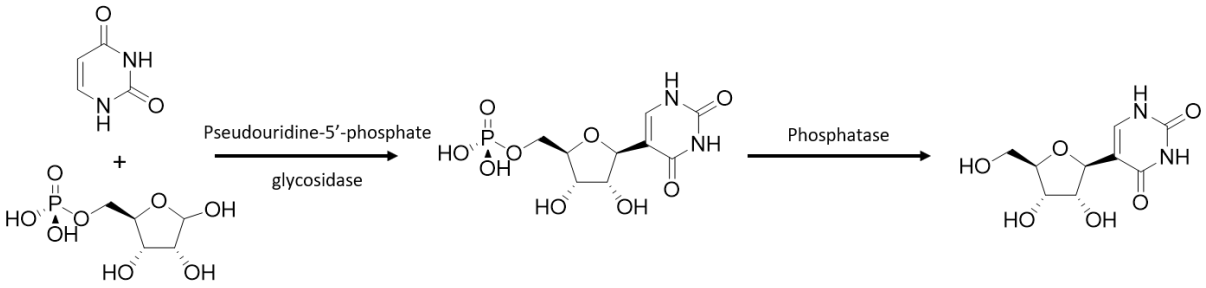


Figure 5.15: Reaction between ribose-5'-phosphate and uracil catalysed by YeiN, followed by dephosphorylation using calf intestinal phosphatase.

The use of recombinant PsmH has already been described above. Other recombinant enzymes required for the *in vitro* assays were: flavin-dependent GMC oxidase PsmM; aminotransferase PsmO; and putative ligases PsmK and PsmI. All these genes were cloned for *E. coli* expression in pET28a as follows (see also Figure 5.15). First, primers were designed to amplify each gene from the PUM cluster, and provide overhanging 5' ends, homologous to the adjacent sequence when assembled. The purified PCR products were assembled together with linearized, *Nde*I- and *Eco*RI-cut pET28a using the Gibson isothermal assembly method. The products were used to transform *E. coli* DH10B, and colonies were screened by PCR for the presence of the desired insert.

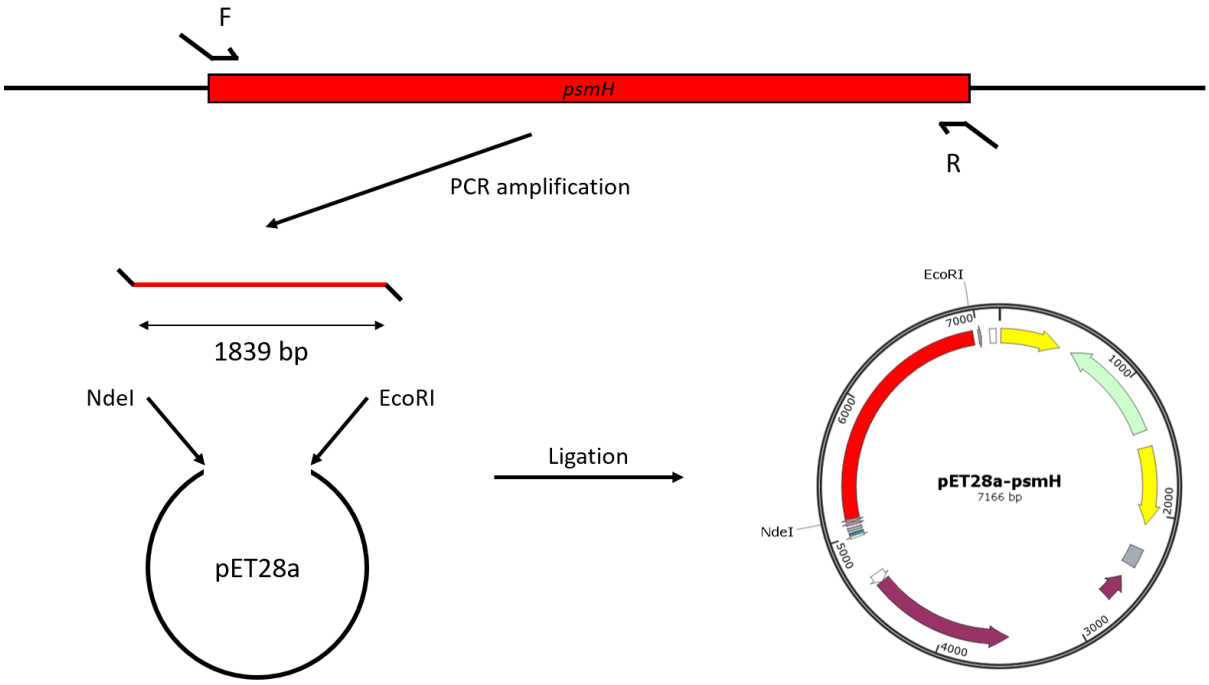


Figure 5.16: Cloning of *psm* genes in pET28a for expression in *E. coli*

The identity of the PCR products and the final vector inserts was confirmed by gel electrophoresis and sequencing. The details of the procedure are given in section 2.3. Figure 5.17 and Figure 5.18 show images of the results of gel electrophoresis from the initial PCR and from the screening of *E. coli* clones, respectively, for the cloning of *psmH* into pET28a. The other genes cloned gave comparable results. Figure 5.19 shows the results of screening of *E. coli* clones from each of the five gene expression vectors constructed. The analogously-constructed pET28a-based vector for YeiN was the kind gift of Dr Hui Hong.

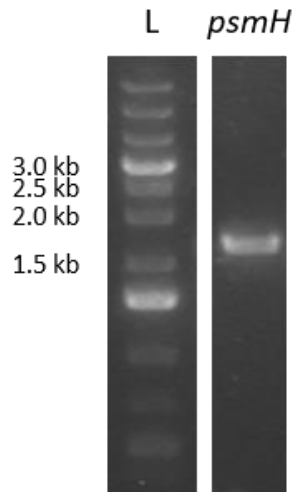


Figure 5.17: Gel electrophoresis image of the *psmH* amplified PCR product. The 'L' lane contains the DNA ladder (1 Kb Plus DNA Ladder). The numbers to the side of the ladder refer to the size of each relevant band in the ladder. The *psmH* lane contains the PCR product amplified for the construction of the PsmH expression vector.

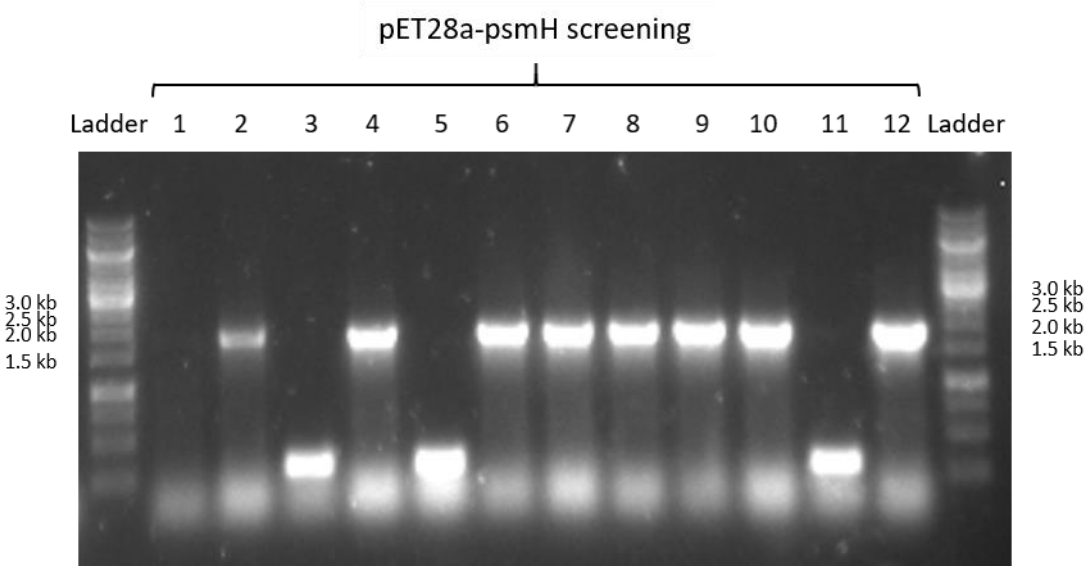


Figure 5.18: Gel electrophoresis image of the cPCR screening to confirm correct assembly of the PsmH expression vector. The 'Ladder' lane contains the DNA ladder (1 Kb Plus DNA Ladder). The numbers to the side of each ladder refer to the size of each relevant band in the ladder. Each numbered lane refers to a colony of E. coli that was screened for the correct expression vector. The lanes 2,4,6-10, and 12 contain bands with the correct size for the PsmH expression vector.

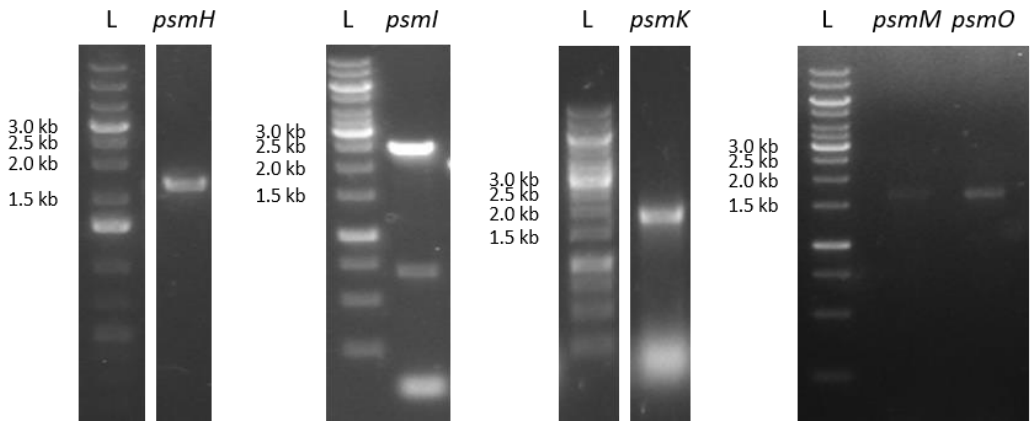


Figure 5.19: Gel electrophoresis images of the inserts from each of the four *psm* expression vectors. The 'L' lane contains the DNA ladder (1 Kb Plus DNA Ladder). The numbers to the side of each ladder refer to the size of each relevant band in the ladder. The *psmH-O* lanes contain the PCR product amplified for the construction of each expression vector.

Once the inserts had been successfully cloned into pET28a vectors they could be transformed into *E. coli* BL21 and induced with IPTG to overexpress the desired protein, which was then recovered and purified as described in section 2.3.12,13. Figure 5.20 shows an SDS-PAGE gel of the purification of PsmM and Figure 5.21 shows the elution fractions collected and concentrated from each protein, to be used in the *in vitro* assays.

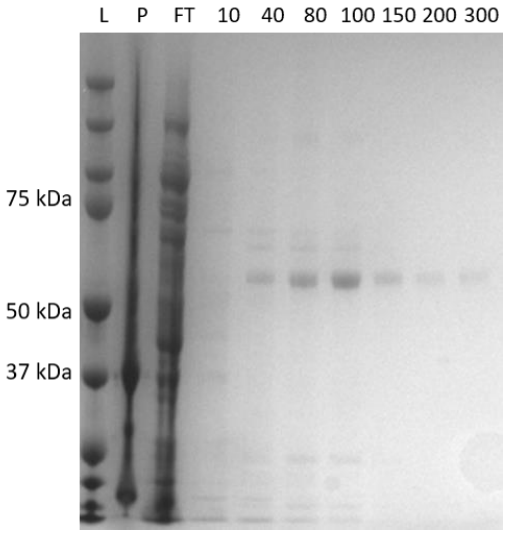


Figure 5.20: SDS-PAGE gel image showing the process of purifying PsmM from the expression culture. The 'L' lane contains the protein size ladder (Precision Plus Protein Standards). The numbers to the side of the ladder refer to the size of each relevant band in the ladder. 'P' and 'FT' refer to the insoluble (pellet) and soluble (flow-through) fractions of the purification, respectively. The numbers above each of the remaining lanes refer to the concentration of imidazole in the buffer used to elute the fraction that has been loaded onto each lane.

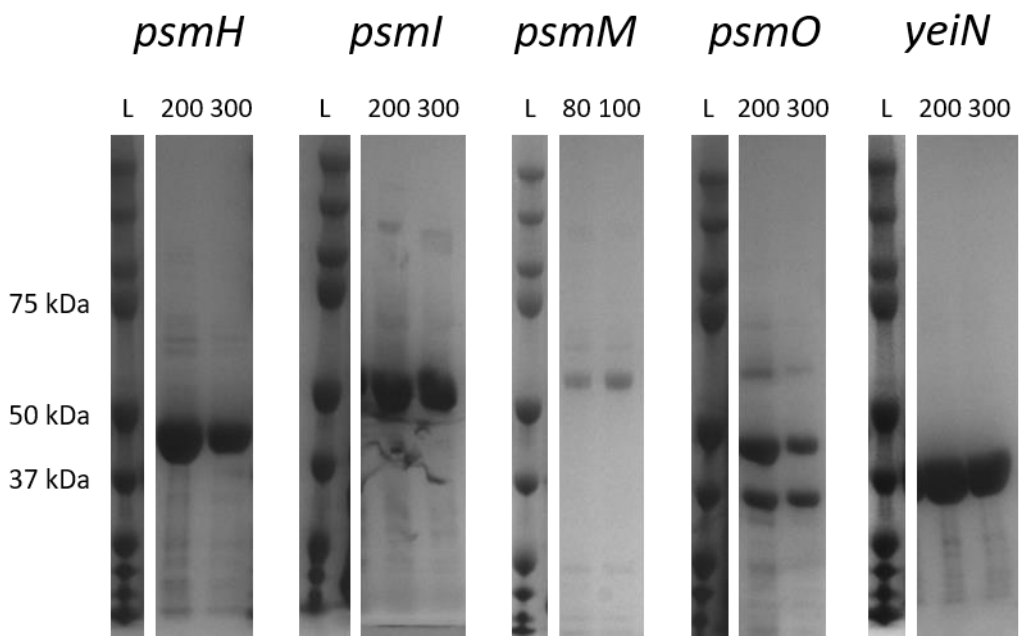


Figure 5.21: SDS-PAGE gel images showing the elution fractions collected for each of the six proteins for the *in vitro* assays. The 'L' lane contains the protein size ladder (Precision Plus Protein Standards). The numbers to the side of each ladder refer to the size of each relevant band in the ladder. The numbers above each pair of lanes refer to the concentration of imidazole in the buffer used to elute the fraction that has been loaded onto each lane. All enzymes were recovered in high yield. PsmM was recovered in the lowest yield, but could be concentrated to 50 µM, which is more than enough for the required assays.

Each protein eluted in 2 ml of the imidazole wash buffer. From the size of the bands on the SDS-PAGE gel, it is clear that some expressed better than others. YeiN consistently expressed very well and yielded higher concentrations of protein than the others, but this is unsurprising as it is an *E. coli* protein. PsmM tended to express poorly, and PsmO always eluted with another smaller protein of approximately 37 kDa. The smaller protein was not identified but may be a truncated version of PsmO. For the highly expressed proteins, such as YeiN, a single elution fraction was collected and concentrated using a 30K Merck Millipore centrifugal filter unit. For poorly-expressed proteins, several fractions were combined before concentration. The wash buffer was replaced with a suitable reaction buffer through a series of concentration and dilution steps using the desired buffer. Unfortunately, PsmK did not express in *E. coli* BL21, despite several attempts, using different expression conditions. Sequencing confirmed the cloning had been correct, so the reason for this is unclear. Figure 5.22 shows an SDS-PAGE gel analysis of the soluble and insoluble fractions of a culture of BL21 transformed with pET28a-psmK.

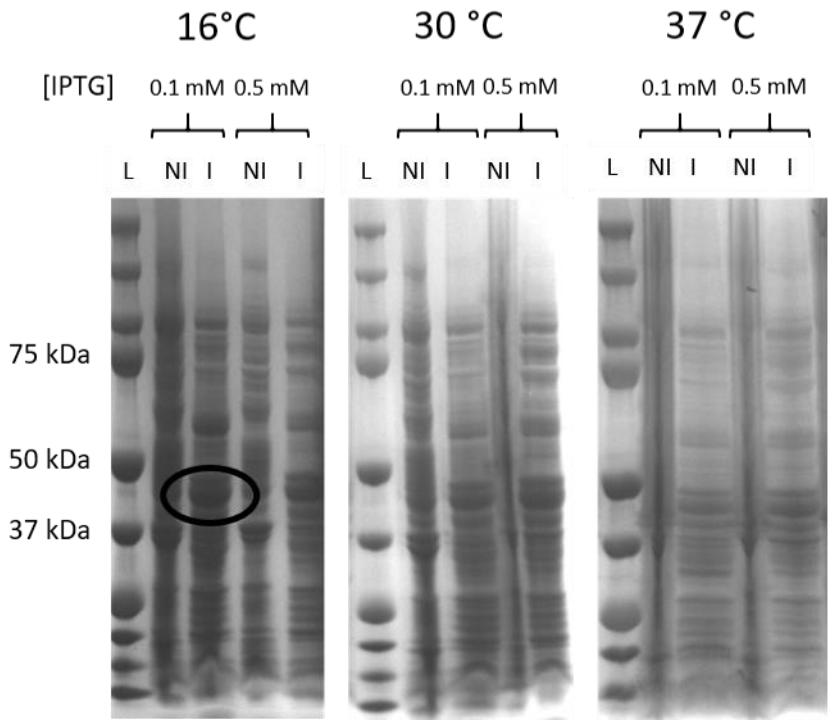


Figure 5.22: SDS-PAGE gel image of the soluble and insoluble fractions of a culture of BL21 transformed with pET28a-psmK. The 'L' lane contains the protein size ladder (Precision Plus Protein Standards). The numbers to the side of each ladder refer to the size of each relevant band in the ladder. 'NI' and 'I' lanes refer to 'Not Induced' and 'Induced' with IPTG at the concentrations indicated (0.1 mM and 0.5 mM). The temperatures refer to the temperature the expression culture was incubated at.

Attempted formation of 5'-amino-pseudouridine (APU) in vitro

It was thought best to combine the action of the flavin-dependent GMC oxidase PsmM with that of aminotransferase PsmO, to prevent problems, such as over-oxidation or product inhibition, arising from the build-up of the reactive aldehyde product of the oxidase. Different amino acids needed to be tried as amino donors in the PsmO-catalysed reaction. The substrate for the oxidase would be generated *in situ* by the joint action of YeiN and calf intestinal phosphatase. The principle of the assay is summarised in Figures 5.11 and 5.12. The reaction conditions used are summarised in Table 5.3, and the reaction was performed in a 0.05M phosphate buffer with 0.1M NaCl at pH 7.4. The reaction mixture was incubated at 30°C for 2 hours for each step. Between each reaction chloroform was added to precipitate the protein. The supernatant was then removed and used in the next reaction. The progress of the reaction was monitored by following the UV absorbance of the nucleobase, and by NMR. Figures 5.23 to 5.27 summarise the results of the NMR and HPLC analysis of these reactions.

Table 5.3: Reaction conditions for attempted enzymatic synthesis of APU

Component	Volume (µl)	Final concentration
YeiN reaction, illustrated in Figure 5.24.		
uracil	5	1 mM
ribose-5'-phosphate	5	1 mM
YeiN	5	7 µM
Phosphate buffer	485	0.05 M
Phosphatase reaction, illustrated in Figure 5.25.		
Calf intestinal phosphatase	5	50 units
10X Reaction buffer	50	1X
Coupled oxidation-aminotransferase reaction, illustrated in Figure 5.26		
PsmM	5	3 µM
PLP	50	1 mM
PsmO	5	5 µM
Amino acid donor (Ala/Asp/Glu)	5	10 mM

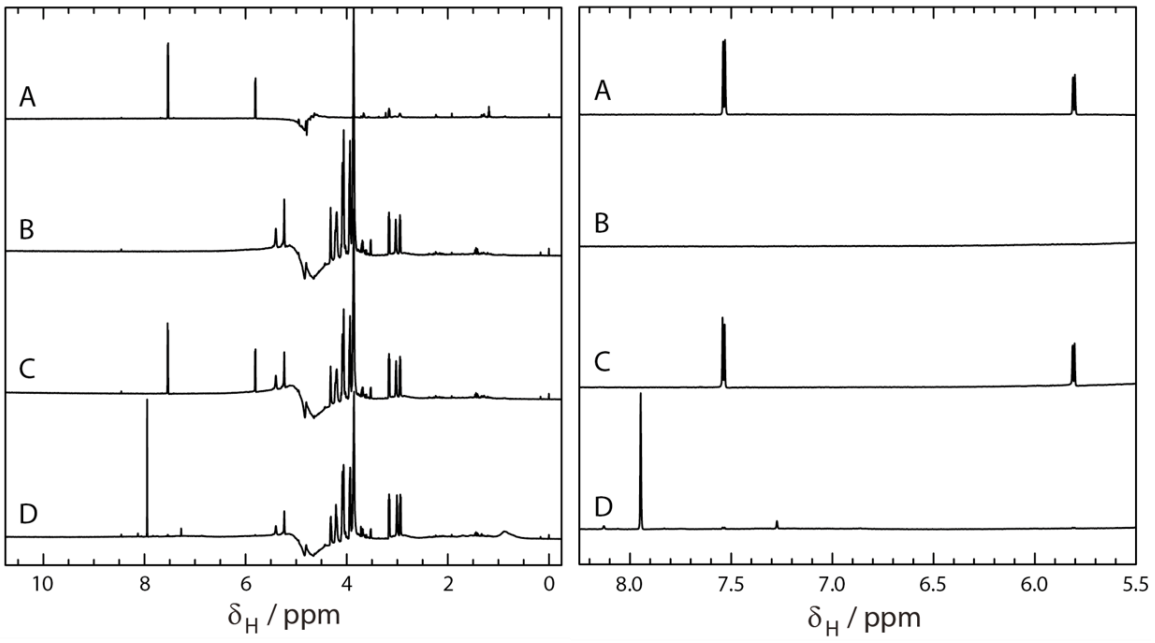


Figure 5.23: NMR spectra of the reaction between uracil and ribose-5'-phosphate, catalysed by YeiN. The left-hand panel shows the entire spectrum, while the right-hand panel focusses on the high field region where the aromatic protons signals can be found. (A) ^{1}H NMR spectrum of uracil. **(B)** ^{1}H NMR spectrum of R5P. **(C)** ^{1}H NMR spectrum of uracil and R5P. **(D)** ^{1}H NMR spectrum of uracil, R5P and YeiN.

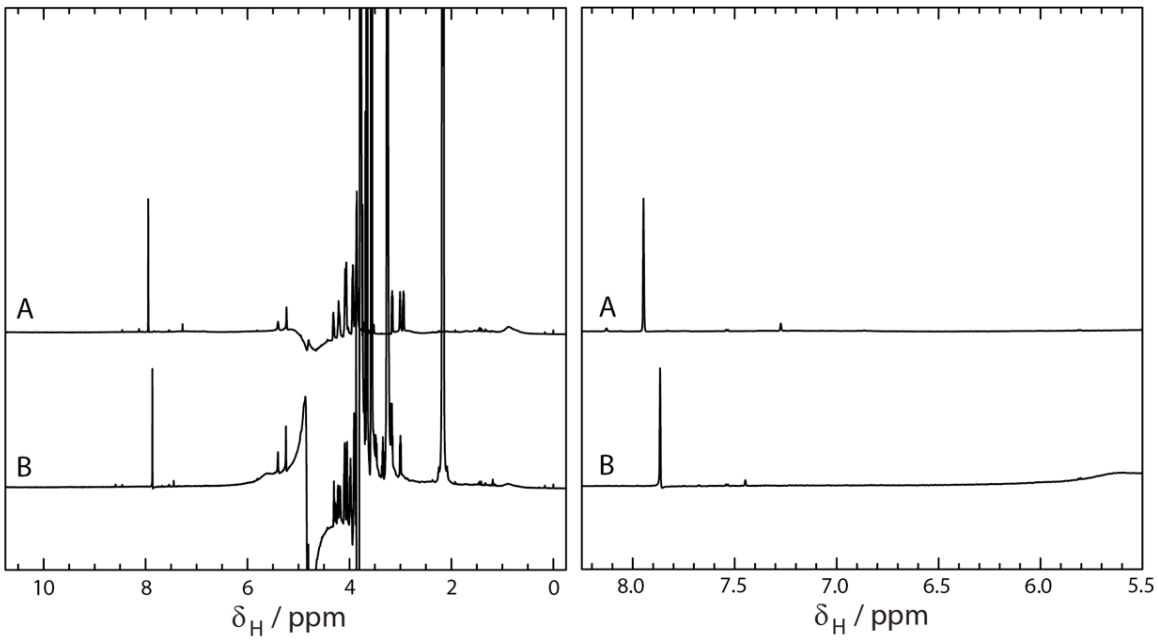


Figure 5.24: NMR spectra of the dephosphorylation reaction of pseudouridine-5'-phosphate, catalysed by Calf Intestinal Phosphatase. The left-hand panel shows the entire spectrum, while the right-hand panel focusses on the high field region where the aromatic protons signals can be found. (A) 1H NMR spectrum of uracil, R5P and *yeiN*. **(B)** 1H NMR spectrum of (A) after the subsequent addition of Calf Intestinal Phosphatase. The additional peaks in (B) are the result of the Tris-HCl buffer

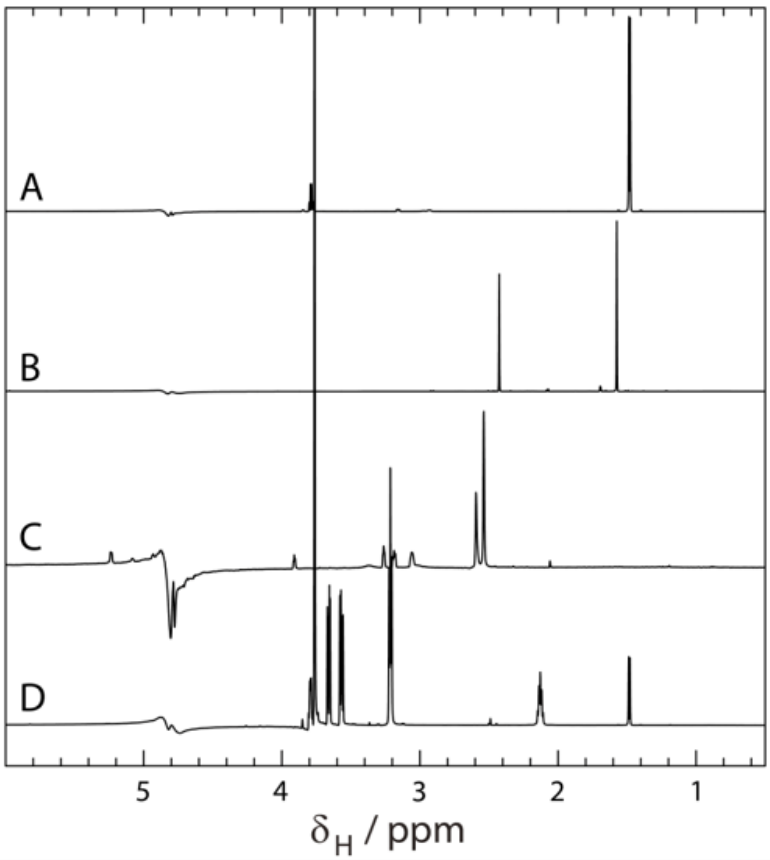
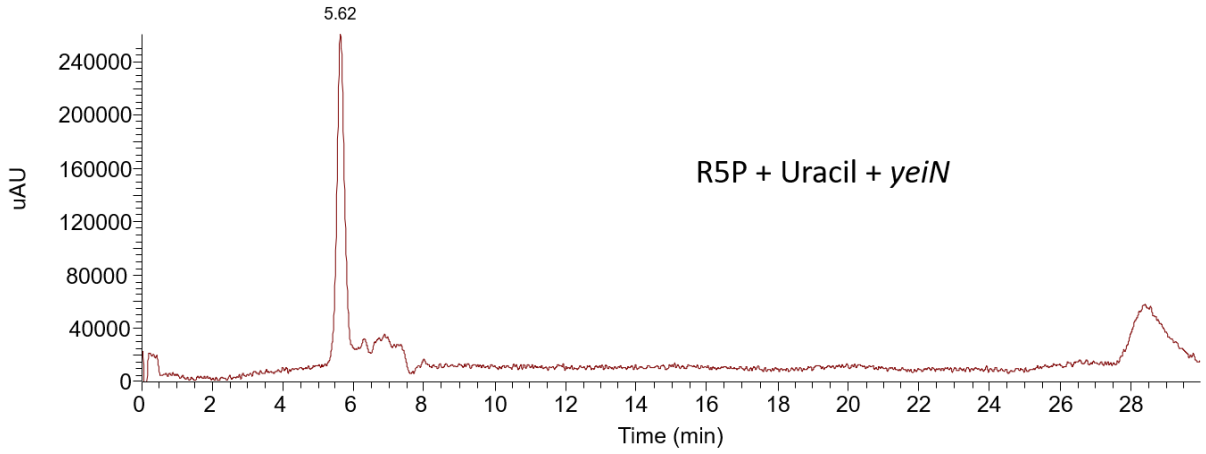


Figure 5.25: NMR spectra of the coupled oxidation and aminotransferase reaction of pseudouridine, catalysed by PsmM and PsmO, respectively. (A) ^{1}H NMR spectrum of alanine; **(B)** ^{1}H NMR spectrum of pyruvate; **(C)** ^{1}H NMR spectrum of PLP (cofactor to PsmO); **(D)** ^{1}H NMR spectrum of reaction mixture (see Table 5.3) using alanine as an amino donor. The small signal in (D) at approximately 2.5 ppm could potentially have been caused by the production of a small amount of pyruvate, indicating the coupled reaction worked to some extent. The large signal in (D) at approximately 3.8 ppm is an unknown contaminant.

(A)



(B)

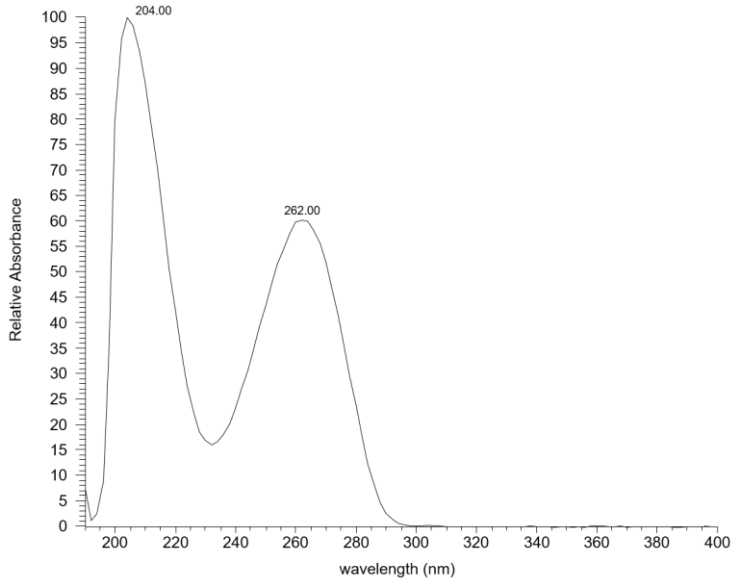


Figure 5.26: UV absorption spectrum showing the product of the reaction between ribose-5'-phosphate and uracil, catalysed by YeiN. (A) The peak at 5.62 min is Ψ produced by the YeiN catalysed reaction. Absorbance measured at 262 nm; **(B)** The characteristic absorption signature of the uracil moiety at 262 nm is the basis for the UV/Vis characterisation of Ψ and related compounds.

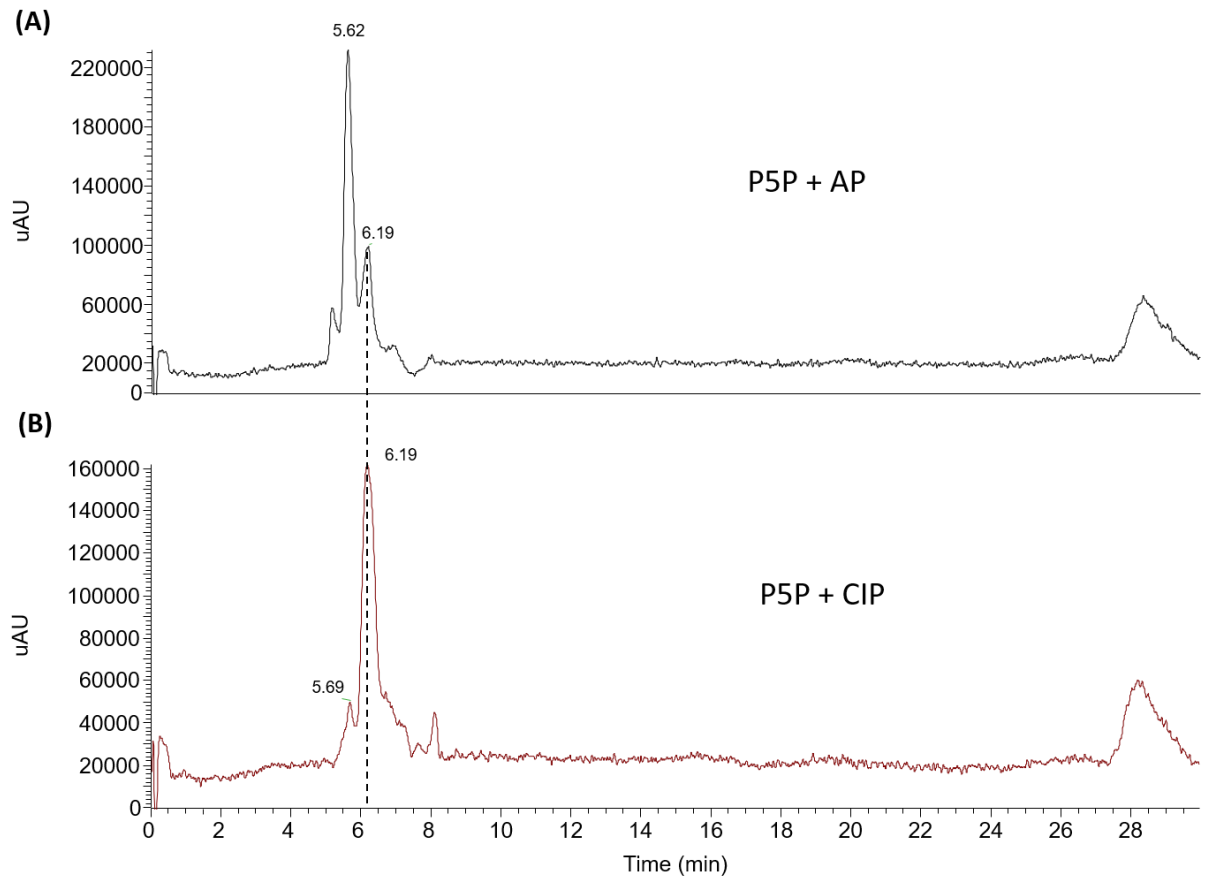


Figure 5.27: UV absorption spectra showing the product of the dephosphorylation reaction of pseudouridine-5'-phosphate, catalysed by commercial phosphatases. (A) Dephosphorylation of P5P to Ψ by antarctic phosphatase. P5P has not been fully converted. **(B)** Dephosphorylation of P5P to Ψ by calf intestinal phosphatase. Absorbance measured at 262 nm.

The NMR and HPLC results showed that the first two reactions, to generate P5P and Ψ from uracil went to completion between combining the components and putting the sample into the NMR spectrometer. Figure 5.23 shows that when uracil and R5P are combined with YeiN, peaks at 5.8 ppm and 7.5 ppm disappear and in their place a peak at 7.95 ppm appears, consistent with the predicted changes in chemical shift. Figure 5.24 shows that when CIP is also present the signal at 7.95 ppm shifts to 7.85 ppm, consistent with Ψ having been produced. This is also supported by the HPLC evidence in Figure 5.26 and 5.27, tracked by the shifts in retention time of peaks with an absorbance maximum of 262 nm. Figure 5.25 shows that when alanine was used as an amino donor, there is potentially a small amount of pyruvate generated (peak at 2.5 ppm), indicative of an aminotransferase reaction having occurred. However, in the absence of an APU standard it is difficult to tell if any APU was generated at the same time. In contrast, there was no evidence from HPLC for conversion of Ψ into the amine

APU. The HPLC spectra show some changes when PsmM and PsmO are added to the reaction mixture, however no mass associated with APU or any intermediate is observed. The additional signals that appear at 3.2 ppm and between 3.5 to 3.8 ppm in Figure 5.24B are due to the presence of the CutSmart buffer which contains Tris-acetate and the CIP which is in a Tris-HCl buffer with 50% glycerol. This makes the ^{1}H NMR spectra messy and difficult to interpret in the affected region, however the buffer was used to maximise the chances of the dephosphorylation reaction to generate Ψ. The production of pyruvate in spectra D of Figure 5.25 should indicate that alanine has been turned over by an aminotransferase and was used as a measure of the progress of the coupled reaction. Pyruvate and alanine were used as standards in place of APU because of the lack of availability of APU from chemical suppliers.

Kinetic studies of oxidoreductase, PsmM

It was important to determine whether the recombinant GMC oxidase PsmM is enzymatically active in the absence of the aminotransferase. If the oxidase uses oxygen as the oxidant then H_2O_2 should be produced and can be detected using components of a commercial glucose oxidase assay kit, namely horseradish peroxidase (HRP) and the dye Amplex Red (AR). In the presence of H_2O_2 the horseradish peroxidase catalyses reaction between H_2O_2 and Amplex Red to form the red-fluorescent oxidation product resorufin. A range of substrate concentrations (25 mM to 0.05 mM final concentrations) for Ψ were mixed with a solution containing HRP and AR. These mixtures were then added to PsmM and the absorption at 570 nm (A_{570}) was measured at 30 second intervals over 5 minutes. To test whether the oxidase could also accept uridine as a substrate, the same experiment was done with uridine substituted for Ψ. The initial rates were calculated by plotting A_{570} against time, and those initial rates were plotted against substrate concentration to generate the plots in Figure 5.28. With the assistance of Elisabeth Chen, a Michaelis-Menten model was fitted to the data and the V_{max} and K_m values for PsmM with Ψ and uridine were calculated, along with their associated standard errors. The values for these kinetic constants are shown in Table 5.4.

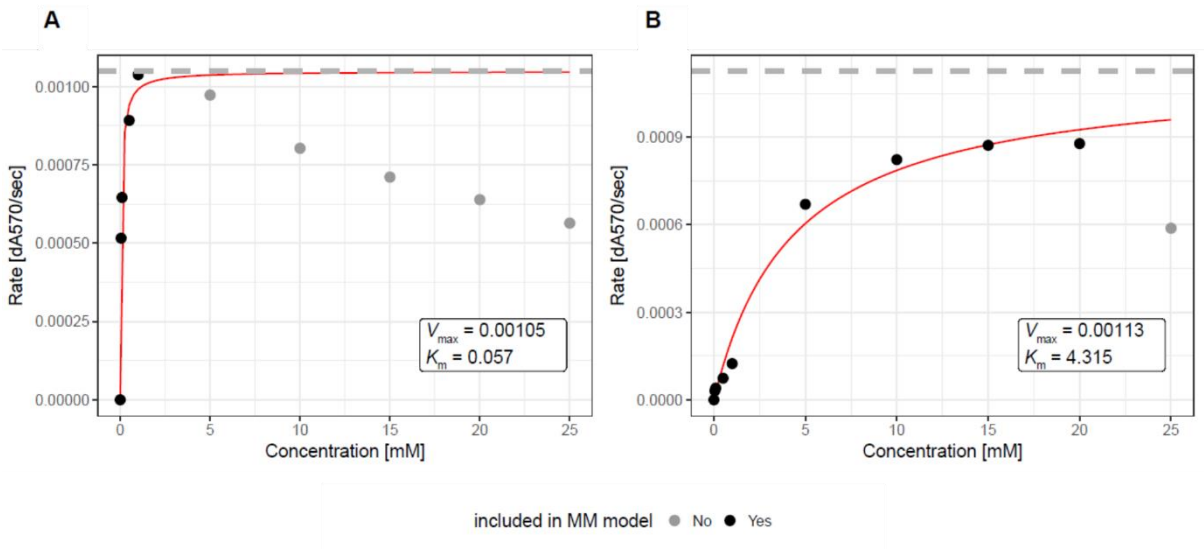


Figure 5.28: Michaelis-Menten plots showing the rate of reaction for PsmM using pseudouridine and uridine as substrates. (A) Rate vs substrate concentration for the oxidation of pseudouridine by PsmM (0.6 µM final concentration). The five substrate concentrations between 5 mM and 25 mM were omitted from the Michaelis-Menten model because at substrate concentrations higher than 1 mM substrate inhibition effects begin to reduce the rate of reaction. **(B)** Rate vs substrate concentration for the oxidation of uridine by PsmM (0.6 µM final concentration). The substrate concentration of 25 mM was omitted from the Michaelis-Menten model.

Table 5.4: Michaelis-Menten kinetic analysis of the oxidation of pseudouridine and uridine by PsmM.

Pseudouridine		**Uridine**	
V_{max} (s^{-1})	1.05×10^{-3}	V_{max} (s^{-1})	1.13×10^{-3}
Standard error	4.23×10^{-5}	Standard error	7.57×10^{-5}
K_m (M)	5.70×10^{-5}	K_m (M)	4.32×10^{-3}
Standard error	1.00×10^{-5}	Standard error	9.40×10^{-4}
k_{cat} (s^{-1})	1.75×10^{3}	k_{cat} (s^{-1})	1.88×10^{3}
k_{cat}/K_m ($s^{-1}M^{-1}$)	3.07×10^{7}	k_{cat}/K_m ($s^{-1}M^{-1}$)	4.37×10^{5}

These data confirm that the recombinant GMC oxidase uses oxygen as the oxidant and that it has the appropriate activity against the pseudouridine substrate. Further, an analysis of the kinetics shows PsmM favours pseudouridine by a factor of almost 2 orders of magnitude compared to uridine. This suggests PsmM as a gatekeeper ensuring the specificity of PUM production by S. albus A980 even in the presence of intracellular uridine. However, it also suggests that if pseudouridine concentrations

were to fall there is a possibility that normal uridine might be utilised by the PUM biosynthetic pathway (see section 5.2.5 below).

The success of this work to prove the activity of PsmM, with a Ψ substrate, in isolation could be repeated with the remaining enzymes. This would further increase confidence in the *in vitro* approach, and the hypothesised function of each enzyme. This would require pure intermediate substrates, which could either be prepared by enzymatic reaction, or purchased from a chemical supplier.

5.2.3: Attempted Inactivation of Genes in the PUM cluster

The in-frame deletion of individual genes within a biosynthetic gene cluster is a powerful approach to gaining insight into their role. For example, it can be used to help define cluster boundaries; it can be used to confirm the predicted role of the gene; and if an intermediate or shunt product is produced instead of PUM its structure gives valuable mechanistic clues. It is also the best way to interrogate the role of genes which have only non-specific annotation. At the start of work on the PUM cluster, there was no information on whether the strain was transformable. I identified PsmQ and PsmU as candidates for the N-hydroxylation of deoxy-PUM; and PsmN, annotated as an adenylate kinase, as an enzyme potentially involved in the biosynthesis of APU. Finally, PsmB, annotated as a hydrolase, was selected for knockout studies to interrogate its role, if any, in PUM biosynthesis.

While my work was in progress, an investigation by Sosio *et al*. [296] into the biosynthesis of PUM by knocking out key biosynthetic genes was published. This reported PsmQ as the enzyme uniquely responsible for the N-amide hydroxylation of deoxy-PUM. Because of this finding I discontinued work on *psmQ* and continued with putative hydrolase *psmB,* cytochrome P450 *psmU* and adenylate kinase *psmN*. This section describes the construction of appropriate vectors and attempts - unfortunately unsuccessful - to complete the deletion of *psmB*, *psmU* and *psmN*. In parallel, Dr Hui Hong undertook the in-frame deletion of *psmL*, encoding the *truD*-like pseudouridine synthase.

For the creation of the Δ*psmN*, Δ*psmB*, and Δ*psmU* mutants a homologous recombination-based approach was used. The strategy is outlined in Figure 5.29. Primers were designed to amplify 2 kbp regions of DNA either side of each target gene (left and right flanking arms). The primers were designed to have overhanging 5’ ends, homologous to the adjacent sequence when assembled. The amplified PCR products were purified and assembled together with linearized pYH7 using the Gibson isothermal assembly method. The mixtures were used to transform *E. coli* DH10B and colonies were

screened by PCR for the desired insert. The identity of the PCR products and the final vector inserts was confirmed by gel electrophoresis analysis (Figure 5.30 and Figure 5.31). The procedure is described in detail in section 2.3. Figure 5.32 shows the gel electrophoresis from step 6 from each of the cloned vectors.

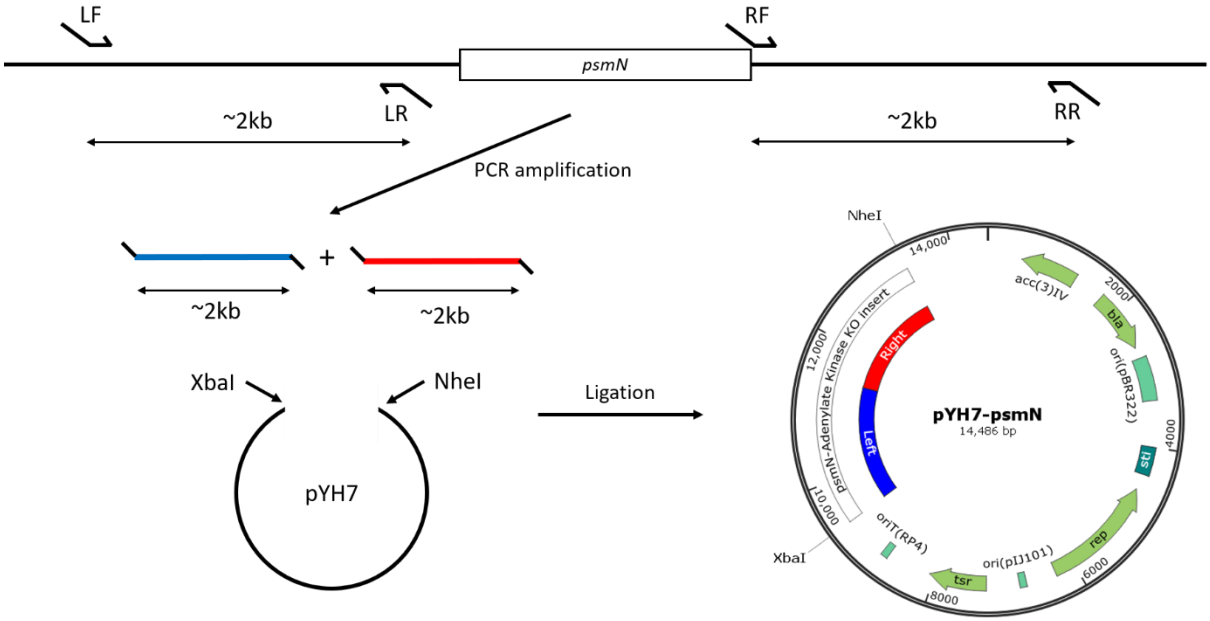


Figure 5.29: The strategy for inactivation of *psmN*, *psmB*, *psmQ*, and *psmU* in *S. albus* A980.

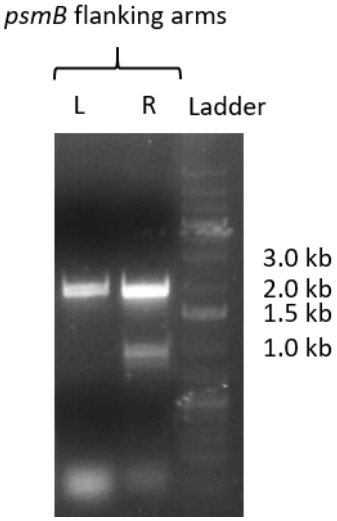


Figure 5.30: Gel electrophoresis image of PCR products used to construct the *psmB* knockout vector. The 'Ladder' lane contains the DNA ladder (1 Kb Plus DNA Ladder). The numbers to the side of the ladder refer to the size of each relevant band in the ladder. The 'L' and 'R' lanes refer to the left and right flanking arms, respectively.

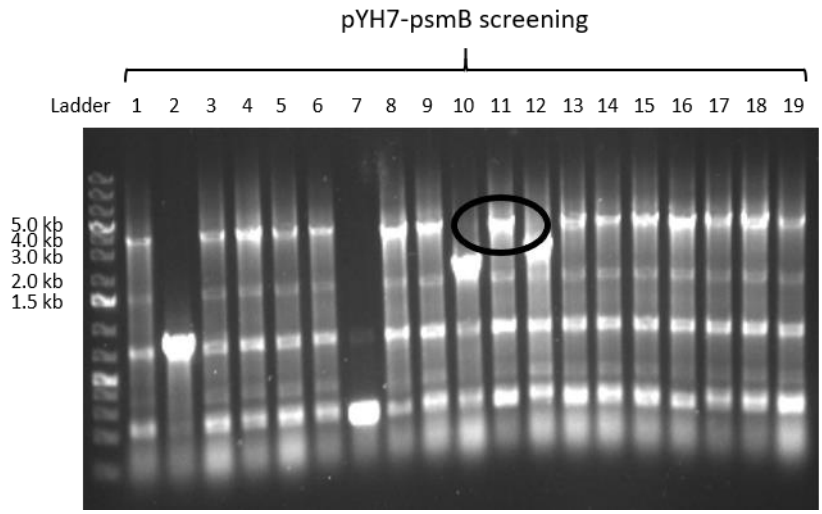


Figure 5.31: Gel electrophoresis image of the cPCR screening to confirm correct assembly of the *psmB* knockout vector. The 'Ladder' lane contains the DNA ladder (1 Kb Plus DNA Ladder). The numbers to the side of the ladder refer to the size of each relevant band in the ladder. Each numbered lane refers to a colony of E. coli that was screened for the correct knockout vector. The lanes 1,3-6,8,9,11, and 13 to 19 contain bands with the correct size for the *psmB* knockout vector. The circled lane (lane 11) indicates the colony used to purify the *psmB* knockout vector in preparation for transformation into A980.

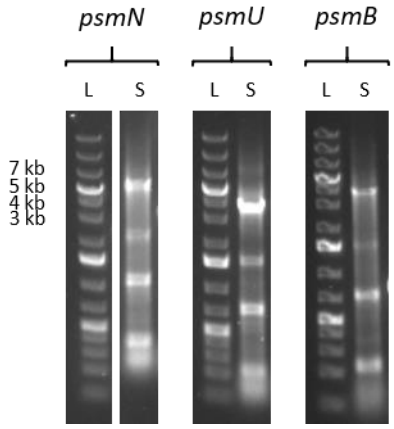


Figure 5.32: Gel electrophoresis images of the inserts from each of the three *psm* knockout vectors. The 'L' lanes contains the DNA ladder (1 Kb Plus DNA Ladder). The numbers to the side of the ladder refer to the size of each relevant band in the ladder. The 'S' lanes refer to a colony used to confirm the size of the assembled insert for each of the *psm* knockout vectors.

Once the knockout vectors for *psmN*, *psmU*, and *psmB* had been obtained they were transformed into A980 by intergeneric conjugation following the protocol outlined in section 2.2.6. Each conjugation plate was incubated at 30°C for up to 1 week. By then several colonies had begun to grow. These were streaked onto selective media to confirm the presence of the vector. After two more rounds of restreaking on selective media the putative exconjugants were free of any contamination and ready for the relaxation step. At this point I attempted unsuccessfully to use *Streptomyces* cPCR to gain additional evidence that these exconjugants contained the integrated vector. I considered the growth of the colonies on selective media evidence enough for the integration of the vector and continued with the relaxation steps. Three apramycin resistant colonies were selected from each conjugation plate, and these were relaxed by three rounds of inoculation and growth in liquid SFM, incubated at 30°C. After the third round a sample of each culture was taken, diluted and streaked across 2TY agar. Single colonies (24 for each knockout) were selected. These were patched onto pairs of non-selective and selective plates to test of the loss of antibiotic resistance. All colonies tested had lost resistance. Each relaxant was then grown in liquid culture and genomic DNA was extracted from each relaxant. A screen for the presence of a second crossover identified that all colonies were wildtype. As this screen was performed towards the end of my PhD I did not have time to repeat the experiment.

Fortunately, Dr Hui Hong had meanwhile succeeded in obtaining an in-frame deletion in the *psmL* gene (the *truD* homologue). She analysed the fermentation culture by LCMS. The HPLC spectra from A980 wildtype and A980 Δ*psmL* showed a clear reduction in the signal representing PUM (see Figure 5.33). Unexpectedly, the Δ*psmL* spectrum also revealed a UV absorption signal corresponding to a molecule with the same mass as PUM, but with a different retention time. Further investigation of this unexpected peak by MS/MS analysis revealed similar fragmentation patterns to PUM, consistent with the idea that the peak is produced by an N-nucleoside isomer of PUM.

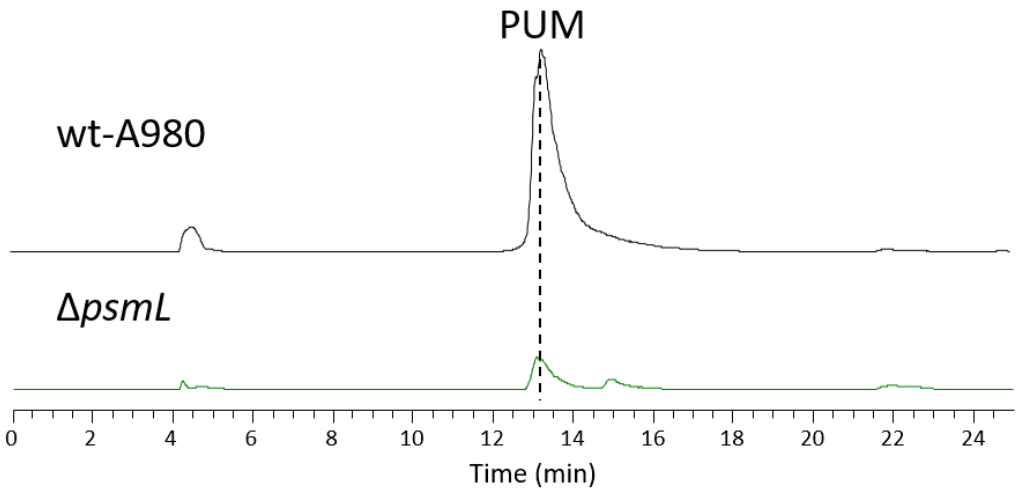


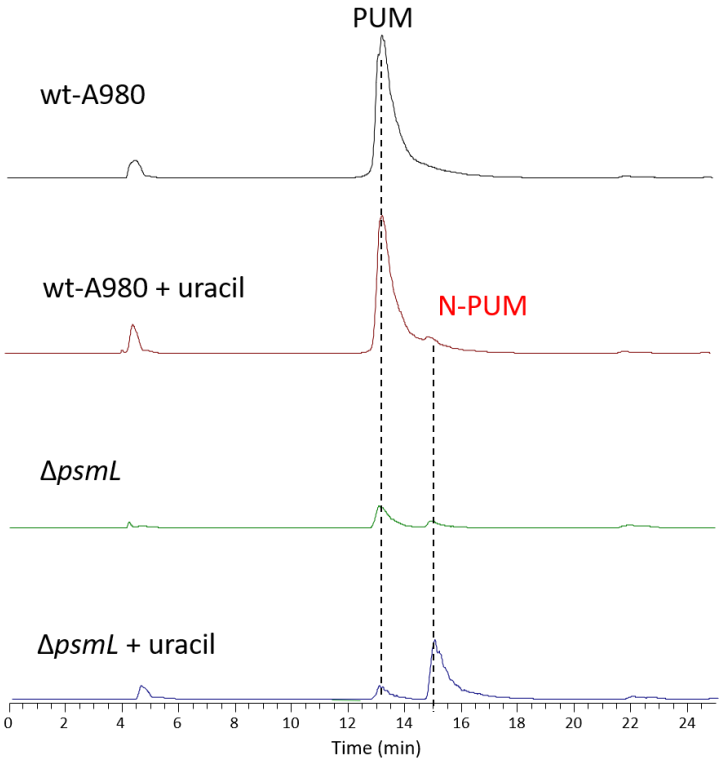
Figure 5.33: LCMS analysis of wild type A980 and A980 Δ*psmL*. (Experiment carried out by Dr Hui Hong). The spectra are normalised to the wild type A980 spectrum. The residual PUM production in the Δ*psmL* strain is likely due to free Ψ found in the cell from primary metabolism. The small peak seen in the Δ*psmL* spectrum at approximately 15 mins has the same characteristic UV absorption and mass as PUM.

Sosio and colleagues [296] also created a *truD* mutant of their *Streptomyces* sp. strain, and they stated categorically that in this mutant PUM production was abolished and no *N*-linked PUM was produced, as might have been expected if pseudouridine production was abolished. When they fed uridine to the Δ*psmL* mutant they did not see the accumulation of any compound.

In contrast, Dr Hui Hong's analysis appeared to show that in *S. albus* A980, there is detectable redirection of biosynthesis towards an *N*-linked version of PUM when the *truD* -like gene *psmL* is deleted. This was supported by the kinetic data obtained here on the key GMC oxidase enzyme which accepted uridine as a substrate, albeit with much lower efficiency. To confirm this finding of the tolerance of the PUM biosynthetic pathway in *S. albus* A980, I used the *S. albus* A980-Δ*psmL* kindly provided by Dr Hong. A high concentration of uracil was fed to the Δ*psmL* mutant strain: a 2-day pre-culture of A980-Δ*psmL* was used to inoculate 50 ml SFM containing 50 mg of uracil in a 250 ml flask with a spring to help aeration. After 5 days the cells were harvested by centrifugation and the supernatant was used for HPLC and MS analysis. Figure 5.34 shows the results of that analysis and compares them against wild type A980. The data shown in Figure 5.34 were kindly provided by Dr. Hui Hong. The peak at *m*/*z* = 259.36 is characteristic of the *N*-PUM isomer and is generated by a cleavage of the C-N nucleoside bond, cleavage of the guanidinoacetate-glutamate amide bond, and loss of ammonia.

The analysis clearly showed that formation of *N*-PUM in the *S. albus* A980-Δ*psmL* mutant is boosted by feeding with uracil. It is difficult to be sure why Sosio and colleagues obtained different results. It might be a strain difference that means no *N*-PUM is visible in their *truD* mutant. Also, feeding of uridine may be less effective than feeding with uracil. Meanwhile, direct comparison of the intensity of the signals for PUM and *N*-PUM shows that the *N*-nucleoside is produced in levels lower than those of *C*-nucleoside in the wild type, but that it is present in quantities that should allow its future isolation and testing. This result encourages the view that there is some flexibility in the PUM biosynthetic pathway that might allow the future creation of interesting PUM analogues.

(A)



(B)

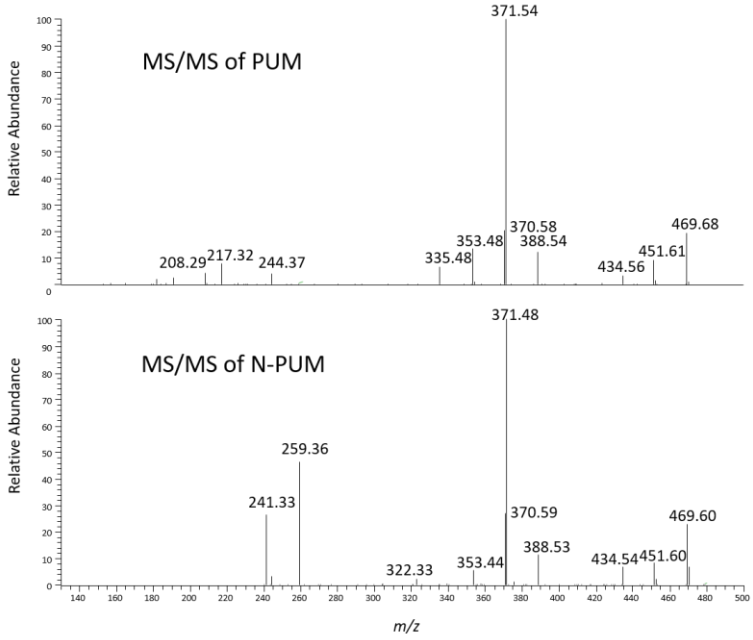


Figure 5.34: HPLC spectra and mass spectrometry data comparing the fermentation products of wild type A980 and A980-Δ*psmL*. (A) LCMS analysis of PUM and its *N*-isomer. The spectra have been normalised to wild type PUM production. **(B)** MS/MS analysis of PUM and its *N*-isomer. The peak at *m*/*z* = 259.36 is characteristic of the *N*-PUM isomer.

5.3: Concluding remarks

The work described in this chapter has identified the biosynthetic gene cluster for the remarkable *N*-nucleoside antibiotic pseudouridimycin in *Streptomyces albus* DSM 40763 (A980). It has defined the key genes required for the biosynthesis of PUM and characterised the amidinotransferase by *in vitro* assay. A two-step *in vitro* enzymatic reaction combining a pseudouridine-5'-phosphate glycosidase followed by a dephosphorylation reaction by a commercial phosphatase successfully generated Ψ from R5P and uracil. Unfortunately, the *in vitro* assays for the remainder of the pathway to APU were unsuccessful. Nevertheless, it was possible to show that the GMC oxidase PsmM shows the predicted ability to oxidise Ψ. Further, the enzyme accepted uridine albeit with much lower efficiency. Unfortunately, knockouts of *psmN*, *psmU*, and *psmB* were not obtained. The publication by Sosio *et al*. implicates PsmQ as the enzyme responsible for the N-amide hydroxylation, but the hydrolase, PsmB, and the p450, PsmU, still have no defined role in the cluster. The adenylate kinase, PsmN, is still an enticing target for knockout studies as its role in the biosynthesis of PUM cannot be easily predicted, other than the possibility of its involvement in the generation of Ψ.

The recent work by Sosio et al. [296] offers complementary insights into the biosynthesis of PUM. The Sosio gene cluster has differences in both arrangement and type of genes present when compared to the PUM cluster in A980. The work by Sosio and colleagues matches my predictions of the functions of the *psm* cluster genes. However, there is an interesting difference between the Sosio PUM producer and A980. When the *truD* homologue is knocked out in the Sosio strain the only PUM related compound that is observed is guanidinoacetate, and when the mutant strain is fed with uridine, it is not incorporated into any PUM related compound. This is in contrast to A980, where the Δ*TruD* (Δ*psmL*) is able to take up uracil and incorporate it into an *N*-PUM, a novel nucleoside. Identifying and understanding the differences between the two PUM clusters may help characterise the pathway more fully.

The important question that remains unanswered regards the origin of Ψ as a substrate for PsmM. The TruD homologue, psmL, has the function of generating Ψ residues in tRNA polymers, but that Ψ must be released from the tRNA before it can be incorporated into PUM. It is possible that PsmL is capable of converting a metabolite related to uridine into Ψ but the direct conversion of uridine into Ψ has been attempted by Dr Hong with recombinant enzyme without success (Hong *et al*., 2018, in the press). Another possibility is that the PUM cluster does not have a dedicated pathway to produce free Ψ and must rely on primary metabolism to degrade converted Ψ in tRNA before incorporating the nucleoside into PUM. It has been well documented that *Streptomyces* spp.

undergoes significant RNA degradation, coinciding with important morphological changes during transition to stationary phase. In *E. coli*, tRNA has been shown to become significantly unstable as a stress response [300], so it is plausible that the entry of Streptomyces into the stationary phase, which can trigger secondary metabolite production, might also accelerate tRNA breakdown. It may be possible to probe this idea by measuring production of PUM across several timepoints and comparing the production of PUM to the growth cycles of A980.

Chapter 6: Conclusions and Suggestions for Future Work

The work described in this thesis has addressed the current paradigm for genome mining, biosynthetic gene cluster characterisation, and engineering of biosynthetic gene clusters towards generation of novel biologically active compounds in *Streptomyces* spp. It has also explored new approaches towards PKS bioengineering in the form of Accelerated Evolution.

The main conclusion that can be drawn from the work in Chapter 3 is that while the engineering technology Accelerated Evolution shows promise in the rapamycin PKS system, further work is required to understand the mechanisms behind AE before it can be generalised and applied to other systems. In its current state, AE can produce a large library of rapamycin derived compounds, but besides a single putative exception, no derivative compounds were observed in the three systems AE was applied to in this work. The failure to produce the desired response may be due to the absence of an important factor found in the rapamycin producer, or it might be the result of an inadequate protocol that can't produce the desired recombination effect. This effect has only been recorded once, other than the paper by Wlodek *et al*. suggests that AE is possible in other strains but is extremely rare. I believe in the absence of a better understanding of the mechanism, it relied too heavily on luck to achieve an effect that had only been observed in a very small number of polyketide producing strains. Future work must focus on characterising the mechanism behind AE in *Streptomyces rapamycinicus*. There are several approaches that could contribute to this goal. Identifying genes that may be involved in the homologous recombination between modules and systematically knocking them out in an effort to break AE in the knockouts may provide a starting point for identifying other strains to which AE can be applied. This may be difficult as homologous recombination is not as well understood in *Streptomyces* spp. as it is in other species, but it could provide vital information on the mechanism behind AE and so should still be pursued. Another approach would be to study the genome sequence of the recombinant *S. rapamycinicus* strains. The recombination points may provide clues on the most effective entry points to be targeted for future attempts at applying AE.

In addition to understanding the mechanism behind AE, future work should aim to create several approaches that assist with screening. The nature of AE experiments, especially in the work described above, produce large number of exconjugant and recombinant strains. Screening for desired traits takes considerable time and effort. The application of colony PCR for *Streptomyces* colonies saved a lot of work, but it is far from perfect. Any method that can ease the screening process should be explored and if successful, embraced.

The work in Chapter 4 focused on the genome mining of *S. albus* DSM40763. The conclusions that can be made from this work are firstly that the strain appears to be rich in uncharacterised biosynthetic potential, including rarer clusters such as a putative hybrid cyclodipeptide-terpene and an enormous 21 module PKS. The second conclusion that can be made is that an essentially complete genome sequence allows significant advantages over the more widespread fragmented genomes in the time and effort required to identify clusters of interest. Assembly of shotgun fragments can be difficult in strains such as *Streptomyces*, known for their high GC% content and large repeating sequences, but it is essential to giving an accurate estimate of biosynthetic potential. A further conclusion is that automated annotation tools such as antiSMASH cannot be used without manual curation. The program tends to annotate much wider limits to gene clusters than would be annotated manually, leading to the tool predicting hybrid clusters, and other misleading results. The tool also failed to identify certain classes of clusters, for example the PUM cluster was not correctly identified by the latest version of antiSMASH (4.0).

Chapter 5 discussed the *in vitro* and *in vivo* studies towards characterising the PUM biosynthetic gene cluster found in *S. albus* DSM40763. PUM is a potentially valuable therapeutic agent, and its biosynthetic pathway shows promise as a platform for generating a library of related compounds. This was shown by the ability of the PUM gene cluster to generate an *N*-nucleoside isomer of PUM. The *in vivo* work initially identified this compound, with feeding experiments showing the presence of uracil (1 mg/ml) increased production of *N*-PUM. The *in vitro* work complemented this observation by showing that the oxidoreductase PsmM, which is thought to be the first enzyme acting on the Ψ substrate in the PUM biosynthetic pathway, could turnover both Ψ and U, with an approximate hundred-fold preference for Ψ. The promiscuity of PsmM and presumably the downstream enzymes suggest they may be tolerant for other substrates. An attempt to reconstitute the entire PUM biosynthetic pathway *in vitro* was halted by a failure to generate the 5'-amino-pseudouridine (APU) component of PUM, however the guanidinoacetate component was generated *in vitro* by PsmH, and PsmM was shown to convert Ψ, which generates the precursor to APU. Future work in this area should focus on proving the ability of the aminotransferase, PsmO, to turn over the product of PsmM. This would be helped greatly by access to pure substrates which could be prepared by scaling up the enzymatic reactions described above and purified by preparative HPLC. Finally, the heterologous expression of the ligase, PsmK, should be investigated as it would be the last barrier to a full *in vitro* reconstitution of the biosynthesis of PUM.

References

1. Avery, M. A., Chong, W. K. M. & Jennings-White, C. Stereoselective total synthesis of (+)-artemisinin, the antimalarial constituent of Artemisia annua L. *J. Am. Chem. Soc.* **114,** 974–979 (1992).
2. Dhanjee, H. H. *et al.* Total Syntheses of (+)- and (−)-Tetrapetalones A and C. *J. Am. Chem. Soc.* **139,** 14901–14904 (2017).
3. Inai, M., Asakawa, T. & Kan, T. Total synthesis of natural products using a desymmetrization strategy. *Tetrahedron Lett.* **59,** 1343–1347 (2018).
4. Koehn, F. E. & Carter, G. T. The evolving role of natural products in drug discovery. *Nat. Rev. Drug Discov.* **4,** 206–220 (2005).
5. Harvey, A. L. Natural products in drug discovery. *Drug Discov. Today* **13,** 894–901 (2008).
6. Shen, B. A New Golden Age of Natural Products Drug Discovery. *Cell* **163,** 1297–1300 (2015).
7. All natural. *Nat. Chem. Biol.* **3,** 351 (2007).
8. Steyn, P. S. & Bu'Lock, J. D. *The Biosynthesis of Mycotoxins*. (Acedemic Press, 1980). doi:10.1016/B978-0-12-670650-5.X5001-8
9. Bu'Lock, J. D., Nisbet, L. J., Winstanley, D. J. & Microbiology, S. for G. *Bioactive microbial products : search and discovery*. (London ; New York : Published for the Society for General Microbiology by Academic Press, 1982).
10. Williams, D. H., Stone, M. J., Hauck, P. R. & Rahman, S. K. Why Are Secondary Metabolites (Natural Products) Biosynthesized? *J. Nat. Prod.* **52,** 1189–1208 (1989).
11. Firn, R. D. & Jones, C. G. The evolution of secondary metabolism – a unifying model. *Mol. Microbiol.* **37,** 989–994 (2000).
12. Cragg, G. M. & Newman, D. J. Biodiversity: A continuing source of novel drug leads. *Pure Appl. Chem.* **77,** 7–24 (2005).
13. Zhong, G. & Wan, F. [An outline on the early pharmacetical development before Galen]. *Zhonghua Yi Shi Za Zhi Beijing China 1980* **29,** 178–182 (1999).
14. Dev, S. Ancient-modern concordance in Ayurvedic plants: some examples. *Environ. Health Perspect.* **107,** 783–789 (1999).
15. Dias, D. A., Urban, S. & Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2,** 303–336 (2012).
16. Hong, J. Natural product diversity and its role in chemical biology and drug discovery. *Curr. Opin. Chem. Biol.* **15,** 350–354 (2011).
17. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18,** 380–416 (2001).
18. Schatz, A., Bugie, E. & Waksman, S. A. Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. 1944. *Clin. Orthop.* 3–6 (2005).
19. Davies, J. Where have all the Antibiotics Gone? *Can. J. Infect. Dis. Med. Microbiol.* **17,** 287–290 (2006).
20. Schueffler, A. & Anke, T. Fungal natural products in research and development. *Nat. Prod. Rep.* **31,** 1425–1448 (2014).
21. Götz, K., Liermann, J. C., Thines, E., Anke, H. & Opatz, T. Structure elucidation of hypocreolide A by enantioselective total synthesis. *Org. Biomol. Chem.* **8,** 2123–2130 (2010).
22. Shiono, Y. *et al.* A new benzoxepin metabolite isolated from endophytic fungus Phomopsis sp. *J. Antibiot. (Tokyo)* **62,** 533–535 (2009).
23. Izumikawa, M. *et al.* JBIR-37 and -38, novel glycosyl benzenediols, isolated from the sponge-derived fungus, Acremonium sp. SpF080624G1f01. *Biosci. Biotechnol. Biochem.* **73,** 2138–2140 (2009).
24. McChesney, J. D., Venkataraman, S. K. & Henri, J. T. Plant natural products: Back to the future or into extinction? *Phytochemistry* **68,** 2015–2022 (2007).

25. Anarat-Cappillino, G. & Sattely, E. S. The chemical logic of plant natural product biosynthesis. *Curr. Opin. Plant Biol.* **19,** 51–58 (2014).

26. Cragg, G. M. Paclitaxel (Taxol): a success story with valuable lessons for natural product drug discovery and development. *Med. Res. Rev.* **18,** 315–331 (1998).

27. Paddon, C. J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496,** 528–532 (2013).

28. Watve, M. G., Tickoo, R., Jog, M. M. & Bhole, B. D. How many antibiotics are produced by the genus *Streptomyces*? *Arch. Microbiol.* **176,** 386–390 (2001).

29. Baltz, R. H. Antimicrobials from Actinomycetes: Back to the Future. 7

30. Mannanov, R. N. & Sattarova, R. K. Antibiotics Produced by Bacillus Bacteria. *Chem. Nat. Compd.* **37,** 117–123 (2001).

31. Hamdache, A., Lamarti, A., Aleu, J. & Collado, I. G. Non-peptide Metabolites from the Genus Bacillus. *J. Nat. Prod.* **74,** 893–899 (2011).

32. Challinor, V. L. & Bode, H. B. Bioactive natural products from novel microbial sources. *Ann. N. Y. Acad. Sci.* **1354,** 82–97 (2015).

33. Reichenbach, H. Myxobacteria, producers of novel bioactive substances. *J. Ind. Microbiol. Biotechnol.* **27,** 149–156 (2001).

34. Reichenbach, H., Gerth, K., Irschik, H., Kunze, B. & Höfle, G. Myxobacteria: a source of new antibiotics. *Trends Biotechnol.* **6,** 115–121 (1988).

35. Mahenthiralingam, E. *et al.* Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic Burkholderia ambifaria Genomic Island. *Chem. Biol.* **18,** 665–677 (2011).

36. Partida-Martinez, L. P. & Hertweck, C. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* **437,** 884–888 (2005).

37. Oka, M. *et al.* Glidobactins A, B and C, new antitumor antibiotics. *J. Antibiot. (Tokyo)* **41,** 1338–1350 (1988).

38. Duplessis, C. & Crum-Cianflone, N. F. Ceftaroline: A New Cephalosporin with Activity Against Methicillin-Resistant Staphylococcus aureus (MRSA). *Clin. Med. Rev. Ther.* **2011,** 1–17 (2011).

39. Scheeren, T. W. Ceftobiprole medocaril in the treatment of hospital-acquired pneumonia. *Future Microbiol.* **10,** 1913–1928 (2015).

40. Antoraz, S., Santamaría, R. I., Díaz, M., Sanz, D. & Rodríguez, H. Toward a new focus in antibiotic and drug discovery from the Streptomyces arsenal. *Front. Microbiol.* **6,** (2015).

41. Embley, T. M. & Stackebrandt, E. The molecular phylogeny and systematics of the actinomycetes. *Annu. Rev. Microbiol.* **48,** 257–289 (1994).

42. Chater, K. F., Biró, S., Lee, K. J., Palmer, T. & Schrempf, H. The complex extracellular biology of Streptomyces. *FEMS Microbiol. Rev.* **34,** 171–198 (2010).

43. Wibberg, D. *et al.* Complete genome sequence of Streptomyces reticuli, an efficient degrader of crystalline cellulose. *J. Biotechnol.* **222,** 13–14 (2016).

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

45. Chater, K. F. Genetics of differentiation in Streptomyces. *Annu. Rev. Microbiol.* **47,** 685–713 (1993).

46. Chater, K. F. & Horinouchi, S. Signalling early developmental events in two highly diverged Streptomyces species. *Mol. Microbiol.* **48,** 9–15 (2003).

47. Angert, E. R. Alternatives to binary fission in bacteria. *Nat. Rev. Microbiol.* **3,** 214–224 (2005).

48. Mayer, A. F. & Deckwer, W. D. Simultaneous production and decomposition of clavulanic acid during Streptomyces clavuligerus cultivations. *Appl. Microbiol. Biotechnol.* **45,** 41–46 (1996).

49. Theobald, U., Schimana, J. & Fiedler, H. P. Microbial growth and production kinetics of Streptomyces antibioticus Tü 6040. *Antonie Van Leeuwenhoek* **78,** 307–313 (2000).

50. Torres-Bacete, J. *et al.* Optimization of culture medium and conditions for penicillin acylase production by Streptomyces lavendulae ATCC 13664. *Appl. Biochem. Biotechnol.* **126,** 119–132 (2005).

51. Yagüe, P., Lopez-Garcia, M. T., Rioseras, B., Sanchez, J. & Manteca, A. New insights on the development of Streptomyces and their relationships with secondary metabolite production. *Curr. Trends Microbiol.* **8,** 65–73 (2012).

52. Vecht-Lifshitz, S. E., Sasson, Y. & Braun, S. Nikkomycin production in pellets of Streptomyces tendae. *J. Appl. Bacteriol.* **72,** 195–200 (1992).

53. Yang, Y. K. *et al.* Image analysis of mycelial morphology in virginiamycin production by batch culture of Streptomyces virginiae. *J. Ferment. Bioeng.* **81,** 7–12 (1996).

54. Blake, K. L. & O'Neill, A. J. Transposon library screening for identification of genetic loci participating in intrinsic susceptibility and acquired resistance to antistaphylococcal agents. *J. Antimicrob. Chemother.* **68,** 12–16 (2013).

55. Liu, A. *et al.* Antibiotic Sensitivity Profiles Determined with an Escherichia coli Gene Knockout Collection: Generating an Antibiotic Bar Code. *Antimicrob. Agents Chemother.* **54,** 1393–1403 (2010).

56. Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. V. Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* **13,** 42–51 (2015).

57. Wang, W. *et al.* High-level tetracycline resistance mediated by efflux pumps Tet(A) and Tet(A)-1 with two start codons. *J. Med. Microbiol.* **63,** 1454–1459 (2014).

58. Vargiu, A. V. & Nikaido, H. Multidrug binding properties of the AcrB efflux pump characterized by molecular dynamics simulations. *Proc. Natl. Acad. Sci.* **109,** 20637–20642 (2012).

59. Abraham, E. P. & Chain, E. An enzyme from bacteria able to destroy penicillin. 1940. *Rev. Infect. Dis.* **10,** 677–678 (1988).

60. Sanders, C. C. & Sanders, W. E. Emergence of resistance to cefamandole: possible role of cefoxitin-inducible beta-lactamases. *Antimicrob. Agents Chemother.* **15,** 792–797 (1979).

61. D'Costa, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477,** 457–461 (2011).

62. Gardete, S. & Tomasz, A. Mechanisms of vancomycin resistance in Staphylococcus aureus. *J. Clin. Invest.* **124,** 2836–2840 (2014).

63. Brown, E. D. & Wright, G. D. Antibacterial drug discovery in the resistance era. *Nature* **529,** 336–343 (2016).

64. Osbourn, A. Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet.* **26,** 449–457 (2010).

65. Zheng, Y., Szustakowski, J. D., Fortnow, L., Roberts, R. J. & Kasif, S. Computational Identification of Operons in Microbial Genomes. *Genome Res.* **12,** 1221–1230 (2002).

66. Dhillon, N., Hale, R. S., Cortes, J. & Leadlay, P. F. Molecular characterization of a gene from Saccharopolyspora erythraea (Streptomyces erythraeus) which is involved in erythromycin biosynthesis. *Mol. Microbiol.* **3,** 1405–1414 (1989).

67. Malpartida, F. & Hopwood, D. A. Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. *Nature* **309,** 462–464 (1984).

68. Bibb, M. J. Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.* **8,** 208–215 (2005).

69. Wietzorrek, A. & Bibb, M. A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold. *Mol. Microbiol.* **25,** 1181–1184 (1997).

70. Sekurova, O. N. *et al.* In Vivo Analysis of the Regulatory Genes in the Nystatin Biosynthetic Gene Cluster of Streptomyces noursei ATCC 11455 Reveals Their Differential Control Over Antibiotic Biosynthesis. *J. Bacteriol.* **186,** 1345–1354 (2004).

71. Hesketh, A., Sun, J. & Bibb, M. Induction of ppGpp synthesis in Streptomyces coelicolor A3(2) grown under conditions of nutritional sufficiency elicits actII-ORF4 transcription and actinorhodin biosynthesis. *Mol. Microbiol.* **39,** 136–144 (2001).

72. Takano, E. γ-Butyrolactones: Streptomyces signalling molecules regulating antibiotic production and differentiation. *Curr. Opin. Microbiol.* **9,** 287–294 (2006).

73. Kato, J., Chi, W.-J., Ohnishi, Y., Hong, S.-K. & Horinouchi, S. Transcriptional Control by A-Factor of Two Trypsin Genes in Streptomyces griseus. *J. Bacteriol.* **187,** 286–295 (2005).

74. Stone, M. J. & Williams, D. H. On the evolution of functional secondary metabolites (natural products). *Mol. Microbiol.* **6,** 29–34 (1992).

75. Konishi, M. *et al.* Dynemicin A, a novel antibiotic with the anthraquinone and 1,5-diyn-3-ene subunit. *J. Antibiot. (Tokyo)* **42,** 1449–1452 (1989).

76. Yu, J. Current Understanding on Aflatoxin Biosynthesis and Future Perspective in Reducing Aflatoxin Contamination. *Toxins* **4,** 1024–1057 (2012).

77. Nakanaga, K. *et al.* Buruli Ulcer and Mycolactone-Producing Mycobacteria. *Jpn. J. Infect. Dis.* **66,** 83–88 (2013).

78. Rohm, B., Scherlach, K. & Hertweck, C. Biosynthesis of the mitochondrial adenine nucleotide translocase (ATPase) inhibitor bongkrekic acid in Burkholderia gladioli. *Org. Biomol. Chem.* **8,** 1520–1522 (2010).

79. Koehn, F. E. & Carter, G. T. Rediscovering natural products as a source of new drugs. *Discov. Med.* **5,** 159–164 (2005).

80. Koehn, F. E. High impact technologies for natural products screening. *Prog. Drug Res. Fortschritte Arzneimittelforschung Prog. Rech. Pharm.* **65,** 175, 177–210 (2008).

81. Li, J. W.-H. & Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **325,** 161–165 (2009).

82. Weissman, K. J. Chapter 1 Introduction to Polyketide Biosynthesis. in *Methods in Enzymology* (ed. Hopwood, D. A.) **459,** 3–16 (Academic Press, 2009).

83. Yu, D., Xu, F., Zeng, J. & Zhan, J. Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* **64,** 285–295 (2012).

84. Austin, M. B. & Noel, J. P. The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* **20,** 79–110 (2003).

85. Waters, E. R. Molecular adaptation and the origin of land plants. *Mol. Phylogenet. Evol.* **29,** 456–463 (2003).

86. Stewart, C., Vickery, C. R., Burkart, M. D. & Noel, J. P. Confluence of structural and chemical biology: plant polyketide synthases as biocatalysts for a bio-based future. *Curr. Opin. Plant Biol.* **16,** 365–372 (2013).

87. Austin, M. B., O'Maille, P. E. & Noel, J. P. Evolving biosynthetic tangos negotiate mechanistic landscapes. *Nat. Chem. Biol.* **4,** 217–222 (2008).

88. Weiss, R. B. The anthracyclines: will we ever find a better doxorubicin? *Semin. Oncol.* **19,** 670–686 (1992).

89. Rix, U., Fischer, C., Remsing, L. L. & Rohr, J. Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Nat. Prod. Rep.* **19,** 542–580 (2002).

90. Matharu, A. L., Cox, R. J., Crosby, J., Byrom, K. J. & Simpson, T. J. MCAT is not required for in vitro polyketide synthesis in a minimal actinorhodin polyketide synthase from Streptomyces coelicolor. *Chem. Biol.* **5,** 699–711 (1998).

91. Arthur, C. J. *et al.* Self-malonylation is an intrinsic property of a chemically synthesized type II polyketide synthase acyl carrier protein. *Biochemistry* **44,** 15414–15421 (2005).

92. Keatinge-Clay, A. T., Maltby, D. A., Medzihradszky, K. F., Khosla, C. & Stroud, R. M. An antibiotic factory caught in action. *Nat. Struct. Mol. Biol.* **11,** 888–893 (2004).

93. Cox, R. J. Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Org. Biomol. Chem.* **5,** 2010 (2007).

94. Nguyen, H.-C., Darbon, E., Thai, R., Pernodet, J.-L. & Lautru, S. Post-PKS Tailoring Steps of the Spiramycin Macrolactone Ring in Streptomyces ambofaciens. *Antimicrob. Agents Chemother.* **57,** 3836–3842 (2013).

95. Tang, L., Fu, H., Betlach, M. C. & McDaniel, R. Elucidating the mechanism of chain termination switching in the picromycin/methymycin polyketide synthase. *Chem. Biol.* **6,** 553–558 (1999).

96. Bevitt, D. J., Cortes, J., Haydock, S. F. & Leadlay, P. F. 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. *Eur. J. Biochem.* **204,** 39–49 (1992).

97. Caffrey, P. *et al.* An acyl-carrier-protein – thioesterase domain from the 6-deoxyerythronolide B synthase of Saccharopolyspora erythraea. *Eur. J. Biochem.* **195,** 823–830 (1991).

98. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. Modular organization of genes required for complex polyketide biosynthesis. *Science* **252,** 675–679 (1991).

99. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. *Nature* **348,** 176–178 (1990).

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

101. Aparicio, J. F. *et al.* Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus: Analysis of the enzymatic domains in the modular polyketide synthase. *Gene* **169,** 9–16 (1996).

102. Stinear, T. P. *et al.* Giant plasmid-encoded polyketide synthases produce the macrolide toxin of Mycobacterium ulcerans. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 1345–1349 (2004).

103. Katz, L. Chapter 6 The DEBS Paradigm for Type I Modular Polyketide Synthases and Beyond. in *Methods in Enzymology* **459,** 113–142 (Elsevier, 2009).

104. Weissman, K. J. Polyketide stereocontrol: a study in chemical biology. *Beilstein J. Org. Chem.* **13,** 348–371 (2017).

105. Marsden, A. F. *et al.* Stereospecific acyl transfers on the erythromycin-producing polyketide synthase. *Science* **263,** 378–380 (1994).

106. Wilson, M. C. & Moore, B. S. Beyond ethylmalonyl-CoA: The functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Nat. Prod. Rep.* **29,** 72–86 (2011).

107. Ray, L. & Moore, B. S. Recent advances in the biosynthesis of unusual polyketide synthase substrates. *Nat. Prod. Rep.* **33,** 150–161 (2016).

108. Kellenberger, L. *et al.* A polylinker approach to reductive loop swaps in modular polyketide synthases. *Chembiochem Eur. J. Chem. Biol.* **9,** 2740–2749 (2008).

109. Keatinge-Clay, A. T. The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29,** 1050–1073 (2012).

110. Wang, H. *et al.* Genetic Screening Strategy for Rapid Access to Polyether Ionophore Producers and Products in Actinomycetes. *Appl. Environ. Microbiol.* **77,** 3433–3442 (2011).

111. Mai, T. T. *et al.* Salinomycin kills cancer stem cells by sequestering iron in lysosomes. *Nat. Chem.* **9,** 1025–1033 (2017).

112. Sala, C. *et al.* Simple Model for Testing Drugs against Nonreplicating Mycobacterium tuberculosis. *Antimicrob. Agents Chemother.* **54,** 4150–4158 (2010).

113. Luhavaya, H., Williams, S. R., Hong, H., Gonzaga de Oliveira, L. & Leadlay, P. F. Site-Specific Modification of the Anticancer and Antituberculosis Polyether Salinomycin by Biosynthetic Engineering. *ChemBioChem* **15,** 2081–2085 (2014).

114. Gallimore, A. R. The biosynthesis of polyketide-derived polycyclic ethers. *Nat. Prod. Rep.* **26,** 266–280 (2009).

115. Westley, J. W. *et al.* Isolation and characterization of the first halogen containing polyether antibiotic X-14766A, a product of Streptomyces malachitofuscus subsp. downeyi. *J. Antibiot. (Tokyo)* **34,** 139–147 (1981).

116. Westley, J. W. Polyether Antibiotics — Biosynthesis. in *Biosynthesis* (ed. Corcoran, P. J. W.) 41–73 (Springer Berlin Heidelberg, 1981).

117. Haney, M. E. & Hoehn, M. M. Monensin, a new biologically active compound. I. Discovery and isolation. *Antimicrob. Agents Chemother.* **7,** 349–352 (1967).

118. Leadlay, P. F. *et al.* Engineering of complex polyketide biosynthesis — insights from sequencing of the monensin biosynthetic gene cluster. *J. Ind. Microbiol. Biotechnol.* **27,** 360–367 (2001).

119. Hüttel, W., Spencer, J. B. & Leadlay, P. F. Intermediates in monensin biosynthesis: A late step in biosynthesis of the polyether ionophore monensin is crucial for the integrity of cation binding. *Beilstein J. Org. Chem.* **10,** 361–368 (2014).

120. Gallimore, A. R. *et al.* Evidence for the Role of the monB Genes in Polyether Ring Formation during Monensin Biosynthesis. *Chem. Biol.* **13,** 453–460 (2006).

121. Harvey, B. M. *et al.* Evidence that a novel thioesterase is responsible for polyketide chain release during biosynthesis of the polyether ionophore monensin. *Chembiochem Eur. J. Chem. Biol.* **7,** 1435–1442 (2006).

122. Bruheim, P. *et al.* Chemical Diversity of Polyene Macrolides Produced by Streptomyces noursei ATCC 11455 and Recombinant Strain ERD44 with Genetically Altered Polyketide Synthase NysC. *Antimicrob. Agents Chemother.* **48,** 4120–4129 (2004).

123. Zotchev, S. B. Polyene macrolide antibiotics and their applications in human therapy. *Curr. Med. Chem.* **10,** 211–223 (2003).

124. Fjærvik, E. & Zotchev, S. B. Biosynthesis of the polyene macrolide antibiotic nystatin in Streptomyces noursei. *Appl. Microbiol. Biotechnol.* **67,** 436–443 (2005).

125. Abu-Salah, K. M. Amphotericin B: an update. *Br. J. Biomed. Sci.* **53,** 122–133 (1996).

126. Whitfield, G. B., Brock, T. D., Ammann, A., Gottlieb, D. & Carter, H. E. Filipin, an Antifungal Antibiotic: Isolation and Properties. *J. Am. Chem. Soc.* **77,** 4799–4801 (1955).

127. Payero, T. D. *et al.* Functional analysis of filipin tailoring genes from Streptomyces filipinensis reveals alternative routes in filipin III biosynthesis and yields bioactive derivatives. *Microb. Cell Factories* **14,** (2015).

128. Aparicio, J. F., Mendes, M. V., Antón, N., Recio, E. & Martín, J. F. Polyene macrolide antibiotic biosynthesis. *Curr. Med. Chem.* **11,** 1645–1656 (2004).

129. Gimpl, G. & Gehrig-Burger, K. Probes for studying cholesterol binding and cell biology. *Steroids* **76,** 216–231 (2011).

130. DiLella, A. G. Chromosomal assignment of the human immunophilin FKBP-12 gene. *Biochem. Biophys. Res. Commun.* **179,** 1427–1433 (1991).

131. Rowe, C. J. *et al.* Engineering a polyketide with a longer chain by insertion of an extra module into the erythromycin-producing polyketide synthase. *Chem. Biol.* **8,** 475–485 (2001).

132. Gregory, M. A. *et al.* Mutasynthesis of Rapamycin Analogues through the Manipulation of a Gene Governing Starter Unit Biosynthesis. *Angew. Chem. Int. Ed.* **44,** 4757–4760 (2005).

133. Gregory, M. A. *et al.* Structure guided design of improved anti-proliferative rapalogs through biosynthetic medicinal chemistry. *Chem. Sci.* **4,** 1046–1052 (2013).

134. Haydock, S. F. *et al.* Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett.* **374,** 246–248 (1995).

135. Tang, Y., Kim, C.-Y., Mathews, I. I., Cane, D. E. & Khosla, C. The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 11124–11129 (2006).

136. Reid, R. *et al.* A Model of Structure and Catalysis for Ketoreductase Domains in Modular Polyketide Synthases. *Biochemistry* **42,** 72–79 (2003).

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

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

139. Bonnett, S. A. *et al.* Structural and stereochemical analysis of a modular polyketide synthase ketoreductase domain required for the generation of a cis-alkene. *Chem. Biol.* **20,** 772–783 (2013).

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

141. Labonte, J. W. & Townsend, C. A. Active Site Comparisons and Catalytic Mechanisms of the Hot Dog Superfamily. *Chem. Rev.* **113,** 2182–2204 (2013).

142. Palaniappan, N., Alhamadsheh, M. M. & Reynolds, K. A. cis-Δ2,3-Double Bond of Phoslactomycins Is Generated by a Post-PKS Tailoring Enzyme. *J. Am. Chem. Soc.* **130,** 12236–12237 (2008).

143. Mamoun M. Alhamadsheh, Nadaraj Palaniappan, Suparna DasChouduri, ‡ and & Reynolds*, K. A. Modular Polyketide Synthases and cis Double Bond Formation: Establishment of Activated cis-3-Cyclohexylpropenoic Acid as the Diketide Intermediate in Phoslactomycin Biosynthesis. (2007). doi:10.1021/ja068818t

144. Kwan, D. H. *et al.* Prediction and Manipulation of the Stereochemistry of Enoylreduction in Modular Polyketide Synthases. *Chem. Biol.* **15,** 1231–1240 (2008).

145. Zheng, J., Gay, D. C., Demeler, B., White, M. A. & Keatinge-Clay, A. T. Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol.* **8,** 615–621 (2012).

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

147. Wilkinson, B. *et al.* Novel octaketide macrolides related to 6-deoxyerythronolide B provide evidence for iterative operation of the erythromycin polyketide synthase. *Chem. Biol.* **7,** 111–117 (2000).

148. Moss, S. J., Martin, C. J. & Wilkinson, B. Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Nat. Prod. Rep.* **21,** 575–593 (2004).

149. Olano, C. *et al.* Evidence from engineered gene fusions for the repeated use of a module in a modular polyketide synthase. *Chem. Commun.* **0,** 2780–2782 (2003).

150. Xu, W. *et al.* An Iterative Module in the Azalomycin F Polyketide Synthase Contains a Switchable Enoylreductase Domain. *Angew. Chem. Int. Ed Engl.* **56,** 5503–5506 (2017).

151. Weissman, K. J. The structural basis for docking in modular polyketide biosynthesis. *Chembiochem Eur. J. Chem. Biol.* **7,** 485–494 (2006).

152. Keatinge-Clay, A. Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* **384,** 941–953 (2008).

153. Donadio, S., McAlpine, J. B., Sheldon, P. J., Jackson, M. & Katz, L. An erythromycin analog produced by reprogramming of polyketide synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 7119–7123 (1993).

154. Katz, L. & Ashley, G. W. Translation and Protein Synthesis: Macrolides. *Chem. Rev.* **105,** 499–528 (2005).

155. Kuhstoss, S., Huber, M., Turner, J. R., Paschal, J. W. & Rao, R. N. Production of a novel polyketide through the construction of a hybrid polyketide synthase. *Gene* **183,** 231–236 (1996).

156. Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed Engl.* **48,** 4688–4716 (2009).

157. Weissman, K. J. & Leadlay, P. F. Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Microbiol.* **3,** 925–936 (2005).

158. Helfrich, E. J. N. & Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **33,** 231–316 (2016).

159. Musiol, E. M. *et al.* Supramolecular Templating in Kirromycin Biosynthesis: The Acyltransferase KirCII Loads Ethylmalonyl-CoA Extender onto a Specific ACP of the trans-AT PKS. *Chem. Biol.* **18,** 438–444 (2011).

160. Stevens, D. C., Wagner, D. T., Manion, H. R., Alexander, B. K. & Keatinge-Clay, A. T. Methyltransferases excised from *trans*-AT polyketide synthases operate on *N*-acetylcysteamine-bound substrates. *J. Antibiot. (Tokyo)* **69,** 567–570 (2016).

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

162. Lohman, J. R. *et al.* Structural and evolutionary relationships of "AT-less" type I polyketide synthase ketosynthases. *Proc. Natl. Acad. Sci.* **112,** 12693–12698 (2015).

163. Cole, S. T. *et al.* Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393,** 537–544 (1998).

164. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **417,** 141–147 (2002).

165. Ikeda, H. *et al.* Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. *Nat. Biotechnol.* **21,** 526–531 (2003).

166. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437,** 376–380 (2005).

167. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–59 (2008).

168. Oliynyk, M. *et al.* Complete genome sequence of the erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338. *Nat. Biotechnol.* **25,** 447–453 (2007).

169. Ohnishi, Y. *et al.* Genome Sequence of the Streptomycin-Producing Microorganism Streptomyces griseus IFO 13350. *J. Bacteriol.* **190,** 4050–4060 (2008).

170. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323,** 133–138 (2009).

171. Feng, Y., Zhang, Y., Ying, C., Wang, D. & Du, C. Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinformatics* **13,** 4–16 (2015).

172. Schadt, E. E., Turner, S. & Kasarskis, A. A window into third-generation sequencing. *Hum. Mol. Genet.* **19,** R227-240 (2010).

173. Gomez-Escribano, J. P., Alt, S. & Bibb, M. J. Next Generation Sequencing of Actinobacteria for the Discovery of Novel Natural Products. *Mar. Drugs* **14,** (2016).

174. Nett, M., Ikeda, H. & Moore, B. S. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **26,** 1362–1384 (2009).

175. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).

176. Boddy, C. N. Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. *J. Ind. Microbiol. Biotechnol.* **41,** 443–450 (2014).

177. Medema, M. H. *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39,** W339–W346 (2011).

178. Blin, K. *et al.* antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45,** W36–W41 (2017).

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

180. Claesen, J. & Bibb, M. J. Biosynthesis and regulation of grisemycin, a new member of the linaridin family of ribosomally synthesized peptides produced by Streptomyces griseus IFO 13350. *J. Bacteriol.* **193,** 2510–2516 (2011).

181. Jiang, Y. *et al.* Identification and characterization of the cuevaene A biosynthetic gene cluster in streptomyces sp. LZ35. *Chembiochem Eur. J. Chem. Biol.* **14,** 1468–1475 (2013).

182. Kaeberlein, T., Lewis, K. & Epstein, S. S. Isolating 'Uncultivable' Microorganisms in Pure Culture in a Simulated Natural Environment. *Science* **296,** 1127–1129 (2002).

183. Feng, Z., Kallifidas, D. & Brady, S. F. Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 12629–12634 (2011).

184. Chang, F.-Y. & Brady, S. F. Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 2478–2483 (2013).

185. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517,** 455–459 (2015).

186. Gaisser, S. *et al.* Direct production of ivermectin-like drugs after domain exchange in the avermectin polyketide synthase of Streptomyces avermitilis ATCC31272. *Org. Biomol. Chem.* **1,** 2840–2847 (2003).

187. Martin, C. J. *et al.* Heterologous expression in Saccharopolyspora erythraea of a pentaketide synthase derived from the spinosyn polyketide synthase. *Org. Biomol. Chem.* **1,** 4144–4147 (2003).

188. Del Vecchio, F. *et al.* Active-site residue, domain and module swaps in modular polyketide synthases. *J. Ind. Microbiol. Biotechnol.* **30,** 489–494 (2003).

189. Petkovic, H. *et al.* A novel erythromycin, 6-desmethyl erythromycin D, made by substituting an acyltransferase domain of the erythromycin polyketide synthase. *J. Antibiot. (Tokyo)* **56,** 543–551 (2003).

190. Zhang, Z. *et al.* Non-benzoquinone geldanamycin analogs trigger various forms of death in human breast cancer cells. *J. Exp. Clin. Cancer Res. CR* **35,** (2016).

191. McDaniel, R., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. Engineered biosynthesis of novel polyketides. *Science* **262,** 1546–1550 (1993).

192. Kao, C. M., Katz, L. & Khosla, C. Engineered biosynthesis of a complete macrolactone in a heterologous host. *Science* **265,** 509–512 (1994).

193. Komatsu, M., Uchiyama, T., Ōmura, S., Cane, D. E. & Ikeda, H. Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 2646–2651 (2010).

194. Jones, A. C. *et al.* Phage P1-Derived Artificial Chromosomes Facilitate Heterologous Expression of the FK506 Gene Cluster. *PLOS ONE* **8,** e69319 (2013).

195. Rodriguez, E. *et al.* Rapid engineering of polyketide overproduction by gene transfer to industrially optimized strains. *J. Ind. Microbiol. Biotechnol.* **30,** 480–488 (2003).

196. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science* **291,** 1790–1792 (2001).

197. Peirú, S., Menzella, H. G., Rodríguez, E., Carney, J. & Gramajo, H. Production of the potent antibacterial polyketide erythromycin C in Escherichia coli. *Appl. Environ. Microbiol.* **71,** 2539–2547 (2005).

198. Wattanachaisaereekul, S., Lantz, A. E., Nielsen, M. L. & Nielsen, J. Production of the polyketide 6-MSA in yeast engineered for increased malonyl-CoA supply. *Metab. Eng.* **10,** 246–254 (2008).

199. Kevin R., O., Vo, K. T., Michaelis, S. & Paddon, C. Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Res.* **25,** 451–452 (1997).

200. Joska, T. M., Mashruwala, A., Boyd, J. M. & Belden, W. J. A universal cloning method based on yeast homologous recombination that is simple, efficient, and versatile. *J. Microbiol. Methods* **100,** 46–51 (2014).

201. Wlodek, A. *et al.* Diversity oriented biosynthesis via accelerated evolution of modular gene clusters. *Nat. Commun.* **8,** 1206 (2017).

202. Cortes, J. *et al.* Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science* **268,** 1487–1489 (1995).

203. Lowry, B., Walsh, C. T. & Khosla, C. In Vitro Reconstitution of Metabolic Pathways: Insights into Nature's Chemical Logic. *Synlett Acc. Rapid Commun. Synth. Org. Chem.* **26,** 1008–1025 (2015).

204. Carreras, C. W. & Khosla, C. Purification and in vitro reconstitution of the essential protein components of an aromatic polyketide synthase. *Biochemistry* **37,** 2084–2088 (1998).

205. Zawada, R. J. & Khosla, C. Domain analysis of the molecular recognition features of aromatic polyketide synthase subunits. *J. Biol. Chem.* **272,** 16184–16188 (1997).

206. Zawada, R. J. & Khosla, C. Heterologous expression, purification, reconstitution and kinetic analysis of an extended type II polyketide synthase. *Chem. Biol.* **6,** 607–615 (1999).

207. Bierman, M. *et al.* Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. *Gene* **116,** 43–49 (1992).

208. Sun, Y. *et al.* Organization of the biosynthetic gene cluster in Streptomyces sp. DSM 4137 for the novel neuroprotectant polyketide meridamycin. *Microbiology* **152,** 3507–3515 (2006).

209. MacNeil, D. J. *et al.* Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111,** 61–68 (1992).

210. Li, W. *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **43,** W580-4 (2015).

211. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinforma. Oxf. Engl.* **23,** 1289–1291 (2007).

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

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

214. Muth, G., Farr, M., Hartmann, V. & Wohlleben, W. Streptomyces ghanaensis Plasmid pSG5: Nucleotide Sequence Analysis of the Self-Transmissible Minimal Replicon and Characterization of the Replication Mode. *Plasmid* **33,** 113–126 (1995).

215. Cook, D. M. & Farrand, S. K. The oriT region of the Agrobacterium tumefaciens Ti plasmid pTiC58 shares DNA sequence identity with the transfer origins of RSF1010 and RK2/RP4 and with T-region borders. *J. Bacteriol.* **174,** 6238–6246 (1992).

216. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).

217. Brautaset, T. *et al.* Hexaene Derivatives of Nystatin Produced as a Result of an Induced Rearrangement within the nysC Polyketide Synthase Gene in S. noursei ATCC 11455. *Chem. Biol.* **9,** 367–373 (2002).

218. Cobb, R. E., Wang, Y. & Zhao, H. High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. *ACS Synth. Biol.* (2014). doi:10.1021/sb500351f

219. Jack Lin's CRISPR/Cas9 gRNA Finder. Available at: http://spot.colorado.edu/~slin/cas9.html. (Accessed: 10th June 2016)

220. Huang, H., Zheng, G., Jiang, W., Hu, H. & Lu, Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in Streptomyces. *Acta Biochim. Biophys. Sin.* **47,** 231–243 (2015).

221. Sun, F.-H., Luo, D., Shu, D., Zhong, J. & Tan, H. Development of an Intergeneric Conjugal Transfer System for Xinaomycins-Producing Streptomyces noursei Xinao-4. *Int. J. Mol. Sci.* **15,** 12217–12230 (2014).

222. Bhatt, A. *et al.* Accumulation of an E,E,E-Triene by the Monensin-Producing Polyketide Synthase when Oxidative Cyclization is Blocked. *Angew. Chem. Int. Ed.* **44,** 7075–7078 (2005).

223. Rodriguez, E. *et al.* Rapid engineering of polyketide overproduction by gene transfer to industrially optimized strains. *J. Ind. Microbiol. Biotechnol.* **30,** 480–488 (2003).

224. Seeger, K. *et al.* The biosynthetic genes for prenylated phenazines are located at two different chromosomal loci of Streptomyces cinnamonensis DSM 1042. *Microb. Biotechnol.* **4,** 252–262 (2011).

225. Du, D. *et al.* Genome engineering and direct cloning of antibiotic gene clusters via phage φBT1 integrase-mediated site-specific recombination in Streptomyces. *Sci. Rep.* **5,** (2015).

226. Simon, J. R. & Moore, P. D. Homologous recombination between single-stranded DNA and chromosomal genes in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **7,** 2329–2334 (1987).

227. Yoo, H.-G. *et al.* Characterization of 2-octenoyl-CoA carboxylase/reductase utilizing pteB from Streptomyce avermitilis. *Biosci. Biotechnol. Biochem.* **75,** 1191–1193 (2011).

228. Zhang, L. *et al.* The adnAB Locus, Encoding a Putative Helicase-Nuclease Activity, Is Essential in Streptomyces. *J. Bacteriol.* **196,** 2701–2708 (2014).

229. Hoff, G., Bertrand, C., Piotrowski, E., Thibessard, A. & Leblond, P. Implication of RuvABC and RecG in homologous recombination in Streptomyces ambofaciens. *Res. Microbiol.* **168,** 26–35 (2017).

230. Hong, H., Samborskyy, M., Usachova, K., Schnatz, K. & Leadlay, P. F. Sulfation and amidinohydrolysis in the biosynthesis of giant linear polyenes. *Beilstein J. Org. Chem.* **13,** 2408–2415 (2017).

231. *The Prokaryotes: A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications*. (Springer-Verlag, 1992).

232. Pesic, A. *et al.* Isolation and structure elucidation of the nucleoside antibiotic strepturidin from Streptomyces albus DSM 40763. *J. Antibiot. (Tokyo)* **67,** 471–477 (2014).

233. Maffioli, S. I. *et al.* Antibacterial Nucleoside-Analog Inhibitor of Bacterial RNA Polymerase. *Cell* **169,** 1240-1248.e23 (2017).

234. Landwehr, W., Wolf, C. & Wink, J. Actinobacteria and Myxobacteria-Two of the Most Important Bacterial Resources for Novel Antibiotics. *Curr. Top. Microbiol. Immunol.* **398,** 273–302 (2016).

235. Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinforma. Oxf. Engl.* **16,** 944–945 (2000).

236. Labeda, D. P. *et al.* Phylogenetic relationships in the family Streptomycetaceae using multi-locus sequence analysis. *Antonie Van Leeuwenhoek* **110,** 563–583 (2017).

237. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci. CABIOS* **8,** 275–282 (1992).

238. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33,** 1870–1874 (2016).

239. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* 9 (2004).

240. Skinnider, M. A., Johnston, C. W., Merwin, N. J., Dejong, C. A. & Magarvey, N. A. Global analysis of prokaryotic tRNA-derived cyclodipeptide biosynthesis. *BMC Genomics* **19,** (2018).

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

242. van Heel, A. J., de Jong, A., Montalbán-López, M., Kok, J. & Kuipers, O. P. BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res.* **41,** W448–W453 (2013).

243. Louis, P. & Galinski, E. A. Characterization of genes for the biosynthesis of the compatible solute ectoine from Marinococcus halophilus and osmoregulated expression in Escherichia coli. *Microbiology* **143,** 1141–1149 (1997).

244. Yu, Y., Zhang, Q. & van der Donk, W. A. Insights into the evolution of lanthipeptide biosynthesis. *Protein Sci. Publ. Protein Soc.* **22,** 1478–1489 (2013).

245. Maksimov, M. O. & Link, A. J. Prospecting genomes for lasso peptides. *J. Ind. Microbiol. Biotechnol.* **41,** 333–344 (2014).

246. Itoh, T. *et al.* Actinorhodin Biosynthesis: Structural Requirements for Post-PKS Tailoring Intermediates Revealed by Functional Analysis of ActVI-ORF1 Reductase. *Biochemistry* **46,** 8181–8188 (2007).

247. Du, L., Sánchez, C. & Shen, B. Hybrid Peptide–Polyketide Natural Products: Biosynthesis and Prospects toward Engineering Novel Molecules. *Metab. Eng.* **3,** 78–95 (2001).

248. Gaisser, S., Trefzer, A., Stockert, S., Kirschning, A. & Bechthold, A. Cloning of an avilamycin biosynthetic gene cluster from Streptomyces viridochromogenes Tü57. *J. Bacteriol.* **179,** 6271–6278 (1997).

249. Ahlert, J. *et al.* The Calicheamicin Gene Cluster and Its Iterative Type I Enediyne PKS. *Science* **297,** 1173–1176 (2002).

250. Zhang, Q., Doroghazi, J. R., Zhao, X., Walker, M. C. & Donk, W. A. van der. Expanded Natural Product Diversity Revealed by Analysis of Lanthipeptide-Like Gene Clusters in Actinobacteria. *Appl. Environ. Microbiol.* **81,** 4339–4350 (2015).

251. Minowa, Y., Araki, M. & Kanehisa, M. Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes. *J. Mol. Biol.* **368,** 1500–1517 (2007).

252. Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat Prod Rep* **24,** 162–190 (2007).

253. Lai, C.-Y. & Cronan, J. E. β-Ketoacyl-Acyl Carrier Protein Synthase III (FabH) Is Essential for Bacterial Fatty Acid Synthesis. *J. Biol. Chem.* **278,** 51494–51503 (2003).

254. Wu, K., Chung, L., Revill, W. P., Katz, L. & Reeves, C. D. The FK520 gene cluster of Streptomyces hygroscopicus var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. *Gene* **251,** 81–90 (2000).

255. Dorrestein, P. C. *et al.* The bifunctional glyceryl transferase/phosphatase OzmB belonging to the HAD superfamily that diverts 1,3-bisphosphoglycerate into polyketide biosynthesis. *J. Am. Chem. Soc.* **128,** 10386–10387 (2006).

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

257. Thompson, T. B., Katayama, K., Watanabe, K., Hutchinson, C. R. & Rayment, I. Structural and functional analysis of tetracenomycin F2 cyclase from Streptomyces glaucescens. A type II polyketide cyclase. *J. Biol. Chem.* **279,** 37956–37963 (2004).

258. Terui, Y. *et al.* Xantholipin, a novel inhibitor of HSP47 gene expression produced by Streptomyces sp. *Tetrahedron Lett.* **44,** 5427–5430 (2003).

259. Tang, Y., Tsai, S.-C. & Khosla, C. Polyketide Chain Length Control by Chain Length Factor. *J. Am. Chem. Soc.* **125,** 12708–12709 (2003).

260. Taguchi, T. *et al.* Biosynthetic conclusions from the functional dissection of oxygenases for biosynthesis of actinorhodin and related Streptomyces antibiotics. *Chem. Biol.* **20,** 510–520 (2013).

261. Werner, G., Hagenmaier, H., Drautz, H., Baumgartner, A. & Zähner, H. Metabolic products of microorganisms. 224. Bafilomycins, a new group of macrolide antibiotics. Production, isolation, chemical structure and biological activity. *J. Antibiot. (Tokyo)* **37,** 110–117 (1984).

262. Vilches, C., Hernandez, C., Mendez, C. & Salas, J. A. Role of glycosylation and deglycosylation in biosynthesis of and resistance to oleandomycin in the producer organism, Streptomyces antibioticus. *J. Bacteriol.* **174,** 161–165 (1992).

263. Thibodeaux, C. J., Melançon, C. E. & Liu, H. Natural Product Sugar Biosynthesis and Enzymatic Glycodiversification. *Angew. Chem. Int. Ed Engl.* **47,** 9814–9859 (2008).

264. Handa, N. *et al.* Crystal structure of a novel polyisoprenoid-binding protein from Thermus thermophilus HB8. *Protein Sci.* **14,** 1004–1010 (2005).

265. Chen, S. *et al.* Organizational and Mutational Analysis of a Complete FR-008/Candicidin Gene Cluster Encoding a Structurally Related Polyene Complex. *Chem. Biol.* **10,** 1065–1076 (2003).

266. Gondry, M. *et al.* A Comprehensive Overview of the Cyclodipeptide Synthase Family Enriched with the Characterization of 32 New Enzymes. *Front. Microbiol.* **9,** (2018).

267. Mishra, A. K., Choi, J., Choi, S.-J. & Baek, K.-H. Cyclodipeptides: An Overview of Their Biosynthesis and Biological Activity. *Mol. Basel Switz.* **22,** (2017).

268. Spenkuch, F., Motorin, Y. & Helm, M. Pseudouridine: still mysterious, but never a fake (uridine)! *RNA Biol.* **11,** 1540–1554 (2014).

269. Hamma, T. & Ferré-D'Amaré, A. R. Pseudouridine Synthases. *Chem. Biol.* **13,** 1125–1135 (2006).

270. Kaya, Y., Del Campo, M., Ofengand, J. & Malhotra, A. Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold. *J. Biol. Chem.* **279,** 18107–18110 (2004).

271. Isono, K. Nucleoside antibiotics: structure, biological activity, and biosynthesis. *J. Antibiot. (Tokyo)* **41,** 1711–1739 (1988).

272. Niu, G. & Tan, H. Nucleoside antibiotics: biosynthesis, regulation, and biotechnology. *Trends Microbiol.* **23,** 110–119 (2015).

273. Winn, M., Goss, R. J. M., Kimura, K. & Bugg, T. D. H. Antimicrobial nucleoside antibiotics targeting cell wall assembly: Recent advances in structure–function studies and nucleoside biosynthesis. *Nat. Prod. Rep.* **27,** 279–304 (2010).

274. Nix, D. E., Swezey, R. R., Hector, R. & Galgiani, J. N. Pharmacokinetics of Nikkomycin Z after Single Rising Oral Doses. *Antimicrob. Agents Chemother.* **53,** 2517–2521 (2009).

275. Holden, W. M., Fites, J. S., Reinert, L. K. & Rollins-Smith, L. A. Nikkomycin Z is an effective inhibitor of the chytrid fungus linked to global amphibian declines. *Fungal Biol.* **118,** 48–60 (2014).

276. Bormann, C., Mattern, S., Schrempf, H., Fiedler, H. P. & Zähner, H. Isolation of Streptomyces tendae mutants with an altered nikkomycin spectrum. *J. Antibiot. (Tokyo)* **42,** 913–918 (1989).

277. Lauer, B. *et al.* Molecular characterization of co-transcribed genes from Streptomyces tendae Tü901 involved in the biosynthesis of the peptidyl moiety and assembly of the peptidyl nucleoside antibiotic nikkomycin. *Mol. Gen. Genet. MGG* **264,** 662–673 (2001).

278. Evans, D. R. *et al.* The biosynthesis of nikkomycin X from histidine in Streptomyces tendae. *Tetrahedron Lett.* **36,** 2351–2354 (1995).

279. Isono, K., Sato, T., Hirasawa, K., Funayama, S. & Suzuki, S. Biosynthesis of the nucleoside skeleton of polyoxins. *J. Am. Chem. Soc.* **100,** 3937–3939 (1978).

280. Maffioli, S. I. *et al.* Antibacterial Nucleoside-Analog Inhibitor of Bacterial RNA Polymerase. *Cell* **169,** 1240-1248.e23 (2017).

281. Schonell, M., Dorken, E. & Grzybowski, S. Rifampin. *Can. Med. Assoc. J.* **106,** 783–786 (1972).

282. McClure, W. R. & Cech, C. L. On the mechanism of rifampicin inhibition of RNA synthesis. *J. Biol. Chem.* **253,** 8949–8956 (1978).

283. Campbell, E. A. *et al.* Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell* **104,** 901–912 (2001).

284. Ezekiel, D. H. & Hutchins, J. E. Mutations affecting RNA polymerase associated with rifampicin resistance in Escherichia coli. *Nature* **220,** 276–277 (1968).

285. Lin, W. *et al.* Structural Basis of Transcription Inhibition by Fidaxomicin (Lipiarmycin A3). *Mol. Cell* **70,** 60-71.e15 (2018).

286. Davis, F. F. & Allen, F. W. Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* **227,** 907–915 (1957).

287. Johnson, L. & Soll, D. In vitro Biosynthesis of Pseudouridine at the Polynucleotide Level by an Enzyme Extract from Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **67,** 943–950 (1970).

288. Arena, F., Ciliberto, G., Ciampi, S. & Cortese, R. Purification of pseudouridylate synthetase I from Salmonella typhimurium. *Nucleic Acids Res.* **5,** 4523–4536 (1978).

289. Samuelsson, T. & Olsson, M. Transfer RNA pseudouridine synthases in Saccharomyces cerevisiae. *J. Biol. Chem.* **265,** 8782–8787 (1990).

290. Cortese, R., Kammen, H. O., Spengler, S. J. & Ames, B. N. Biosynthesis of Pseudouridine in Transfer Ribonucleic Acid. *J. Biol. Chem.* **249,** 1103–1108 (1974).

291. Ericsson, U. B., Nordlund, P. & Hallberg, B. M. X-ray structure of tRNA pseudouridine synthase TruD reveals an inserted domain with a novel fold. *FEBS Lett.* **565,** 59–64 (2004).

292. KAYA, Y. & OFENGAND, J. A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *RNA* **9,** 711–721 (2003).

293. Humm, A., Fritsche, E., Steinbacher, S. & Huber, R. Crystal structure and mechanism of human L-arginine:glycine amidinotransferase: a mitochondrial enzyme involved in creatine biosynthesis. *EMBO J.* **16,** 3373–3385 (1997).

294. Llewellyn, N. M. & Spencer, J. B. Biosynthesis of 2-deoxystreptamine-containing aminoglycoside antibiotics. *Nat. Prod. Rep.* **23,** 864–874 (2006).

295. Huang, Y. *et al.* Pyridoxal-5′-phosphate as an oxygenase cofactor: Discovery of a carboxamide-forming, α-amino acid monooxygenase-decarboxylase. *Proc. Natl. Acad. Sci.* **115,** 974-979 (2018).

296. Sosio, M. *et al.* Analysis of the Pseudouridimycin Biosynthetic Pathway Provides Insights into the Formation of C-nucleoside Antibiotics. *Cell Chem. Biol.* **25,** 540-549.e4 (2018).

297. Li, M., Chen, L., Deng, Z. & Zhao, C. Characterization of AmtA, an amidinotransferase involved in the biosynthesis of phaseolotoxins. *FEBS Open Bio* **6,** 603–609 (2016).

298. Bhushan, R. & Brückner, H. Marfey's reagent for chiral amino acid analysis: a review. *Amino Acids* **27,** 231–247 (2004).

299. Thapa, K., Oja, T. & Metsä-Ketelä, M. Molecular evolution of the bacterial pseudouridine-5′-phosphate glycosidase protein family. *FEBS J.* **281,** 4439–4449 (2014).

300. Svenningsen, S. L., Kongstad, M., Stenum, T. S., Muñoz-Gómez, A. J. & Sørensen, M. A. Transfer RNA is highly unstable during early amino acid starvation in Escherichia coli. *Nucleic Acids Res.* **45,** 793–804 (2017).

Appendix

Primer	Sequence
Primers for nystatin gene cluster rearrangement work	
NysC-flank-F	atttgttcagaacgctcggttgccgccgggcgtttttttatgatcatatagtgcgggctgg
NysC-flank-R	cacagcccgatcatgtgc
NysA-flank-F	tttccgagccacggaattcattgcacatgatcgggctgtgggtgacccttccttccgg
NysA-flank-R	acgccagcaacgcggccttttttacggttcctggcctctagctttccgaagtacaggccc
sgRNA insert	gagacatctttgaagacaaacgctccctttctttagaagaggcgttttagagctagaaatagcaagttaa aataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgctttttttagcataacccccttggg gcctctaaacgggtcttgaggggttttttggctgctccttcggtcggacgtgcgtctacgggcaccttaccg cagccgtcggctgtgcgacacggacggatcgggcgaactggccgatgctgggagaagcgcgctgctgta cggcgcgcaccgggtgcggagcccctcggcgagcggtgtgaaacttctgtgaatggcctgttcggttgct ttttttatacggctgccagataaggcttgcagcatctgggcggctaccgctatgatcggggcgttcctgca attcttagtgcgagtatctgaaaggggatacgcactagtgattccccggtccggtttaagtcttctttcacg tggc
g-check-F	gtcattgtcggcgcaactat
g-check-R	ttttgtgatgctcgtcaggg
Primer for Accelerated Evolution work	
NysKS6-F	acagctatgacatgattacgcacggtctctgctgctcc
NysKS6-R	cgcggccgcgcgcgatatcggagttgttcggcgcggtc
EryKS6-F	acagctatgacatgattacgcgacagcagtacgggcac
EryKS6-R	cgcggccgcgcgcgatatcgttcgactccatgaccgcc
EryKS-AT6-F	gcttgggctgcaggtcgactaggtcctggaaacgacgtg
EryKS-AT6-R	acagctatgacatgattacgaagacgttccgccagtcc
EryAT-KR6-F	gcttgggctgcaggtcgactctgaccgacctgctcagc
EryAT-KR6-R	acagctatgacatgattacggtgatgtgcagttcctcgac
EryACP-KS6-F	gcttgggctgcaggtcgactgaactgcacatcaccctcga
EryACP-KS6-R	acagctatgacatgattacgcgacagcagtacgggcac
PreKS2-F	acagctatgacatgattacgggtcaccaactcacgcaaac
PreKS2-R	cgcggccgcgcgcgatatcgcaccctcaccttcgactacc
PreKS-AT2-F	gcttgggctgcaggtcgactcatgacggggcaggactac
PreKS-AT2-R	acagctatgacatgattacgctgctggcctcgatgaaga
PreACP-KS-F	gcttgggctgcaggtcgactctccccacgtatcccttcc
PreACP-KS-R	acagctatgacatgattacgggtcaccaactcacgcaaac
FilKS11-F	gcttgggctgcaggtcgacttgctcagctgctcggttc
FilKS11-R	acagctatgacatgattacgcgacaacacccacggcac
FilKS-AT11-F	gcttgggctgcaggtcgacttgttcgtcggcatgtcctac
FilKS-AT11-R	acagctatgacatgattacgggtctccagccagtaccg
FilAT-KR11-F	gcttgggctgcaggtcgactcatgtggcgggtgtgttg
FilAT-KR11-R	acagctatgacatgattacggaaccgctcccagtccac
FilACP-KS11-F	gcttgggctgcaggtcgactgcgttcgtcatgttctcctc
FilACP-KS11-R	acagctatgacatgattacgcgacaacacccacggcac
pKC-check-F	gttgtaaaacgacggccagt
pKC-check-R	cttccggctcgtatgttgtg

Primers for heterologous expression of pseudouridimycin biosynthetic genes

psmH-F	ctggtgccgcgcggcagccatatgaaccaaccggcgaccct
psmH-R	ttgtcgacggagctcgaattcgacggacggggagcggct
psmI-F	ctggtgccgcgcggcagccatatgcccctcacccagcgg
psmI-R	ttgtcgacggagctcgaattccgctcagacgccggcct
psmK-F	ctggtgccgcgcggcagccatatggcgctgctgctgctcaa
psmK-R	ttgtcgacggagctcgaattcgccgcccagtagagccggt
psmL-F	ctggtgccgcgcggcagccatatgacgaccggcaacccg
psmL-R	ttgtcgacggagctcgaattcagagggtgtcgatgagtcc
psmR-F	ctggtgccgcgcggcagccatatgtcctcgaccacttcccc
psmR-R	ttgtcgacggagctcgaattcggtggaggttcttcgcaatg
psmM-F	ctggtgccgcgcggcagccatatgccgcacggcttcga
psmM-R	ttgtcgacggagctcgaattcgttcgtggtgtgtgtggtg
psmO-F	ctggtgccgcgcggcagccatatgcgtcagcccgccga
psmO-R	ttgtcgacggagctcgaattcggcaagcggtcctactgg

Primers for knocking out pseudouridimycin biosynthetic genes

psmNL-F	ggggacctgcaggtcgactctagaagcagctgttcggtcacccc
psmNL-R	acccgagaggaacgcgcccgtgacaccccgagaacatcctgcgc
psmNR-F	cgtagcgcaggatgttctcggggtgtcacgggcgcgttcctctc
psmNR-R	cgccgaaagttcctcgaagctagcgcgcgggcctggtgaacata
psmUL-F	ggggacctgcaggtcgactctagacgcctactcccgggtccact
psmuL-R	gtccgtccgggtgatgtcgagccggtcttacaccggggcgcact
psmUR-F	gccaagtgcgccccggtgtaagaccggctcgacatcacccggac
psmUR-R	cgccgaaagttcctcgaagctagcagggcgacaaggagggcatg
psmBL-F	ggggacctgcaggtcgactctagacttcctcggtcgtgcggtcc
psmBL-R	gtgcgcgcgaagacctgccgttcggcaccctagacgagcacccc
psmBR-F	ccgaggggtgctcgtctagggtgccgaacggcaggtcttcgcg
psmBR-R	cgccgaaagttcctcgaagctagccagctcaccactgccggcca
psmNLcheck-F	gccgaagcgctgggtgtcgt
psmNLcheck-R	ggcgtgccggtcgtttccca
psmNRcheck-F	cgaaccggcacccaccacga
psmNRcheck-R	ccggttggtgggcggaggat
psmBLcheck-F	tcgggagggctcttcggtgc
psmBLcheck-R	tgaggtgagcgccccgtcga
psmBRcheck-F	gccctgcgccaccttctcct
psmBRcheck-R	tcggtgcagaaggtggcgtcc
psmULcheck-F	aacggctcgccctccaaccc
psmULcheck-R	acacgccgcggtccacttcg
psmURcheck-F	gccaagtgcgccccggtgta
psmURcheck-R	cggtgatctggaacgcgcgg
psmNKO-F	gttggtccagtggtcgtgcg
psmNKO-R	gagaacagcacacggtcggc
psmUKO-F	cgttcccgtaatcccgctgc
psmUKO-R	tccacttcgacacgccaggt
psmBKO-F	tgcgccaccttctccttcgt
psmBKO-R	agggccccggagatgttgtc