# Genome mining of rare actinomycetes

Eleni Vikeli

John Innes Centre

Department of Molecular Microbiology

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

September 2018

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

# Abstract

The rise of antibiotic-resistant microbes, combined with an absence of novel antimicrobials in the development pipeline, raises the spectre of a post-antibiotic era and has prompted a resurgence of interest in antibiotic discovery and development.

This project focuses on combining advanced genomics and bioinformatics with traditional bioassays and untargeted metabolomics, with the aim to explore the natural product repertoire of three *Saccharopolyspora* strains isolated from the cuticles of the Kenyan plant ants, *Tetraponera penzigi*. The three strains were sequenced using the PacBio RSII platform. Bioinformatics suggested the presence of at least 23 biosynthetic gene clusters (BGCs) in each strain. Interestingly, several of these are predicted to encode novel and uncommon members of chemotypes known to possess potent biological activity, including several potential anti-infective agents. We observed that the genomes of all three strains encode a cinnamycin-like BGC that proved silent despite culturing under a varied range of conditions. The isolate *Saccharopolyspora* sp. KY21 proved most tractable under laboratory conditions and was chosen as the basis for further study.

I report the activation of this BGC by expression of two native genes under a constitutive promoter, *kyaL* and *kyaR1*. *kyaL* encodes a phosphatidyl ethanolamine methyl transferase gene which represents a self-immunity mechanism; resistance – *kyaR1* encodes a pathway specific positive regulator of the SARP family. This led to the isolation and characterisation of kyamicin, a novel type B lantibiotic with activity against a range of Gram-positive bacteria. Furthermore, I describe the engineering of a platform for expression of type B lantibiotics, based on the kyamicin machinery.

Additionally, I report the discovery of a novel NRPS-derived siderophore via the exploitation of untargeted metabolomics data, as well as preliminary attempts to isolate an antifungal compound via a bioassay-guided approach.

Finally, I discuss the biosynthesis of sporeamicin from *Saccharopolyspora* sp. L53-18, an erythromycin-like macrolide antibiotic.

## Acknowledgements

I would like to express my vast gratitude to my primary supervisor Prof. Barrie Wilkinson for his unlimited support and faith since the beginning of my project. It was a great challenge for me completing this journey and I couldn't have asked for a better supervisor and mentor. I explored many different projects during my PhD and his guidance was indispensable in order to focus my science and also to follow my passion in science communication. I would also like to thank my secondary supervisor Prof. Matthew Hutchings for his advice and input on my projects as well as his group at UEA for our collaboration. In particular, Neil Holmes for helping me with the ant strains and being next to me at so many outreach events.

My thesis would not be possible without the constant support and help of my friends and colleagues, both past and present, from the Molecular Microbiology department. I would like to specifically thank Dr Silke Alt, Dr Siobhan Dorai-Raj, Dr Juan-Pablo Gomez-Escribano and Dr Govind Chandra for their help with the genome related questions and Dr Daniel Heine for helping me with all my chemistry related problems.

Special thanks to Dr. Javier Santos-Aberturas, Dr Sibyl Batey, Dr Natalia Miguel Vior and Dr David Widdick for their constant help and encouragement. Without them, this thesis wouldn't be possible.

I would also like to acknowledge Dr. Gerhard Saalbach and Dr Lionel Hill for their help with the analysis of my samples as well as Elaine Barclay for the SEM microscopy.

I would like to thank my mum Maria and my sister Dimitra for all their support and for putting up with me missing family holidays because I had to work in the lab. They missed me a great deal, but they were always sending positive thoughts.

Finally, I would like to thank my fiancé and soon-to-be husband, Oscar Gonzalez, who has been encouraging me constantly to pursue my dreams. He is always there for me in good times and bad times, helping me to put everyday life in perspective and to believe in myself.

## Author’s declaration

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

# Abbreviations

| | |
|---|---|
| **BGC** | Biosynthetic gene cluster |
| **2,3-DHBA** | 2,3-Dihydroxybenzoate |
| **DNA** | Deoxyribonucleic acid |
| **MIC** | Minimum inhibitory concentration |
| **NP** | Natural Product |
| **NRP** | Nonribosomal peptide |
| **RiPP** | Ribosomally synthesised and post-translationally modified peptide |
| **OSMAC** | One Strain–Many Compounds |

## NRPS/PKS domains

| | |
|---|---|
| **A** | Adenylation |
| **ACP** | Acyl carrier protein |
| **ArCP** | Aryl carrier protein |
| **C** | Condensation |
| **Cy** | Cyclisation |
| **DEBS** | erythromycin B synthetase |
| **DH** | Dehydratase |
| **E** | Epimerisation |
| **ER** | Enoyl reductase |
| **KR** | Ketoreductase |
| **KS** | Ketosynthase |
| **MT** | Methyltransferase |
| **PCP** | Peptidyl carrier protein |
| **TE** | Thioesterase |

# Table of Contents

# Index of Figures

# Index of Tables

# Chapter 1:

Introduction

# 1 Introduction

## 1.1 Antimicrobial natural products

Natural products (NPs) are a large group of metabolites that originate from bacteria, fungi, plants and some types of animals and have a broad range of functions. Strictly speaking NPs include both primary metabolites, which are essential for survival for the producer organism, and secondary metabolites, which are extremely specialised and in principle dispensable, but more commonly the term NP is reserved to this second type of molecules (Katz and Baltz, 2016). These organisms commonly produce NPs to establish an advantage in response to specific environmental circumstances, such as competition for resources, defence against predators, or as virulence factors in the case of pathogenic organisms (Mun et al., 2017, Saha et al., 2016, Baldeweg et al., 2018).

NP research explores the varied activity of these metabolites, with a focus on those with antimicrobial properties (Katz and Baltz, 2016). Antimicrobial describes compounds that kill or inhibit the growth of microbes, encompassing both antibiotics and antifungals. These are now fundamental to modern medicine, and practices such as orthopaedic prosthetic surgeries or transplants would be impossible without them. The rise in resistance to antimicrobials has driven the need for novel NPs with new mechanisms of action (Martens and Demain, 2017).

### 1.1.1 Brief history of antimicrobial NP discovery

In the beginning of the 20$^{th}$ century, Paul Erhlic coined the term ‘selective toxicity’ and introduced the concept of the ‘chemotherapeutic index’ (Krantz, 1974, Ehrlich, 1901). This was defined as the minimum effective dose of a given medicine divided by the maximum tolerated dose. Ehrlich developed a theory that specific microbes, harmful to humans, could be killed without harming the body itself, with the use of a hypothetical agent. He named the hypothetical agent Zauberkugel; the magic bullet (Witkop, 1999). The first so called magic bullet was Salvarsan, an antimicrobial organoarsenic compound developed by Ehrlich in 1909 for the treatment of syphilis. This discovery also led to the finding that microorganisms could become resistant to drugs (Kasten, 1996).

The great revolution in modern medicine and the use of antimicrobials occurred after 1928, when Sir Alexander Fleming discovered penicillin. Penicillin was the first NP antibiotic and it is produced by *Penicillium* fungi. In the 1930s Howard Florey and Ernst Chain purified penicillin at an industrial scale (Chain et al., 1940) and its structure was elucidated in 1949 by Dorothy Hodgkin (Hodgkin, 1949). Around the same time as

Fleming's work, the microbiologist Selman Waksman focused his research on the study of soil microorganisms and how they contribute to soil health and fertility. His extensive studies concentrated on actinomycetes bacteria (Waksman, 1961) and particularly on *Streptomyces*, which proved to be prolific producers of antibiotics. Waksman's work yielded 20 antibiotics and one in particular, streptomycin (Schatz et al., 1944) was the first antibiotic effective against tuberculosis (Waksman et al., 1946), leading to him being awarded a Nobel Prize in 1952 . After these major advances, the approach of scaling up antibiotic production by fermentation of microorganisms was widely exploited. The 1940s-1960s were known as the 'Golden Age' of antimicrobial discovery, where the major classes of antibiotics were discovered and introduced to the clinic (Jones et al., 2017). This led to a drop in infection-associated mortality and overall increase in life expectancy thanks to its help in advancing other areas of medicine. However, the antibiotic discovery production line slowed down dramatically at the end of the 1960s. In the following decades, only a few novel antibiotics such as the lipopeptide daptomycin, have been discovered and approved for clinical use (Laxminarayan, 2014). A timeline of antibiotics discovery is presented in Figure 1.1.

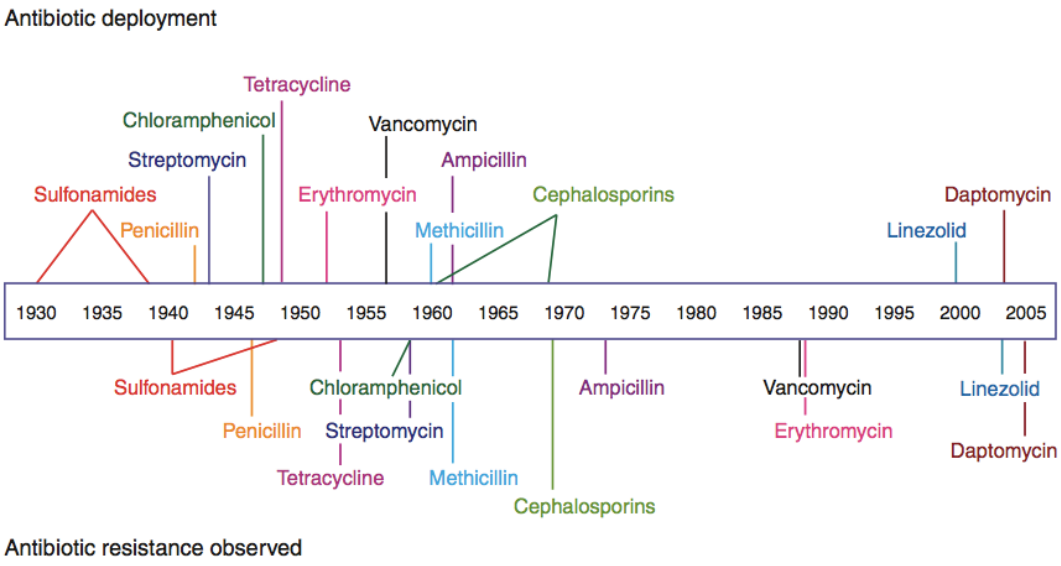



**Figure 1.1: Timeline showing the introduction of new antibiotics** and the emergence of resistance against them (Clatworthy et al., 2007). Used with permission from *Nature Chemical Biology*.

### 1.1.2 Challenges in antimicrobial NP discovery

One of the major issues that emerged soon after the 'Golden Age' previously described was the re-discovery of compounds. It was found that regardless of the location of soil sampling and the isolation techniques that followed, screening programs were mostly identifying known antibiotics (Baltz, 2006). This forced the introduction of dereplication techniques that, relying on mass spectrometry and nuclear magnetic resonance, would determine whether the identified compounds were truly novel (Lewis, 2013). Therefore, the risk of rediscovery and the need for dereplication imposed a time consuming and expensive bottleneck in natural product discovery that still to this day we try to overcome with the use of new tools and technologies (Gaudêncio and Pereira, 2015).

As a result of these obstacles, pharmaceutical companies switched their focus to synthetic chemistry for the creation of novel antibiotics. This proved expensive and largely unsuccessful, with few antibiotics making it to the clinic. One notable exception are the quinolones, which interfere with DNA replication by targeting DNA gyrase and are active against Gram-negative and Gram-positive bacteria (Heeb et al., 2011). Another approach for the development of antimicrobials is the use of semi-synthetic derivatives, as it has been routinely done with the β-lactams, for which we are currently between the fourth and fifth generation of derivatives (Page, 2012), or the antibiotics clarithromycin and telithromycin, semisynthetic derivatives of erythromycin generated in an effort to develop alternatives with improved pharmacokinetic properties (Watanabe et al., 1993, Scheinfeld, 2004). Even then, the cost and time required to get a new antibiotic to the market have had a negative impact on the discovery and development efforts of the big pharmaceutical companies over the last years. As a result, most of their natural product departments and screening programs have either been eliminated or significantly reduced, further slowing the discovery process (Martens and Demain, 2017).

In parallel with this deceleration in antibiotic discovery, bacteria continued to evolve under the selective pressure of the extended use and abuse of antibiotics in the clinic and in agriculture (Bérdy, 2012, Laxminarayan, 2014). This resulted to the continuous increase of resistance, with Gram-negative bacteria tending to be more resistant (Fernandes, 2006). Resistance to new antibiotics appeared a short time after their introduction into the clinic (Clatworthy et al., 2007), and the emergence of multidrug-resistant (MDR) and extensively-drug-resistant (XDR) bacteria is currently a serious threat for human health even in the developed world (Sherry and Howden, 2018, Martens and Demain, 2017, Laxminarayan, 2014). Groups of microorganisms such as the

ESKAPE pathogens (a group of Gram-positive and Gram-negative bacteria responsible for life-threatening nosocomial infections with widespread drug resistance mechanisms (Santajit and Indrawattana, 2016)) are part of the Global Priority List that the World Health Organisation (WHO) has compiled to help prioritize the research and development of new antibiotics against the most pressing bacterial threats (Organization, 2017).

The expense and difficulty associated with antimicrobial discovery led to a dearth in the identification of classes of compounds. Combined with the rapidly rising threat of AMR, this has created an urgent need for the discovery of novel NPs, with new modes of action to overcome current resistance mechanisms (Lewis, 2013, Santajit and Indrawattana, 2016).

## 1.2 Approaches for natural product discovery

### 1.2.1 Bioassay-guided discovery

Perhaps the most traditional method for searching for new antibiotics is bioassay-guided discovery. This is based on testing extracts from microbial cultures against bioindicator strains to reveal antimicrobial activity. Usually the organism of interest is cultured under a range of different media conditions, as compound production is often media dependent. If activity is observed, separation and purification techniques are subsequently utilised to identify which compound(s) is responsible for the activity. The major drawback of this approach is the problem of rediscovery, as discussed above. The chance of finding the same compounds from different strains or different locations of sampling is quite high and it leads to misdirected efforts (Baltz, 2008, Silver, 2011).

Despite its limitations, this approach is still widely applied today. Recent examples include: the discovery of amycomicin from *Amycolatopsis* sp. AA4, a highly modified fatty acid that inhibits *Staphylococcus aureus* (Pishchany et al., 2018); a *Streptomyces*-derived auxin with activity against phytopathogenic fungi and dermatophytes (Saravana Kumar et al., 2018), and six different NPs from the endophytic fungus *Emericella* sp. TJ29 (He et al., 2017).

### 1.2.2 Genome mining

Advances in genomic technologies over the last 20 years have had a great impact on the field of NP discovery (Bachmann et al., 2014). In 2001, the first genome sequence

of a *Streptomyces* strain was published: *Streptomyces coelicolor* A3(2). This revealed the presence of more than 20 biosynthetic gene clusters (BGCs) with no known products in laboratory conditions (Bentley et al., 2002). These clusters were termed 'silent' and they represented 75% of the potential NPs encoded by the strain. The subsequent sequencing projects of *Streptomyces avermitilis* (Ikeda et al., 2003) and of *Saccharopolyspora erythraea* NRRL23338 (Oliynyk et al., 2007) gave further insight into the untapped biosynthetic potential of even the most well-studied strains. In the years that followed, the speed and accuracy of genome sequencing has been rapidly accelerated, whereas the cost has decreased dramatically leading to genome sequencing becoming a routine laboratory procedure (Gomez-Escribano et al., 2016, Loman and Pallen, 2015). As a result, the NCBI genome (https://www.ncbi.nlm.nih.gov/genome/) database currently contains more than 160,000 bacterial genome sequences, 16,000 of which correspond to Actinobacteria.

The growing availability of genome sequences led to the development of tools with which to mine the data and identify biosynthetic gene clusters (BGCs). Tools such as antiSMASH (Blin et al., 2017), NaPDoS (Ziemert et al., 2012), GNP/PRISM (Skinnider et al., 2015) or BAGEL (van Heel et al., 2018), now allow for the rapid identification and annotation of BGCs (Monciardini et al., 2014). The revolution in genome mining enabled the acceleration of the NPs discovery pipeline, with the identification of many novel families of compounds (Bachmann et al., 2014). The many silent (also known as cryptic) BGCs identified encode for secondary metabolites that are often selectively expressed in response to very specific stimuli and under strict regulation, due to the high energy cost to the bacterial producer, and hence why they are not seen in laboratory conditions (Bibb, 2005). The activation of silent BGCs is one of the major goals for NP discovery, with expression in heterologous hosts being one of the main techniques employed (Rutledge and Challis, 2015, Gomez-Escribano et al., 2016).

### 1.2.3 Untargeted metabolomics

More recently, advances in genomic technologies have been matched with advances metabolomic data analysis. Reductions in the costs of processing many samples, along with advances in the sensitivity and accuracy of mass spectrometers has given rise to untargeted metabolomics approaches. This entails profiling the entire metabolome of bacteria or fungi species, commonly analysing a variety of different media and culture conditions. Crude extracts are subjected to ultra-performance liquid chromatography (UPLC) analysis coupled to spectrometry analyses (MS) (Baltz, 2017). This method

follows the OSMAC (One Strain–Many Compounds) approach (Bode et al., 2002) to unearth the great metabolic diversity often present in single strains (Senges et al., 2018).

With untargeted metabolomics, tandem MS/MS data are also analysed and metabolic networks can be constructed to provide a deeper understanding of the chemical diversity and structural relationships within the detected metabolomes (Chavali et al., 2012, Nothias et al., 2018). This allows for rapid identification of NPs, including pathway intermediates that might be dismissed in other analyses (Crone et al., 2016, Eyles et al., 2018, Watrous et al., 2012). All these data can now be deposited in molecular networks such as Global Natural Products Social (GNPS) based on the sharing of raw, processed or identified tandem mass (MS/MS) spectrometry data (Wang et al., 2016).

### 1.2.4 Exploration of rare environmental niches

As the majority of antimicrobial NPs have been isolated from soil-dwelling microorganisms, it has been proposed that, one approach to counter the rediscovery of the same antimicrobial compounds is to explore more unusual environmental niches, under the reasoning that biological diversity will lead to chemical diversity (Lyddiard et al., 2016, Challinor and Bode, 2015). These niches include marine sediments (Jang et al., 2013), deserts (Goodfellow et al., 2018), arctic regions (Wietz et al., 2012), and other extreme environments (Chen et al., 2014, Molloy and Hertweck, 2017, Simmons et al., 2008, Yu et al., 2010). One difficulty present both in soil and in any of these cases is that the vast majority of environmental bacteria cannot be cultured under laboratory conditions and therefore accessing their promising biosynthetic potential is more challenging (Rappé and Giovannoni, 2003, Lewis, 2013).

To overcome this problem, recent advances have been made in the field of culturing recalcitrant bacteria. One characteristic example is the work of Yang Bai and colleagues where they studied leaf and root microbiota of flowering plants (Bai et al., 2015). They combined three bacterial isolation procedures: colony picking from agar plates, limiting dilution in liquid media in 96-well microtiter plates and microbial cell sorting. As a result, they managed to characterize a total of 21 leaf-derived strains (phyllosphere bacteria) although these were undetectable in the culture-independent leaf community profiling.

Another example of ways to overcome the issue of unculturable bacteria is the Isolation chip (or ichip) method of culturing bacteria (Nichols et al., 2010). The ichip cultures bacterial species within their soil environment by diluting the soil in agar and nutrients. In that way a single cell grows in the ichip's small compartments or wells. The chip is wrapped with a semipermeable membrane and is then buried back in the soil to allow

interaction with nutrients not available in lab conditions. More than half of the original bacterial population can survive by this procedure. The most characteristic example was the discovery of the bacteria species *Eleftheria terrae*, producer of the antibiotic antibiotic teixobactin that targets drug-resistant strains like MRSA, in 2015 (Ling et al., 2015).

One potentially rich source of antimicrobial compounds comes from symbiotic niches. In these systems extended coevolution between the symbionts selects for the production of compounds that often aid in defence or confer other sorts of advantages that support the interaction between organisms  This sort of approach has yielded some interesting compounds coming from symbionts of marine organisms such as sponges or bryozoans, endophytic bacteria, or microorganisms that form part of insect or even human microbiomes (Akbar et al., 2018, Donia et al., 2014).

## 1.3 Actinomycetes as prolific producers of natural products

As it has been mentioned before, almost all antibiotics in clinical use today have been isolated from natural sources, mainly from microorganisms (Bérdy, 2012, Newman and Cragg, 2016). Among them, the more prolific antimicrobial producers in history are the actinomycetes, and it has been estimated that they are the source of around 60% of the known bioactive NPs, with around 80% of them being produced by the genus *Streptomyces* (Butler, 2008, Bérdy, 2012). Fungi produce approximately 30% the remaining NPs and the rest are made by other types of bacteria. Actinomycetes are Gram-positive bacteria with high G-C content and are very diverse morphologically, but many are characterised for having hyphal growth that develops in the form of substrate and aerial mycelium, and for undergoing a differentiation process that culminates in spore formation (Flärdh and Buttner, 2009, Chaudhary et al., 2013). These are environmental bacteria that inhabit mainly the soil, although screening studies in novel niches as mentioned before has revealed their presence in all sorts of environments, including the ocean (Zotchev, 2012) or extreme environments like the Atacama desert (Gomez-Escribano et al., 2015).

Similar to other microorganisms, symbiotic interactions are a promising source of antimicrobial producing actinomycetes. One of the best documented cases is that of leaf cutter ants, which farm fungi of the *Lepiotaceae* family that they use as food source and engage in a mutualistic relationship with different genus of actinomycetes, mainly *Streptomyces* and *Pseudonocardia*, which produce antifungal and antibacterial compounds that help protect the colony against pathogenic fungi and bacteria (Heine et al., 2018).

### 1.3.1 *Saccharopolyspora*, another talented NP producer genus

*Saccharopolyspora* is a genus of bacteria within the family *Pseudonocardiaceae*, in the class of Actinobacteria. The etymology of the genus comes from Saccharum, a generic name of sugar cane, and the Greek words ‘polus’ meaning many and ‘spora’ meaning seed, or spore. Hence, *Saccharopolyspora* stands for the many spored organism from sugar cane (Lacey and Goodfellow, 1975). It consists of aerobic, high GC content, Gram-positive species that produce aerial hypha with long chains of spores in spiny spirals (Zhou et al., 1998). *Saccharopolyspora species* have not been as widely studied as other actinobacteria and most of the known species have been isolated from diverse sources. Based on the taxonomy browser of NCBI, more than 35 *Saccharopolyspora* species have been reported in total.

This rare actinomycete genus is the source of many unique natural products, Figure 1.2. Most notably, the medically and agrochemically important antibiotic erythromycin (Oliynyk et al., 2007) and the environmentally benign insecticide spinosyn (Kirst, 2010), (Pan et al., 2011). Novel chemistry and biology is associated with other *Saccharopolyspora* species studied in any detail, for example, *Saccharopolyspora sp.* L53-18 produces the novel macrolide antibiotic sporeamicin (Yaginuma et al., 1992) and the thermophile *Saccharopolyspora rectivirgula* is a causative agent of hypersensitivity pneumonitis, an inflammatory lung disease (Farmers lung) (Pettersson et al., 2014).

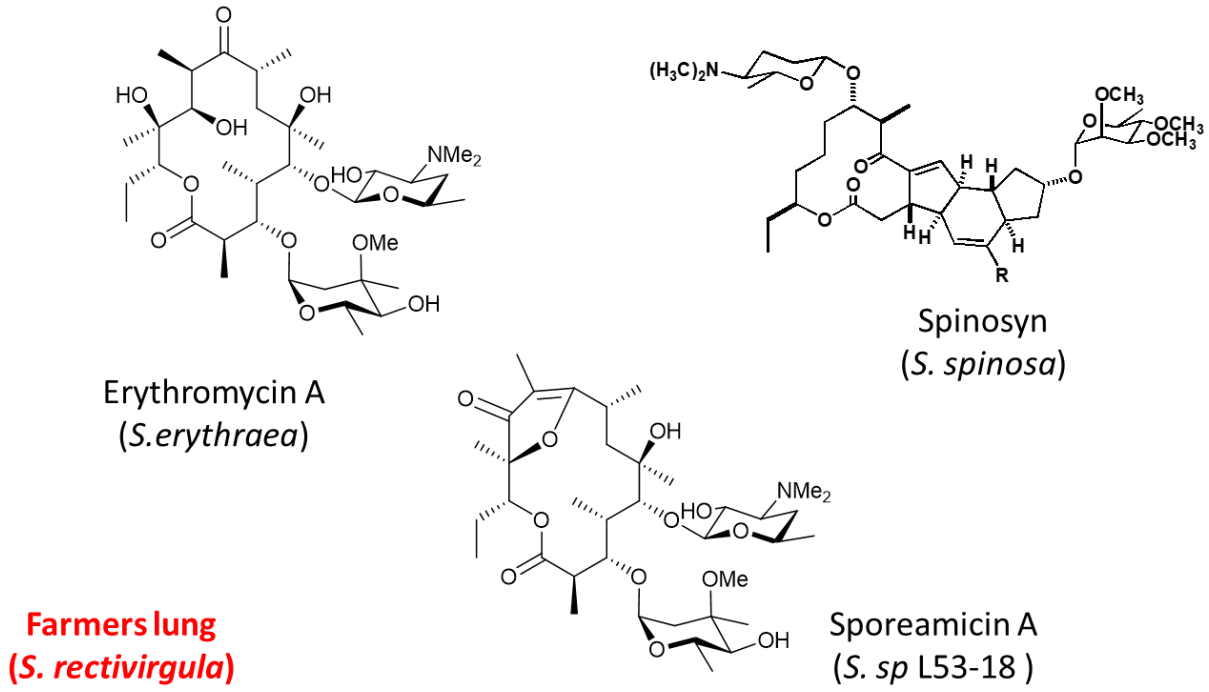



**Figure 1.2: Structures of NPs isolated from *Saccharopolyspora* sp.**

Analysis of the BGCs from published *Saccharopolyspora* species suggest they have the capacity to produce many more NPs. Below I outline the main classes of NPs, with a focus on those that will be discussed at some point in this thesis.

## 1.4 Types of natural products

### 1.4.1 Non-ribosomal peptides

Non-ribosomal peptides (NRPs) are a class of NPs that are synthesized by non-ribosomal peptide synthetases (NRPSs). NRPSs are large multidomain enzymatic complexes that can incorporate and process several hundreds of structurally different monomers, including non-proteinogenic amino acids, to give very structurally diverse cyclic or linear polypeptides (Marahiel, 2016, Payne et al., 2017, Fischbach and Walsh, 2006).

NRPSs have a modular architecture that allows biosynthesis to proceed in an assembly line fashion in which each module is in charge of incorporating a new residue to the assembled molecule. Therefore, each of these modules consists of a set of core catalytic domains that are required for catalysing the activation, incorporation of a specific amino acid into the growing peptide chain. The minimum set of domains required for one elongation cycle are Adenylation (A), Thiolation or Peptidyl Carrier Protein (T/PCP), and Condensation (C) domains (Figure 1.3).

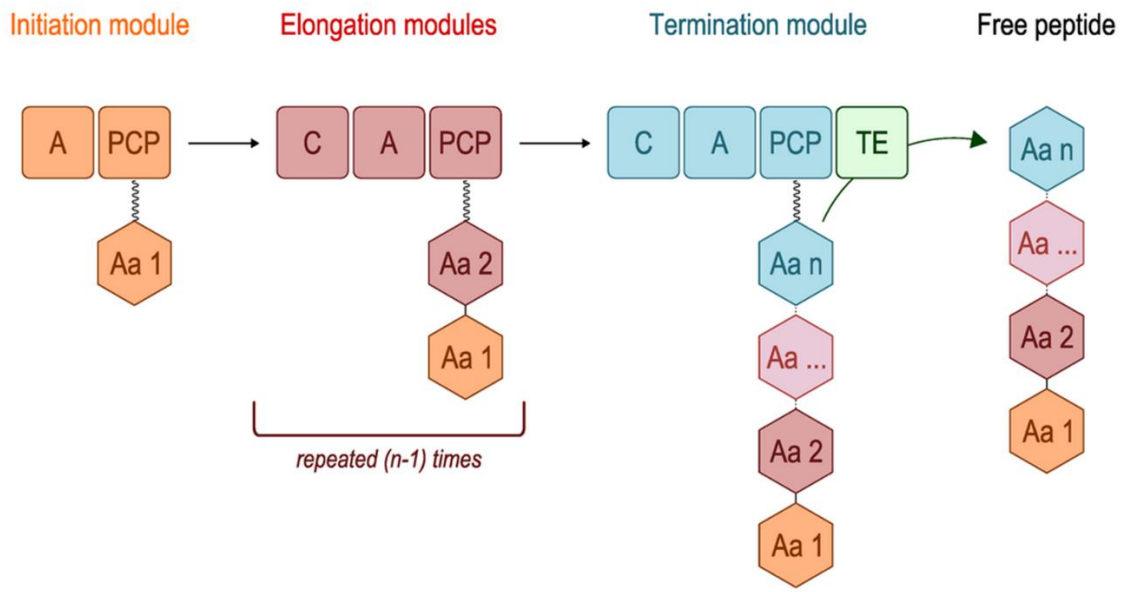



**Figure 1.3: Illustration of NRP biosynthesis.** Aa: amino acid; A: adenylation domain; C: condensation domain; PCP: peptidyl carrier protein; and TE: thioesterase domain (Desriac et al., 2013) Used with permission from *Marine drugs* journal.

The peptide assembly is initiated by a loading module, with A and PCP domains only. The A domain selects a specific amino acid substrate, activates it with ATP to give an amino-acyl adenylate and subsequently tethers it to a phosphopantetheine (PPant) prosthetic group on the adjacent PCP via a thioester bond. A domains typically consist of 550 aa and the PCP-domain of approximately 80 aa. After the loading module, a variable number of elongation modules are in charge of extending the molecule through the addition of the aminoacids activated by their A domains. In these modules, the C domain (approximately 450 aa) is responsible for the formation of peptide bonds between the carboxyl group of the forming peptide and the aa carried by the adjoining module. Then the growing peptide chain moves to the following module (Martínez-Núñez and López, 2016). Finally, A thioesterase (TE) domain usually present in the last module of the NRPS, is responsible for the release of the peptide from the backbone after all the required modifications (Fischbach and Walsh, 2006). Other domains that can be a part of an NRPS assembly line are: Cyclization (Cy) domains that cyclize peptides into thiazoline or oxazolines; Oxidation (Ox) domains that produce thiazoles or oxazoles; Reduction (Red) domains that give thiazolidines or oxazolidines; terminal Reduction (R) domains to give a terminal aldehyde alcohol; Epimerization (E) domains that produce D-amino acids and N-methylation (NMT) domains.

#### 1.4.1.1 Siderophores

An example of a type of compound that is commonly (although with some very notable exceptions (Carroll and Moore, 2018)) NRPS-derived is siderophores. Siderophores (from the Greek: "iron carrier") are molecules that strongly complex metals for absorption into microbes (Saxena et al., 1986). Notably, pyochelin is a siderophore of *Pseudomonas aeruginosa* that chelates copper and zinc as well as iron (Brandel et al., 2012). Siderophores are synthesized by bacteria and fungi when the availability of the iron is low in the environment, as this metal is essential for several cellular processes, such as respiration or DNA synthesis. Although iron is extremely abundant in geological terms, its bioavailability is seriously limited by the poor solubility of the $Fe^{3+}$ ion at physiological pH. Thus, in many environments its incorporation into the cells requires its solubilization and absorption in a chelated form (Saha et al., 2016). Based on the type of the iron ligand they contain, NRPS- based siderophores may contain one, two, or three chelating oxygen ligands such as hydroxamates (Yamanaka et al., 2005), catecholates (Harris et al., 1979), or α-hydroxycarboxylates (Figure 1.4) (Berti and Thomas, 2009), (Kadi and Challis, 2009).

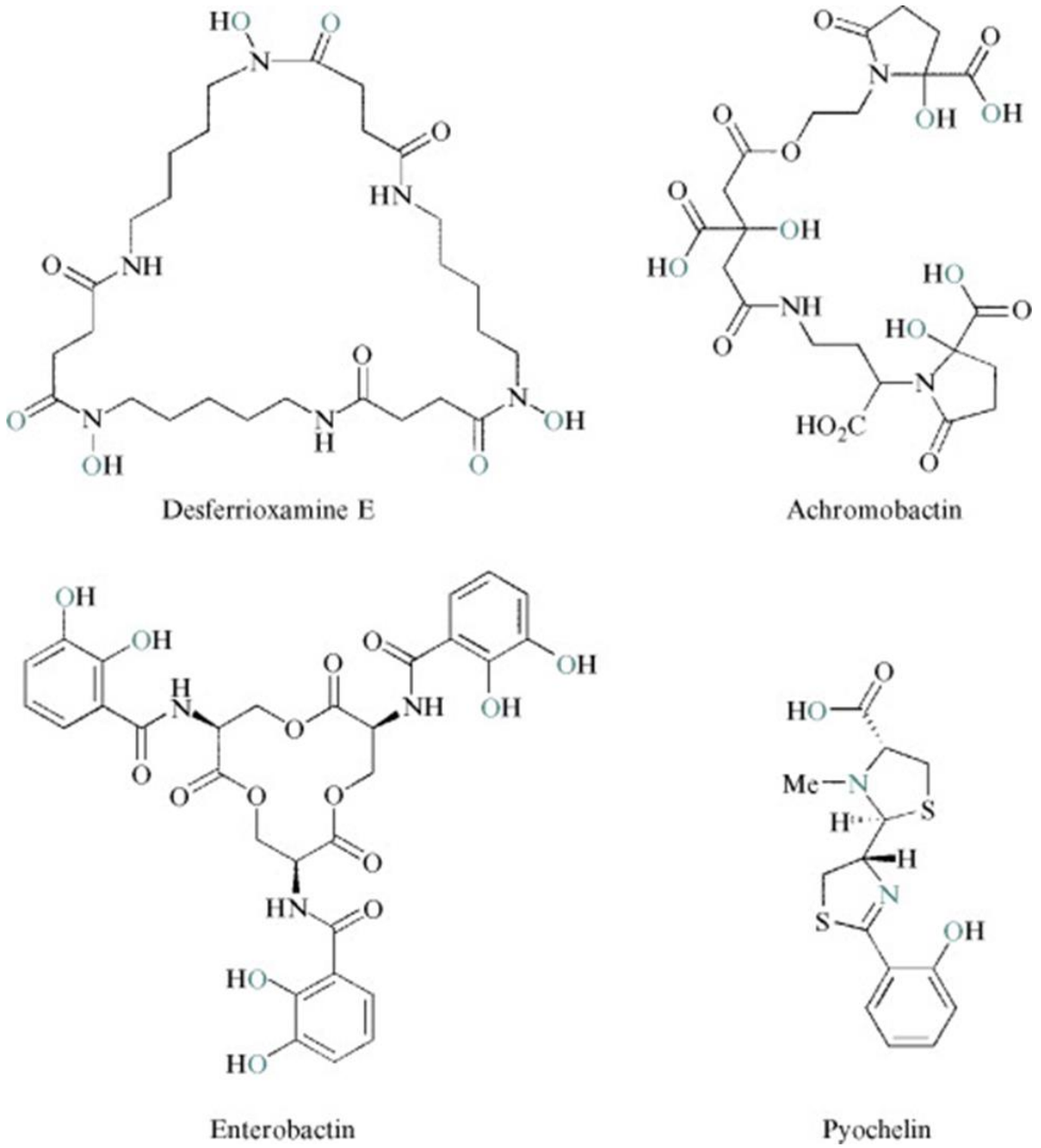



**Figure 1.4: Structures of siderophores containing different types of ferric-iron-chelating functional groups.** Hydroxamate in desferrioxamine E, α-hydroxycarboxylate in achromobactin, catecholate in enterobactin, and nitrogen heterocycle/ phenolate/ carboxylate in pyochelin. The atoms that bind to $Fe^{3+}$ in each structure are highlighted in gray. Note: desferroxamine is not synthesised by a NRPS. Used with permission from *Methods in Enzymology* journal (Kadi and Challis, 2009).

#### 1.4.1.2 Medically relevant NRPS-derived molecules

There are numerous NRPS-derived molecules that are medically important and are being used successfully to treat numerous infections (Felnagle et al., 2008). Characteristically, β-lactams are a large class of antibiotics that come from linear NRPSs. They include the penicillins and the cephalosporins that in the beginning of the 20$^{th}$ century were one of the main classes of antibiotics prescribed around the world (Elander, 2003). Another characteristic class is the glycopeptides such as vancomycin, an NRP used widely to treat life-threatening infections for patients that have an allergy to β-lactams. Vancomycin is used against some MRSA infections and other severe infections from *Corynebacterium* and *Streptococcus* (Nicas and Cooper, 1997). The other clinically relevant glycopeptide, teicoplanin, is a semisynthetic antibiotic with a spectrum of activity similar to vancomycin and is used in Europe in cases of vancomycin-resistant enterococci (Kahne et al., 2005). Finally, lipopeptides, with daptomycin as the main representative, are an important member of the NRPs and constitute one of the newest classes of antibiotics that was introduced the last 40 years(Huber et al., 1988). Daptomycin is used to treat mainly skin infections caused by resistant pathogens including MRSA and VRE bacteria (Kirkpatrick et al., 2003). It is worth mentioning the cyclosporins (Dreyfuss et al., 1976), a product of linear NRPSs from fungal species that have anti-inflammatory and immunosuppressant activity and are being used to prevent rejection of organ transplants (Morris, 1984) and for autoimmune diseases (Borel and Hiestand, 1999).

### 1.4.2 Polyketides

Polyketides are a very diverse class of molecules generated by the successive condensation and modification of short carboxylic acid chains that are activated in the form of acyl-CoA extender units. There are several classes of enzymes responsible for their biosynthesis, all known as polyketide synthases (PKSs), that can be divided in three main groups according to their organisation. Type I PKSs are large, multi-domain enzymes with a similar architecture to that of NRPSs in which each module is responsible for the addition of a new extender unit (Weissman, 2015). Type II PKSs are composed by several monodomain proteins which act iteratively through several rounds of chain extension, in order to generate the final molecule (Hertweck et al., 2007). Type III PKS on the other hand are homodimers of ketosynthases which catalyze the condensation of one or several molecules of extender units in an iterative fashion (Katsuyama and Ohnishi, 2012).

The modules of type I PKS and the minimal module of type II PKSs catalyse the addition and a-/b-carbon modification of an extender unit to the growing PK chain (Minowa et al., 2007). PKSs catalyse decarboxylative Claisen-like thioester condensation reactions to connect the extender units together (Staunton and Weissman, 2001). The minimum set of domains required for chain elongation are an acyltransferase (AT), an acyl carrier protein (ACP) and a ketosynthase (KS) domain. ATs select an extender malonyl-CoA unit that gets transferred onto the PPant group of the ACP domain, via a covalent thioester bond. Similar to the NRPS biosynthesis, type I PKSs start with a loading module consisting of an AT and ACP domain that incorporate a starter unit. The following modules contain a KS domain that receives the growing PK chain and catalyses its condensation with the acyl extender group loaded by the AT domain and attached to the ACP domain in that same module (Weissman, 2015). Further modifications at the PK chain can be catalysed by accessory domains such as ketoreductase (KR), dehydratase (DH), enoyl reductase (ER) and C-methyltransferase (MT). Typically, a TE releases the PK chain from the PKS assembly line. A hypothetical PKS assembly line is presented in Figure 1.5.

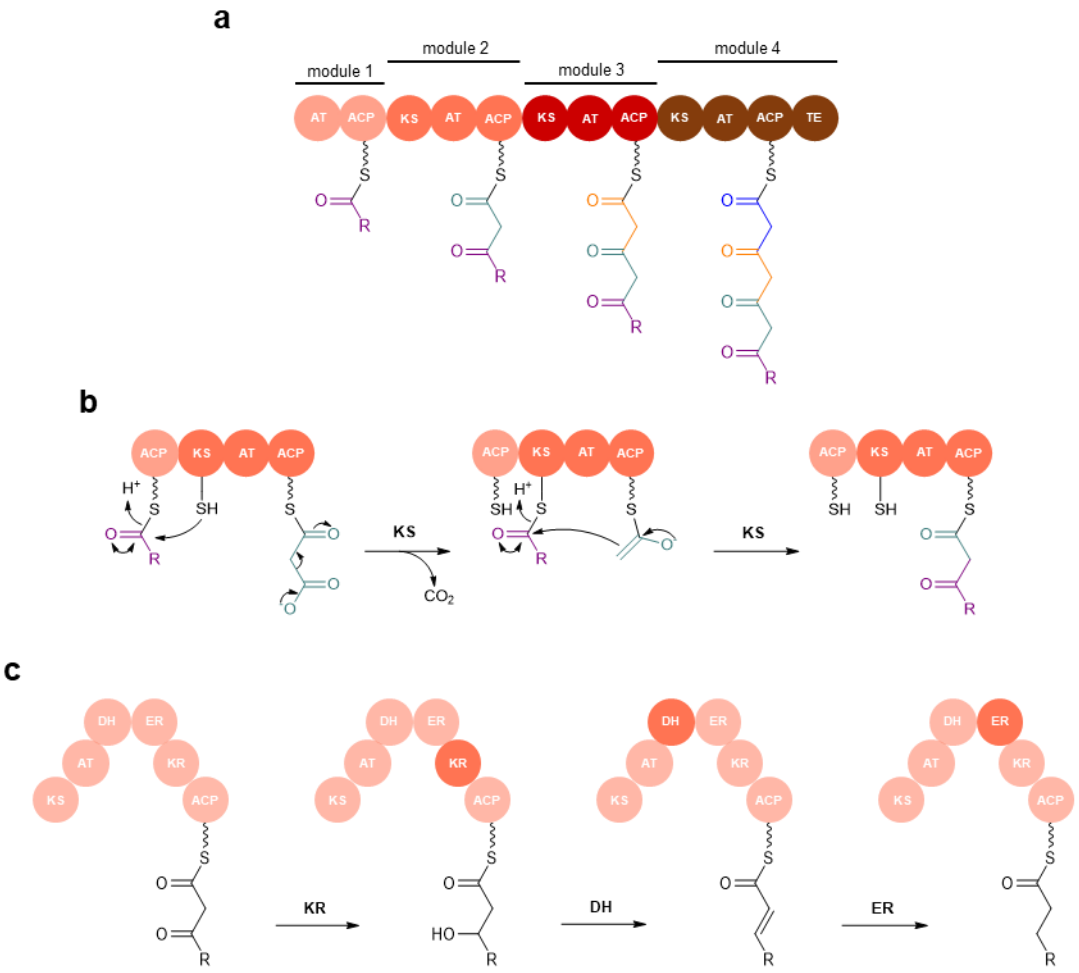



**Figure 1.5: Illustration of PKS assembly line biosynthesis.** (**a**) PK modular chain elongation on a hypothetical assembly line (**b**) The Claisen condensation catalysed during PKS biosynthesis. A C–C bond formation occurs between two esters or an ester and another carbonyl compound in the presence of a strong base, resulting in a β-keto ester or a β-diketone. Decarboxylation of the downstream (methyl)malonyl-*S*-ACP yields a nucleophilic thioester enolate, which attacks the upstream acyl-*S*-ACP thioester to form a C-C bond. (**c**) β-Carbon processing. The product of the α-ketoacyl-S-T condensation is reduced by the KR domain to form a β-hydroxyacyl-S-T and is dehydrated by the DH domain. This reaction yields an β-enoyl-S-T and reduction of this species by the ER domain results in a fully saturated acyl-S-T. Adapted from Scott (2017) and reproduced here with permission.

### 1.4.2.1 Erythromycin

The type I PKS macrolide antibiotic erythromycin was first isolated from the actinomycete *Saccharopolyspora erythraea* (formerly *Streptomyces erythreus*) and was used in the clinic as early as 1952 (Haight and Finland, 1952). Erythromycin A is a member of macrolide family of compounds, which are characterized by big macrocyclic lactone rings that can be 14, 15, or 16 membered (Zhang et al., 2008). It is used in clinical medicine against infections caused by gram-positive bacteria and also for many pulmonary infections, such as Legionnaire's disease and as an alternative for patients allergic to penicillins (Staunton and Wilkinson, 1997). Erythromycin also acts as agonist of the motilin receptor and a semi-synthetic derivative is under development as a potential treatment for gastric motility disorder (Butler, 2008).

As illustrated in Figure 1.6, erythromycin is derived from a PKS named 6-deoxyerythronolide B synthetase (DEBS). This PKS is split in three different proteins, DEBS 1, DEBS 2, and DEBS 3 (Caffrey et al., 1992), encoded by genes designated as. *eryAI, eryAII, and eryAIII* respectively. Each DEBS protein contains two modules and each module has domains that encode for an AT, ACP, and KS. Module four contains a DH, ER and KR domain. The TE domain in module six releases the aglycon product of DEBS 3, which is cyclized and gives 6-deoxyerythronolide B (DEB). Additional tailoring steps result in the production of erythromycin A.

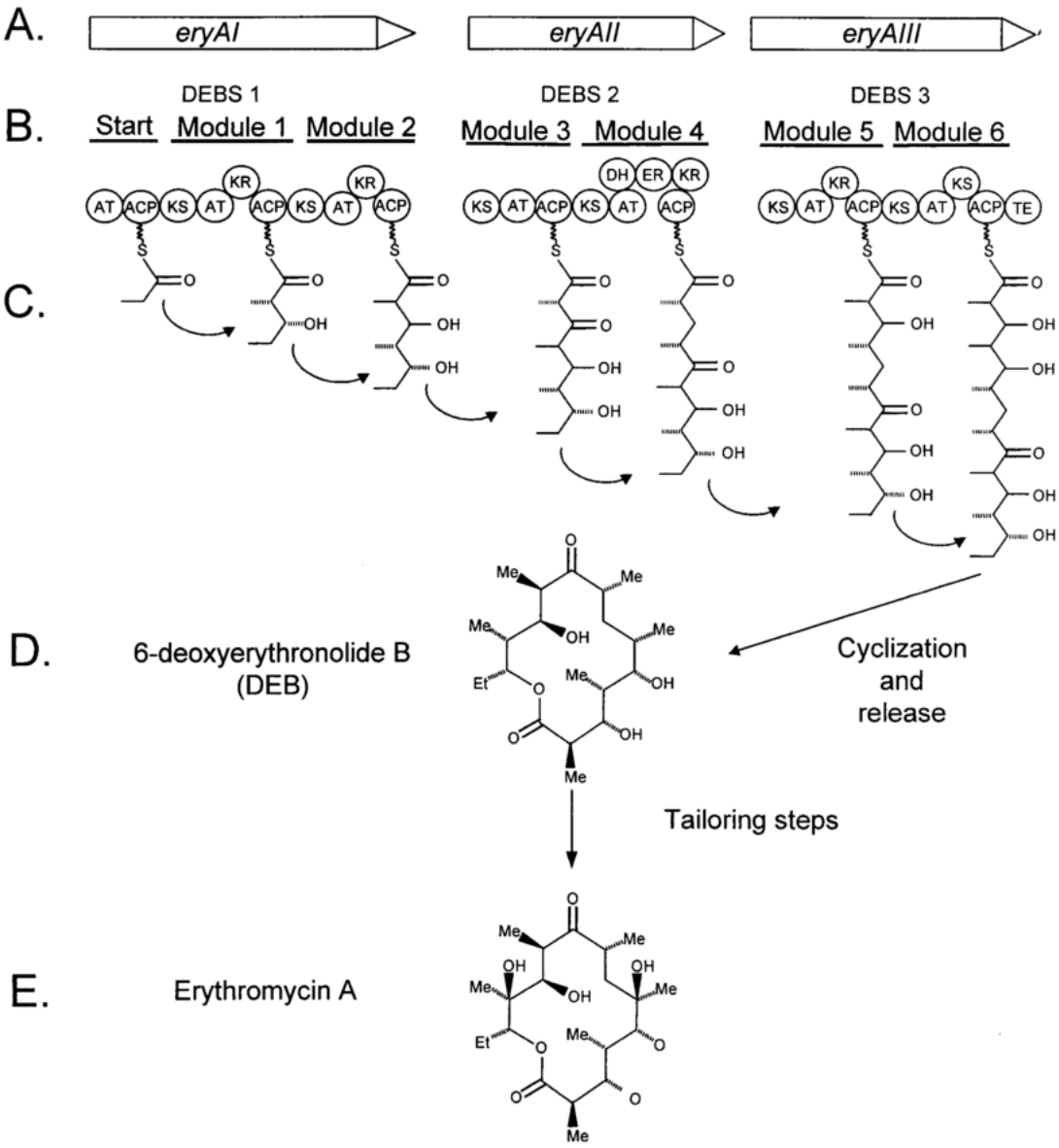



**Figure 1.6: Biosynthesis of erythromycin A.** Three ORFs designated for the biosynthesis of erythromycin: *eryAI*, *eryAII*, and *eryAIII*. B. 6-deoxyerythronolide B synthetase (DEBS) consists of three proteins DEBS 1, DEBS 2, and DEBS 3, encoded by A. C. Modules of each DEBS protein. D. Cyclization and release step leads to 6-deoxyerythronolide B (DEB). E. Additional tailoring steps produce Erythromycin. Reproduced with permission from *Microbiology and molecular biology reviews* (Bender et al., 1999).

### 1.4.3 Ribosomally synthesised and post-translationally modified peptides

Ribosomally synthesised and post-translationally modified peptides (RIPPs) are NP originated from peptides that are encoded by small structural genes. Despite their proteinogenic nature, these compounds exhibit a great structural diversity thanks to the extensive and varied modifications they undergo (Skinnider et al., 2016, Hudson and Mitchell, 2018). The full-length product of the RiPP structural genes is usually denominated precursor peptide and designated as “A”, while the gene encoding it is designated as *lanA* (Arnison et al., 2013). This precursor peptide has two parts, the core peptide, which is the part of the peptide that constitutes the final product and is the substrate for the posttranslational modifications, and the leader peptide, that usually precedes the core. The leader peptide does not form part of the final molecule and it is cleaved at some point during RiPP maturation (Figure 1.7). This leader peptide often has recognition domains where the tailoring enzymes bind to modify the core peptide. In most cases it is not clear whether the cleavage happens right at the end of the pathway or earlier in the process and is a characteristic that varies between RiPPs (Arnison et al., 2013).

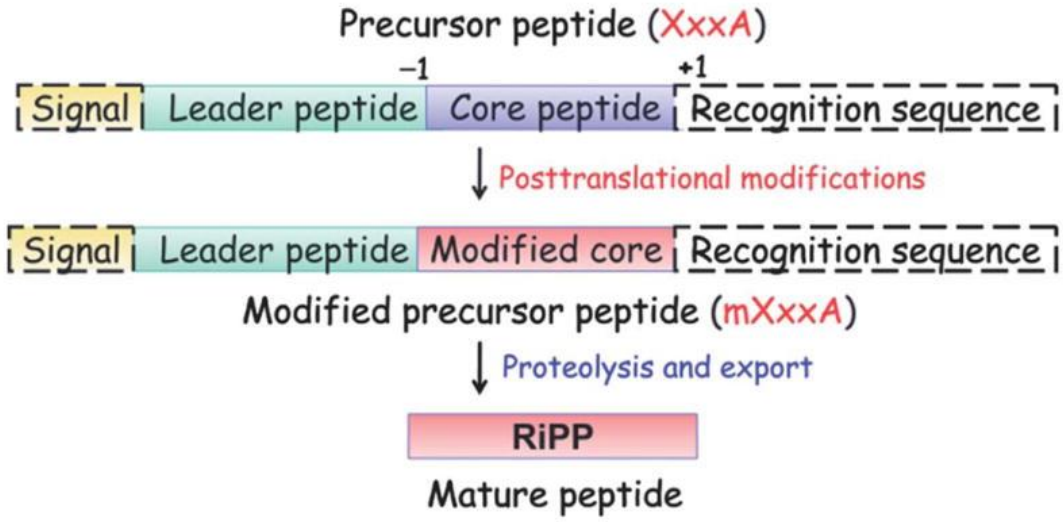



**Figure 1.7: General biosynthetic pathway for RiPPs** (Arnison et al., 2013). Reproduced with permission of the Royal Society of Chemistry.

### 1.4.3.1 Lantibiotics

Lantibiotics are a group of RiPPs characterized by the presence of lanthionine type thioester bonds which crosslink specific amino acid side-chains of the core peptide. Lantibiotics are produced by many Gram-positive bacteria such as lactic acid bacteria, *Bacillus*, *Enterococcus* and actinomycetes, among others (Cooper et al., 2010), (Figure 1.8), and their biosynthesis requires extensive posttranslational modifications (Nagao, 2009). Based on the structure and the antimicrobial properties, lantibiotics can be subdivided into type A & B.

Type-A lantibiotics include molecules such as nisin (Shin et al., 2016), subtilin (Banerjee and Hansen, 1988) and gallidermin (Kellner et al., 1988). They are elongated peptides up to 34 residues in length and their lanthionine bridges are arranged in a similar way (Figure 1.8). Type-B lantibiotics include related molecules, such as cinnamycin, duramycin & ancovenin and they are globular peptides, up to 19 residues long. They disrupt enzymatic function, such as inhibition of the cell wall biosynthesis (McAuliffe et al., 2001). Their mode of action is based on the disruption of the membrane integrity of their target (Chatterjee et al., 2005).

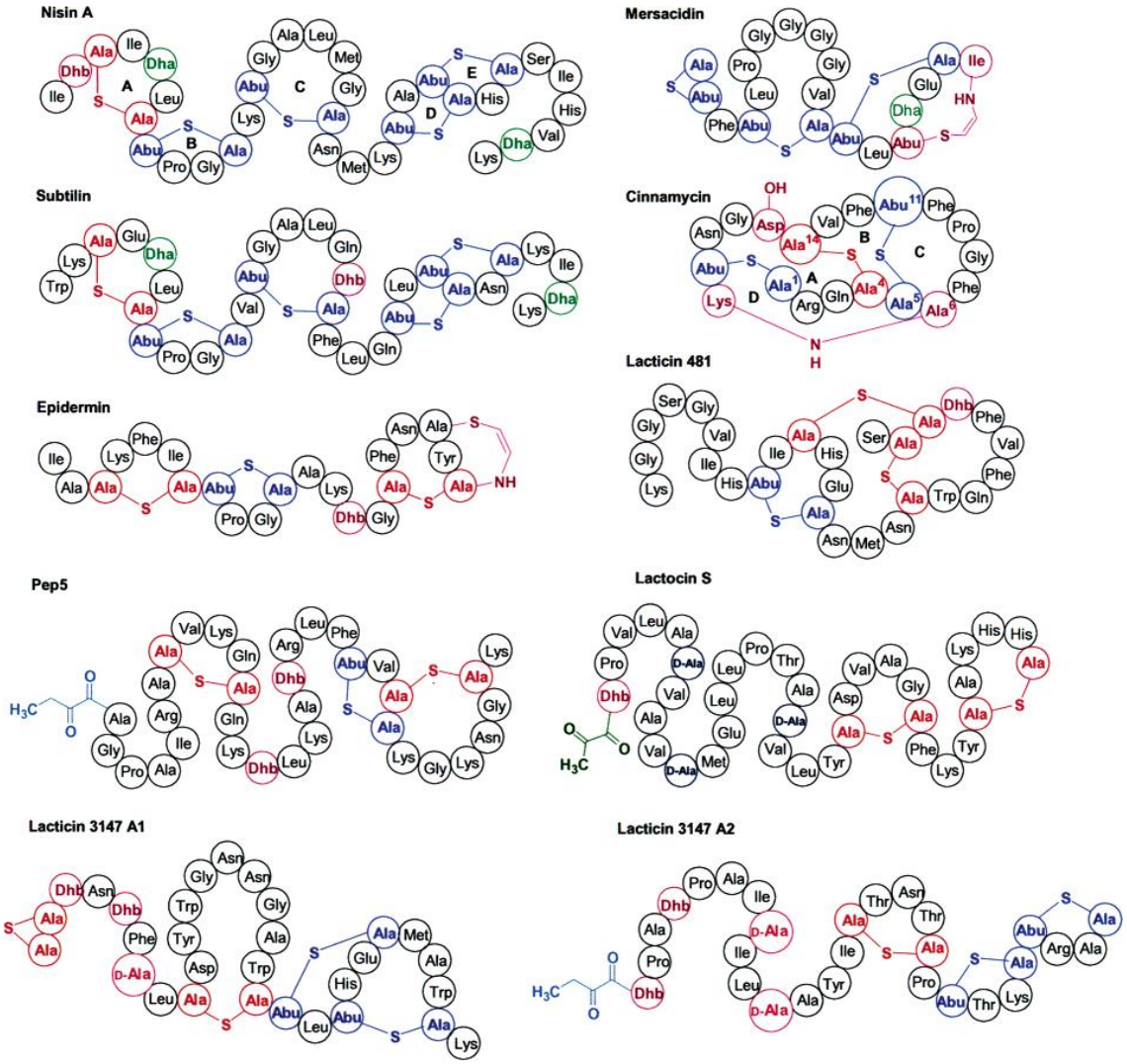



**Figure 1.8: Structures of known lantibiotics.** Reproduced with permission from *Chemical Reviews* (Chatterjee et al., 2005).

The first type B lantibiotic to be characterised was cinnamycin from *Streptomyces cinnamoneus* DSM 40646 (Figure 1.9). It is active against a broad range of gram-positive bacteria, primarily against *Bacillus subtilis*. Cinnamycin acts by binding to the membrane amino phospholipid, phosphatidylethanolamine (PE). This induces transbilayer lipid movement (TLM) which compromises the structural integrity of the membrane leading to cell lysis (Hullin-Matsuda et al., 2016). This mechanism is unique to the cinnamycin family of lantibiotics (Widdick et al., 2003).

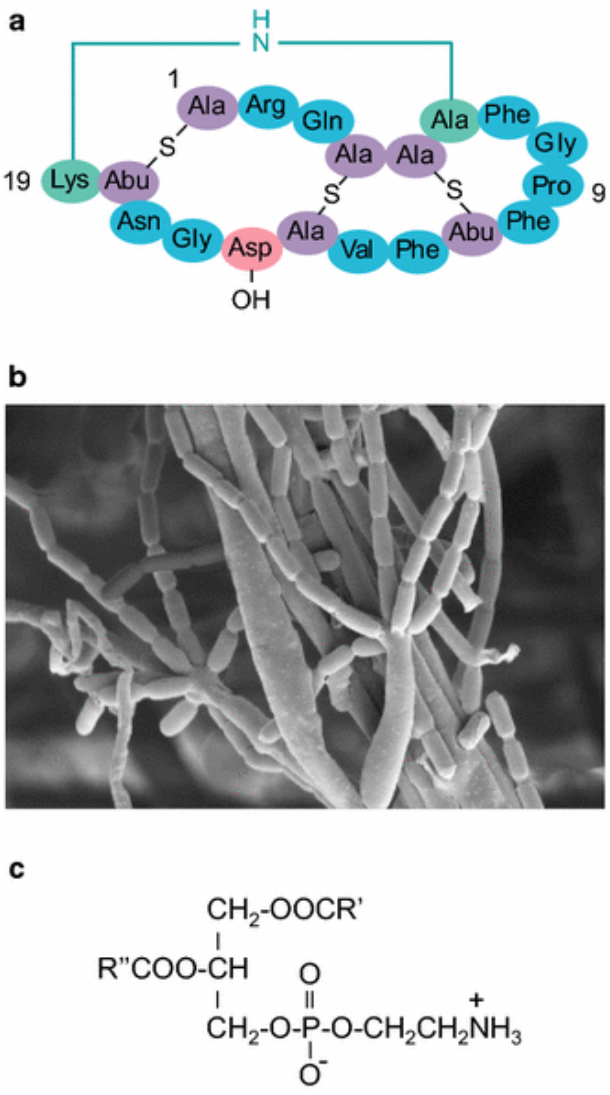



**Figure 1.9: Profile of Cinnamycin** A. Cinnamycin structure. B. Scanning electron micrograph of *S. cinnamoneus*. C. Structure of PE. Reproduced with permission from *Journal of industrial microbiology & biotechnology* (O'Rourke et al., 2017).

Duramycin is closely related to cinnamycin, and they differ by a single amino acid in position 2: arginine for cinnamycin and lysine for duramycin. Duramycin is a NP produced by various streptomycetes such as *Streptomyces cinnamoneus* ATCC 12686. Duramycin was recently reported to effectively block TIM1, a cofactor associated with the ZIKV (Zika virus) infection in primary placental cells and chorionic villus explants. This is of great interest since it suggests a possible medical application to blocking the virus to infect the foetus through the mother (Tabata et al., 2016). Additionally, duramycin is in phase II clinical trials for the treatment of cystic fibrosis (Oliynyk et al., 2010).

#### 1.4.3.1.1 Biosynthesis

Lantibiotics undergo post translational modifications of specific amino acids inside the cell, get transported across the cytoplasmic membrane and there, the leader peptides are cleaved to release the mature peptide that has a specific activity (McAuliffe et al., 2001). Lanthionine bridges are composed of two alanine residues that are crosslinked on their β-carbon atoms by a thioether link (Figure 1.10). Thioether links are formed first by a dehydration of T4, T11, T18 and S6 by LanM to form dehydrobutyrine (Dhb) and dehydroalanine (Dha) residues, respectively. After thioether cyclization by LanM, Dhb becomes S-linked Abu and Dha becomes S-linked Ala. D15 is hydroxylated by LanX and the lysinoalanine bridge is formed between Dha6 and K19 catalysed by LanN. After the core peptide is fully modified, the leader peptide is cleaved proteolytically to yield the mature lantibiotic.

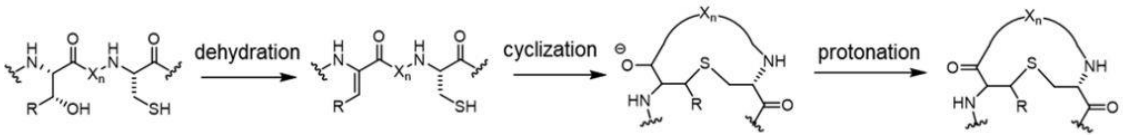



**Figure 1.10: General biosynthetic scheme of thioether linkage in lanthipeptides**. R = H, Me; Xn = amino acids. Reproduced with permission from *Appl Environ Microbiol* journal (Chen et al., 2017).

#### 1.4.3.1.2 Regulation

The genes involved in the regulation of lantibiotics are presented in Table 1.1, proposed by cinnamycin and duramycin regulation. In detail, *lanA* encodes for the precursor peptide, *lanM* is the lanthionine synthetase that introduces the lanthionine and methyl-lanthionine bridges, *lanN* encodes the formation of lysinoalanine bridge, *lanX* is a hydroxylase that modifies D15, *lanTH* encodes an ABC transporter involved in the export

of the lantibiotic, *lanKR* encodes a two-component regulatory system of unknown function, *lanL* encodes immunity and *lanR1* the activator of the putative operon encoding the lantibiotic biosynthetic machinery (O'Rourke et al., 2017, An et al., 2018).

**Table 1.1: Proteins encoded by lantibiotics BGCs.**

| Lantiobiotic | Proposed Function |
|---|---|
| LanA | Precursor peptide |
| LanN | Formation of lysinoalanine bridge |
| LanM | Formation of lanthionine residues |
| LanX | Hydroxylation of D15 |
| LanT | Export |
| LanH | Export |
| LanR | Regulation |
| LanK | Regulation |
| LanL | Immunity |
| LanR1 | Regulation |

## 1.5 Project Objectives

This project will explore the diversity and biological activity of new NPs produced by underexploited *Saccharopolyspora* species, including new strains isolated from fungus-growing ant colonies.

Firstly, I investigate three *Saccharopolyspora sp.* strains named KY, from *Tetraponera penzigi* fungus farming ants. All three strains, KY3, KY7 and KY21 were isolated from bacteria from the ants located in the domatia, hollow swellings at base of thorns of acacia trees in Kenya. By using a tripartite approach that combines genomic, untargeted metabolomic and bioassay data, the diversity of the NP capacity of the strains is presented.

Also, I investigate a known *Saccharopolyspora* strain, the producer of sporeamicin, a 14-membered erythromycin-like antibiotic. The aim of this part of the project is to identify new enzymes and mechanisms that lead to the biosynthesis of sporeamicin.

Overall this project aims to:

1. Use genome mining approaches to identify putative silent BGCs of interest in KY strains and express them.
2. Explore the potential of engineering a platform to express type B lantibiotics.
3. Identify novel compounds by untargeted metabolomics and bioassay data and match them to a genetic context.
4. Investigate the biosynthesis of sporeamicin and identify novel enzymes.

# Chapter 2:

## Materials and methods

# 2 Materials and Methods

## 2.1 Lab supplies

All reagents were obtained from commercial suppliers (Sigma Aldrich and Alfa Aesar, BD Biosciences) and were used without further purification. Solvents were obtained from Fisher Scientific in HPLC grade. A list of commercially available enzymes, kits, reagents and size markers for DNA used are in shown in **Table 2.1**.

**Table 2.1: Lab supplies used in this work**

| **Suppliers** | **Enzymes and kits** |
|---|---|
| GE Healthcare Life Sciences | Whatman Grade AA discs |
| New England Biolabs | 1 kb and 2-log DNA ladders<br>Gibson Assembly® Master Mix<br>Q5® High-Fidelity DNA polymerase<br>T4 DNA ligase |
| MP Biomedicals | FastDNA™ SPIN KIT |
| Promega, Southampton | GoTaq® G2 Green Master Mix |
| Roche | Deoxynucleoside Triphosphate (dNTPs) Set PCR Grade<br>rAPid Alkaline Phosphatase |
| Qiagen | QIAquick® Gel Extraction Kit<br>QIAquick® PCR Purification Kit<br>QIAprep® Spin Miniprep Kit |

## 2.2 Bacterial strains, growth conditions, plasmids and primers

### 2.2.1 Bacterial strains and growth conditions

All *Saccharopolyspora* strains were grown on SF+M medium with appropriate antibiotics at 30°C. All *E. coli* and bioindicator strains were grown at LB medium with appropriate antibiotics at 37°C, unless stated differently. The strains are listed in Table 2.2. The bioindicator strains used in bioassays in this study are listed in Table 2.3.

**Table 2.2: Strains used in this work**

| Strain | Description | Reference |
|---|---|---|
| *E. coli* ATCC 25922 | Bioassay strain; WT | ATCC, USA |
| *E. coli* DH5α | Cloning strain; $F^-$, *endA1*, *glnV44*, *thi-1*, *recA1*, *relA1*, *gyrA96*, *deoR*, *nupG*, *ϕ80dlac*Δ*(lacZ)M15*, Δ(*lacIZYA-argF)U169*, *hsdR17*($r_K^-$ $m_K^+$), λ– | Hanahan, 1983 |
| *E. coli* ET12567/pUZ8002 | E. coli ET12567 containing plasmid pUZ8002, a not self-transmissible plasmid which can mobilize other plasmids | Mc Neil et *al*, 1992; Flett, 1997 |
| *Saccharopolyspora* sp. KY3 | Strain from the cuticles of *Tetraponera penzigi* | This work |
| *Saccharopolyspora* sp. KY7 | Strain from the cuticles of *Tetraponera penzigi* | This work |
| *Saccharopolyspora* sp. KY21 | Strain from the cuticles of *Tetraponera penzigi* | This work |
| *Saccharopolyspora* sp. L53-18/ FERM BP-2231 | Producer of sporeamicin | Yaginuma, S. *et al.*, 1992 |
| *Streptomyces cinnamoneus* DSM 40005 | Producer of cinnamycin | DSMZ, Germany |
| *Streptomyces coelicolor* M1152 | antibiotic producing superhost [ΔactΔred Δcpk Δcda rpoB(C1298T)] | Escribano et *al*, 2011 |
| *Streptomyces coelicolor* M1152/pWDW63 | M1152 carrying the biosynthetic genes for kyamicin | This work |
| *Streptomyces coelicolor* M1152/pEVK6 | M1152 carrying pIJ10257/R1L | This work |
| *Streptomyces coelicolor* M1152/pEVK7 | M1152 carrying pIJ10257/LR1 | This work |
| *Streptomyces coelicolor* M1152/pWDW63/pEVK6 | M1152 carrying the biosynthetic genes for kyamicin plus pIJ10257/R1L | This work |
| *Streptomyces coelicolor* M1152/pWDW63/pEVK7 | M1152 carrying the biosynthetic genes for kyamicin plus pIJ10257/LR1 | This work |
| *Streptomyces coelicolor* M1152/pEVK6/pWDW70 | M1152 carrying pIJ10257/R1L & pWDW70 | This work |
| *Streptomyces coelicolor* M1152/pEVK6/pEVK10 | M1152 carrying pIJ10257/R1L & pWDW70 carrying the synthetic kyamicin core peptide | This work |

**Table 2.3: Bioassay strains used in this study**

| Bioassay strains | Description/Genotype | Reference |
|---|---|---|
| *Enterococcus faecium* 6295 (VRE) | Clinical Isolate from Norfolk and Norwich University Hospital. Professor Andrew Hart, Dr Catherine Tremlett and Ashleigh Crane. | NNUH |
| *Bacillus subtilis* ESKAPE | ESKAPE strain | Handelsman lab |
| *Staphylococcus aureus* (MRSA) | Clinical Isolate from Dr Justin O'Grady. | UEA Medical School |
| *Staphylococcus epidermidis* | ESKAPE strain, 14990$^{TM}$ | ATCC, USA |
| *Bacillus subtilis* EC1524 | Bioassay strain; trpC2, Subtilin BGC deleted | Widdick et *al.*, 2003 |
| *Streptomyces antibioticus,* TU1798 | Gram-positive bioassay strain | Tübingen |
| *Candida albicans* CA-6 | opportunistic pathogenic yeast | Maconi et *al.,*1976 |

### 2.2.2 Vectors

Plasmids used and constructed in this study are listed in Table 2.4.

**Table 2.4: Plasmids used in this work**

| Plasmid | Genotype/description | Reference |
|---|---|---|
| pGP9 | pSET152-derived φBT-based integrative expression vector | Gregory et al., 2003 |
| pSET152 | ϕC31 attP-conjugative vector | Gregory et al., 2003 |
| pOSV556t | integrative plasmid, ermE* promoter | Sydor et *al., 2012* |
| pIJ10257 | oriT, φBT1 attB-int, Hygr, ermEp* | Hong et *al.,* 2005 |
| pKC1132 | conjugative suicide vector | Bierman et al., 1992 |
| pEVK1 | pUC57/pR1L 1 | GenScript™ |
| pEVK2 | pUC57/pLR1 1 | GenScript™ |
| pEVK3 | pUC57/KY3 | GenScript™ |
| pEVK4 | pGP9/R1L | This work |
| pEVK5 | pGP9/LR1 | This work |
| pEVK6 | pIJ10257/R1L | This work |
| pEVK7 | pIJ10257/LR1 | This work |
| pEVK10 | pWDW70/KYA 4 | This work |
| pADW11 | pOSV55t based vector with a pSAM2 integrase and and an *ermE** promoter. It carries a thiostrepton resistance gene for selection in *Streptomyces* and a carbenicillin resistance gene for *E. coli* | This work |
| pADW19 | Same as pADW11 but with additional *ermE**p::*kyaR1* and the *kyaN* promoter | This work |

| | | |
|---|---|---|
| pWDW60 | pUC57/KY3 based lantibiotic biosynthetic cluster subcloned EcoRI/XbaI into pBlueScriptIIKS cut the same. This was made because I think there is a concatenation of at least two copies of the cluster in the GenScript construct that has been causing trouble with Redirect. | This work |
| pWDW63 | pSET152/KY3. KY3 stands for set of biosynthetic genes of kyamicin BGC | This work |
| pWDW55 | pUC57/KY3 lantibiotic biosynthetic cluster transformed into BW25113pIJ790 and targeted with a cinA replacement cassette generated from pIJ10700 (hyg) using the following primers | This work |
| pWDW59 | pWDW55, carrying the KY3 lantibiotic with *cin*A replaced with a cassette generated from pIJ10700 (hyg), transformed into DH5αBT340 and grown at 42C so that the cassette is flipped out and replaced by a SCAR. | This work |
| pWDW61 | kyaA generated by PCR using the GenScript plasmid as template and Q5 polymerase | This work |
| pWDW65 | The EcoR1/KpnI fragment of pWDW59 cloned into the vector fragment of EcoRI/KpnI cut pWDW64 to reconstitute the biosynthetic gene cluster but with kyaA replaced with a scar. | This work |
| pWDW66 | pADW19 with *kyaA* and thiostrepton resistance genes, under the control of *kyaN* promoter | This work |
| pWDW68 | A PCR covering the leader peptide and upstream region of the GenScript Biosynthetic cluster in pWDW60 cut with EcoRI/KpnI and cloned into pBSIIKS | This work |
| pWDW69 | A PCR covering the downstream region from the end of the leader peptide of the GenScript Biosynthetic cluster in pWDW60 cut with StuI/KpnI and cloned into pWDW68 also cut StuI/KpnI. | This work |
| pWDW70 | The EcoRI/KpnI fragment from pWDW69 carrying the 5' end of the kyamicin cluster was cloned into the vector portion of a EcoRI/KpnI digested pWDW63. This restores most of the cluster but the *kyaA* propeptide region has been replaced with a StuI site. | This work |

### 2.2.3 Primers

The primers used in this study are presented in Appendix 1.

## 2.3 Culture and production media

### 2.3.1 Culture media

**Instant Mash Agar (IMA)** for sporulation of Actinomycetes

| | |
|---|---|
| SMASH | 20.0 g |
| Agar | 20.0 g |
| Tap water to 1 L | |

**SNA**

| | |
|---|---|
| Difco Nutrient Broth | 8.0 g |
| Formedium | 7.0 g |
| Distilled water to 1 L | |

**SF+M** for growth and conjugation of actinomycetes

| | |
|---|---|
| Soya flour | 20.0 g |
| Mannitol | 20.0 g |
| Lab M $N_o$ 1 agar | 20.0 g |
| Tap water to 1 L | |

**SV2** for seed cultures and for production of mycelia

| | |
|---|---|
| Glucose | 15.0 g |
| 100% Glycerol solution | 15.0 g |
| Soy Peptone | 15.0 g |
| NaCl | 3.0 g |
| $CaCO_3$ | 1.0 g |

**Yeast Extract-Malt Extract (YEME)** for growth of *Saccharopolyspora* sp. L53-18 (Yaginuma et al., 1992)

| | |
|---|---|
| Yeast extract | 3.0 g |
| Bacto-peptone | 5.0 g |
| Malt extract | 3.0 g |
| Glucose | 10.0 g |
| Sucrose | 340.0 g |

Distilled water to 1 L
For YEME agar addition of 15 g Difco Bacto agar
After autoclaving, addition of $MgCl_2 \cdot 6H_2O$ to a 5 mM final concentration

### 2.3.2 Production media

All the SM Media protocols were provided from GlaxoSmithKline (formerly Glaxo Wellcome) and they are generally screening media for metabolite production in actinomycetes.

**ERY-P Medium** Suggested for production of Erythromycin

| | |
|---|---|
| Glucose | 50.0 g |
| Soy flour | 30.0 g |
| Ammonium sulphate | 3.0 g |
| Sodium chloride | 5.0 g |
| Calcium carbonate | 6.0 g |
| Distilled water to 1 L | |
| Adjust pH to 7.0 | |

**R5** for all overlay bioassays

| | |
|---|---|
| Sucrose | 103.0 g |
| $K_2SO_4$ | 0.25 g |
| $MgCl_2 \cdot 6H_2O$ | 10.12 g |
| Glucose | 10.0 g |
| Casamino acid | 0.1 g |
| Yeast Extract | 5.0 g |
| TES | 5.73 g |
| Trace elements solution | 2.0 mL |
| Distilled water to 1 L | |
| Adjust to pH 7.0 with 5M NaOH | |

For solid medium this was dispensed in 95 mL aliquots and add 22 g of Difco Bacto Agar (Becton & Dickinson 214030)

**Trace Elements for R5**

| | |
|---|---|
| $ZnCl_2$ | 0.04 g |
| $FeCl_2 \cdot 6H_2O$ | 0.2 g |
| $CuCl_2 \cdot 2H_2O$ | 0.01 g |
| $MnCl_2 \cdot 4H_2O$ | 0.01 g |
| $Na_2B_40_7 \cdot 10H_2O$ | 0.01 g |
| $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$ | 0.01 g |

**SM3**

| | |
|---|---|
| Glucose | 5.0 g |
| MD30E Maltodextrin | 50.0 g |
| Arkasoy soya flour | 25.0 g |
| Molasses (beet) | 3.0 g |
| $K_2HPO_4$ | 0.25 g |
| $CaCO_3$ | 2.5 g |
| Distilled water to 1 L | |
| Adjust pH to 7.0 with KOH | |

**SM5**

| | |
|---|---|
| Peptone (Oxoid L34) | 20.0 g |
| Lab Lemco (Oxoid L29) | 8.0 g |
| Glucose | 15.0 g |
| Glycerol soln. | 10.0 g |
| $CaCO_3$ | 0.4 g |
| Distilled water to 1 L | |
| Adjust pH to 7.2 with KOH | |

**SM6**

| | |
|---|---|
| Corn Steep Liquor | 40.0 g |
| MD30E Maltodextrin | 20.0 g |
| NaCl | 2.5 g |
| $MgSO_4$ | 0.5 g |
| Tap water to 1 L | |
| Adjust pH to 7.0 with KOH | |

**SM7**

| | |
|---|---|
| MOPS | 20.9 g |
| L-Proline | 15.0 g |
| Glycerol soln. | 20.0 g |
| Sucrose | 2.5 g |
| L-Glutamate | 1.5 g |
| NaCl | 0.5 g |
| $K_2HPO_4$ | 2.0 g |
| *0.2M $MgSO_4$ | 10.0 mL |
| *0.02 M $CaCl_2$ | 10.0 mL |
| *Trace salts No. 1 (CAB) | 5.0 mL |

*NOTE to be added pre-autoclaving

Distilled water to 1 L

Adjust pH to 6.5 with KOH

**SM12**

| | |
|---|---|
| Arkasoy soya flour | 10.0 g |
| Glucose | 50.0 g |
| Peptone (Oxoid L34) | 4.0 g |
| Lab Lemco (Oxoid) | 4.0 g |
| Yeast Extract (Oxoid) | 1.0 g |
| NaCl | 2.5 g |
| $CaCO_3$ | 5.0 g |

Tap water to 1 L

Adjust pH to 7.6 with KOH

**SM14**

| | |
|---|---|
| Glucose | 10.0 g |
| Soy Peptone (Lab M) | 20.0 g |
| Lab Lemco (Oxoid) | 5.0 g |
| NaCl | 5.0 g |
| $ZnSO_4 \cdot 7H_2O$ | 0.01 g |

Distilled water to 1 L

Adjust pH to 7.0 with KOH

**SM15**

| | |
|---|---|
| MOPS | 20.9 g |
| Casamino acids | 11.5 g |
| Glycerol soln. | 23.0 g |
| NaCl | 0.5 g |
| $K_2HPO_4$ | 0.52 g |
| EDTA | 0.25 g |
| *$MgSO_4 \cdot 7H_2O$ | 0.49 g |
| *$CaCl_2 \cdot 2H_2O$ | 0.029 g |
| *Trace salts No.1 | 5.0 mL |

*added pre-autoclaving

Distilled water to 1 L

Adjust pH to 6.5 with KOH

**SM18**

| | |
|---|---|
| Glucose | 15.0 g |
| Soluble starch (BDH) | 40.0 g |
| Pharmedia | 25.0 g |
| Molasses (beet) | 20.0 g |
| $CaCO_3$ | 8.0 g |

Tap water to 1 L

Natural pH 7.2 pre-sterilisation

**SM19**

| | |
|---|---|
| Tomato paste | 40.0 g |
| Oat flour (Avenaflo) | 15.0 g |
| Glucose | 2.0 g |

Tap water to 1 L

Natural pH 6.0 pre-sterilisation

**SM20**

| | |
|---|---|
| Maltose | 20.0 g |
| Peptone (Oxoid L37) | 5.0 g |
| Lab Lemco (Oxoid) | 5.0 g |
| Yeast extract (Oxoid) | 3.0 g |

| | |
|---|---|
| NaCl | 3.0 g |
| $MgSO_4 \cdot 7H_2O$ | 1.0 g |
| Tap water to 1 L | |
| Natural pH 7.2 pre-sterilisation | |

**SM25**

| | |
|---|---|
| Peptone (Oxoid L34) | 10.0 g |
| Malt extract (Oxoid) | 21.0 g |
| Glycerol soln. | 40.0 g |
| Distilled water to 1 L | |
| Natural pH 6.3-6.5 | |
| After autoclaving pH is 6.3 | |

**SM30**

| | |
|---|---|
| Tomato paste | 40.0 g |
| Oat flour (Avenaflo) | 15.0 g |
| Glucose | 2.0 g |
| Tap water to 1 L | |
| Adjust pH to 4.5 with HCl | |

**SM32**

| | |
|---|---|
| Peptone (Oxoid L34) | 10.0 g |
| Malt extract (Oxoid) | 21.0 g |
| Glycerol soln. | 40.0 g |
| Distilled water to 1 L | |
| Adjust pH to 4.5 with HCl | |

**Sporeamicin seed medium** for *Saccharopolyspora* sp. L53-18 (Yaginuma et al., 1992)

| | |
|---|---|
| Glucose | 10.0 g |
| Dextrin | 10.0 g |
| Yeast extract | 5.0 g |
| Casein hydrolysate | 5.0 g |
| Calcium carbonate | 1.0 g |
| Tap water to 1 L | |
| Adjust pH to 6.5 | |

**Sporeamicin production medium** for production of sporeamicin (Yaginuma et al., 1992)

| | |
|---|---|
| Glucose | 30.0 g |
| Corn steep liquor | 10.0 g |
| Dry yeast | 6.0 g |
| Cobalt chloride | 0.1 g |

Tap water to 1 L

Adjust pH to 7

## 2.4 Molecular biology methods

### 2.4.1 *Saccharopolyspora* genomic DNA preparation

*Saccharopolyspora* cultures were prepared by inoculating spores in 50 mL SV2 in 250 mL Erlenmeyer flask and subsequently grown for 3 days at 30°C (or until sufficient mycelium was observed). When the use of genomic DNA was for PCR applications, the isolation was performed by using the FastDNA™ SPIN KIT, (MP Biomedicals). High quality genomic DNA for sequencing was isolated by the salting-out procedure.

**Salting out procedure for the isolation of genomic DNA**

50 mL cultures were pelleted at 4000 g for 5 min, the supernatant was removed, and the mycelium was divided in 15 mL falcon tubes. The mycelium was resuspended in 10 mL of 10.3% sucrose solution and pelleted at 4000 rpm for 5 min. It was again resuspended in 5 mL SET buffer. 200 µL of lysozyme solution were added and subsequently 40 µL of RNAse ONE™ solution. The mix was incubated at 37°C at 125 rpm for 1 h. After 1 h, extra 100 µL of lysozyme were added and repeated if needed, until lysis was complete. 140 µL of proteinase K solution was added, mixed and 300 µL of SDS 20% was added, mixed by inversion and the mixture was incubated for 1 h at 55°C, with occasional inversion. Then 2 mL 5M NaCl were added and it was mixed by inversion, before letting to cool down to 37°C. Subsequently, 6 mL of chloroform were added with a glass pipette at the mixture, mixed by inversion for 30 min at 20°C. The mixture was pelleted at 4500 g for 15 min at 20°C. The supernatant was transferred to 4 falcon tubes and isopropanol was added up to 13 mL and mixed by inversion. After 3 min the aggregated DNA was spooled with a crooked glass pipette and transferred to a sterile microcentrifuge tube with 1 mL of 70% ethanol. DNA was store in -4°C till further use. When dried for use, DNA was stored in 1 mL TE buffer.

### 2.4.2 Polymerase chain reaction (PCR)

PCR reactions were performed in a T100™ Thermal Cycler (Bio-Rad). The general protocol for 50 µL reaction and the PCR conditions are presented (Table 2.5). The general protocol for colony PCR with the GoTaq master mix is presents at Table 2.6.

**Table 2.5: Protocol for 50 µL Q5 PCR reaction**

| Component | 50 µL reaction | Final Concentration |
|---|---|---|
| 5X Q5 Reaction Buffer | 10 µL | 1X |
| 10 mM dNTPs | 1 µL | 200 µM |
| 10 µM Forward Primer | 2.5 µL | 0.5 µM |
| 10 µM Reverse Primer | 2.5 µL | 0.5 µM |
| Template DNA | variable | < 1,000 ng |
| Q5 High-Fidelity DNA Polymerase | 0.5 µL | 0.02 U/µL |
| 5X Q5 High GC Enhancer | 10 µL | 1X |
| Nuclease-Free Water | to 50 µL | |

| Step | Temperature [°C] | Time | Repeat (×) |
|---|---|---|---|
| Initial denaturation | 98 | 3 min | 1 |
| Denaturation | 98 | 30 s | 30 |
| Annealing | 50-72 | 30 s | |
| Elongation | 72 | 45 s/kb | |
| Final elongation | 72 | 7 min | 1 |
| Hold | 12 | ∞ | 1 |

**Table 2.6: Protocol for 50 µL Colony GoTaq PCR reaction**

| Component | 25 µL reaction | Final Concentration |
|---|---|---|
| 2X GoTaq Green PCR Mix | 12.5 µL | 1X |
| 10 µM Forward Primer | 0.5 µL | 0.5 µM |
| 10 µM Reverse Primer | 0.5 µL | 0.5 µM |
| Template DNA | variable | < 250 ng |
| Nuclease-Free Water | to 25 µL | |

| Step | Temperature [°C] | Time | Repeat (×) |
|---|---|---|---|
| Initial denaturation | 95 | 5 min | 1 |
| Denaturation | 95 | 30 s | 30 |
| Annealing | 50-72 | 30 s | |
| Elongation | 72 | 45 s/kb | |
| Final elongation | 72 | 7 min | 1 |
| Hold | 12 | ∞ | 1 |

### 2.4.3 Agarose gel electrophoresis

Agarose gels were prepared by adding agarose to TBE buffer to a final concentration between 0.8 and 1.0%, depending on the size of DNA fragments for analysis. 10 µg/mL ethidium bromide and DNA were prepared by addition of loading dye before applying it to the gel. 1 kb or 2-log ladders (NEB) was applied on the gel. Gel electrophoresis was performed in TBE buffer using a PowerPac™ Universal Power Supply (Bio-Rad) run between 80-120 V, until separation of DNA fragments. A UV Transilluminator and Gel Documentation System (UVP) was used to visualise the samples ran on the gel.

### 2.4.4 Cloning

Standard molecular biology procedures were applied. DNA fragments were cut with the appropriate restriction endonucleases in 50 µL reaction volumes and incubated at 37°C for 2 h. 1 µL of alkaline phosphatase (AP) was added along with 6 µL of AP buffer and 3 µL dd$H_2O$ and incubated for a further 60 min at 37°C. Digests were purified by 0.8-1.0% agarose gel electrophoresis. Purified digest products were combined for a ligation for 16-18 h at room temperature in a 3:1 insert to vector ratio. At the end of the reaction, 5 µL of ligation mixture was transformed in competent *E. coli* DH5α cells.

### 2.4.5 Electrocompetent cell preparation

An overnight culture of 10 mL LB *E. coli* ET12567/pUZ8002 cells was prepared, with added chloramphenicol and kanamycin. 1 mL of the overnight culture was used to inoculate 100 mL of LB containing the appropriate antibiotics and incubated at 37°C and 250 rpm until $O.D._{600}$ reached approximately 0.4. After reaching desired OD, the content of the flask was split in two falcon tubes and centrifuged at 4°C and 7000 rpm for 5 min. The supernatant was discarded, and the pellet was resuspended in 50 mL of sterile, ice-cold milliQ water. The centrifugation was repeated at 4°C and 7000 rpm for 5 min and the supernatant was discarded once more. The contents of both falcon tubes were pooled together by resuspending one in 50 mL of sterile, ice-cold milliQ water and then transferring it to the other. The wash step was repeated 2 times in total and after the final wash step the cells were resuspended in a smaller volume of sterile, ice-cold 10% Glycerol (between 1-5 mL depending on cell density), pelleted at 4°C and 7000 rpm for 5 min. Finally, the pellet was resuspended with no more than 1 mL of sterile, ice-cold 10% Glycerol to make a very dense cell suspension and was dispensed in 50 µL aliquots. These were used either immediately or stored at -80°C for later use.

### 2.4.6 Bacterial transformation

#### *2.4.6.1* Transformation of chemically competent *E. coli*

50 µL cells were mixed with either 5 µL of ligation mix, 2 µL Gibson Assembly mix or 1 µL of vector DNA and incubated on ice for 30 min. Then cells were heat shocked at 42°C for 45 s before returning to ice for a further 2 min. 950 µL of LB medium were then added to the transformed cells and recovery was performed at 37°C, 250 rpm for 1 h. Cells were finally plated on LB agar with appropriate selection and incubated O/N at 37°C. Resulting colonies were screened by colony PCR to identify successful colonies.

#### *2.4.6.2* Transformation of electrocompetent *E. coli*

The electroporation cuvettes where prechilled on ice. The electrocompetent *E. coli* cells were thawed on ice, 50 µL per sample and transferred in the cuvette. 5-10 µL of a ligation reaction or 1 µL of plasmid DNA were transferred and mixed gently by shaking the cuvette. A BIORAD™ pulse controller was set to 25 µF, 1.5 kV, 800 Ω. The samples in the cuvettes were electroporated and immediately 950 µL of LB were added and mixed by gentle pipetting. The cells in Lb were transferred in 1.5 mL microcentrifuge tubes and incubated for 1 h at 220 rpm/ 37°C. After incubation, 10 or 100 µL of the cells were plated on LB agar containing an antibiotic. The plates were incubated o/n at 37°C. The following day 5 mL of LB with antibiotic were inoculated with a single colony for consequent plasmid extraction.

### 2.4.7 Plasmid extraction and Sanger sequencing

The 10 mL of *E. coli* LB cultures with appropriate antibiotic were established from single colonies on plates as described in 2.4.6 or from -80°C 20% glycerol stocks and were grown O/N at 37°C, 250 rpm. Plasmid DNA was purified using a QIAprep® Spin Miniprep Kit (Qiagen). All constructs generated were sequenced by Sanger sequencing (Eurofins Genomics).

### 2.4.8 Conjugation in Actinomycetes

This protocol is based to the conjugation protocol from Kieser et al. (2000) but it has modifications to enhance conjugation effectiveness in *Saccharopolyspora* species. Electro-competent cells of *E. coli* ET12567/pUZ8002 with the *oriT* containing plasmid were transformed and selected for the incoming plasmid only, as described. A colony from the plate was inoculated into 5 mL of LB containing chloramphenicol (25 µg/mL

final), kanamycin (50 μg/mL final) and the antibiotic used to select for the *oriT*-containing plasmid and inoculated at 37ºC overnight. A well grown overnight culture was diluted to 1:100 (500μl) in fresh 50 mL LB and appropriate antibiotics in half concentration than the overnight in a 250 mL flask and grown at 37ºC, shaking at 250 rpm until O.D. $_{600}$ = 0.4-0.6 (normally after 4-5 h). The growth of *E. coli* cells was stopped by placing the flask on ice. The whole broth was transferred in a falcon tube and the cultures were pelleted at 2500 rpm at 4ºC. The supernatant was discarded, and the pellet was washed with 50 mL LB. The wash step was repeated twice and then the *E. coli* cells were resuspended in 1 mL LB. At the same time, 150 μL from the thick spore stock solution (instead of commonly used 50 μL) of the strain we aim to conjugate into, were spun to remove 20% glycerol, then resuspended in 0.5 mL of LB and heat shocked at 50°C for 10 min. 0.5 mL *E. coli* cells and 0.5 mL spores solution were mixed and plated in various dilutions on SF+ M plates containing 10 mM $MgCl_2$. The plates were left to dry in the cabinet and then incubated at 28-30ºC for 16-20 h. The 30 mL agar plates were overlayed with 1 mL water containing the appropriate amount of antibiotics (nalidixic acid 12.5 μg/mL final concentration in the plate and antibiotic to select for the incoming plasmid, e.g. apramycin 50 μg/mL final concentration). Incubation was continued until exconjugants appeared.

### 2.4.9 Gibson Assembly

Reactions were assembled following manufacturer's instructions containing 10 μL of 2×Gibson Assembly Master Mix (Gibson et al., 2009), (NEB) and a calculated ratio of PCR products, made up with ddH2O to a final reaction volume of 20 μL. Samples were then incubated at 50°C for 15 min in a T100™ Thermal Cycler (Bio-Rad), before transformation of the ligation mix into NEB® 5-alpha Competent (High Efficiency) *E. coli*.

## 2.5 Expression, mutagenesis and complementation experiments

### 2.5.1 Ordering operons from GenScript®

The synthetic operons pEVK1, pEVK2 and pEVK3 described in chapters 3 and 4 were ordered from GenScript® and received in a pUC57 vector.

### 2.5.2 Disruption of *spoAII* in *Saccharopolyspora* sp BR-2231

In order to disrupt *spoAII* gene in *Saccharopolyspora* sp BR-2231, primers eryAf and eryAr were designed to amplify 2433 bp of the gene, for cloning as a HindIII-EcoRI fragment into the suicide vector pKC1132. This would then be used to disrupt the WT chromosomal copy of the gene by single cross-over. A set of sequencing primers were used for verification of the construct by sequencing (eryAII_Int1a_forw, eryAII_Int1b_forw, eryAII_Int2a_forw, eryAII_Int2b_forw, eryAII_Int3a_rev, eryAII_Int3b_rev, eryAII_Int4a_for, eryAII_Int4b_for, eryAII_Int4c_rv).

### 2.5.3 Construction of a *kyaA* deletion plasmid

Two PCR primers, kyaAdelF and kyaAdelR, were used to generate a replacement cassette using pIJ10700 as a template. This generated a replacement cassette with a hygromycin resistance gene, flanked with FLP recombinase recombination sites. This was used to replace *kyaA* in pEVK3. The kyaAdelF and kyaAdelR primers possess 39 bp 5' extensions that correspond to either 39 bp upstream of and including the start codon of *kyaA* in the case of kyaAdelF or 39 bp downstream of and including the stop codon of *kyaA* in the case of kyaAdelR. These 5' extensions were used by the lambda phage recombinase to replace the *kyaA* gene with the PCR generated hygromycin cassette via a double crossover. The product of this process, pWDW55, was then used to transform DH5αBT340. This strain has an FLP recombinase gene carried on BT340, a plasmid with a temperature sensitive replicon. Transformants were then grown for 16-18 h in 10 mL of LB at 42°C without any antibiotic selection. Next day a plasmid extraction was carried out from those cultures and the plasmids were used to transform DH5α. Transformants were then selected for carbenicillin resistance and hygromycin sensitivity. This resulted in the plasmid pWDW59 which carries a 81 bp scar in place of *kyaA*. This was confirmed by sequencing using primers deltaKyaAseqF and deltaKyaAseqR. A 4.6 kb EcoRI/KpnI fragment of pWDW59 covering the 5' end of the biosynthetic cluster was cloned into the 8 kb vector fragment of EcoRI/KpnI cut pWDW63 to reconstitute the BGC but with *kyaA* replaced with a scar, called pWDW65.

### 2.5.4 Complementation of Δ*kyaA* with an ermE* promoter plasmid

A PCR product was generated using the GenScript® plasmid pEVK3 as template and Q5 polymerase with the set of primers cinAcompF and cinAcompR that covered the whole coding region of the *kyaA* gene. That was digested by *Nde*I/*Xba*I and cloned into

pADW11 also cut *Nde*I/*Xba*I, under the control of an *erm*E* promoter. The vector pADW11 is based on pOSV556t, a plasmid that integrates at a pSAM2 integration site (Sydor and Challis, 2012), carries a thiostrepton resistance gene for selection in *S. coelicolor* and a carbenicillin resistance gene for selection in *E. coli*. This resulted in plasmid pWDW61.

### 2.5.5 Complementation of Δ*kyaA* with a kyaR1 promoter plasmid

A *Nde*I/*Eco*RI fragment from pWDW61 carrying the *kyaA* and thiostrepton resistance genes was cloned into *Nde*I/*Eco*RI cut pADW19. In this plasmid, the *kyaA* gene is under the control of *kyaN* promoter, which is activated by an erm*E**p::*kyaR1* gene also located on pADW19.

### 2.5.6 Complementation of Δ*kyaA* in cis

A partial deletion of *kyaA* in which the core peptide coding region of *kyaA* was replaced with a *Stu*I restriction site, was constructed. A PCR was generated (with primer set KY3smNdeI-F and KY3smStuI-R) covering from the *Eco*RI site to the second base pair of the codon for *KyaA* Val 59. A *Stu*I site was created using the AGG sequence from the last base pair of the E58 codon and the first two base pairs from the V59 codon with CCT added to give AGGCCT the StuI site. Six random base pairs were inserted between the *Stu*I site and a *Kpn*I site located at the end of the PCR fragment. This was cut with *Eco*RI/*Kpn*I and cloned into *Eco*RI/*Kpn*I cut pBlueScriptIIKS to yield pWDW68. A further PCR was generated (with primers set KY3lgStuI-F and KY3lgNdeI-R2) with a *Stu*I site at the end of the PCR corresponding to where the core peptide coding region and stop codon of *kyaA* would have been and with the *Kpn*I site located in the *kyaX* gene. This PCR was cut with *Stu*I/*Kpn*I and cloned into *Stu*I/*Kpn*I cut pWDW68, to give pWDW69. The *Eco*RI/*Kpn*I fragment from this PCR was then subcloned into the 8 kb fragment of *Eco*RI/*Kpn*I cut pWDW63, to give pWDW70.

This construct has a *Stu*I site were the core peptide coding region had been. A synthetic double-stranded operon encoding for the kyamicin core peptide was ordered and cloned into pWDW70 via GIBSON. The pWDW70 carrying the synthetic kyamicin core peptide was named pEVK10.

## 2.6 Bioassays

All bioindicator strains used in bioassays were grown from a single colony in 10 mL LB for 16-18 h at 37°C. 500 µL of the overnight culture was transferred to 10 mL fresh LB and grown till $O.D._{600}$ = 0.4 - 0.6.

### 2.6.1 Overlay bioassays

For each *Saccharopolyspora* strain to be tested either 5 µL or a streak from the spore stock were applied in the centre of an SF+M plate and left to grow for seven days at 30°C.

500 µL of the exponential culture of the bioindicator strain was mixed with 5 mL of soft nutrient agar (SNA). The mixture was used to overlay the *Saccharopolyspora* plate, left to air dry and incubated at room temperature overnight. The result was observed the following morning, with a zone of clearance around the *Saccharopolyspora* streak indicating the production of an inhibitory compound.

### 2.6.2 Disk and agar diffusion assays

For assays with crude extracts 20 µL of each dilution were applied to sterile disks (6 mM), which were allowed to dry and transferred to SNA plates infused with bioindicator strains as described above.

In order to determine the minimum inhibitory concentration (MIC) values of lantibiotic compounds a stock solution of 1000 µg/mL in sterile water was prepared, along with serial dilutions from 256 – 8 µg/mL. As disk assays resulted in small inhibition zones, direction application of the compounds to the agar was trialled. 5 µL of the dilutions were applied directly to the agar which allowed for better diffusion of the compounds and clear inhibition zones. The MIC was defined as the lowest concentration for which a clear zone of inhibition was observed.

## 2.7 Extraction protocols

Unless stated otherwise, all small-scale extractions described in chapters 5 and 6, were done by mixing equal volumes of solvent and sample, followed by vortexing and shaking for 20 min. The mixture was subject to centrifugation and resultant supernatant was then analysed by HPLC.

### 2.7.1 Extraction from liquid for isolation of kyamicin

M1152/pWDW63/pEVK6 was grown in 6 L of tryptic soy broth (TSB) at 28°C for seven days. The cells were harvested and extracted at with methanol/water (1:1; 500 mL), with ultrasonication for 2 h and subsequent shaking for 16 h. After centrifugation, the supernatant was filtered and concentrated under vacuum giving 613.4 mg of crude material. This was redissolved in a minimum volume of 50% methanol and loaded on to reversed phase solid phase extraction (SPE) cartridges. These were then washed with water to remove polar compounds and eluted with 50% methanol. The resulting fractions were combined, concentrated under vacuum and then further purified by semi-preparative HPLC to yield pure kyamicin (2.5 mg).

### 2.7.2 Extraction from agar plates

R5 agar plates were inoculated with M1152/pWDW63/pEVK6 and grown at 30°C for 14-20 days. The plates were frozen at -20°C for 16 h, thawed and chopped, before adding an equal volume of methanol and leaving for 1 h. The supernatant was collected, the methanol extraction was repeated, and the collective supernatant was dried down and processed as above.

### 2.7.3 Extraction from overlay bioassay plates

All overlay bioassay plates were analysed in the same way. Agar plugs of ¼ inch diameter, were cut with a cork borer from the zone of clearing (or close to the growth streak in the case of negative control) and transferred to microcentrifuge tubes. Samples were frozen at -80°C for 10 min and allowed to thaw. 300 µL of 5% formic acid was added to each tube, which was vortexed and shaken for 20 min. The same was subject to centrifugation at 13000 rpm for 15 min. The supernatants were carefully collected and transferred to filter vials (HSTL labs) for HPLC analysis.

## 2.8 Chemical analysis, isolation and characterisation of natural products

### 2.8.1 Analytical LCMS

#### 2.8.1.1 UHPLC-MS

Measurements were performed on a Nexera X2 liquid chromatograph (LC-30AD) LCMS system (Shimadzu) connected to an autosampler (SIL-30AC), a Prominence column oven (CTO-20AC) and a Prominence photo diode array detector (SPD-M20A). A Kinetex® 1.7 µm C18 100 Å, 100×2.1 mm column (Phenomenex) was used with a gradient of water (0.1% formic acid)/methanol. Starting conditions: 90/10, hold at 90/10 for 1 min, to 0/100 within 9.00 min, hold for 2.00 min, to 90/10 from within 0.5 min, hold at 90/10 for 0.5 min. The UHPLC-System was connected to a LCMS-IT-TOF Liquid Chromatograph mass spectrometer (Shimadzu).

#### 2.8.1.2 UHPLC-HRMS

Data were acquired with an Acquity UHPLC system (Waters Corporation) equipped with an ACQUITY UHPLC® HSS T3 1.8 µm, 2.1 × 100 mm column (Waters Corporation) connected to a Synapt G2-Si high-resolution mass spectrometer (Waters Corporation).

For analytical UHPLC a gradient between mobile phase A ($H_2O$ with 0.1 % formic acid) and mobile phase B (acetonitrile with 0.1% formic acid) at a flow rate of 200 µL/min was used. Initial conditions were 1% B for 1 min, ramped to 60% B within 9 min, ramped to 99% B within 1 min, held for 2 min, returned to 1 % B within 0.1 min and held for 4.9 min.

MS spectra were acquired with a scan time of 1 s in the range of $m/z$ = 50 - 1200 in positive $MS^E$ resolution mode. The following parameters were used: capillary voltage of 3 kV, sampling Cone 40, source offset: 80, source temperature of 120°C, desolvation temperature of 350°C, desolvation gas flow of 800 L/h.

A solution of sodium formate was used for calibration. A solution of leucine encephalin ($H_2O$/MeOH/formic acid: 49.95/49.95/0.1) was used as lock mass and has been injected every 20 s during the runs. The lock mass has been acquired with a scan time of 0.3 s and 3 scans to average. The lock mass (556.2766) has been applied during data acquisition.

### 2.8.2 Semi-preparative HPLC

For purification of EV60 semi-preparative HPLC was performed on a 1100 Series System (Agilent Technologies) equipped with a Synergi™ 4 µm Fusion-RP 80 Å LC column 250×4.6 mm (Phenomenex). The following gradient was used: water (0.1% formic acid)/ACN, starting conditions: 90/10, hold for 1 min, to 68/32 within 12 min, to 0/100 within 0.5 min, hold for 3.5 min, to 90/10 within 0.5 min, hold for 1.5 min. For recovery of the material, the collected fractions were concentrated under reduced pressure to remove acetonitrile, then loaded onto an SPE cartridge, washed with water and finally eluted with methanol and the solvent reduced under reduced pressure.

For kyamicin semi-preparative HPLC was carried out over a Phenomenex Gemini-NX reversed-phase column (C18, 110 Å, 150 × 21.2 mm) using a Thermo Scientific Dionex UltiMate 3000 HPLC system. A gradient was used with mobile phases of A: $H_2O$ (0.1% formic acid) and B: methanol; 0–1 min 10% B, 1–35 min 10-85% B, 35–40 min 85-100% B, 40–45 min 100% B, 45-45.1 min 100-10% B, 45.1-50 min 10% B; flowrate 20 mL/min; injection volume 1000 µL. Absorbance was monitored at 215 nm and fractions (20 mL) were collected and analysed by LCMS. Kyamicin was observed in fractions 22-25 which were combined and concentrated to yield the pure compound as an off-white solid (2.5 mg).

### 2.8.3 Reduction of kyamicin

Kyamicin (1 mg) was dissolved in methanol (0.5 mL) and added to an aqueous solution of 20 mg/mL $NiCl_2$ (0.5 mL). The solution was added to a tube containing 5 mg of $NaBH_4$, resulting in the generation of hydrogen gas and the formation of a black $Ni_2B$ precipitate. The tube was immediately sealed, and the mixture was stirred at 55°C. The reaction progress was monitored by UHPLC-HRMS as described above, for which a peak at 8.45 min with an *m/z* of 899.3556 was observed for kyamicin ($[M + 2H]^{2+}$ *m/z* calculated for $C_{76}H_{108}N_{20}O_{25}S_3$ = 899.3551). The successive formation of peaks with the following masses were observed: *m/z* = 884.3767 ($[M + 2H]^{2+}$ m/z calculated for $C_{76}H_{110}N_{20}O_{25}S_2$ = 884.3768), retention time 8.38 min; *m/z* = 869.3990 ($[M + 2H]^{2+}$ *m/z* calculated for $C_{76}H_{112}N_{20}O_{25}S$ = 869.3987), retention time 8.01 min and 854.4203 ($[M + 2H]^{2+}$ *m/z* calculated for $C_{76}H_{114}N_{20}O_{25}$ = 854.4204), retention time 7.91 min corresponding to the successive reduction of the three thioether bridges in the molecule. After 5 h only the 854.42 *m/z* ion could be observed, indicating that the kyamicin starting material had been completely reduced. The precipitate was collected by centrifugation at 13,000 rpm for 10 min. As the reaction supernatant contained only trace amounts of the desired product, a

fresh solution of MeOH/$H_2O$ 1:1 (0.5 mL) was added to the precipitate and it was subject to ultrasonication for 30 min. Reduced kyamicin was then detected in sufficient quantities for the following fragmentation experiments.

### 2.8.4 Fragmentation of reduced kyamicin

For ESI MS2 fragmentation the mass of interest (854.42) was selected using an inclusion list and fragmented using data directed analysis (DDA) with the following parameters: top3 precursor selection (inclusion list only); MS2 threshold: 50,000; scan time 0.5 s without dynamic exclusion. Collision energy (CE) was ramped between 15-20 at low mass (50 m/z) and 40-100 at high mass (2000 m/z). Further increase of the CE to 20-30/60-120 led to complete fragmentation.

For MALDI-TOF analysis the samples were mixed with α-cyano-4-hydroxycinnamic acid as matrix and analysed on an AutoflexTM Speed MALDI-TOF/TOF mass spectrometer (Bruker DaltonicsTM GmbH, Coventry, UK). The instrument was controlled by a flexControlTM (version 3.4, Bruker) method optimised for peptide detection and calibrated using peptide standards (Bruker). For sequence analysis fragments produced by PSD were measured using the LIFT method (Bruker). All spectra were processed in flexAnalysisTM (3.4, Bruker).

### 2.8.5 NMR

NMR measurements were performed on both Bruker AVANCE III 400 MHz and Bruker AVANCE III 800 MHz spectrometers. Chemical shifts are reported in parts per million (ppm) relative to the solvent residual peaks of DMSO-$d_6$ ($^1$H: 2.50 ppm, quintet; $^{13}$C: 39.52 ppm, septet).

# Chapter 3:

# Discovery and characterisation of kyamicin, a new lantibiotic from plant-ant derived *Saccharopolyspora* strains

# 3 Discovery and characterisation of kyamicin, a new lantibiotic from plant-ant derived *Saccharopolyspora* strains

## 3.1 Introduction

The current need for novel antimicrobial compounds focused our research on the investigation of three *Saccharopolyspora* strains that were isolated from fungus farming *Tetraponera penzigi* ant samples from various locations from Kenya. Previous investigation of new *Streptomyces* species isolated from this underexplored ecological niche identified antimicrobial polyketide natural products with activity against multidrug-resistant microorganisms (Qin et al., 2017).

The three strains were sequenced using the PacBio RSII platform and bioinformatics analysis suggested the presence of approximately 23 BGCs in each case. Many of these appear to encode new or unusual members of chemotypes known to possess potent biological activity, including several potentially anti-infective molecules. The isolate *Saccharopolyspora sp.* KY21 proved most tractable under laboratory conditions and was chosen as the basis for further study.

We noticed that all three strains contain a cinnamycin-like BGC (Widdick et al., 2003) that proved silent despite culturing under a wide range of conditions. Cinnamycin is a type B lantibiotic produced by *Streptomyces cinnamoneus* DSM 40005 and is active against a broad range of Gram-positive bacteria (Hans-Georg Sahl and Bierbaum, 1998). Therefore, the aim of this work was to activate the cinnamycin-like BGC and subsequently isolate and characterise the compound herein referred to as kyamicin (in reference to the Kenyan origin). We eventually activated the cinnamycin-like BGCs by heterologous expression of two native genes under a constitutive promoter: these were a PE methyl transferase resistance gene, and a pathway-specific positive regulator, two essential genes for the expression of lantibiotics as described in literature (O'Rourke et al., 2017). This yielded the novel antibiotic kyamicin.

## 3.2 Objectives

The main objectives of this project were to:

- Investigate the cinnamycin-like BCG present in three novel *Saccharopolyspora* strains.
- Activate the kyamicin cluster in the native host.
- Heterologously express of the kyamicin cluster in *Streptomyces coelicolor* M1152.
- Scale up the production of kyamicin, to isolate material for characterisation of its chemical structure and to test its bioactivity.

## 3.3 Genome sequencing of KY strains

The *Saccharopolyspora* strains KY3, KY7 and KY21 were isolated from the domatia of *T. penzigi* fungus farming plant ants isolated from two different locations in Kenya. The *T. penzigi* workers were provided by Naomi E. Pierce (Harvard, U.S.A.) and were sampled from Kitengela and Ngong Hills from southern Kenya, as well as Mpala Road from central Kenya as shown in **Figure 3.1** (Seipke et al., 2013). *Tetraponera* microbiomes contain more than 75% Proteobacteria but interestingly only 5% of Actinobacteria.

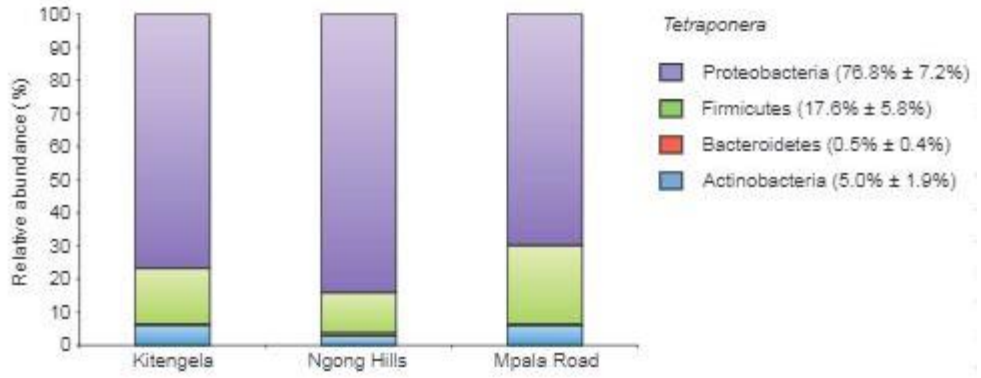



**Figure 3.1: Relative pyrotag abundance of phyla in cuticular microbiomes of *Tetraponera* ants.** The cuticles of the ants are dominated by Proteobacteria. Used with permission from *MicrobiologyOpen* (Seipke et al., 2013).

We explored those three fungus-growing, ant-associated strains using a combined approach of genomics, unbiased metabolomics, and bioassays. The strains were firstly genetically analysed.

After amplifying 16S rDNA and Sanger sequencing with the universal primers 533F and 1492R it was shown that KY3 and KY7 were identical, while KY21 had a single base pair difference. (Genbank accession numbers JX306001, JX306003, JX306004, respectively). All three strains identify with *Saccharopolyspora* 16S rDNA sequences available in public databases by 99%.

Genomic DNA was then isolated from the KY strains and sequenced at the Earlham Institute (Norwich, UK) via PacBio with SMRT sequencing technology and assembled using the HGAP3 and HGAP2 pipeline (Chin et al., 2013).

The genome analysis was performed by Dr Juan Pablo Gomez-Escribano. Assembly of the sequence data for strain KY21 gave a contig of 6370354 bp that represents a circular chromosome; this also contains a plasmid of 50300 bp. The strain KY3 is organized in a

circular chromosome of 6328480 bp. For strain KY7 we obtained two contigs of 4932021 and 1392929 bp that do not share any overlapping sequence and most likely represent a split circular chromosome.

After performing an alignment between the KY3 and KY7 genome sequences with RAST SEED Viewer and BLAST dot plot (Aziz et al., 2008, Boratyn et al., 2012),a full synteny was shown along their genomes with 99-100% sequence identity at the nucleotide level suggesting that KY3 and KY7 are the same strain and are different to KY21. A visual representation of the genome of KY21 is presented in Figure 3.2.

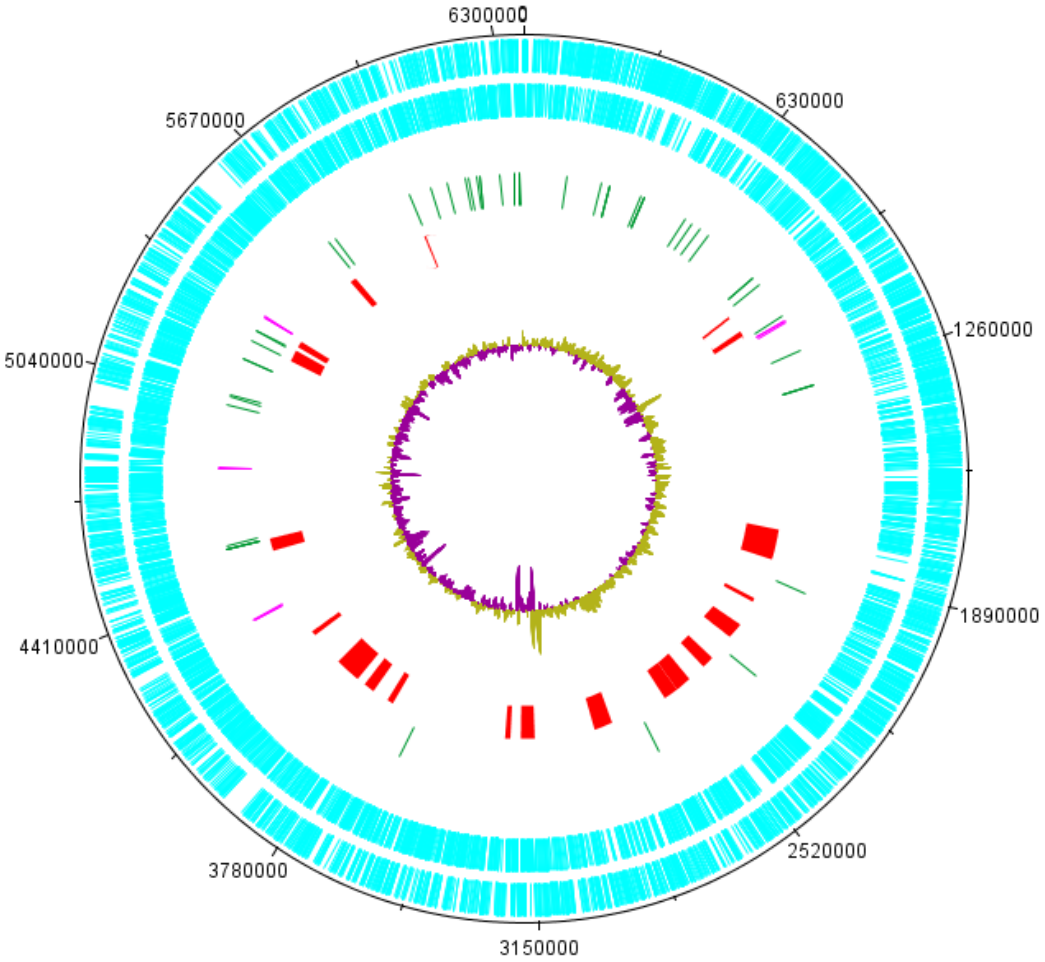



**Figure 3.2: *Saccharopolyspora* KY21 genome as a circular plot.** The tracks from the outside to the inside represent: Forward CDS; Reverse CDS; rRNA - green and tRNA – pink; clusters – red; GC skew [(GC)/(G+C)].

After annotating the sequencing data of the strains with antiSMASH (Weber et al., 2015) it was observed that the three genomes encode for 23 BGCs that mostly overlap (Table 3.1). This project is focused on a cinnamycin-like lantipeptide BGC (highlighted) present in all three strains.

Table 3.1: antiSMASH output of strain *Saccharopolyspora* sp. KY21. List of BGCs with prediction of known compounds based on structural patterns. Wherever the similarity was lower than 20 %, it has been annotated as no significant similarities. The lantipeptide cinnamycin-like BGC has been highlighted.

| **antiSMASH Cluster No.** | **BGC type** | **Position** | | **BGCs of compounds with highest similarity (%)** |
|---|---|---|---|---|
| | | **from** | **to** | |
| 1 | Ectoine | 956878 | 967267 | Ectoine (75%) |
| 2 | Terpene | 1280939 | 1302951 | - |
| 3 | Terpene | 1603794 | 1625917 | No significant similarities |
| 4 | NRPS | 1630963 | 1692484 | No significant similarities |
| 5 | T1PKS | 2399316 | 2448369 | No significant similarities |
| 6 | Aminoglycoside/ aminocyclitol | 2798692 | 2819888 | No significant similarities |
| 7 | Lassopeptide | 2966632 | 2988078 | Chaxapeptin (28%) |
| 8 | Trans-AT PKS | 2976109 | 3056206 | Dorrigocin/migrastatin (81%) |
| 9 | T2PKS | 3090576 | 3133037 | Tetracenomycin (33%) |
| 10 | Terpene | 3197311 | 3219704 | No significant similarities |
| 11 | Terpene | 3681982 | 3703022 | Isorenieratene (57%) |
| 12 | NRPS | 3741605 | 3797270 | Streptothricin (83%) |
| 13 | Other | 4030163 | 4074212 | No significant similarities |
| 14 | Terpene | 4072067 | 4098200 | Hopene (30%) |
| 15 | NRPS-T1PKS | 4331679 | 4388369 | SGR_Polycyclic tetramate macrolactam (50%) |
| 16 | NRPS | 4391798 | 4456904 | Coelibactin (54%) |
| 17 | Arylpolyene | 4528923 | 4570269 | No significant similarities |
| **18** | **Bacteriocin-Lantipeptide-NRPS** | **4693417** | **4743949** | **Cinnamycin (47%)** |
| 19 | Terpene | 4840778 | 4861980 | No significant similarities |
| 20 | T1PKS | 5022943 | 5142561 | FR-008 (76%) |
| 21 | Terpene | 5926466 | 5949174 | Brasilicardin_A (45%) |
| 22 | Butyrolactone | 6010332 | 6022041 | - |

## 3.4 Annotation of the kyamicin BGC

As mentioned above, in all three *Saccharopolyspora* KY strains, we observed the same cinnamycin-like BGC. In collaboration with Dr David Widdick, the cluster was studied, and annotated. The kyamicin BGC includes all the genes essential for the production of the lantibiotic and it resembles the cinnamycin cluster as shown in Figure 3.3.

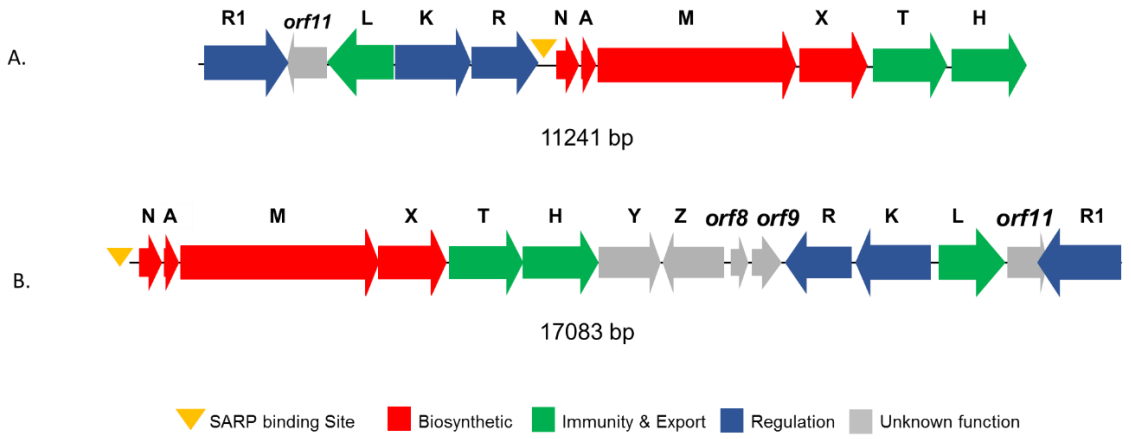



**Figure 3.3: Comparison of A. kyamicin (this study) and the B. cinnamycin (Widdick et al., 2003) BGCs.**

When we compared the kyamicin BCG to the cinnamycin one, we saw that all the biosynthetic and regulatory genes necessary for the production and export of the lantibiotic were present in the kyamicin cluster and homologous to the ones for cinnamycin. We also noted that although *cinorf11* is not essential in the heterologously expressed cluster, it is still present in these clusters. Also, in Figure 3.4, I present the comparison between the amino acid sequence of the core peptides of kyamicin and the other known lantibiotics of this type: cinnamycin (Widdick et al., 2003), cinnamycin B (Kodani et al., 2016), duramycin (Huo et al., 2017) and mathermycin (Chen et al., 2017).

The amino acid identity between proteins encoded by each BGC is presented in Table 3.2 and it applies to all three KY strains. Both cinnamycin and kyamicin contain 19 amino acids but they differ in seven positions. Whereas cinnamycin has a globular structure with one lanthionine (Lan) and two methyllanthionine (MeLan) bridges, kyamicin has three MeLan bridges, as do the most closely related molecules, mathermycin and cinnamycin B (Figure 3.4).

**Table 3.2: Amino acid identity (%) between proteins encoded by the cinnamycin and kyamicin BGCs.**

| Kyamicin | Cinnamycin | Amino Acid Identity | Function |
|---|---|---|---|
| KyaN (123aa) | CinN (119aa) | 79/119 (66%) | Formation of lysinoalanine bridge |
| KyaA (78aa) | CinA (78aa) | 50/79 (63%) | Precursor peptide |
| KyaM (1065aa) | CinM (1088aa) | 627/1080 (58%) | Formation of lanthionine residues |
| KyaX (302aa) | CinX (325aa) | 175/322 (54%) | Hydroxylation of aspartate 15 |
| KyaT (327aa) | CinT (309aa) | 202/292 (69%) | Export |
| KyaH (294aa) | CinH (290) | 187/291 (64%) | Export |
| Not Present | CinY | - | Not essential |
| Not present | CinZ | - | Not essential |
| Not present | Cinorf8 | - | Not essential |
| Not present | Cinorf9 | - | Not essential |
| KyaR (216aa) | CinR (216aa) | 160/209 (77%) | Regulation |
| KyaK (372aa) | CinK (354aa) | 195/344 (57%) | Regulation |
| KyaL (226aa) | CinL (236aa) | 131/220 (60%) | Immunity |
| Kyaorf11 (295aa) | Cinorf11 (396aa) | 94/178 (53%) | Not essential |
| KyaR1 (260aa) | CinR1 (261aa) | 129/252 (51%) | Regulation |

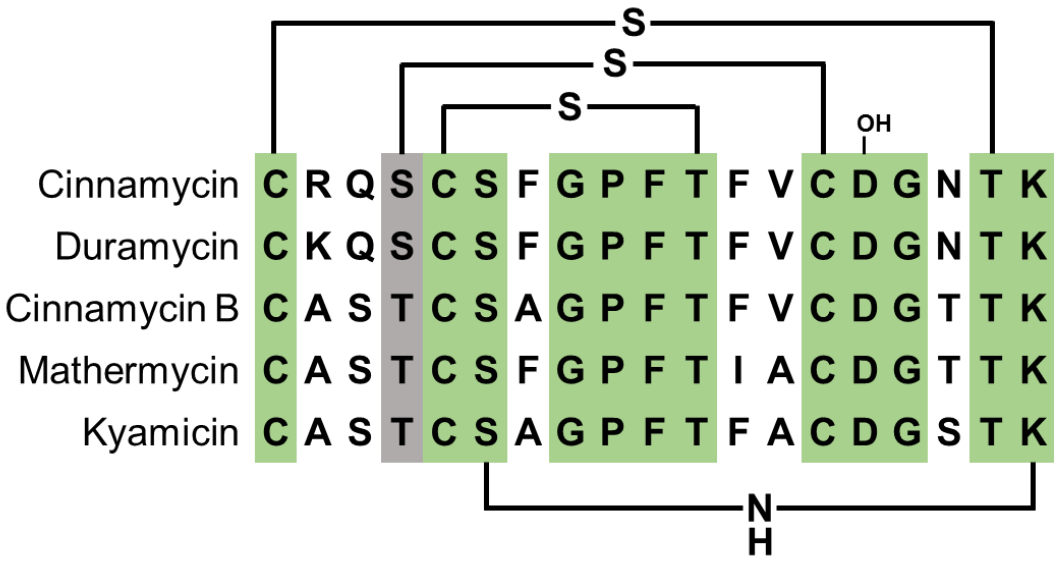



**Figure 3.4: Amino acid sequence comparison of the core peptides of lantibiotics discussed in this thesis.** Sequence alignment of the core peptides of cinnamycin, duramycin, cinnamycin B, mathermycin and kyamicin, highlighting the positions of posttranslational modifications present in the mature molecules. Lanthionine and methyllanthionine rings are indicated by thioether linkages, between cysteines and dehydrated serines or threonines, respectively. The hydroxylated aspartate residue, D15, and the lysinoalanine bridge between K19 and S6 are indicated.

Based on the published biosynthesis of cinnamycin (O'Rourke et al., 2017), in Figure 3.5 the proposed biosynthesis of kyamicin is presented.

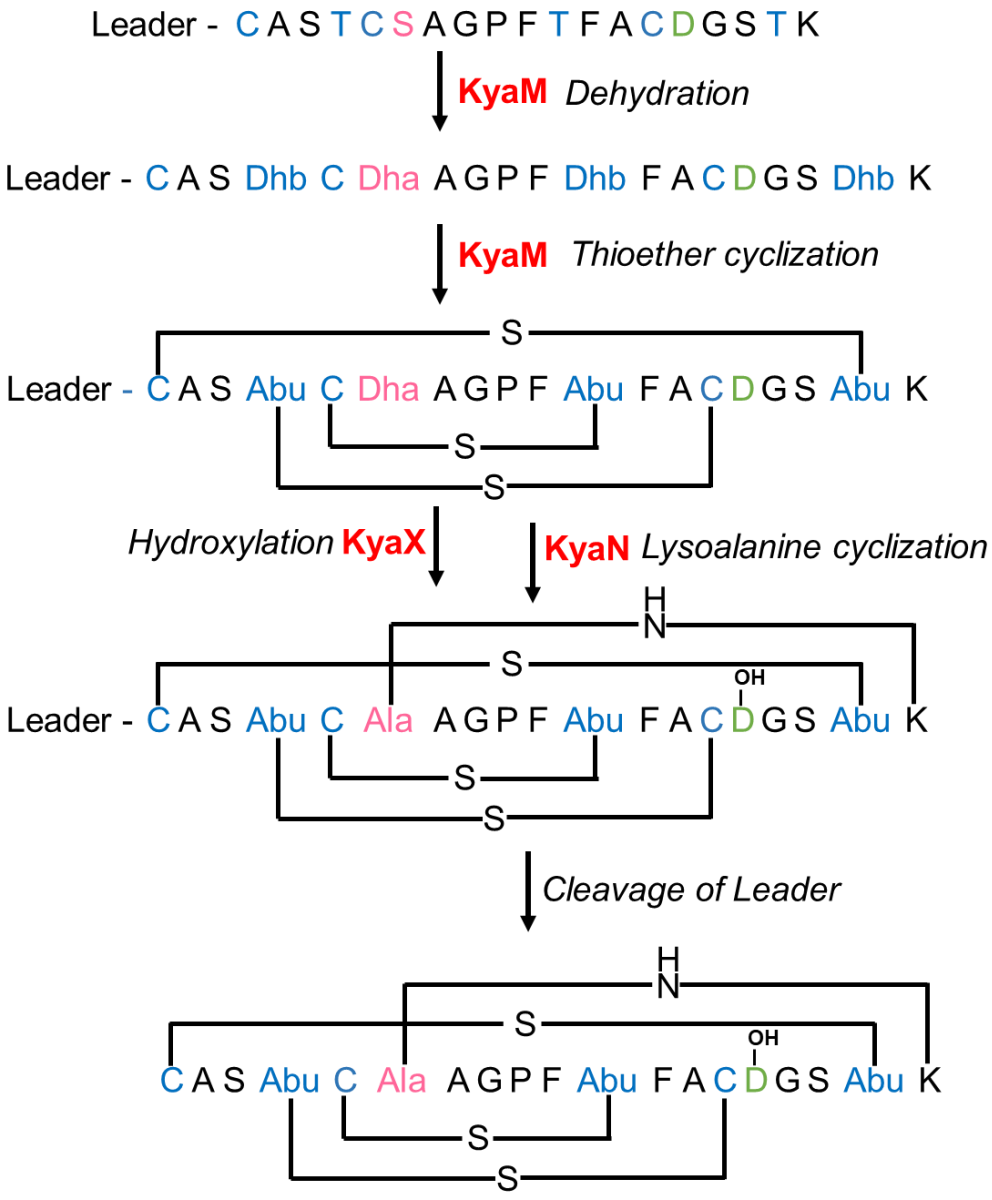



**Figure 3.5: Proposed biosynthesis of kyamicin.** The thioether bridges are formed first by a dehydration of T4, T11, T18 and S6 by KyaM to form dehydrobutyrine (Dhb) and dehydroalanine (Dha) residues, respectively. After thioether cyclization by KyaM, Dhb becomes S-linked Abu and Dha becomes S-linked Ala. D15 is hydroxylated by KyaX and the lysinoalanine bridge is formed between Dha6 and K19 catalysed by KyaN. After the core peptide is fully modified, the leader peptide is cleaved proteolytically to yield mature kyamicin.

The postulated biosynthetic pathway predicts the expected molecular formula of kyamicin as $C_{76}H_{108}N_{20}O_{25}S_3$. This corresponds to the loss of four $H_2O$ molecules in the dehydration steps and one addition of O from the hydroxylation of D15. The structure of kyamicin after post-translational modifications is shown in **Figure 3.6**.

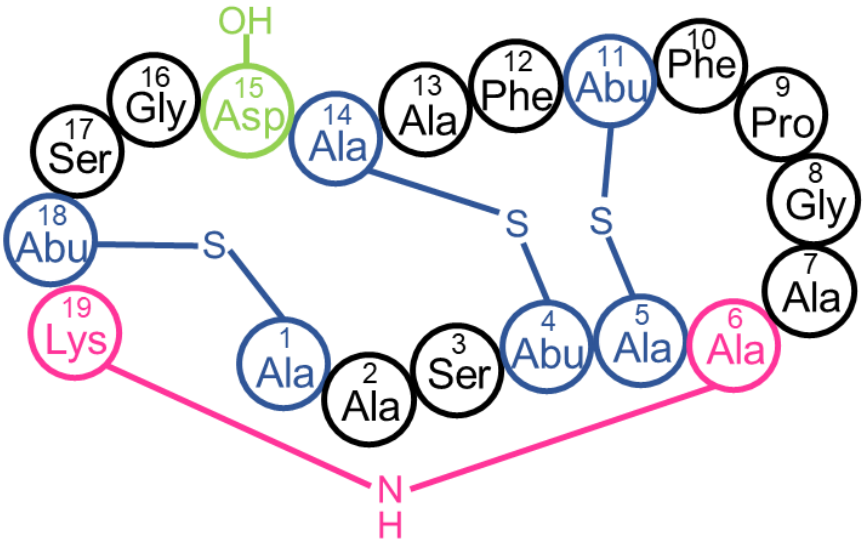



**Figure 3.6: Structure of kyamicin after the post-translational modifications.** Methyllanthionine bridges are indicated by blue thioether linkages, between cysteines and dehydrated threonines that become S-linked alanine and aminobutyric acid (Abu) residues, respectively. The hydroxylated aspartate residue, D15, is indicated in green and the lysinoalanine bridge between K19 and S6 are indicated in pink.

## 3.5 Production of the kyamicin BGC by *Saccharopolyspora* sp. KY strains

### 3.5.1 Growth trials for kyamicin detection

As cinnamycin and duramycin are known to inhibit Gram positive bacteria, the indicator strain *B. subtilis* EC1524 (Widdick et al., 2003) was used to perform overlay assays to detect the production of kyamicin.

None of the KY strains had activity against *B. subtilis* in preliminary tests, contrary to what would be expected for the production of kyamicin. These tests were based on growing the strain in a variety of production media (Chapter 2.3.1) and performing disk diffusion bioassays. These production media are commonly used in industry as screening media for the identification of natural products from actinomycetes. They were brought to our lab by Prof. Barrie Wilkinson from GlaxoWellcome, and they are described as SM media in Chapter 2. The selection of these 13 media, out of a larger group of media available, was made based on differences in composition, pH and additives. The aim of this selection was to vary as much as possible the potential metabolic profile of the strain.

### 3.5.2 Design and construction of synthetic operons for kyamicin activation

In order to activate the transcription of the kyamicin BGC, we designed and ordered two synthetic artificial operons of *kyaL*, the PE-methyl transferase that provides resistance and *kyaR1*, the *Streptomyces* antibiotic regulatory protein (SARP). SARPs are confined to and have been studied largely in streptomycetes. Their sole role appears to be the regulation of antibiotic production (stationary phase onset) (Wietzorrek and Bibb, 1997). They act as transcriptional activators and appear to recognise hexameric repeats every 11 bp. SARPs are also activators of their dependent promoters (Barreales et al., 2018).

This cloning strategy was based on the reported role of these two genes at the onset of cinnamycin biosynthesis (O'Rourke et al., 2017). The two constructs differed in the order of the genes (*kyaR1-kyaL* or *kyaL-kyaR1*), but both had the same short intergenic region carrying a *kyaN* ribosome binding site (RBS) separating the genes (Figure 3.7). The reason why the *kyaN* ribosome binding site was chosen to be part of the design, is because it is an almost perfect RBS. The synthetic constructs were received as pUC57 based vectors with pUC57/pR1L, named pEVK1 and pUC57/pLR1, named pEVK2.

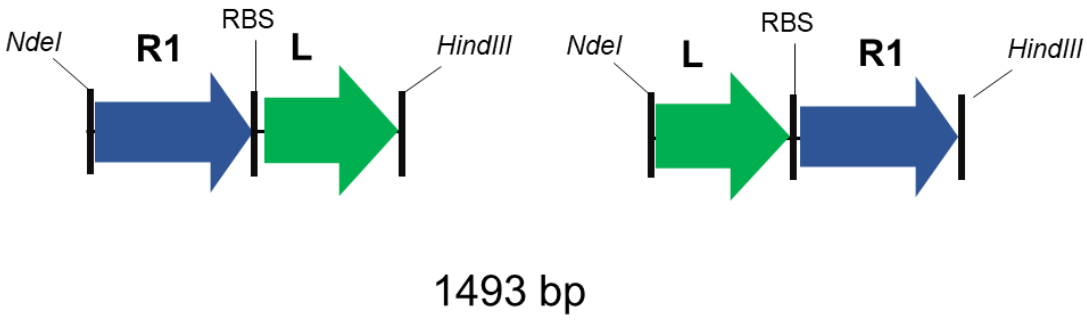



**Figure 3.7: Schematic of the two synthetic artificial operons of *kyaL*, the PE-methyl transferase that provides resistance, and *kyaR1*, the *Streptomyces* antibiotic regulatory protein (SARP)- the homologs of *cinR1* and *cinL*.**

We hypothesised that by sub-cloning these genes into vectors with constitutive promoters we could bypass the natural regulation of the SARP, therefore constitutively expressing the SARP and the resistance gene.

Each of the two synthetic operons were cloned into the *φ*BT1-based integrative expression vector pGP9 as a *NdeI/HindIII* fragment under the control of the ActII-ORF4/PactI activator/promoter system (Kuščer et al., 2007) to express it constitutively in all three *Saccharopolyspora* KY strains. This activator/promoter system is commonly used in *Saccharopolyspora* strains to drive the expression of genes, in a constitutive way

(Rowe et al., 1998). The plasmid pGP9/R1L was named pEVK4 and the plasmid pGP9/LR1 was named pEVK5.

The resulting plasmids were used to transform *E. coli* ET12567/pUZ8002, a *dam/dcm*-minus strain used in interspecific conjugations with actinomycetes, and subsequently mated into *Saccharopolyspora* KY strains.

### 3.5.3 Activation of kyamicin BGC in *Saccharopolyspora* KY21

Only in the strains carrying the constructs described in section 3.5.2, a zone of clearing of *Bacillus subtilis* was observed, regardless of the arrangement of the operon. The wild-type and the strain carrying the empty pGP9 vector showed no inhibition (Figure 3.8).

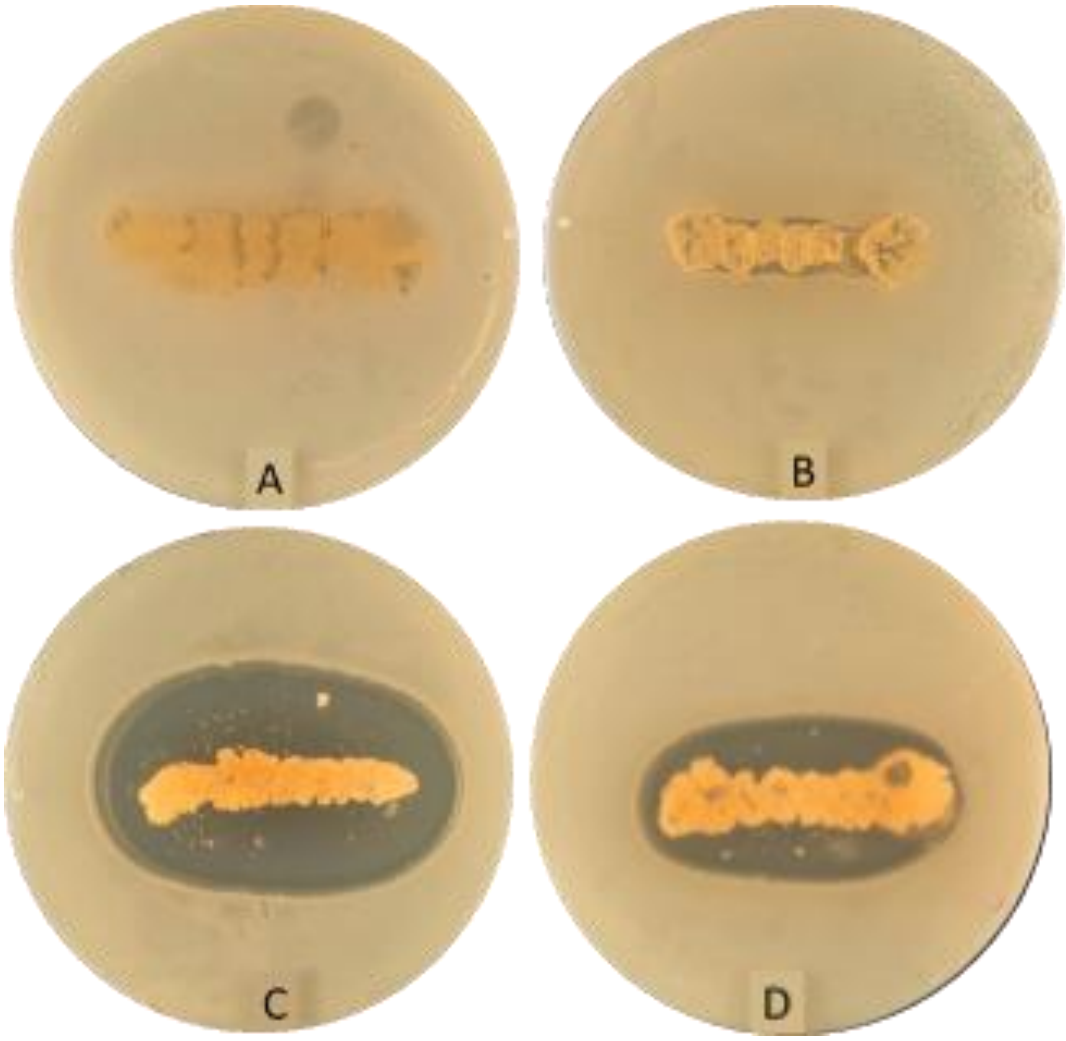



**Figure 3.8: Activation of kyamicin cluster in *Saccharopolyspora* KY21.** Undersides of plates streaked with *Saccharopolyspora* strains overlaid with *B. subtilis* EC1524. Lantibiotic production is indicated by a zone of inhibition of *B. subtilis* EC1524 around the culture - A: KY21 B:KY21/pGP9 C: KY21/pEVK4 D: KY21/ pEVK5.

To confirm the production of kyamicin, agar plugs were taken from the bioassay plates, (from the clearing zone of positive plates and close to the streak from negative plates) extracted and analysed by UHPLC-MS as described in Chapter 2. According to the expected molecular formula of kyamicin, we anticipated a monoisotopic $[M + H]^+$ *m/z* of 1797.7029 and an $[M + 2H]^{2+}$ *m/z* of 899.3551. We observed a peak of 899.3556 (Δ = 0.5560 ppm) corresponding to the doubly charged molecule in either order of the genes in the activation cassette (Figure 3.9). This peak was not present in the negative controls.

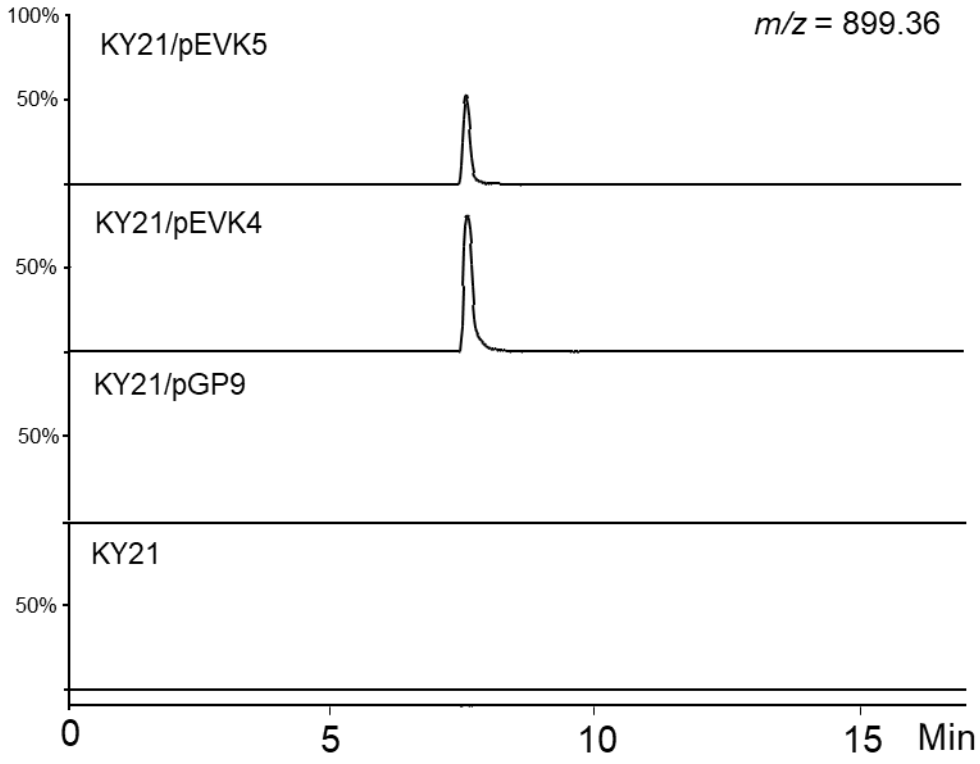



**Figure 3.9: Extracted Ion Chromatograms (XIC) of extracts from KY21 kyamicin bioassay plates.** Agar plugs were taken adjacent to the *Saccharopolyspora* streak, extracted with 5% formic acid and analysed by UHPLC-MS (ESI). A peak corresponding to the kyamicin $[M + 2H]^{2+}$ of 899.36 (*m/z* observed = 899.3556; *m/z* calculated = of 899.3551; Δ = 0.5560 ppm) was detected only in KY21/pEVK4 and KY21/pEVK5. XICs are scaled such that 100% = a signal intensity of 20,000,000.

In order to purify kyamicin for further analysis and characterisation scale-up of *Saccharopolyspora sp*. KY21 cultures, in solid SF+M or in liquid SV2, was attempted but the material collected was not sufficient to be isolated for further study. Further details of the isolation of kyamicin are reported in section 3.8.

### 3.5.4 Activation of kyamicin BGC in *Saccharopolyspora* KY3 and KY7

In order to determine whether the same constructs carrying the activation cassettes, described above and shown in Figure 3.7, plasmids pEVK1 and pEVK2 were separately mated in the *Saccharopolyspora* KY3 and KY7 to test if the kyamicin cluster could be activated in the other KY strains as well. In the case of KY3, the results of the inhibition of *B. subtilis* were very similar to the results of KY21, with the only difference of a slightly smaller inhibition zone. This is possibly due to KY3 exhibiting less growth than KY21 (Figure 3.10).

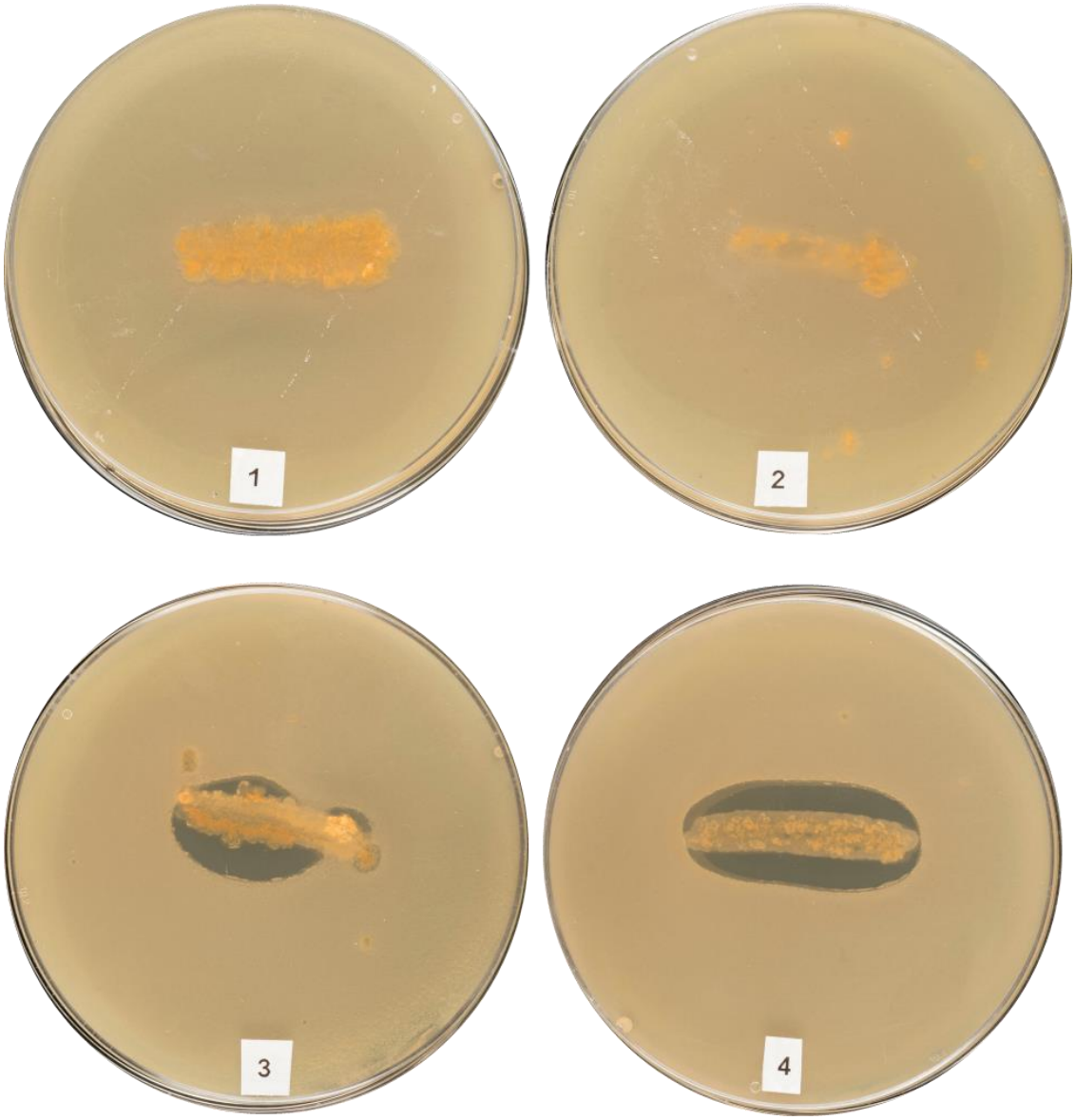



**Figure 3.10: Activation of kyamicin cluster in *Saccharopolyspora* KY3** - Undersides of plates streaked with various *Saccharopolyspora KY3* overlaid with *B. subtilis* EC1524.

Lantibiotic production is indicated by a zone of inhibition of *B. subtilis* EC1524 around the culture- 1: KY3 2: KY3pGP9 3: KY3/pEVK4 4: KY3/pEVK5.

The KY7 strain also grows poorly on agar plates, compared to KY21. In this case, intriguingly, only the strain carrying the synthetic operon with *kyaR1* first and *kyaL* second showed inhibition against *B.subtilis* (Figure 3.11). HPLC-MS analysis was carried out in a similar way as above and the presence of kyamicin was confirmed in the inhibition zones and not in the negative samples.

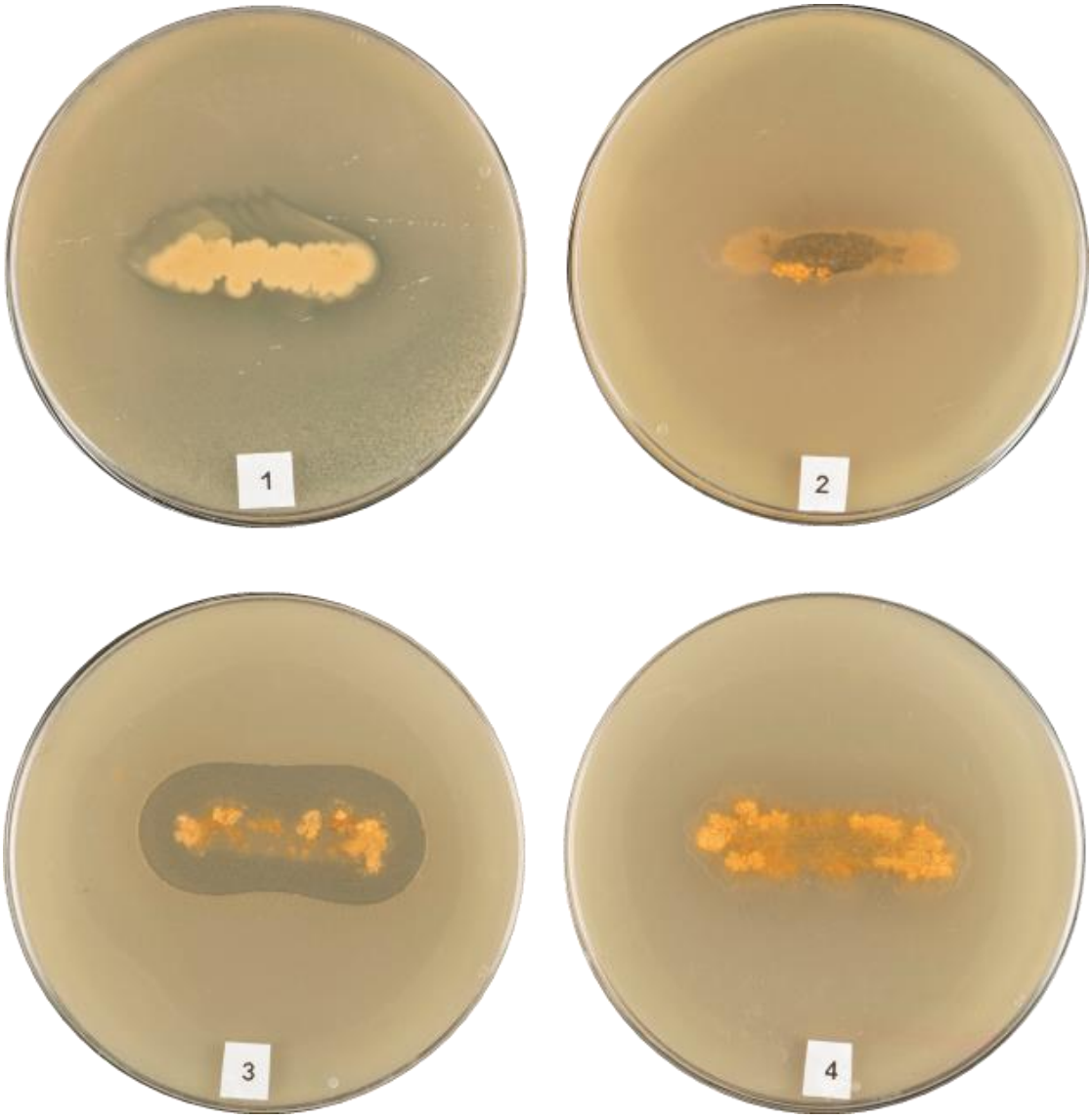



**Figure 3.11: Activation of kyamicin cluster in *Saccharopolyspora* KY7.** - Undersides of plates streaked with various *Saccharopolyspora KY7* overlaid with *B. subtilis* EC1524 Lantibiotic production is indicated by a zone of inhibition of *B. subtilis* EC1524 around the culture. 1: KY7 2:KY7/pGP9 3: KY7/pEVK4 4: KY7/pEVK5.

## 3.6 Heterologous expression of kyamicin BGC

Having successfully activated the kyamicin BGC in the *Saccharopolyspor*a KY strains, the next objective was to express it in a different host. *Streptomyces coelicolor* M1152 was chosen as a widely used superhost for expression of actinomycete genes (Gomez-Escribano and Bibb, 2011, Widdick et al., 2018).

### 3.6.1 Design and construction of synthetic operons for kyamicin expression

In this part of the study, we collaborated with Dr David Widdick, in the design and construction of vectors, hence the naming of some plasmids as pWDW (referring to D. Widdick) in contrast to pEVK (referring to E. Vikeli).

In order to achieve the heterologous expression of the kyamicin BGC in M1152, we designed a synthetic operon encoding the genes *kyaN* to *kyaH* as an *Eco*RI/*Xba*I fragment as shown in Figure 3.12. This was purchased from GenScript in a pUC57-based plasmid and named pEVK3. This contains the whole of the intergenic region between *kyaR* and *kyaN* so as to include the SARP binding site (SARP BS). As mentioned above the SARP BS is a binding motif of five to six base pairs, repeated every 11 base pairs, which is a characteristic of SARPs (Wietzorrek and Bibb, 1997). A comparison of the SARP BSs of kyamicin, cinnamycin and duramycin is presented in Chapter 4. Transcription is initiated by the cognate SARP, in this case, *kyaR1*, from the *kyaN* promoter on the 3'- side of the binding site. It also contains the intergenic region downstream of *kyaH*, in case there is a transcription terminator in that region. The synthetic operon was cloned into the *φ*C31, integrative, conjugative vector pSET152 (Kieser et al., 2000) that confers apramycin resistance, yielding pWDW63. This is orthogonal to the activation cassette plasmids pEVK6 (pIJ10257/R1L) and pEVK7 (pIJ10257/LR1). These were constructed based on the two GenScript synthetic artificial operons of *kyaL* and *kyaR1* mentioned in section 3.4.2. They were cloned as *NdeI/HindIII* fragments into pIJ10257 a *φ*BT1 based integrative, conjugative, expression vector with a hygromycin resistance marker ,and an *ermE** promoter (Bibb et al., 1985) to express the constructs constitutively in M1152.


**Figure 3.12: Schematic of synthetic operon carrying genes *kyaN* to *kyaH*** as an *Eco*RI/*Xba*I fragment. These genes were expected to be essential for the biosynthesis of kyamicin.

### 3.6.2 Expression of kyamicin BGC in *Streptomyces coelicolor* M1152

The resulting plasmids described in section 3.6.1 were used to transform *E. coli* ET12567/pUZ8002 individually and subsequently mated one after the other into *S. coelicolor* M1152. Several apramycin/hygromycin resistant ex-conjugants were selected. When transferred onto bioassay plates and overlaid with *B. subtilis* EC1524, a zone of clearing was observed only in the strain carrying both the BGC and the activator cassette with *kyaR1* first and *kyaL* second, pEVK6 (Figure 3.13).

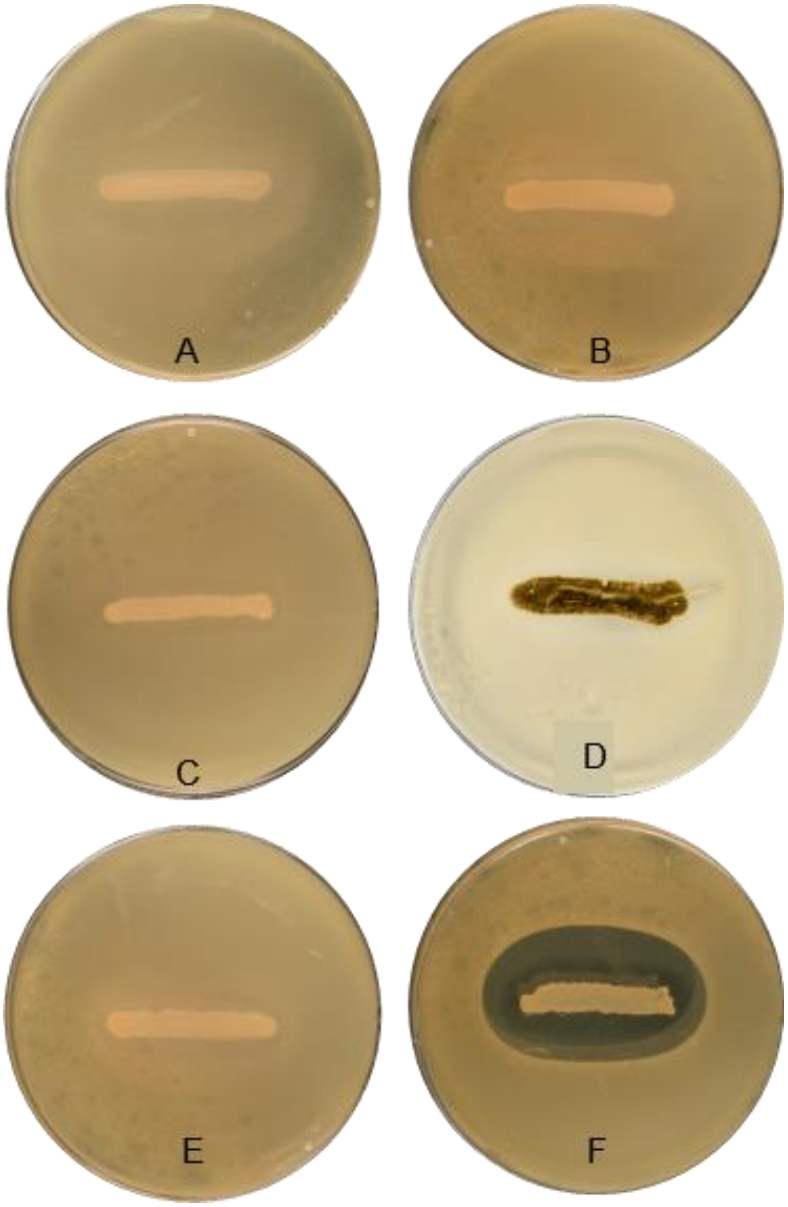



**Figure 3.13: Undersides of plates streaked with various *Streptomyces coelicolor* strains overlaid with *B. subtilis* EC1524.** Lantibiotic production is indicated by a zone of inhibition of *B. subtilis* EC1524 around the culture. A: M1152, B: M1152/pWDW63, C: M1152/pEVK7, D: M1152/pEVK6, E: M1152/pWDW63/pEVK7, F: M1152/pWDW63/pEVK6.

Agar plugs were taken from the zone of clearing or close to the streak in the absense of clearing, extracted and analysed as before by UHPLC-MS(ESI). This showed the expected $[M + 2H]^{2+}$ ion for kyamicin only in the strain carrying both the BGC and the activator cassette with *kyaR1* first and *kyaL* second, pEVK6 (*m/z* observed = 899.3547; *m/z* calculated = of 899.3551; Δ = -0.4448 ppm). Kyamicin was not present in the relevant controls (Figure 3.14).

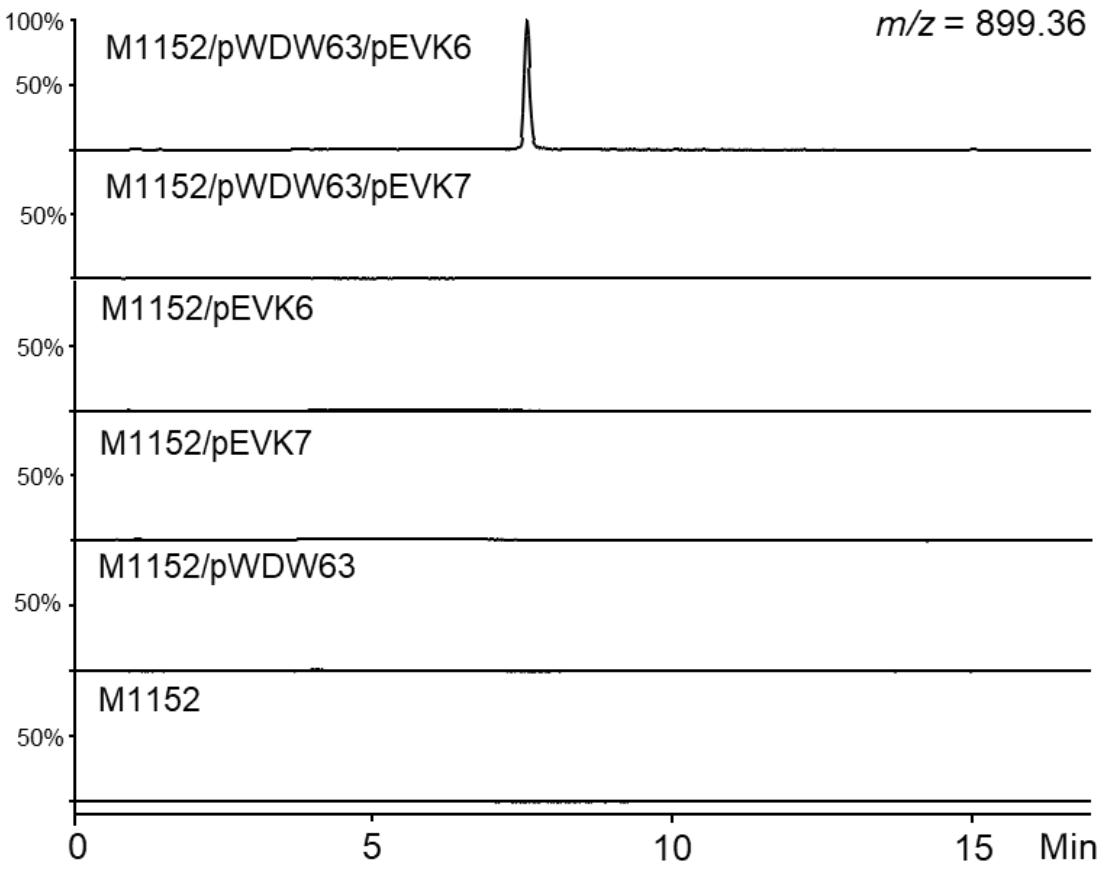



**Figure 3.14: Extracted Ion Chromatograms (XIC) of extracts from M1152 kyamicin bioassay plates.** Agar plugs were taken adjacent to the *Saccharopolyspora* streak, extracted with 5% formic acid and analysed by UHPLC-MS (ESI). A peak corresponding to the kyamicin $[M + 2H]^{2+}$ of 899.36 (*m/z* observed = 899.3547; *m/z* calculated = of 899.3551; Δ = -0.4448 ppm) was detected only in M1152/pWDW63/pEVK6 with *kyaR1* first and *kyaL* second. XICs are scaled such that 100% = a signal intensity of 1,500,000.

### 3.6.3 Expression of kyamicin BGC in *Saccharopolyspora erythraea* superhost

To further investigate the possibility of scaling up the production of kyamicin, I also cloned both of these vectors (pWDW63 and pEVK6) in an industrially optimised version of *Saccharopolyspora erythraea,* kindly provided by Isomerase Therapeutics (Cambridge, UK).

This strain is missing the erythromycin BGC and it has been modified to carry an artificial construct with four phage integration sites that allow simultaneous insertion of four orthogonal vectors. It also has high conjugation efficiency and all of the above characteristics make it an ideal strain for large scale fermentation of natural products.

Based on the previous observations, I only proceeded with the cloning of vector pEVK6 that has the R1L order. The colonies of the conjugation had either a red or a white phenotype, which has been reported previously in the literature for *Saccharopolyspora* strains (Cortés et al., 2002). The conjugation resulted in only 2 colonies, one white and one red, so I selected one colony of each phenotype to perform the bioassay.

As seen previously, only the *Saccharopolyspora* strain that carried both constructs showed inhibition against *B. subtilis* EC1524 (Figure 3.15). Future work will investigate the production of kyamicin from this superhost further.

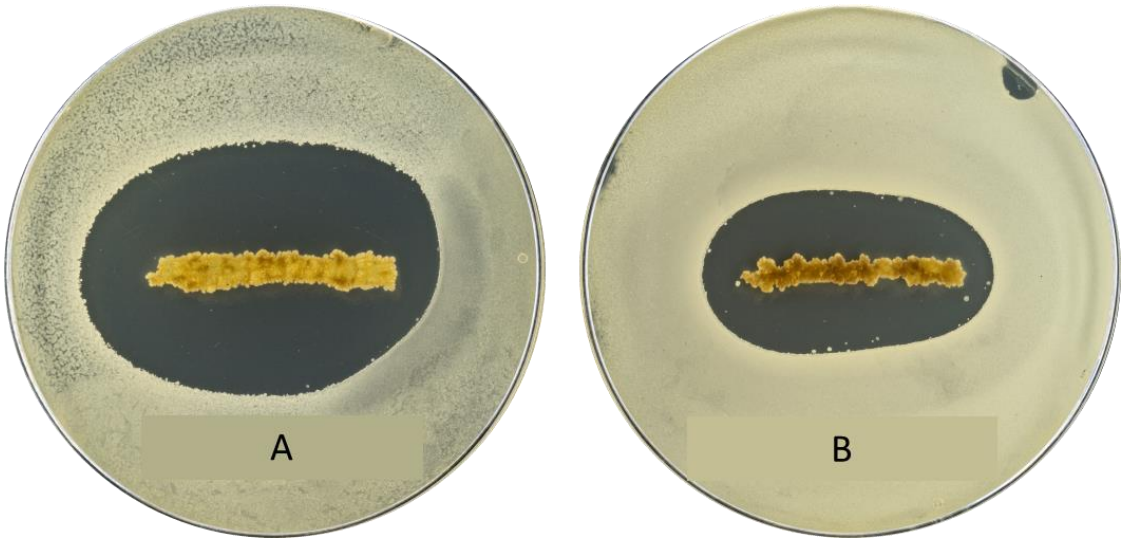



**Figure 3.15: *Saccharopolyspora erythraea* superhost carrying the kyamicin BGC in an overlay bioassay against *B. subtilis* EC1524.**
A. *Saccharopolyspora erythraea*/pWDW63/pEVK6 red phenotype colony
B. *Saccharopolyspora erythraea*/pWDW63/pEVK6 white phenotype colony.

## 3.7 Disruption and complementation of kyamicin BGC

To verify the role of *kyaA* in the kyamicin pathway we deleted it and attempted to complement it in the heterologous host. This would also enable us to investigate whether we can replace *kyaA* with other structural analogues, which could lead to the production of other type B lantibiotics.

### 3.7.1 Disruption of kyamicin BGC and abolishment of production

*lanA* genes are the structural genes that encode the precursor peptides of lantibiotics. Deletion of these genes abolishes lantibiotic production (Widdick et al., 2003). For that reason, deletion of *kyaA* was performed by the REDIRECT method of Gust et al. (2002).

### 3.7.2 Construction of a *kyaA* deletion plasmid

Dr David Widdick performed the construction of the deletion plasmids described below.

In order to achieve the deletion of *kyaA,* the original pUC57 based construct from GenScript, pEVK3, that carries the *bla* resistance gene was used to transform *E. coli* BW25113/pIJ790. This is a strain that carries pIJ790 a plasmid with the arabinose-inducible recombinase genes from the lambda phage. Two PCR primers, kyaAdelF and kyaAdelR, were used to generate a replacement cassette using pIJ10700 as a template. This generated a replacement cassette with a hygromycin resistance gene, flanked with FLP recombinase recombination sites. This was used to replace *kyaA* in pEVK3. The kyaAdelF and kyaAdelR primers possess 39 base pair 5' extensions that correspond to either 39 base pairs upstream of and including the start codon of *kyaA* in the case of kyaAdelF or 39 base pairs downstream of and including the stop codon of *kyaA* in the case of kyaAdelR. These 5' extensions were used by the lambda phage recombinase to replace the *kyaA* gene with the PCR generated hygromycin cassette via a double crossover. The product of this process, pWDW55, was then used to transform DH5αBT340. This strain has an FLP recombinase gene carried on BT340, a plasmid with a temperature sensitive replicon. Transformants were then grown for 16-18 hours in 10 mL of LB at 42 °C without any antibiotic selection. Next day, a plasmid extraction was carried out from those cultures and the plasmids were used to transform DH5α. Transformants were then selected for carbenicillin resistance and hygromycin sensitivity. This resulted in the plasmid pWDW59 which carries a scar in place of *kyaA*. This was confirmed by sequencing using primers deltaKyaAseqF and deltaKyaAseqR. A 4.6 kb *EcoR*I/*Kpn*I fragment of pWDW59 covering the 5' end of the biosynthetic cluster was cloned into the 8 kb vector fragment of *EcoR*I/*Kpn*I cut pWDW63 to reconstitute the BGC but with *kya*A replaced with a scar, called pWDW65.

### 3.7.3 *KyaA* is essential for the production of kyamicin

*E. coli* ET12567/pUZ8002 was transformed with plasmid pWDW65, the *kyaA* deletion construct, and then pWDW65 was mated into *S. coelicolor* M1152 carrying the activator

cassette, pEVK6. The absence of kyamicin production was confirmed by the lack of *B.subtilis* inhibition in bioassays (Figure 3.16).

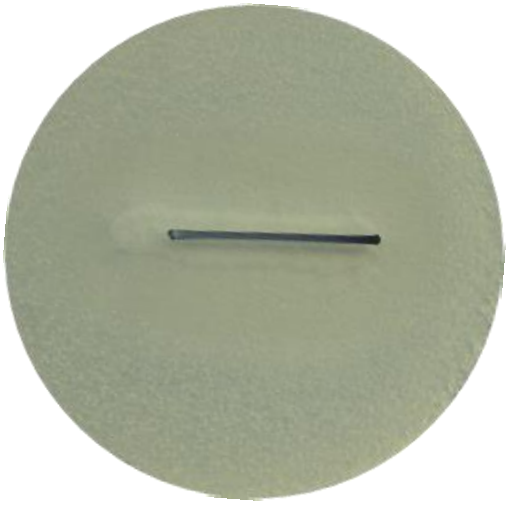


**Figure 3.16: Underside of plate streaked with *Streptomyces coelicolor* M1152/pWDW65/pEVK6 strain overlaid with *B. subtilis* EC1524.** Kyamicin production is abolished, as shown by the lack of a zone of inhibition of *B. subtilis* EC1524 around the culture.

### 3.7.4 Attempts to complement Δ*kyaA* in trans

#### 3.7.4.1 Complementation of Δ*kyaA* with an *erm*E* promoter plasmid

To further prove the necessity of *kyaA* to the production of kyamicin, attempts were made to complement the deletion described in section 3.6.2 and 3.6.3.

A PCR product was generated that covered the whole coding region of the *kya*A gene. That was digested by *Nde*I/*Xba*I and cloned into pADW11 also cut with *Nde*I/*Xba*I, under the control of an *erm*E* promoter. The vector pADW11 is a pOSV556t based vector (Sydor and Challis, 2012) that has a pSAM2 integrase (Raynal et al., 2002) and carries a thiostrepton resistance gene for selection in *S. coelicolor* and a carbenicillin resistance gene for selection in *E. coli*. This resulted in plasmid pWDW61.

When this plasmid was introduced into *S. coelicolor* M1152pWDW65pEVK6, the expected inhibition of *B. subtilis* was not observed. Agar plugs were taken and extracted as before and no kyamicin was detected. pWDW61 did not complement the Δ*kyaA* of pWDW65. A reason for absence of kyamicin could be that pWDW65 is not producing enough transcript to make enough *kyaA* for kyamicin to be detected in bioassays or that

the Δ*kyaA* mutation is having a polar effect on the transcription of the remaining downstream *kya* biosynthetic genes.

### 3.7.4.2 Reconstitution of Δ*kyaA* with a *kyaR1* promoter plasmid

Based on the negative results of section 3.7.4.1, we cloned the same PCR fragment encoding *kyaA* used previously, in pADW19, aiming to produce more *kyaA* transcript and complement the Δ*kyaA* phenotype. pADW19 was built by Dr David Widdick the same way as pADW11 with the exception that it has an *ermE**p::*kyaR1* promoter and a *kyaN* promoter. In pADW19 then, we introduced an *Nde*I/*Eco*RI fragment from pWDW61 carrying the *kyaA* and thiostrepton resistance genes, under the control of *kyaN* promoter, which is activated by the mentioned *ermE**p::*kyaR1* gene. This resulted in plasmid pWDW66.

When this plasmid was introduced into *S. coelicolor* M1152/pWDW65/pEVK6, again the expected inhibition of *B. subtilis* was not observed. Agar plugs were taken and extracted but no kyamicin was detected. For this reason, we changed approach and tried complementing a partial Δ*kyaA* in cis.

### 3.7.5 Successful reconstitution of partial Δ*kyaA* in cis

A diagram of this experiment can be seen in Figure 3.17.

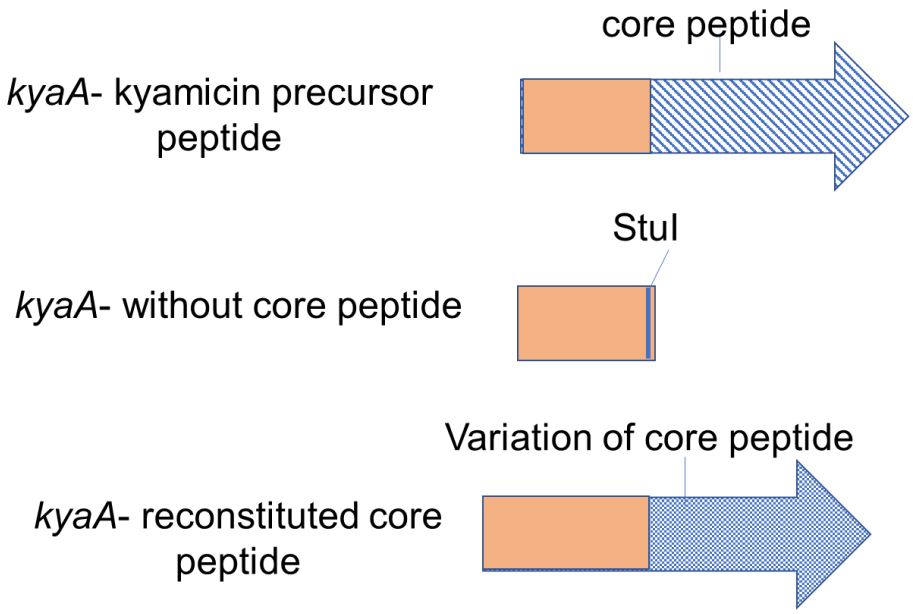



**Figure 3.17: Diagram of experimental design for partial complementation of Δ*kyaA* and introduction of other type B lantibiotic core peptides.**

A partial deletion of *kyaA* in which the core peptide coding region of *kyaA* was replaced with a *Stu*I restriction site was constructed. A PCR was generated covering from an *Eco*RI site introduced to the first base pair of the codon Val 59 of *kyaA*. A *Stu*I site was created using the AGG sequence from the last two base pairs of the E58 codon and the first base pair from the V59 codon and added CCT to give AGGCCT, the *Stu*I site. Six random base pairs were inserted between the *Stu*I site and a *Kpn*I site located at the end of the PCR fragment. This was cut with *Eco*RI/*Kpn*I and cloned into *Eco*RI/*Kpn*I cut pBlueScriptIIKS to yield pWDW68. A further PCR was generated with a *Stu*I site at the end of the PCR corresponding to where the core peptide coding region and stop codon of *kyaA* would have been and with the *Kpn*I site located in the *kya*X gene. This PCR was cut with *Stu*I/*Kpn*I and cloned into *Stu*I/*Kpn*I cut pWDW68, to give pWDW69. The *Eco*RI/*Kpn*I fragment from this PCR was then subcloned into the 8 kb fragment of *Eco*RI/*Kpn*I cut pWDW63, to give pWDW70. This construct has a *Stu*I site were the core peptide coding region had been. A synthetic double-stranded oligo encoding for the kyamicin core peptide was ordered and cloned into pWDW70 via GIBSON (Gibson, 2010) and mated into M1152/pEVK6 along with empty pWDW70 as negative control. The pWDW70 carrying the synthetic kyamicin core peptide was named pEVK10.

As a result, the strain carrying the activator cassette and the synthetic kyamicin core peptide with the remainder of the leader, complemented the mutation of Δ*kyaA* and the production of kyamicin was restored (Figure 3.18 and Figure 3.19).

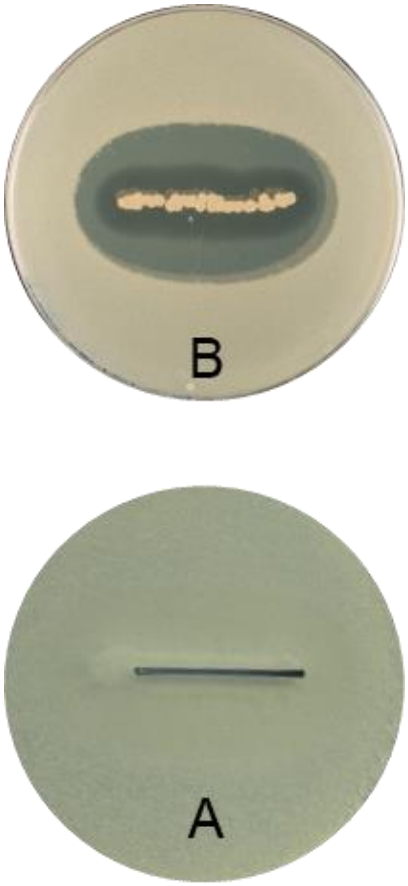



**Figure 3.18: Undersides of plates streaked with strains of *S. coelicolor* M1152 overlaid with *B. subtilis* EC1524 after complementing Δ*kyaA*.** A. M1152/pEVK6/pWDW70, B. M1152/pEVK6/pEVK10. Inhibition was observed only for the strain carrying the activator cassette and the vector carrying the synthetic kyamicin core peptide.

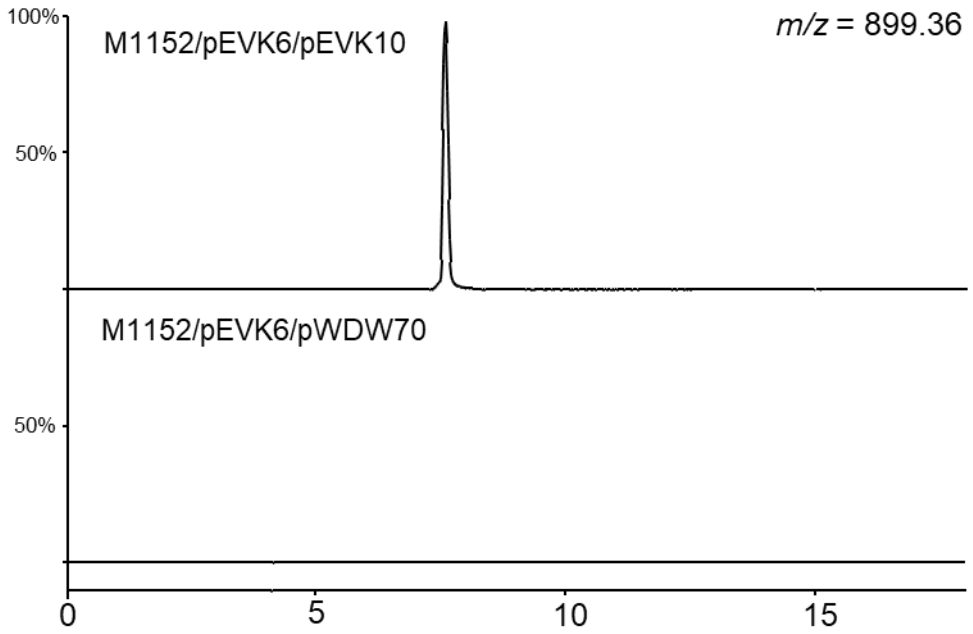



**Figure 3.19: Extracted ion chromatograms (XICs) of extracts from kyamicin complementation bioassay plates.** Agar plugs from the plates in Figure 3.18 were extracted as described previously and analysed by UHPLC-MS (ESI). A peak corresponding to the kyamicin $[M + 2H]^{2+}$ of 899.36 (*m/z* observed = 899.3556; *m/z* calculated = of 899.3551; Δ = 0.5560 ppm) was observed only in M1152/pEVK6/pEVK10. XICs are scaled such that 100% = a signal intensity of 10,000,000.

## 3.8 Scale-up and isolation of kyamicin

In order to fully characterise the kyamicin, it was necessary to scale-up the production to isolate sufficient quantities of the molecule. As mentioned above, scale-up of *Saccharopolyspora sp*. KY21 cultures, in solid SF+M or in liquid SV2, did not give enough material for further study.

In collaboration with Dr Sibyl Batey, we tested the kyamicin production of *S. coelicolor* M1152/pWDW63/pEVK6 in small volumes of tryptic soy broth (TSB) liquid media, which *S.coelicolor* is known to grow very well in (Gomez-Escribano and Bibb, 2011). Kyamicin was being produced in detectable amounts in small-scale trials, so we proceeded with the scale-up growth of the heterologous host strain.

### 3.8.1 Purification of kyamicin

*S. coelicolor* M1152/pWDW63/pEVK6, was grown in 12 flasks of size 2500 mL with 500 mL TSB in each, (giving a total of 6 L) at 28˚C for seven days. After that, the cells were harvested and extracted with 500 mL of methanol/water in ratio 1:1 by sonicating for two hours and shaking for 16 to 18 h. After centrifugation, the cell pellet was discarded, and the supernatant was filtered and concentrated. The resulting residue was re-dissolved and passed-through a SPE column and eluted with a gradient of water, to remove polar compounds such as sugars. The fractions that contained kyamicin were combined and concentrated before further purification. The resulting extract was subjected to two repetitions of semi-preparative HPLC to yield 2.5 mg of pure kyamicin.

### 3.8.2 Chemical characterisation of kyamicin

The chemical characterisation of kyamicin was performed with the aid of Dr Sibyl Batey.

#### 3.8.2.1 Reduction and Fragmentation

Tandem mass spectrometry (MS/MS) is commonly used to characterise peptide natural products. In the case of lantibiotics, the lanthionine bridges hamper the fragmentation impeding the characterisation of these molecules. In order to circumvent this issue, it has been reported (Martin et al., 2004, He et al., 2007) that chemical reduction of the lanthionine bridges is possible using $NaBH_4$-$NiCl_2$. A modified version of the published protocols was carried out, as described in section 2.8.3. LCMS(ESI) analysis of the product showed an $[M + 2H]^{2+}$ ion at 854.4203 *m/z*, in agreement with the calculated *m/z* of 854.4204 ($\Delta$ = -0.11704) corresponding to the loss of three sulphur atoms and gain of six hydrogen atoms, giving the molecular formula $C_{76}H_{114}N_{20}O_{25}$, Figure 3.20.

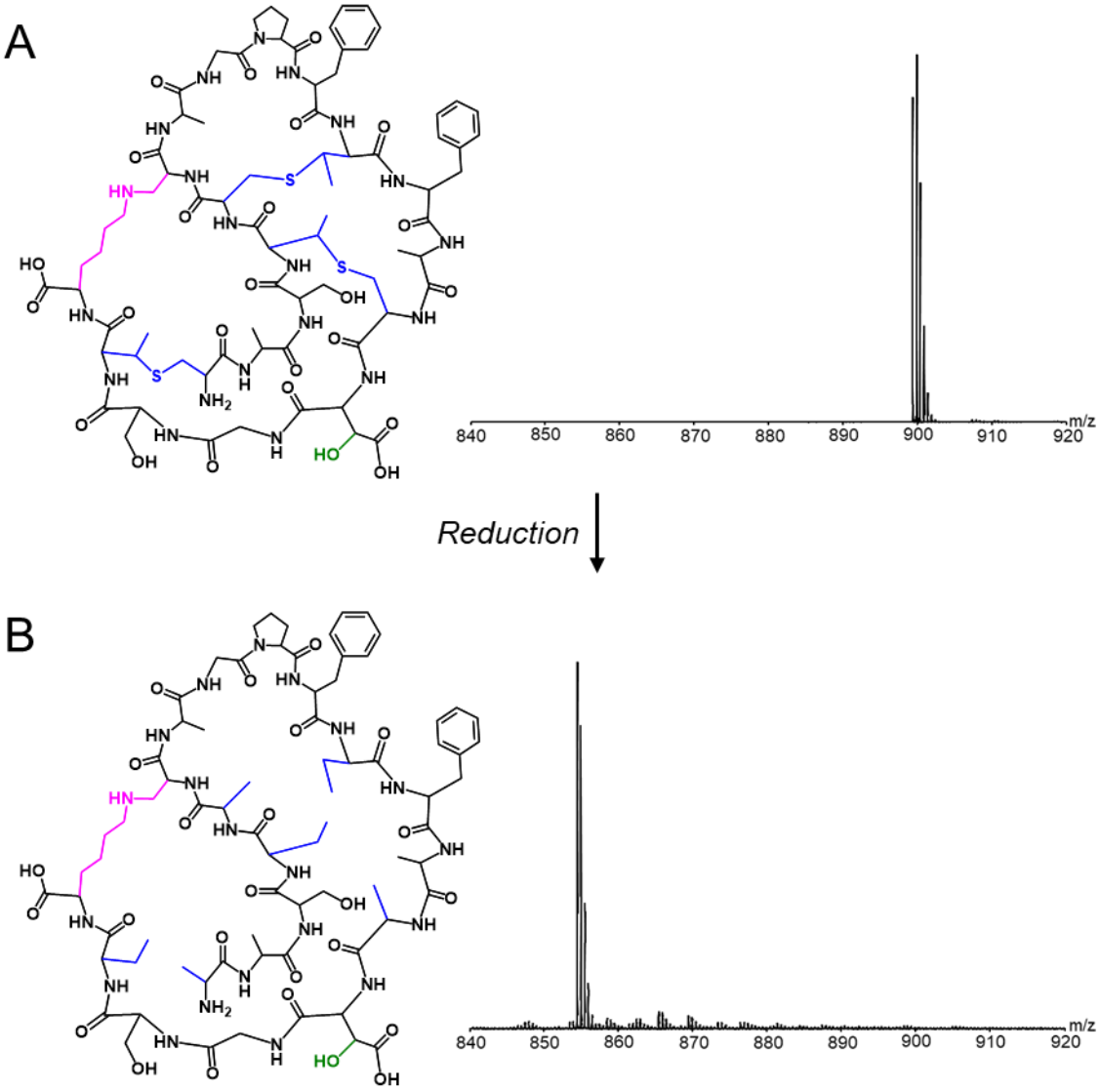



**Figure 3.20: Reduction of kyamicin.** A: Chemical structure of kyamicin corresponding to a molecular formula of $C_{76}H_{108}N_{20}O_{25}S_3$ (*m/z* calculated: 899.3551, *m/z* observed for purified kyamicin: 899.3556). B: Chemical structure of kyamicin after reduction of the three thioether bonds corresponding to the loss of three sulfur atoms and gain of six hydrogen atoms to give a molecular formula of $C_{76}H_{114}N_{20}O_{25}$; *m/z* calculated: 854.4204, *m/z* observed: 854.4203 (Δ = -0.11704). Intact and reduced thioether bridges are shown in blue; the lysinoalanine bridge is shown in pink, the hydroxylation of D15 is shown in green.

To confirm the connectivity of the peptide, the MS/MS analysis was carried out using both ESI and matrix-assisted LASER desorption ionization (MALDI) methods. The clearest spectrum was achieved with MALDI MS/MS, showing the complete ion series for the peptide, Figure 3.21. As had been suggested previously (Kodani et al., 2016), fragmentation of the lysinoalanine bridge occurred via a rearrangement resulting in a glycine at residue six and N=C double bond at the end of the K19 side chain.

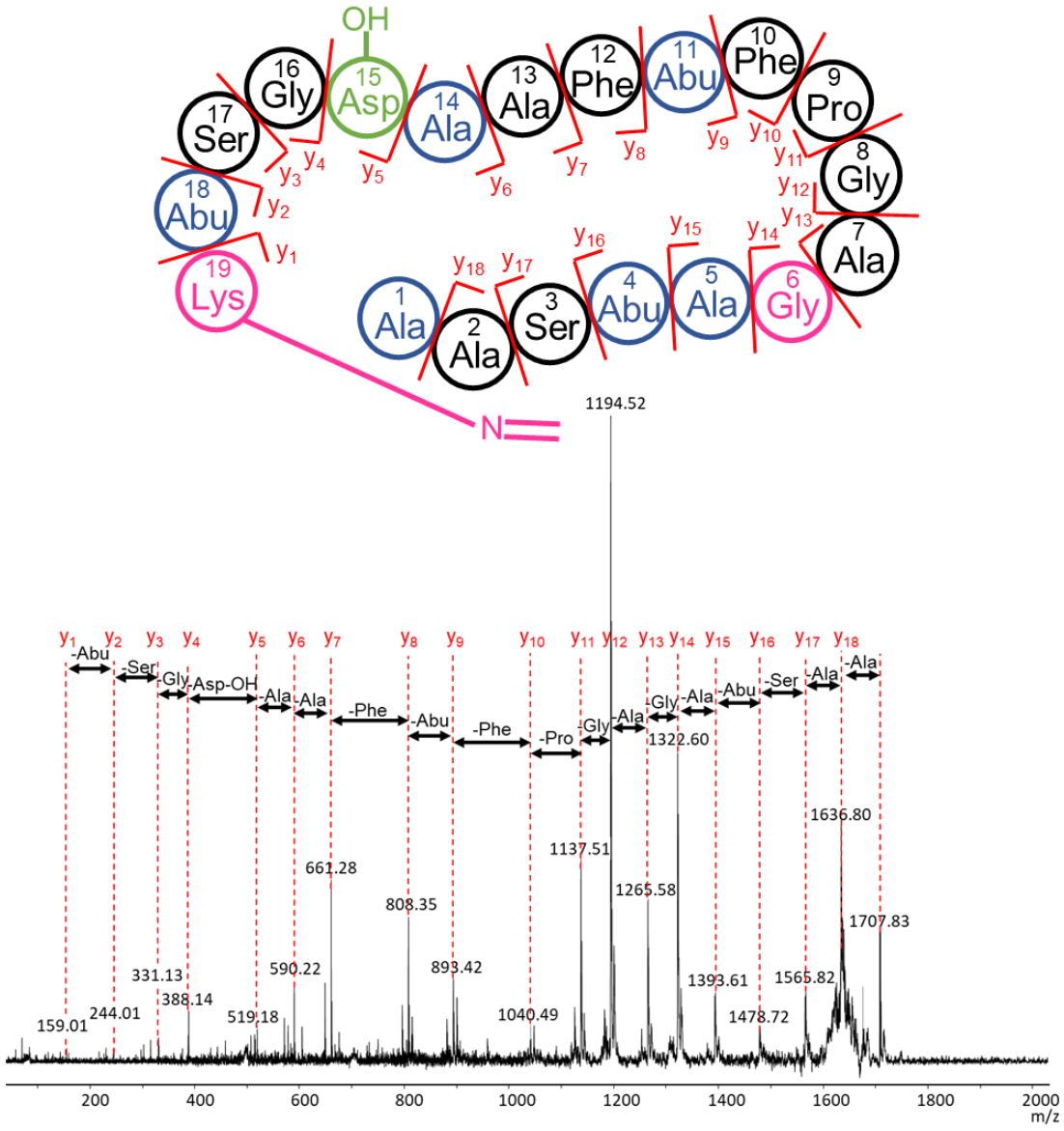



**Figure 3.21: Fragmentation of reduced kyamicin.** MALDI MS/MS fragmentation showed the y ion series of the reduced peptide. The fragmentation of the lysinoalanine bridge occurred via a rearrangement resulting in a glycine at residue 6 and N=C in the side chain of K19. The expected connectivity for the complete peptide was observed.

### 3.8.2.2 NMR

Whilst the MS fragmentation results matched the predicted structure, the gold standard of small molecule analysis is NMR as it provides information about the structural and spatial relationships between different atoms. Therefore, to confirm the purity and connectivity of kyamicin, NMR analyses were carried out by Dr. Sibyl Batey. The sample was dissolved in deuterated dimethylsulfoxide (DMSO) and a $^{1}$H NMR spectrum was recorded at 800 MHz, Figure 3.22. Two-dimensional NMR experiments were then carried out, including HSQC, TOCSY, and NOESY, which enabled the identification of proton-carbon correlations. In total, 13 of the 18 anticipated amide protons could be identified. These correspond to 13 spin systems in the TOCSY spectrum, which is consistent with what is observed for cinnamycin B (Kodani et al., 2016). These spin systems could be putatively assigned based on their spatial relationship determined from the NOESY spectrum. Coupling in the HSQC spectrum then allowed identification of several C atoms in the molecule. All the two-dimensional spectra can be found in Appendix 2.

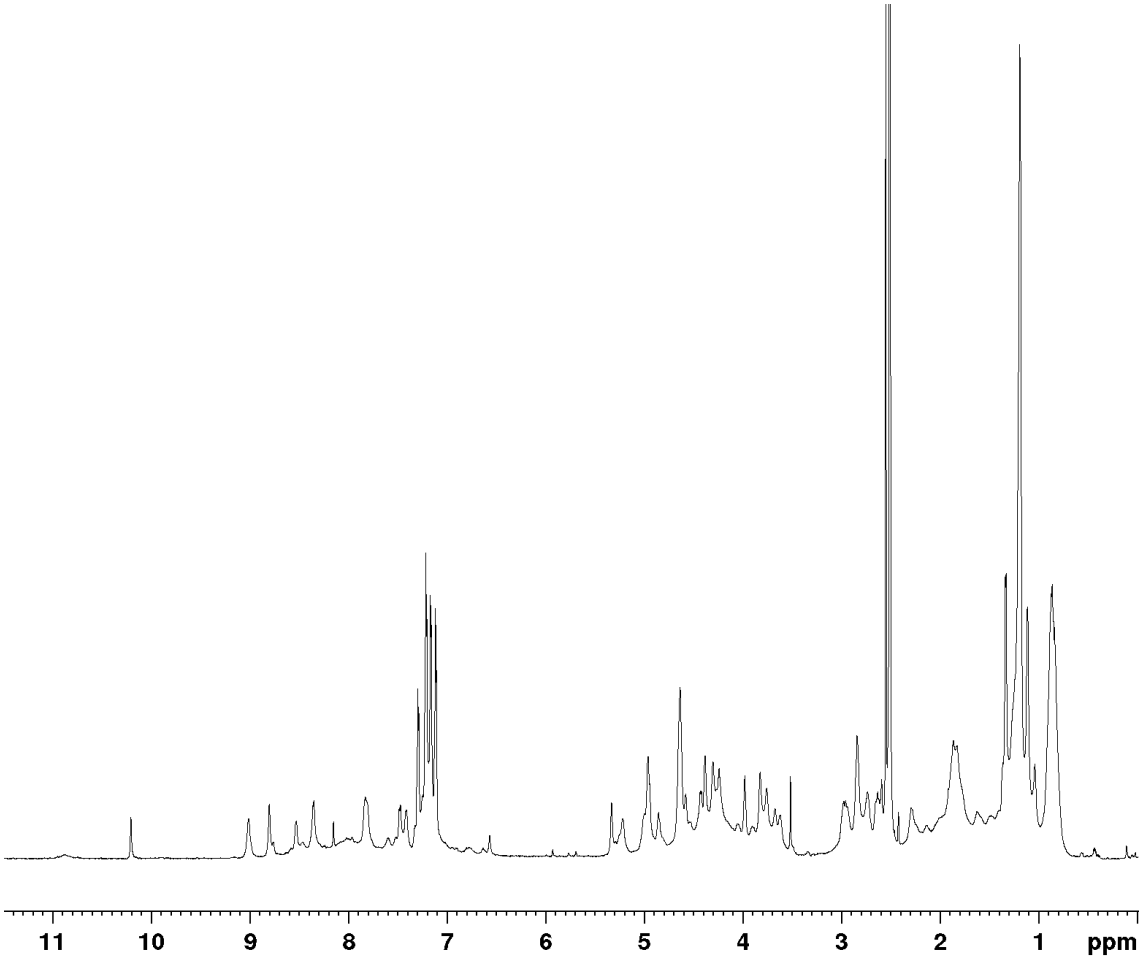



**Figure 3.22: $^{1}$H NMR spectrum of kyamicin in $d_6$-DMSO.**

## 3.9 Bioactivity of kyamicin

After isolating pure kyamicin, the activity of the compound against various bioindicator strains was compared with the commercially available compounds cinnamycin and duramycin.

### 3.9.1 Kyamicin inhibits *B.subtilis* at a Minimum Inhibitory Concentration (MIC) of 128 µg/mL

The application of pure kyamicin in preliminary disk diffusion assays showed weak inhibition and so direct application of the compound on the agar was trialled. The zone of inhibition was greater, indicating that the compound binds to the disks as described for other compounds (Scott, 2017). As shown in Figure 3.23, kyamicin inhibits *B. subtilis* EC1524 at a Minimum Inhibitory Concentration (MIC) of 128 µg/mL, whereas duramycin inhibits at 32 µg/mL and cinnamycin at 16 µg/mL.

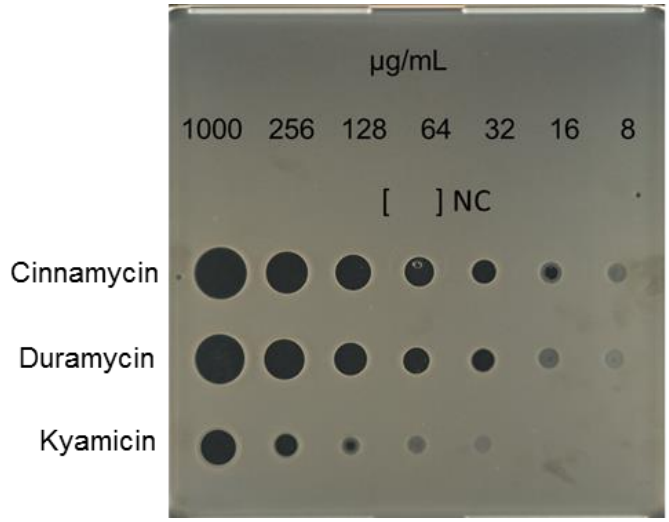



**Figure 3.23: Comparative bioassay of kyamicin, duramycin and cinnamycin against *B. subtilis* EC1524.** Direct application of serial dilutions of the compounds in water to determine the MIC of each substance. NC = $H_2O$ as a negative control.

### 3.9.2 Kyamicin inhibits a range of Gram-positive bacteria in overlay assays

*S. coelicolor* M1152/pWDW63/pEVK6 was tested in overlay bioassays against a variety of clinical bacterial strains. Inhibition was only observed against the Gram-positive indicator strains as anticipated (Figure 3.24).

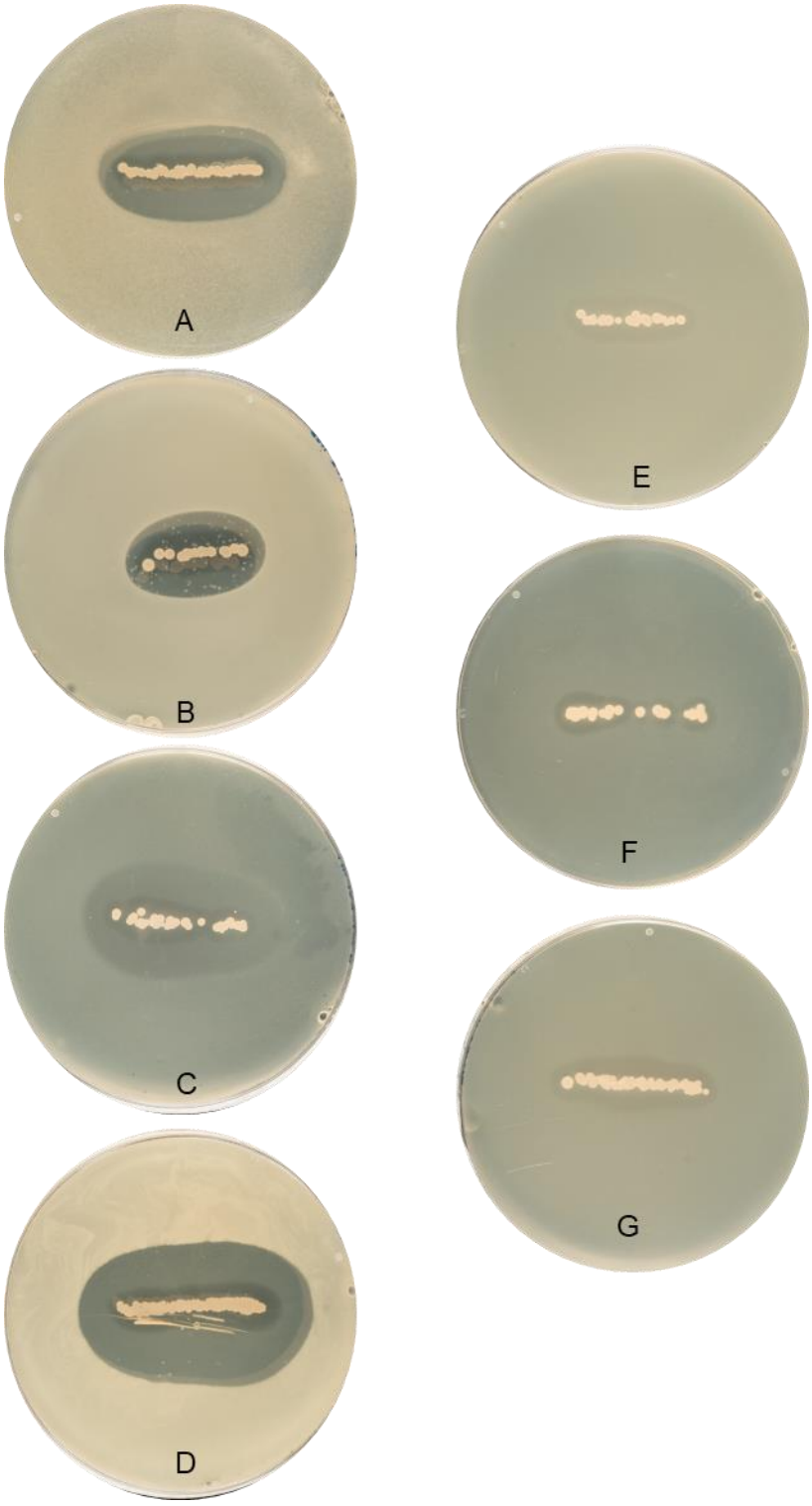



**Figure 3.24: Undersides of plates streaked with M1152pWDW63pEVK6 , overlaid with** A. *Micrococcus luteus*, B. *B. subtilis* ESKAPE, C. *B. subtilis* 168 wild-type, D. *Staphylococcus epidermidis* ESKAPE, E. Methicillin-resistant *S. aureus* (MRSA), F. Vancomycin-Resistant *Enterococcus faecium* 6295 (VRE), G. *E. coli* ATCC 25922.

### 3.9.3 Kyamicin inhibits *Streptomyces antibioticus*

Cinnamycin B is closely related lantibiotic to kyamicin with only two different amino acids in its core peptide sequence. This compound was reported to exhibit activity against *Streptomyces antibioticus* but not against *B.subtilis* (Kodani et al., 2016). The activity of kyamicin against *S. antibioticus* TU1798 was therefore investigated. The inhibition zone observed was much smaller than the equivalent zone from duramycin and cinnamycin (Figure 3.25). This fact raises interesting questions about the diversity of lantibiotic activity and how small amino acid variations impact on bioactivity.

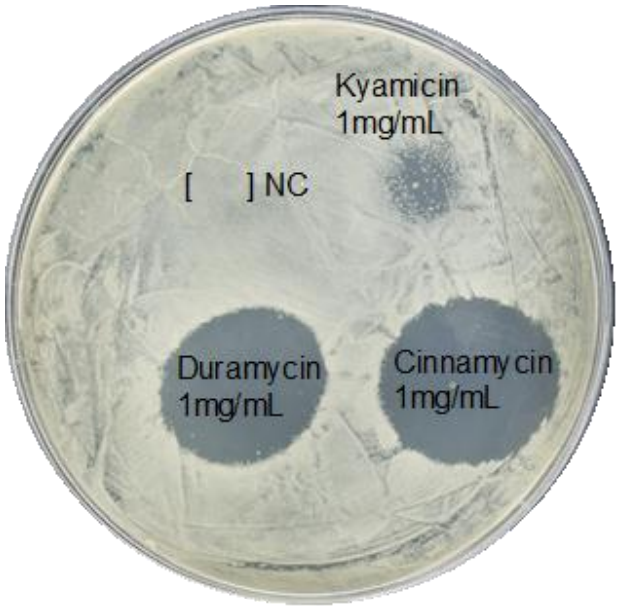



**Figure 3.25: Comparative bioassay of kyamicin, duramycin and cinnamycin against *Streptomyces antibioticus* TU1798.** Direct application on agar of 5 µL of 1 mg/mL of each compound. NC = $H_2O$ as a negative control.

## 3.10 Discussion

The first objective of this project was to study the cinnamycin-like BGC in the *Saccharopolyspora* strains KY3, KY7 and KY21. Based on the analysis of the genomes, it was shown that the three strains are quite similar and the predicted BGCs are overlapping, despite having been sampled from two different locations. This also makes the existence of the same kyamicin BGC in all three strains very intriguing and serves to illustrate the importance of investigating unusual ecological niches in the search for novel antibiotics. Kyamicin is a new addition to the lantibiotics family, a class of antibiotics that target phosphatidylethanolamine (PE) in bacterial cell walls. The importance of lantibiotics has been emphasized in recent years. For example, duramycin is in phase II of clinical trials for the treatment of cystic fibrosis (Oliynyk et al., 2010).

The second objective was to activate the kyamicin cluster in the native host. This was achieved by designing an activation cassette comprised of a SARP and a methyltransferase, which functions to provide resistance via the methylation of PE in the host membranes. This is the first time that the utilisation of the SARP and the methyltransferase has been reported to successfully switch on a silent lantibiotic cluster. The order of the genes at the operon did not matter for the activation in KY21.

When expressing the kyamicin cluster in *Streptomyces coelicolor* M1152, only the construct with the SARP first (pEVK6) was operational. There are many possible reasons for this discrepancy between the heterologous host and the native strain. One explanation is that the RBS from *kyaN* that was introduced did not function properly. Another observation that supports the theory that the resistance *kyaL* gene is not being expressed, is the fact that the strain producing M1152/pWDW63/pEVK6 doesn't grow particularly well on the bioassay plates. Future work will be done to try to express just the SARP on its own to investigate if that alone can activate kyamicin production. A different intragenic region could also be tried or the downstream genes of *kyaR1* or *kyaL* could be tagged to check their expression by using Western Blotting.

In terms of the scale up of the production of kyamicin, future work will take advantage of the successful expression of kyamicin in the *Saccharopolyspora* superhost to produce large quantities. This will be useful for further bioactivity studies or further characterisation of the kyamicin molecule, such as three dimensional structural studies.

After resolving the chemical structure of kyamicin, the bioactivity of this new lantibiotic was compared to the known representatives of the lantibiotic family. The most interesting finding was the differences in the activity of kyamicin against Gram-positive bacteria

compared to the closest in structure (only two differences) reported lantibiotic, cinnamycin B. Since the key residues for PE binding are conserved, the differences between the two molecules, are unlikely to affect how they bind PE in the membrane of the cell, but they could affect how the lantibiotics disrupt the membrane. This could explain the differences between the bioactivity of kyamicin, duramycin and cinnamycin as well. Cinnamycin might exhibit the highest activity against Gram-positive bacteria of the three because of its better ability to penetrate the cell as it is comprised of more hydrophobic amino acids.

Overall, this work showcases how genome-led chemistry and the activation of silent BGCs enables the discovery of new antibiotics.

# Chapter 4:

# Engineering a platform for expression of type B lantibiotics

# 4 Engineering a platform for expression of type B lantibiotics

## 4.1 Introduction

A variety of type B lantibiotics have been isolated, and they are most commonly produced by *Streptomyces* spp. and *Bacillus* spp. (Chatterjee et al., 2005). The rise in readily available genomic data has resulted in a plethora of published genomes encoding more putative lantibiotic BGCs, suggesting many other type B lantibiotic structures are possible. However, there is a discrepancy between this apparent genomic wealth of potential lantibiotics and the modest number that have been isolated and characterised. The main obstacles to the investigation of these predicted lantibiotics are difficulties growing the bacterial species and/or the lack of production under laboratory conditions.

Given the importance of lantibiotics as potential therapeutic agents with multiple modes of action (Shin et al., 2016), the exploration of alternative structures with altered bioactivity is highly desirable. Representatives of this family of antibiotics have made it to clinical trials, and include duramycin for cystic fibrosis (Oliynyk et al., 2010) and cinnamycin that has been suggested as an alternative treatment for atherosclerosis (Hansson, 2005, Marki et al., 1991).

In Chapter 3, I presented the discovery of a new member of the family of type B lantibiotics, kyamicin. This is a circular, 19-membered lantibiotic with activity against several Gram-positive bacteria. The process of activating the originally silent kyamicin cluster involved two different parts. Firstly, in order to awaken the silent BGC in the *Saccharopolyspora* native host, an activation cassette was engineered that comprised of *kyaL*, the PE-methyl transferase resistance gene, and a pathway-specific positive regulator, the SARP *kyaR1*.

In parallel, we engineered the machinery for the expression of kyamicin in the heterologous host *S. coelicolor* M1152, using an activator cassette (above) in combination with a vector containing the kyamicin biosynthetic genes. That resulted in the expression of kyamicin. In the same heterologous host, we deleted *kyaA*, the structural gene of kyamicin with the scope of substituting it with *lanA* structural homologues. This complementation efforts proved especially challenging. For that reason, we created a construct with a partial deletion of *kyaA* in which the core peptide coding region of *kyaA* was replaced with a *Stu*I restriction site. During our efforts to complement the knockout of *kyaA* in this partial construct, we hypothesised that this tool

might allow us to construct a platform for the heterologous expression of an array of other type B lantibiotics, including those that are new-to-nature.

## 4.2 Objectives

- Express duramycin and cinnamycin B with kyamicin-based activator cassettes as described in chapter 3.
- Identify type B lantibiotics BGCs in genomic databases based on the similarity of the core peptides sequences.
- Investigate the expression of other potential lantibiotics with kyamicin-based activator cassette.

## 4.3 Expression of duramycin BGC in *S. coelicolor* M1152

Having used the activator cassette in the heterologous expression of kyamicin in M1152, we investigated whether the same kyamicin activator cassette can be used to activate the expression of another type B lantibiotic in *S. coelicolor* M1152.

Fortuitously, a construct containing the putative duramycin biosynthetic genes had previously been constructed and was available to us (Dr. David Widdick, unpublished data). This contained the genes *durN* to *durZ* that had been identified in *Streptomyces cinnamoneus cinnamoneus* DSM 40005 and inserted into a pOJKKH vector (pOJ436 based). However, this lacked the immunity and regulatory genes *durL*, *durR1* and *durRK*. These were unintentionally omitted as the presence of a phage insertion site adjacent to *durZ* (Figure 4.1) meant they were not isolated and identified at the time of the work. The absence of the regulatory genes meant that heterologous expression of duramycin in *S. lividans* was not achieved, and this construct was not used further. The cloning of the duramycin cluster is partially described in Widdick et al, 2003.

More recently, a paper was published describing the organisation of the duramycin cluster in *Streptomyces cinnamoneus* ATCC 12686 (also known as DSM40005) (Huo et al., 2017). We compared the sequence of the duramycin BGC in their duramycin producing strain to the one described above and found a homology of 99% between the two, as well as the same phage insertion site corresponding to a drop in GC content and the presence of a recombinase/integrase adjacent to *dur*Z. Again, the duramycin regulatory/immunity genes were not included in their duramycin gene cluster, but we could identify them in a similar arrangement as seen above, analogous to kyamicin and cinnamycin, at a remote location in the genome (Figure 4.1).

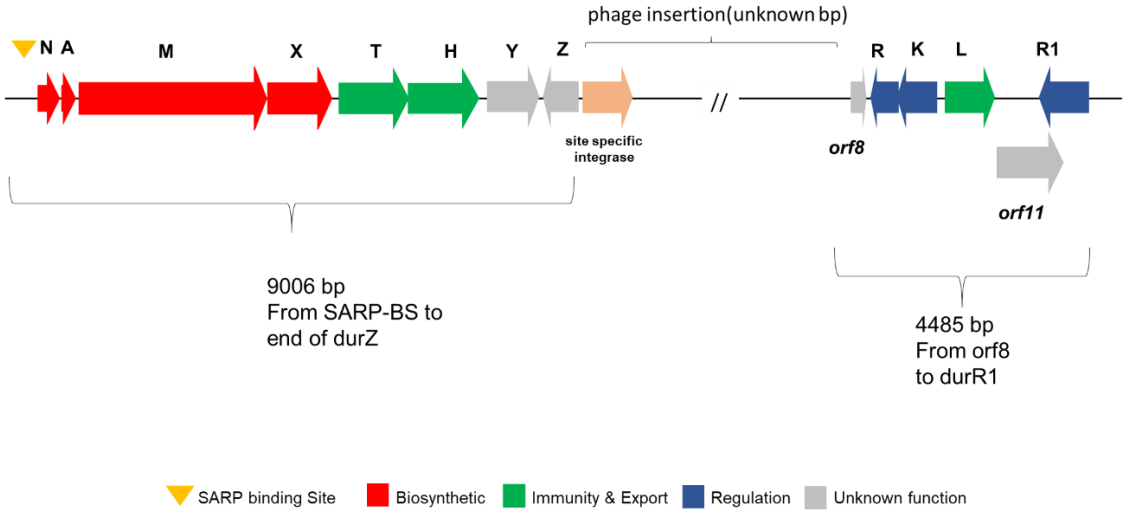



**Figure 4.1: Schematic of duramycin cluster including biosynthetic and regulatory genes.** The size of the phase insertion site could not be defined because of the organisation of the published genome of *Streptomyces cinnamoneus* ATCC 12686 in two different contigs.

Based on the above and the fact that the duramycin BCG has a SARP binding site that looks almost identical to that of the kyamicin (as well as cinnamycin) as shown in Figure 4.2, the plasmids pEVK6 (R1L) and pEVK7 (LR1) carrying the activation cassettes, as described in chapter 3.6.1, were used to transform *E. coli* ET12567/pUZ8002 and mated sequentially into *S. coelicolor* M1152 (Gomez-Escribano and Bibb, 2011) along with the pOJKHH vector that carries the biosynthetic genes for the production of duramycin.

```
kyamicin        GCGGTGAAACAGCTTTGAAGGCGGACTGAAACCACCGCCCCTAGCGTCGGTCGC      55
cinnamycin      CTCCTGAAAGCGGAGTGAAACCGTAGTGAAAGCGGACGCTCCTAGTGTCGTTCTC     55
duramycin       TCCCTGAAAGCGGACTGAAACCGTAGTGAAAGTGGGCGCCCTAGTGTCTTTCTC      55
                    *****  *   ****  ** * *****       * ***** ***  ** *

kyamicin        TGAAACAGCTTTGAAGGCGGACTGAAACC       29
cinnamycin      TGAAAGCGGAGTGAAACCGTAGTGAAAGC       29
duramycin       TGAAAGCGGACTGAAACCGTAGTGAAAGT       29
                *****  *   ****  ** * *****
```

**Figure 4.2: Sequence alignment of putative SARP binding sites of kyamicin, cinnamycin and duramycin amino acid sequences.** The alignment was performed with ClustalX2.

We observed inhibition of *B. subtilis* EC1524 in overlay bioassays only for the pEVK6 carrying strain, in contrast to the pEVK7 strain and the relevant controls (Figure 4.3).

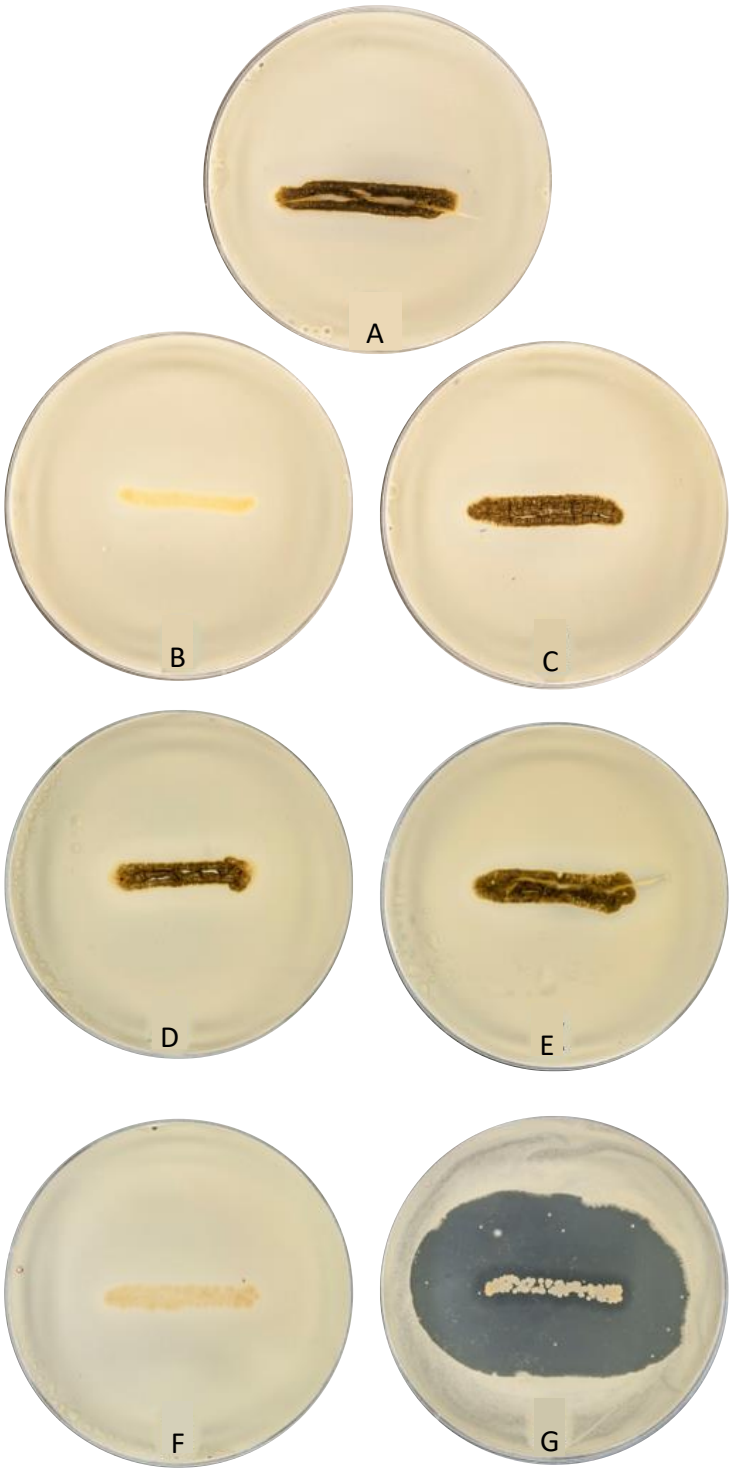



**Figure 4.3: Undersides of plates streaked with *S. coelicolor* strains overlaid with *B. subtilis* EC1524.** Duramycin production is indicated by a zone of inhibition of *B. subtilis* EC1524 around the culture. A: M1152, B: M1152/pOJKHH, C: M1152/pIJ10257, D: M1152/pEVK7, E: M1152/pEVK6, F. M1152/pOJKHH/pEVK7, G: M1152/pOJKHH/pEVK6.

Agar plugs were taken from the zone of the clearing or from near the streak, extracted and analysed as before by UHPLC-MS(ESI). For the strain with both the duramycin

biosynthetic genes and the pEVK6 plasmid an ion with *m/z* $[M + 2H]^{2+}$ = 1006.9229, corresponding to the expected ion for duramycin (Calculated m/z $[M + 2H]^{2+}$ for $C_{89}H_{125}N_{23}O_{25}S_3$ = 1006.9262; Δ = -3.2773 ppm). Duramycin was not present in the any of the relevant controls or with the pEVK7 activator cassette, with the regulation and immunity genes in the alternative order (Figure 4.4).

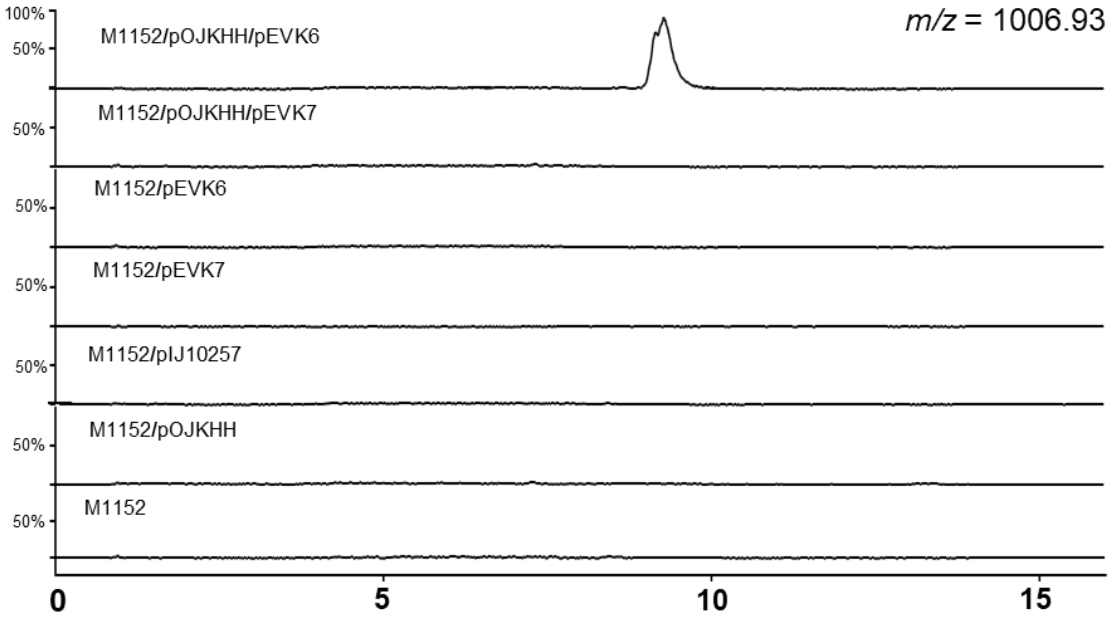



Figure 4.4: Extracted Ion Chromatograms (XIC) for duramycin bioassay plates. Agar plugs were taken from the bioassay plates in Figure 4.3 as described previously and analysed by UHPLC-MS (ESI). A peak corresponding to the duramycin $[M + 2H]^{2+}$ of 1006.93 (*m/z* observed = 1006.9229; *m/z* calculated = of 1006.9262; Δ = -3.2773 ppm) was detected only for M1152/pOJKHH/pEVK6. XICs are scaled such that 100% = a signal intensity of 8,000.

## 4.4 Genome mining to identify cryptic BGCs for type B lantibiotics

Next, we wanted to investigate whether the system used for the expression of the duramycin BGC can be used with other lantibiotic type B core peptides. As discussed in section 4.1, we have constructed a vector containing one the kya biosynthetic genes but with part of the core peptide sequence of *kyaA* replaced with a *Stu*I site, pWDW70 (section 3.7.5). We reasoned that we could use this plasmid in strain M1152pEVK6 as a platform for the expression of other type B lantibiotics by inserting core peptide

sequences into the *Stu*I site. We used public databases to find orphan type B lantibiotic BGCs, using as reference *cinA* and *cinN*. All the BGCs found were associated with genome sequencing projects.

21 BGCs were identified, mainly in actinomycetes, but also in cyanobacteria. This is an interesting observation since lantibiotics are usually produced by Gram-positive bacteria, whereas cyanobacteria are classified as Gram-negative. The list includes the three KY strains described in this study as well as three unpublished genomes of *Actinomadura* spp., kindly shared with us by collaborators (NAI698, NAI711, NAI716).

The core peptides from the strains identified were organised in an alignment made using Clustal Omega and the output from Clustal Omega was run through Boxshade (**Figure 4.5**).

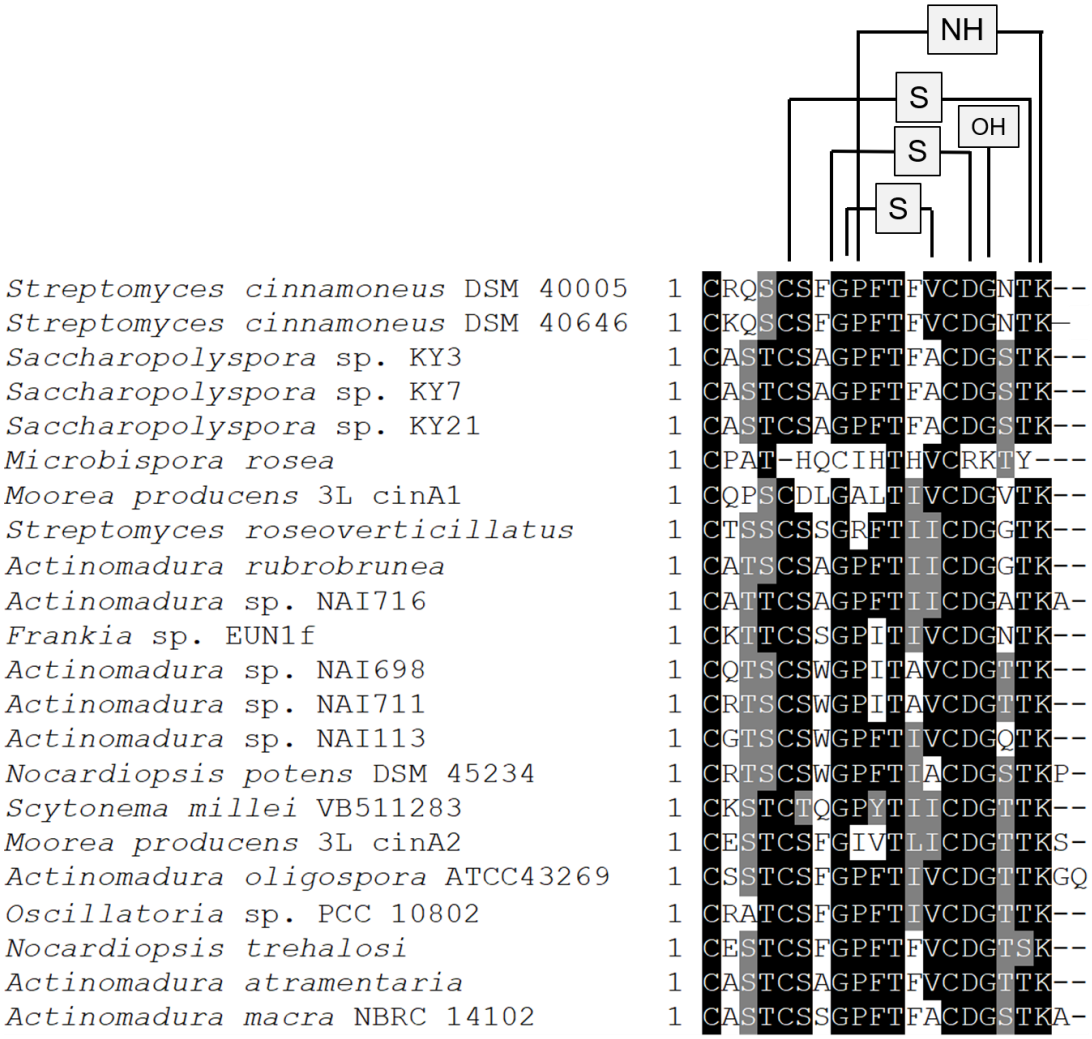



**Figure 4.5: Alignments of core peptides of all the positives hits at the databases on the search based on the *lan*N, the whole precursor peptide & the core peptide.**

From this alignment we can make many different observations. The *Microbispora rosea* core peptide bears no resemblance to the cinnamycin family core peptides. It has no possibility for a lysino-alanine ring or a hydroxy aspartate residue and all but one of the lanthionine rings cannot be formed. It is also the shortest of the potential core peptides at 15aa in length.

The other clearly atypical *lanA* is from the cyanobacteria *Moorea producens* 3L. The arrangement of the BGC in this organism is unusual in that there are two *lanA* genes associated with it. The *lanA1* core peptide cannot form the inner of the lanthionine bridges as there is no serine or threonine, but an aspartate. The core peptide from *Scytonema millei* VB511283, a cyanobacterium, lacks an obvious *lanN* even though it has residues that could form a lysine-alanine ring, although it is the only one to have a threonine instead of a serine. Strain *Actinomadura* sp. NAI716 is one of five with additional amino acids after the N-terminal lysine of the core peptide with a post-lysine alanine residue.

The core peptide from *Actinomadura macra* NBRC 14102 also possesses an extra alanine after the N-terminal lysine. The core peptide from *Nocardiopsis potens* DSM 45234 possesses a post-lysine proline, the core peptide *lanA2* from *Moorea producens* 3L has a post-lysine serine residue, and the *lanA* from *Actinomadura oligospora* ATCC 43269 has two post-lysine amino acids, a glycine and a glutamine. Assuming these extra amino acids are not sequencing artefacts, it could suggest that they do not interfere with the lysino-alanine ring formation.

After the analysis above we chose two lantibiotic core peptides with which to apply the expression system. Firstly, that of *Actinomadura atramentaria* NBRC 14695, which produces the previously characterised lantibiotic cinnamycin B (Kodani et al., 2016). Secondly, that of *N. potens* DSM 45234, which has not previously been characterised and is more dissimilar to kyamicin.

## 4.5 Expression of cinnamycin B BGC in *S. coelicolor*

The recently reported lantibiotic cinnamycin B (Kodani et al., 2016) is produced by *A. atramentaria* NBRC 14695. Cinnamycin B differs from kyamicin only in two amino acid residues, yet the reported bioactivity profile is very different. Whilst kyamicin is active against a range of Gram-positive bacteria (section 3.9.2), cinnamycin B was reported to be active only against *S. antibioticus* and to have no activity against *B. subtilis* or other standard indicator strains. This was therefore chosen as a target to test the expression system described in section 3.7.5.

### 4.5.1 Design and construction of synthetic operon for cinnamycin B expression

A synthetic double-stranded oligo encoding for the cinnamycin B core peptide was ordered and cloned into pWDW70 via GIBSON (Gibson, 2010) and was independently mated into M1152/pEVK6, along with empty pWDW70 as a negative control. The pWDW70 carrying the synthetic cinnamycin B core peptide was named pEVK11.

The strain carrying the activator cassette and the synthetic cinnamycin B core peptide showed inhibition against *B.subtilis* EC1524 (Figure 4.6). Agar plug extracts confirmed the presence of cinnamycin B with an *m/z* $[M + 2H]^{2+}$ ion of 920.3760 (calculated for $C_{79}H_{114}N_{20}O_{25}S_3$ $[M + 2H]^{2+}$ *m/z* = 920.3786; Δ = -2.8249 ppm) (Figure 4.7). Interestingly, this result is opposing what was previously reported. In order to verify the antibacterial activity of cinnamycin B against Gram-positive bacteria, it is required to prepare cultures of the producer strain, extract them and isolate the compound and attempt fragmentation analysis to reveal the connectivity of the lantibiotic.

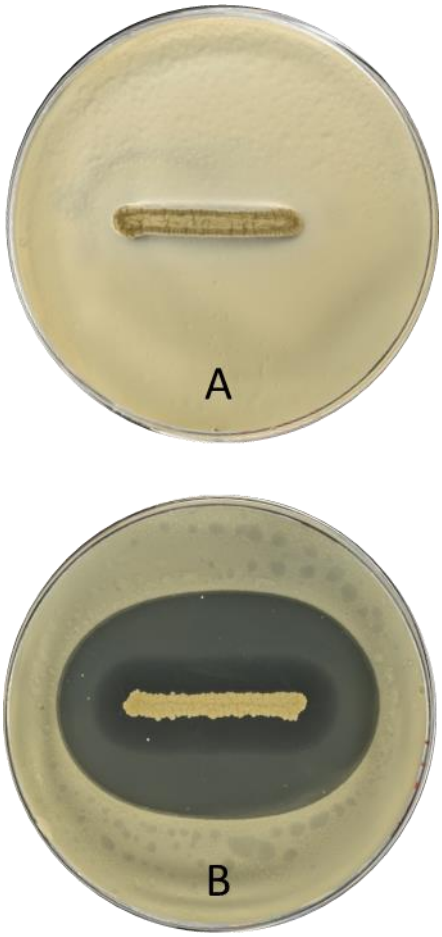



**Figure 4.6: Underside of plates with streaked strains of *S. coelicolor* overlaid with *B. subtilis* EC1524.** A: M1152/pEVK6/pWDW70, B: M1152/pEVK6/pEVK11. Inhibition was observed only at the strain carrying the activator cassette and the vector carrying the synthetic cinnamycin B core peptide.

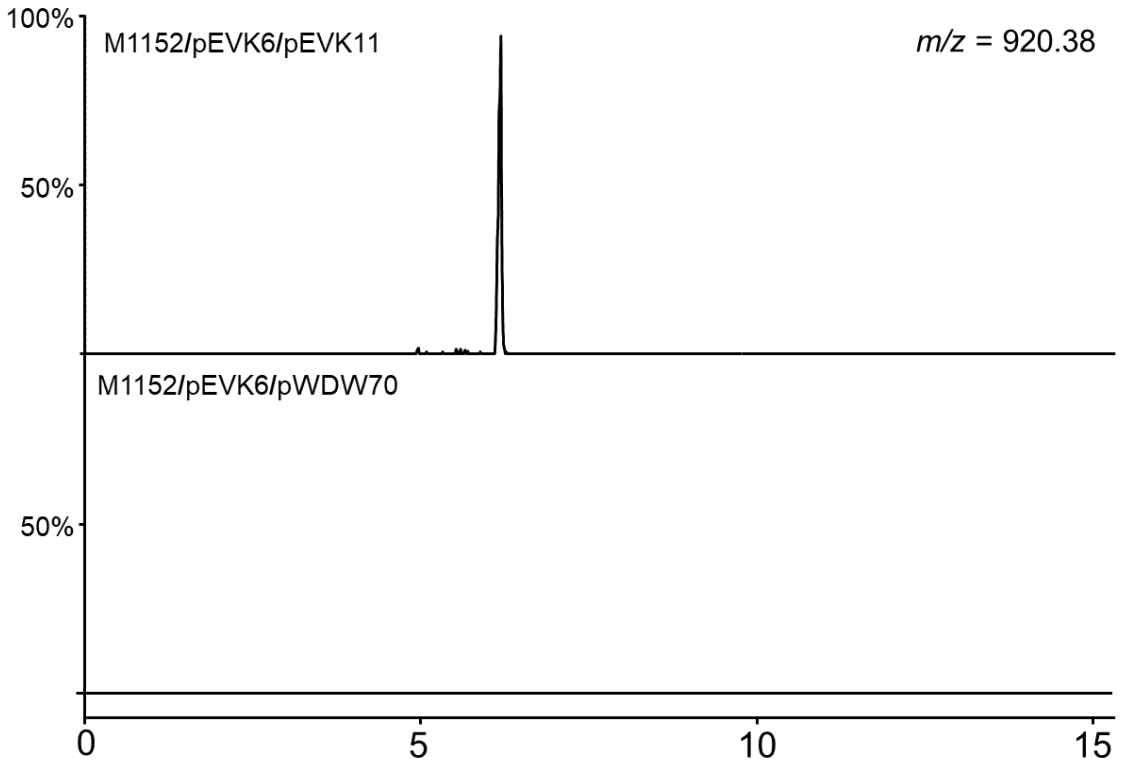



**Figure 4.7: Extracted Ion Chromatograms (XIC) for Cinnamycin B bioassay plates.** Agar plugs were taken from the bioassay plates in Figure 4.6 as described previously and analysed by UHPLC-MS (ESI). A peak corresponding to the cinnamycin B $[M + 2H]^{2+}$ of 920.38 (*m/z* observed = 920.3760; *m/z* calculated = of 920.3786; Δ = -2.8249 ppm) was detected only for M1152/pEVK6/ pEVK11. XICs are scaled such that 100% = a signal intensity of 2,500,000.

There might be various reasons leading to the discrepancy between the published and observed effects of cinnamycin B on *B.subtilis*. Firstly, different strains of *B.subtilis* were used in the two assays. Specifically, I used strain EC1524 in this work whereas *Bacillus subtilis* subsp *subtilis* NBRC 13719$^T$ was used in the Kodani publication. Secondly, they are referring to a spot-on-lawn method of assaying antibiotics, where 5 µg of cinnamycin B (0.5µL, 10 mg/mL solution) was spotted on an agar plate of the indicator strain. In my assay, the producing strain was streaked in the middle of the plate and left to grow for seven days before overlaying soft agar containing the indicator strain. The quantity of cinnamycin B produced on the plate cannot be accurately quantified without extraction of this plate but it is possible that it is greater than in the publication hence this could make a potential low activity more obvious. Lastly, we could not exclude the probability of an experimental error.

In order to identify the source of the discrepancy explained above, a potential experiment would involve acquiring the producing strain *Actinomadura atramentaria* DSM 43919$^T$

(=NBRC 14695$^{T}$). Growing the strain in the specified medium and conditions reported in Kodani et al., as well as following the exacts methods for cinnamycin B isolation, would lead to a quantity of the pure compound that could be tested further.

## 4.6 Efforts to express other potential lantibiotics from genetically distant families

From the database search (section 4.4), we identified some more distant strains that are carrying the *lanA* gene encoding for a potential lantibiotic. We selected the core peptide from *N. potens* DSM 45234 because it possesses a post-lysine proline. The thought behind this choice was to investigate whether the system could tolerate a longer by one amino acids core peptide sequence (**Figure 4.5**).

### 4.6.1 Design and construction of synthetic operon for expression of putative lantibiotic from *N. potens* DSM 45234

A synthetic double-stranded oligo encoding for the *N. potens* DSM 45234 putative type B lantibiotic was ordered and cloned into pWDW70 via GIBSON assembly (Gibson, 2010) and mated independently into M1152/pEVK6 along with empty pWDW70 as negative control. The pWDW70 carrying the synthetic *N. potens* lantibiotic core peptide was named pEVK12. Two colonies were transferred onto bioassay plates and grown for 7 days at 30°C.

### 4.6.2 Attempted bioassay of M1152/pEVK6/pEVK12 against *B. subtilis* EC1524

No inhibition against *B. subtilis* was observed after performing overlay bioassays of M1152/pEVK6/pEVK12 (Figure 4.8). Agar plugs were taken from the bioassay plates, extracted and analysed by HPLC-MS to test whether traces of the predicted compound were present. No *m/z* ion matching the expected molecular formula of $C_{89}H_{129}N_{25}O_{26}S_3$ was observed ($[M + 2H]^{2+}$ calculated as 1030.9424; $[M + 3H]^{3+}$ calculated as 687.6307).

**Figure 4.8: Underside of plate with streaked strain of *S. coelicolor* M1152/pEVK12 overlaid with *B. subtilis* EC1524.** The strain showed no inhibition against *B. subtilis*.

As described above, the thought behind selecting the core peptide from *N. potens* DSM 45234 was to investigate the elasticity and tolerance of the expression system. This specific core peptide sequence is longer by one amino acid, a proline, after the final lysine. The absence of a positive bioassay could either mean that the expression system cannot accommodate the extra amino acid or that downstream enzymes modified the peptide and did not allow for the production and release of the anticipated lantibiotic. Further genetic checks need to be performed to investigate whether the modified plasmid was properly conjugated into M1152. Nevertheless, pEVK12 was checked and verified by Sanger sequencing before conjugation and it also conferred the appropriate resistance to the ex-conjugants.

## 4.7 Discussion

Using the kyamicin activation cassette, we managed to express the duramycin cluster in *S. coelicolor* M1152 and verified production of the compound with overlay bioassays and MS. This demonstrates that the activation cassette can be used with a lantibiotic BGC from different species. This suggests that the regulation and immunity mechanisms of type B lantibiotics are a general phenomenon, and SARPs and immunity genes are interchangeable between the different BGCs. Again, pEVK7 (LR1) did not activate expression, which may be due to *kya*L not being expressed, as discussed previously.

We then explored the lantibiotic machinery in a different way with the intention to engineer a platform for expression of type B lantibiotics. Firstly, cinnamycin B was successfully expressed by using this approach. This demonstrates that the kyamicin biosynthetic proteins are promiscuous enough to tolerate the modification of alanine to valine at position 13 and serine to threonine at position 17. However, as these changes are relatively small, further investigation is required to determine the limits of the changes tolerated in the core peptide. The next candidates to try are the mathermycins and cinnamycin itself.

The next steps for developing the lantibiotic expression platform further are to explore different heterologous hosts, such as the *Saccharopolyspora* superhost. Other possibilities are trying different biosynthetic machinery or activator cassettes, to find the most promiscuous biosynthetic proteins and the most widely applicable regulation and immunity genes.

The expression of the putative lantibiotic of *N.potens* DSM 45234 was not successful. The major difference between this core peptide and kyamicin is an extra amino acid in the former. It may be that changes in peptide length are less tolerated than amino acid substitutions. However, as this lantibiotic is only predicted from genomic data, we cannot rule out that it does not encode a functional lantibiotic in the original strain.

We also showed that cinnamycin B is active against *B. subtilis* EC1524, contrary to previously published data. In this work a different strain, of *B. subtilis* was used, namely *B. subtilis subsp subtilis* NBRC 13719$^{T}$, which could account for the different results observed. However, as cinnamycin B and kyamicin are so similar it seems unlikely that they would have markedly different bioactivity profile.

Future work on this project will consist of various approaches to establish a robust expression system based on the kyamicin machinery. Firstly, we will focus to more

similar genera to cinnamycin/kyamicin. Then, the database search will be repeated in more depth in order to identify new additions to putative lantibiotic BGCs and attempt to express them, regardless of the genera.

Finally, attempts will be done to introduce variations in the established lantibiotics core peptides in order to create randomised analogues of type B lantibiotics that do not exist in nature.

Overall, we demonstrated that firstly the activator cassette described in this work can be used for interspecies expression of BGC's and secondly that we can engineer a platform for the expression of orphan type B lantibiotic BGCs. These novel expression systems have a lot of potential in the search for novel antibiotics and enable us to gain invaluable insight into the structure-function relationship of lantibiotics.

# Chapter 5:

# Discovery of a novel siderophore and antifungal compound by untargeted metabolomics and bioassay-guided approaches

# 5 Discovery of a novel siderophore and antifungal compound by untargeted metabolomics and bioassay-guided approaches

## 5.1 Introduction

The practise of combining genomics with untargeted data-dependent tandem MS is an effective way of identifying antimicrobial compounds (Senges et al., 2018). Basically, it consists of growing the strain in question in an array of different media with varying composition, and the analysis of the culture extracts using UHPLC coupled with MS or tandem MS (MS2) (Baltz, 2017). The data are analysed either with specialised software and metabolic networks can be constructed to provide a deeper understanding of the chemical diversity and structural relationships within the detected metabolomes (Chavali et al., 2012, Nothias et al., 2018). The appearance of each compound can be correlated with the presence of biological activities in the samples.

One of the most traditional approaches to identify promising molecules for antimicrobial use is the bioassay-guided discovery and purification, combined with chemical structure elucidation (Choudhary et al., 2017). Although this approach harbours the risk of the re-discovery of known compounds, it is the fastest way to identify antibacterial and antifungal molecules, by using indicator strains commonly available in the laboratory (Silver, 2011). A previous knowledge about the biosynthetic potential of the strain based on genomic information it is extremely useful to reduce the likelihood of rediscovery events.

Overall, this tripartite approach of genomics, metabolomics and bioassays allows us to pin down novel antimicrobials and to explore the potential of talented strains from rare environmental sources, such as the KY strains.

*Saccharopolyspora sp.* KY21 is one of the KY strains isolated from fungus farming *T. penzigi* ants as presented in Chapter 3. This strain was the most tractable under laboratory conditions and was chosen as the basis for further studies. Previously I outlined how the activation of a silent gene cluster enabled the discovery of a new lantibiotic from this strain. Here I describe how this strain was further probed for novel natural products using a combination of untargeted metabolomics and bioassay guided approaches.

## 5.2 Objectives

- Growth of *Saccharopolyspora* KY21 in multiple media and generation of crude extracts.
- Identification of novel compounds by untargeted metabolomics.
- Bioassay-guided identification of antimicrobial compounds.

## 5.3 Growth trials and analysis of crude extracts

In order to investigate which metabolites *Saccharopolyspora* KY21 can produce, the strain was grown in 13 different liquid media (SM media described in section 2.3.2) for seven days. These production media are commonly used in industry as screening media for the identification of natural products from actinomycetes. The selection of these 13 media, out of a larger group of media available, was made based on differences in composition, pH and additives. The aim of this selection was to vary as much as possible the potential metabolic profile of the strain. Spores of the strain KY21 were used to inoculate centrifuge tubes containing 10 mL of each medium in duplicate. Samples were taken at various timepoints over a period of seven days of incubation and extracted with both methanol and ethyl acetate. The crude extracts were both tested for antimicrobial activity and analysed by UHPLC-MS.

## 5.4 Discovery of the novel NRPS-derived siderophore EV60

The collective LC-MS data were analysed by Profiling Solutions (Shimadzu Corporation, Japan). This is a molecule discovery analysis software for data produced by LCMS systems, commonly used for untargeted compound analysis. The software provides the ability to combine data from multiple LCMS files and to generate comparative abundance profiling studies. Specifically, it processes multiple data files to perform peak picking and listing of the *m/z* values, retention times, and signal intensities for each peak. The output of this analysis is a collective table that allows for comparison and overview of the data. The signal intensities are displayed with colour-coding and that allows for rapid identification of peaks that appear in one medium and not at the rest of the media.

Based on this, it became apparent that a compound of *m/z* = 428.1663 with predicted formula $C_{18}H_{25}N_3O_9$ for the $[M + H]^+$ adduct, was present in the crude extracts deriving from only medium SM7 and not in any of the others tested. Analysis of the crude extract from SM7 showed that this *m/z* corresponded to a peak at a retention time of 5.3 min in the LC (Figure 5.1). Searching for this specific *m/z* in Reaxys, a database that provides access to experimentally measured data, didn't give any matches, suggesting it was novel. The compound was named EV60 and investigated further.

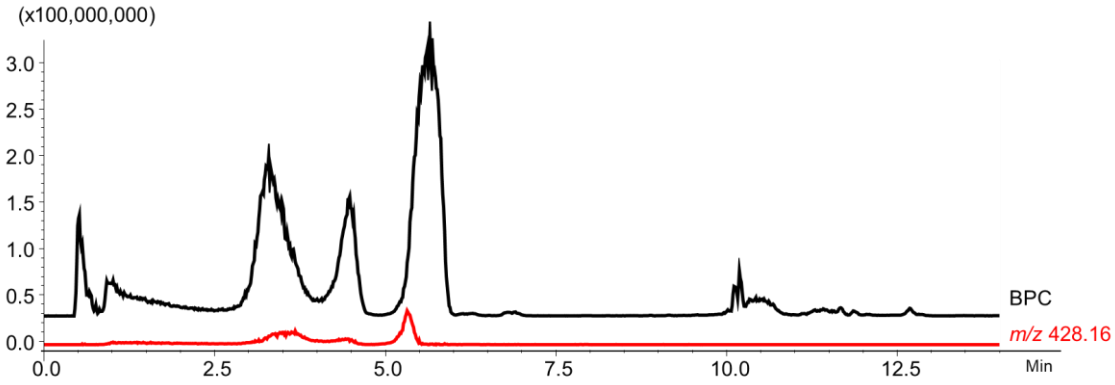



**Figure 5.1: Base peak chromatogram of crude ethyl acetate extract of culture of KY21 in SM7** (shown in black line).The XIC of EV60 (red line) shows a peak corresponding to the *m/z* of 428.16.

### 5.4.1 Isolation of EV60

In order to investigate the nature of compound EV60, spores of strain KY21 were used to inoculate a preculture of SV2 (40 mL) for three days. This preculture was used to inoculate SM7 medium (1 L divided in 4 x 500 mL flasks) and the culture was incubated at 28°C for six days. The whole culture was extracted with ethyl acetate (2 × 500 mL), and the combined organic fractions were concentrated under reduced pressure, loaded onto a SPE column, washed with water and eluted with 30%, 50% and 100% methanol. EV60 was found in the fraction eluted with 30% methanol and resulted in 3 mg of pure compound, requiring no further purification. The collected material was subject to chemical characterisation and efforts were made to match it with a predicted BGC.

### 5.4.2 EV60 Structural elucidation

To gain further insight into the EV60 structure, NMR analysis, connectivity of the molecule and elucidation of the stereochemistry were carried out by Dr. Daniel Heine. The sample was dissolved in deuterated DMSO and $^{1}H$ and $^{13}C$ NMR spectra were recorded. Two-dimensional NMR spectra were then carried out, including HMBC, HSQC and COSY. Details of the NMR data are provided in Appendix 2.

Based on the MS and NMR data we hypothesised that EV60 is a pseudo-tripeptide composed of 2,3-dihydroxybenzoic acid (2,3-DHBA), threonine and $N^5$-hydroxyornithine (Figure 5.2). Additionally, the hydroxylated amino group of the $N^5$-hydroxyornithine side chain features an $N^5$-acetylation. This structure, consisting of a peptide backbone comprising 2,3-DHBA and ornithine, a non-proteogenic amino acid, strongly suggests that the compound could be the product of an NRPS assembly line.

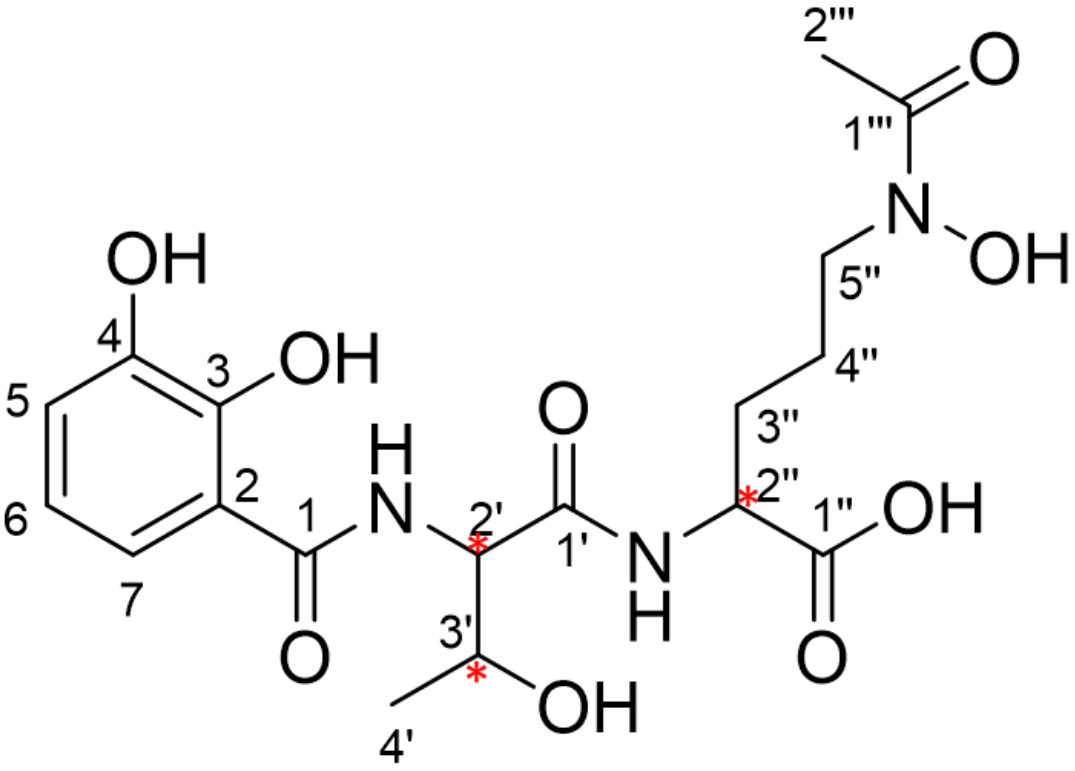



**Figure 5.2: Planar structure of EV60 composed of 2,3-DHBA, threonine and ornithine.** Observed *m/z* = 428.1663 for $[M + H]^+$, calculated *m/z* for $C_{18}H_{25}N_3O_9$ $[M + H]^+$ = 428.1669 (Δ = -1.4013 ppm). Stereocenters indicated by a red asterisk.

Having determined the atomic connectivity of EV60, we wanted to define the stereochemistry of the three stereocenters in the molecule. The stereochemistry of the threonine residue could be determined by Marfey's analysis (Marfey, 1984), employing acidic hydrolysis of the compound and derivatisation of the hydrolysed peptide using 1-fluoro-2-4-dinitrophenyl-5-L-alanine amide (FDAA). This analysis was performed by Dr Daniel Heine. The retention time of the derivatised hydrolysis product was compared to that of the FDAA-derivatives of authentic standards of L- and D-threonine, demonstrating the D- configuration of the threonine residue of the backbone peptide (Figure 5.3). Due to the lack of a commercial standard for D- or L-$N^5$-hydroxyornithine it was not possible to use the same method to determine the stereochemistry at this position. The partial stereochemistry of EV60 is shown in Figure 5.4.

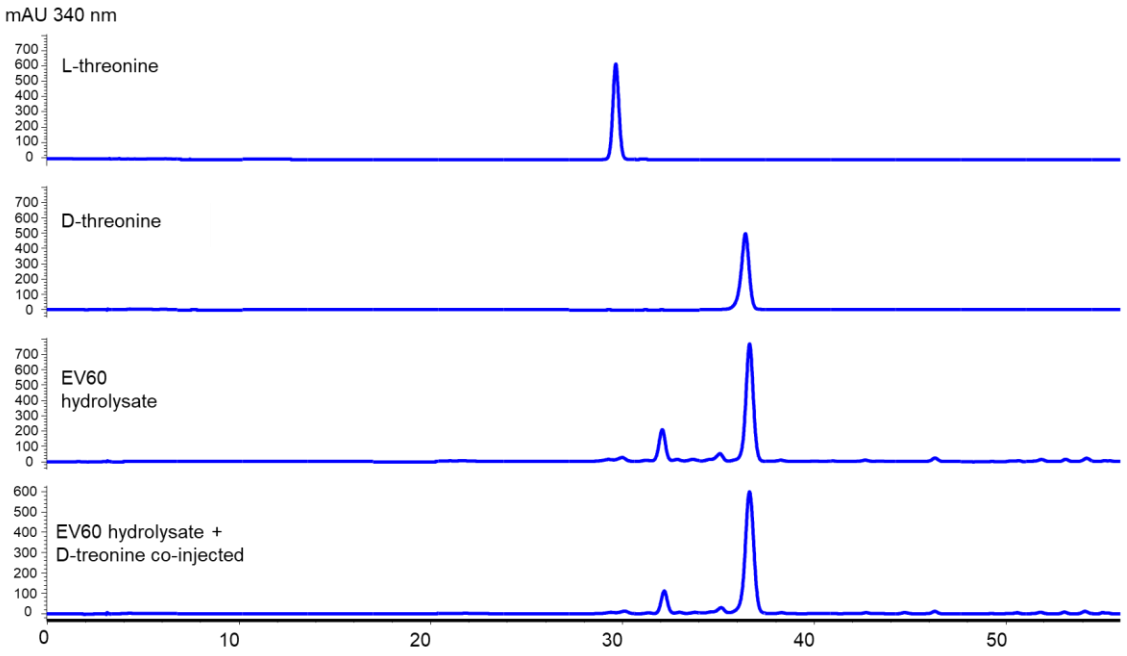



**Figure 5.3: Marfey's analysis for EV60.** Top LC: L-threonine, Second LC: D-threonine, Third LC: EV60 hydrolysate, Bottom LC: EV60 hydrolysate + D-threonine co-injected. Retention times: L-Thr: 29.6 min, D-Thr: 36.5 min, EV60: 36.5 min.

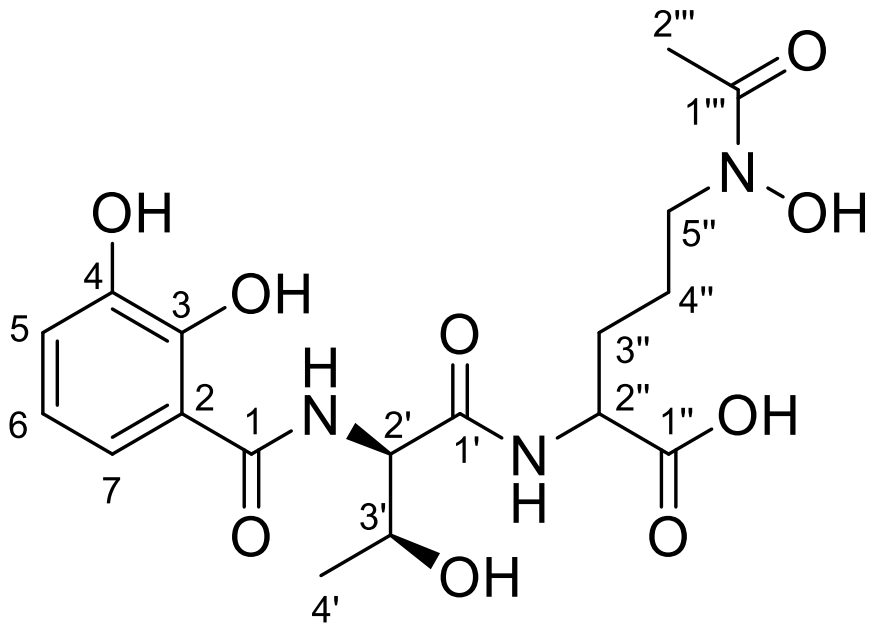



**Figure 5.4: EV60 structure showing stereochemistry for part of the molecule.** The stereochemistry of the threonine residue is the D- configuration. No commercial standard was available for D- or L-$N^5$-hydroxyornithine in order to determine the stereochemistry at this position.

### 5.4.3 Biosynthetic origins of EV60

The structure of EV60 suggested that it is a siderophore (Neilands, 1995). Siderophores are a structurally diverse group. They can be classified depending on their metal binding groups as catecholates (Patzer et al., 2003), hydroxamates (Powell et al., 1980), or carboxylates. EV60 features all those characteristic chemical moieties. Among catecholates, 2,3-DHBA is a natural phenol that is incorporated into various siderophores, and $N^5$-hydroxyornithine would represent an hydroxamate. In addition, EV60 also displays a carboxylic acid derived from the α- carboxylate group of the $N^5$-hydroxyornithine residue of the molecule.

In the most parsimonious scenario, the structure of the molecule would suggest that it is the product of an NRPS assembly line consisting of three modules with adenylation domain specificities for 2,3-DHBA, L- threonine (with no epimerization domain in module 2) and $N^5$-hydroxy-L-ornithine, respectively.

As mentioned before, the antiSMASH output of strain KY21 revealed 23 putative BGCs and five of them contain NRPS assembly lines, Table 5.1. Interestingly, BGC number 4 includes a bimodular NRPS with adenylation domain specificities that strongly suggest the incorporation of an L- threonine by the N-terminal module of the NRPS, and possibly an ornithine by the C-terminal module (as suggested by Stachelhaus code (Marahiel et al., 1997), one out of the four prediction tools implemented in antiSMASH). Together with the presence of a predicted lysine/ornithine-N-monoxygenase located within the cluster, this prediction could mean that in fact this C-terminal module is incorporating $N^5$-hydroxyornithine in the NRPS product. The presence of a thioesterase domain at the end of the bimodular NRPS indicates the end of the NRPS assembly line and the release of its product as a carboxylic acid, thus matching the requirements of the elucidated structure for EV60.

However, this BGC 4 lacks an NRPS module with a predicted adenylation domain specificity for 2,3-DHBA. There is a standalone adenylation domain (Abe et al., 2017) with predicted specificity for small non-polar amino acids (for example gly, ala, val, leu) that is unlikely to justify the incorporation of 2,3-DHBA into the NRPS assembly line.

**Table 5.1: Putative BGCs in KY21 that contain NRPS assembly lines.** Note that BGC18 is the BGC that encodes for kyamicin, as described in chapter 3.

| antiSMASH Cluster No. | BGC type | Position | | BGCs of compounds with highest similarity (%) |
|---|---|---|---|---|
| | | from | to | |
| 4 | NRPS | 1630963 | 1692484 | No significant similarities |
| 12 | NRPS | 3741605 | 3797270 | Streptothricin (83%) |
| 15 | NRPS-T1PKS | 4331679 | 4388369 | SGR_Polycyclic tetramate macrolactam (50%) |
| 16 | NRPS | 4391798 | 4456904 | Coelibactin (54%) |
| 18 | Bacteriocin-Lantipeptide-NRPS | 4693417 | 4743949 | Cinnamycin (47%) |

Intriguingly, another NRPS-containing BGC (cluster 16) does include a standalone adenylation domain showing the expected substrate specificity for 2,3-DHBA. The same cluster contains an isochorismate synthase (required for the biosynthesis of 2,3-DHBA), thus supporting the specifity of this adenylation domain. In addition to this, high-identity homologues of the two additional proteins required for the biosynthesis of 2,3-DHBA (isochorismatase and 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase) have been found close to each other elsewhere in the KY21 genome. It has been previously reported that in some cases the biosynthesis of siderophores requires the cross-talk between two or more BGCs spread across a genome (Lazos et al., 2010, Bosello et al., 2011). Therefore, it could be possible that EV60 is the product of cross-talk between BGC4 and BGC16.

The absence of epimerization domain in the threonine module of BGC4 is surprising, given the D-configuration of the threonine residue. Additionally, the threonine-specific module of the BGC4 bimodular NRPS does not feature the conserved amino acid motif characteristic of dual condensation-epimerization domains (Rausch et al., 2007), which could have been a different explanation for the D configuration of the threonine residue in the final product. A non-canonical NRPS biochemistry or the action of a hypothetical NRP tailoring epimerase could explain our observations, but additional experimental evidence and detailed bioinformatics analysis would be required to test these hypotheses. Mutational experiments will be performed to test whether the biosynthesis of EV60 is a result of cross-talk between multiple BGCs.

## 5.5 Bioassay-guided discovery of an antifungal

The antiSMASH output of strain KY21 suggested some BGCs encoding possible antifungal activity. As shown in Table 5.2, there are two characteristic predictions that could encode for an antifungal; a BGC matching a FR-008/Candicidin BGC (Chen et al., 2003) as well as a SGR_PTM, a polycyclic tetramate macrolactam compound firstly isolated and described from *Streptomyces griseus* (Luo et al., 2013). In order to investigate whether strain KY21 could produce an antifungal compound, we employed the bioassay-guided approach.

**Table 5.2: Putative BGCs in KY21 that could lead to an antifungal compound**

| antiSMASH Cluster No. | BGC type | Position | | BGCs of compounds with highest similarity (%) |
|---|---|---|---|---|
| | | from | to | |
| 5 | T1PKS | 2399316 | 2448369 | No significant similarities |
| 7 | Lassopeptide | 2966632 | 2988078 | Chaxapeptin (28%) |
| 15 | NRPS-T1PKS | 4331679 | 4388369 | SGR_Polycyclic tetramate macrolactam (50%) |
| 20 | T1PKS | 5022943 | 5142561 | FR-008 (76%) |

### 5.5.1 Bioassay results with KY21 on agar plates

KY21 cultures were therefore assayed for anti-fungal activity using the bioindicator strain *Candida albicans* CA-6, kindly provided by Matthew Hutchings lab, UEA (Marconi et al., 1976). SFM plates were firstly inoculated with strain KY21 spores in the centre of the plate and left to grow for seven days at 30 °C. These plates were subsequently overlaid with cultures of *C. albicans* CA-6 mixed with SNA and left at room temperature for 18 h. A zone of clearing was observed around the KY21 culture, demonstrating that strain KY21 has antifungal activity (Figure 5.5).

**Figure 5.5: Strain KY21 overlay bioassay against *C. albicans* CA-6.** Clearing can be noticed around the growth of spores of strain KY21 on an SF+M plate, after 7 days of growth.

In order to recapitulate the overlay bioassay result, strain KY21 was cultured in the selection of the 13 liquid media as described in section 5.3. The crude KY21 methanolic and ethyl acetate extracts were then tested using disk diffusion assays to identify the antifungal compound. Out of the 13 production media, antifungal activity was observed only from the crude methanolic extracts and the water phase of the ethyl acetate extractions of media SM12 and SM18. The solvent phase of the ethyl-acetate extractions had no activity (Figure 5.6**)**. SM12 was chosen for further study as it is an easier medium to handle for extractions.

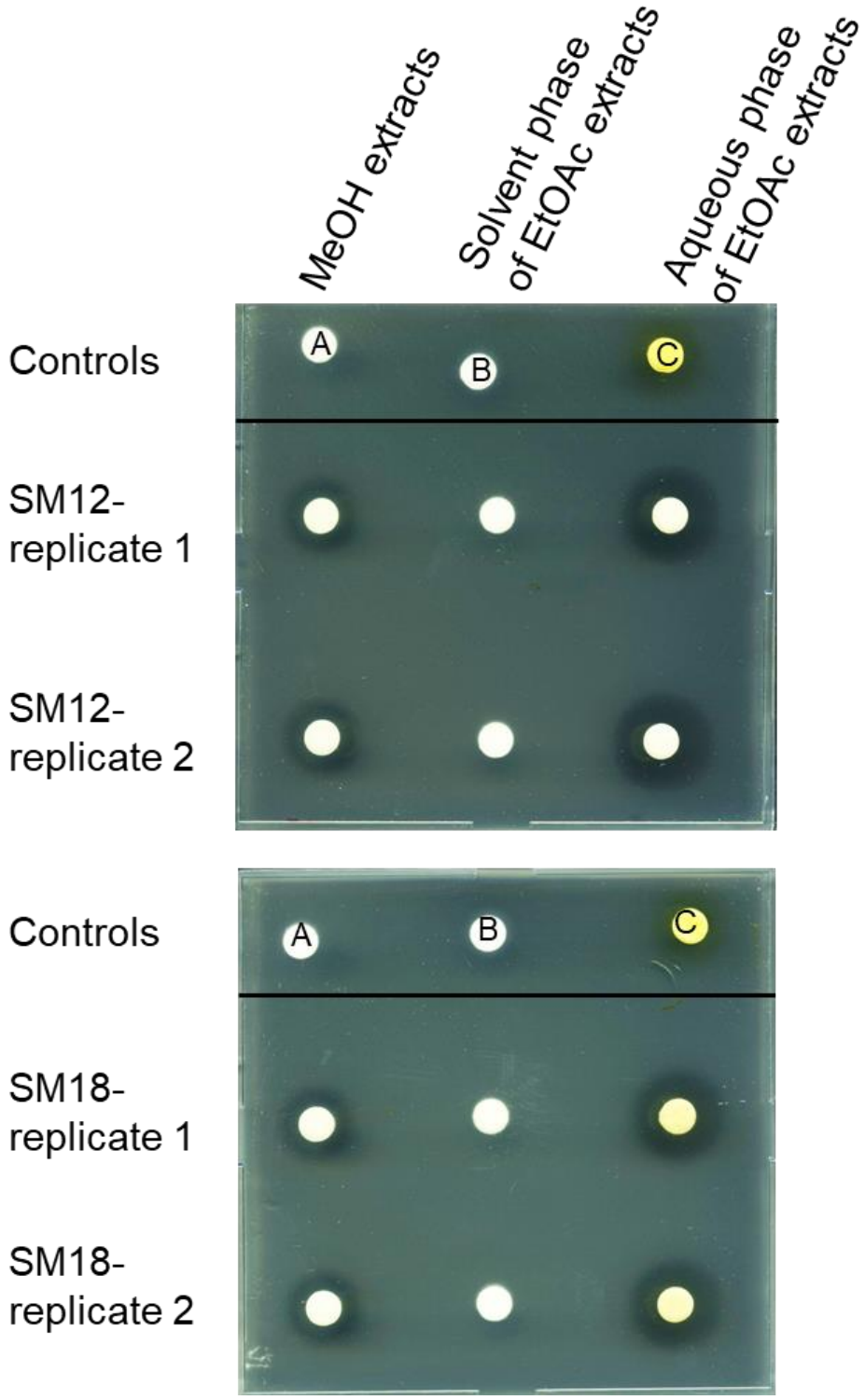



**Figure 5.6: Disk diffusion bioassay testing crude methanolic and ethyl acetate extracts of strain KY21 against *C. albicans* CA-6.** For extracts from both SM12 and SM18 a zone of clearing is observed at the methanolic extracts and the aqueous phase of the ethyl acetate extracts. On the top of each plate negative and positive controls have been included respective to the solvent tested. A: methanol and B: ethyl acetate as negative controls and C: nystatin as positive control

Preliminary metabolomics analysis was carried out, but the metabolic profiles of the crude extracts were too complicated to be able to assign a peak present to all positive bioassay samples. For that reason, the scale up of strain KY21 in SM12 was carried out for the isolation of antifungal compound production. A 2 L culture of strain KY21 in medium SM12 was grown for seven days and the cells were harvested by centrifugation. The supernatant was extracted with ethyl acetate to give an organic and aqueous phase, and the cell pellet was extracted with methanol. All the extracts were processed and tested in bioassays against *C. albicans* CA-6 (Figure 5.7).

In contrast to the smaller scale extractions described earlier, the crude ethyl acetate extract showed antifungal activity this time and with a larger inhibition zone than before. Preliminary efforts were carried out in order to fractionate the crude extract and to identify the active fraction (Figure 5.7).

Further efforts to fractionate and identify the antifungal were paused by inconsistent bioactivity of the extracts, low yields of crude material and isoflavone contaminants from the soy-based media. Investigation of this antifungal compound will be continued in future experiments.

Specifically, the next steps of this project will be repeating scale-up of the culture and attempting isolation of the bioactive fraction of the crude extract. After acquiring the fraction in question, a detailed LCMS analysis can reveal a compound with a specific mass and a UV-vis absorption spectrum. It is known from literature that the published antifungals presented as possible matches of KY21 BGCs in table 5.2, have characteristic UV-vis absorption spectra. Specifically, the FR-008/Candicidin BGC (Chen et al., 2003) presents three distinct peaks in the range of 350-420 nm, standard absorption pattern of polyenes (Szwarc et al., 2015). Similarly, based on the prediction of SGR_PTM, a polycyclic tetramate macrolactam firstly isolated and described from Streptomyces griseus (Luo et al., 2013), we can search for UV maxima around 220 and 323 nm, characteristic for polycyclic tetramate macrolactams (Blodgett et al., 2010).

Further whole genome bioinformatics analysis is also necessary in order to study further possible BGCs and their architecture and ascertain whether they could lead to a natural product. The existing computational tools for BGC identification, such as antiSMASH (Kautsar et al., 2017, Weber et al., 2015) aim primarily to the identification of known types of BGCs and are being constantly updated. Without doubt, this has been important in natural product discovery so far but they lack the ability to predict novel types of BGCs (van der Lee and Medema, 2016).

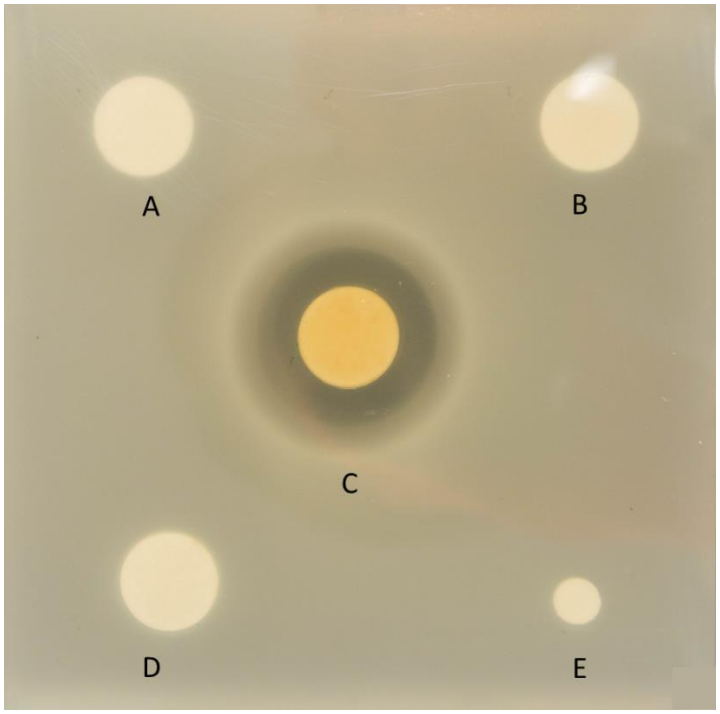



**Figure 5.7: EtOAc extracts of KY21 from 2L liquid culture in SM12 against *C. albicans* CA-6**. A and B are two different fractions of C, that have no activity against *C. albicans* whereas C is the crude ethyl acetate extract collected. D: EtOAc_12 is a sample from the extraction line of falcon tubes, that hadn't given any inhibition, so here is used as an additional negative control next to E: Ethyl Acetate as negative control since all samples used in this bioassay were redissolved in ethyl acetate.

## 5.6 Discussion

In this chapter, two different approaches for NP discovery are described. Firstly, untargeted metabolomics, where the entire metabolome of the bacterial strains grown under a variety of media and culture conditions. This allows for rapid identification of NPs, including pathway intermediates that might be dismissed in other analyses (Watrous et al., 2012). Via this approach a novel NRPS-derived siderophore was identified and its stereochemistry was partially elucidated. Siderophores are molecules that strongly bind metal ions for absorption into microbes (Saxena et al., 1986). It was not possible to find a putative BGC responsible for all the parts of this pseudo-tripeptide. We hypothesise that this compound is the result of the cross talk between two BGCs. Characteristic reported examples of this phenomenon are the putative siderophore erythrochelin, that its production require crosstalk between two separate NRPS BGCs (Lazos et al., 2010) as well as rhodochelin, that requires the coordinated expression of three independent gene clusters in *Rhodococcus jostii* RHA1(Bosello et al., 2011). Cross-talk between clusters is a very powerful evolutionary strategy since it offers more flexibility and more chemical diversity, meaning that with the same number of building blocks, more than one compound can be produced (Wandersman and Delepelaire, 2004).

Mutational experiments will be conducted in order to identify the building blocks that lead to the compound. Specifically, disruption of the adenylation domains in both BGC4 and BGC16 with a single recombination event, will reveal if they are actually responsible for the NRPS, hence whether the cross-talk hypothesis is accurate. Preliminary mutational manipulations of the strain have been attempted, firstly for the adenylation domain outside BGC4, and in BGC16, in order to test the theory of cross-talk between clusters. Disruption of the NRPS from BGC16 of strain KY21 with the adenylation specificity for 2,3-DHBA resulted in the mutant strain not growing in minimal media. Growth of the mutants in rich media was possible, suggesting that this siderophore is essential for the health of the bacteria (unpublished data).

In terms of the novel siderophore identified, more analyses can be done in the future using siderophore binding assays, such as the chromeazurol S (CAS) assay (Alexander and Zuberer, 1991) to monitor iron binding. The mode of action of the molecule can also be investigated to connect metal chelation to potential starvation of bacteria.

In terms of the antifungal compound, further work will be carried out in order to pin down its structure. Firstly, the scale-up experiments and extractions will be repeated to

establish the most robust way of growth of culture and suitable solvent, to solve the issue with inconsistency. Once adequate amount of the compound is isolated, NMR analysis will be carried out, so we can primarily assign the NP to a putative BGC from the KY21 antiSMASH output. In order to facilitate this assignment, we can employ the knowledge we have from antifungal-related literature, specifically in terms of expected UV-vis absorption spectra. Finally, the possibility that activity guided fractionation may lead to the isolation of two or more different antifungal molecules should not be dismissed. When performing a step-by-step separation of compounds, we base it on differences in their physicochemical properties. When we perform a bioassay-guided fractionation, the aim is to assess the biological activity of such compounds. Since this procedure usually begins with a crude extract that shows activity in a preliminary bioassay, the likelihood of simultaneous isolation of more than one NP is high. A thorough testing of the isolated fractions can help to identify whether this is the case with strain KY21 and match the compounds to the corresponding predicted BGCs.

# Chapter 6:

# Investigation of the biosynthesis of sporeamicin A

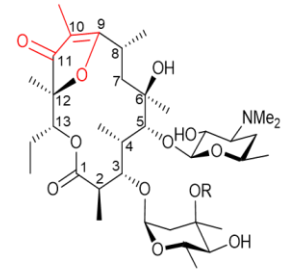
# 6 Investigation of the biosynthesis of sporeamicin A

## 6.1 Introduction

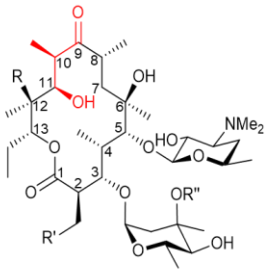
Sporeamicin is a 14-membered macrolide antibiotic, structurally related to erythromycin (Figure 6.1). Sporeamicin A is produced by *Saccharopolyspora* sp. L53-18 (FERM BP-2231) (Yaginuma et al., 1992). This strain is a Gram-positive, fast-growing actinomycete that sporulates easily in liquid and solid media. It was first isolated from a soil sample at Setouchi-cho, Kagoshima Prefecture, Japan (Yaginuma et al., 1992).

Based on literature data, the molecular formula of sporeamicin A is $C_{37}H_{63}NO_{12}$ and it exhibits a strong UV absorption peak at 276 nm. The aglycone of sporeamicin A is highly similar to that of erythromycin A, but with an ether bridge between C12 and C9, a double bond between C10 and C9 and a keto group instead of a hydroxyl group at C11.

Although sporeamicin A has two-fold less bioactivity than erythromycin A against Gram-positive bacteria, it is more acid stable than erythromycin A, which leads to its higher oral bioavailability and better *in vivo* efficacy in bacterial infections. It is active *in vitro* against Gram-positive bacteria & macrolide resistant strains such as *S. aureus* (Morishita et al., 1992). Interestingly, sporeamicin A was first chemically synthesized from erythromycin A, only and later was it isolated from *Saccharopolyspora* sp. L53-18 (Freiberg, 1989). Its structure suggests biosynthetic enzymes with different substrate tolerance compared to the erythromycin pathway, or a novel pathway. Intriguingly, the original report on the production of sporeamicin A suggests it is the major product of *Saccharopolyspora* sp. L53-18 and makes no comment on the co-production of erythromycin A or its congeners (Yaginuma 1992).



Sporeamicin A: R = $CH_3$
Sporeamicin B: R = H



Erythromycin A: R = OH, R' = $CH_3$, R'' = H
Erythromycin B: R = R'' = H, R' = $CH_3$,
Erythromycin C: R = OH, R' = R'' = H
Erythromycin D: R = R' = R'' = H
Erythromycin F: R = R'' = OH, R' = $CH_3$

**Figure 6.1: Comparison of the chemical structures of Sporeamicin and Erythromycin variants.** Key differences are highlighted in red (Omura, 2002).

The structural differences between sporeamicin A and erythromycin A, and the mechanisms behind the production of sporeamicin A, might reveal enzymes valuable for the bioengineering of new erythromycin analogues.

Taking inspiration from the semi-synthesis of sporeamicin A from erythromycin A described above (Freiberg, 1989, Faghih et al., 1996), we reasoned that a key intermediate in the biosynthesis would require at some point a C11-keto group. This would increase the ability for dehydration of the hemiketal intermediate formed after attack of the C12-hydroxy group on the C9-keto group.

Our first biosynthetic idea was that a C11-keto group could be formed by action of a modified erythromycin-like polyketide synthase. Formation of the dihydrofuran moiety could then occur after post-PKS modification to introduce a C12-hydroxyl group (Figure 6.2). Given the natural order of events that take place during the biosynthesis of erythromycin A, and the high substrate specificity, this would imply the equivalent genes for making sporeamicin A would behave differently meaning they would be useful for bioengineering.

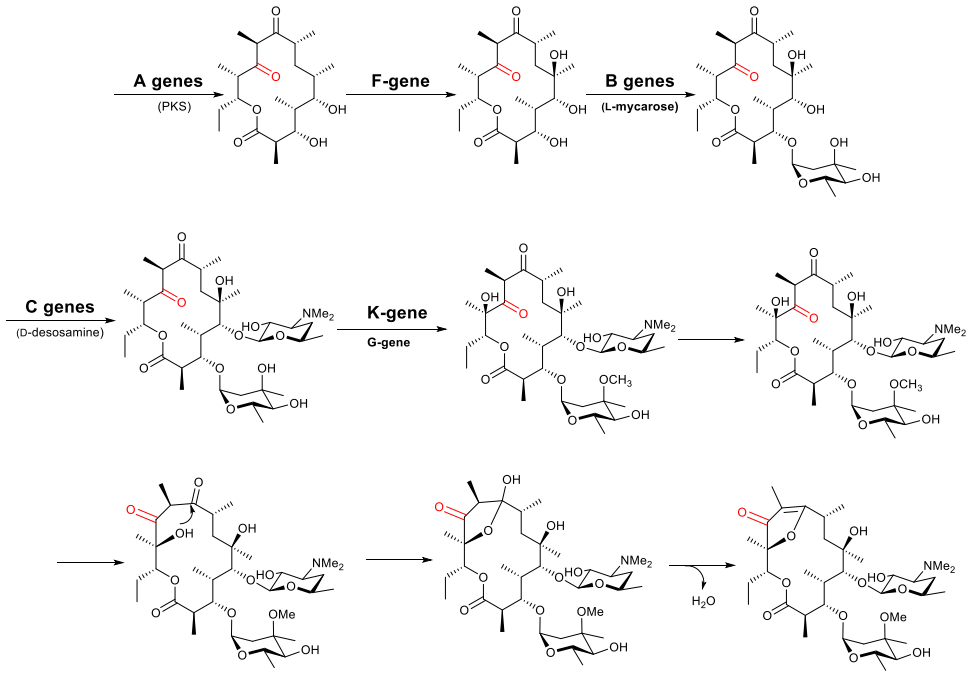



**Figure 6.2: Proposed scheme for biosynthesis of sporeamicin A based on the known biosynthesis of erythromycin A.** Biosynthetic genes are described in Chapter 1.

Our second hypothesis was that erythromycin A is the original product of the pathway and sporeamicin A is generated after oxidation of the C11-hydroxyl group to a keto-group by oxidative enzymes such as a short-chain dehydrogenase/reductase (SDR)-like or a cytochrome P450 monooxygenase (Figure 6.3).

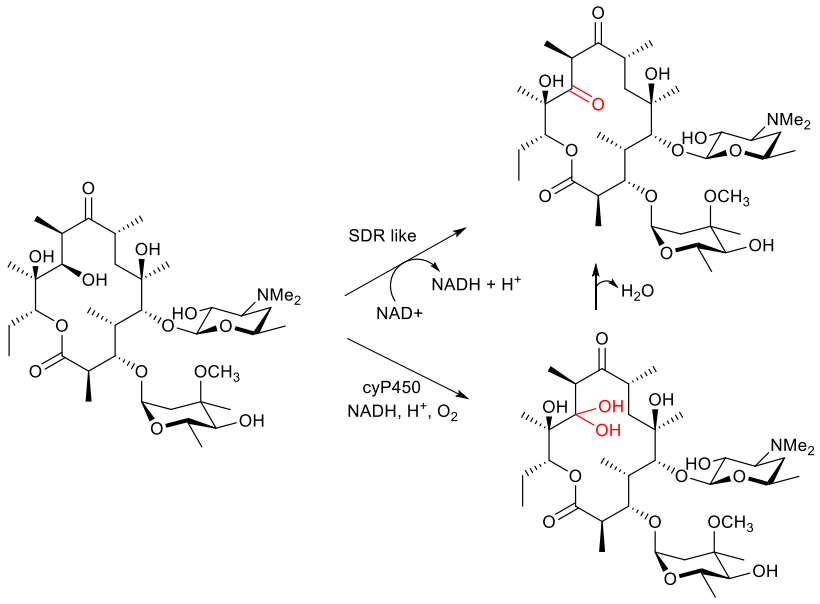



**Figure 6.3: Proposed scheme of post-PKS based synthesis of sporeamicin A.** Erythromycin A is oxidised by either SDR-like or cytochrome P450 enzymes, with subsequent ring formation and dehydration to form sporeamicin A.

Either of these pathways would be interesting and to distinguish between our two hypotheses we firstly needed to establish the production of sporeamicin A in the lab. Then we aimed to use a genomics-led approach to identify the basis for the differences between the two compounds by sequencing *Saccharopolyspora* sp. L53-18 and identifying the BGC.

## 6.2 Objectives

- Analyse the production of sporeamicin A by *Saccharopolyspora* sp. L53-18 (FERM BP-2231).
- Sequence the genome of *Saccharopolyspora* sp. L53-18 and identify the sporeamicin A BGC.
- Compare the sporeamicin A and erythromycin A BGCs to identify the novel enzymes that lead to sporeamicin A production.

## 6.3 Characterisation of *Saccharopolyspora* sp. L53-18

*Saccharopolyspora* sp. L53-18 was deposited in the Japanese Culture Collection as strain FERM BP-2231 (Yaginuma et al., 1992). Due to its inclusion in granted European Patent EP 0379 375 we were able to obtain a sample of the strain after application to the European Patent Office.

Prior to arrival of the strain from Japan we acquired a sample of the strain from Isomerase Therapeutics (Cambridge, UK) to speed up our work. Strain L53-18 grows readily on SFM solid medium where it produces colonies with white aerial mycelium that later develop spores. After several days it produces a diffusible red pigment that is characteristic of other *Saccharopolyspora* strains such as *S. erythraea* (NRRL2338), the wild type producer of erythromycin A. This red pigment has been identified as flaviolin (Cortés et al., 2002). Also like *S. erythraea*, strain L53-18 exhibits a relatively high rate of spontaneous pleotropic mutation that leads to mixed populations of grey and red colonies (Cortés et al., 2002).

We next investigated the morphology of strain L53-18 in more detail using scanning electron microscopy (SEM). The SEM analysis of the strains was performed by Mrs Elaine Barclay. We could then compare this to other *Saccharopolyspora* strains such as *S. erythraea* or *S. spinosa* (ATCC 49460$^T$), the producer of the insecticide spinosyn A (Kirst, 2010) (Figure 6.4), with images of other *Saccharopolyspora* strains provided from the Digital Atlas of Actinomycetes 2 (Society for Actinomycetes, 2013). In this way we could observe that these three strains share the presence of a characteristic spiny sheath covering their spore chains. The spore chains variate in the thickness of the hyphae and the length of the spores but all 3 strains share this appearance of thorns.

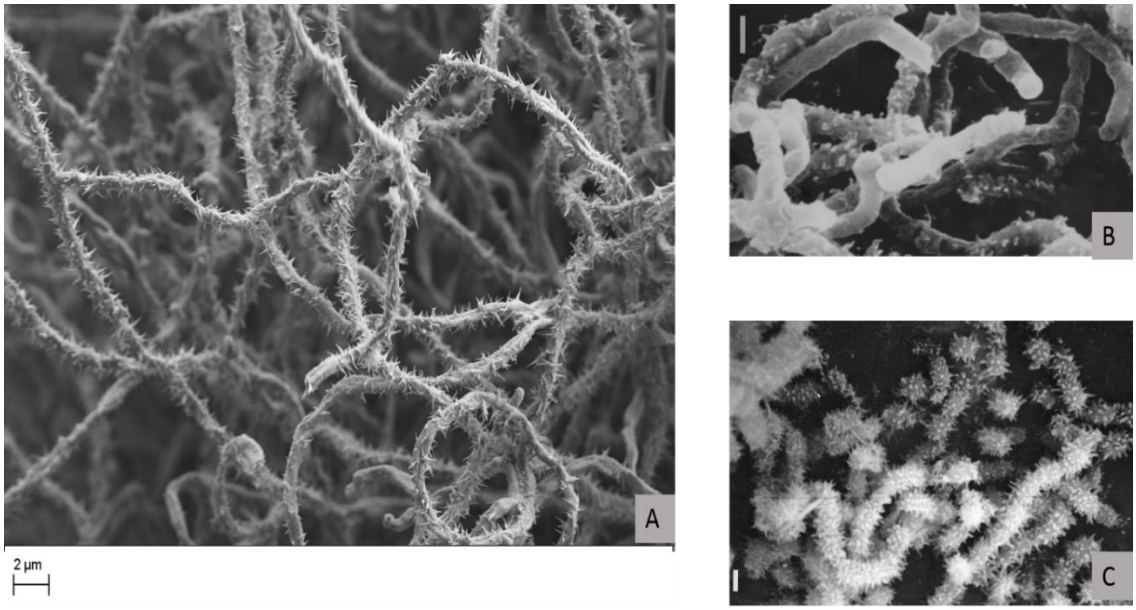



**Figure 6.4: Scanning electron microscopy (SEM) images of A: *Saccharopolyspora sp.* L53-18 (producer of sporeamicin A), imaged by Mrs Elaine Barclay (JIC) from a sample on R5 agar; B: *Saccharopolyspora erythraea* ATCC49460$^T$ by K. Suzuki; C: *Saccharopolyspora spinosa* ATCC49460$^T$ by R.C. Yao & F.P. Mertz.** Images B and C provided by Society for Actinomycetes (2013), http://atlas.actino.jp/. Through these images we observe that the *Saccharopolyspora* strains have a distinct spiny wreath covering their spore chains. The spiny wreath varies in length and width, depending on the size of the spore chains.

## 6.4 Growth of strain L53-18 and analysis of sporeamicin A production

Cultures of *S.* sp. L53-18, were grown in sporeamicin seed medium for three days and transferred to sporeamicin production medium (both used in Yaginuma et al., 1992), in 10 mL duplicates and grown at 30 °C for seven days. The samples were prepared by extraction with methanol and analysed by HPLC-MS (Figure 6.5). The total ion chromatograph (TIC) produced revealed the presence of two distinct peaks at retention times shortly after 6.5 and 7.2 min. The first peak appeared to be erythromycin A (*m/z* calculated for $C_{37}H_{67}NO_{13}$ $[M + H]^+$ = 734.4685; observed *m/z* = 734.4717; Δ = 4.3569 ppm) and was not expected based on the literature report. The second peak corresponded to sporeamicin A (calculated *m/z* for $C_{37}H_{63}NO_{12}$ $[M + H]^+$ = 714.4419; observed *m/z* = 714.4463; Δ = 6.1587 ppm).

Surprisingly, sporeamicin A was not the main metabolite being produced by the strain with erythromycin A being the major metabolite detected, at an approximate ratio of 2:1 with respect to sporeamicin A.

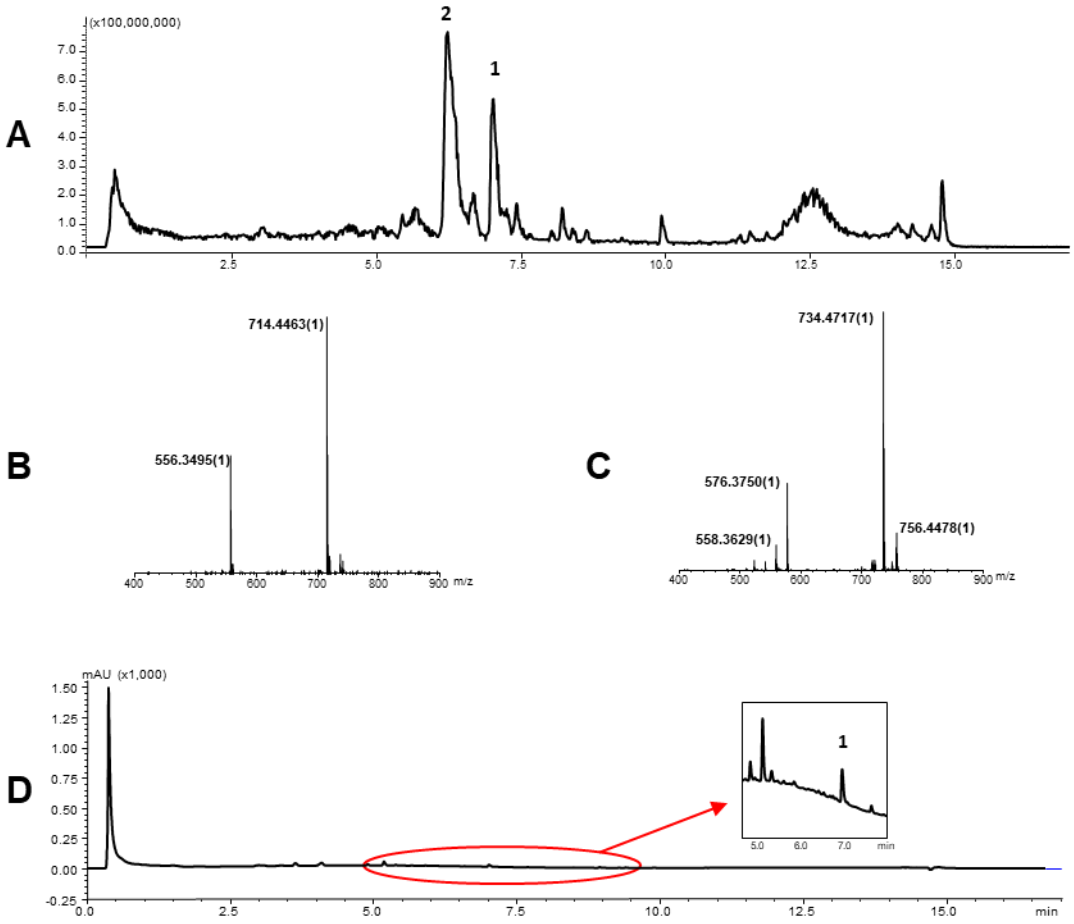



**Figure 6.5 Production of sporeamicin A and erythromycin A by strain L53-18.** A: Total Ion Chromatogram (TIC) of an extract shows peaks corresponding to sporeamicin A (peak 1) and erythromycin A (peak 2). B: Mass spectrum of sporeamicin A with a *m/z* of 714.4463 corresponding to the $[M + H]^+$ adduct. C: Mass spectrum of erythromycin A with a *m/z* of 734.4717 corresponding to the $[M + H]^+$ adduct. D: DAD signal at 276 nm. The inset shows absorbance at 276 nm between 5 and 8 minutes with a peak at 7.2 min that corresponds to sporeamicin A (peak 1). The results were reproducible.

After arrival of authentic strain L53-18 from the Japan Culture Collection named FERM BP-2231, we cultured it in parallel with the L53-18 strain under the same conditions and observed the same results.

As noted above there was no indication of the co-production of erythromycin A in the primary reports describing the strain L53-18 and identification of sporeamicin A as a fermentation product. This was surprising, and we therefore decided to understand if sporeamicin A might also be a product of the erythromycin producer *S. erythraea* NRRL2338. An authentic isolate of this strain was obtained from Isomerase Therapeutics and grown alongside strain L53-18 using both sporeamicin A (Yaginuma et al., 1992) and erythromycin A (Pacey et al., 1998) production media and conditions. Under all conditions tested *S. erythraea* NRRL2338 produced only erythromycin A, whereas the two isolates of the sporeamicin producer (L53-18 and BP-2231) produced a mixture of sporeamicin A and erythromycin A (Figure 6.6).

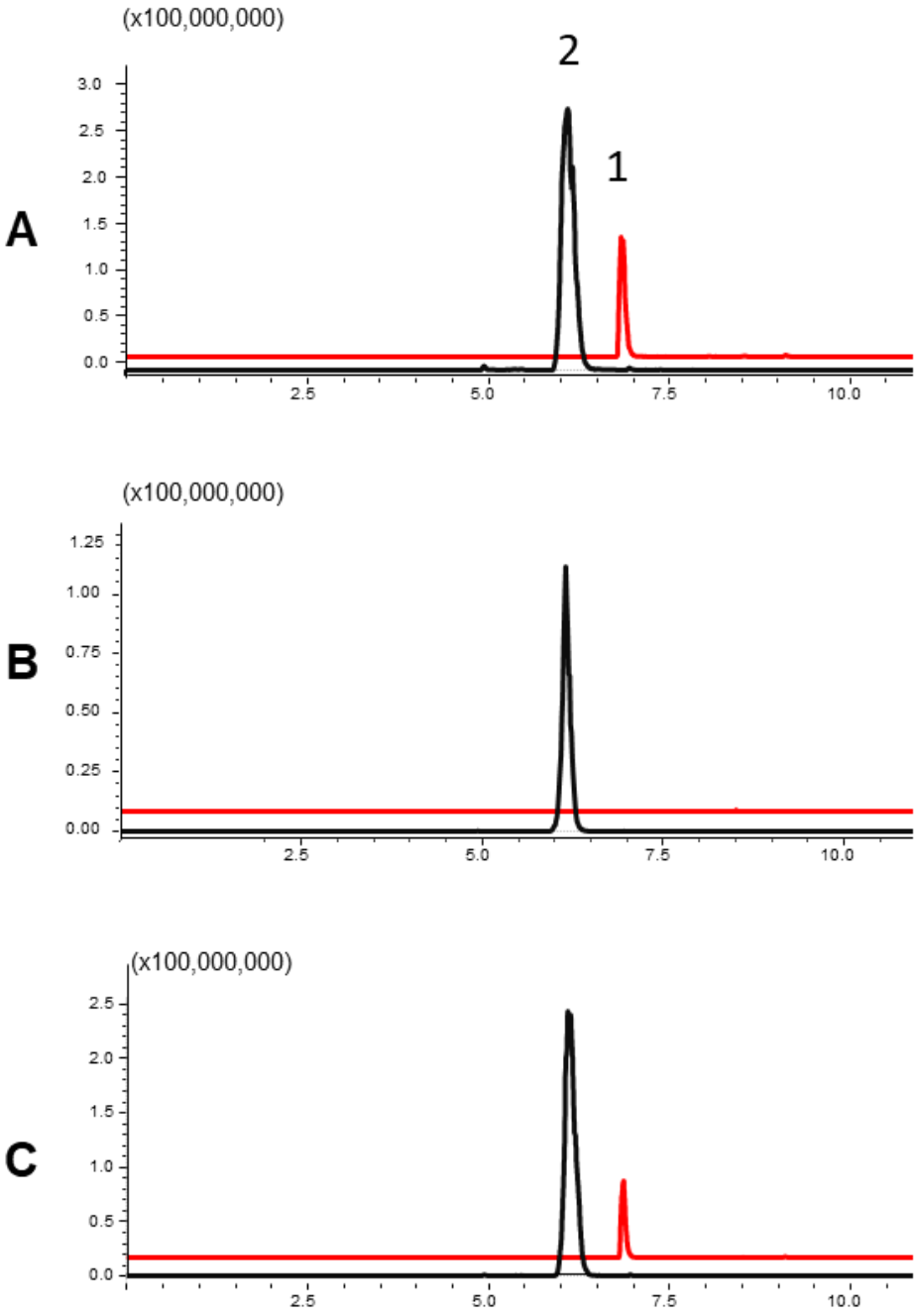



**Figure 6.6: Extracted Ion Chromatograms (XIC) of fermentation extracts from A: *Saccharopolyspora* sp. L53-18 (Isomerase isolate); B: *S. erythraea* NRRL2338 wild type and C: FERM BP-2231 (Japanese isolate)** containing peaks corresponding to sporeamicin A in red (1) and erythromycin A (2) in black. *S. erythraea* wild type produced only erythromycin A, whereas the two sporeamicin producers (L53-18 and BP-2231) produced a mixture of sporeamicin A and erythromycin A.

This behaviour was observed consistently even after testing different culture volumes, sampling time points and crossing over the production media (growing *S. erythraea* in strain L53-18 media and vice versa). We then subjected the two L53-18 isolates to two rounds of single colony isolation followed by cultivation and extract analysis. However, once again there was no change in the ratio of erythromycin A to sporeamicin A. Since strain L53-18 produces erythromycin A as well as sporeamicin A, it suggested that both metabolites have the same origin and that sporeamicin A is likely to be a product of post-PKS modification of erythromycin A, supporting our second hypothesis. On this basis we focused all further work on the authentic isolate of strain L53-18 (FERM BP-2231) obtained directly from Japan.

## 6.5 Sequencing of the *Saccharopolyspora* sp. L53-18 (FERM BP-2231) genome

High molecular weight genomic DNA of strain L53-18 was extracted as described in Chapter 2 (Kieser et al., 2000). This was sent to the Earlham Institute (Norwich, UK) for sequencing on the Pacific Biosciences (PacBio) platform using SMRT technology and assembled using the HGAP3 and HGAP2 pipelines (Chin et al., 2013). Both assembly algorithms offered comparable results. The assembly generated with HGAP3 was selected for further bioinformatic analyses. This yielded two contigs of 8335144 bp and of 29864 bp. The larger contig represents the whole circular chromosome of the bacterium, whereas the smaller contig does not overlap with the larger contig or contain the type of repetitive regions usually associated to misassembled sequences (Gomez-Escribano et al., 2016) meaning it was unlikely to be a sequencing artefact. A contig representing a circular molecule is easily identifiable because one end is directly repeated at the other end, indicating an overlapping stretch of sequence. In order to obtain a full non-repetitive sequence representing the circular chromosome the direct repeat of the ends was removed.

## 6.6 Comparison of *Saccharopolyspora* sp. L53-18 and *S. erythraea* genome sequences

### 6.6.1 Genome comparison

A comparative analysis was performed between the L53-18 genome sequence and the genome sequence of *S. erythraea* NRRL2338 which was previously deposited under accession number NC_009142.1 (Oliynyk et al., 2007). BLASTn analysis showed a high level of synteny with NC_009142.1. Strain L53-18 appears to have two inversions from approximately 2.5M to 3.4M and from 5.3M to 5.8M bases, as shown in Figure 6.7. These inversions reflect the fact that two DNA regions on L53-18 have been inverted in relation to NC_009142.1. 16S ribosomal RNA data showed 99% similarity to the erythromycin A producer. Strain L53-18 has an average GC percentage of 71.1% which is nearly identical to *S. erythraea* (reported 71.15 GC%). *S. erythraea* contains the erythromycin BGC and strain L53-18 has an erythromycin-like cluster with 91% similarity based on the antiSMASH output.

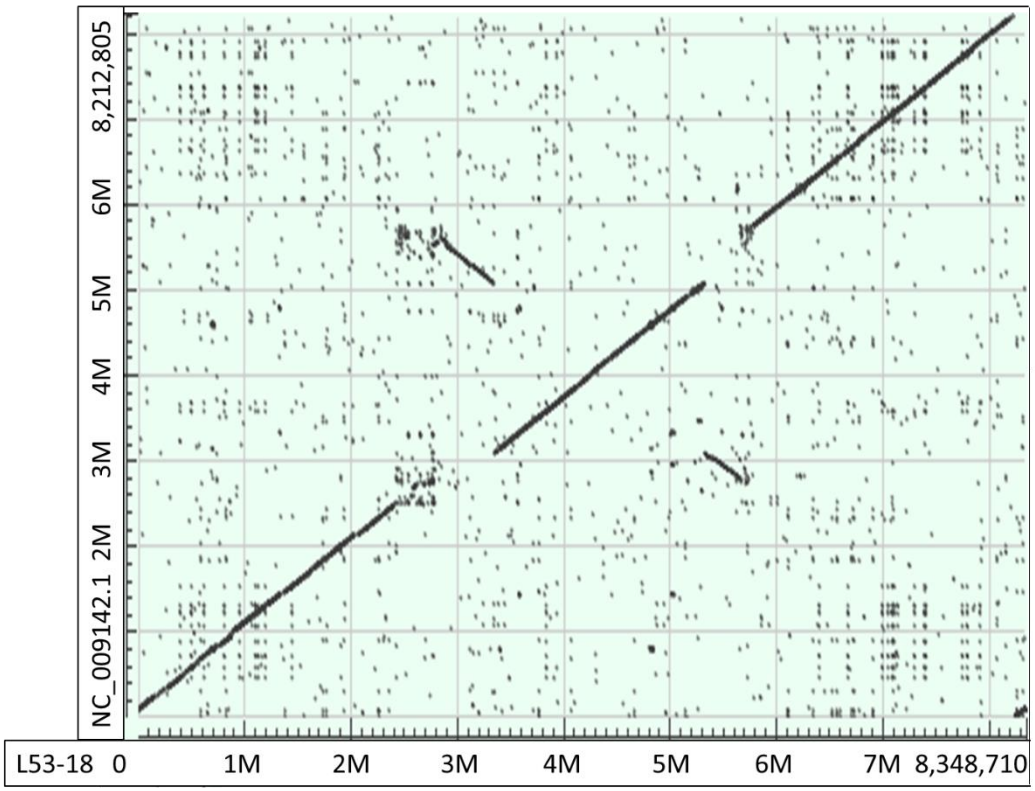



**Figure 6.7: BLASTn generated dot plot comparing *Saccharopolyspora* sp. L53-18 to *S. erythraea* NRRL2338.** This dot matrix shows regions of similarity based upon the BLASTn results comparing the two sequences. The query sequence of *Saccharopolyspora* sp. L53-18 is represented on the x-axis, the subject *S. erythraea* NRRL2338 is represented on the y-axis and the numbers represent the bases of the

query and the subject respectively. Matching regions are shown in the plot as lines. Plus strand matches are slanted from the bottom left to the upper right corner, minus strand matches are slanted from the upper left to the lower right. The number of lines shown in the plot is the same as the number of matching regions found by BLASTn.

A detailed analysis of the erythromycin-like BGC encoded by strain L53-18 showed that it is nearly identical to that published for erythromycin A in *S. erythraea* NRRL2338 (Figure 6.8). The same genes are present in both cases and they are organised in the same way. The domain and module organisation of the PKS region is conserved between the two strains, including the distribution and nature of the β-ketoreductase (KR) domains within them. Furthermore, the ketoreductase (KR) domains from the sporeamicin A cluster are highly similar to their erythromycin A counterparts, with the active site residues all conserved. This similarity extends to KR3, which in the sporeamicin cluster presents the same mutations in the active residues (particularly an asparagine to serine substitution) that render KR3 from the erythromycin PKS inactive (Starcevic et al., 2007) (Figure 6.9). The acyltransferase (AT) domains of each module also possess the same substrate specificity, as calculated by MINOWA prediction. For example, the active site motifs correspond to selection of malonyl and 2S-methylmalonyl-CoA substrates (Minowa et al., 2007). These observations suggested that it is likely that the sporeamicin PKS would generate 6-deoxyerythronolide B, the aglycone intermediate of the erythromycin biosynthetic pathway. After checking the genes surrounding the sporeamicin BGC, no other gene that is part of the cluster encodes an oxidative enzyme that might be responsible for sporeamicin production.

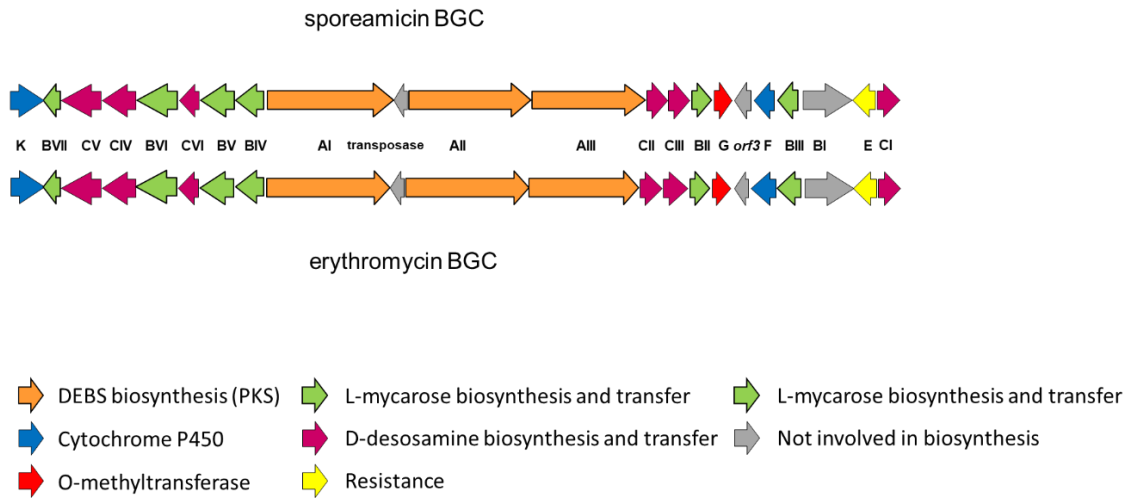



**Figure 6.8: Organisation of the sporeamicin A and erythromycin BGCs.**

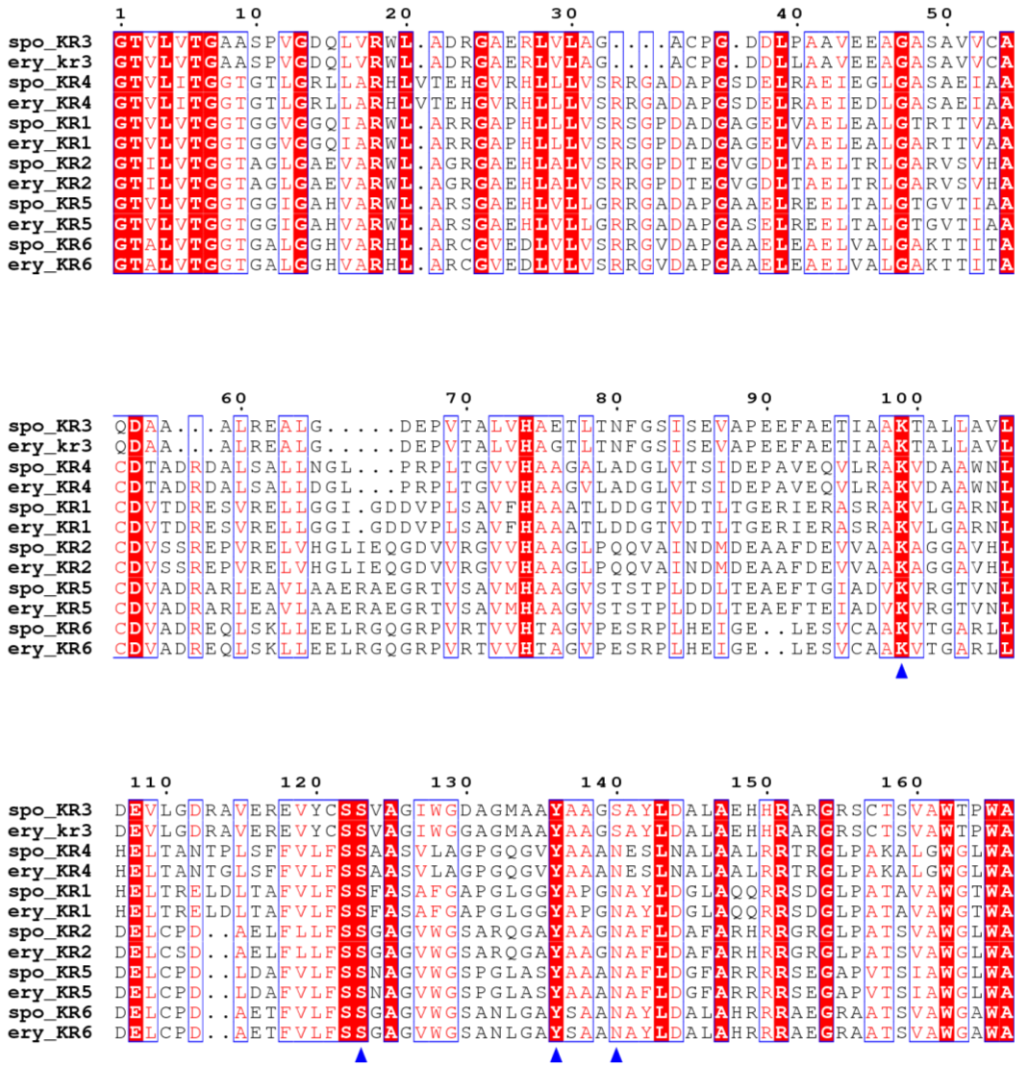



**Figure 6.9: Multiple alignment of the ketoreductase (KR) domains of the erythromycin (ery) and sporeamicin A (spo) PKSs.** White letters over red background represent full conservation of the residue across all sequences. Blue arrows indicate the residues from the active site of the domain. Note that the conserved asparagine residue at position 140 has been mutated to serine in both KR3 domains. This alignment was generated with the MUSCLE tool (Edgar, 2004) from the EMBL-EBI server (Li et al., 2015) and displayed with ESPript 3.0 (Robert and Gouet, 2014).

## 6.7 Disruption of sporeamicin A biosynthesis

I proceeded to investigate the biosynthetic pathway of sporeamicin A in strain L53-18 through mutational analysis and biotransformation experiments.

### 6.7.1 Mutant generation

Disruption mutants in the sporeamicin PKS were generated by standard single-crossover methods. A 2.3 kb fragment from the centre of the *spoAII* (equivalent to *ery*AII) gene was amplified by PCR and cloned into the suicide vector pKC1132 (pOJ260 based) that carries the apramycin marker, as described in Chapter 2 (Bierman et al., 1992). The resulting plasmid, pEVK13, should accomplish the disruption of the target gene by single crossover homologous recombination of the cloned region and the sporeamicin PKS leading to insertion of the suicide vector into the chromosomal DNA. pEVK13 was introduced in strain L53-18 via intergeneric conjugation from *E. coli* ET12567/pUZ8002. Of the 12 ex-conjugants observed, nine of them formed white-grey colonies and three formed red colonies. This is a phenomenon unrelated to the mutation as the red variants represent a class of spontaneous pleiotropic mutants, as mentioned above (Cortés et al., 2002). Six of the ex-conjugants, two red and four white, were taken through for further analysis.

### 6.7.2 Metabolite analysis of the PKS disrupted mutants

The six selected isolates from section 6.7.1 were grown in the same seed and production media used previously for sporeamicin A production. After seven days of growth, samples of fermentation broth were extracted and analysed by LCMS. We observed that the production of sporeamicin A and erythromycin A was abolished in these strains. As expected, this confirms that that both sporeamicin A and erythromycin A share the same biosynthetic origin.

To determine whether the mutants could be chemically complemented and gain more insights in the biosynthesis of sporeamicin A, they were fed with a sample of deoxyerythronolide B, the direct product of the erythromycin PKS, to yield a final concentration of 100 µM (Gaisser et al., 2000). Surprisingly, the production of sporeamicin A and erythromycin A was not restored, suggesting that the compound could not be either internalised or incorporated into the pathway. A likely explanation could be that the mutation had resulted in undesired polar effects. Insertional inactivation often affects the expression of the downstream genes in a cluster and this could be especially possible in this case since *ery*AI, *ery*AII, *ery*AIII, *ery*BII, *ery*CII, *ery*CIII, and *ery*G genes are primarily co-transcribed from a promoter upstream of *ery*AI (Reeves et al., 1999) and we suggest that the same is likely to happen in the sporeamicin cluster. If this is the case, then disruption of the transcriptional unit would affect the expression genes involved in the biosynthesis or attachment of both the L-mycarose (B genes) and

D-desosamine (C genes) deoxysugars (Figure 6.2). Such a transcriptional scenario would mean that exogenously added 6-deoxyerythronolide would not be further modified as observed. Therefore, further work is required to generate a mutant without undesired polar effects. Despite this, our experiment was successful in yielding a mutant lacking the ability to produce sporeamicin A or erythromycin A directly.

## 6.8 Discussion

Sporeamicin A had previously been reported as an erythromycin-like macrolide antibiotic produced by *S.* sp. L53-18 with improved clinical properties with respect to erythromycin (Morishita et al., 1992). Also, the structural differences between sporeamicin A and erythromycin suggested that the biosynthetic enzymes for producing sporeamicin A might have altered substrate specificity that would make them valuable for the bioengineering of new analogues of erythromycin. Therefore, it was of great interest to us to characterise the biosynthesis of sporeamicin A. After following the growth conditions published for the production of sporeamicin A by strain L53-18 we confirmed its production. However, we were surprised by the finding that erythromycin A was actually the main metabolite produced in this experiment. Following this discrepancy, which had not been mentioned before in the literature (Yaginuma et al., 1992), we focused our efforts on a genomic investigation to gain further insight into the biosynthesis of sporeamicin A.

Genome sequencing of strain L53-18 identified an erythromycin-like BGC (*spo* BGC) that we presumed would be responsible for the biosynthesis sporeamicin A. Comparative bioinformatics analysis revealed that the *spo* BGC is essentially identical to the BGC for production of erythromycin A in *S. erythraea*. Based on *in silico* analysis there was no evidence to support that this cluster would produce an aglycone scaffold different to the one that leads to erythromycin A. We also examined the sequence flanking the *spo* BGC and no genes were present that would indicate expression of an oxidative enzyme for the conversion of erythromycin A into sporeamicin A. In fact, there was no obvious difference at the genome level in these flanking regions – this strongly suggested that any gene product responsible for the production of sporeamicin A from erythromycin A must be encoded somewhere else in the genome. The absence of any other compatible BGC in the genome indicated that both sporeamicin A and erythromycin A should come from the same (*spo*) BGC.

In accordance with this hypothesis insertional disruption of the *spo* PKS abolished the production of both sporeamicin A and erythromycin A. Whilst this confirmed their common biosynthetic origin, it was unclear whether sporeamicin A derives from erythromycin A or a common intermediate, and how this modification is introduced. The exogenous addition of 6-deoxyerythronolide B, the first product of the erythromycin PKS, to growing cultures of the *spo* PKS mutant failed to initiate biosynthesis of erythromycin A, although this is probably due to a polar effect on downstream genes which mean the strain can no longer make or attach the deoxysugars L-mycarose (B genes) and D-

desosamine (C genes) (Staunton and Wilkinson, 1997). Further experiments will aim to construct a mutant without polar effects.

Finally, there is an opportunity to TAR clone the sporeamicin A BGC and transfer it to a heterologous host ideal for scale up, such as the *S. erythraea* superhost described earlier in Chapter 3. This might enable the isolation of sporeamicin A if the PKS has a subtle difference that we cannot detect, or it will allow us to further investigate how other modifications occur to the chemical structure.

# Chapter 7:

# Discussion

# 7 Discussion

The focus of this thesis project was to investigate the biosynthetic potential of the *Saccharopolyspora* KY strains isolated from Kenyan fungus-farming *Tetraponera* plant ants. Throughout the project a tripartite exploration of genomic, metabolomic and bioassay-guided approaches was exploited. Based on my investigations four major themes were developed:

1. Activation of the kyamicin BGC and characterisation of its product
2. Engineering of a synthetic platform for the expression of type B lantibiotics
3. Isolation and structural determination of the new pseudo-peptide EV60 produced only in minimal media
4. Investigation of the biosynthesis of sporeamicin and how it differs from erythromycin

## 7.1 Kyamicin and platform for expression of type B lantibiotics

The external microbiome of the *Tetraponera* ant cuticles is heterogeneous and dominated mainly by proteobacteria and firmicutes. Actinobacteria make up only a small percentage of the microbiome and that was the intriguing point for the beginning of this research based on the three *Saccharopolyspora* sp. strains: KY3, KY7 and KY21 (Seipke et al., 2013). This finding was particularly interesting given that *Saccharopolyspora* is a rare genus that contains strains which are prolific producers of important anti-infective agents.

I set out to the study this rare genus from this underexploited ecological niche, firstly using a genomic-led approach. The three genomes were found to be highly similar and encode approximately 23 BGCs with very significant overlap. As described, all three encoded a BGC with identical architecture, for a cinnamycin-like type lanthipeptide that became the focus of the genomic-led investigation. The fact that three different strains, sampled from two different locations, possess clusters for the same antibiotic suggests that either these compounds, or a closely linked trait, confer an advantage in this environment, or that the ants are selecting them for these traits.

I reported the genomics-directed discovery of kyamicin. Kyamicin is a new member of the type B lantibiotic family, which includes cinnamycin and duramycin. As discussed in

this study, this is a potentially valuable finding since lantibiotics are a family of NPs with important activities as antibiotics, inhibitors of viral entry or even therapeutics for cystic fibrosis (Willey and Van Der Donk, 2007, An et al., 2018).

The activation of the kyamicin BGC in the native host and heterologously was crucial to examine the potential for cross species activation of similar BGCs and worked as a proof of concept for this work. This led to the engineering of a type B lantibiotics platform that allowed for the expression of the duramycin and the cinnamycin B clusters, as described in chapter 4. The ability to generate analogues of cinnamycin-like type B lantibiotics with modified properties is of great importance and value, given the significant potential of such NPs in numerous and demanding therapeutic contexts.

Expanding the application of this platform will be the focus of future research. The work described in this study provides a basis for the identification of additional natural lantibiotics, whose biosynthesis is cryptic in the host strain, and for the diversification of their chemical structures to generate new-to-nature molecules. Future lines of research include determining how many of the lantibiotic core peptide sequences uncovered by out bioinformatics analysis can be expressed using this platform, and how far can we stretch the system with unprecedented modifications of the core peptide sequences using a synthetic approach. If the synthetic lantibiotics deriving from this work can be made in useful quantities, and prove to have potentially useful properties, then they may be moved forward to preclinical assessment. One day they may even expand the toolbox of novel antimicrobials.

## 7.2 Untargeted metabolomics

After following an untargeted metabolomics approach, I reported the discovery of EV60, an NRPS-derived siderophore composed of 2,3-dihydroxybenzoic acid, D-threonine and $N^5$-hydroxy-$N^5$-acetyl-ornithine (stereochemistry to be determined).

Parallel to our work, colleagues from Dr Andrew Truman's group were working with the strain *Actinomadura atramentaria* DSM43919. While investigating the biosynthetic pathway of the matlystatins, metalloproteinase inhibitors produced by the strain and reported in previous work (Leipoldt et al., 2017), they made a peculiar observation. During the course of a series of nucleophile feeding experiments, they observed intense phenotypic changes. Subsequent metabolomics analysis revealed the production of several new peaks that could be related to each other by metabolic networking analysis. The further purification and structural elucidation of four of these compounds revealed a pseudo-tripeptide scaffold identical to EV60 but carrying several additional post-NRPS

modifications. Unfortunately, *A. atramentaria* DSM43919 is not genetically tractable. For that reason, we passed onto them the strain KY21 in order to study the biosynthesis of this new family of molecules.

## 7.3 Bioassay-guided fractionation

Bioassay-guided fractionation proved less effective for us in identifying and isolating an antifungal compound that was suggested from overlay bioassays. Our efforts towards scaling up the production of this compound were not successful. Nevertheless, it was a potentially valuable finding that will be investigated further in the future. Isolating new antifungal molecules is of great significance since there is an urgent need for antifungal agents. The number of NPs available to treat life-threatening, invasive fungal infections available is significantly smaller than the ones with antibacterial activity (Roemer and Krysan, 2014). Even though the field of antifungal discovery has not been completely static, since early last century only three classes of antifungal drugs have been developed (Krysan, 2017). Those are the polyenes such as amphotericin B (Herbrecht et al., 2002), the azoles such as fluconazole (Goodman et al., 1992), and the echinocandins such as micafungin (Pfaller et al., 2008). For these reasons, it is of absolute importance to investigate a potential new antifungal agent, deriving from the underexploited environmental niche of the *Tetraponera* ants.

## 7.4 Sporeamicin

The genomic-led approach for investigating the biosynthesis of sporeamicin gave unclear results. Next steps will include exogenously feeding the mutant strain described in chapter 6 with erythromycin and monitoring the production results. The most probable scenario is that after providing erythromycin, some of it will be biotransformed to sporeamicin, supporting our theory that the difference between the two is the result of a post-PKS enzymatic modification. Other approaches we can employ in the future are to try deleting the whole BGC and/or transferring the BGC into a heterologous host such as the *Saccharopolyspora* superhost described in chapter 3. In the second scenario, the strain carrying just the erythromycin-like *spo* BGC, should produce only erythromycin if our hypothesis that the gene responsible for sporeamicin formation is encoded outside of the *spo* BGC.

## 7.5 Conclusions

Based on the experience from this work some broader comments can be made about the three main approaches utilised for NP discovery. Firstly, the genomic-led approach usually starts from a point of many different BGCs that are predicted to encode for a variety of NPs, immediately offering many targets to follow. Importantly, it gives a full picture of the biosynthetic capacity of the strain, revealing the ability for NPs that might be considered silent or cryptic by the other approaches. This then allows for genetic strategies for the activation of cryptic BGCs, for example by the artificial expression of positive activators, or by heterologous expression, as demonstrated in this thesis.

Usually, the re-discovery problem can be avoided by excluding BGCs that encode known products. Furthermore, since predictions are based on known BGCs, this can bias efforts towards the investigation of different variations of known compounds. This could potentially limit the discovery of compounds arising from novel genetic architecture not identified with BGC prediction tools, which would be likely to encode compounds with different structures and/or activities. This issue can be overcome by the design of bespoke scripts to find genes predicted to be responsible for the synthesis a compound of interest. As BGC prediction tools advance this problem may become less of an issue.

Alternatively, the traditional bioassay-guided isolation of NPs ensures that efforts are focused on compounds with the desired bioactivities, which is not guaranteed by the other two approaches discussed here. When utilising the bioassay-guided approach it is important to investigate as many different culture media as possible to avoid missing compounds that are only expressed in certain environmental conditions. Likewise, a number of different bioindicator strains should be used so that all potential bioactivities of interest are covered. However, as mentioned previously, the major drawback of this method is the re-discovery of known compounds.

Untargeted metabolomic approaches circumvents the potential biases of the approaches above, by considering all metabolites regardless of bioactivity or putative BGCs. As with the bioassay-guided approach, it is important to utilise a wide range of different media to try and cover the full metabolomic capacity of the strain of interest. It is also important to carefully check data to exclude *m/z* peaks that are just in the noise or arise from media components, and to consider all possible adducts that might be observed. Although modern UHPLC-HRMS techniques have very high levels of accuracy and sensitivity, it is important to bear in mind that the size and chemical properties of a molecule can affect their detection by MS. One other consideration for this approach is that the identification

of a NP for an *m/z* of interest can be challenging, mainly due to limitations in the available MS databases. The latest addition to untargeted metabolomics is metabolic networking, which will certainly help advance this approach, although again it can be limited by current *m/z* databases, which are still not sophisticated enough to cover the plethora of metabolic data available.

Whilst it is clear these three approaches can work together and complement each other, matching a NP identified by bioactivity or mass to a BGC is not straightforward. This is especially true if mutational experiments in the producing organism are challenging, such as in strains that are not genetically tractable. In many cases, heterologous expression of the putative BGC is necessary to confirm it encodes a particular NP and this can be a lengthy process.

To tackle the rising spectre of a post-antibiotic era, it is clear that we must use all the tools at our disposal in order to discover novel NPs more efficiently.

# Bibliography

**Abe, T., Hashimoto, Y., Sugimoto, S., Kobayashi, K., Kumano, T. & Kobayashi, M.** 2017. Amide compound synthesis by adenylation domain of bacillibactin synthetase. *J Antibiot (Tokyo),* 70**,** 435-442.

**Akbar, N., Siddiqui, R., Iqbal, M., Sagathevan, K. & Khan, N. A.** 2018. Gut bacteria of cockroaches are a potential source of antibacterial compound (s). *Letters in applied microbiology,* 66**,** 416-426.

**Alexander, D. & Zuberer, D.** 1991. Use of chrome azurol s reagents to evaluate siderophore production by rhizosphere bacteria. *Biology and Fertility of soils,* 12**,** 39-45.

**An, L., Cogan, D., Navo, C., Jimenez-Oses, G., Nair, S. & Van Der Donk, W.** 2018. Substrate-assisted enzymatic formation of lysinoalanine in duramycin. *bioRxiv*.

**Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K. D., Fischbach, M. A., Garavelli, J. S., Goransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Muller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J., Rebuffat, S., Ross, R. P., Sahl, H. G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Sussmuth, R. D., Tagg, J. R., Tang, G. L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M. & Van Der Donk, W. A.** 2013. Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Nat Prod Rep,* 30**,** 108-60.

**Aziz, R. K., Bartels, D., Best, A. A., Dejongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., Mcneil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. & Zagnitko, O.** 2008. The rast server: Rapid annotations using subsystems technology. *BMC Genomics,* 9**,** 75-75.

**Bachmann, B. O., Van Lanen, S. G. & Baltz, R. H.** 2014. Microbial genome mining for accelerated natural products discovery: Is a renaissance in the making? *J Ind Microbiol Biotechnol,* 41**,** 175-84.

**Bai, Y., Müller, D. B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., Dombrowski, N., Münch, P. C., Spaepen, S. & Remus-Emsermann, M.** 2015. Functional overlap of the arabidopsis leaf and root microbiota. *Nature,* 528**,** 364.

**Baldeweg, F., Hoffmeister, D. & Nett, M.** 2018. A genomics perspective on natural product biosynthesis in plant pathogenic bacteria. *Natural product reports*.

**Baltz, R. H.** 2006. Marcel faber roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *Journal of Industrial Microbiology and Biotechnology,* 33**,** 507-513.

**Baltz, R. H.** 2008. Renaissance in antibacterial discovery from actinomycetes. *Current Opinion in Pharmacology,* 8**,** 557-563.

**Baltz, R. H.** 2017. Gifted microbes for genome mining and natural product discovery. *Journal of industrial microbiology & biotechnology,* 44**,** 573-588.

**Banerjee, S. & Hansen, J. N.** 1988. Structure and expression of a gene encoding the precursor of subtilin, a small protein antibiotic. *Journal of Biological Chemistry,* 263**,** 9508-9514.

**Barreales, E. G., Vicente, C. M., De Pedro, A., Santos-Aberturas, J. & Aparicio, J. F.** 2018. Promoter engineering reveals the importance of heptameric direct repeats for DNA-binding by sarp-lal regulators in streptomyces. *Applied and environmental microbiology***,** AEM. 00246-18.

**Bender, C. L., Alarcón-Chaidez, F. & Gross, D. C.** 1999. *Pseudomonas syringae* phytotoxins: Mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiology and molecular biology reviews,* 63**,** 266-292.

**Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N., James, K. D., Harris, D. E., Quail, M. A., Kieser, H. & Harper, D.** 2002. Complete genome sequence of the model actinomycete *streptomyces coelicolor* a3 (2). *Nature,* 417**,** 141.

**Bérdy, J.** 2012. Thoughts and facts about antibiotics: Where we are now and where we are heading. *The Journal of antibiotics,* 65**,** 385.

**Berti, A. D. & Thomas, M. G.** 2009. Analysis of achromobactin biosynthesis by pseudomonas syringae pv. Syringae b728a. *Journal of bacteriology,* 191**,** 4594-4604.

**Bibb, M. J.** 2005. Regulation of secondary metabolism in streptomycetes. *Current opinion in microbiology,* 8**,** 208-215.

**Bibb, M. J., Janssen, G. R. & Ward, J. M.** 1985. Cloning and analysis of the promoter region of the erythromycin resistance gene (*erm*e) of *streptomyces erythraeus*. *Gene,* 38**,** 215-226.

**Bierman, M., Logan, R., O'brien, K., Seno, E., Rao, R. N. & Schoner, B.** 1992. Plasmid cloning vectors for the conjugal transfer of DNA from *escherichia coli* to *streptomyces* spp. *Gene,* 116**,** 43-49.

**Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Suarez Duran, H. G., De Los Santos, E. L., Kim, H. U. & Nave, M.** 2017. Antismash 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic acids research,* 45**,** W36-W41.

**Blodgett, J. a. V., Oh, D.-C., Cao, S., Currie, C. R., Kolter, R. & Clardy, J.** 2010. Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 11692-11697.

**Bode, H. B., Bethe, B., Höfs, R. & Zeeck, A.** 2002. Big effects from small changes: Possible ways to explore nature's chemical diversity. *ChemBioChem,* 3**,** 619-627.

**Boratyn, G. M., Schaffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J. & Madden, T. L.** 2012. Domain enhanced lookup time accelerated blast. *Biol Direct,* 7**,** 12.

**Borel, J. F. & Hiestand, P. C.** 1999. Immunomodulation: Particular perspectives. *Transplant Proc,* 31**,** 1464-71.

**Bosello, M., Robbel, L., Linne, U., Xie, X. & Marahiel, M. A.** 2011. Biosynthesis of the siderophore rhodochelin requires the coordinated expression of three independent gene clusters in *rhodococcus jostii* rha1. *J Am Chem Soc,* 133**,** 4587-95.

**Brandel, J., Humbert, N., Elhabiri, M., Schalk, I. J., Mislin, G. L. & Albrecht-Gary, A. M.** 2012. Pyochelin, a siderophore of *pseudomonas aeruginosa*: Physicochemical characterization of the iron(iii), copper(ii) and zinc(ii) complexes. *Dalton Trans,* 41**,** 2820-34.

**Butler, M. S.** 2008. Natural products to drugs: Natural product-derived compounds in clinical trials. *Natural product reports,* 25**,** 475-516.

**Caffrey, P., Bevitt, D. J., Staunton, J. & Leadlay, P. F.** 1992. Identification of debs 1, debs 2 and debs 3, the multienzyme polypeptides of the erythromycin-producing polyketide synthase from saccharopolyspora erythraea. *FEBS letters,* 304**,** 225-228.

**Carroll, C. S. & Moore, M. M.** 2018. Ironing out siderophore biosynthesis: A review of non-ribosomal peptide synthetase (nrps)-independent siderophore synthetases. *Critical reviews in biochemistry and molecular biology***,** 1-26.

**Chain, E., Florey, H. W., Gardner, A. D., Heatley, N. G., Jennings, M. A., Orr-Ewing, J. & Sanders, A. G.** 1940. Penicillin as a chemotherapeutic agent. *The lancet,* 236**,** 226-228.

**Challinor, V. L. & Bode, H. B.** 2015. Bioactive natural products from novel microbial sources. *Annals of the New York Academy of Sciences,* 1354**,** 82-97.

**Chatterjee, C., Paul, M., Xie, L. & Van Der Donk, W. A.** 2005. Biosynthesis and mode of action of lantibiotics. *Chemical Reviews,* 105**,** 633-684.

**Chaudhary, H. S., Soni, B., Shrivastava, A. R. & Shrivastava, S.** 2013. Diversity and versatility of actinomycetes and its role in antibiotic production. *Journal of Applied Pharmaceutical Science,* 3**,** S83-S94.

**Chavali, A. K., D'auria, K. M., Hewlett, E. L., Pearson, R. D. & Papin, J. A.** 2012. A metabolic network approach for the identification and prioritization of antimicrobial drug targets. *Trends Microbiol,* 20**,** 113-23.

**Chen, D., Feng, J., Huang, L., Zhang, Q., Wu, J., Zhu, X., Duan, Y. & Xu, Z.** 2014. Identification and characterization of a new erythromycin biosynthetic gene cluster in actinopolyspora erythraea yim90600, a novel erythronolide-producing halophilic actinomycete isolated from salt field. *PloS one,* 9**,** e108129.

**Chen, E., Chen, Q., Chen, S., Xu, B., Ju, J. & Wang, H.** 2017. Mathermycin, a lantibiotic from the marine actinomycete *marinactinospora thermotolerans*. *Applied and Environmental Microbiology,* 83.

**Chen, S., Huang, X., Zhou, X., Bai, L., He, J., Jeong, K. J., Lee, S. Y. & Deng, Z.** 2003. Organizational and mutational analysis of a complete fr-008/candicidin gene cluster encoding a structurally related polyene complex. *Chemistry & Biology,* 10**,** 1065-1076.

**Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W. & Korlach, J.** 2013. Nonhybrid, finished microbial genome assemblies from long-read smrt sequencing data. *Nature Methods,* 10**,** 563.

**Choudhary, A., Naughton, L. M., Montanchez, I., Dobson, A. D. W. & Rai, D. K.** 2017. Current status and future prospects of marine natural products (mnps) as antimicrobials. *Mar Drugs,* 15.

**Clatworthy, A. E., Pierson, E. & Hung, D. T.** 2007. Targeting virulence: A new paradigm for antimicrobial therapy. *Nature chemical biology,* 3**,** 541.

**Cooper, L. E., Li, B. & Van Der Donk, W. A.** 2010. Biosynthesis and mode of action of lantibiotics. *Comprehensive natural products ii: Chemistry and biology.* Elsevier Ltd.

**Cortés, J., Velasco, J., Foster, G., Blackaby, A. P., Rudd, B. a. M. & Wilkinson, B.** 2002. Identification and cloning of a type iii polyketide synthase required for diffusible

pigment biosynthesis in saccharopolyspora erythraea‡. *Molecular Microbiology,* 44**,** 1213-1224.

**Crone, W. J., Vior, N. M., Santos-Aberturas, J., Schmitz, L. G., Leeper, F. J. & Truman, A. W.** 2016. Dissecting bottromycin biosynthesis using comparative untargeted metabolomics. *Angewandte Chemie International Edition,* 55**,** 9639-9643.

**Desriac, F., Jégou, C., Balnois, E., Brillet, B., Chevalier, P. L. & Fleury, Y.** 2013. Antimicrobial peptides from marine proteobacteria. *Marine drugs,* 11**,** 3632-3660.

**Donia, M. S., Cimermancic, P., Schulze, C. J., Brown, L. C. W., Martin, J., Mitreva, M., Clardy, J., Linington, R. G. & Fischbach, M. A.** 2014. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell,* 158**,** 1402-1414.

**Dreyfuss, M., Härri, E., Hofmann, H. E. A., Kobel, H., Pache, W. & Tscherter, H.** 1976. Cyclosporin a and c. *European journal of applied microbiology and biotechnology,* 3**,** 125-133.

**Edgar, R. C.** 2004. Muscle: A multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics,* 5**,** 113.

**Ehrlich, P.** 1901. Die seitenkettentheorie und ihre gegner. *Münch. Med. Wschr,* 18**,** 2123-2124.

**Elander, R.** 2003. Industrial production of β-lactam antibiotics. *Applied microbiology and biotechnology,* 61**,** 385-392.

**Eyles, T. H., Vior, N. M. & Truman, A. W.** 2018. Rapid and robust yeast-mediated pathway refactoring generates multiple new bottromycin-related metabolites. *ACS synthetic biology,* 7**,** 1211-1218.

**Faghih, R., Freiberg, L., Leonard, J., Plattner, J. J. & Lartey, P. A.** 1996. Synthesis and antibacterial activity of 6-deoxysporeamicin a. *The Journal of antibiotics,* 49**,** 493-495.

**Felnagle, E. A., Jackson, E. E., Chan, Y. A., Podevels, A. M., Berti, A. D., Mcmahon, M. D. & Thomas, M. G.** 2008. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Molecular pharmaceutics,* 5**,** 191-211.

**Fernandes, P.** 2006. Antibacterial discovery and development—the failure of success? *Nature Biotechnology,* 24**,** 1497.

**Fischbach, M. A. & Walsh, C. T.** 2006. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. *Chemical reviews,* 106**,** 3468-3496.

**Flärdh, K. & Buttner, M. J.** 2009. Streptomyces morphogenetics: Dissecting differentiation in a filamentous bacterium. *Nature Reviews Microbiology,* 7**,** 36.

**Freiberg, L. a. E., C. M.; Bacino D. J. ; Seif, L. ; Lartey, P. A and Whittern, D.** Synthesis of amino analogs of (9s,11s)-9-9-deoxo-12-deoxy-9,12-epoxyerythromycin a (a-63483) and (9s)-9-deoxo-11,12-dideoxy-9,12-epoxy-11-oxoerythromycin a (a-63881). ICAAC, 1989 Abbott Laboratories, Abbott Park, IL, 60064, U.S.A.

**Gaisser, S., Reather, J., Wirtz, G., Kellenberger, L., Staunton, J. & Leadlay, P. F.** 2000. A defined system for hybrid macrolide biosynthesis in *saccharopolyspora erythraea*. *Molecular microbiology,* 36**,** 391-401.

**Gaudêncio, S. P. & Pereira, F.** 2015. Dereplication: Racing to speed up the natural products discovery process. *Natural product reports,* 32**,** 779-810.

**Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd & Smith, H. O.** 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods,* 6**,** 343-5.

**Gibson, D. G. E. A.** 2010. *Nature Methods. 901-903.*

**Gomez-Escribano, J. P., Alt, S. & Bibb, M. J.** 2016. Next generation sequencing of actinobacteria for the discovery of novel natural products. *Mar Drugs,* 14.

**Gomez-Escribano, J. P., Castro, J. F., Razmilic, V., Chandra, G., Andrews, B., Asenjo, J. A. & Bibb, M. J.** 2015. The streptomyces leeuwenhoekii genome: De novo sequencing and assembly in single contigs of the chromosome, circular plasmid psle1 and linear plasmid psle2. *BMC genomics,* 16**,** 485.

**Gomez-Escribano, J. P. & Bibb, M. J.** 2011. Engineering *streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microbial Biotechnology,* 4**,** 207-215.

**Goodfellow, M., Nouioui, I., Sanderson, R., Xie, F. & Bull, A. T.** 2018. Rare taxa and dark microbial matter: Novel bioactive actinobacteria abound in atacama desert soils. *Antonie van Leeuwenhoek***,** 1-18.

**Goodman, J. L., Winston, D. J., Greenfield, R. A., Chandrasekar, P. H., Fox, B., Kaizer, H., Shadduck, R. K., Shea, T. C., Stiff, P. & Friedman, D. J.** 1992. A controlled trial of fluconazole to prevent fungal infections in patients undergoing bone marrow transplantation. *New England Journal of Medicine,* 326**,** 845-851.

**Gust, B., Kieser, T. & Chater, K.** 2002. Redirect technology: Pcr-targeting system in *streptomyces coelicolor*. *The John Innes Centre, Norwich, United Kingdom*.

**Haight, T. H. & Finland, M.** 1952. Laboratory and clinical studies on erythromycin. *New England Journal of Medicine,* 247**,** 227-232.

**Hans-Georg Sahl & Bierbaum, G.** 1998. Lantibiotics: Biosynthesis and biological activities of uniquely modified peptides from gram-positive bacteria. *Annual Review of Microbiology,* 52**,** 41-79.

**Hansson, G. K.** 2005. Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med,* 352**,** 1685-95.

**Harris, W. R., Carrano, C. J., Cooper, S. R., Sofen, S. R., Avdeef, A. E., Mcardle, J. V. & Raymond, K. N.** 1979. Coordination chemistry of microbial iron transport compounds. 19. Stability constants and electrochemical behavior of ferric enterobactin and model complexes. *Journal of the American Chemical Society,* 101**,** 6097-6104.

**He, Y., Hu, Z., Li, Q., Huang, J., Li, X.-N., Zhu, H., Liu, J., Wang, J., Xue, Y. & Zhang, Y.** 2017. Bioassay-guided isolation of antibacterial metabolites from *emericella* sp. Tj29. *Journal of Natural Products,* 80**,** 2399-2405.

**He, Z., Kisla, D., Zhang, L., Yuan, C., Green-Church, K. B. & Yousef, A. E.** 2007. Isolation and identification of a paenibacillus polymyxa strain that coproduces a novel lantibiotic and polymyxin. *Applied and Environmental Microbiology,* 73**,** 168-178.

**Heeb, S., Fletcher, M. P., Chhabra, S. R., Diggle, S. P., Williams, P. & Cámara, M.** 2011. Quinolones: From antibiotics to autoinducers. *FEMS microbiology reviews,* 35**,** 247-274.

**Heine, D., Holmes, N. A., Worsley, S. F., Santos, A. C. A., Innocent, T. M., Scherlach, K., Patrick, E. H., Douglas, W. Y., Murrell, J. C. & Vieria, P. C.** 2018. Chemical warfare between leafcutter ant symbionts and a co-evolved pathogen. *Nature communications,* 9**,** 2208.

**Herbrecht, R., Denning, D. W., Patterson, T. F., Bennett, J. E., Greene, R. E., Oestmann, J.-W., Kern, W. V., Marr, K. A., Ribaud, P. & Lortholary, O.** 2002. Voriconazole versus amphotericin b for primary therapy of invasive aspergillosis. *New England Journal of Medicine,* 347**,** 408-415.

**Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A.** 2007. Type ii polyketide synthases: Gaining a deeper insight into enzymatic teamwork. *Natural product reports,* 24**,** 162-190.

**Hodgkin, D. C.** 1949. The x-ray analysis of the structure of penicillin. *Advancement of science,* 6**,** 85-89.

**Huber, F., Pieper, R. & Tietz, A.** 1988. The formation of daptomycin by supplying decanoic acid to *streptomyces roseosporus* cultures producing the antibiotic complex a21978c. *Journal of Biotechnology,* 7**,** 283-292.

**Hudson, G. A. & Mitchell, D. A.** 2018. Ripp antibiotics: Biosynthesis and engineering potential. *Curr Opin Microbiol,* 45**,** 61-69.

**Huo, L., Okesli, A., Zhao, M. & Van Der Donk, W. A.** 2017. Insights into the biosynthesis of duramycin. *Appl Environ Microbiol,* 83.

**Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. & Ōmura, S.** 2003. Complete genome sequence and comparative analysis of the industrial microorganism *streptomyces avermitilis*. *Nature biotechnology,* 21**,** 526.

**Jang, K. H., Nam, S. J., Locke, J. B., Kauffman, C. A., Beatty, D. S., Paul, L. A. & Fenical, W.** 2013. Anthracimycin, a potent anthrax antibiotic from a marine-derived actinomycete. *Angewandte Chemie International Edition,* 52**,** 7822-7824.

**Jones, M. B., Nierman, W. C., Shan, Y., Frank, B. C., Spoering, A., Ling, L., Peoples, A., Zullo, A., Lewis, K. & Nelson, K. E.** 2017. Reducing the bottleneck in discovery of novel antibiotics. *Microbial ecology,* 73**,** 658-667.

**Kadi, N. & Challis, G. L.** 2009. Chapter 17 siderophore biosynthesis: A substrate specificity assay for nonribosomal peptide synthetase-independent siderophore synthetases involving trapping of acyl-adenylate intermediates with hydroxylamine. *Methods in enzymology.* Academic Press.

**Kahne, D., Leimkuhler, C., Lu, W. & Walsh, C.** 2005. Glycopeptide and lipoglycopeptide antibiotics. *Chemical Reviews,* 105**,** 425-48.

**Kasten, F. H.** 1996. Paul ehrlich: Pathfinder in cell biology. 1. Chronicle of his life and accomplishments in immunology, cancer research, and chemotherapy. *Biotech Histochem,* 71**,** 2-37.

**Katsuyama, Y. & Ohnishi, Y.** 2012. Type iii polyketide synthases in microorganisms. *Methods in enzymology.* Elsevier.

**Katz, L. & Baltz, R. H.** 2016. Natural product discovery: Past, present, and future. *Journal of industrial microbiology & biotechnology,* 43**,** 155-176.

**Kautsar, S. A., Suarez Duran, H. G., Blin, K., Osbourn, A. & Medema, M. H.** 2017. Plantismash: Automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res,* 45**,** W55-W63.

**Kellner, R., Jung, G., Horner, T., Zahner, H., Schnell, N., Entian, K. D. & Gotz, F.** 1988. Gallidermin: A new lanthionine-containing polypeptide antibiotic. *European Journal of Biochemistry,* 177**,** 53-59.

**Kieser, T., Foundation, J. I., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A.** 2000. *Practical streptomyces genetics*, John Innes Foundation.

**Kirkpatrick, P., Raja, A., Labonte, J. & Lebbos, J.** 2003. Daptomycin. *Nat Rev Drug Discov,* 2**,** 943-4.

**Kirst, H. A.** 2010. The spinosyn family of insecticides: Realizing the potential of natural products research. *J Antibiot (Tokyo),* 63**,** 101-11.

**Kodani, S., Komaki, H., Ishimura, S., Hemmi, H. & Ohnishi-Kameyama, M.** 2016. Isolation and structure determination of a new lantibiotic cinnamycin b from *actinomadura atramentaria* based on genome mining. *J Ind Microbiol Biotechnol,* 43**,** 1159-65.

**Krantz, J. C.** 1974. Historical medical classics involving new drugs.

**Krysan, D. J.** 2017. The unmet clinical need of novel antifungal drugs. *Virulence,* 8**,** 135-137.

**Kuščer, E., Coates, N., Challis, I., Gregory, M., Wilkinson, B., Sheridan, R. & Petković, H.** 2007. Roles of raph and rapg in positive regulation of rapamycin biosynthesis in *streptomyces hygroscopicus*. *Journal of bacteriology,* 189**,** 4756-4763.

**Lacey, J. & Goodfellow, M.** 1975. A novel actinomycete from sugar-cane bagasse: *Saccharopolyspora hirsuta* gen. Et sp. Nov. *Microbiology,* 88**,** 75-85.

**Laxminarayan, R.** 2014. Antibiotic effectiveness: Balancing conservation against innovation. *science,* 345**,** 1299-1301.

**Lazos, O., Tosin, M., Slusarczyk, A. L., Boakes, S., Cortes, J., Sidebottom, P. J. & Leadlay, P. F.** 2010. Biosynthesis of the putative siderophore erythrochelin requires unprecedented crosstalk between separate nonribosomal peptide gene clusters. *Chem Biol,* 17**,** 160-73.

**Leipoldt, F., Santos-Aberturas, J., Stegmann, D. P., Wolf, F., Kulik, A., Lacret, R., Popadić, D., Keinhörster, D., Kirchner, N., Bekiesch, P., Gross, H., Truman, A. W. & Kaysser, L.** 2017. Warhead biosynthesis and the origin of structural diversity in hydroxamate metalloproteinase inhibitors. *Nature Communications,* 8**,** 1965.

**Lewis, K.** 2013. Platforms for antibiotic discovery. *Nature reviews Drug discovery,* 12**,** 371.

**Li, W., Cowley, A., Uludag, M., Gur, T., Mcwilliam, H., Squizzato, S., Park, Y. M., Buso, N. & Lopez, R.** 2015. The embl-ebi bioinformatics web and programmatic tools framework. *Nucleic acids research,* 43**,** W580-W584.

**Ling, L. L., Schneider, T., Peoples, A. J., Spoering, A. L., Engels, I., Conlon, B. P., Mueller, A., Schäberle, T. F., Hughes, D. E. & Epstein, S.** 2015. A new antibiotic kills pathogens without detectable resistance. *Nature,* 517**,** 455.

**Loman, N. J. & Pallen, M. J.** 2015. Twenty years of bacterial genome sequencing. *Nature Reviews Microbiology,* 13**,** 787.

**Luo, Y., Huang, H., Liang, J., Wang, M., Lu, L., Shao, Z., Cobb, R. E. & Zhao, H.** 2013. Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. *Nat Commun,* 4**,** 2894.

**Lyddiard, D., Jones, G. L. & Greatrex, B. W.** 2016. Keeping it simple: Lessons from the golden era of antibiotic discovery. *FEMS microbiology letters,* 363.

**Marahiel, M. A.** 2016. A structural model for multimodular nrps assembly lines. *Natural product reports,* 33**,** 136-140.

**Marahiel, M. A., Stachelhaus, T. & Mootz, H. D.** 1997. Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chemical reviews,* 97**,** 2651-2674.

**Marconi, P., Bistoni, F., Boncio, L., Bersiani, A., Bravi, P. & Pitzurra, M.** 1976. Utilizzazione di soluzione salina ipertonica di cloruro di potassio (3m kc1) per l'estrazione di antigeni solubili da candia albicans. *Ann Sclavo,* 18**,** 61-66.

**Marfey, P.** 1984. Determination of d-amino acids. Ii. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Research Communications,* 49**,** 591.

**Marki, F., Hanni, E., Fredenhagen, A. & Van Oostrum, J.** 1991. Mode of action of the lanthionine-containing peptide antibiotics duramycin, duramycin b and c, and cinnamycin as indirect inhibitors of phospholipase a2. *Biochem Pharmacol,* 42**,** 2027-35.

**Martens, E. & Demain, A. L.** 2017. The antibiotic resistance crisis, with a focus on the united states. *The Journal of antibiotics,* 70**,** 520.

**Martin, N. I., Sprules, T., Carpenter, M. R., Cotter, P. D., Hill, C., Ross, R. P. & Vederas, J. C.** 2004. Structural characterization of lacticin 3147, a two-peptide lantibiotic with synergistic activity. *Biochemistry,* 43**,** 3049-56.

**Martínez-Núñez, M. A. & López, V. E. L. Y.** 2016. Nonribosomal peptides synthetases and their applications in industry. *Sustainable Chemical Processes,* 4**,** 13.

**Mcauliffe, O., Ross, R. P. & Hill, C.** 2001. Lantibiotics: Structure, biosynthesis and mode of action. *FEMS Microbiology Reviews,* 25**,** 285-308.

**Minowa, Y., Araki, M. & Kanehisa, M.** 2007. Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes. *J Mol Biol,* 368**,** 1500-17.

**Molloy, E. M. & Hertweck, C.** 2017. Antimicrobial discovery inspired by ecological interactions. *Current opinion in microbiology,* 39**,** 121-127.

**Monciardini, P., Iorio, M., Maffioli, S., Sosio, M. & Donadio, S.** 2014. Discovering new bioactive molecules from microbial sources. *Microbial biotechnology,* 7**,** 209-220.

**Morishita, A., Mutoh, N., Ishizawa, K., Suzuki, T., Yokoiyama, S. & Yaginuma, S.** 1992. Sporeamicin a, a new macrolide antibiotic. Iii. Biological properties. *J Antibiot (Tokyo),* 45**,** 613-7.

**Morris, P. J.** 1984. The impact of cyclosporin a on transplantation. *Adv Surg,* 17**,** 99-127.

**Mun, W., Kwon, H., Im, H., Choi, S. Y., Monnappa, A. K. & Mitchell, R. J.** 2017. Cyanide production by chromobacterium piscinae shields it from bdellovibrio bacteriovorus hd100 predation. *mBio,* 8**,** e01370-17.

**Neilands, J.** 1995. Siderophores: Structure and function of microbial iron transport compounds. *Journal of Biological Chemistry,* 270**,** 26723-26726.

**Newman, D. J. & Cragg, G. M.** 2016. Natural products as sources of new drugs from 1981 to 2014. *Journal of natural products,* 79**,** 629-661.

**Nicas, T. & Cooper, R.** 1997. Vancomycin and other glycopeptides: Biotechnology of antibiotics. *Ed. Strohl WR New York.: Marcel Dekker, Inc.***,** 363.

**Nichols, D., Cahoon, N., Trakhtenberg, E. M., Pham, L., Mehta, A., Belanger, A., Kanigan, T., Lewis, K. & Epstein, S. S.** 2010. Use of ichip for high-throughput *in situ* cultivation of "uncultivable" microbial species. *Applied and Environmental Microbiology,* 76**,** 2445-2450.

**Nothias, L. F., Nothias-Esposito, M., Da Silva, R., Wang, M., Protsyuk, I., Zhang, Z., Sarvepalli, A., Leyssen, P., Touboul, D., Costa, J., Paolini, J., Alexandrov, T., Litaudon, M. & Dorrestein, P. C.** 2018. Bioactivity-based molecular networking for the

discovery of drug leads in natural product bioassay-guided fractionation. *J Nat Prod,* 81**,** 758-767.

**O'Rourke, S., Widdick, D. & Bibb, M.** 2017. A novel mechanism of immunity controls the onset of cinnamycin biosynthesis in *streptomyces cinnamoneus* dsm 40646. *Journal of industrial microbiology & biotechnology,* 44**,** 563-572.

**Oliynyk, I., Varelogianni, G., Roomans, G. M. & Johannesson, M.** 2010. Effect of duramycin on chloride transport and intracellular calcium concentration in cystic fibrosis and non-cystic fibrosis epithelia. *Apmis,* 118**,** 982-990.

**Oliynyk, M., Samborskyy, M., Lester, J. B., Mironenko, T., Scott, N., Dickens, S., Haydock, S. F. & Leadlay, P. F.** 2007. Complete genome sequence of the erythromycin-producing bacterium *saccharopolyspora erythraea* nrrl23338. *Nat Biotechnol,* 25**,** 447-53.

**Omura, S.** 2002. *Macrolide antibiotics: Chemistry, biology, and practice*, Elsevier.

**Organization, W. H.** 2017. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. *Geneva: World Health Organization*.

**Pacey, M. S., Dirlam, J. P., Geldart, R. W., Leadlay, P. F., Mcarthur, H. A., Mccormick, E. L., Monday, R. A., O'connell, T. N., Staunton, J. & Winchester, T. J.** 1998. Novel erythromycins from a recombinant saccharopolyspora erythraea strain nrrl 2338 pig1. *The Journal of antibiotics,* 51**,** 1029-1034.

**Page, M. I.** 2012. *The chemistry of β-lactams*, Springer Science & Business Media.

**Pan, Y., Yang, X., Li, J., Zhang, R., Hu, Y., Zhou, Y., Wang, J. & Zhu, B.** 2011. Genome sequence of the spinosyns-producing bacterium *saccharopolyspora spinosa* nrrl 18395. *J Bacteriol,* 193**,** 3150-1.

**Patzer, S., Baquero, M., Bravo, D., Moreno, F. & Hantke, K.** 2003. The colicin g, h and x determinants encode microcins m and h47, which might utilize the catecholate siderophore receptors fepa, cir, fiu and iron. *Microbiology,* 149**,** 2557-2570.

**Payne, J. A., Schoppet, M., Hansen, M. H. & Cryle, M. J.** 2017. Diversity of nature's assembly lines–recent discoveries in non-ribosomal peptide synthesis. *Molecular BioSystems,* 13**,** 9-22.

**Pettersson, B. M., Behra, P. R., Manduva, S., Das, S., Dasgupta, S., Bhattacharya, A. & Kirsebom, L. A.** 2014. Draft genome sequence of *saccharopolyspora rectivirgula*. *Genome Announc,* 2.

**Pfaller, M., Boyken, L., Hollis, R., Kroeger, J., Messer, S., Tendolkar, S. & Diekema, D.** 2008. In vitro susceptibility of invasive isolates of candida spp. To anidulafungin, caspofungin, and micafungin: Six years of global surveillance. *Journal of clinical microbiology,* 46**,** 150-156.

**Pishchany, G., Mevers, E., Ndousse-Fetter, S., Horvath, D. J., Paludo, C. R., Silva-Junior, E. A., Koren, S., Skaar, E. P., Clardy, J. & Kolter, R.** 2018. Amycomicin: A potent and specific antibiotic discovered with a targeted interaction screen. *bioRxiv*.

**Powell, P., Cline, G., Reid, C. & Szaniszlo, P.** 1980. Occurrence of hydroxamate siderophore iron chelators in soils. *Nature,* 287**,** 833.

**Qin, Z., Munnoch, J. T., Devine, R., Holmes, N. A., Seipke, R. F., Wilkinson, K. A., Wilkinson, B. & Hutchings, M. I.** 2017. Formicamycins, antibacterial polyketides produced by *streptomyces formicae* isolated from african *tetraponera* plant-ants. *Chemical Science,* 8**,** 3218-3227.

**Rappé, M. S. & Giovannoni, S. J.** 2003. The uncultured microbial majority. *Annual Reviews in Microbiology,* 57**,** 369-394.

**Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H.** 2007. Phylogenetic analysis of condensation domains in nrps sheds light on their functional evolution. *BMC Evolutionary Biology,* 7**,** 78.

**Raynal, A., Friedmann, A., Tuphile, K., Guerineau, M. & Pernodet, J. L.** 2002. Characterization of the *att*p site of the integrative element psam2 from *streptomyces ambofaciens*. *Microbiology,* 148**,** 61-7.

**Reeves, A. R., English, R. S., Lampel, J. S., Post, D. A. & Vanden Boom, T. J.** 1999. Transcriptional organization of the erythromycin biosynthetic gene cluster of saccharopolyspora erythraea. *J Bacteriol,* 181**,** 7098-106.

**Robert, X. & Gouet, P.** 2014. Deciphering key features in protein structures with the new endscript server. *Nucleic acids research,* 42**,** W320-W324.

**Roemer, T. & Krysan, D. J.** 2014. Antifungal drug development: Challenges, unmet clinical needs, and new approaches. *Cold Spring Harbor Perspectives in Medicine,* 4**,** a019703.

**Rowe, C. J., Cortés, J., Gaisser, S., Staunton, J. & Leadlay, P. F.** 1998. Construction of new vectors for high-level expression in actinomycetes. *Gene,* 216**,** 215-223.

**Rutledge, P. J. & Challis, G. L.** 2015. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature reviews microbiology,* 13**,** 509.

**Saha, M., Sarkar, S., Sarkar, B., Sharma, B. K., Bhattacharjee, S. & Tribedi, P.** 2016. Microbial siderophores and their potential applications: A review. *Environmental Science and Pollution Research,* 23**,** 3984-3999.

**Santajit, S. & Indrawattana, N.** 2016. Mechanisms of antimicrobial resistance in eskape pathogens. *BioMed research international,* 2016.

**Saravana Kumar, P., Yuvaraj, P., Gabrial Paulraj, M., Ignacimuthu, S. & Abdullah Al-Dhabi, N.** 2018. Bio-prospecting of soil *streptomyces* and its bioassay-guided isolation of microbial derived auxin with antifungal properties. *Journal de Mycologie Médicale,* 28**,** 462-468.

**Saxena, B., Modi, M. & Modi, V.** 1986. Isolation and characterization of siderophores from azospirillum lipoferum d-2. *Microbiology,* 132**,** 2219-2224.

**Schatz, A., Bugle, E. & Waksman, S. A.** 1944. Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. *Proceedings of the Society for Experimental Biology and Medicine,* 55**,** 66-69.

**Scheinfeld, N.** 2004. Telithromycin: A brief review of a new ketolide antibiotic. *Journal of drugs in dermatology: JDD,* 3**,** 409-413.

**Scott, T. A.** 2017. *Biosynthesis and mode of action of the β-lactone antibiotic obafluorin.* Doctor of Philosophy Thesis, University of East Anglia.

**Seipke, R. F., Barke, J., Heavens, D., Yu, D. W. & Hutchings, M. I.** 2013. Analysis of the bacterial communities associated with two ant-plant symbioses. *Microbiologyopen,* 2**,** 276-83.

**Senges, C. H. R., Al-Dilaimi, A., Marchbank, D. H., Wibberg, D., Winkler, A., Haltli, B., Nowrousian, M., Kalinowski, J., Kerr, R. G. & Bandow, J. E.** 2018. The secreted metabolome of *streptomyces chartreusis* and implications for bacterial chemistry. *Proceedings of the National Academy of Sciences,* 115**,** 2490-2495.

**Sherry, N. & Howden, B.** 2018. Emerging gram negative resistance to last-line antimicrobial agents fosfomycin, colistin and ceftazidime-avibactam–epidemiology, laboratory detection and treatment implications. *Expert review of anti-infective therapy,* 16**,** 289-306.

**Shin, J. M., Gwak, J. W., Kamarajan, P., Fenno, J. C., Rickard, A. H. & Kapila, Y. L.** 2016. Biomedical applications of nisin. *Journal of applied microbiology,* 120**,** 1449-1465.

**Silver, L. L.** 2011. Challenges of antibacterial discovery. *Clinical Microbiology Reviews,* 24**,** 71-109.

**Simmons, T. L., Coates, R. C., Clark, B. R., Engene, N., Gonzalez, D., Esquenazi, E., Dorrestein, P. C. & Gerwick, W. H.** 2008. Biosynthetic origin of natural products isolated from marine microorganism–invertebrate assemblages. *Proceedings of the National Academy of Sciences,* 105**,** 4587-4594.

**Skinnider, M. A., Dejong, C. A., Rees, P. N., Johnston, C. W., Li, H., Webster, A. L., Wyatt, M. A. & Magarvey, N. A.** 2015. Genomes to natural products prediction informatics for secondary metabolomes (prism). *Nucleic acids research,* 43**,** 9645-9662.

**Skinnider, M. A., Johnston, C. W., Edgar, R. E., Dejong, C. A., Merwin, N. J., Rees, P. N. & Magarvey, N. A.** 2016. Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining. *Proceedings of the National Academy of Sciences,* 113**,** E6343-E6351.

**Society for Actinomycetes, J.** 2013. *Digital atlas of actinomycetes 2* [Online]. Available: http://atlas.actino.jp/ [Accessed].

**Starcevic, A., Jaspars, M., Cullum, J., Hranueli, D. & Long, P. F.** 2007. Predicting the nature and timing of epimerisation on a modular polyketide synthase. *Chembiochem,* 8**,** 28-31.

**Staunton, J. & Weissman, K. J.** 2001. Polyketide biosynthesis: A millennium review. *Natural Product Reports,* 18**,** 380-416.

**Staunton, J. & Wilkinson, B.** 1997. Biosynthesis of erythromycin and rapamycin. *Chemical reviews,* 97**,** 2611-2630.

**Sydor, P. K. & Challis, G. L.** 2012. Oxidative tailoring reactions catalyzed by nonheme iron-dependent enzymes: Streptorubin b biosynthesis as an example. *Methods in enzymology.* Elsevier.

**Szwarc, K., Szczeblewski, P., Sowiński, P., Borowski, E. & Pawlak, J.** 2015. The stereostructure of candicidin d. *The Journal of antibiotics,* 68**,** 504.

**Van Der Lee, T. a. J. & Medema, M. H.** 2016. Computational strategies for genome-based natural product discovery and engineering in fungi. *Fungal Genetics and Biology,* 89**,** 29-36.

**Van Heel, A. J., De Jong, A., Song, C., Viel, J. H., Kok, J. & Kuipers, O. P.** 2018. Bagel4: A user-friendly web server to thoroughly mine ripps and bacteriocins. *Nucleic acids research*.

**Waksman, S. A.** 1961. Classification, identification and descriptions of genera and species. *The actinomycetes,* 2.

**Waksman, S. A., Reilly, H. C. & Johnstone, D. B.** 1946. Isolation of streptomycin-producing strains of *streptomyces griseus*. *Journal of Bacteriology,* 52**,** 393-397.

**Wandersman, C. & Delepelaire, P.** 2004. Bacterial iron sources: From siderophores to hemophores. *Annu. Rev. Microbiol.,* 58**,** 611-647.

**Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A. & Luzzatto-Knaan, T.** 2016. Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nature biotechnology,* 34**,** 828.

**Watanabe, Y., Morimoto, S., Adashi, T., Kashimura, M. & Asaka, T.** 1993. Chemical modification of erythromycins. Ix. 1. *The Journal of antibiotics,* 46**,** 647-660.

**Watrous, J., Roach, P., Alexandrov, T., Heath, B. S., Yang, J. Y., Kersten, R. D., Van Der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J. M., Moore, B. S., Laskin, J., Bandeira, N. & Dorrestein, P. C.** 2012. Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences,* 109**,** E1743-E1752.

**Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., Lee, S. Y., Fischbach, M. A., Muller, R., Wohlleben, W., Breitling, R., Takano, E. & Medema, M. H.** 2015. Antismash 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res,* 43**,** W237-43.

**Weissman, K. J.** 2015. The structural biology of biosynthetic megaenzymes. *Nature chemical biology,* 11**,** 660.

**Widdick, D., Dodd, H., Barraille, P., White, J., Stein, T., Chater, K., Gasson, M. & Bibb, M.** 2003. Cloning and engineering of the cinnamycin biosynthetic gene cluster from *streptomyces cinnamoneus cinnamoneus* dsm 40005. *Proc Natl Acad Sci U S A,* 100**,** 4316-4321.

**Widdick, D., Royer, S. F., Wang, H., Vior, N. M., Gomez-Escribano, J. P., Davis, B. G. & Bibb, M. J.** 2018. Analysis of the tunicamycin biosynthetic gene cluster of *streptomyces chartreusis* reveals new insights into tunicamycin production and immunity. *Antimicrob Agents Chemother,* 62.

**Wietz, M., Månsson, M., Bowman, J. S., Blom, N., Ng, Y. & Gram, L.** 2012. Wide distribution of closely related, antibiotic-producing arthrobacter strains throughout the arctic ocean. *Applied and environmental microbiology***,** AEM. 07096-11.

**Wietzorrek, A. & Bibb, M.** 1997. A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an ompr-like DNA-binding fold. *Molecular Microbiology,* 25**,** 1181-1184.

**Willey, J. M. & Van Der Donk, W. A.** 2007. Lantibiotics: Peptides of diverse structure and function. *Annu. Rev. Microbiol.,* 61**,** 477-501.

**Witkop, B.** 1999. Paul ehrlich and his magic bullets, revisited. *Proceedings of the American Philosophical Society,* 143**,** 540-557.

**Yaginuma, S., Morishita, A., Ishizawa, K., Murofushi, S., Hayashi, M. & Mutoh, N.** 1992. Sporeamicin a, a new macrolide antibiotic. I. Taxonomy, fermentation, isolation and characterization. *J Antibiot (Tokyo),* 45**,** 599-606.

**Yamanaka, K., Oikawa, H., Ogawa, H.-O., Hosono, K., Shinmachi, F., Takano, H., Sakuda, S., Beppu, T. & Ueda, K.** 2005. Desferrioxamine e produced by *streptomyces griseus* stimulates growth and development of *streptomyces tanashiensis*. *Microbiology,* 151**,** 2899-2905.

**Yu, H., Zhang, L., Li, L., Zheng, C., Guo, L., Li, W., Sun, P. & Qin, L.** 2010. Recent developments and future prospects of antimicrobial metabolites produced by endophytes. *Microbiological research,* 165**,** 437-449.

**Zhang, H., Wang, Y. & Pfeifer, B. A.** 2008. Bacterial hosts for natural product production. *Molecular pharmaceutics,* 5**,** 212-225.

**Zhou, Z.-H., Liu, Z.-H., Qian, Y.-D., Kim, S. B. & Goodfellow, M.** 1998. *Saccharopolyspora spinosporotrichia* sp. Nov., a novel actinomycete from soil. *International Journal of Systematic and Evolutionary Microbiology,* 48**,** 53-58.

**Ziemert, N., Podell, S., Penn, K., Badger, J. H., Allen, E. & Jensen, P. R.** 2012. The natural product domain seeker napdos: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PloS one,* 7**,** e34064.

**Zotchev, S. B.** 2012. Marine actinomycetes as an emerging resource for the drug development pipelines. *Journal of biotechnology,* 158**,** 168-175.

# Appendix 1:

Primers

# Appendix 1

| Primer | Sequence 5'-3' |
|---|---|
| 16S_forw | AGAGTTTGATCATGGCTCAG |
| 16S_rev | TACGGTTACCTTGTTACGACTT |
| eryAf | CCAAGCTTCGGGAGGTCTACTGCTCGT |
| eryAr | CGGAATTCGAGCAACTTCCAGTCGGTGT |
| eryAII_Int1a_forw | ACTCGACTCGCTCAACGC |
| eryAII_Int1b_forw | TGAACGGCCAGTCCTACA |
| eryAII_Int2a_forw | GATGGCCCTGCGCAAGCGCCT |
| eryAII_Int2b_forw | GCAGCTGCTGGCCGGTGAGG |
| eryAII_Int3a_rev | AGCTCGTCGGCGTGCAACGACC |
| eryAII_Int3b_rev | AATCCCAGTCGATGTGCG |
| eryAII_Int4a_for | GATTTCGACCCGGGCTTCTT |
| eryAII_Int4b_for | GTCTTCGTCGGCATGAACGG |
| eryAII_Int4c_rv | AAGTCGACGAAGGTGTACGG |
| ermE-int_forw | AACGCTACGAGTCGATGGTC |
| ermE-int_rev | ATTCCTCTCGACCAACAGGC |
| cinAdelF | ATGATCCGAACGCACACCGACATCCAGGAGGAGATCATGATTCCGGGGATCCGTCGACC |
| cinAdelR | CGGCGAGCCCGGAGGCTCACCGCGTCGATGCTCAGCTCATGTAGGCTGGAGCTGCTTC |
| chkapraFout | AGATGATCGAGGATGCCAGG |
| chkapraFin1 | CGGATGCAGGAAGATCAACG |
| chkapraFin2 | GAGAAGTACCTGCCCATCGA |
| chkapraR | GCCCGTAGAGATTGGCGAT |
| pGP9chkF | GACGATGACGACGACCACC |
| pGP9chkR | TGGATCTATCAACAGGAGTCCA |
| RILint | GTTCCACGAGCGGTTCTG |
| deltaCinAseqF | GGCCGAGCGCTCGTGGAGTTG |
| deltaCinAseqR | GGGCGGCCAGATCGAGTGCTC |
| pSET152 F | GGCCGAGCGCTCGTGGAGTTG |
| pSET152 R | AAATACCGCACAGATGCGTAAG |
| kyaseqF1 | GACCTCCAGCGCCACCTGCAC |
| kyaseqR1 | GTCATGGTGATCTCCTTCGTC |
| kyaseqF2 | ATGCCGGTGCCGCAGATCGTC |
| kyaseqR2 | CCCGTGCAGCAGGTCCAGGTG |
| kyaseqF3 | CGCACCGGTGCGTTCGCCGAC |

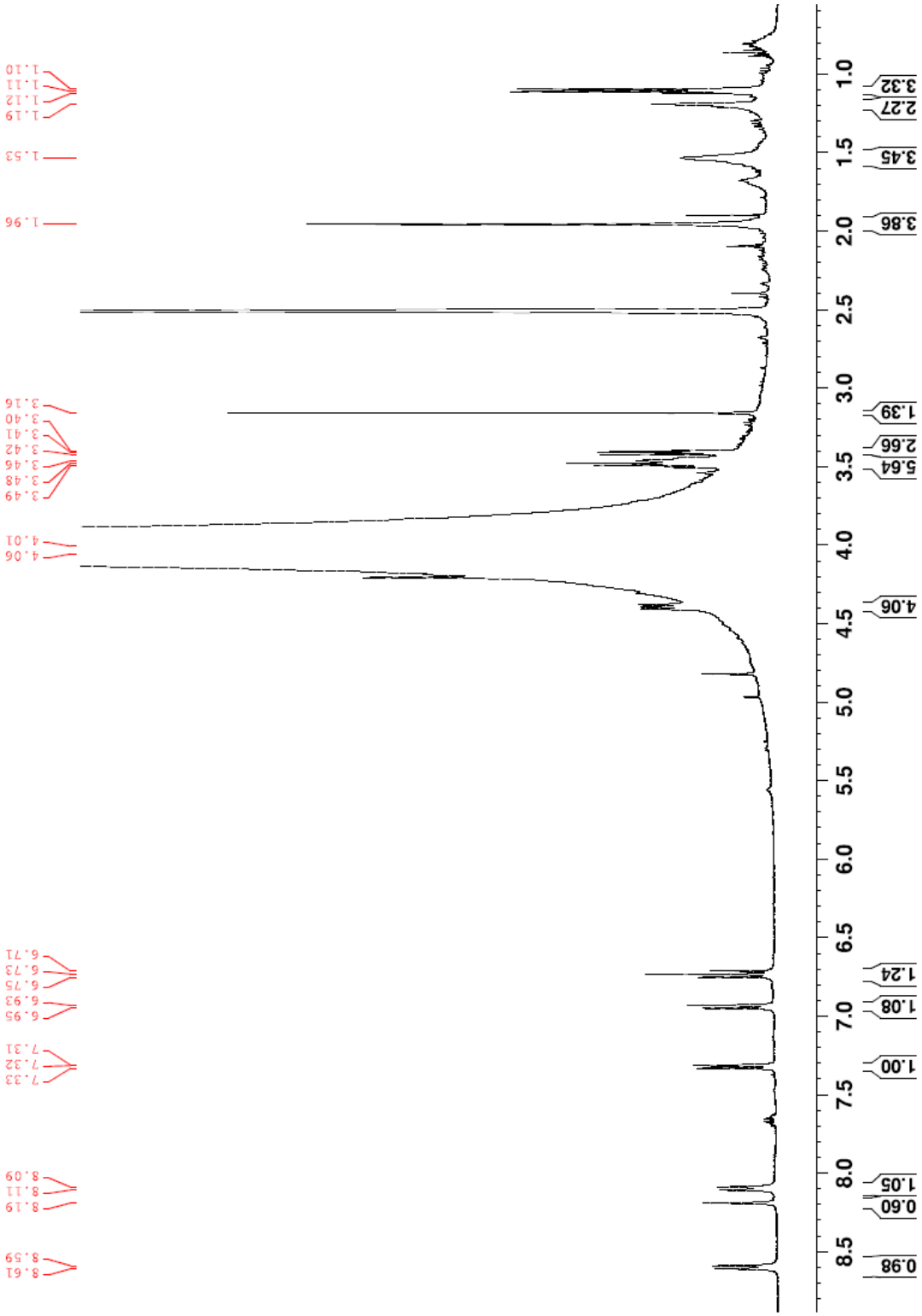
| | |
|---|---|
| kyaseqR3 | GATGGACTCCAGGTGCTCCTG |
| kyaseqR4 | GTAGGCGAGGTAGTCGTCCTC |
| deltaCinXseqF | CCCGCCGCGACCGTGAGCACC |
| deltaCinXseqR | CGATCGCGGGTGCGTCCTCGG |
| 152seqF | TAATGCAGCTGGCACGACAGG |
| cinAcompF | GGCGCGCATATGTCCGACACCATTCTTCG |
| cinAcompR | GCCGCGTCTAGATCACTTGGTCGAGCCGTCGCAC |
| 533F | GTGCCAGCGCCGCGGTAA |
| 1492R | CGGTTACCTTGTTACGACTT |
| kyaAdelF | ATGATCCGAACGCACACCGACATCCAGGAGGAGATCATGATTCCGGGGATCCGTCGACC |
| kyaAdelR | CGGCGAGCCCGGAGGCTCACCGCGTCGATGCTCAGCTCATGTAGGCTGGAGCTGCTTC |
| deltaKyaAseqF | GGCCGAGCGCTCGTGGAGTTG |
| deltaKyaAseqR | GGGCGGCCAGATCGAGTGCTC |
| cinAcompF | GGCGCGCATATGTCCGACACCATTCTTCG |
| cinAcompR | GCCGCGTCTAGATCACTTGGTCGAGCCGTCGCAC |
| KY3smNdeI-F | GCGGAATTCCGAAGCCGCGGATCTCCTGCCGTGG |
| KY3smStuI-R | GCGGGTACCTTTAAAAGGCCTCGTTGGCCGCGATCCCCTTGG |
| KY3lgStuI-F | GCGAGGCCTAGCATCGACGCGGTGAGCCTCC |
| KY3lgNdeI-R2 | CCGATAGCTGTTCGGCGGGTACCAC |

# Appendix 2:

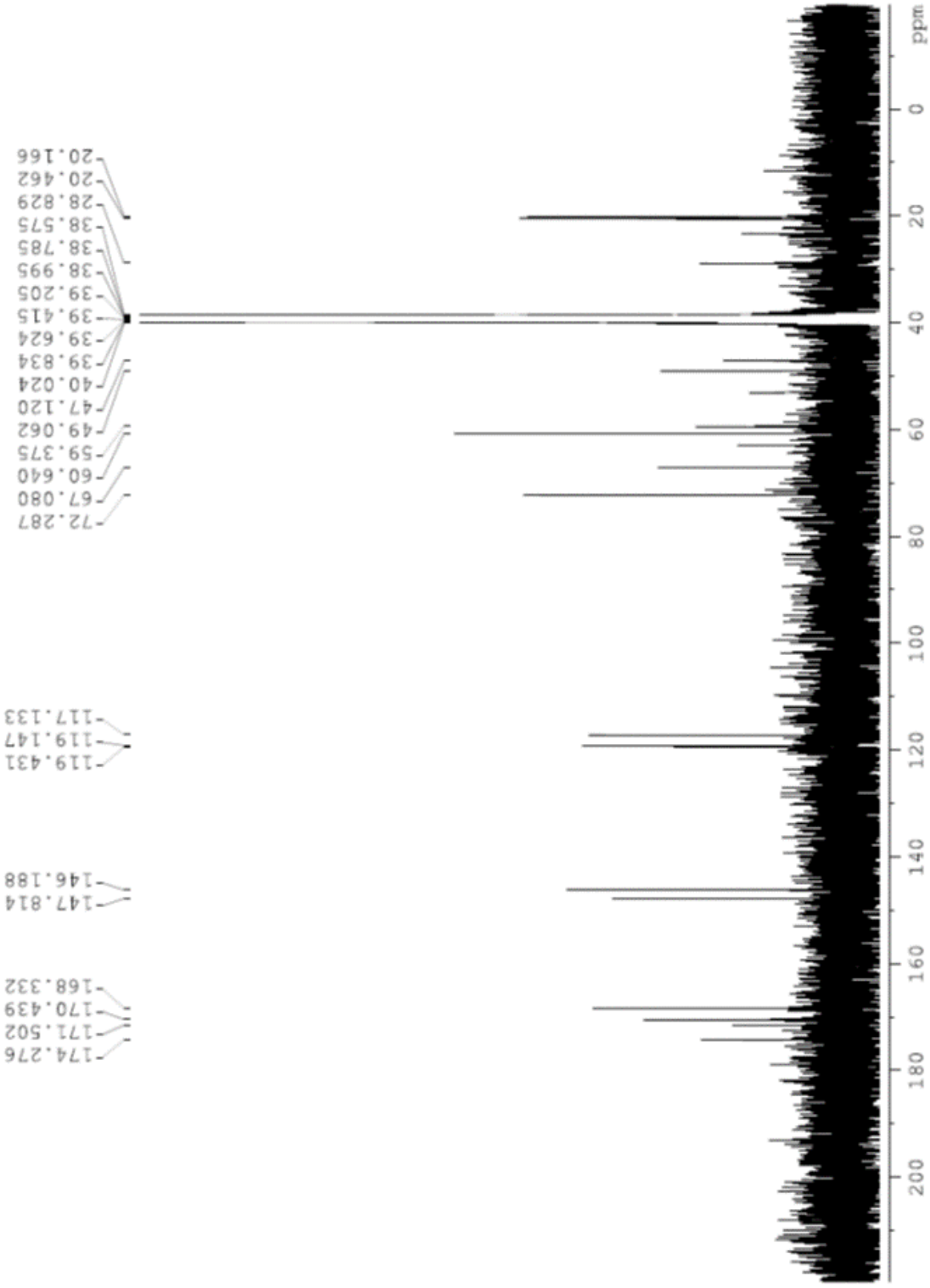
Supplementary Figures and Tables

# Appendix 2

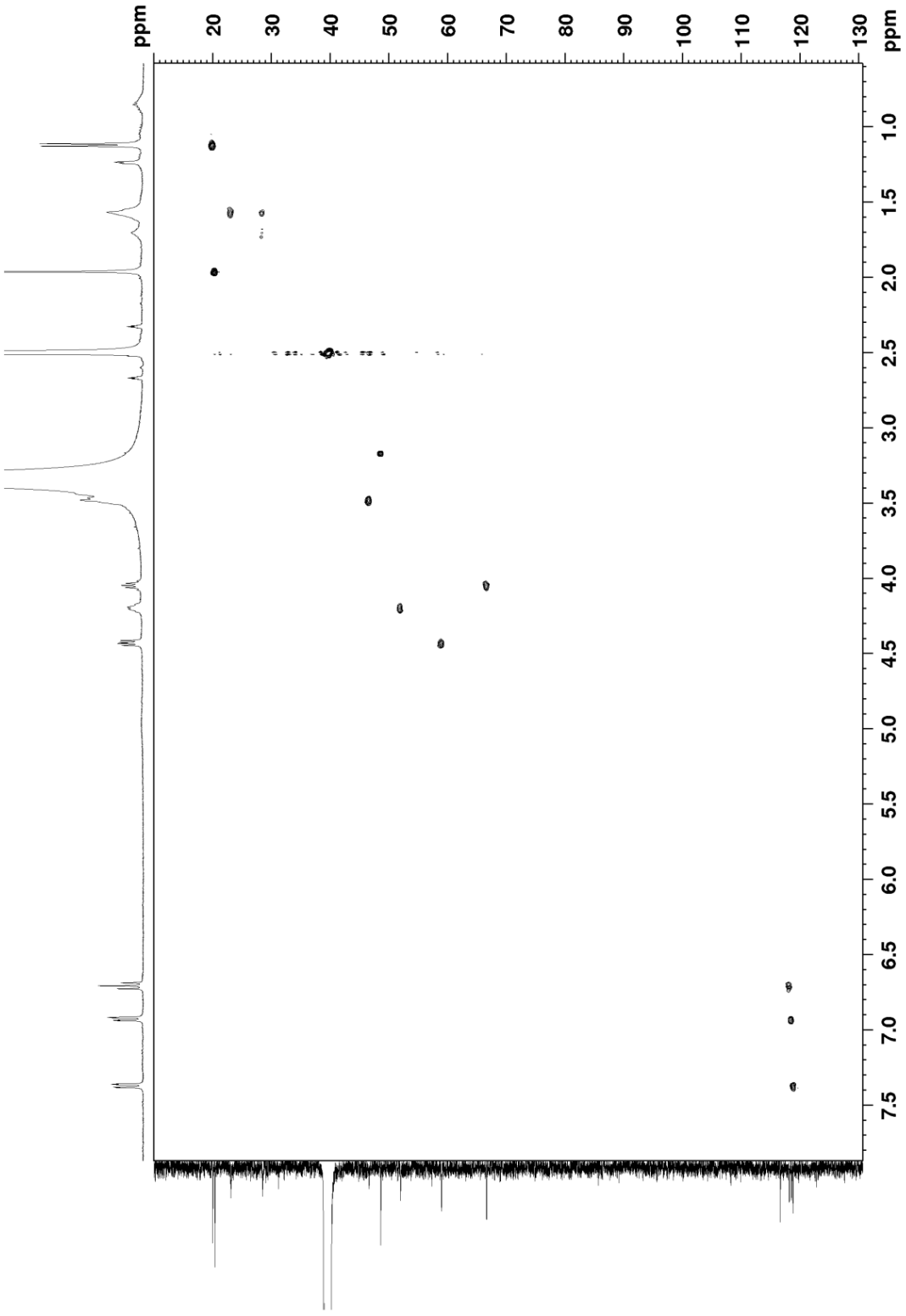
Supplementary Figure 1. $^{1}$H NMR spectrum (DMSO, 400 MHz) of EV60.



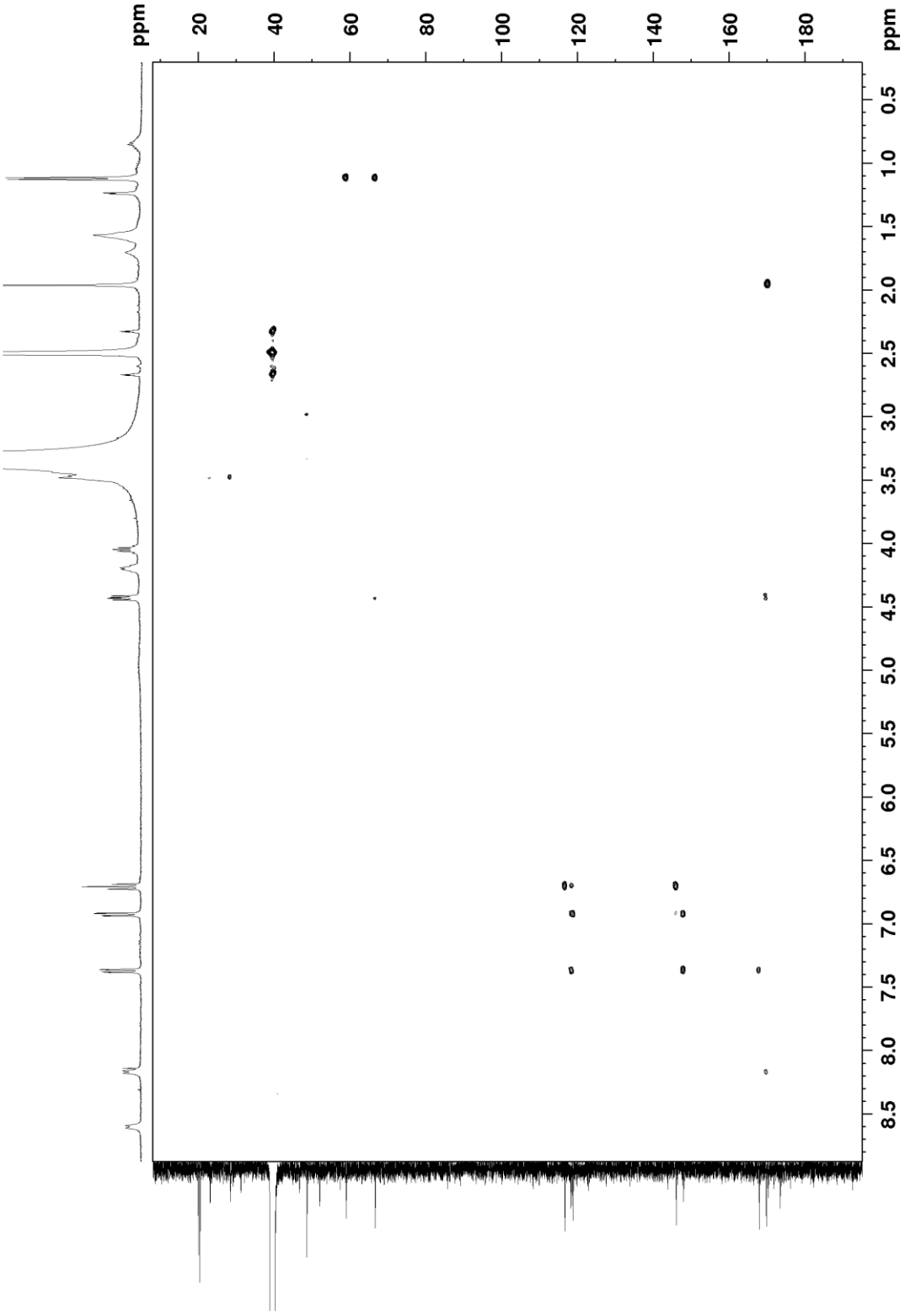
Supplementary Figure 2. $^{13}C$ NMR spectrum (DMSO, 100 MHz) of EV60.



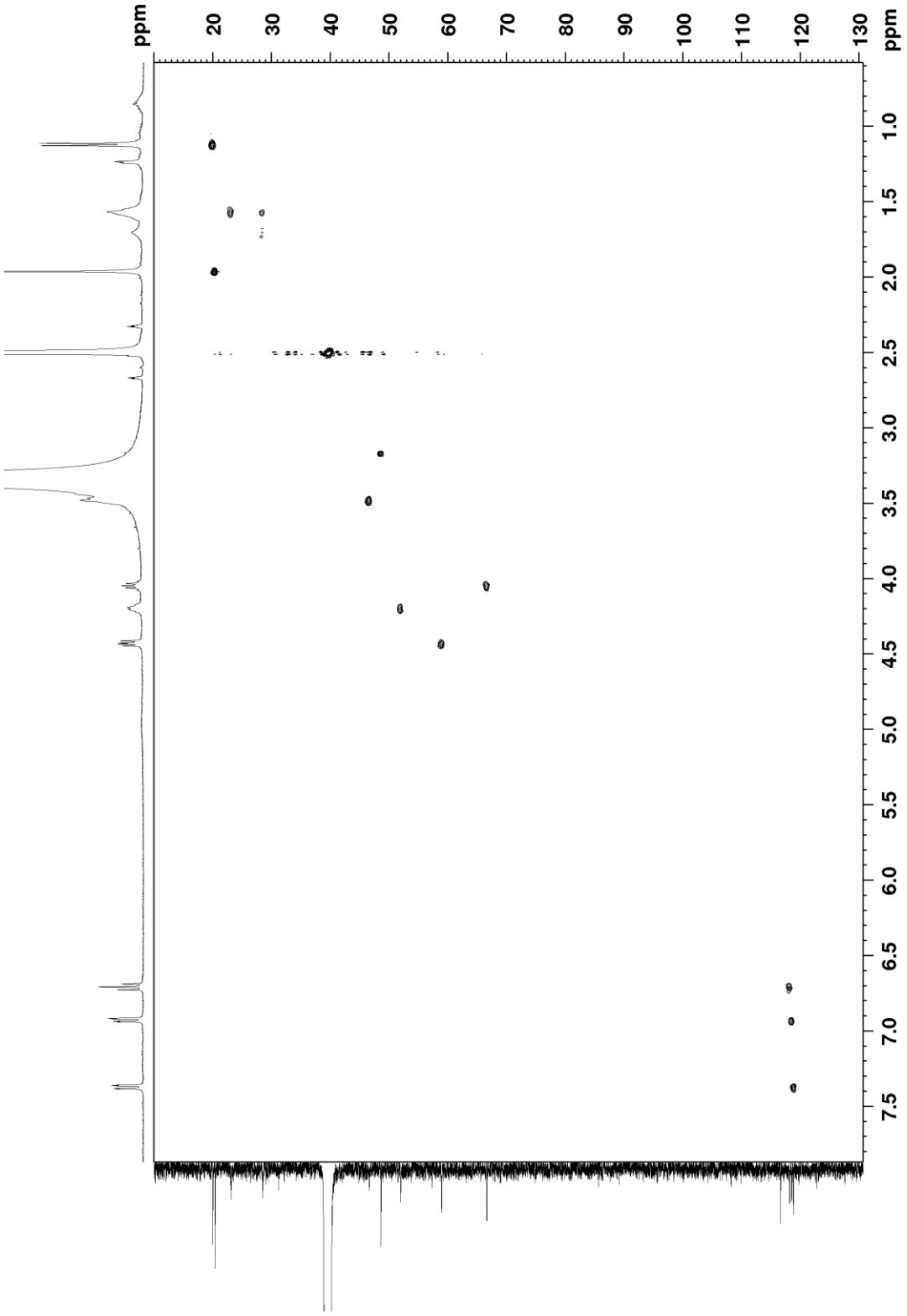
Supplementary Figure 3. COSY NMR spectrum (DMSO, 100 MHz) of EV60



Supplementary Figure 4. HMBC NMR spectrum (DMSO, 100 MHz) of EV60.



Supplementary Figure 5. HSQC NMR spectrum (DMSO, 100 MHz) of EV60.



Supplementary Table 1: $^{1}$H NMR chemical shift ppm table.

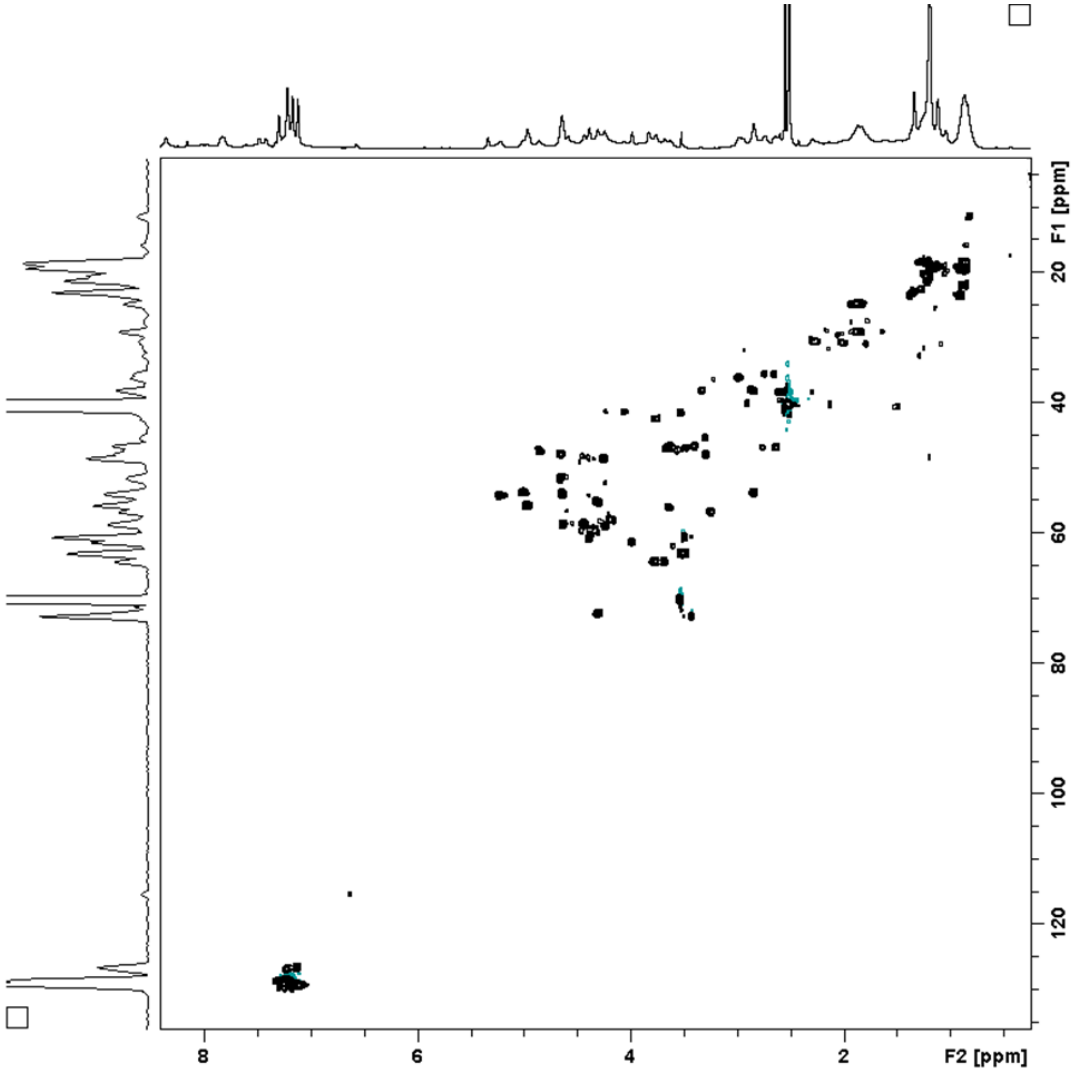
| | **EV60** ***δ* in ppm** | |
|---|---|---|
| **number** | **$^{1}$H (*J* in Hz)** | **$^{13}$C** |
| **1** | - | 167.9 |
| **2** | - | 116.7 |
| **3** | - | 148.0 |
| **4** | - | 146.0 |
| **5** | 6.93, dd (7.8, 1.4) | 118.5 |
| **6** | 6.71, dd (8.0, 7.8) | 118.2 |
| **7** | 7.37, dd (8.0, 1.4) | 118.8 |
| **2'-NH** | 8.59, d (8.0) | - |
| **1'** | - | 169.8 |
| **2'** | 4.43, dd (7.9, 5.1) | 58.9 |
| **3'** | 4.05, m | 66.6 |
| **4'** | 1.12, d (6.3) | 19.9 |
| **2''-NH** | 8.18, d (7.5) | - |
| **1''** | - | 173.4 |
| **2''** | 4.20, m | 52.0 |
| **3''** | 1.51-1.77, m | 28.5 |
| **4''** | 1.51-1.77, m | 23.1 |
| **5''** | 3.47, m | 46.6 |
| **1'''** | - | 170.3 |
| **2'''** | 1.96, s | 20.3 |

Analytical data for EV60

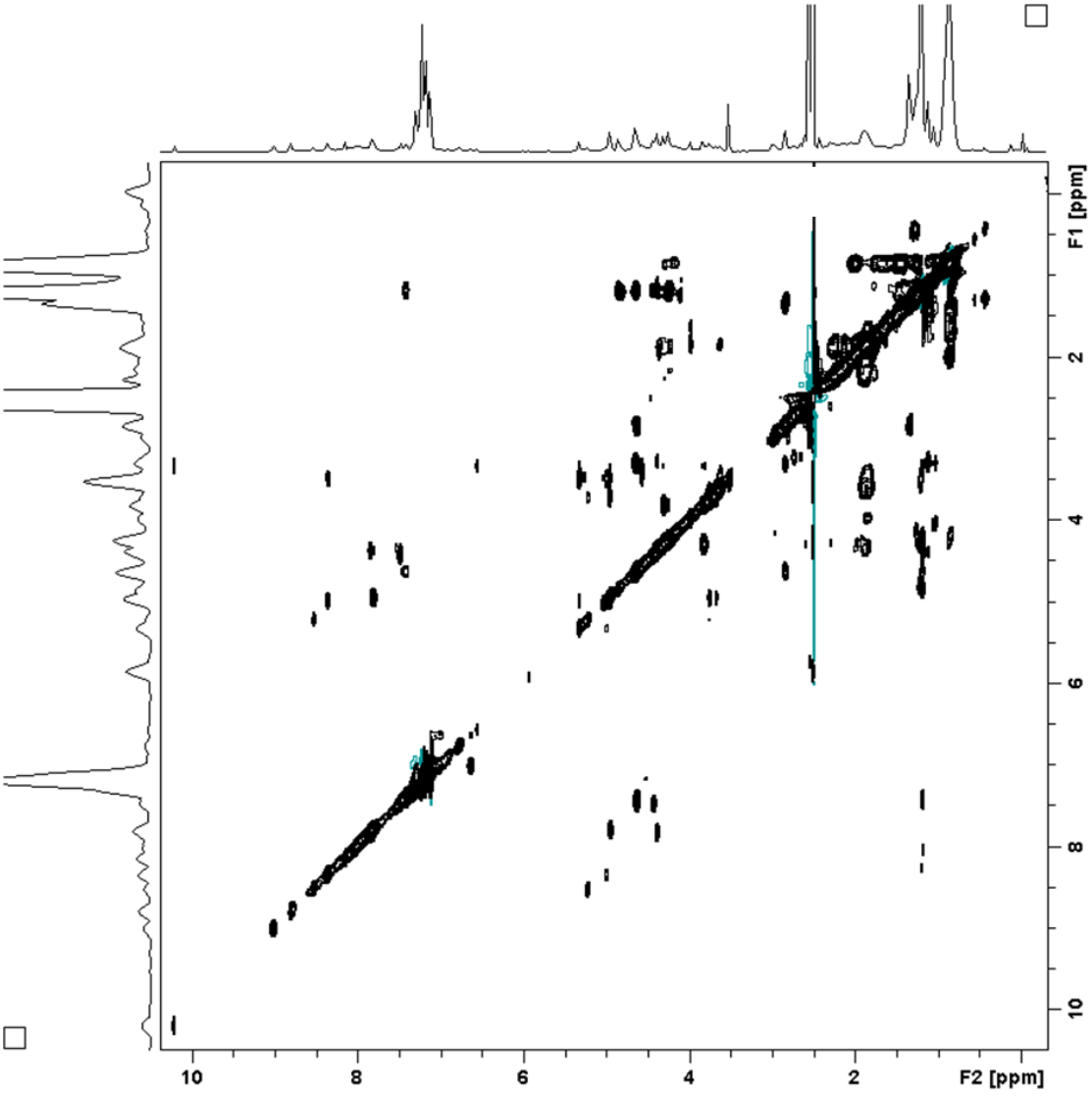
UV/Vis (MeOH): $\lambda_{max}$ (log ε) = 205 (3.53), 249 (2.72), 313 (2.35) nm.

$[\alpha]_D^{20} =$ -10.93

Supplementary Figure 6. HSQC NMR spectrum (DMSO, 800 MHz) of kyamicin.



Supplementary Figure 7. TOCSY NMR spectrum (DMSO, 800 MHz) of kyamicin.



Supplementary Figure 8. NOESY NMR spectrum (DMSO, 800 MHz) of kyamicin.