THE INVESTIGATION INTO THE SYNTHESIS OF 2,5-DIPHENYLOXAZOLE IN *Streptomyces polyantibioticus* SPR^{T}

BY

DARREN EDWARD STEGMANN

THESIS PRESENTED FOR THE DEGREE OF MASTER OF SCIENCE

IN THE DEPARTMENT OF MOLECULAR AND CELL BIOLOGY

FACULTY OF SCIENCE

UNIVERSITY OF CAPE TOWN

JANUARY 2011

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would first and foremost like to thank Dr Paul Meyers for his supervision, constant guidance and advice during the course of this dissertation.

I would also like to extend my gratitude to Dr Gareth Everest for his helpful suggestions and opinions with experimental procedures and writing.

My thanks also go out to Dr Marilize Le Roes-Hill for her work which set the foundation for this project.

I also want to extend my appreciation to the rest of the laboratory, past and present, including Bobby, Jeffery, Iulia, Iric and Saeed for the encouragement and interesting times shared.

I am also indebted to Di James and Bruna Galvão for DNA sequencing and the National Research Foundation and the Medical Research Council for funding.

My greatest thanks go out to my family for their love, support and patience during my studies.

ABSTRACT

As part of an antibiotic-screening programme, an actinomycete, *Streptomyces polyantibioticus* SPR^T, was isolated from soil collected from the banks of the Umgeni River, KwaZulu-Natal Province, South Africa. It exhibited antibiosis against *M. tuberculosis* H37Rv^T, prompting interest in its antibiotic production. An antibiotic produced by *S. polyantibioticus* SPR^T was isolated and its structure determined by nuclear magnetic resonance (NMR) and X-ray crystallography to be 2,5-diphenyloxazole (DPO). Of great interest is the independent confirmation of the antibiotic activity of DPO and extension of the data to show activity against non-replicating persistent cells of *M. tuberculosis.* It seems likely that 2,5-DPO is synthesized non-ribosomally by *S. polyantibioticus* SPR^T. It is proposed that DPO is synthesised from the starting units of benzoic acid and β-hydroxyphenylalanine or phenylalanine, undergoing peptide bond formation followed by cyclization and decarboxylation to form DPO. A nonribosomal peptide synthetase (NRPS) possessing an adenylation domain specific for phenylalanine or the existence of cyclization domains could indicate its ability to synthesis DPO. The presence of a phenylalanine ammonia-lyase (PAL) coding gene within *S. polyantibioticus* SPR^T would allow it to synthesise benzoyl-CoA for DPO production. Three unique adenylation domains were isolated from *S. polyantibioticus* SPR^T and were designated DS-D, PJ15 and PJ73. It was shown by the NRPSpredictor program that DS-D is specific for glycine, alanine or valine; PJ15 for ornithine, lysine or arginine; PJ73 for serine, threonine or dehydrothreonine. These adenylation domains were used as probes in Southern hybridization experiments to identify larger DNA fragments from their respective NRPS genes. 28 adenylation domain fragments and flanking sequences were detected via Southern hybridization, but it proved challenging to clone these fragments. Polymerase Chain Reaction (PCR) primers were designed to amplify the *encP* gene (for phenylalanine ammonia-lyase) from '*Streptomyces maritimus*' DSM 41777^T and it was shown by PCR and Southern hybridization that *encP* could be detected in '*S. maritimus*'. However, an *encP*-homologue could not be detected within the genome of *S. polyantibioticus* SPR^T by these methods. Cyclization-domain PCR primers were designed based on sequences from characterised *Streptomyces* and *Streptoalloteichus* thiazole and oxazole producers. No cyclization domains could be amplified from *S. polyantibioticus* SPR^T or the control strains. Future work should involve the identification or confirmation of a PAL-like gene within *S. polyantibioticus* SPR^T and further attempts at cloning adenylation domain sequences detected via Southern hybridization. Cyclization domains and an adenylation domain specific for phenylalanine may still be found within *S. polyantibioticus* SPR^T. Knowledge of the gene cluster responsible for production of DPO will provide much insight into its synthesis and allow combinatorial biosynthesis studies to commence and create novel derivatives of DPO for the treatment of different types of drug resistant tuberculosis (DR-TB).

CHAPTER 1

INTRODUCTION

CONTENTS

CHAPTER 1

INTRODUCTION

Since the discovery of the antibiotic, penicillin, by Sir Alexander Fleming in 1928 (Nobel Foundation, 2010), humanity and microbial pathogens have been in a continuous arms race. The relentless evolution of drug resistant strains of pathogens is placing methods for their treatment under serious pressure. In turn, for mankind to ensure its survival, it must adapt to meet this threat. The most promising solution is to accelerate the discovery of novel antibiotics against which pathogens would have no resistance. Furthermore these novel antibiotics and their molecular scaffolds may be able to yield novel derivatives, through direct chemical reactions or by tailoring the genes which code for the antibiotic itself. This would substantially increase the size of the arsenal of effective antibiotics which we so desperately need.

1.1 ANTIBIOTIC RESISTANCE

In recent years, three main classes of antibiotic-resistant pathogens have emerged as major threats to public health. The first is methicillin-resistant *Staphylococcus aureus* (MRSA) (Figure 1.1). It causes a high mortality rate and its rising prevalence increases the likelihood that vancomycin-resistant *S. aureus* (VRSA) may emerge, an even more challenging pathogen to treat (Fischbach & Walsh, 2010).

The second class is the multidrug-resistant (MDR) and pan-drug-resistant (PDR) Gram-negative bacteria such as strains of *Acinetobacter baumannii* (Figure 1.1), *Escherichia coli*, *Klebsiella pneumoniae*, and *Pseudomonas aeruginosa* (Figure 1.1). This class is less prevalent than MRSA, but many infections may be untreatable resulting in death. The class is notoriously difficult to treat, because their outer membrane blocks the entry of some antibiotics, and efflux pumps expel many that successfully enter the cell (Fischbach & Walsh, 2010).

The third class consists of MDR and extensively-drug-resistant (XDR) strains of *Mycobacterium tuberculosis* (Figure 1.1). This pathogen is responsible for the disease tuberculosis (TB). It is transmitted by respirable droplets generated during forceful expiratory actions such as coughing (Konstantinos & Health, 2010). Tuberculosis normally attacks the lungs of the infected individual, but can also attack other parts of the body such as the kidneys, spine, and brain (CDC, 2010). Symptoms of TB include a chronic cough with blood-tinged sputum, fever, night sweats, and weight loss (Konstantinos & Health, 2010). MDR-TB and XDR-TB are a rising threat in the developing

world (Fischbach & Walsh, 2010). This can mostly be attributed to patients not completing their course of antibiotic therapy and to co-infection by the human immunodeficiency virus (HIV), which causes acquired immunodeficiency syndrome (AIDS). Patients suffering from AIDS are extremely susceptible to further infections due to their compromised immune systems. Furthermore, they are defenceless against forms of drug resistant TB, which often prove fatal (CDC, 2010). In addition to a high HIV prevalence, patients in the developing world also suffer from malnutrition, poverty and overcrowding, further increasing their susceptibility to TB (Cape Gateway, 2006).

MDR-TB and XDR-TB arise from the incorrect or irregular use of prescribed drugs for ordinary TB infections. The usage of inappropriate drug combinations, using single drugs for ordinary TB, clinics running out of drug stocks, inadequate counselling of patients leading to patients not taking their treatment correctly, poor treatment compliance or patients not returning for treatment all may result in the development of forms of drug resistant TB (Cape Gateway, 2006).

MDR-TB is defined as being resistant to at least two of the best anti-TB drugs, isoniazid and rifampicin. These drugs are considered first-line drugs and are used to treat anyone who has contracted TB. XDR-TB is defined as TB that is resistant to isoniazid and rifampin, plus resistant to any fluoroquinolone and at least one of three injectable second-line drugs (i.e., amikacin, kanamycin, or capreomycin) (CDC, 2010).

The rise of HIV infection coupled with the emergence of MDR-TB and XDR-TB is a serious concern. In Africa, HIV is the single most important factor contributing to the increase in the incidence of TB since 1990 (WHO, 2010) and the World Health Organization (WHO) ranks South Africa the 5th most TB burdened country in the world, having an average incidence rate of 948 tuberculosis cases per 100 000 people in the population per annum and 218 TB related deaths per 100 000 people in the population per annum. It is now more important than ever that novel drugs are discovered or created against MDR-TB and XDR-TB (WHO Global TB report, 2009).

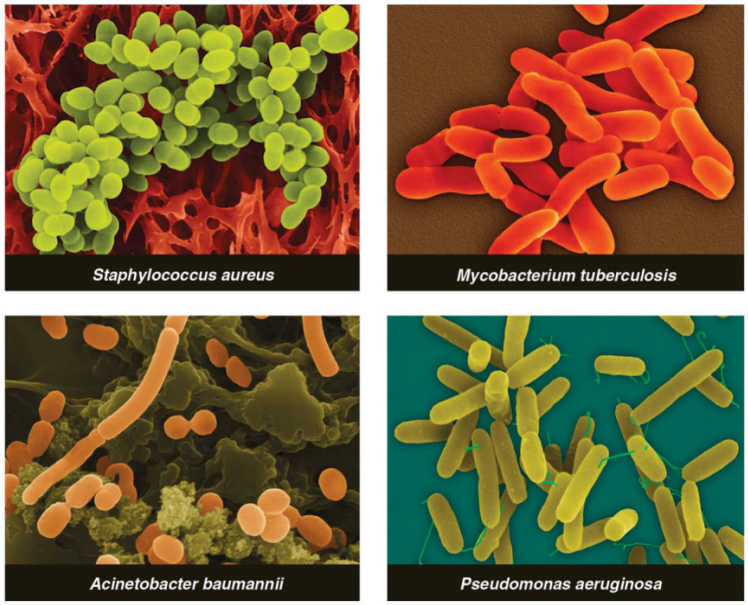

Figure 1.1 False-colour scanning electron micrographs of multidrug-resistant bacteria. Taken from Fischbach & Walsh, 2010.

1.2 ANTIBIOTIC DISCOVERY

Antibiotics (from the Ancient Greek: ἀντί – *anti*, "against", and βίος – *bios*, "life") (Davey *et al.*, 2000) are defined as substances or compounds that kill bacteria or inhibit their growth. The term antibiotic was coined by Selman Waksman in 1942 to describe any substance produced by a microorganism that is antagonistic to the growth of other microorganisms (Waksman, 1947). The value of antibiotics to medicine in combating bacterial infections was recognised, which lead to the increased interest in the isolation and discovery of a variety of antibiotics.

Today, more than two-thirds of the antibiotics used for the treatment of human infections are isolated from microbes as natural products or are semisynthetic derivatives of these molecules (Fischbach, 2009). Natural products are defined as chemical molecules obtained from living organisms (Lefevre *et al.*, 2008) and they are mainly divided into two classes. The first class is the primary metabolites. They are essential molecules that are globally shared by almost all living species and include proteins, fatty acids, sugars and nucleotides. The second class is secondary metabolites, which are lower molecular weight molecules of extremely diverse chemical structures that are produced by specific species of microbes. Secondary metabolites are generally produced after growth has stopped. Many of these secondary metabolites possess significant value to medicine in the form of antibiotics, immunosuppressants, anti-inflammatories, toxins, siderophores and anticancer agents (Zhang *et al.*,

2009; Yanai *et al.,* 2006). Microbial secondary metabolites account for 10% of all bioactive natural products known today (Lefevre *et al.*, 2008).

In particular, most natural product antibiotics have been isolated from soil actinomycetes which are described as Gram-positive, aerobic bacteria that produce branching filamentous hyphae and also possess a high GC (guanine plus cytosine) content in their DNA (Lechevalier & Lechevalier, 1981; Fischbach & Walsh, 2010). These antibiotics from actinomycetes and other microbes have been evolving for roughly 1 billion years and their effectiveness is proven by their ability to penetrate other microbes and inhibit target enzymes, macromolecules or structures (Baltz, 2008).

Soils have been screened for many years for bioactive-compound producing bacteria and only a miniscule portion of the surface of the planet has been sampled so far, resulting in only a small fraction of actinomycete taxa being discovered. Furthermore, only the tip of the iceberg of soil actinomycetes and their antibiotic products have been sampled in the past 50 years. Pharmaceutical companies typically screen tens of thousands of actinomycetes per year because of their vast untapped wealth of novel antibiotics and other secondary metabolites. It is very likely that there are still more novel antibiotics produced by novel actinomycetes waiting to be discovered (Baltz, 2008). It is therefore imperative to search underexplored ecological niches and identify new bacterial taxa, which may also provide increased antibiotic diversity. This approach has been successful and has revealed many novel natural products. Marine niches, such as deep-sea sediments (Bull & Stach**,** 2007; Gulder & Moore, 2009) and actinomycetes associated with marine sponges have proven fruitful in natural product discovery (Gandhimathi *et al.*, 2008). Terrestrial symbioses are also promising ecological goldmines. Some recent studies on the bacterial symbionts of insects, ascidians, and fungi have yielded many new natural products (Fischbach & Walsh, 2010).

Actinomycete genome sequencing has revealed that actinomycetes have large genomes when compared to their bacterial cousins and they use a significant fraction of their coding capacity (5-10%) for the production of secondary metabolites, often antibiotics (Baltz, 2008). Approximately 100 genomes of archaea and bacteria have been sequenced so far as part of the Genomic Encyclopedia of Bacteria and Archaea (GEBA) project. This brings the total number of sequenced actinobacterial genomes to 125 (as at 29 December 2010; http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi). GEBA has shown that with the sequencing of the genomes of diverse bacterial genera, novel genes are being revealed (Wu *et al.,* 2009).

1.3 ANTIBIOTIC PRODUCTION

Antibiotics vary greatly in structure and the mechanisms by which they are synthesised are also diverse. Peptide antibiotics are known to include many unusual residues and modifications. Some of these scarce residues include D-amino acids, β-amino acids, hydroxyl-amino acids and N-methylated residues, with a few examples of the common modifications to these peptides being acylation, glycosylation and heterocyclic ring formation (Kleinkauf & van Döhren, 1990, Marahiel *et al.,* 1997, Marahiel, 1997, Konz & Marahiel, 1999). Their biosynthetic origin has been linked to an enzymatic system termed as the 'thiotemplate multienzymic mechanism' (Kleinkauf & van Döhren, 1990), which produces these peptides non-ribosomally.

This enzymatic system is more commonly referred to as a non-ribosomal peptide synthetase (NRPS). The biological role of NRPSs in their producing organisms is often unclear, however an interesting diversity of pharmacological and medicinal properties can be linked to the non-ribosomal peptides produced by them. Many important antibiotics are synthesized by NRPSs, including penicillins, cephalosporins and glycopeptides (Stachelhaus *et al.*, 1999). In addition to antibiotics, many nonribosomal peptides are cytostatic, antiviral, antitumor, immunosuppressive or enzyme inhibitory compounds (Marahiel, 1997).

Polyketide synthetases (PKSs) are also linked to antibiotic production, specifically the complex carbon skeletons of polyketides of both the macrolide class of large glycosylated esters, such as erythromycin A, and the polycyclic (aromatic) class, such as oxytetracycline. . There are three major types of PKSs: Type I, Type II (aromatic) and Type III PKSs. Type I PKSs are characterised by the linking of the multiple catalytic domains to form megasynthase polypeptides. In comparison, Type II PKSs have separate catalytic domains that form subunits (Walsh *et al.*, 1997). Type III PKSs do not have acyl carrier protein (ACP) domains. The biosyntheses of polyketides share striking similarities with fatty acid biosynthesis (Khosla *et al*., 1999).

Polyketides and nonribosomal peptides are two large families of natural products. They are biosynthesized with comparable logic by multimodular enzymes acting in assembly line arrays. Copolymerization via mixed modules of PKS and NRPS machinery may be used to assemble hybrid polyketide-nonribosomal peptide molecules with useful therapeutic activity (Walsh, 2004).

It is well established that species with a high amount of secondary metabolite production often possess a high number of PKS and NRPS genes. The discoveries from genome sequencing support this fact. The genome of the avermectin producer, *Streptomyces avermitilis,* encodes 29 secondary

metabolites, including 8 nonribosomal peptides and 11 polyketides. Furthermore, the genome of the erythromycin producer, *Saccharopolyspora erythraea*, encodes for 25 potential secondary metabolites, which include 7 nonribosomal peptides and 11 polyketides (Baltz, 2008). The number of potential metabolites suggests that natural product biosynthetic capacity has been severely underestimated, because only some of these predicted molecules have been detected in culture (Walsh, 2004).

Many important antibiotic pharmacophore and medicinal chemical derivatives from nonribosomal and polyketide molecules such as vancomycin, erythromycin, tobramycin, capreomycin, daptomycin, chloroeremomycin, echinocandin B and many β-lactam antibiotics were discovered and developed by Eli Lilly and Company, (Baltz, 2005). In addition, the natural product templates of actinonin, pleuromutilin, ramoplanin and tiacumicin B, which are compounds undergoing clinical evaluation, represent novel templates not found in currently marketed antibacterial drugs and show promise for future drug development, (Butler & Buss, 2006). NRPSs and PKSs provide invaluable scaffolds for future drug design. With the understanding of the enzymatic processes involved and the genes that code for them, it is possible to tailor these enzymes to produce novel derivatives.

Daptomycin, produced via an NRPS mechanism in *Streptomyces roseosporus*, is a prime example of combinatorial biosynthesis to exploit its scaffold and obtain novel molecules. Combinatorial biosynthesis is defined as the use of mutation or genetic engineering of biosynthetic pathways, or precursor feeding to generate libraries of compounds related to, but different from, known secondary metabolites, increasing their structural diversity (Baltz, 2008; Floss, 2006). The biosynthetic pathway of *S. roseosporus* has been engineered by gene disruption, mega-gene replacements, and module exchanges, coupled with natural lipidations to produce novel derivatives, many of which are very potent antibiotics (Baltz, 2008).

1.4 NONRIBOSOMAL PEPTIDE SYNTHETASES

These large multifunctional enzymes range from 100-1700 kDa in size (Marahiel *et al.*, 1997). The NRPSs all have a unique modular structure in which each particular module has a specific function in recognition, activation or modification of a single substrate amino acid residue in the final non-ribosomal peptide product (Marahiel *et al.,* 1997). The number of these modules is equal to the number of amino acids found in the backbone of the corresponding product peptide (Kopp *&* Marahiel, 2007). Each module in the enzyme consists of many domains which are involved in the substrate adenylation and acylation. They are aligned in such a way as to be collinear with the sequence of the peptide product (Marahiel *et al.,* 1997) allowing the amino acid sequence to be predicted from the NRPS gene sequence. This is often termed the colinearity rule. The modules operate independently from each other, but they work together to catalyze the formation of the final peptide (Stachelhaus *et al.,* 1995). It is important to note that there are several exceptions to the colinearity rule, particularly for NRPSs that assemble siderophores (Challis & Naismith, 2004), which are high-affinity iron chelating compounds.

In bacteria, the genes coding for NRPS are organised into operons, which are usually 6-45 kilobases in length and each polypeptide synthetase (PPS) consists of between 1 and 8 modules (Marahiel *et al.,* 1997). Bacterial protein templates are often much smaller than their fungal counterparts, which can reach up to 46 kilobases in length, e.g. the cyclosporine synthetase gene from *Tolypocladium niveum* (Weber *et al.,* 1994). More importantly, each NRPS module consists of many defined domains which catalyze specific reactions within the sequence of nonribosomal peptide synthesis (Stachelhaus *et al.*, 1995, Marahiel *et al.*, 1997, Von Döhren *et al.,* 1997). These domains are sometimes described as the 'tool box' of PPSs (Konz & Marahiel*,* 1999). Each domain has in possession, a highly conserved signature sequence, consisting of homologous repeating units, which bestows upon the domain a unique fingerprint (Marahiel, 1997). This repetitive arrangement of functional units creates unique sequences and contains important residues, which are directly involved in the reaction (Konz & Marahiel*,* 1999) and are able to recognise and activate specific amino acids.

There are many different kinds of domains present in each module. The most common domains are adenylation, thiolation (also known as peptidyl carrier protein - PCP), tailoring/editing, condensation, cyclization, thioesterase, oxidation, reductase and communication-mediating domains. The minimum number of domains a module must consist of is three and these domains must at least consist of an adenylation (A), condensation (C) and thiolation (PCP) domain in the order C-A-PCP (Challis *et al*., 2000; Tanovic *et al*., 2008) (Figure 1.2).

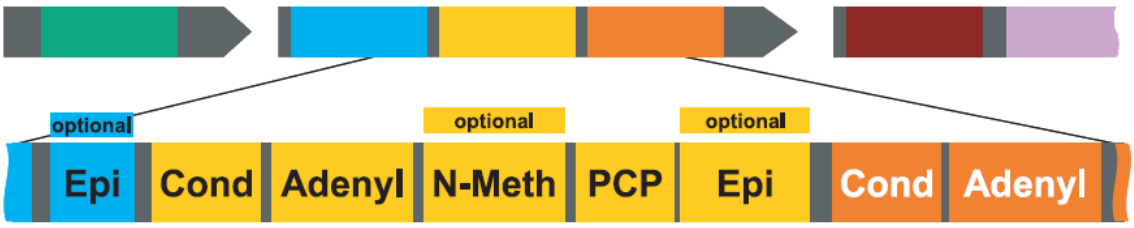

Figure 1.2 Module and domain structure of NRPSs. Top-middle image: one complete NRPS consisting of three modules. Below: enzymatic domains that are contained in a complete module: **Cond**: condensation domain, **Adenyl**: adenylation domain, **N-Meth**: N-methylation domain (optional, does not appear in all NRPSs), **PCP**: thiolation domain (Peptidyl Carrier Protein domain), **Epi**: epimerization domain (also optional). Other optional domains are cyclization, oxidation, reduction, thioesterase and communication-mediating domains. Adapted from Rausch *et al.* (2005).

The formation of successful peptide bonds involves a reaction sequence to be carried out on each module. The sequence proceeds as follows; "ATP-dependent amino- or carboxylic-acid activation and the release of pyrophosphate, followed by transfer of the acyl adenylate-AMP to a specific thiol group and formation of a carboxy-thioester-bound intermediate" (Marahiel, 1997). The thiol group belongs to 4'-phosphopantetheine (4'-PP), which is a cofactor. 4'-PP is covalently bound to each thiolation (PCP) domain. After this reaction has been completed, the thioesterified substrates on the modules are transpeptidised (Marahiel, 1997). Thus the peptide chain is assembled (Figure 1.3).

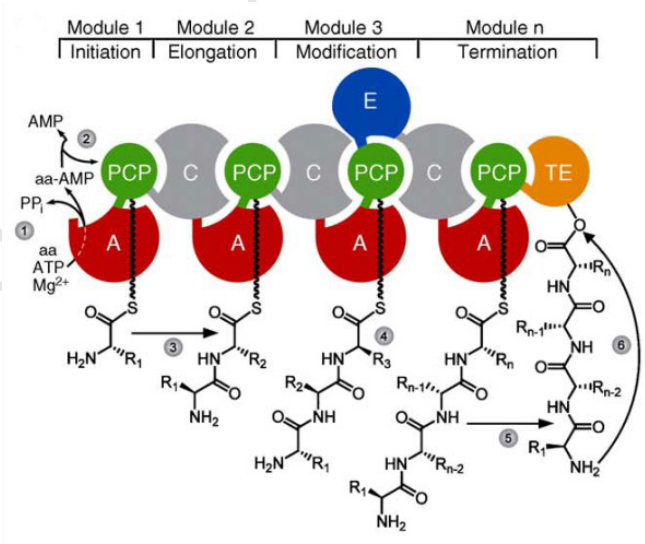

Figure 1.3 A Simplified mechanism of nonribosomal peptide (NRP) synthesis. **(1)** The amino acid is activated as an aminoacyl-AMP by the adenylation domain (A). **(2)** Transfer of the amino acid onto the PCP domain (PCP). **(3)** Condensation of PCP-bound amino acids. **(4)** Possibility of amino acid modifications, for example by epimerization domains (E). **(5)** Transesterification of the peptide chain from the terminal PCP onto the TE domain (TE). **(6)** TE catalyzed product release by either hydrolysis or macrocyclization. The number of modification domains and modules are variable. Adapted from Strieker *et al.* 2010.

1.4.1 THE ADENYLATION DOMAIN

The adenylation domain is often called the essential region of the module, as it contains the substrate-recognition and ATP-binding sites and is thus responsible for the activation of the specific substrate amino acid and formation of its acyl adenylate through ATP hydrolysis. The adenylation domain is a member of a large superfamily of adenylate-forming enzymes. The members of this superfamily have highly conserved signature sequences of about 550 amino acids and activate their substrate amino or carboxylic acids as the corresponding acyl adenylates (Marahiel, 1997). Adenylation domains are thought to operate as functionally independent units in addition to acting in concert with the surrounding domains in the NRPS (Konz & Marahiel, 1999).

The crystal structures of the adenylation domains of two members of the adenylation superfamily have been solved. These domains belong to firefly luciferase of *Photinus pyralis* and the phenylalanine-adenylation (PheA) domain of GrsA (Conti *et al.*, 1997). The crystal structures have very similar topologies, even though they only share 16% amino acid sequence identity and their substrates are structurally different. We can conclude that all adenylation domains of PPS have a similar 3-D structure if they are at least 30-60% similar to PheA. PheA has two subdomains, a large amino-terminus and a smaller, compact carboxy-terminus. Water molecules mediate interactions between the subunits. Most of the amino acid residues involved in substrate binding are associated with the amino-terminal domain, however, a lysine residue, located in the carboxy terminal domain, provides an important interaction in coordinating the α-carboxyl group of the substrate amino acid. There is the possibility of making specific alterations of the residues in the binding sites to modify enzyme activities (Marahiel, 1997).

1.4.2 THE THIOLATION (PCP) DOMAIN

The adenylation domain is followed by an equally important thiolation (PCP) domain, which is about 100 amino acids in length. The thiolation domain shares a high homology with the acyl carrier protein (ACP) involved in fatty acid synthesis. Therefore the thiolation domain has earned the name peptidyl carrier protein (PCP) (Stachelhaus *et al.*, 1996). Carrier proteins may either be freestanding or embedded in enzymatic systems. Another kind of carrier protein is the aryl carrier protein (ArCP), which is commonly found in siderophore NRPSs (Qiao *et al*., 2007).

The 4'-PP is bound to the side chain of a serine residue located in a highly conserved signature sequence in the thiolation domain (Lambalot *et al.*, 1996). When the 4'-PP is bound, the PPS is in its active *holo* form and when unbound it is in its inactive *apo* form (Marahiel, 1997). This modification of 4'-PP is catalyzed by enzymes belonging to the superfamily of 4'-PP transferases (Konz &

Marahiel, 1999). These transferase enzymes promote nucleophilic attack of the serine hydroxyl group on the pyrophosphate bridge of CoA. This results in a transfer of the 4'-PP to the thiolation domain and thus liberation of 3,5-ADP (Lambalot *et al.*, 1996). The PCP domain is responsible for the transport of energy-rich thioester-bound substrates and all elongation intermediates between the catalytic active domains, via the 4'-PP attached to it (Kopp & Marahiel, 2007).

PCP domains shuffle reaction intermediates among different catalytic NRPS domains and are capable of adopting alternative conformational states. Adenylation and condensation domains associate closely to form a catalytic platform, with their active sites on the same side of the platform. The peptidyl carrier protein domain is flexibly tethered to this platform and thus can move with its substrate-loaded 4´-PP arm between the active site of the adenylation domain and the donor side of the condensation domain (Tanovic *et al*., 2008).

1.4.3 THE CONDENSATION DOMAIN

Condensation domains direct the condensation between adjacent thioesterified intermediates and are approximately 450 amino acids in length (De Crecy-Lagard *et al.*, 1995). They are normally found upstream of the adenylation domains in modules involved in peptide elongation or, in other words, between two modules mediating peptide elongation. A critical point is that they are absent from modules involved in peptide initiation. Even though there is not sufficient biochemical evidence of their function, this difference in location supports the idea of their hypothesised role (Smith *et al.*, 1990, Turgay *et al.*, 1992). There is also little information concerning the exact mechanism of peptide elongation and how the interaction of modules affects the direction of polymerization. It is thought that the condensation domain catalyzes the nucleophilic attack of the downstream PCP-bound acceptor amino acid with its α-amino group on the activated thioester of the upstream PCP-bound donor amino acid or peptide (Stachelhaus *et al.*, 1999; Bergendahl *et al.*, 2002). This intermediate is then passed on to other, subsequent condensation or modification domains for elongation and structural changes (Kopp & Marahiel, 2007).

Condensation domains contain many core motifs termed C1 to C7 and also share a highly conserved core sequence with a group of acyl transferases, which includes dihydrolipoyl transacetylases and chloramphenicol acetyltransferases (Marahiel *et al.*, 1997). The best conserved sequence motif, called C3, is crucial in non-ribosomal peptide synthesis and has been identified as HHXXXDG (where X is any amino acid residue) (Keating *et al*., 2002). In these acyl transferases, the second histidine of the HHXXXDG motif is thought to act as the general base promoting nucleophilic attack of the hydroxy moiety of chloramphenicol and of the thiol group of CoA on the carbonyl carbon atom of the acetyl thioester (Marahiel *et al*., 1997). Thus, in NRPSs, it can be assumed that the second histidine of the conserved motif promotes nucleophilic attack of the free α-amino group of the

downstream PCP-bound acceptor amino acid on the activated thioester of the upstream PCP-bound donor amino acid/peptide. Additionally, dual epimerisation/condensation domains, which incorporate D-amino acids into peptides despite the absence of epimerisation domains, exhibit a second His-motif, HH[I/L]XXXXGD, which is located close to the N-terminus of the domain (Rausch *et al*., 2007).

Condensation domains may control the elongation of peptides and perhaps the termination of peptide elongation, however, there is not enough evidence to support this. Possible termination reactions may involve cyclization, thioesterase-driven hydrolysis (via the thioesterase domain) or transfer of the peptide chain to other functional groups. A domain responsible for this might be an extra condensation domain (or equivalent) at either end of the module in comparison to normal condensation domains found between modules (Konz & Marahiel, 1999).

1.4.4 THE EDITING/TAILORING DOMAIN

For the incorporation of less common amino acids, such as D-amino acids or N-methylated amino acids, modules include certain domains known as editing or tailoring domains, between their adenylation and thiolation domains (Marahiel, 1997, von Döhren *et al.*, 1997). More specifically they are called epimerisation or methylation domains, which are 400 amino acids and 420 amino acids in length, respectively. The methylation domain has been shown to contain 3 core motifs including a glycine-rich sequence, which is very similar to the S-adenosylmethionine (SAM) binding site in a heterologous class of cosubstrate-dependent methyltransferases. It has been shown that this domain catalyzes the N-methylation on the thioester prior to peptide-bond formation (Haese *et al.*, 1993). These domains also possess a conserved HHXXXDG motif, as seen in condensation domains (Keating *et al*., 2002).

Similar conditions are found in modules making use of D-amino acids, where they have the epimerisation domain added to the carboxy terminus of the thiolation domain. However, there are known modules that are able to incorporate D-amino acids without an epimerization domain, e.g. chloroeremomycin and cyclosporin. In these examples, the adenylation domain only recognises D-amino acids that have been provided by an external epimerase (Hoffman *et al.*, 1994). It has been noted that some tailoring enzymes do not form part of the NRPS, but act externally. They may act as separate subunits, *in trans*, or may be embedded within the assembly line at relevant modules, acting *in cis* (Walsh *et al.*, 2001, Koglin & Walsh, 2009).

1.4.5 THE THIOESTERASE DOMAIN

Thioesterase domains are involved in incorporating the last amino acid into the final peptide and are of 250 amino acids in length (Marahiel, 1997, Marahiel *et al.*, 1997, von Döhren *et al.*, 1997). Thioesterase domains have a homologous sequence to thioesterases and also have a specific signature sequence. The fully lengthened peptide, which is bound to the last thiolation domain, is transferred to the hydroxyl group of the serine residue (part of the conserved signature sequence) in the thioesterase domain, to create an acyl-O-enzyme intermediate. The residues of serine at position 80 (Ser80), histidine at position 207 (His207) and aspartate at position 107 (Asp107) form the catalytic triad of residues in the active site of the thioesterase domain.

This acyl-O-enzyme bond is then cleaved by acyltransfer to water, which releases the peptide from the module via hydrolysis (Cane *et al.*, 1998; Tanovic *et al.*, 2008) or a reaction with an intramolecular nucleophile results in a cyclic structure. This macrocyclization dominates among the bulk of NRPSs (Kopp *&* Marahiel, 2007). Thioesterase domains function in disconnecting the covalent linkage between the full-length peptidyl chain and the 4'-PP thiol arm of the most downstream thiolation domain (Bruner *et al*., 2002). The thioesterase domain is normally positioned on the C-terminal of the NRPS module where it acts as a terminator of peptide elongation (Tanovic *et al*., 2008).

1.4.6 THE REDUCTION DOMAIN

Reduction domains catalyze the reductive cleavage of the associated thiolation domain's acyl group, which results in the release of a linear aldehyde. They are located on the carboxy-terminal ends of modules and are about 350 amino acids in length. They also contain a NAD(P)H binding site to allow their reductive properties. It is possible that the reduction domain could serve as a tool to allow for peptide chain termination. Other possible domains which could allow peptide chain termination are condensation and cyclization domains (Konz & Marahiel, 1999).

1.4.7 THE CYCLIZATION DOMAIN

Analysis of the structure of the formed products from certain condensation domains suggests that they are involved in peptide chain termination and cyclization. This cyclization may be the result of the formation of an amide bond or the formation of an ester bond and the domains involved in this must therefore be able to catalyze two different types of nucleophilic attack. Condensation domains that are involved with cyclization share a high similarity with cyclization domains. Cyclization and condensation domains have both been shown to carry out their reactions independently (Duerfahrt *et al.*, 2004). In cyclization domains the HHXXXDG motif seen in condensation domains is replaced by

DXXXXD. The conserved aspartate residues are critical for both condensation and heterocyclization (Keating *et al*., 2002).

Cyclization domains are involved in the formation of heterocyclic rings, such as oxazolines and thiazolines, in the peptide backbone (Konz & Marahiel, 1999). These domains are able to substitute condensation domains at the amino terminus inside a module and are able to incorporate serine, threonine or cysteine residues and thus conduct heterocyclization during the peptide bond formation (Gehring & Walsh, 1998, Quadri *et al.*, 1998, Reimmann *et al.*, 1998). This heterocyclization results in molecules with remarkable biochemical properties. These five atom rings are able to chelate metal ions and interact with proteins, DNA, and RNA. Thus peptides containing heterocyclic rings are of great importance to pharmacy and medicine as many important compounds, such as antibiotics, contain heterocyclic rings (Ikai *et al.*, 1992, Du *et al.*, 2003, Walsh, 2004).

1.4.8 THE OXIDATION DOMAIN

Another kind of domain is the oxidation domain. It is thought that this domain is involved in the oxidation of dehydro heterocycles, such as cysteines, serines, threonines, thiazolines and oxazolines, by two-electron transfer to yield heteroaromatic thiazoles or oxazole rings (Schneider *et al.,* 2003; Duerfahrt *et al.*, 2004; Walsh, 2004). Oxidations at the end of peptide synthesis can also create cross-linked and cyclized structures that are often seen in antibiotics, as it increases the compound's affinity for biological targets (Walsh, 2004).

1.4.9 THE COMMUNICATION-MEDIATING (COM) DOMAIN

It is possible that there may be undesired interactions between non-partner modules in an NRPS when modules interact with each other, regardless of their order in the biosynthetic complex. There must therefore be some kind of selective interaction between partner NRPS modules that prevents the false contact between nonpartner modules. Short terminal sequence motifs were found to mediate the specific channeling of reaction intermediates between partner modules, because they contain communication-mediating (COM) domains (Hahn & Stachelhaus, 2004).

There are two types of COM domains. A donor COM domain (COM^D) located at the C terminus of an aminoacyl- or peptidyl-donating NRPS module and an acceptor COM domain (COM^A) located at the N terminus of the accepting partner NRPS module form a matching (compatible) set, required for the proper intermolecular interaction between adjacent modules (Hahn & Stachelhaus, 2006). The structure of the molecular interactions between COM domains has resolved. A helix (COM helix) protrudes from a COM^D on the epimerization domain of the module and a hand-shaped motif (COM hand) from a COM^A on the condensation domain of the N-terminal. The COM hands on condensation

domains act as docking sites for COM helices on the epimerization domains. It is thought that the interaction between the COM helix and the COM hand structures facilitates NRPS-NRPS interactions (Tanovic *et al*., 2008).

Without the selectivity provided by different sets of COM domains, the modules of an NRPS complex would form random biosynthetic products. Hence, these domains are essential for the formation of a defined biosynthetic product and are important for combinatorial biosynthesis in allowing modules to interact with each other correctly by ensuring that their COM domains are compatible (Hahn & Stachelhaus, 2004).

1.5 NRPS SPECIFICITY

Biochemical studies have shown that adenylation domains are the gate keeper units of repeated modules due to their specificity to recognise only certain amino acids. Hence they determine what molecules can be synthesized by the NRPS and act in a 'lock and key' process (von Döhren *et al.*, 1999). Knowledge about which substrates are bound by specific adenylation domains increases the potential use of NRPSs as scaffolds for combinatorial biosynthesis. Mutagenesis or additional reactions would allow the generation of novel derivatives of the nonribosomal peptide and ideally novel, more potent antibiotics.

Numerous methods have been designed to disclose the specificity of adenylation domains. A traditional method for analysis of adenylation domains involves aminoacyl-AMP formation and detection via a radioactive ATP-[^{32}P]pyrophosphate (PP_i) exchange assay. Aminoacyl-AMP and subsequent pyrophosphate formation indicates recognition of the amino acid as a substrate by the adenylation domain (Otten *et al.*, 2007). A non-radioactive method has also been formulated using malachite green to measure orthophosphate (P_i) concentrations after degradation by inorganic pyrophosphatase of the PP_i released during aminoacyl-AMP formation (McQuade *et al.*, 2008).

1.5.1 PREDICTIVE SUBSTRATE BINDING POCKET CODE

The most practical method to determine specificity of adenylation domains, is to examine the residues that interact with the substrate directly. Analysis of the crystal structures of firefly luciferase and the phenylalanine-adenylation (PheA) domain of GrsA have shown that adenylation domains contain some highly conserved core motifs which surround the active site where the substrate binds. These conserved motifs are denoted A1 to A10 and are spread out over approximately 450 amino acids (Figure 1.4a). Site-directed mutagenesis and photoaffinity labelling confirmed that these motifs play a key role in ATP binding and hydrolysis. In contrast, the residues lining the substrate(phenylalanine)-binding pocket of GrsA were found to be located within a ~100 amino acid stretch and possess a lower overall similarity when comparing across adenylation domains (Stachelhaus *et al.*, 1999).

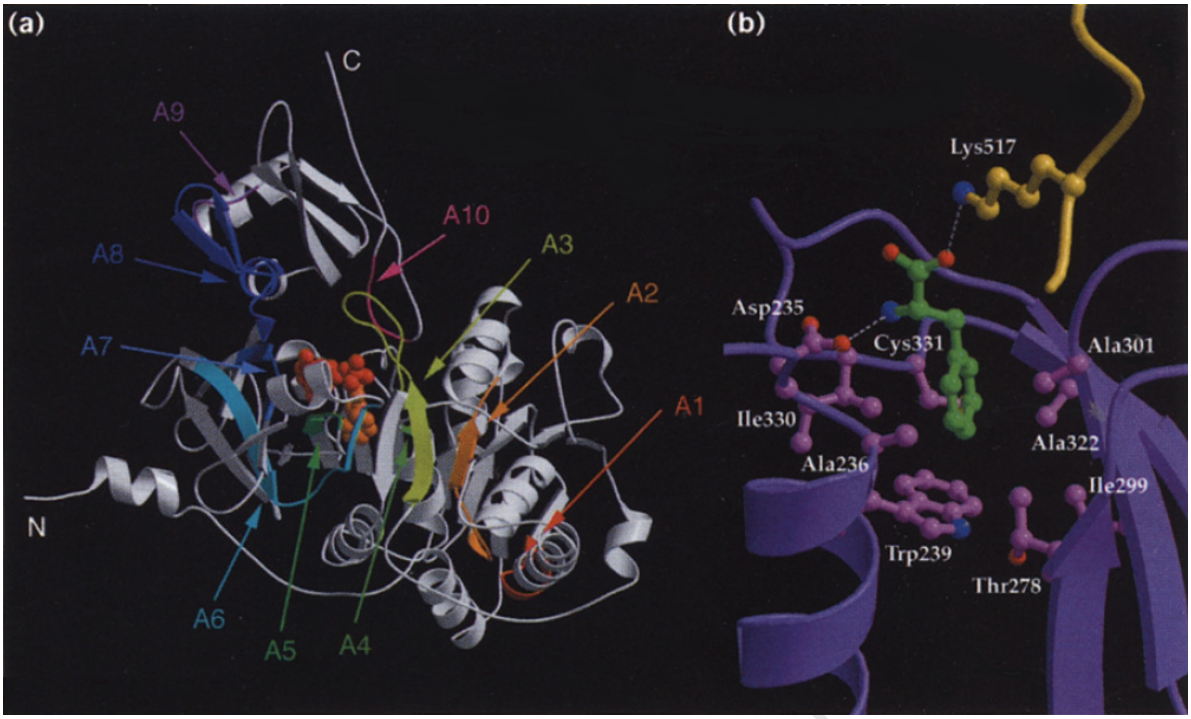

Figure 1.4 Structural basis for phenylalanine activation by PheA, Conti *et al.*, 1997. **(a)** The ribbon diagram of PheA shows how the A domain is folded into a large amino-terminal and a smaller carboxy-terminal subdomain. The AMP (red) and phenylalanine (orange), bound at the interface between the N- and C-terminal domains, are shown in space-filling presentation. The locations of the highly conserved core motifs are denoted as A1 to A10. **(b)** The phenylalanine-specific binding pocket consists of ten residues, Asp235, Ala236, Trp239, Thr278, Ile299, Ala301, Ala322, Ile330, Cys331 and Lys517. Asp235 and Lys517 mediate electrostatic interactions (shown as dotted lines). Adapted from Stachelhaus *et al.*, 1999.

Individual residues in adenylation domains that directly interact with the substrate amino acid are not highly conserved across adenylation domains. Therefore the substrate pockets are specific for each amino-acid substrate. Analysis of the residues in the binding pockets has revealed, "A correlation between the polarity of the substrate sidechain and the sidechains of some residues forming the hypothetical substrate pocket" (Marahiel, 1997). This has allowed us to enhance our understanding of the specific workings and specificity of adenylation domains. The crystal structure of the PheA domain of GrsA has allowed determination of the exact residues that interact with phenylalanine in GrsA (Conti *et al.*, 1997).

In PheA, residue 235, aspartate (Asp235) and residue 517, lysine (Lys517) stabilise the amino acid substrate by electrostatic interactions between the α-amino and α-carboxyl groups of phenylalanine. These 2 residues are located in the conserved core motifs of A4 (Asp235) and A10 (Lys517). Since these residues are located in conserved motifs, they remain conserved. The rest of the specificity (phenylalanine) binding pocket for PheA was determined to be alanine at 236 (Ala236), isoleucine at 330 (Ile330) and cysteine at 331 (Cys331) on one side of the pocket and alanine at 322 (Ala322) and

at 301 (Ala301), isoleucine at 299 (Ile299) and threonine at 278 (Thr278) on the other side of the pocket. These sides are separated by the indole ring of tryptophan-239 at the bottom of the pocket (Figure 1.4b). Because there is a high degree of sequence identity between NRPS adenylation domains, the GrsA structure represents a model that all NRPS adenylation domains could fit (Conti *et al.*, 1997). Therefore, amino-acid residues that correspond to the key positions lining the PheA binding pocket would likewise mediate substrate specificity in other adenylation domains (Stachelhaus *et al.*, 1999).

Alignments of consensus adenylation domain residues with the PheA domain of GrsA revealed that all adenylation domains exhibit the same kind of binding pockets, with different key residues that interact with different amino-acid substrates (Figure 1.5a). More importantly, these alignments have defined sets of conserved recognition templates for particular binding pockets, always specific for a particular amino-acid substrate (Challis *et al.*, 2000) (Figure 1.5b). These signature sequences can be interpreted as the 'code' for NRPSs (Stachelhaus *et al.*, 1999). These rules for specificity sequences have been used to predict substrates from primary sequences of several uncharacterised or unknown adenylation domains. The reliability of this code has been demonstrated in the successful mutation of all the residues in the key positions of the PheA binding pocket, which resulted in an alteration or relaxation in its substrate specificity. Furthermore, a single mutation of histidine-322 to glutamic acid-322 in an aspartate-activating domain (AspA) was predicted and shown to alter the specificity to asparagine (Stachelhaus *et al.*, 1999).

However, the code does feature some redundancy. For example, there are two unique signature sequences for cysteine and four for leucine. Furthermore, not all adenylation domains specific for phenylalanine have the exact specificity binding pocket sequence as GrsA, e.g. BarG of the barbamide synthesis gene cluster from *Lyngbya majuscula* (GenBank Protein Sequence accession number: AAN32981). Additionally, examples of codes, for which the predicted substrate specificity does not correspond to the actual activated amino acid, have been reported (McQuade *et al.*, 2008).

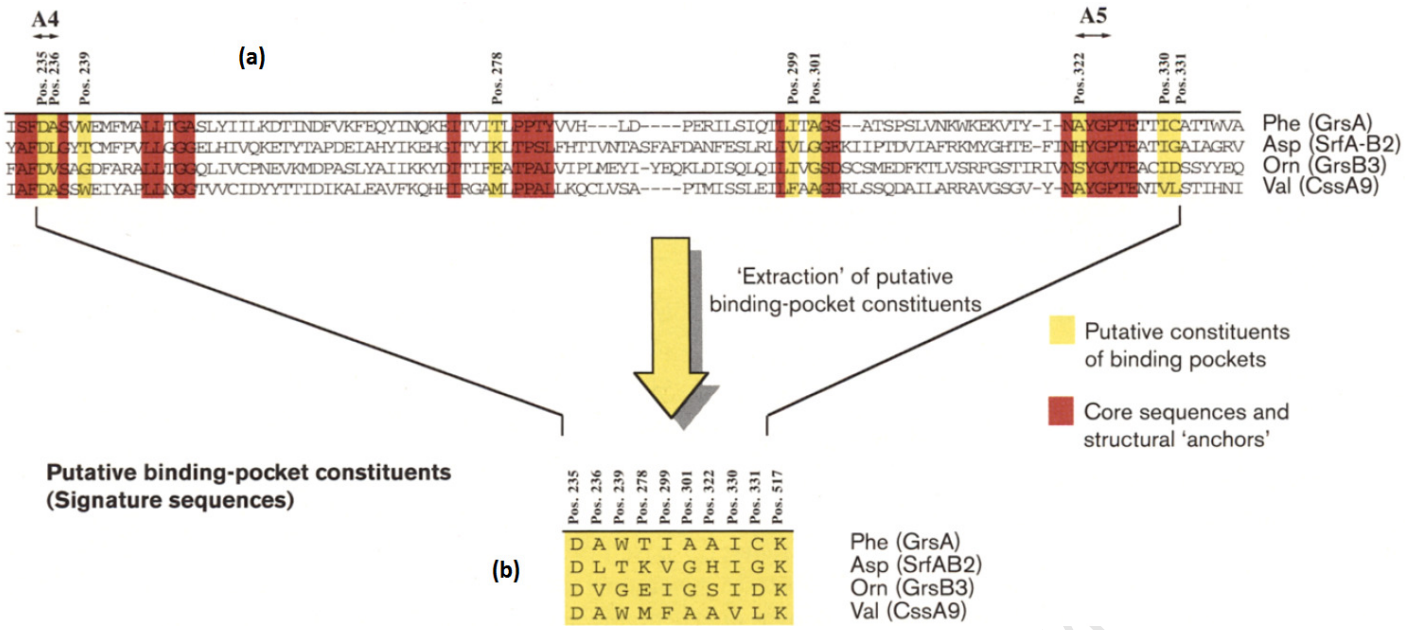

Figure 1.5 An alignment of adenylation domain primary amino acid sequences to determine binding-pocket constituents. **(a)** PheA (Top Row) was aligned against three examples, AspA from SrfA-B2, OrnA from GrsB3 and ValA from CssA9. The sequence of ~100aa between core motifs A4 and A5 (see also Figure 1.4b) contains the signature sequence of an adenylation (A) domain. Yellow residues indicate those involved in the binding pocket positions and brown residues indicate conserved motifs which anchor the alignment. **(b)** Signature sequences involved in the binding pockets of the aligned adenylation domains. The missing residue, Lys517, is highly conserved within motif A10, which is not shown in the protein sequence. This alignment was extended to 160 different adenylation domains to confirm accuracy. Adapted from Stachelhaus *et al.*, 1999.

Furthermore, certain positions may be more variable or be more 'wobble'-like than others among different adenylation domains. Positions 235 and 517 are considered invariant, 236, 301 and 330 are moderately variant and 239, 278, 299, 322 and 331 are highly variant. This variability reflects each position's importance in contributing to substrate specificity, but also causes dissimilarity between adenylation domains specific for identical substrate amino acids (Stachelhaus *et al.*, 1999). Predictions of substrate specificities might be impeded by the limited amount of sequence information available. The accumulation of more sequence and biochemical data will result in the discovery of additional signatures for given specificities. Specificity selection rules in a newly discovered system could be quite different from known adenylation domains and an accurate prediction could be impossible due to insufficient baseline data (Stachelhaus *et al.*, 1999). The specificity of uncharacterised adenylation domains must therefore be inferred based on the available code for domains with known specificity or based on consensus sequences for each substrate amino acid (Rausch *et al.*, 2005). However, there are possible deviations from the specificity code and limitations in its usage. Numerous adenylation domain specificities have been deciphered and a large collection of these is slowly being accumulated (Table 1.1).

TABLE 1.1 Consensus specificity code for substrates from various adenylation domains

	Position											
Domain	235	236	239	278	299	301	322	330	331	517	Biosynthetic template	Similarity
Aad	E	P	R	N	I	V	E	F	V	K	AcvA	94%
Ala	D	L	L	F	G	I	A	V	L	K	CssA, Hts1	55%
Asn	D	L	T	K	L	G	E	V	G	K	BacA, CepA, Dae, Glg1, TycC	90%
Asp	D	L	T	K	V	G	H	I	G	K	BacC, SrfAB, LicB, LchAB	100%
Cys(1)	D	H	E	S	D	V	G	I	T	K	AcvA	96%
Cys(2)	D	L	Y	N	L	S	L	I	W	K	BacA, HMWP2	88%
Dab	D	L	E	H	N	T	T	V	S	K	SyrE	100%
Dhb/Sal	P	L	P	A	Q	G	V	V	N	K	EntE, DhbE, MbtA, PchD, VibE, YbtE	83%
Gln	D	A	Q	D	L	G	V	V	D	K	LicA, LchAA	100%
Glu(1)	D	A	W	H	F	G	G	V	D	K	FenA, FenC, FenE, PPS1, PPS3, PPS4	95%
Glu(2)	D	A	K	D	L	G	V	V	D	K	BacC, SrfAA	95%
Ile (1)	D	G	F	F	L	G	V	V	Y	K	BacA, BacC, LicC, LchAC	92%
Ile (2)	D	A	F	F	Y	G	I	T	F	K	FenB, PPS5	100%
Leu(1)	D	A	W	F	L	G	N	V	V	K	BacA, LicA, LchAA, LicB, LchAB, SrfAA, SrfAB	99%
Leu(2)	D	A	W	L	Y	G	A	V	M	K	CssA	100%
Leu(3)	D	G	A	Y	T	G	E	V	V	K	GrsB, TycC	100%
Leu(4)	D	A	F	M	L	G	M	V	F	K	LicA, LchAA, SrfAA	97%
Orn(1)	D	M	E	N	L	G	L	I	N	K	FxbC	100%
Orn(2)	D	V	G	E	I	G	S	I	D	K	BacB, FenC, GrsB, PPS1, TycC	98%
Phe	D	A	W	T	I	A	A	V	C	K	GrsA, SnbDE, TycA, TycB	88%
Phg/hPhg	D	I	F	L	L	G	L	L	C	K	CepB, CepC, SnbDE	80%
Pip/Pip-@	D	F	Q	L	L	G	V	A	V	K	FkbP, RapP, SnbA, SnbDE	75%
Pro	D	V	Q	L	I	A	H	V	V	K	GrsB, FenA, PPS4, SnbDE, TycB	87%
Ser	D	V	W	H	L	S	L	I	D	K	EntF, SyrE	90%
Thr/Dht	D	F	W	N	I	G	M	V	H	K	AcmB, Fxb, PPS2, PyoD, SnbC, SyrB, SyrE	91%
Tyr(1)	D	G	T	I	T	A	E	V	A	K	FenA, PPS2, PPS4	100%
Tyr(2)	D	A	L	V	T	G	A	V	V	K	TycB, TycC	80%
Tyr(3)	D	A	S	T	V	A	A	V	C	K	BacC, CepA, CepB	78%
Val(1)	D	A	F	W	I	G	G	T	F	K	GrsB, FenE, LicB, LchAB, PPS3, SrfAB, TycC	96%
Val(2)	D	F	E	S	T	A	A	V	Y	K	AcvA	94%
Val(3)	D	A	W	M	F	A	A	V	L	K	CssA	95%
Variability	3%	16%	16%	39%	52%	13%	26%	23%	26%	0%	Wobble-like positions	

Consensus sequences for the recognition of various amino-acid substrates were determined from clusters of signature sequences derived from adenylation domains activating the same substrates. The biosynthetic templates of origin, along with overall similarity of the signature sequences from each template are depicted. Red residues denote variable constituents within a codon. 'Wobble'-like positions, revealing a large variability throughout all codons determined (variabilities >36%), are indicated in cyan. Aad, δ(L-α-aminoadipic acid); Dab, 2,3-diamino butyric acid; Dhb, 2,3-dihydroxy benzoic acid; Sal, salicylate; Phg, L-phenylglycine; hPhg, 4-hydroxy-L-phenylglycine; Pip, L-pipecolinic acid; Dht, dehydrothreonine; '@' indicates a modification of the residue. Adapted from Stachelhaus *et al.*, 1999.

1.5.2 PHYSICO-CHEMICAL PROPERTIES OF THE BINDING POCKET

The predictive method described above is based on the high structural conservation of binding pockets and multiple sequence alignment to GrsA, which can be extracted to form a 'specificity-conferring code'. However a new method has been developed for predicting the specificity of adenylation domains by machine learning using the physico-chemical fingerprint of the residues lining the binding pocket.

Residues are encoded into transductive support vector machines (TSVMs) for machine learning based on the physico-chemical properties of the amino acids and utilize a continuously updated dataset of adenylation domains with known specificity. Physico-chemical properties analysed include the number of hydrogen bond donors, polarity, volume, secondary structure preferences, hydrophobicity and isoelectric point (Rausch *et al.*, 2005). Specificities for very similar substrates that frequently show cross-specificities were pooled to generate the so-called composite specificities and predictive models were built for them. The reliability of the models was confirmed in cross-validations and in comparison with the currently used sequence comparison based method described above. This new method was able to give a specificity prediction in an additional 18% of cases. However a combination of both detection methods creates a powerful new prediction tool that should be used to determine specificity of NRPSs as accurately as possible (Rausch *et al.*, 2005).

1.6 PHENYLALANINE AMMONIA-LYASE (PAL)

This section serves to provide important background information necessary to understand the experimental strategy described in Chapter 2.

In eukaryotic systems, benzoic acid is a common metabolite and is a component of many important natural products such as salicylic acid, cocaine, taxol and the zaragozic acids. The ubiquitous phenylalanine ammonia-lyase (PAL) plant enzymes and chalcone synthases (CHSs) are key biosynthetic catalysts in phenylpropanoid and flavonoid assembly, respectively, and were thought to be largely restricted to plants (Moore *et al*., 2002). In prokaryotic systems, benzoic acid is a rare starter unit for biosynthetic pathways such as soraphen, thiangazole and benzoyl α-L-rhamno-pyranoside synthesis (Hertweck & Moore, 2000). Benzoyl coenzyme A (benzoyl-CoA) is a rare bacterial metabolite that serves as a starter unit for the biosynthesis of the polyketides enterocin and soraphen and as a central intermediate of anaerobic aromatic molecule metabolism. The biosynthesis of benzoyl-CoA involves at least two oxidative pathways from the amino acid phenylalanine (Xiang & Moore, 2003).

'Streptomyces maritimus' produces benzoyl-CoA via cinnamic acid in a plant-like manner from phenylalanine by a PAL-mediated reaction (catalysed by EncP), during the biosynthesis of the polyketide bacteriostatic agent enterocin (Figure 1.6) (Moore *et al*., 2002). The initial biochemical reaction involves the conversion of L-phenylalanine to *trans*-cinnamic acid by a novel bacterial PAL. Activation of cinnamic acid to its CoA thioester is followed by a single round of β-oxidation yielding benzoyl-CoA. In '*S. maritimus*', this primes the enterocin type II polyketide synthase for chain extension with seven molecules of malonyl-CoA (Xiang & Moore, 2005).

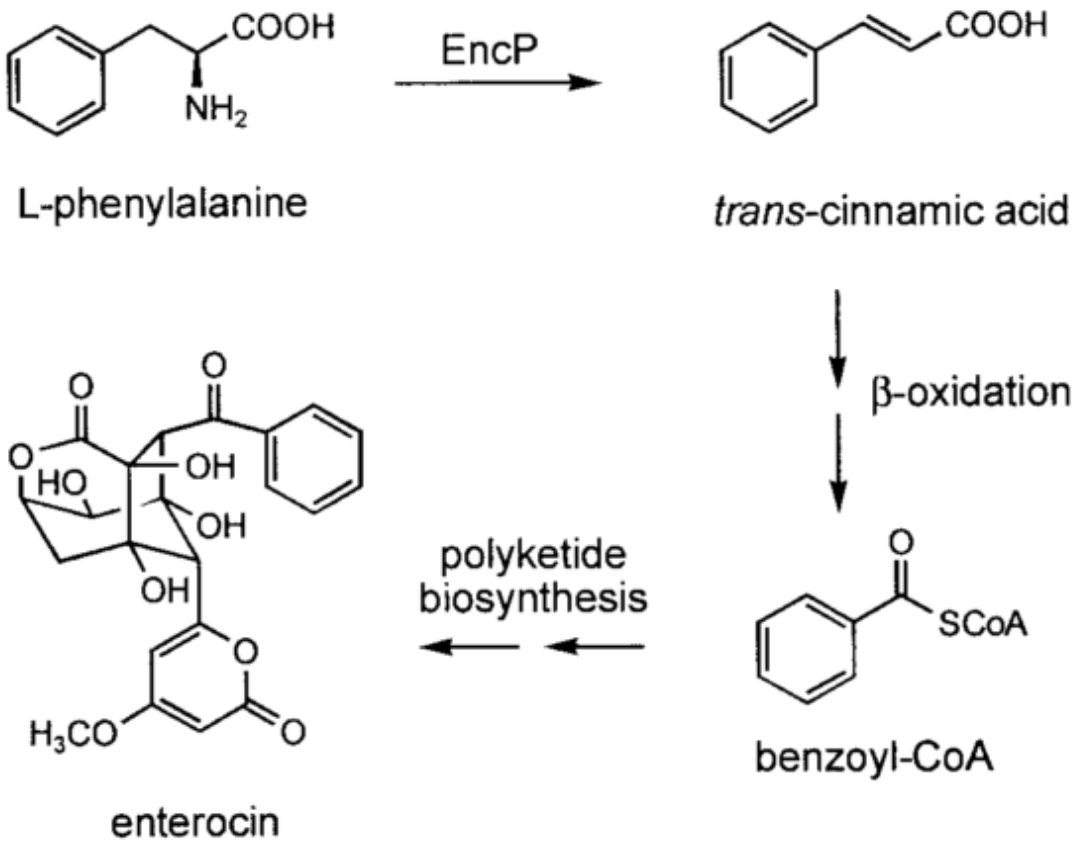

Figure 1.6 EncP-catalyzed conversion of phenylalanine to *trans*-cinnamic acid and biosynthesis of benzoyl-CoA-derived enterocin in '*S. maritimus*'. Taken from Xiang & Moore 2005.

Although PAL is an ubiquitous higher-plant enzyme that functions in an identical manner to bacterial PAL, it has only been encountered in a few other bacteria, where it is involved in benzoyl-CoA biosynthesis such as in *Sorangium cellulosum* and in the biosynthesis of cinnamamide in *Streptomyces verticillatus* (Xiang & Moore, 2005). Besides PAL, the only other well characterised and described bacterial benzoate pathways are anaerobic and involve various starter units (Hertweck & Moore, 2000). The presence of the PAL-encoding gene *encP* has been shown to be absolutely required for benzoyl-CoA formation in '*S. maritimus*'. Thus, *encP* is a prime candidate to screen for when searching for benzoyl-CoA biosynthetic potential (Xiang & Moore, 2003).

1.7 AIM OF THIS STUDY

As part of an antibiotic-screening programme, an actinomycete, *S. polyantibioticus* SPR^T, was isolated from soil collected from the banks of the Umgeni River, KwaZulu-Natal Province, South Africa (Le Roes-Hill & Meyers, 2009). It exhibited antibiosis against various Gram-positive and Gram-negative bacteria, including *Enterococcus faecium* VanA (a vancomycin-resistant strain), *Mycobacterium aurum* A+ (shares similar antibiotic susceptibility to *M. tuberculosis*, but is not pathogenic; Chung *et al.*, 1995) and *Escherichia coli* ATCC 25922, (le Roes, 2005). Further testing showed it also exhibited activity against *M. tuberculosis* H37Rv^T, prompting interest in its antibiotic production.

An antibiotic produced by *S. polyantibioticus* SPR^T was isolated and its structure determined by nuclear magnetic resonance (NMR) and X-ray crystallography to be 2,5-diphenyloxazole (Figure 1.7)(le Roes, 2005). Of great interest, was independent confirmation of the antibiotic activity of DPO and extension of the data to show activity against non-replicating persistent cells of *M. tuberculosis* (Giddens *et al.*, 2005). It seems likely that 2,5-DPO is synthesized non-ribosomally by *S. polyantibioticus* SPR^T. DPO is unusual in that it is a 2,5-disubstituted oxazole, whereas most other disubstituted oxazoles from biological sources are 2,4-substituted.

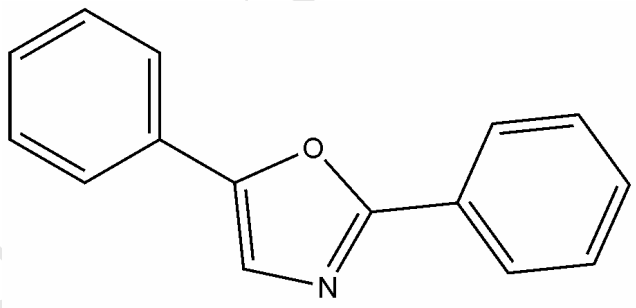

Figure 1.7 The structure of 2,5-diphenyloxazole (DPO)

This study's aims were to characterize and identify the genes involved in the production of DPO within *S. polyantibioticus* SPR^T, to confirm whether DPO is synthesized by an NRPS and furthermore to determine the number of NRPSs within the genome of *S. polyantibioticus* SPR^T. This genetic information from *S. polyantibioticus* SPR^T will allow for combinatorial biosynthesis of the NRPS genes responsible for DPO production to be performed, allowing for the production of derivatives of DPO. This will allow the synthesis of new antibiotics, some of which could become candidates for development as novel drugs to treat drug resistant tuberculosis.

1.8 REFERENCES

Baltz, R.H. (2005). Natural product discovery and development at Eli Lilly and Company: one scientist's view. *SIM News*, **55:** 5-16.

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

Bergendahl, V., Linne, U., & **Marahiel, M.A. (2002).** Mutational analysis of the C-domain in nonribosomal peptide synthesis. *European Journal of Biochemistry,* **269:** 620–629.

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

Bull, A.T., & Stach, J.E.M. (2007). Marine actinobacteria: new opportunities for natural product search and discovery. *Trends in Microbiology*, **15:** 491-499.

Butler, M.S. & Buss, A.D. (2006). Natural products--the future scaffolds for novel antibiotics? *Biochemical Pharmacology*, **71:** 919-929.

Cane, D.E., Walsh, C.T. & Khosla, C. (1998). Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science*, **282:** 63-68.

Cape Gateway, Department of Health (Provincial Government of the Western Cape) (2006), Last accessed November 2010. http://www.capegateway.gov.za/eng/pubs/public_info/M/115549

Centers for Disease Control and Prevention (CDC) (2010). Last accessed November 2010. http://www.cdc.gov/TB/

Challis, G.L. & Naismith, J.H. (2004). Structural aspects of non-ribosomal peptide biosynthesis. *Current Opinion in Structural Biology*, **14:** 748-56.

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

Chung, G. A. C., Aktar, Z., Jackson, S. & Duncan, K. (1995). High-throughput screen for detecting antimycobacterial agents. *Antimicrobial Agents and Chemotherapy*, **39**: 2235-2238.

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

Davey, P.G., Ledingham, J.G.G. & Warrell, D.A. (2000). "Antimicrobial chemotherapy". *Concise Oxford Textbook of Medicine*, 1475.

De Crecy-Lagard, V., Marliere, P. & Saurin, W. (1995). Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation. *Les Comptes Rendus de l'Académie des Sciences III,* **318:** 927-936.

Duerfahrt, T., Eppelmann, K., Müller, R., & Marahiel, M.A. (2004). Rational Design of a Bimodular Model System for the Investigation of Heterocyclization in Nonribosomal Peptide Biosynthesis. *Chemistry & Biology*, **11:** 261–271.

Du, L., Chen, M., Zhang, J., & Shen, B. (2003). BlmIII and BlmIV nonribosomal peptide synthetase-catalyzed biosynthesis of the bleomycin bithiazole moiety both in cis and in trans aminoacylation. *Biochemistry*, **42:** 9731–9740.

Fischbach, M.A. (2009). Antibiotics from microbes: converging to kill. *Current Opinion in Microbiology*, **12:** 520-527.

Fischbach, M.A. & Walsh, C.T. (2010). Antibiotics For Emerging Pathogens. *Science*, **325:** 1089-1093.

Floss, H.G. (2006). Combinatorial biosynthesis--potential and problems. *Journal of Biotechnology*, **124:** 242-357.

Gandhimathi, R., Arunkumar, M., Selvin, J., Thangavelu, T., Sivaramakrishnan, S., Kiran, G., Shanmughapriya, S. & Natarajaseenivasan, K, (2008). Antimicrobial potential of sponge associated marine actinomycetes. *Journal de Mycologie Médicale / Journal of Medical Mycology*, **18:** 16-22.

Gehring, A.M. & Walsh C.J. (1998). Iron acquisition in the plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis. Chemistry & Biology,* **5:** 573-586.

Giddens, A. C., Boshoff, H. I. M., Franzblau, S. G., Barry, C. E. III & Copp, B. R. (2005). Antimycobacterial natural products: synthesis and preliminary biological evaluation of the oxazole-containing alkaloid texaline. *Tetrahedron Letters,* **46:** 7355-7357.

Gulder, T.A.M. & Moore, B.S., (2009). Chasing the treasures of the sea – bacterial marine natural products. *Current Opinion in Microbiology*, **12:** 252-260

Haese, A., Schubert, M., Herrmann, M. & Zocher, R. (1993). Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme catalysing N-methyldepsipeptide formation in *Fusarium scirpi*. *Molecular Microbiology*, **7:** 905-914.

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

Hahn, M. & Stachelhaus, T. (2006). Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *Proceedings of the National Academy of Sciences of the United States of America*, **103:** 275-280.

Hertweck, C. & Moore, B.S. (2000). A Plant-like Biosynthesis of Benzoyl-CoA in the Marine Bacterium `*Streptomyces maritimus*'. *Tetrahedron*, **56:** 9115-9120.

Hoffman, K., Schneider-Scherzer, E., Kleinkauf, H. & Zocher, R. (1994). Purification and characterisation of eucaryotic alanine racemase acting as key enzyme in cyclosporin biosynthesis. *Journal of Biological Chemistry*, **269:** 12710-12714.

Ikai, Y., Oka, H., Hayakawa, J., Harada, K.I., & Suzuki, M. (1992). Structural characterization of bacitracin components by Frit-fast atom bombardment (FAB) liquid chromatography/masspectrometry (LC/MS). *Journal of Antibiotics (Tokyo)*, **45:** 1325–1334.

Keating, T.A., Marshall, C.G., Walsh, C.T. & Keating, A.E. (2002). The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nature Structural Biology*, **9:** 522-6.

Khosla, C., Gokhale, R.S., Jacobsen, J.R. & Cane, D.E. (1999). Tolerance and specificity of polyketide synthases. *Annual Review of Biochemistry*, **68:** 219-253.

Kleinkauf, H. & van Döhren, H. **(1990).** Nonribosomal biosynthesis of peptide antibiotics. *FEBS Journal*, **192:** 1-15.

Koglin, A. & Walsh, C.T. (2009). Structural insights into nonribosomal peptide enzymatic assembly lines. *Natural Product Reports*, **26:** 987-1000.

Konstantinos, A. & Health, Q. (2010). Diagnostic tests, Testing for tuberculosis. *Australian Prescriber*, **33**: 12–18.

Konz, D. & Marahiel, M.A. (1999). How do peptide synthetases generate structural diversity? *Chemistry & Biology*, **4:** R39-R48.

Kopp, F. & Marahiel, M.A. (2007). Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *The Royal Society of Chemistry*, **24:** 735–749.

Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C. & Christopher T. Walsh (1996). A new enzyme superfamily – the phosphopantetheinyl transferases. *Chemistry & Biology*, **3:** 923-936.

Lechevalier, H. A. & Lechevalier, M. P. (1981). Introduction to the order Actinomycetales. In: The Prokaryotes: a handbook on habitats, isolation and identification of bacteria, **2:** 1915-1922. Edited by M. P. Starr, H. Stolp, H. G. Trüper, A. Balows, & H. G. Schlegel. Berlin, Germany: Springer-Verlag.

Lefevre, F., Robe, P., Jarrin, C., Ginolhac, A., Zago, C., Auriol D., Vogel T.M., Simonet P. & Nalin, R. (2008). Drugs from hidden bugs: their discovery via untapped resources. *Research in Microbiology*, **159:** 153-61.

Le Roes, M. (2005). Selective isolation characterization and screening of actinomycetes for novel anti-tubercular antibiotics. PhD Thesis, Department of Molecular and Cell Biology, University of Cape Town. **Chapter 5**, 167-182.

Le Roes-Hill, M. & Meyers, P. R. (2009). *Streptomyces polyantibioticus* sp. nov., isolated from the banks of a river. *International Journal of Systematic and Evolutionary Microbiology*, **59:** 1302–1309.

Marahiel, M.A. (1997). Protein templates for the biosynthesis of peptide antibiotics. *Chemistry & Biology*, **4:** 561-567.

Marahiel, M.A., Stachelhaus, T. & Mootz, H.D. (1997). Modular peptide synthetases involved in non-ribosomal peptide synthesis. *Chemical Reviews*, **97:** 2651−2673

McQuade, T.J. Shallop, A.D., Sheoran, A., Delproposto, J. E., Tsodikov, O.V. & Garneau-Tsodikova, S. (2008). A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. *Analytical Biochemistry*, **386:** 244-250.

Moore, B.S., Hertweck, C., Hopke, J.N., Izumikawa, M., Kalaitzis, J.A., Nilsen, G., O'Hare, T., Piel, J., Shipley, P.R., Xiang, L. & Austin, M.B. and Noel, J.P. (2002). Plant-like biosynthetic pathways in bacteria: from benzoic acid to chalcone. *Journal of Natural Products*, **65:** 1956-1962.

Nobel Prize Foundation, **4 Nov 2010.** "Sir Alexander Fleming - Biography". Nobelprize.org. Last accessed November 2010. http://nobelprize.org/nobel_prizes/medicine/laureates/1945/fleming-bio.html

Otten, L.G., Schaffer, M.L., Villiers, B.R., Stachelhaus, T. & Hollfelder, F. (2007). An optimized ATP/PP_i-exchange assay in 96-well format for screening of adenylation domains for applications in combinatorial biosynthesis. *Biotechnology Journal*, **2:** 232-240.

Qiao, C., Wilson, D.J., Bennett, E.M., Aldrich, C.C. (2007). A mechanism-based aryl carrier protein/thiolation domain affinity probe. *Journal of the American Chemical Society*, **129:** 6350-6351.

Quadri, L.E. Sello, J., Keating, T.A., Weinreb, P.H. & Walsh, C.T. (1998). Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence conferring siderophore mycobactin. *Chemistry & Biology*, **5:** 631-645.

Rausch, C. Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D.H. (2005). Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Research*, **33:** 5799-808.

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.

Reimmann, C., Serino, L., Beyeler, M. & Haas, D. (1998). Dihydroaeruginoic acid synthetase and pyochelin synthetase, products of the pchEF genes, are induced by the extracellular pyochelin in *Pseudomonas aeruginosa*. *Microbiology*, **144:** 3135-3148.

Schneider, T.L., Shen, B., & Walsh, C.T. (2003). Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry*, **42:** 9722–9730.

Smith, D.J., Alison, J.E. & Turner, G. (1990). The multifunctional peptide synthetase performing the first step of penicillin biosynthesis in Penicillium chrysogenum is a 421073 dalton protein similar to Bacillus brevis peptide antibiotic synthetases. *The EMBO Journal*, **9:** 2743-2750.

Stachelhaus, T., Hüser, A. & Marahiel, M.A. (1996). Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chemistry & Biology*, **3:** 913-921.

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

Stachelhaus, T., Schneider, A. & Marahiel, M.A. (1995). Rational design of peptide synthetases antibiotics by targeted replacement of bacterial and fungal domains. *Science*, **269:** 69-72.

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

Tanovic, A. Samel, S.A., Essen, L.O. & Marahiel, M.A. (2008). Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science (New York, N.Y.)*, **321:** 659-63.

Turgay, K., Krause, M. & Marahiel, M.A. (1992). Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Molecular Microbiology*, **6**: 529-546.

Von Döhren, H., Dieckmann, R. & Pavela-Vrancic, M. (1999). The nonribosomal code. *Chemistry & Biology*, **6:** 273-279.

Von Döhren, H., Keller, U., Vater, J. & Zocher, R. (1997). Multifunctional peptide synthetases. *Chemical Reviews*, **97**: 2675-2705.

Waksman, S.A. (1947). What Is an Antibiotic or an Antibiotic Substance? *Mycologia,* **39**: 565-569.

Walsh, C.T. (2004). Polyketide and Nonribosomal Peptide Antibiotics: Modularity and Versatility. *Science*, **303:** 1805-1810.

Walsh, C.T. Chen, H., Keating, T.A., Hubbard, B.K., Losey, H.C., Luo, L., Marshall, C.G., Miller, D.A., & Patel, H.M. (2001). Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Current Opinion in Chemical Biology*, **5:** 525-534.

Walsh, C.T. Gehring, A.M., Weinreb, P.H., Quadri, L.E. & Flugel, R.S. (1997). Post-translational modification of polyketide and nonribosomal peptide synthases. *Current Opinion in Chemical Biology*, **1:** 309-315.

Weber, G., Schörgendorfer, K., Schneider-Scherzer, E, & Leitner, E. (1994). The peptide synthetase catalyzing cyclosporin production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. *Current Genetics*, **26:** 120-125.

World Health Organisation (WHO), Tuberculosis Fact Sheet (2010). Last accessed November 2010. http://www.who.int/mediacentre/factsheets/fs104/en/index.html

World Health Organisation (WHO) Report 2009, Global Tuberculosis Control.

Wu, D., Hugenholtz. P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., Kunin, V., Goodwin, L., Wu, M., Tindall, B. J., Hooper, S. D., Pati, A., Lykidis, A., Spring, S., Anderson, I. J., D'haeseleer, P., Zemla, A., Singer, M., Lapidus, A., Nolan, M., Copeland, A., Han, C., Chen, F., Cheng, J. F., Lucas, S., Kerfeld, C., Lang, E., Gronow, S., Chain, P., Bruce, D., Rubin, E. M., Kyrpides, N. C., Klenk, H. P. & Eisen, J. A. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature*, **462:** 1056-1060.

Xiang, L. & Moore, B.S. (2003). Characterization of Benzoyl Coenzyme A Biosynthesis Genes in the Enterocin-Producing Bacterium "*Streptomyces maritimus*". *Journal of Bacteriology*, **185**: 399-404.

Xiang, L. & Moore, B.S., (2005). Biochemical Characterization of a Prokaryotic Phenylalanine Ammonia Lyase. *Journal of Bacteriology,* **187:** 4286-4289.

Yanai, K., Murakami, T. & Bibb, M. (2006). Amplification of the entire kanamycin biosynthetic gene cluster during empirical strain improvement of Streptomyces kanamyceticus. *Proceedings of the National Academy of Sciences of the United States of America*, **103:** 9661-9666.

Zhang, W., Li, Z., Miao, X. & Zhang, F. (2009). The screening of antimicrobial bacteria with diverse novel nonribosomal peptide synthetase (NRPS) genes from South China sea sponges. *Marine Biotechnology (New York, N.Y.)*, **11:** 346-55.

CHAPTER 2

THE INVESTIGATION INTO THE SYNTHESIS OF 2,5- DIPHENYLOXAZOLE

CONTENTS

CHAPTER 2

THE INVESTIGATION INTO THE SYNTHESIS OF 2,5-DIPHENYLOXAZOLE

2.1 SUMMARY

Three unique adenylation domains were isolated from *S. polyantibioticus* SPR^T and were designated DS-D, PJ15 and PJ73. It was shown by the NRPSpredictor program that DS-D is specific for glycine, alanine or valine; PJ15 for ornithine, lysine or arginine; PJ73 for serine, threonine or dehydrothreonine. These adenylation domains were used as probes in Southern hybridization experiments to identify larger DNA fragments from their respective NRPS genes. 28 adenylation domain fragments and flanking sequences were detected via Southern hybridization, but it proved challenging to clone these fragments. PCR primers were designed to amplify the *encP* gene (for phenylalanine ammonia-lyase) from '*Streptomyces maritimus*' DSM 41777^T and it was shown by PCR and Southern hybridization that *encP* could be detected in '*S. maritimus*'. However, an *encP*-homologue could not be detected within the genome of *S. polyantibioticus* SPR^T by these methods. Cyclization-domain PCR primers were designed based on sequences from characterised *Streptomyces* and *Streptoalloteichus* thiazole and oxazole producers. No cyclization domains could be amplified from *S. polyantibioticus* SPR^T or the control strains.

2.2 INTRODUCTION

2.2.1 BIOSYNTHESIS OF DPO

DPO is known as a scintillator, which is a material which exhibits a flash of light or luminescence when excited by ionizing radiation (Leo, 1994). DPO is currently only known to be synthesised chemically (Adrova *et al*., 1956), therefore its discovery from a biological origin is of great interest.

A suggested biochemical pathway for the formation of DPO is shown in Figure 2.1. It is proposed that DPO is synthesised from the starting units of benzoic acid (Figure 2.1A) and β-hydroxyphenylalanine or phenylalanine (Figure 2.1B). A peptide bond is formed between the carboxyl group of benzoic acid and the amino group of β-hydroxyphenylalanine resulting in the formation of benzoyl-β-hydroxyphenylalanine (Figure 2.1C).

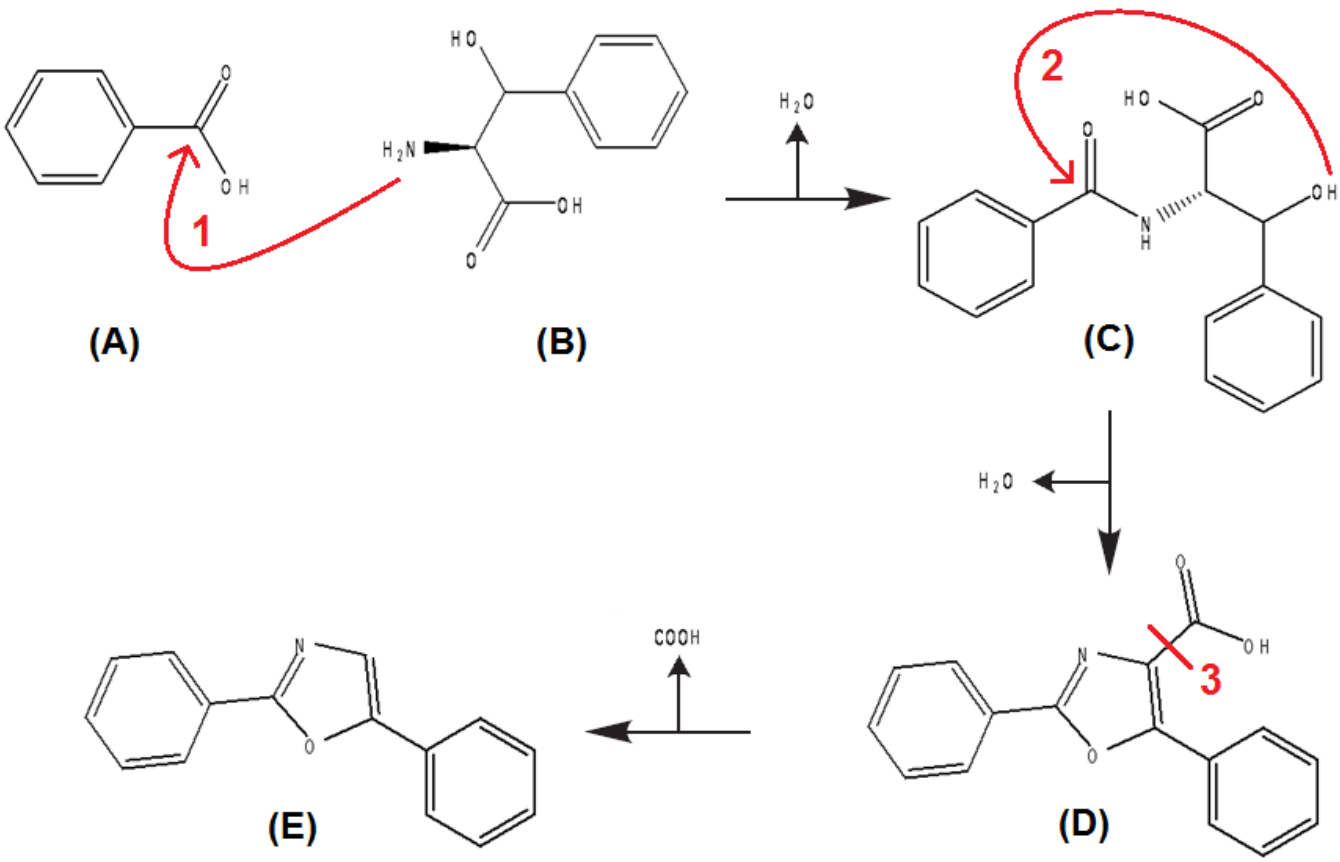

Figure 2.1 Proposed reaction scheme for the synthesis of DPO in *S. polyantibioticus* SPR^T. **(A)** Benzoic Acid, **(B)** β-Hydroxyphenylalanine, **(C)** Benzoyl-β-Hydroxyphenylalanine, **(D)** 4-Carboxy 2,5-Diphenyloxazole, **(E)** DPO. Reactions are shown by red arrows or lines, **1**- peptide bond formation, **2**- Cyclization, **3**- Decarboxylation. An NRPS is proposed to catalyse the condensation of **(A)** and **(B)**, as well as the proposed heterocyclization of **(C)** to form **(D)**.

Following the formation of benzoyl-β-hydroxyphenylalanine, the molecule is proposed to undergo a cyclization or cyclodehydration reaction to form an oxazoline intermediate and then oxidised to form the oxazole. This involves a nucleophilic attack by the phenylalanine β-hydroxy group on the carbonyl group of the amide. This cyclization results in the formation of 4-carboxy 2,5-diphenyloxazole (Figure 2.1D). The decarboxylation of the oxazole group results in the formation of DPO as the final product (Figure 2.1E).

Due to the condensation between benzoic acid and β-hydroxyphenylalanine and the heterocyclization across the peptide bond to form 4-carboxy 2,5-diphenyloxazole, it is highly probable that this reaction cascade would be catalysed by an NRPS. The NRPS responsible for the biosynthesis of DPO is proposed to have ArCP, Cyclization, Adenylation, Oxidation, PCP and Thioesterase domains (Figure 2.2).

The ArCP domain acts as a docking site for the benzoyl intermediate and the PCP domain for the β-hydroxyphenylalanyl intermediate. The cyclization domain would be involved in the condensation of benzoic acid and β-hydroxyphenylalanine to form benzoyl-β-hydroxyphenylalanine and would also play a role in the cyclization of benzoyl-β-hydroxyphenylalanine to 4-carboxy 2,5-diphenyloxazole. The thioesterase domain would be involved in the termination of peptide elongation by disconnection of the covalent linkage between DPO and the 4'-phosphopantetheine (4'-PP) thiol arm. The oxidation domain would play a role in the oxidation of the oxazoline to form the oxazole in the cyclization reaction. It need not be part of the NRPS and may act as an external tailoring enzyme acting *in trans*. A decarboxylase would be involved in the decarboxylation of 4-carboxy 2,5-diphenyloxazole and could also be an external enzyme acting *in trans*, similar to curacin A biosynthesis in *L. majuscula* (Gu *et al*., 2006). The adenylation domain would function in the recognition of β-hydroxyphenylalanine or phenylalanine as a substrate and pass it to the PCP domain. It is not clear whether the adenylation domain would bind phenylalanine (which would subsequently be modified to β-hydroxyphenylalanine before the cyclization reaction) or whether it would bind β-hydroxyphenylalanine directly. If phenylalanine is the substrate, the DPO NRPS would be expected to have an additional domain for a P450 monooxygenase to allow for the β-hydroxylation reaction, although this activity could also be provided *in trans*.

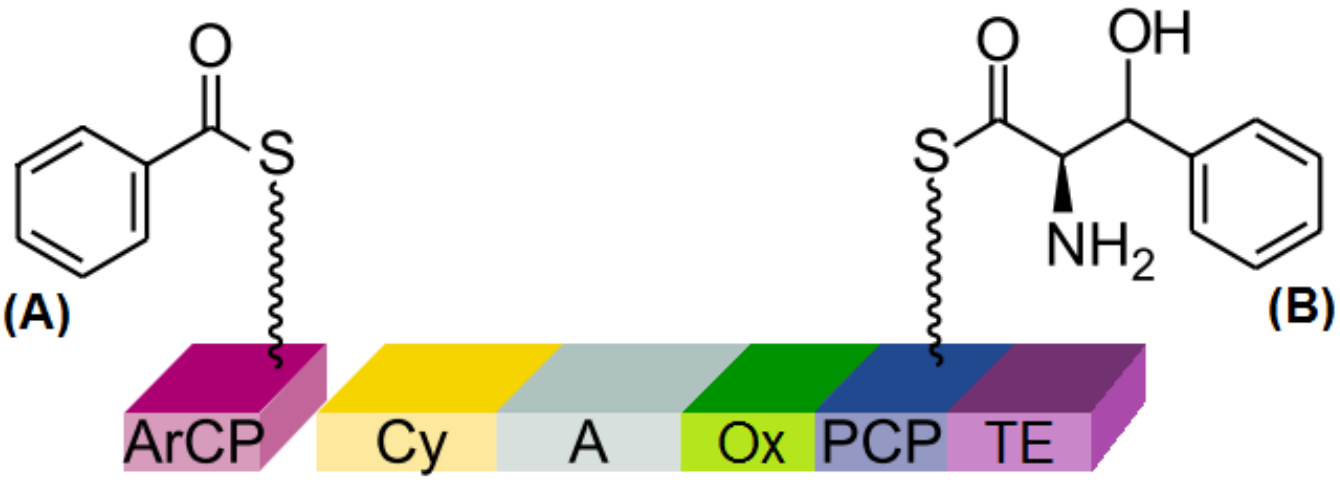

Figure 2.2 Proposed module arrangement for a DPO producing NRPS. Benzoate and β-hydroxyphenylalanine bind to the serine residue belonging to the 4'-phosphopantetheine (4'-PP) arm of their respective carrier protein domains to form an activated thioester derivative. Benzoate binds to the 4'-PP of ArCP forming a benzoyl intermediate **(A)** and β-hydroxyphenylalanine binds to the 4'-PP of PCP forming a β-hydroxyphenylalanyl intermediate **(B)**. Domains: **ArCP** - aryl carrier protein, **Cy** – cyclization domain, **A** – adenylation domain, **Ox** – oxidation domain, **PCP** - peptidyl carrier protein, **TE** – thioesterase domain.

2.2.2 GENE CLUSTER ISOLATION

Identifying an adenylation domain within the genome of *S. polyantibioticus* SPR^T with a binding pocket substrate specificity for phenylalanine or β-hydroxyphenylalanine would be expected to indicate an NRPS responsible for the synthesis of DPO. The location of this NRPS within the genome may then be discovered by Southern blotting, which will aid in the isolation of the rest of the gene coding for the NRPS, as well as the other genes for the biosynthesis of DPO.

The adenylation domains of the many possible NRPSs within *S. polyantibioticus* SPR^T can be amplified using the Polymerase Chain Reaction (PCR) and suitable primers. The amplified adenylation domains may then be interrogated by cloning, sequencing and identifying their amino-acid binding specificities by comparison with the binding-pocket specificities of adenylation domains for which the amino-acid substrates are known. Larger portions of the gene cluster may then be detected by Southern hybridization. Detected larger portions of the NRPS gene may then be further cloned and sequenced, gathering vital genetic information. Cyclization domains may be analysed in a similar fashion. A cyclization domain present in an NRPS is an indicator of cyclization reactions in non-ribosomal peptides. PCR primers specific for oxazole and thiazole producing cyclization domains may be used to amplify such domains in the genome.

In addition to searching for NRPS genes, the presence of PAL within the genome of *Streptomyces polyantibioticus* SPR^T would be of great interest, not only for its rarity in bacterial systems, but also because *S. polyantibioticus* SPR^T is proposed to utilize benzoic acid produced in this way in the synthesis of 2,5-diphenyloxazole (DPO). Such a PAL could be amplified from the genome of *S. polyantibioticus* SPR^T using suitable PCR primers. Bands detected on a Southern hybridization could also be an indication of the presence of PAL or a similar enzyme within the genome.

2.3 MATERIALS AND METHODS

2.3.1 STRAINS, MEDIA AND GROWTH CONDITIONS

'*S. maritimus*' DSM 41777^T, *Streptomyces virginiae* NRRL B-1446^T and *Streptomyces mobaraensis* DSM 40903 were grown in yeast extract-malt extract broth (YEME) (Shirling & Gottlieb, 1966) and *S. polyantibioticus* SPR^T was grown in *Streptosporangium* Medium (SM) (Pfefferle *et al.*, 2000). '*S. maritimus*', *S. virginiae* and *S. polyantibioticus* SPR^T were grown at 30^{o}C for 7 days with shaking and *S. mobaraensis* DSM 40903 at 30^{o}C for 3 days with shaking. Gram stains and single-colony streaking onto YEME or SM plates were performed to check for contamination.

2.3.2 GENOMIC DNA EXTRACTION

Total genomic DNA was extracted from bacterial cell mass using the genomic DNA extraction method of Wang *et al.* (1996), with the following modifications: 25mg lysozyme/ml was used instead of 5mg/ml and the cells were incubated in the lysozyme buffer at 37^{o}C overnight instead of for 30min; isoamyl alcohol was omitted from the chloroform extraction step and the final precipitation of DNA was performed with the addition of one tenth of a volume of 3M sodium acetate (pH 5.2) and one volume of room-temperature isopropanol. The concentration of DNA was measured spectrophotometrically using a Nanodrop® ND-1000 spectrophotometer. To confirm that the extracted genomic DNA was of good quality, the 16S rRNA gene was PCR amplified as detailed in Cook & Meyers (2003).

2.3.3 DESIGN OF OLIGONUCLEOTIDE PRIMERS

All primers used in this study are listed in Table 2.1, accompanied by their binding positions and primer sequences. The amplification of a 700bp region of the adenylation domain of an NRPS gene cluster was achieved using the A3F forward primer and A7R reverse primer (Ayuso-Sacido & Genilloud, 2005?). These primers were designed to bind to the conserved motifs of actinomycete adenylation domains, named A3 and A7, respectively. Additionally, these primers are degenerate to suit sequences containing a high G+C content, common to actinomycete genomes. pJET1.2 forward and reverse primers were used to amplify inserts cloned into the pJET1.2/blunt vector. pUC/M13 forward and reverse primers were used to amplify inserts cloned into pGEM®-T Easy and pUC18 vectors.

TABLE 2.1. Oligonucleotide primers used in this study

Primer Name	Binding Position	Primer Sequence (5'→3')
A3F	1120-1142 [a]	GCSTACSYSATSTACACSTCSGG
A7R	1820-1801 [a]	SASGTCVCCSGTSCGGTAS
Cy-F	35-60 [b]	AGCCITTCYCSCTSACSSMBSTSCAG
Cy-R	1077-1054 [b]	GICSAGSWISSWSGTGAASACSAC
Smob Cy-R	1077-1054 [c]	AGGCAGGTCGGAGGTGAAGACGAC
EncP-F	766-787	GACTCGCACCTGGCGGTCAAC
EncP-R	1486-1465	GTAGTCGGTGATGGTCTCGTC
pJET1.2-F	310-332	CGACTCACTATAGGGAGAGCGGC
pJET1.2-R	428-405	AAGAACATCGATTTTCCATGGCAG
pUC/M13-F	2949-2972	CGCCAGGGTTTTCCCAGTCACGAC
pUC/M13-R	197-176	TCACACAGGAAACAGCTATGAC

[a] Based on the *Brevibacillus brevis GrsA1* sequence (accession number: D00519).

[b] Based on the *S. virginiae virH* sequence (Table 2.2).

[c] Based on the '*S. verticillus*' ATCC 15003 *blmIV* sequence (Table 2.2).

Primers use standard ambiguity codes for nucleotides; I = inosine.

'F' denotes forward primer and 'R' denotes reverse primer.

Cy-F forward and Cy-R reverse primers were designed to amplify an approximately 1050bp fragment from Cyclization domains in NRPS gene clusters. A multiple amino-acid sequence alignment of the cyclization-domain regions of five oxazole and thiazole producing NRPSs from various bacterial strains (Table 2.2) was performed and Cy-F and Cy-R were designed based on conserved regions. Smob Cy-R reverse primer was based on the *blmIV* gene sequence (bleomycin) only and binds at exactly the same nucleotide positions as Cy-R. EncP-F forward and EncP-R reverse primers were designed to amplify *encP* from the biosynthetic gene cluster for benzoyl-CoA-derived enterocin in '*S. maritimus*' (AF254925) and produce a 721bp fragment.

TABLE 2.2. Bacterial species and genes used to design cyclization primers

Species name	Accession Number	Gene	Aromatic Compound
Streptoalloteichus hindustanensis ATCC 31158	EF032505	*tlmIV*	Tallysomycin
Streptomyces atroolivaceus (no strain number given)	AF48456	*lnmI*	Leinamycin
Streptomyces flavoviridis ATCC 21892	EU670723	*zbmIV*	Zorbamycin
'*Streptomyces verticillus*' ATCC 15003	AF210249	*blmIV*	Bleomycin
Streptomyces virginiae	AB283030	*virH*	Virginiamycin M_1

2.3.4 POLYMERASE CHAIN REACTION (PCR) PROTOCOLS

2.3.4.1 *encP* AMPLIFICATION

The *encP* amplification protocol consisted of the following: initial denaturation at 96 °C for 2 minutes, followed by 30 cycles of denaturation at 96 °C for 45 seconds, annealing at 60 °C for 30 seconds and elongation at 72 °C for 70 seconds, with a final elongation at 72 °C for 5 minutes. PCR reactions contained: 50-200ng of DNA, 1U Supertherm *Taq* polymerase (JMR Holdings, USA), 0.5µM of each primer, 200µM of each dNTP and 4mM $MgCl_2$, in a total volume of 50µl.

2.3.4.2 ADENYLATION DOMAIN AMPLIFICATION

The PCR protocol for the amplification of the adenylation domains of the NRPS regions was as detailed in Ayuso-Sacido & Genilloud (2005): initial denaturation at 95 °C for 5 minutes, followed by 35 cycles of denaturation at 95 °C for 30 seconds and elongation at 72 °C for 4 minutes, with a final elongation at 72 °C for 10 minutes. An annealing temperature of 60 °C for 2 minutes was used instead of 56 °C. PCR reactions contained: 50-200ng of DNA, 2.6U Expand High Fidelity *Taq* polymerase as part of the Expand High Fidelity PCR System (Roche), 2.5µM of each primer, 200µM of each dNTP, 4mM $MgCl_2$, and 3% glycerol in a total volume of 50µl.

2.3.4.3 CYCLIZATION DOMAIN AMPLIFICATION

The PCR protocol for the amplification of the cyclization domains was identical to the PCR protocol for the amplification of the adenylation domains (section 2.3.4.2), except that an annealing temperature of 56 °C for 2 minutes was used instead of 60 °C. The concentrations of the PCR reactants were also identical, except that 0.5µM of each primer was used instead of 2.5µM.

2.3.4.4 COLONY PCR

A colony PCR protocol was used to amplify adenylation domain fragments and Southern-hybridization identified fragments from recombinant plasmids cloned into transformants. The PCR protocol for the amplification of these inserts was identical to the protocol used for the amplification of the adenylation domains. PCR reactions contained: toothpick-tip size amount of cell mass from a transformant colony, 2.5U KAPA2G Robust DNA Polymerase as part of the KAPA2G™ Robust PCR Kit (KAPABIOSYSTEMS, Cape Town, South Africa), 0.5µM of each primer, 200µM of each dNTP and 1.5mM $MgCl_2$ in a total volume of 20µl. Primers used in each reaction were dependent on the vector used (Table 2.1).

All PCR products, including probes for Southern hybridisations, were electrophoresed alongside a λ-*Pst*I molecular marker on 0.8% agarose gels, containing 0.8µg/ml ethidium bromide, and visualized on a GelDoc System (BioRad). The amplified products were purified using the MSB® Spin PCRapace kit (Invitek, Germany).

2.3.5 CLONING

The 700bp PCR-amplified segments of the adenylation domains were ligated into pGEM-T Vectors as part of the pGEM®-T and pGEM®-T Easy Vector System kit (Promega) and pJET1.2/blunt suicide vectors as part of GeneJET™ PCR Cloning Kit, (Fermentas Life Sciences). Competent *Escherichia coli* DH5α and HB101 cells, were prepared by the $CaCl_2$ shock treatment method (Dagert & Ehrlich, 1979) and were used for the transformations, performed as described by Sambrook *et al.* (1989). Transformants were plated onto Luria-Bertani (LB) agar (Sambrook *et al.*, 1989) containing 100µg/ml ampicillin and incubated at 37°C for 24hrs. Transformants were subcultured onto fresh plates to allow confirmation of transformation. The presence of the desired inserts was determined by performing colony PCR using the A3F forward and A7R reverse primers for DNA fragments cloned into the pGEM®-T Easy vector and using pJET1.2 forward and reverse sequencing primers for DNA fragments cloned into pJET1.2/blunt. The PCR protocol used was that described in section 2.3.4.4.

Digested *S. polyantibioticus* SPR^T genomic DNA was excised from gels and purified using the Favorgen GEL/PCR Purification Mini Kit (Favorgen™, Germany) and cloned into blue/white selection vectors pUC18 (Yanisch-Perron *et al*., 1985) and pGEM-T Easy and also into suicide vectors pJET1.2/blunt and pEcoR251 (Zappe *et al*., 1986). Ligation was performed using the Sticky-End Protocol for pJET1.2/blunt in accordance with the kit protocol. pUC18 was digested with *Kpn*I and pEcoR251 was digested separately with *Bgl*II and *Pst*I. Digestions were performed overnight at 37°C with 1.5U of each restriction endonuclease along with the appropriate restriction buffers and 1µg plasmid DNA. pUC18 and pEcoR251 were dephosphorylated before ligation using rAPid Alkaline Phosphatase as part of the rAPid Alkaline Phosphatase kit (Roche), but with an overnight incubation at 37°C. Plasmids were blunted where necessary using the Klenow fragment or a blunting enzyme included in the GeneJET™ PCR Cloning Kit (Fermentas).

Transformation was carried out as described in this section (2.3.5). pJET1.2/blunt and pEcoR251 suicide vector transformants were obtained by positive selection, meaning that all transformants must contain a plasmid with an insert. The transformants were plated onto LB agar containing 100µg/ml ampicillin and incubated at 37°C for 24hrs. pGEM®-T Easy Vector and pUC18 transformants were plated onto LB agar containing 100µg/ml ampicillin, 0.5mM IPTG and 80µg/ml X-Gal and incubated at 37°C for 24hrs after which blue/white selection was employed. Transformants were subcultured onto fresh LB/Amp/IPTG/X-Gal plates to allow confirmation of transformation. pUC18 transformants were screened by performing colony PCR using the M13 forward and reverse primers (Table 2.1 and section 2.3.4.4).

2.3.6 DNA SEQUENCING AND SEQUENCE ANALYSIS

All sequencing was carried out using dye termination reactions on an Applied Biosystems Big Dye terminator v3.1 DNA sequencer using BIOLINE Half Dye Mix. All plasmid samples in this study were prepared for sequencing using the Invisorb Spin Plasmid Mini Two kit (Invitek, Germany) and peqGOLD Plasmid Miniprep KitI (peqlab Biotechnologie GmbH, Germany). Inserts cloned into pGEM-T Easy and pJET1.2/blunt were sequenced with the A3F forward and A7R reverse sequencing primer pair and pJET1.2 forward and reverse sequencing primer pair, respectively. All sequence chromatograms were viewed and edited using Chromas v2.01, (Technelysium, Australia). DNA alignments, translations, *in silico* digestions and other analyses were performed using DNAMAN v5.2.9, (Lynnon Biosoft, U.S.A.).

The Basic Local Alignment Search Tool (BLAST) was used to compare the sequence similarity of isolated DNA sequences (and their translated protein sequences) to known published sequences in the NCBI database (Altschul *et al.*, 1997). The NRPSpredictor program (Rausch *et al.*, 2005) available online (http://www-ab.informatik.uni-tuebingen.de/software/NRPSpredictor), was used to determine the signature sequences of the binding pockets of the cloned adenylation domains. The nucleotide sequence of each adenylation domain was translated and provided to the program, which identified the key binding-pocket residues by an alignment with GrsA and determined the probable specificity of the adenylation domain based on the biochemical and biophysical properties of the binding-pocket residues.

2.3.7 SOUTHERN HYBRIDIZATION

2.3.7.1 RESTRICTION ENDONUCLEASE DIGESTION

For Southern hybridizations using adenylation domain probes, total genomic DNA of *S. polyantibioticus* SPR^T was digested with the following pairwise combinations of restriction endonucleases: *Sph*I & *Stu*I, *Pst*I & *Pvu*I, *Xho*I & *Pvu*I, *Pvu*II & *Sph*I, *Pst*I & *Kpn*I and *Avr*II & *Stu*I.

For the Southern hybridization using the *encP* probe, DNA of *S. polyantibioticus* SPR^T was digested with the following pairwise combinations of restriction endonucleases: *Sph*I & *Stu*I, *Pvu*II & *Sph*I, *Avr*II & *Sph*I and *Avr*II & *Stu*I. '*S. maritimus*' genomic DNA was digested with a pairwise combination of *Sph*I & *Stu*I.

Reaction volumes varied from 20µl to 50µl and contained approximately 20µg of DNA, 1.5U of each restriction endonuclease along with the appropriate restriction buffer. Digestions were performed overnight at 37^{o}C.

2.3.7.2 PROBE PREPARATION

The 700bp NRPS adenylation-domain fragments amplified from *S. polyantibioticus* SPR^T used as Southern-hybridization probes were cloned either into pGEM-T Easy or into pJET1.2/blunt. Clone DS-D is in pGEM-T Easy and clones PJ15 and PJ73 are in pJET1.2/blunt. These transformed vectors served as templates for probe preparation used for Southern hybridizations. Additionally, PCR-amplified *encP* from '*S. maritimus*' served as a template for probe preparation and was also used in a Southern hybridization.

These templates were used to create probes by PCR using the PCR DIG Probe Synthesis Kit (Roche). Adenylation domain probes were synthesised using the adenylation domain PCR protocol and plasmid specific primers (section 2.3.4.2). The amplified *encP* gene probe was synthesized using the above mentioned kit, but using the *encP* PCR protocol (section 2.3.4.1) and EncP-F forward primer and EncP-R reverse primer.

2.3.7.3 SOUTHERN BLOT HYBRIDIZATION

Restriction endonuclease digested *S. polyantibioticus* SPR^T and '*S. maritimus*' genomic DNA was electrophoresed on 0.8% agarose gels (0.7% agarose for PJ15 and PJ73 Southern hybridizations), containing 0.8µg/ml ethidium bromide and blotted onto Hybond N$^+$ membranes (Amersham). The *encP* and DS-D probe Southern hybridizations were blotted using capillary transfer with 0.4M NaOH/1M NaCl, modified from Reed & Mann (1985). The PJ15 and PJ73 probes for Southern hybridizations were blotted using a Trans-blot® SD Semi-dry Transfer Cell (BIO-RAD according to the accompanying protocol.

Each membrane was then air dried and placed into a sealable plastic bag with DIG Easy Hybridization Buffer (Roche) (approximately 10ml/100cm^2 of membrane). This was followed by incubation with gentle shaking for 2 hours at 55.2^{o}C for the DS-D probe, 53.8^{o}C for the *encP* probe and 48 ^{o}C for probes PJ15 and PJ73, which are the hybridization temperatures used for each probe. Hybridization for probes PJ15 and PJ73 was performed at a temperature 5^{o}C lower than the calculated hybridization temperature to reduce stringency. The hybridization solution was prepared by adding 40µl denatured probe to 13.3ml Hybridization buffer (3µl/ml). Hybridization was performed at the hybridization temperature overnight with gentle shaking. The temperature of hybridization was calculated by the following formula:

$T_{hyb} = T_M - (20^oC \text{ to } 25^oC)$

Where, $T_M = 49.82 + 0.41(\%GC) - 600/\ell$

T_M – melting point of probe-target hybrid
%GC – percentage G+C residues in the probe sequence
T_{hyb} – optimal temp for hybridization of the probe to target in DIG Easy Hybridization
ℓ – length of the probe/hybrid in base pair

The membrane was then washed twice in low stringency buffer, containing 2x standard saline citrate (SSC) and 0.1% sodium dodecyl sulfate (SDS), at room temperature for 5 minutes with shaking, followed by washing twice in high stringency buffer (0.1x SSC; 0.1% SDS) at 68^{o}C for 15min with gentle shaking. This was followed by washing with washing buffer (1x maleic acid pH 7.5; 0.3% Tween 20) and blocking with 2% skim milk.

Following the blocking step, the membrane was incubated with the anti-DIG Alkaline Phosphatase conjugate (Roche) at room temperature with shaking for 30 minutes. The membrane was then washed twice with washing buffer for 15 minutes at room temperature with shaking to remove excess antibody.

Before detection, the membrane was equilibrated for 3 min in Detection Buffer (0.1M Tris-Cl, 0.1M NaCl pH 9.5) at room temperature. Detection was achieved by incubating the membrane in 20ml Detection buffer containing 0.175mg/ml 5-bromo-4-chloro-indolyl- phosphate (BCIP) and 0.25mg/ml 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) in the dark for 0.5-16 hours. The reaction was stopped by rinsing the membrane with TE buffer (10mM Tris-Cl, 1mM EDTA, pH 7.6) for 5 minutes and then the result was digitally captured.

2.4 RESULTS AND DISCUSSION

2.4.1 SCREENING FOR *encP*

PCR amplification was performed using the designed *encP* primers (Table 2.1) on genomic DNA extracted from *S. polyantibioticus* SPR^T and '*S. maritimus*'. 16S rRNA gene PCR amplification was performed in parallel to confirm the quality of the extracted DNA and to ensure amplification was possible. Gel electrophoresis was used to quantify the results of the PCR (Figure 2.3).

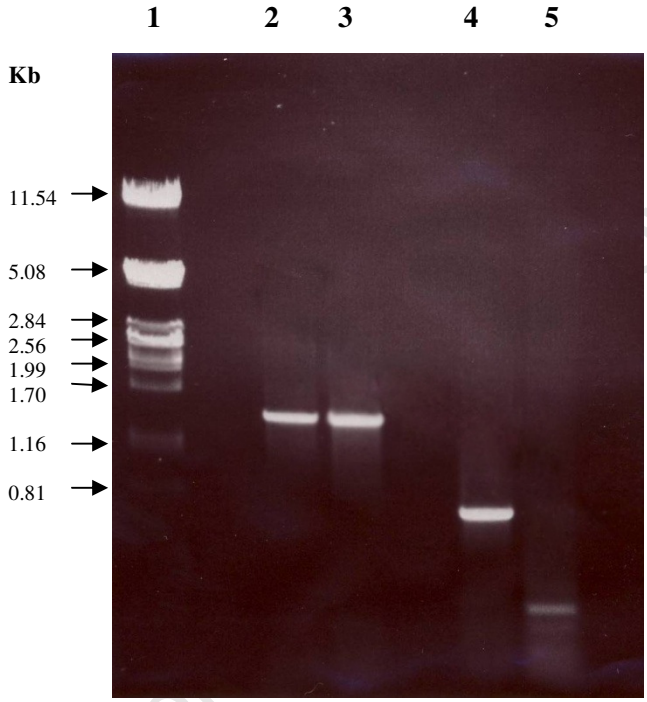

Figure 2.3 Gel Electrophoresis of PCR amplified *encP* and16S rDNA from *S. polyantibioticus* SPR^T and '*S. maritimus*'. Lanes: **1**- λ-*Pst*I molecular marker, **2**- '*S. maritimus*' amplified 16S rDNA, **3**- *S. polyantibioticus* SPR^T amplified 16S rDNA, **4**- '*S. maritimus*' amplified *encP*, **5**- *S. polyantibioticus* SPR^T with no amplified *encP*.

Gel electrophoresis of the PCR amplified genomic DNA showed amplification for the 16S rRNA gene controls for both strains resulting in a single 1.5kb band for each. PCR amplification using the *encP* primers only showed amplification for '*S. maritimus*' genomic DNA resulting in a clear, single band of about 0.7kb. There was no similar band observed for *encP* amplification from the *S. polyantibioticus* SPR^T genomic DNA, but some smaller bands were seen which could be the result of primer dimers. *EncP* single-primer amplifications of *S. polyantibioticus* SPR^T and '*S. maritimus*' genomic DNA resulted in no bands, confirming the specificity of these primers (data not shown).

This result suggested that there is no *encP* or similar gene in *S. polyantibioticus* SPR^T. However, it is possible that there was no amplification because the *S. polyantibioticus* SPR^T homologue of the gene does not have the specific sequences complementary to the primers used (the primer design was based on the single bacterial PAL gene sequence available). Following this experiment, Southern hybridization was used as an alternative way of detecting *encP* in *S. polyantibioticus* SPR^T.

A Southern hybridization experiment was performed using various pairwise restriction endonuclease digestions of *S. polyantibioticus* SPR^T genomic DNA along with a single pairwise restriction endonuclease digestion of '*S. maritimus*' genomic DNA (Figure 2.4). In addition, amplified *encP* from '*S. maritimus*' and unlabelled probe DNA were used as positive controls.

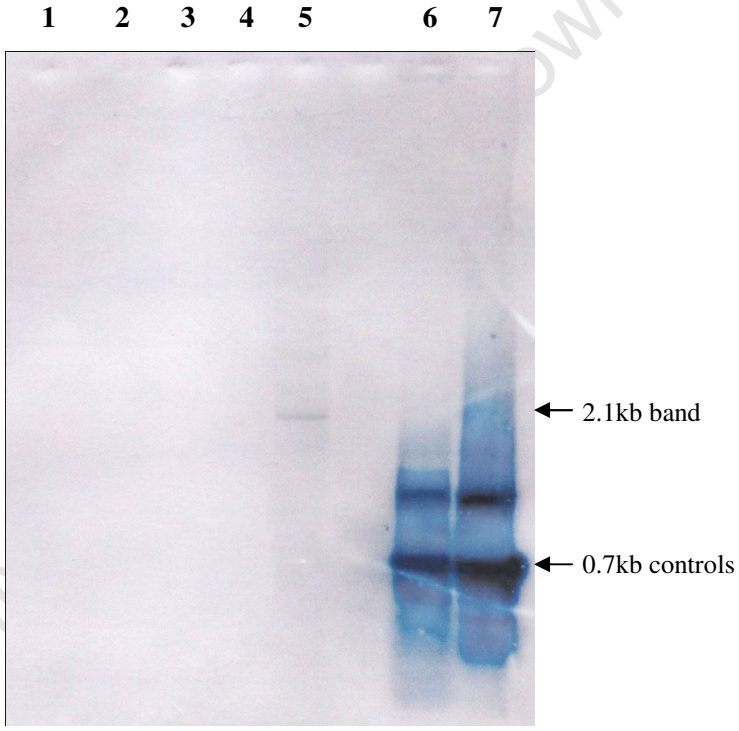

Figure 2.4 Southern hybridization of restriction endonuclease digested *S. polyantibioticus* SPR^T genomic DNA using the '*S. maritimus*' *encP* probe. Lanes 1-4 are pairwise endonuclease digestions of 20 µg *S. polyantibioticus* SPR^T genomic DNA with the following combinations: **1**-*Sph*I *& Stu*I, **2**- *Pvu*II *& Sph*I, **3**- *Avr*II *& Stu*I, **4**- *Avr*II *& Sph*I. 20 µg '*S. maritimus*' genomic DNA is in Lane **5** (digested with *Sph*I *& Stu*I). Lanes **6** and **7** are the amplified *encP* control and an unlabelled probe control, respectively (positive controls).

There was no hybridization observed for any of the *S. polyantibioticus* SPR^T genomic digests (Figure 2.4, lanes 1-4). A weak hybridization band of 2.1kb was observed for the '*S. maritimus*' genomic digest (Figure 2.4, lane 5). This result showed that *encP* can be detected in '*S. maritimus*' genomic DNA via Southern hybridization when using *encP* as a probe. Both positive controls worked (Figure 2.4, lanes 6 and 7) This result agrees with that for the PCR amplification experiment (Figure 2.3) and suggests that there is no *encP* or similar gene in *S. polyantibioticus* SPR^T.

The evidence presented suggests that there is no PAL-like gene in *S. polyantibioticus* SPR^T . There may, however, be a novel PAL-like enzymatic process in *S. polyantibioticus* SPR^T, which is able to synthesise benzoic acid. This novel PAL gene may have a very dissimilar nucleotide sequence to the 'S. *maritimus*' *encP* gene, which would affect its detection by Southern hybridization and PCR amplification via the *encP* primers. Another possibility is that benzoic acid needs be provided to *S. polyantibioticus* SPR^T externally (which seems unlikely, as benzoic acid was not added to the growth medium), or the hypothesis that benzoic acid is a starting metabolite in the production of DPO may be incorrect and a different mechanism is used in its biosynthesis.

2.4.2 ADENYLATION DOMAIN AMPLIFICATION AND CLONING

NRPS adenylation domains were amplified from *S. polyantibioticus* SPR^T by primers specific for adenylation domain conserved motifs. These amplified products were subsequently cloned into pGEM-T Easy and pJET1.2/blunt, which were transformed into competent *E. coli* cells. The adenylation domains had to be cloned before sequencing because of the high possibility of there being more than one NRPS gene and, therefore, several adenylation domains within the genome of *S. polyantibioticus* SPR^T. Direct sequencing of the PCR products would have lead to mixed sequencing results.

These adenylation domain clones were screened by colony PCR to confirm the presence of inserts. After confirmation, the clones were sequenced. The sequences were subjected to nucleotide and protein BLAST analyses to confirm their identity as adenylation domains. In total 194 clones were obtained, of which only 59 were found to contain inserts of the correct length. These 59 clones were sequenced using appropriate primers. Only 10 adenylation domain sequences were of good quality. Out of these sequences, three unique sequences were identified, belonging to clones DS-D, PJ15 and PJ73 (all just under 700 bp in length). The sequences were translated and the correct open reading frame was identified in each case. The first 9 amino acid residues corresponding to the specificity binding code of adenylation domains were identified for each unique sequence. The identification of the amino acid residues was carried out by aligning the protein sequences of each clone against that of

the PheA (phenylalanine activating) domain of GrsA (1AMU_A), in a similar manner to Figure 1.5 in Chapter 1.

Additionally, these protein sequences were then fed into the NRPSpredictor program (Rausch *et al.*, 2005), which performs an alignment against the PheA domain of GrsA, but also analyses the physico-chemical fingerprint of the residues lining the specificity binding pocket to give an indication of the most probable specificity of the binding pocket. This specificity is dependent on two variables: either several substrate amino acids have the same properties or these substrate amino acids have properties that only a few share. BLAST analysis was used on the sequences of these adenylation domains to locate the closest relating sequences and also cross-check their specificity. The specificity binding code and likely amino acid substrate specificity for the three unique adenylation domains isolated from *S. polyantibioticus* SPR^T are listed in Table 2.3.

TABLE 2.3. Specificity Binding Pocket Code and Amino Acid Specificity

Adenylation domain Source	Specificity Binding Pocket Code Residue and Position[a]										Amino Acid Specificity[b]
	235	**236**	**239**	**278**	**299**	**301**	**322**	**330**	**331**	**517**	
Clone DS-D	D	M	V	Q	F	G	L	V	Y	-	Gly, Ala, Val
Clone PJ15	D	I	N	Y	W	G	G	I	G	-	Orn, Lys, Arg
Clone PJ73	D	F	W	N	V	G	M	V	H	-	Ser, Thr, Dht
GrsA	D	A	W	T	I	A	A	I	C	K	Phe, Trp, Phg
BarG[c]	D	A	W	T	V	A	A	V	C	K	Phe, Trp, Phg
Variability[d] %	3	16	16	39	52	13	26	23	26	0	

[a] According to GrsA numbering (1AMU_A).
[b] Predictions in order of decreasing preference, assuming that several substrate amino acids have the same properties as opposed to a few. If only a few amino acids share the same properties, there are less possible candidate amino acids in the binding pocket that can interact with a particular substrate, resulting in fewer possible substrates as binding pockets would be more specific.
[c] Part of the barbamide gene cluster in *Lyngbya majuscule* (AAN32981), a cyanobacterium.
[d] Proposed 'wobble'-like positions across adenylation domains and common specificities (Stachelhaus *et al.*, 1999).
Residue position 517 was not obtained for *S. polyantibioticus* SPR^T adenylation domain clones, but based on the high conservation of this residue, it is most likely lysine. Amino acid abbreviations use standard one and three letter codes. Dht - dehydrothreonine, Phg - L-phenylglycine.

The PheA domain of GrsA is a well known phenylalanine-specific adenylation domain. The adenylation domain of BarG is also specific for phenylalanine. The NRPSpredictor supports the specificity of phenylalanine for both these proteins based on their sequences and extends it to other non-polar aromatic amino acids such as tryptophan and L-phenylglycine. When comparing the two sequences, the key differences are at the 299 and 330 positions where BarG differs from GrsA with

valine residues instead of isoleucine. These positions are areas of high to medium variability, so this difference does not have a large impact on their substrate affinity. Residues with high variability may not have direct (or only very little) interaction with the substrate amino acid, so their absence may have little impact on substrate specificity. Isoleucine and valine are both non-polar amino acids, which may make them interchangeable, as they could perform the same function in the binding pocket interacting with non-polar phenylalanine. GrsA and BarG have very little similarity in binding pocket residues to DS-D, PJ15 and PJ73, which supports the evidence that suggests that these *S. polyantibioticus* SPR^T adenylation domains are specific for different substrate amino acids.

Clone DS-D was shown by the NRPSpredictor to have a high substrate specificity for glycine, which is considered weakly polar and alanine and valine, which are non-polar. All 3 residues are hydrophobic, having relatively small side chains and are not considered bulky, which could explain their similarity as substrates. Protein BLAST analysis of the amino acid sequence showed a high similarity of clone DS-D to an NRPS *in Paenibacillus polymyxa* E681 (YP_003871373) with an amino acid similarity of 58%, BaeJ in *Bacillus amyloliquefaciens* FZB42 (YP_001421292) with an amino acid similarity of 59% and Ta1 in *Myxococcus xanthus* (CAB38084) with an amino acid similarity of 57%, which all have similar substrate specificities. None of these BLAST matches is to an actinobacterium.

The NRPSpredictor showed Clone PJ15 to have a high substrate specificity for ornithine, lysine and arginine. Ornithine is not commonly found in proteins, however NRPSs and PKSs are known to include ornithine in the synthesis of molecules such as siderophores, which are Fe^{3+} carrying molecules (Challis & Ravel, 2000, Eisendle *et al.*, 2003, Garvey & Keller, 2010). Ornithine, lysine and arginine all have long, basic side chains, which would lead them to having similar molecular interactions within the binding pocket of the adenylation domain. Protein BLAST analysis of the amino acid sequence showed a high similarity to adenylation domains *in Streptomyces ambofaciens* ATCC 23877 (CAJ88262), *Streptomyces coelicolor* A3(2) (NP_624809) and a peptide synthetase from *Streptomyces hygroscopicus* ATCC 53653 (ZP_05511991), all of which share the same substrate specificity as PJ15 and identical amino acid similarities of 92%.

Clone PJ73 was shown by the NRPSpredictor to have a high substrate specificity for serine, threonine and dehydrothreonine, however, this only applies if many substrate amino acids have the same properties. If properties amongst the substrate amino acids are less common, threonine is the preferred substrate for Clone PJ73, as suggested by the NRPSpredictor. Various adenylation domains for strains of *Pseudomonas syringae* (ZP_06494975 with an amino acid similarity of 73%, ZP_07265550 also with an amino acid similarity of 73% and ZP_05641707 with an amino acid similarity of 72%), *Ralstonia solanacearum* GMI1000 (NP_522202) with an amino acid similarity of 72% and a

arthrofactin synthetase/syringopeptin synthetase NRPS from *S. hygroscopicus* ATCC 53653 (ZP_05519317) with a similarity of 80%**,** were revealed as the closest relatives to Clone PJ73, determined by protein BLAST analysis of the amino acid sequence. None of these BLAST matches is to an actinobacterium, except *S. hygroscopicus* ATCC 53653 (ZP_05519317).

Serine residues are common starting units for the biosynthesis of oxazoles (Roy *et al*., 1999). It may be possible that the adenylation domain cloned into PJ73 may recognise the starting substrate of serine, instead of phenylalanine, and utilise it for the biosynthesis of DPO. Serine, threonine and dehydrothreonine are all polar amino acids and therefore would interact with the binding pocket in a similar manner. If serine is activated in DPO synthesis, the addition of a phenyl group to carbon 3 of serine would be required.

Based on the analysis conducted, clones DS-D, PJ15 and PJ73, containing adenylation domains from *S. polyantibioticus* SPR^T, are not specific for phenylalanine. These adenylation domains are too dissimilar from known phenylalanine specific adenylation domains, such as PheA in GrsA and BarG. It should be noted that there is a slight chance that these domains are specific for other amino acids than those suggested by the NRPSpredictor.

Out of the 10 clones containing confirmed adenylation domains, five contained the PJ15 (orn/lys/arg) adenylation domain, four contained the DS-D (gly/ala/val) adenylation domain and only one contained the PJ73 (ser/thr/dht) adenylation domain. As actinomycetes are known to contain multiple NRPS genes (Sosio *et al*., 2000), *S. polyantibioticus* SPR^T may contain other adenylation domains that were not amplified by PCR and cloned in this study. Thus, the phenylalanine specific adenylation domain in *S. polyantibioticus* SPR^T is still waiting to be discovered.

2.4.3 SOUTHERN HYBRIDIZATION USING ADENYLATION DOMAIN PROBES

The choice of pairwise restriction endonuclease combinations for the Southern hybridization experiments was made in an attempt to isolate DNA fragments large enough to provide additional sequence data on the *S. polyantibioticus* SPR^T NRPSs. Endonucleases chosen were also based on their ability to digest G+C rich regions in DNA, common in actinomycetes and NRPS sequences. In total, three different Southern hybridizations were performed using three different probes – the adenylation domain clones PJ15, PJ73 and DS-D (Figure 2.5). Many different bands were observed ranging from 0.57kb to 11.81kb in length. The sizes of the fragments obtained are summarised in Table 2.4.

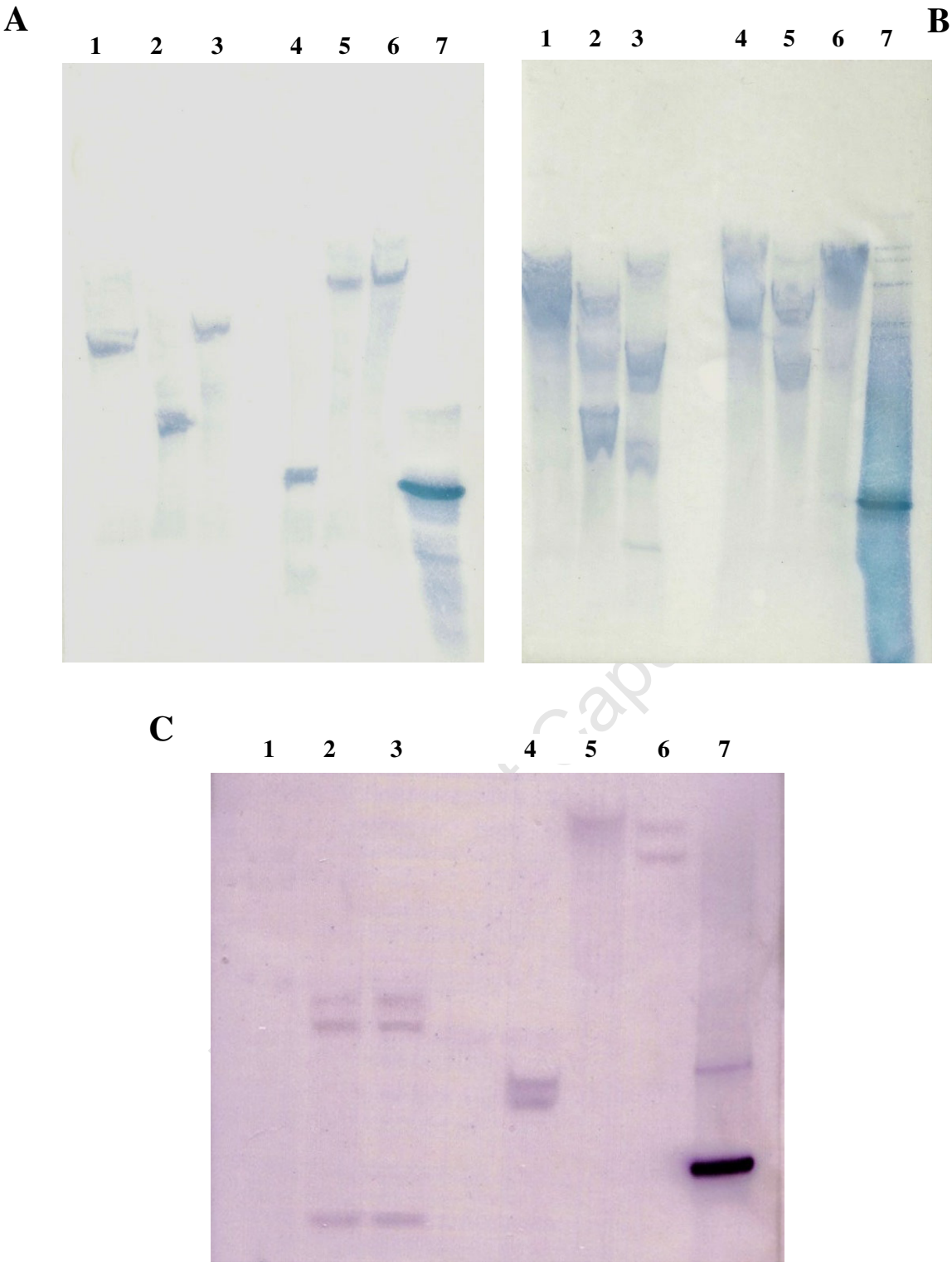

Figure 2.5 Southern hybridization of restriction endonuclease digested *S. polyantibioticus* SPR^T genomic DNA using the PJ15 (**A - Top Left**), PJ73 (**B - Top Right**) and DS-D (**C – Bottom Middle**) adenylation domains as probes. Lanes 1-6 contain identical pairwise endonuclease digests of 20 µg *S. polyantibioticus* SPR^T genomic DNA in each Southern hybridization, with the following combinations; Lanes: **1**- *Sph*I& *Stu*I, **2**- *Pst*I & *Pvu*I, **3**- *Xho*I & *Pvu*I, **4**- *Pvu*II & *Sph*I, **5**- *Pst*I & *Kpn*I, **6**- *Avr*II & *Stu*I, **7**- Unlabelled Probe (positive control).

TABLE 2.4. Summary of Bands and Relative Sizes Observed for Southern Hybridizations

Probe	Lane Number Including Endonuclease Pair and Band Size(s) (kb)						
	1	**2**	**3**	**4**	**5**	**6**	**7**
	***Sph*I & *Stu*I**	***Pst*I & *Pvu*I**	***Xho*I & *Pvu*I**	***Pvu*II & *Sph*I**	***Pst*I & *Kpn*I**	***Avr*II & *Stu*I**	**Unlabelled Probe**
PJ15	2.62	1.23	3.24	0.84	5.29	5.8	0.77
PJ73	4.93	1.52, 4.72	0.57, 1.09, 1.92, 7.91	3.84	1.92, 4.17	7.15, 8.79	0.73
DS-D	No result	0.69, 1.92, 2.55	0.69, 1.92, 2.55	1.19, 1.34	11.81	7.66	0.84

It can clearly be seen that numerous bands were detected in the Southern hybridization experiments. Different enzyme combinations generated DNA fragments of various sizes containing the target adenylation domain. However, the probes may have hybridized to different adenylation domains in the genomic DNA or fragments of the same adenylation domain produced by the digestions. *In silico* digests of the probes revealed DS-D is digested by *Pvu*I at position 190, *Stu*I at 389 and *Sph*I at 279; PJ15 is digested by *Pvu*II at 526 and PJ73 by *Xho*I at 88 and *Pvu*II at 49 and 367. Therefore not all of the bands seen are different adenylation domains. In particular, lanes where some of the additional bands may indicate the same adenylation domain are lanes 2 & 3 for DS-D and lanes 3 & 5 for PJ73. Besides the lanes mentioned with multiple bands, which could be due to the target adenylation domain being cleaved, there are multiple bands present in other lanes. These lanes may contain hybridization bands indicating the presence of other adenylation domains in the genome, perhaps part of a different module on the same NRPS or an entirely different NRPS. Since NRPS genes exist in clusters, it is possible that one of these bands represents a fragment containing one NRPS's adenylation domain from a gene cluster involved in the synthesis of DPO.

Given that the sizes of these bands are known, gel electrophoresis of genomic DNA digested using the same endonuclease pair can be performed and DNA of the appropriate size can be excised from the gel and cloned. Sequencing of these clones would yield sequencing data of DNA flanking the adenylation domains such as condensation, cyclization or thioesterase domains. We may also be able to identify novel adenylation domains within *S. polyantibioticus* SPR^T in this way.

2.4.4 CLONING OF FRAGMENTS DETECTED VIA SOUTHERN HYBRIDIZATION

Various plasmids were used in an attempt to clone the digested genomic DNA, namely, pUC18, pGEM-T Easy, pJET1.2/blunt and pEcoR251 using various sizes of inserts determined by Southern hybridization. DNA fragments in the appropriate size ranges were excised from agarose gels for cloning.

Clone transformation efficiency was relatively low using all plasmids. This was mainly attributed to the low quantity and possible degradation of the genomic DNA yielded by gel extraction. To complicate matters, commercial vectors such as pGEM-T Easy are designed to ligate with inserts produced by PCR with *Taq* polymerase, which creates adenine overhang tails (A-tails). pGEM-T Easy has thymine overhangs in its multiple cloning site, thus requiring inserts to possess A-tails. Therefore if inserts are not produced via PCR, such as gel extracted genomic digests, inserts must then be A-tailed using an A-tailing procedure if they are blunt-ended or blunted firstly using a Klenow fragment if they are sticky ended, then A-tailed. Blunting may also cause vectors to be read out of frame. Suicide vectors that are blunted result in the suicide gene (*Eco*RI endonuclease in pEcoR251 and *Eco*47IR in pJET1.2/blunt) being read out of frame, rendering it ineffective. The manufacturers of pJET1.2/blunt and pGEM-T easy claim that their vectors result in successful clones 90% and 70-100% of the time, respectively. This was certainly not the case in this study, as less than 50% of clones, using both vectors, contained an insert of the correct length. This unexpectedly poor efficiency of both vector systems had a negative effect on the isolation of clones, as many cloning attempts were required to obtain suitable clones for analysis.

Suicide/positive selection vectors, such as pJET1.2/blunt, are supposed to remove the need to screen every clone for an insert due to the expression of the endonuclease gene in non-transformants, which leads to the digestion of the *E. coli* chromosomal DNA and thus kills the host.

When pUC18 was used for cloning, dephosphorylation of the vector was carried out to prevent self ligation, but this seemed to be unsuccessful due to the low number of clones obtained. Dephosphorylation controls showed that an overnight incubation of linearised pUC18 with alkaline phosphatase still resulted in self-ligated (empty) vectors.

Clones were screened by colony PCR to identify the presence of inserts of appropriate sizes. The few clones that were successfully transformed were sent for sequencing. Very few clones produced good sequence data and none had sequences pertaining to adenylation domains, other NRPS domains or their flanking regions. This method to identify larger pieces of DNA flanking adenylation domains is theoretically very useful, but in practice there are many obstacles, which cause some difficulty in obtaining good results. If more suitable vectors are acquired or a more efficient protocol for gel

extraction obtained, the quantity of successfully transformed clones with the correct insert should increase. Cosmid vectors may be suitable for this kind of experiment as they can hold much larger inserts and a genomic library can be created. This library could be screened using adenylation domain primers to identify rapidly cosmids containing sequences coding for NRPSs.

2.4.5 CYCLIZATION DOMAIN AMPLIFICATION

The cyclization domain PCR primers were designed based on an alignment of cyclization domains from known oxazole and thiazole producers (Table 2.2). *S. virginiae* NRRL B-1446^T was used as a positive control strain for these primers, as it is an oxazole producer and the *S. virginiae virH* gene (for virginiamycin M1 biosynthesis) was used in designing the cyclization domain primers. Although an adenylation domain PCR fragment was amplified from *S. virginiae* NRRL B-1446^T (data not shown), which proves the existence of an NRPS within its genome, no cyclization domain fragment was amplified from *S. virginiae* NRRL B-1446^T or *S. polyantibioticus* SPR^T. It is possible that the primers did not work, because the *virH* sequence from GenBank appears to be from a different strain of *S. virginiae*, namely, MAFF 10-06014 (Pulsawat *et al*., 2007). The strain used in this study was the type strain of *S. virginiae*, which may not contain *virH* (antibiotic-producing ability in actinomycetes is known to be strain specific, not species specific).

S. mobaraensis DSM 40903 was then used as a new positive control strain to validate the cyclization primers. *S. mobaraensis* DSM 40903 is the same strain as '*S. verticillus*' ATCC 15003. The name, *Streptomyces verticillus*, has never been validly published (http://www.bacterio.cict.fr/s/streptomycesc.html) and therefore has no standing in bacterial nomenclature. The Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) record for *S. mobaraensis* DSM 40903 indicates that strain DSM 40903 is identical to '*S. verticillus*' strain ATCC 15003 (http://www.dsmz.de/microorganisms/html/strains/strain.dsm040903.html; see also http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=29309). Thus, the *blmIV* gene sequence from '*S. verticillus*' ATCC 15003 that was used in designing the cyclization domain primers (Table 2.2) should be exactly the same strain used as a positive control for the PCR experiments.

No cyclization domain fragment was amplified from *S. mobaraensis* DSM 40903. It was then realised that the Cy-R primer was binding in the incorrect orientation. This lead to the redesign of the reverse cyclization domain primer. Smob Cy-R was the replacement reverse primer and was designed specifically to bind to the *blmIV* sequence of '*S. verticillus*' ATCC 15003. PCR using Cy-F and Smob Cy-R resulted in numerous bands being amplified for *S. mobaraensis* DSM 40903 and no amplification for *S. polyantibioticus* SPR^T. A PCR experiment was performed using each primer alone to confirm specificity (Figure 2.6).

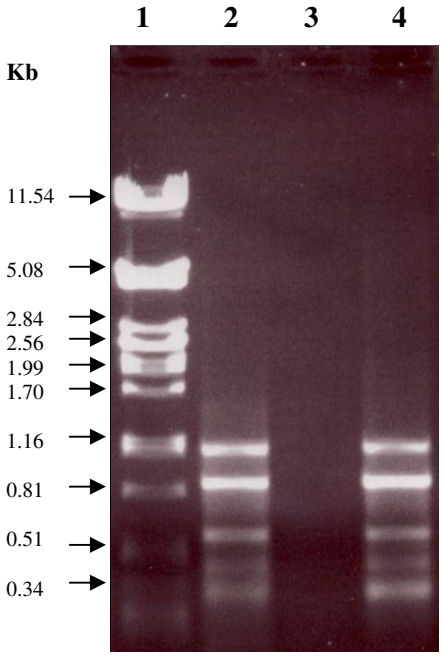

Figure 2.6 Gel electrophoresis of PCR amplified DNA from *S. mobaraensis* DSM 40903 displaying results when using single and paired primers. Lanes: **1**- λ-*Pst*I molecular marker, **2**- Cy-F & Smob Cy-R, **3**- Cy-F only, **4**- Smob Cy-R only.

A total of 5 bands were observed for amplification using both primers and Smob Cy-R individually. This PCR using single primers confirmed that Smob Cy-R was responsible for all the bands seen, indicating amplification of non-target sequences. According to the sequence for the *blmIV* gene of '*S. verticillus*' ATCC 15003, a fragment of 1055 nucleotides should be amplified from the genome. It was very strange that there were no additional bands when both primers were used. This suggested that Cy-F was not binding to its target DNA. We suspected we were working with the incorrect strain, lacking NRPS genes. To test this, adenylation domain amplification in *S. mobaraensis* DSM 40903 was used to confirm the existence of NRPSs in its genome (Figure 2.7).

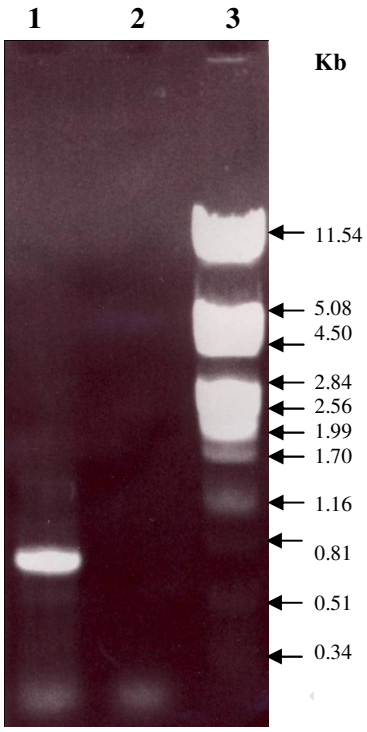

Figure 2.7 Gel electrophoresis of PCR amplified adenylation domains from *S. polyantibioticus* SPR^T and *S. mobaraensis* DSM 40903. Lanes: **1**- *S. polyantibioticus* SPR^T, **2**- *S. mobaraensis* DSM 40903, **3**- λ-*Pst*I molecular marker.

The expected single band of 0.7 kb was observed for *S. polyantibioticus* SPR^T, but there was no adenylation domain amplification for *S. mobaraensis* DSM 40903. The existence of no adenylation domains in *S. mobaraensis* DSM 40903 or no amplifiable adenylation domains is extremely peculiar. No adenylation domains in the genome suggest that there are no NRPSs and therefore no cyclization domains. This would explain the lack of binding by Cy-F and the non-specific amplification by Smob Cy-R. One possibility is that the *S. mobaraensis* DSM 40903 used here is, in fact, ***not*** the same strain as a '*S. verticillus*' ATCC 15003. If this is the case, then the strain used for designing the cyclization domain primers and the positive control strain used to test these primers are different strains and would be expected to give different PCR-amplification results. A mistake with the identification of the strain deposited as *S. mobaraensis* DSM 40903 could have been the cause of this problem.

It is unfortunate that we did not have a reliable control strain to test the cyclization domain primers against. Using '*S. verticillus*' ATCC 15003 directly as a positive control for the primers would be the best step forward, as correct amplification would allow clarification of whether the lack of amplification in *S. polyantibioticus* SPR^T was due to the primer design or the lack of the target sequence.

The successful amplification of cyclization domains would produce slightly larger fragments of DNA than the adenylation domain primers. These fragments could be sequenced and then, by BLAST analysis and comparison with thiazole and oxazole producers, the function of the amplified cyclization domain could be confirmed. Proof of the presence of a cyclization domain in *S. polyantibioticus* SPR^{T} would provide support for the current hypothesis on how DPO is synthesised. It would also allow for an even larger section of the DPO NRPS to be amplified by PCR using the cyclization domain forward primer (Cy-F) and the adenylation domain reverse primer (A7R) (Figure 2.2). This larger fragment could be used as a highly-specific probe for the DPO NRPS in a Southern hybridization experiment.

2.5 CONCLUSION

This study was successful in demonstrating the existence of NRPSs present within the genome of *S. polyantibioticus* SPR^T. Although the specificity of the isolated adenylation domains was predicted, there is still room for error in this regard and these domains could have different specificities from those suggested. It is still uncertain whether DPO ***is*** synthesised by a phenylalanine specific NRPS, let alone whether the correct adenylation domain involved in DPO synthesis was cloned. It is thus important that the biosynthetic hypothesis be confirmed by further efforts. If DPO is being biosynthesized via the proposed mechanism, is it important to understand how benzoic acid may be supplied as a starter unit. It was shown in this study that *S. polyantibioticus* SPR^T most likely does not possess a gene coding for a bacterial PAL involved in synthesising benzoic acid. However, *S. polyantibioticus* SPR^T may contain a rare PAL that is completely novel in bacterial systems. This unique PAL may be undetectable via PCR with *encP* primers or Southern hybridization with *encP* as a probe. Unfortunately, the lack of a PAL gene provides evidence against the currently proposed biosynthetic pathway.

Using cyclization domains along with adenylation domains as anchors to locate the NRPS gene of interest is perfectly feasible. It is crucial that a reliable positive control strain be obtained for testing the cyclization domain PCR primers and for Southern hybridizations. Knowing the sequence and position of 2 different domains within an NRPS gene creates greater potential for further sequencing by primer walking. Once more sequence for the regions immediately surrounding the adenylation domains isolated in this study is obtained, larger portions of the NRPS gene and eventually the entire DPO biosynthetic gene cluster can be uncovered by utilizing primer walking upstream and downstream of the adenylation domain and cyclization domain.

If the correct adenylation domain involved in DPO synthesis has already been isolated (e.g. PJ73), sequencing of adjacent regions by primer walking, followed by BLAST analysis, will reveal the function of the genes sequenced in this way. Further investigation into fine tuning of the techniques in cloning the DNA fragments detected via Southern hybridization may also allow more gene sequences to be obtained. Eventually this will allow combinatorial biosynthesis experimentation to begin and thus create antibiotic derivatives of DPO.

Future work into the characterization of the genes involved in the production of DPO should involve further testing of the biosynthetic-pathway hypothesis and the sequencing of the genomic DNA fragments cloned in this study and their flanking sequences. Further attempts at isolating adenylation domain clones could lead to the isolation of the domain specific for phenylalanine. Furthermore, studying DPO synthesis at the biochemical level in *S. polyantibioticus* SPR^T would gather important data with regard to how DPO is synthesised and would provide support for the current hypothesis or lead to the development of a new hypothesis.

2.6 REFERENCES

Adrova, N.A., Koton, M.M., & Florinsky, F.S. (1956). Preparation of 2,5-Diphenylaoxazole and its Scintillation Efficiency in Plastics. *Institute of Macromolecular Compounds of the Academy of Sciences of the USSR*, 394-395.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & **Lipman, D.J. (1997).** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* **25:** 3389-3402.

Ayuso-Sacido, G. & Genilloud, O. (2005). New PCR Primers for the sequencing of NRPS and PKS-I Systems in Actinomycetes: Detection and Distribution of These Biosynthetic Gene Sequences in Major Taxonomic Groups. *Microbial Ecology*, **49:** 10-24.

Challis, G.L. & Ravel, J. (2000). Coelichelin, a new peptide siderophore encoded by the Streptomyces coelicolor genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. *FEMS Microbiology Letters*, **187:** 111-114.

Cook A.E. & Meyers P.R. (2003). Rapid identification of filamentous actinomycetes to the genus level using genus-specific 16S rRNA gene restriction fragment patterns. *International Journal of Systematic Evolutionary Microbiology*, **6:** 1907-1915.

Dagert, M. & Ehrlich, S.D. (1979). Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. *Gene*, **6:** 23-28.

Eisendle, M. ,Oberegger, H., Zadra, I. & Haas, H. (2003). The siderophore system is essential for viability of *Aspergillus nidulans*: functional analysis of two genes encoding l-ornithine N 5-monooxygenase (*sidA*) and a non-ribosomal peptide synthetase (*sidC*). *Molecular Microbiology*, **49:** 359-375.

Garvey, G.S. & Keller, N.P. (2010). Fungal Secondary Metabolites and Their Fundamental Roles in Human Mycoses. *Current Fungal Infection Reports*, **4:** 256-265.

Gu, L., Jia, J., Liu, H., Håkansson, K., Gerwick, W.H. & Sherman, D.H., (2006). Metabolic coupling of dehydration and decarboxylation in the curacin A pathway: functional identification of a mechanistically diverse enzyme pair. *Journal of the American Chemical Society*, **128:** 9014-9015.

Leo, W. R. (1994). Techniques for Nuclear and particle Physics Experiments, 2nd edition, *Springer-Verlag*, Pg 158.

Pfefferle, C., Theobald, U., Gürtler, H. & Fiedler, H. (2000). Improved secondary metabolite production in the genus *Streptosporangium* by optimization of the fermentation conditions. *Journal of Biotechnology*. **80:** 125-142.

Pulsawat, N., Kitani, S. & Nihira, T. (2007). Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae*. *Gene*, **393:** 31-42.

Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D.H. (2005). Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using Transductive Support Vector Machines (TSVM). *Nucleic Acids Research,* **33:** 5799-5808.

Reed, K.C. & Mann, D.A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Research*, **13:** 7207-7221.

Roy, R.S., Gehring, A.M., Milne, J.C., Belshaw, P.J. & Walsh, C.T., (1999). Thiazole and oxazole peptides: biosynthesis and molecular machinery. *Natural Product Reports*, **16:** 249-63.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Bacterial media, antibiotics and bacterial strains. Molecular Cloning, a laboratory manual, 2nd Ed. Cold Spring Harbour: *Cold Spring Harbour Laboratory Press*.

Shirling, E. B. & Gottlieb, D. (1966). Methods for characterization of *Streptomyces* species. *International Journal of Systematic Bacteriology,* **16:** 313-340.

Sosio, M., Bossi, E., Bianchi, A. & Donadio, S., (2000). Multiple peptide synthetase gene clusters in Actinomycetes. *Molecular and General Genetics*, **264:** 213-221.

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

Wang, Y., Zhang, Z. & Ruan, J. (1996) A proposal to transfer *Microbispora bispora* (Lechevalier, 1965) to a new genus, *Thermobispora* gen. nov., as *Thermobispora bispora* combo. nov. *International Journal of Systematic Bacteriology*, **46:** 933 – 938.

Yanisch-Perron C., Vieira J. & Messing J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33:** 103-119

Zappe, H., Jones, D.T. & Woods, D.R. (1986). Cloning and expression of *Clostridium acetobutylicum* endoglucanase, cellobiase and amino acid biosynthesis genes in *Escherichia coli. Journal of General Microbiology,* **132:** 1367-1372.

CHAPTER 3

GENERAL DISCUSSION

CHAPTER 3

GENERAL DISCUSSION

This study was successful in testing a technique to isolate adenylation domain sequences from NRPS gene clusters in *S. polyantibioticus* SPR^T. A similar technique was extended to cyclization domains, however, the primers designed to amplify these domains still require further analysis and testing. Despite a phenylalanine specific adenylation domain not being isolated from *S. polyantibioticus* SPR^T in this study, three adenylation domains, each with unique amino-acid substrate specificities, were isolated. Further attempts at isolating adenylation domains from *S. polyantibioticus* should be undertaken, as an adenylation domain specific for phenylalanine may still be found.

The specificities of the 3 adenylation domains isolated in this study were identified by firstly aligning the adenylation domain amino acid sequences with the PheA domain of GrsA to locate the binding pocket amino-acid residues. The NRPSpredictor program (Rausch *et al.*, 2005) then analysed the physico-chemical properties of these residues and suggested substrate specificities based on this analysis, along with a comparison with similar binding pockets to confirm this prediction. This approach is mostly accurate, however, examples of codes for which the predicted substrate specificity does not correspond to the actual activated amino acid have been reported (Lombo *et al*., 2006). It may be possible that, in addition to the 10 key residues of the binding pocket, there are other residues involved in dictating the identity of the substrate amino acid activated by an adenylation domain (McQuade *et al*., 2008).

The protein encoded by adenylation domain sequence PJ73 is predicted to be specific for serine, which is a common starting unit for the biosynthesis of oxazoles (Roy *et al*., 1999). If serine is activated in DPO synthesis, the addition of a phenyl group to carbon 3 of serine would be required. This β-phenylation of an activated serine seems less likely than the β-hydroxylation of an activated phenylalanine by a P450 monooxygenase. However, perhaps β-phenylation *is* used by *S. polyantibioticus* (le Roes, 2005).

The protein encoded by adenylation domain sequence PJ15 is predicted to be specific for ornithine, which is a common starter unit for the synthesis of siderophores. This is interesting, as siderophores have been used to create siderophore-antibiotic conjugates called sideromycins, which can be naturally occurring or synthetic. These conjugates act as Trojan horses, where transport of the siderophore into cells is used to allow entry of the antibiotic. This is a highly effective approach against those pathogens possessing an export mechanism to expel antibiotics (Miller *et al*., 1991,

Budzikiewicz, 2001, Braun *et al*., 2009). The PJ15 adenylation domain may be involved in siderophore biosynthesis in *S. polyantibioticus* SPR^T.

Examples of shortened adenylation domains exist in the literature. In α-lipomycin biosynthesis in *Streptomyces aureofaciens* Tü117, the adenylation domain of the α-lipomycin NRPS (*lipNrps*) was described as being 280 amino acids shorter than other previously described NRPS A domains, making it impossible to predict the constituents of the substrate binding pocket. The existence of this domain is crucial to the production of α-lipomycin, as mutants with this gene knocked out did not produce α-lipomycin (Bihlmaier *et al*., 2006). To emphasise the significance of these shortened adenylation domains, the adenylation domains involved in the biosynthesis of streptolydigin in *Streptomyces lydicus* NRRL 2433, S1gN1 and S1gN2 were also shown to be missing core motifs thought to be present in all adenylation domains. S1gN1 is missing motifs A3 and A4, which form part of the deduced binding pocket for adenylation domains. In addition, S1gN2 is missing four motifs, A1, A2, A5 and A6, which makes it impossible to predict the constituents of the substrate binding pocket by comparison to other NRPS systems (Olano *et al*., 2009). Both α-lipomycin and streptolydigin are tetramic acid antibiotics.

If the occurrence of these shortened adenylation domains missing core motifs is common in nature, it is possible that these domains may be present in many NRPSs, but their distinct lack of similarity to available sequences could hamper their identification and discovery. The adenylation domain primers used in this study were designed to bind to the conserved motifs A3 and A7 (Ayuso-Sacido & Genilloud, 2005). Shortened adenylation domains may not be able to be amplified because of the lack of these primer target sequences. Additionally, if PCR products *are* produced, they may be shorter than the predicted size, which may cause them to be discarded or ignored. Adenylation domains within *S. polyantibioticus* containing truncated sequences may not have been amplified and any truncated phenylalanine-specific adenylation domain, would not have been identified.

Previous work forming the foundation for this project involved the exploration and identification of various antibiotics produced by *S. polyantibioticus* SPR^T (le Roes, 2005) and its classification as a novel bacterial species (Le Roes-Hill & Meyers, 2009). This work showed the presence of five anti-*M. aurum* A+ activity spots using bioautography. Unfortunately, only DPO was isolated and described. It is possible that the adenylation domains isolated in this study are involved in the production of these other, unclassified antibiotics if their synthesis involves NRPSs. The adenylation domains identified in this study could also form part of a single NRPS consisting of multiple modules responsible for the synthesis of one peptide antibiotic.

Although cyclization domains could not be amplified from *S. polyantibioticus* SPR^T due to experimental difficulties, other techniques may be employed to sequence regions adjacent to the isolated adenylation domains, including flanking cyclization domains. Primer walking or 'targeted gene walking' has been used successfully for many sequencing and gene characterization studies (Parker *et al*., 1991). This technique involves the use of a single specific PCR primer matching the known sequence and a second, non-specific 'walking' primer. Together with a single step PCR protocol, this method can be used to isolate sequences upstream or downstream from a known sequence. Another method used to identify unknown DNA fragments located adjacent to available sequences is SiteFinding-PCR, which uses a nested-PCR protocol (Tan *et al*., 2005). SiteFinding-PCR uses a gene-specific primer along with a SiteFinder oligonucleotide primer and together they amplify DNA sequences between them exponentially. The SiteFinder oligonucleotide primer is highly degenerate and binds at random positions in the genomic DNA. Importantly, non-target sequences are not amplified exponentially (i.e. between two SiteFinder primer binding sites), owing to the suppression effect of product stem–loop structure formation. Recent work has optimised the sequence of the SiteFinder oligonucleotide primer to bind less frequently in actinobacterial GC-rich DNA, creating larger amplicons for analysis (Gröning *et al*., 2010). Although it is recommended that the SiteFinding-PCR method be attempted in *S. polyantibioticus* SPR^T, the method used in this study to isolate cyclization domain sequences should produce results once a reliable control is obtained on which to test and optimise the designed primers. However, for both methods to yield useful results, the DPO-specific adenylation or cyclization domain must first be identified.

Despite *S. polyantibioticus* SPR^T appearing not to possess a PAL gene similar to *encP*, it may produce benzoyl-CoA via another mechanism. Besides the aerobic process catalysed by PAL in *'S. maritimus'*, other aerobic and anaerobic pathways for the production of benzoyl-CoA and derivatives exist. There are examples of aerobic non-PAL pathways in nature, often with a starting unit of phenylacetic acid (PA). However, not all of these pathways are completely understood. For example, aerobic soraphen A biosynthesis in *Sorangium cellulosum* begins with a phenyl side group that is derived from phenylalanine, but the exact pathway involving its incorporation is unknown - it is assumed that the phenyl ring is integrated via carboxybenzoyl-CoA (Schupp *et al*., 1995, Ligon *et al*., 2002).

PA is produced anaerobically in *Thauera aromatica* via the transamination of phenylalanine to phenylpyruvate, which is then decarboxylated to phenylacetaldehyde followed by dehydrogenation to PA (Schneider *et al*., 1997). It is important to note that this pathway exists for the degradation of the aromatic ring and cannot be considered a biosynthetic pathway. There are also some well established aerobic degradation pathways for phenylacetic acids, all resulting in ring cleavage of the phenyl group (Mohamed *et al*., 2002). Many more examples exist of anaerobic degradation of benzoyl-CoA with

numerous starting units including phenylalanine, vanilline, *p*-cresol, phenol, aniline, benzyl alcohol, toluene and phenylpropionate (Harwood *et al*., 1999).

In the β-subclass proteobacterium *Azoarcus evansii*, PA is degraded via an anaerobic mechanism to benzoyl-CoA (Figure 3.1 B). PA derived from phenylalanine is converted to phenylacetyl-CoA (PA-CoA) by CoA thioesterification, which is common in the anaerobic catabolism of aromatic compounds. PA-CoA is then oxidised to phenylglyoxylate, followed by a final oxidative decarboxylation to benzoyl-CoA. This is similar to the pathway present in *T. aromatica* (Rhee & Fuchs, 1999). It is suspected that an aerobic pathway coexists in *A. evansii* to fully metabolise PA (Figure 3.1 A), but all that has so far been uncovered is that PA-CoA, produced in an identical fashion in anaerobic biosynthesis, may be instead converted by the addition of oxygen to 2-hydroxyphenylacetyl-CoA acid. Benzoyl-CoA may be an intermediate further down this aerobic pathway, but more research is required to provide further clarification (Mohamed, 2000, Mohamed *et al*., 2002).

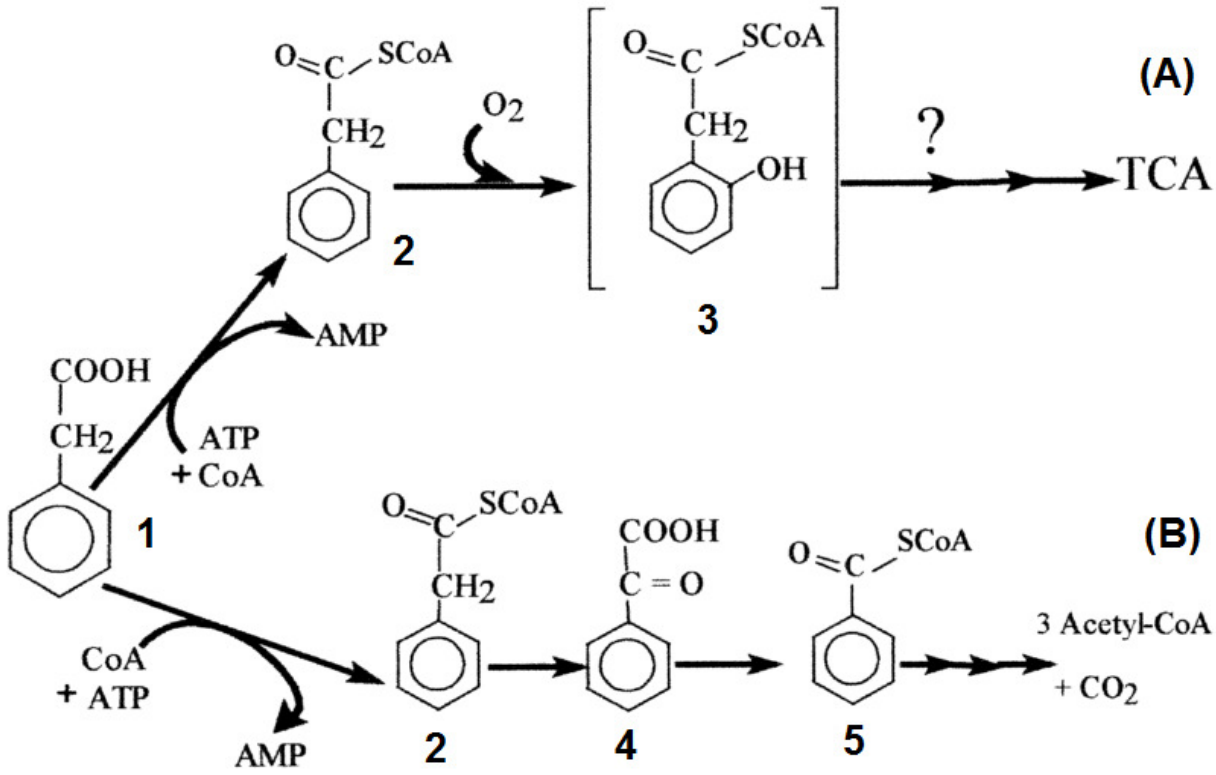

Figure 3.1 Phenylacetic acid (PA) metabolic pathways in *A. evansii*. **(A)** represents the aerobic pathway and **(B)** the anaerobic pathway. Phenylacetate-coenzyme A ligase (PA-CoA ligase) catalyses the CoA thioesterification of PA to PA-CoA in the aerobic pathway and (possibly an isozyme) the identical reaction in the anaerobic pathway. **1**- phenylacetic acid (PA), **2**- phenylacetyl-CoA, **3**- 2-hydroxyphenylacetyl-CoA (parentheses indicates **3** is part of a working hypothesis), **4**- phenylglyoxylate, **5**- benzoyl-CoA. Products of the aerobic pathway may be fed into the TCA cycle. Adapted from Mohamed *et al*. 2002.

The initial step of this PA aerobic pathway in *A. evansii*, involving the activation of PA to PA-CoA, has been reported in other bacterial species including *Pseudomonas putida* and *E. coli*. Phenylacetate-coenzyme A ligase (PA-CoA ligase) has been found to catalyse this initial reaction in *A. evansii*. Furthermore, the gene coding for PA-CoA ligase (*paaK*) has also been identified in *P. putida* and *E. coli* and the PA-CoA ligases share many biochemical and catalytic features, including highly conserved amino acid motifs. The gene encoding this enzyme is 1320 bp long and codes for a protein of 48.75 kDa (440 amino acids).

These PAL-CoA ligase enzymes for the aerobic and anaerobic PA degradation pathways in *A. evansii* are considered to be distinct, as the biochemical and molecular characteristics of PA-CoA ligase for the aerobic pathway are quite different from those of the enzyme catalyzing the same reaction under anaerobic conditions in the same bacterium (Mohamed, 2000, Mohamed *et al*., 2002). Furthermore, PA-CoA ligase was isolated from a genomic library of *A. evansii* by probing with a fragment of the gene encoding the isoenzyme that is induced under anaerobic conditions (Rost *et al*., 2002), suggesting that the genes for PA-CoA ligase and its isoenzyme are not too dissimilar. Thus *A. evansii* has two enzymes recognising identical substrates, having similar genetic sequences, but differing in biochemical properties and the oxygen requirement for their induction.

Although the products and intermediates of the PA aerobic pathway remain unclear, it is possible that benzoate may be the final product in the aerobic metabolism of PA, which would then be metabolised by the cell and used as a carbon source, as seen in *T. aromatica* (Schneider *et al*., 1997). However, neither ring-hydroxylating nor ring-cleaving reactions for PA have been observed in *A. evansii* so far (Mohamed *et al*., 2001), suggesting that the phenyl group could be conserved through the pathway and possibly result in a benzoate-like molecule. Since PA-CoA or an isoenzyme could be involved in either aerobic or anaerobic metabolism, searching for a similar gene to *paaK* (coding for PA-CoA ligase) in *S. polyantibioticus* SPR^T could indicate its ability to synthesise benzoyl-CoA aerobically from PA. Furthermore, searching for the genes coding for the enzymes involved in the conversion of phenylalanine to PA may also be useful. Many enzymes are involved in this process, such as L-phenylalanine:2-oxoglutarate amino transferase, L-phenylalanine NAD+ oxidoreductase (deaminating), phenylpyruvate decarboxylase and phenylacetaldehyde dehydrogenase. The genes coding for these enzymes may be probed for or PCR-amplified from the genome in a similar way to *encP*.

Another approach to validate the current hypothesis of DPO synthesis in *S. polyantibioticus* SPR^T could be to use feeding experiments using ^{13}C-labelled substrates. ^{13}C-Labelled phenylalanine and benzoate can be purchased and supplied to separate cultures of the bacterium. Once these substrates have been allowed time to be metabolised, the DPO can be purified and subjected to ^{13}C nuclear

magnetic resonance (NMR) to determine whether the ^{13}C has been incorporated into the antibiotic. This will give a clear indication of how DPO is synthesised if it carries any ^{13}C in its structure. When DPO was first isolated from *S. polyantibioticus* SPR^{T}, there was no benzoate or any molecules of similar structure in the culture medium, prompting the idea of its synthesis from benzoate and phenylalanine produced by *S. polyantibioticus* SPR^{T}. If ^{13}C labelled benzoate is supplied in this experiment in the culture medium and the cells can take up the extracellular benzoate, the cells should not discriminate and use it equally with self produced benzoate (the same applies to feeding with ^{13}C-phenylalanine). However, there are important considerations before conducting this experiment. Besides being very costly, *S. polyantibioticus* SPR^{T} is likely to only produce DPO under certain conditions. The formation of antibiotics is regulated by nutrients, growth rate, feedback control, enzyme inactivation, and enzyme induction (Demain, 1998). The appropriate conditions will need to be optimised for *S. polyantibioticus* SPR^{T} before feeding experiments begin. Antibiotics are part of secondary metabolism, which is brought on by exhaustion of a nutrients or addition of an inducer, and/or by a growth rate decrease (Demain, 1998). The stationary phase of cell growth is the most likely period for antibiotic production in bacterial cultures, therefore it would be the best time to add the ^{13}C-labelled substrate to the *S. polyantibioticus* SPR^{T} culture.

Further attempts at cloning sequences detected via Southern hybridization is currently the most apparent way forward in the molecular-genetic part of this study. Alternatively, a genomic library could be constructed. Different kinds of vectors, such as cosmids may prove useful in this regard, as streptomycete genomes are large (> 8 Mb). Cosmid vectors are better suited than plasmid vectors for this kind of experiment, as they can hold much larger inserts and a genomic library could be created with fewer clones. This library could be screened by PCR using adenylation domain primers to identify rapidly cosmids containing sequences coding for NRPSs. The gene of interest is guaranteed to be somewhere in this library compared to the current Southern-hybridization-lead mini-genomic library method where not a single clone may contain the gene of interest. The genomic-library route will still require a lot of screening, as *S. polyantibioticus* SPR^{T} was shown in this study to have several NRPS adenylation domains, all of which would be detected in the PCR screening of the genomic library with the A3F and A7R adenylation-domain PCR primers.

In the near future, the isolation and characterization of the gene cluster responsible for the synthesis of DPO will enhance our ability and knowledge in the struggle against drug resistant tuberculosis. Ideally, many derivatives of DPO will be forged from this gene cluster template, which will substantially aid in combating this health crisis.

CHAPTER 3 REFERENCES

Ayuso-Sacido, G. & Genilloud, O. (2005). New PCR Primers for the sequencing of NRPS and PKS-I Systems in Actinomycetes: Detection and Distribution of These Biosynthetic Gene Sequences in Major Taxonomic Groups. *Microbial Ecology*, **49:** 10-24.

Bihlmaier, C., Welle, E., Hofmann, C., Welzel, K., Vente, A., Breitling, E., Müller, M., Glaser, S. & Bechthold, A. (2006). Biosynthetic gene cluster for the polyenoyltetramic acid alpha-lipomycin. *Antimicrobial Agents and Chemotherapy*, **50:** 2113-2121.

Braun, V., Pramanik, A., Gwinner, T., Köberle, M. & Bohn, E. (2009). Sideromycins: tools and antibiotics. *Biometals*, **22:**3–13.

Budzikiewicz, H. (2001). Siderophore-Antibiotic Conjugates Used as Trojan Horses Against *Pseudomonas aeruginosa*. *Current Topics in Medicinal Chemistry*, **1:** 73-82(10).

Demain, A.L. (1998). Induction of microbial secondary metabolism. *International microbiology: the official journal of the Spanish Society for Microbiology*, **1:** 259-264.

Gröning, J.A.D., Tischler, D., Kaschabek, S.R. & Schlömann, M. (2010). Optimization of a genome-walking method to suit GC-rich template DNA from biotechnological relevant Actinobacteria. *Journal of basic microbiology*, **50:** 499-502.

Harwood, C.S., Burchhardt, G., Herrmann, H. & Fuchs, G, (1999). Anaerobic metabolism of aromatic compounds via the benzoyl-CoA pathway. *FEMS Microbiology Reviews*, **22:** 439-458.

Le Roes, M. (2005). Selective isolation characterization and screening of actinomycetes for novel anti-tubercular antibiotics. PhD Thesis, Department of Molecular and Cell Biology, University of Cape Town. **Chapter 5**, 167-182.

Le Roes-Hill, M. & Meyers, P. R. (2009). *Streptomyces polyantibioticus* sp. nov., isolated from the banks of a river. *International Journal of Systematic and Evolutionary Microbiology*, **59:** 1302–1309.

Ligon, J., Ligon, J., Hill, S., Beck, J., Zirkle, R., Molnár, I., Zawodny, J., Money, S. & Schupp, T. (2002). Characterization of the biosynthetic gene cluster for the antifungal polyketide soraphen A from *Sorangium cellulosum* So ce26. *Gene*, **285:** 257-67.

Lombo, F., Velasco, A., Castro, A., de la Calle, F., Brana, A.F., Sanchez-Puelles, J.M., Mendez, C. & Salas, J.A. (2006). Deciphering the biosynthesis pathway of the antitumor thiocoraline from a marine actinomycete and its expression in two Streptomyces species, *ChemBioChem,* **7:** 366–376.

McQuade, T.J. Shallop, A.D., Sheoran, A., Delproposto, J. E., Tsodikov, O.V. & Garneau-Tsodikova, S. (2008). A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. *Analytical Biochemistry*, **386:** 244-250.

Miller, M.J, McKee, J.A., Minnick, A.A. & Dolence, E.K. (1991). The design, synthesis and study of siderophore-antibiotic conjugates Siderophore mediated drug transport. *BioMetals,* **4:** 62-69.

Mohamed, M.E., (2000). Biochemical and molecular characterization of phenylacetate-coenzyme A ligase, an enzyme catalyzing the first step in aerobic metabolism of phenylacetic acid in *Azoarcus evansii*. *Journal of Bacteriology*, **182:** 286-94.

Mohamed, M.E., Zaar, A., Ebenau-jehle, C., Fuchs, G. (2001). Reinvestigation of a New Type of Aerobic Benzoate Metabolism in the Proteobacterium Azoarcus evansii. *Journal of Bacteriology*, **183:** 1899-1908.

Mohamed, M.E., Ismail, W., Heider, J. & Fuchs, G. (2002). Aerobic metabolism of phenylacetic acids in *Azoarcus evansii*. *Archives of Microbiology*, **178:** 180-192.

Olano, C., Gómez, C., Pérez, M., Palomino, M., Pineda-Lucena, A., Carbajo, R. J., Braña, A. F., Méndez, C. & Salas, J. A. (2009). Deciphering biosynthesis of the RNA polymerase inhibitor streptolydigin and generation of glycosylated derivatives. *Chemistry & Biology*, **16:** 1031-1044.

Parker, J.D., Rabinovitch, P.S. & Burmer, G.C. (1991). Targeted gene walking polymerase chain reaction. *Nucleic Acids Research*, **19:** 3055-3060.

Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D.H. (2005). Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using Transductive Support Vector Machines (TSVM). *Nucleic Acids Research,* **33:** 5799-5808.

Rhee, S.K. & Fuchs, G. (1999). Phenylacetyl-CoA:acceptor oxidoreductase, a membrane-bound molybdenum-iron-sulfur enzyme involved in anaerobic metabolism of phenylalanine in the denitrifying bacterium *Thauera aromatica*. *European Journal of Biochemistry / FEBS*, **262:** 507-515.

Rost, R., Haas, S., Hammer, E., Herrmann, H. & Burchhardt, G. (2002). Molecular analysis of aerobic phenylacetate degradation in *Azoarcus evansii*. *Molecular Genetics and Genomics : MGG*, **267:** 656-663.

Roy, R.S., Gehring, A.M., Milne, J.C., Belshaw, P.J. & Walsh, C.T., (1999). Thiazole and oxazole peptides: biosynthesis and molecular machinery. *Natural Product Reports*, **16:** 249-263.

Schneider, S., Mohamed, M.E. & Fuchs, G. (1997). Anaerobic metabolism of L-phenylalanine via benzoyl-CoA in the denitrifying bacterium *Thauera aromatica*. *Archives of Microbiology*, **168:** 310-320.

Schupp, T., Toupet, C., Cluzel, B., Neff, S., Hill, S., Beck, J.J. & Ligon, J.M. (1995). A *Sorangium cellulosum* (myxobacterium) gene cluster for the biosynthesis of the macrolide antibiotic soraphen A: cloning, characterization, and homology to polyketide synthase genes from actinomycetes. *Journal of Bacteriology*, **177:** 3673-3679.

Tan, G. Gao, Y., Shi, M., Zhang, X., He, S., Chen, Z. & An, C. (2005). SiteFinding-PCR: a simple and efficient PCR method for chromosome walking. *Nucleic Acids Research*, **33:** e122.