IDENTIFICATION AND PRELIMINARY CHARACTERIZATION OF THE 2,5-DIPHENYLOXAZOLE BIOSYNTHETIC PATHWAY IN *STREPTOMYCES POLYANTIBIOTICUS* SPR^{T}.

by

Ian Kyle Kemp

A THESIS PRESENTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE DEPARTMENT OF MOLECULAR AND CELL BIOLOGY

FACULTY OF SCIENCE

UNIVERSITY OF CAPE TOWN

JULY 2015

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The following people and organisations are thanked:

- My supervisor, Dr Paul Meyers, for his assistance and support throughout the duration of this project.
- All the members of Lab 202/3 for their continuous support and advice, with special mention to Gareth Everest and Sarah Curtis. It has been a pleasure working with all of you.
- The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.
- The Medical Research Council and the University of Cape Town for the financial support that made this project possible.
- Dr P. Whitney Swain III for providing the plasmid vector pJN100 and *E. coli* ET12567/pUZ8002 that were used in this study.
- Dr Bohdan Ostash for providing the plasmid vector pOJ260 that was used in this study and for the advice on *Streptomyces* transformation.
- My parents, Ian and Anne-Marie, for their love, financial support and guidance, especially during tough times.
- To all academic staff, scientific officers and departmental assistants for all their help and assistance.

IDENTIFICATION AND PRELIMINARY CHARACTERIZATION OF THE 2,5-DIPHENYLOXAZOLE BIOSYNTHETIC PATHWAY IN *STREPTOMYCES POLYANTIBIOTICUS* SPR^T.

by

Ian Kyle Kemp

Department of Molecular and Cell Biology, University of Cape Town, Private Bag, Rondebosch, 7701, South Africa

ABSTRACT

An antibacterial compound produced by the actinomycete, *Streptomyces polyantibioticus* SPR^T, exhibited antibiosis against *Mycobacterium tuberculosis* H37Rv^T (the causative agent of tuberculosis), which prompted interest in its biosynthesis. The antibacterial compound was isolated in a previous study and its structure was determined by X-ray crystallography and nuclear magnetic resonance (NMR) to be 2,5-diphenyloxazole (DPO). Based on the structure of DPO, a biosynthetic scheme for the synthesis of this molecule was proposed, whereby a non-ribosomal peptide synthetase (NRPS) condenses a molecule of benzoic acid with 3-hydroxyphenylalanine. The dipeptide is converted to a diphenyloxazole derivative by heterocyclization and a final decarboxylation step leads to DPO. To determine whether the hypothesis pertaining to the DPO biosynthetic pathway was correct, initial efforts were made to identify the genes coding for benzoic acid synthesis and the DPO NRPS in the *S. polyantibioticus* SPR^T genome using PCR amplification, Southern hybridization and sequencing. This led to the identification of 12 unique adenylation (A) domains (of which one was specific for phenylalanine) and a gene, *paaK*, encoding a phenylacetate CoA-ligase (PA-CoA), putatively involved in benzoic acid biosynthesis. However, no further sequence information could be obtained for the genes encoding the phenylalanine-specific A domain or PA-CoA and similar attempts to identify other NRPS-associated domains, as well as genes involved in benzoic acid synthesis, proved unsuccessful. In light of these difficulties, the *S. polyantibioticus*

SPR^T genome was sequenced and a gene cluster was identified as being responsible for the biosynthesis of DPO using a genome mining approach. However, contrary to the hypothesis that a linear NRPS system for DPO biosynthesis would be identified, the gene cluster exhibited a nonlinear arrangement. The core domains are arranged as A-PCP-C (instead of C-A-PCP) and there is also a stand-alone heterocyclization (Cy) domain, a stand-alone thioesterase (TE) domain and an acyl-CoA synthetase putatively involved in activating benzoic acid. Furthermore, there are two NRPS domains in the gene cluster that are believed to be inactive. A possible biosynthetic pathway for benzoyl-CoA production, encoded by a separate gene cluster, was identified based on the genome analysis of *S. polyantibioticus* SPR^T. In order to confirm the involvement of the identified genes in DPO biosynthesis, an intergeneric conjugation protocol was developed for the introduction of plasmid DNA into *S. polyantibioticus* SPR^T and subsequent gene disruption experiments. The putative DPO biosynthetic genes were insertionally activated via homologous recombination and the method for isolating DPO was carried out on each of the mutant strains, after which the extracts were assayed for activity against *Mycobacterium aurum* A+ using TLC-bioautography analysis. The absence of activity against *M. aurum* A+ in the extracts from mutant strains *S. polyantibioticus* ΔA99, *S. polyantibioticus* ΔCYC and *S. polyantibioticus* ΔACY suggested the involvement of the A domain encoded by gene SPR_53060, the putative Cy domain encoded by gene SPR_53040 and the acyl-CoA synthetase encoded by gene SPR_52860 in the biosynthesis of DPO. However, attempts to identify the genes responsible for benzoic acid biosynthesis proved unsuccessful, as gene disruption did not abolish DPO activity in the *S. polyantibioticus* ΔLAC, *S. polyantibioticus* ΔPAAK and *S. polyantibioticus* ΔCIN mutant strains encoding the putative D-lactate dehydrogenase encoded by gene SPR_60250, the PA-CoA ligase (*paaK*) encoded by gene SPR_46390 and the cinnamate-CoA ligase encoded by gene SPR_60150, respectively. Based on the genome annotation analysis and gene disruption studies, a model for DPO biosynthesis is proposed. At this stage, the model cannot account for the source of benzoic acid, as *in vivo* gene disruption experiments disproved both of the hypotheses on how benzoic acid is synthesized in *S. polyantibioticus* SPR^T. However, alternative hypotheses regarding the mechanism of benzoic acid biosynthesis in *S. polyantibioticus* SPR^T are proposed and are suggested as the place to start in future studies to elucidate the production of this unusual starter unit in DPO biosynthesis. Furthermore, the identification of the gene cluster responsible for DPO biosynthesis has

laid the foundation for future combinatorial biosynthetic studies to create derivatives of DPO that might be used in the treatment of drug resistant tuberculosis. Lastly, the *S. polyantibioticus* SPR^T genome sequence could be explored for the identification of antibiotic gene clusters for other potential antitubercular antibiotics that this organism produces.

ABBREVIATIONS

α	alpha
β	beta
Δ	delta
γ	gamma
λ	lambda
Ω	ohm (s)
Φ	phi
µg	microgram(s)
µl	microlitre(s)
µF	microFarad (s)
µM	micromolar
°C	degrees Celsius
%	percentage
A	adenylation domain
apr^R	apramycin resistant
AIDS	Acquired Immunodeficiency Syndrome
ACP	acyl carrier protein
ArCP	aryl carrier protein
ATP	adenosine triphosphate
AMP	adenosine monophosphate
AMT	aminotransferase
BC	Before Christ
bp	base pair(s)
BLAST	Basic Local Alignment Search Tool
C	condensation domain
Cy	heterocyclization domain
CDC	Centers for Disease Control and Prevention
CDD	Conserved Domain Database
clt	*cis*-acting locus of transfer
COGs	clusters of orthologous groups
COM	communication-mediating
CoA	Coenzyme A
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
ssDNA	single-stranded DNA

dH_20	distilled water
DPO	2,5-diphenyloxazole
dNTP	deoxyribonucleoside triphosphates (dATP, dCTP, dTTP and dGTP)
et al.	and others (*et alii*)
E	epimerization domain
FAS	fatty acid synthase
FMN	flavin mononucleotide
g	gram(s)
gDNA	genomic DNA
GrsA	gramicidin synthetase
HAL	histidine ammonia-lyase
HIV	Human immunodeficiency virus
HM	Hacène's medium
HMM	Hidden Markov Model
HPLC	High-performance liquid chromatography
h	hour(s)
Inc.	Incorporated
IncP	Incompatibility group P
ISP	International *Streptomyces* Project
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilobase(s)
kg	kilogram(s)
kDa	kiloDalton(s)
kV	kiloVolt(s)
l	litre
LA	Luria-Bertani agar
LB	Luria-Bertani broth
M	molar
MDR	multidrug-resistant
MEGA	Molecular Evolutionary Genetics Analysis
ML	maximum likelihood
MP	maximum parsimony
MRSA	methicillin-resistant *Staphylococcus aureus*
MS	mannitol soya
MT	methyltransferase
Mb	mega base(s)

mg	milligram(s)
min	minute(s)
ml	millilitre(s)
mM	millimolar
mm	millimetre(s)
mRNA	messenger RNA
ms	millisecond (s)
mA	milliAmp(s)
NADP	Nicotinamide adenine dinucleotide phosphate
NGS	Next-generation sequencing
NJ	neighbour joining
NRPS	Non-ribosomal peptide synthetase
NRP	Non-ribosomally synthesized peptide
ng	nanogram(s)
nm	nanometre(s)
OD	optical density
ORF	open reading frame
oriT	origin of transfer
Ox	oxidation domain
O/N	overnight
PA	phenylacetate
PA-CoA	phenylacetate-CoA ligase
PAL	phenylalanine ammonia-lyase
PCP	peptidyl carrier protein
PCR	polymerase chain reaction
PDR	pan drug-resistant
PEG	polyethylene glycol
PKS	Polyketide Synthase
PPi	inorganic pyrophosphate
PPTase	phosphopantetheinyltransferase
PP	phosphopantetheine
p	plasmid
RAST	Rapid Annotation using Subsystem Technology
RE	reduction domain
R*f*	retention factor
REase	restriction endonuclease
RSA	Republic of South Africa
R-M	restriction-modification
RNA	ribonucleic acid
RNase	ribonuclease

rpm	revolutions per minute
s	second(s)
SEM	standard error of the mean
SAM	*S*-adenosylmethionine
SDS	sodium dodecyl sulphate
SSC	standard saline citrate
spp.	species (plural)
^{T}	type strain
TB	Tuberculosis
TE	thioesterase domain
TLC	thin layer chromatography
TSB	tryptic soy broth
TSVM	transductive support vector machine
U	unit(s)
USA	United States of America
UK	United Kingdom
UV	ultraviolet
VRSA	vancomycin-resistant *Staphylococcus aureus*
v/v	volume per volume
V	volt(s)
WHO	World Health Organisation
XDR	extensively drug-resistant
YEME	yeast extract-malt extract medium (ISP Medium No.2)
ZMW	zero-mode waveguide

CHAPTER 1

GENERAL INTRODUCTION

CHAPTER 1

GENERAL INTRODUCTION

CHAPTER 1

GENERAL INTRODUCTION

The existence of antimicrobial agents was first described in 1877 when Louis Pasteur and Robert Koch observed that one type of bacterium could prevent the growth of another. The term "antibiosis' was introduced by the French bacteriologist, Vuillemin, as a descriptive name for this phenomenon. In 1942, Selman Waksman, an American microbiologist, devised the term antibiotic to describe any substance produced by a microorganism that is antagonistic to the growth of other microorganisms in high dilution (Waksman, 1947). The value of antibiotics to medicine was soon realized and it lead to the heightened interest in the isolation and discovery of these novel compounds. Indeed, the rapid development of antimicrobial agents over the past century has greatly improved the treatment of infection and disease. However, the increased usage of these drugs in the 1950s sparked the arms race between humanity and microbial pathogens, as increased examples of recalcitrant infections due to the inevitable rise of drug-resistant pathogens became apparent (Davies, 2012; Davies, 2007).

1.1 ANTIBIOTIC RESISTANCE

The use of antibiotics as a means of treating infectious disease in humans and food-producing animals has revolutionized medicine and become widespread since the introduction of penicillin in the 1940s. Regrettably, the extensive misuse of these drugs in both the clinical and agricultural setting has led to the rapid emergence of drug-resistant strains of bacteria and other microbes. This resistance has led to a decrease in the effectiveness of currently available antimicrobial drugs and the problem is further compounded by the emergence of multidrug-resistant microbial strains.

Microorganisms that have developed resistance to several different antibiotics are referred to as multidrug-resistant (MDR) strains.

It was reported in 2004 that more than 70 % of pathogenic bacteria were predicted to be resistant to at least one of the currently available antibiotics. Although we have witnessed a steady increase in resistance in almost every pathogen to most of the currently available antibiotics, not all of these antibacterial agents display the same rate of resistance development. Indeed, single target drugs such as rifampicin are more susceptible to the development of resistance than drugs that inactivate several targets irreversibly, such as penicillin (Demain & Sanchez, 2009; Spratt, 1994).

Global healthcare systems are also encountering extensively drug-resistant (XDR) organisms on a regular basis and these microorganisms are resistant to all antibiotics except colistin, a highly toxic agent, which was abandoned in the 1960s due to its questionable efficacy. In the most severe case, certain microorganisms are pan drug-resistant (PDR), meaning that they are resistant to all available antibiotics (Udwadia & Vendoti, 2013; Udwadia, 2012).

It has been reported that at least 2 million people become infected with antibiotic-resistant bacteria each year in the United States, with at least 23 000 people dying as a direct result of these infections. It is also believed that most deaths related to antibiotic resistance happen in healthcare facilities such as nursing homes and hospitals. One of the most common causes of healthcare-associated infections is methicillin-resistant *Staphylococcus aureus* (MRSA), claiming the lives of more than 11 000 people in the United States in 2011. This MDR, Gram-positive pathogen is resistant to all penicillins and cephalosporins (i.e. all β-lactam antibiotics), as well as often being resistant to clindamycin and quinolones. *S. aureus* bacteria produce a biofilm which protects them from the environment. These biofilms can grow on wounds, scar tissue and medical implants or devices (Dantas & Sommer, 2014). It is estimated that more than 70 % of the bacterial species that produce biofilms are likely to be resistant to at least one of the drugs commonly used in anti-infectious therapy in hospitals (Demain & Sanchez, 2009). Once described as only a hospital-inhabiting or nosocomial infection, the rate of MRSA infections has increased rapidly among the general population in the past

decade, which further exacerbates the urgent need for new therapeutic agents (Centers for Disease Control and Prevention, 2013).

Vancomycin has long been the treatment of choice for MRSA infections, however, new treatments were sought after the Centers for Disease Control and Prevention (CDC) reported the first case of *S. aureus* resistant to both methicillin and vancomycin in 2002. Since then, the number of reported cases of vancomycin-resistant *S. aureus* (VRSA) has remained relatively low and thus the epidemiology and risk factors associated with VRSA are not completely understood. Further research is required in order to understand these aspects, as well as the implications pertaining to clinical and infection control and optimal treatment (Howden *et al.,* 2010; Applebaum, 2006).

Pathogens that are less prevalent than MRSA, but that pose a threat of infection that is genuinely untreatable are those such as the carbapenem-resistant *Enterobacteriaceae*. Within this large family of Gram-negative bacteria, strains of *Acinetobacter baumannii, Escherichia coli, Klebsiella pneumonia* and *Pseudomonas aeruginosa* are currently tormenting global healthcare systems due to their resistance to all β-lactam antibiotics. This resistance is conferred by the ability of their outer membrane to prevent the entry of the antibiotics, in addition to being able to expel, via efflux pumps, the remainder that does successfully enter the cell (Fischbach & Walsh, 2009).

Furthermore, a report released by the CDC in 2013 identified the pathogens *Clostridium difficile* and *Neisseria gonorrhoeae* as antibiotic-resistant bacteria that require urgent public health attention in order to identify infections and limit transmission (CDC, 2013). The report also highlighted the serious threat of MDR and XDR strains of *Mycobacterium tuberculosis,* the causative agent of tuberculosis (TB).

Tuberculosis is a contagious airborne disease and is second only to human immunodeficiency virus (HIV)/AIDS as the leading killer worldwide due to a single infectious agent (Liu *et al.,* 2012). Approximately 9 million people contracted TB in 2013, of which 1.5 million people died from the disease. About 480 000 people developed MDR-TB globally in 2013, an increase of almost 7 % from 2012, of which more than half the cases were in developing countries such as India, China and the

Russian Federation. It was also estimated that 9 % of the MDR-TB cases had XDR-TB (World Health Organisation, 2014)

In the majority of cases, TB is treatable and curable with the available first-line anti-TB drugs (rifampicin, isoniazid and pyrazinamide). However, disease caused by *M. tuberculosis* strains that are resistant to one or more of these standard drugs is far more challenging to treat, largely due to the higher cost, adverse side-effects and longer treatment regimes of the available second-line drugs. MDR-TB strains are defined as being resistant to at least two of the first-line anti-tubercular drugs, isoniazid and rifampicin. XDR-TB strains are defined as being resistant to both these drugs as well as to any fluoroquinolone and at least one of the injectable second-line drugs (i.e. kanamycin, amikacin or capreomycin). Extensively drug-resistant TB has been identified in 92 countries and in all regions of the world (WHO, 2014; CDC, 2013; Rivers & Mancera, 2008; Petrini & Hoffner, 1999).

The primary cause of MDR-TB and XDR-TB is incomplete or incorrect treatment, which is mainly attributed to patients not completing their full course of antibiotic therapy, but is also due to the usage of inappropriate drug combinations, poor treatment compliance, using single drugs for ordinary TB, clinics running out of drug stocks and the use of poor quality medicines (WHO, 2014; Cape Gateway, 2006).

The World Health Organisation (WHO) has predicted that between 2000 and 2020, almost 1 billion people globally will contract tuberculosis and that the disease will claim the lives of approximately 35 million people. Twenty-two high-burden countries were identified, which account for 81 % of all estimated TB cases that occur globally. South Africa has the highest incidence of TB in the world per capita population and this problem is compounded by the high incidence of HIV co-infection, as HIV-positive individuals infected with *M. tuberculosis* have a dramatically increased chance of developing TB compared to TB-infected, HIV-negative individuals (Churchyard *et al.,* 2014; Bloom and Murray, 1992). South Africa also has the largest number of HIV-associated TB cases and the second-largest number of MDR-TB cases, after India. Notable progress has been achieved in reducing TB prevalence and deaths and improving treatment outcomes for new TB cases in recent years, but the burden still remains enormous, especially considering the drastic rise in the numbers of MDR and

XDR strains. New TB drugs and vaccines are urgently required to further accelerate progress towards improved TB control in South Africa and other countries (Churchyard *et al.,* 2014).

Consequently, there is a global healthcare emergency, as new antibiotic discovery and subsequent treatment options have been outpaced by the emergence of drug-resistant microorganisms. In order to control the global TB epidemic, there is an urgent need for new and improved TB drugs. These new drugs should, ideally, be effective against both MDR and XDR *M. tuberculosis* strains, be able to shorten the duration of treatment from the current six month regime and simplify treatment by reducing the number of pills that need to be taken each day, as well as reducing the dosage frequency to once a week. In addition, new TB drugs should be able to be administered simultaneously with HIV drugs and be effective against latent TB (Koul *et al*., 2011; Lamichhane, 2011; Barry *et al*., 2009; Young *et al.*, 2009; Rivers and Mancera, 2008).

In light of this, taxonomy-guided bacterial bioprospecting needs to continue to play an important role in identifying new microbial natural products, such as bacterial secondary metabolites that have already served the pharmaceutical industry well as TB drugs. It has been estimated that less than 10 % of the world's biodiversity has been tested for biological activity, so there is a high likelihood of isolating new antibiotic molecules from bacteria (Ashforth *et al.,* 2010; Harvey, 2000).

In addition, the relentless evolution of antibiotic-resistant strains of microbial pathogens erodes the utility of the classical antibiotics and necessitates the perpetual need for discovery of new antibiotics (Ashforth *et al.,* 2010). Consequently, numerous research groups are now focussing on identifying novel antibiotics in new species of *Streptomyces* and other genera of *Actinobacteria* due to the fact that they are renowned for their ability to produce antibiotics and other bioactive natural products with a wide range of applications in medicine and agriculture. Examples of these compounds include antibacterials such as erythromycin A, vancomycin and daptomycin, antifungals such as amphotericin B, immunosuppressants such as FK-506, anticancer agents such as doxorubicin and epoxomicin, anthelmintics such as avermectin, insecticides such as spinosyn A and herbicides such as phosphinothricin (Figure 1.1) (Challis, 2014).

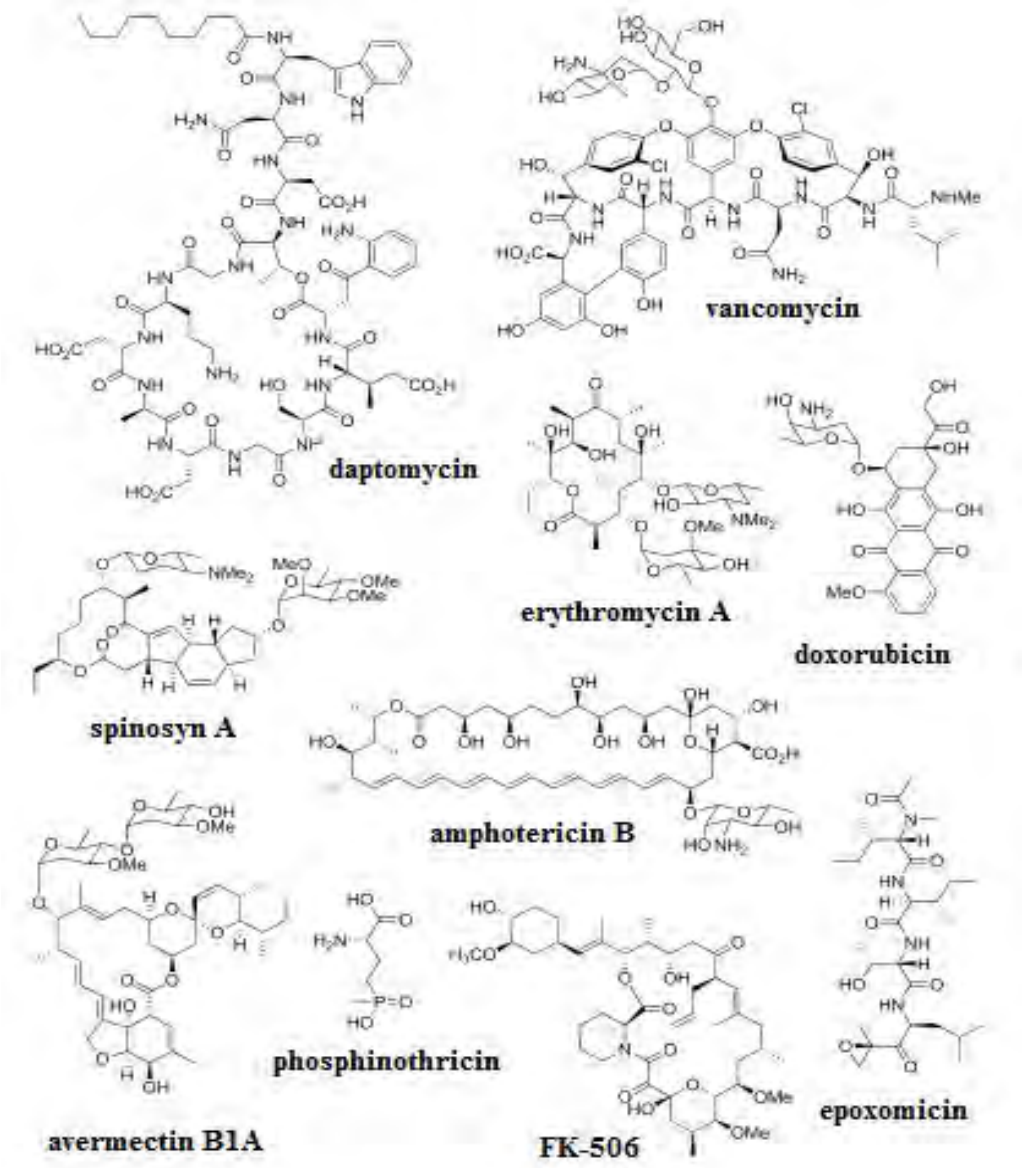

Figure 1.1 Examples of secondary metabolites used in medicine and agriculture produced by *Actinobacteria* (Challis, 2014).

1.2 THE PHYLUM *ACTINOBACTERIA*

The phylum *Actinobacteria* represents one of the largest taxonomic units within the domain *Bacteria* and is predicted to be the most abundant source of small molecule diversity on earth (Miao & Davies, 2010). The phylum consists of Gram-positive bacteria with typically elevated guanosine (G) and cytosine (C) content in their chromosomal DNA, ranging from 50 % in certain corynebacteria to more than 70 % in *Streptomyces* and *Frankia* species (Ventura *et al.,* 2007).

Actinobacteria are inhabitants of soil, the rhizosphere, marine and extreme arid environments and exhibit a wide range of morphologies, including coccoid, rod-shaped and hyphal forms. The filamentous actinobacteria are often referred to as actinomycetes and produce branching hyphae which form a mycelium. In many genera of actinomycetes, some of the vegetative hyphae differentiate into arthrospores. Some genera of actinomycetes exhibit fragmentation in liquid and/or plate culture (e.g. *Amycolatopsis*), a phenomenon in which the vegetative hyphae break down into rod-shaped or coccoid elements. Additionally, actinobacteria are able to produce a variety of extracellular enzymes and secondary metabolites as a result of the regulated expression of gene clusters (MacNeil *et al.,* 1992; Chater, 1990). Interest in the phylum intensified after the discovery of streptomycin in the laboratory of Selman Waksman in 1943 and the subsequent observation that many of the secondary metabolites are indeed potent antibiotics. In particular, *Streptomyces* species have been exploited extensively by the pharmaceutical industry as a primary source of natural products for use as therapeutic agents, but bacteria belonging to the suborders *Micromonosporineae* (e.g. *Micromonospora*), *Pseudonocardineae* (e.g. *Amycolatopsis*) and *Streptosporangineae* (e.g. *Planobispora*) have also been described as abundant producers of novel antibacterial antibiotics (Ashforth *et al*., 2010; Miao & Davies, 2010; Ventura et al., 2007).

Actinobacterial genomes, particularly those of the actinomycetes (order *Actinomycetales*), are larger than most other bacteria, with sizes ranging from 0.93 Mb (*Tropheryma whipplei*) to 11.9 Mb (‘*Streptomyces bingchenggensis*’) and therefore possess a high capacity for secondary metabolite production (Verma *et al*., 2013).

Among the actinomycetes, such as *Streptomyces*, it is estimated that more than 60 % of the secondary metabolites produced are synthesized by enzymatic systems known as non-ribosomal peptide synthetases (NRPSs), polyketide synthases (PKSs) or mixed NRPS-PKS pathways (Baltz, 2014). Moreover, it is widely accepted that species that possess a large number of NRPS and/or PKS genes produce more secondary metabolites. For example, the sequencing of the first antibiotic-producing actinomycete genomes, those of *Streptomyces coelicolor* strain A3(2) (8.6 Mb) and *Streptomyces avermitilis* strain MA-4680^T (9.02 Mb), revealed the potential of both organisms to produce a large number of secondary metabolites, most of which were not synthesized under standard growth conditions (Baltz, 2008).

The genome sequence of *S. avermitilis* MA-4680^T revealed 24 PKS and NRPS clusters for previously unidentified secondary metabolites and in the case of *S. coelicolor* A3(2), a genome mining approach identified the capability of the organism to produce twice as many secondary metabolites as was originally thought (Nett *et al.,* 2009; Bentley *et al.,* 2002). Similarly, annotation of the genomes of *S. avermitilis* MA-4680^T and *Streptomyces griseus* strain NBRC 13350, revealed that 60 % and 75 %, respectively, of the biosynthetic gene clusters in these organisms encoded unidentified secondary metabolites, further suggesting that natural product biosynthetic capacity has been immensely underestimated (Challis, 2014; Nett *et al.,* 2009; Walsh, 2004; Challis & Ravel, 2000).

The sequenced actinobacterial genomes (currently numbering 4059) (GenBank, 2015; Land *et al.,* 2015) are an invaluable resource which has revealed that the more bacterial genomes that are sequenced, the more new gene families are discovered, thus uncovering new genetic diversity and thereby new biosynthetic capabilities (Wu *et al*., 2009). This genetic diversity is directly afforded by NRPS multifunctional enzymatic systems that are able to manufacture structurally diverse natural products with a remarkable variety of useful/significant pharmacological activities (Ashforth *et al.,* 2010; Wu *et al.,* 2009).

1.3 NATURAL PRODUCT DISCOVERY

Natural products, which can be produced from primary or secondary metabolism by living organisms such as bacteria, plants, mammals, marine invertebrates, insects and fungi, are defined as low molecular weight, organic molecules (Davies & Ryan, 2012). The bacterial metagenome contributes to the production of primary metabolites and conversion of small molecules into secondary metabolites, also called "specialized metabolites". Two of the important classes of secondary metabolites classified as natural products are the polyketides and non-ribosomal peptides. Other structural classes include alkaloids, terpenoids, aminoglycosides and shikimate-derived molecules (Baltz, 2014; Davies & Ryan, 2012).

For over half a century, natural products have played an important role as targets of study for analytical and synthetic chemists, as chemical tools for probing biological systems to aid in discovering the roles of individual biomolecules, but most importantly, they have acted as a treasure trove of compounds used to treat infectious diseases in humans, animals and crops (Davies, 2011; Fischbach & Walsh, 2009). Indeed, it has been estimated that approximately 40 % of all medicines in clinical use are either natural products or their semi-synthetic derivatives (John, 2009). This may not come as a surprise, considering that herbal medicine formed the cornerstone of sophisticated traditional medicine practices, with the earliest records dating back to 2600 BC in Mesopotamia, where plant-derived substances such as the oils of *Cupressus semevirens* (cypress), *Cedrus* species (cedar) and *Glycyrrhiza glabra* (licorice) where used to treat ailments such as coughing, parasitic infections and inflammation (Cragg & Newman, 2013; John, 2009). It has also been documented that the Greeks and Romans contributed significantly to the development of the use of herbal medicine in the ancient Western world (Cragg & Newman, 2013).

Furthermore, the serendipitous discovery of morphine in 1805 was the first pharmacologically active pure compound isolated from a plant. This organic alkaloid was derived from the resinous gum secreted by *Papaver somniferum* (the opium poppy) and sparked the study of alkaloid chemistry, which inadvertently accelerated the emergence of the modern pharmaceutical industry. Similar alkaloids from organic

substances were isolated soon after, such as strychnine in 1817, caffeine in 1820 and nicotine in 1828 (Li & Vederas, 2009).

In contrast, the microbial-derived drug era began in the 1920s with the discovery of the antibiotic, penicillin, by Sir Alexander Fleming. Penicillin was later produced industrially as a powder and used as a potent antibacterial compound during World War II. The observation of its broad therapeutic use started what became known as "The Golden Age of Antibiotics" and prompted the massive screening of microorganisms for new antibiotics and related novel bioactive molecules (Cragg & Newman, 2013).

By 1990, 80 % of therapeutic medicines were natural products or analogues inspired by them (Li & Vederas, 2009). Microorganisms remain a prolific source of structurally and chemically diverse bioactive metabolites and these include antibacterial agents, such as the penicillins (from *Penicillium* species), cephalosporins (from *Cephalosporium acremonium*), aminoglycosides, tetracyclines and other polyketides of many structural types (from members of the order *Actinomycetales*); immunosuppressive agents, such as the cyclosporins (from *Trichoderma* and *Tolypocladium* species) and rapamycin (from *Streptomyces* species); cholesterol-lowering agents, such as mevastatin (compactin; from *Penicillium* species) and lovastatin (from *Aspergillus* species); and anthelmintics and antiparasitic drugs, such as the ivermectins (from *Streptomyces* species) (Cragg & Newman, 2013; Demain & Sanchez, 2009; Li & Vederas, 2009).

Historically, the classical screening and bioassay-guided isolation techniques were very successful in providing a route to the discovery of novel natural products for the development of new therapeutic agents (Subramani & Aalbersberg, 2013). Over the past two decades, however, natural product discovery efforts, particularly microbially-produced products, have waned considerably due to the high rediscovery rate of known compounds. Most of the common core structures from which today's antibiotics are derived were introduced between the mid-1930s and the early 1960s (Figure 1.2) (Fischbach & Walsh, 2009). Consequently, pharmaceutical firms have preferred to pursue alternative options in their search for new therapeutic agents, such as chemically tailored derivatives of the classical scaffolds and modern techniques such as *in silico*

screening, combinatorial biosynthesis (Cane *et al.,* 1998) and combinatorial biocatalysis (Watve *et al.,* 2001; Michels *et al.,* 1998).

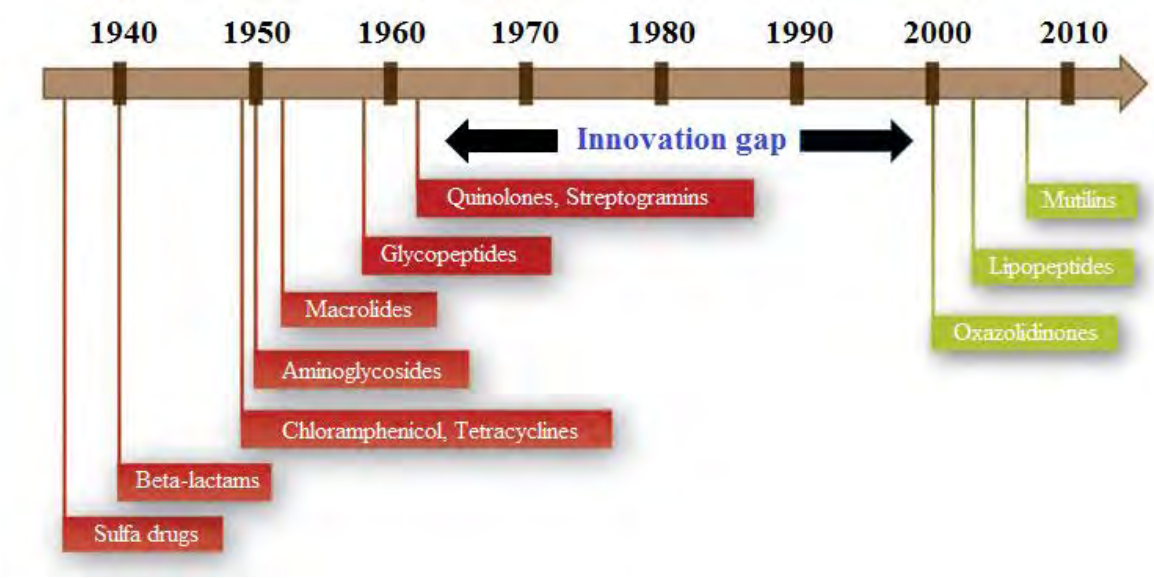

Figure 1.2 No new major classes of antibiotics were introduced between 1962 and 2000 (referred to as an innovation gap) (Fischbach & Walsh, 2009).

Despite this lack of interest, natural products continue to provide greater structural diversity than standard combinatorial chemistry and therefore offer far more opportunities for finding novel chemical scaffolds that are active against a wide range of targets. In addition, the majority of the world's biodiversity remains unexplored and it has become increasingly apparent that approximately only 1 % of the microbial diversity has been cultured and studied experimentally (Van Lanen & Shen, 2006; Harvey, 2000). Attempts at exploring the uncultured or allegedly unculturable bacteria using culture-independent methods such as metagenomics have been investigated and it has become evident that these bacteria may in fact be culturable, thus providing a potentially large source of new bioactive compounds (Watve *et al.*, 2001; Handelsman *et al*., 1998; Hugenholtz *et al*., 1998; Seow *et al.,* 1997).

Nevertheless, it was reported in 2005 that the number of discovered natural products exceeds 1 million and this is mainly due to improvements in screening methods, as well as separation and isolation techniques. Among these compounds, it is estimated that 50-60 % are produced by plants and 5 % are produced by microbes. Furthermore, it is approximated that from the pool of biologically active compounds obtained from microbes, 45 % are produced by actinomycetes, 38 % by fungi and 17 % by unicellular bacteria (Bérdy, 2005).

There is still a gigantic proportion of undiscovered metabolites produced by the actinomycetes and predictive modelling by Watve *et al.* (2001) suggested that more than 150 000 bioactive metabolites are still waiting to be discovered from the members of the genus *Streptomyces* alone. In order to gain access to the biodiversity afforded in these and other actinobacterial suborders, it is imperative to continue efforts to isolate novel actinobacteria from underexplored, diverse natural habitats.

There has been an increasing number of novel secondary metabolites isolated from actinobacteria, specifically from the genus *Streptomyces*, over the past few decades, and one such class is the heterocyclic-ring containing metabolites such as the oxazoles and thiazoles. Several of these metabolites have been shown to exhibit extremely useful biological properties, such as antimycobacterial activity (Walsh & Fischbach, 2010).

1.3.1 Oxazole-containing natural products

Oxazoles are aromatic 5-membered heterocycles which contain both nitrogen and oxygen (Yeh, 2004). They are produced in nature by dramatic chemical modifications performed on elongating peptide chains, most often catalysed by NRPS assembly lines (Walsh *et al*., 2001). Oxazole biosynthesis is achieved via the cyclodehydration of serine or threonine to yield a dihydroheteroaromatic oxazoline, which is then subjected to a two-electron oxidation step and in turn yields the oxazole core structure (Figure 1.3). Reduction of the carbon-nitrogen bond in oxazoline creates an oxazolidine ring (Yeh, 2004; Roy *et al.,* 1999, Milne *et al.,* 1998). All three of these oxidation states can be found in natural products and if the starting residue is cysteine, the corresponding process would yield thiazolines, thiazoles and thiazolidines (Roy *et al.,* 1999; Milne *et al.,* 1998).

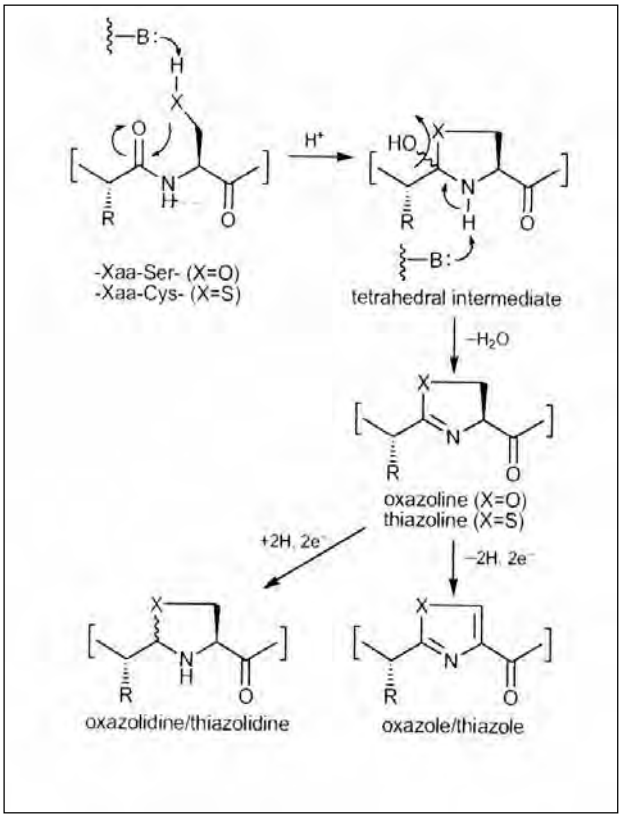

Figure 1.3 Schematic representation of the biosynthetic process involved in oxazole and thiazole production (Roy *et al.,* 1999).

It was initially considered that naturally occurring oxazoles were rare, until heightened research in the 1980s proved their ubiquity in nature (Yeh, 2004). The anticonvulsant alkaloid, pimprinine, was one of the earliest oxazole-containing compounds to be discovered, which was isolated from '*Streptomyces pimprina*' in 1960 (Bhate *et al.,* 1960). Two related alkaloids, pimprinethine and pimprinaphine, were isolated almost twenty years later from *Streptomyces cinnamoneus* and *Streptomyces olivoreticuli*, respectively (Yoshioka *et al.,* 1981). Furthermore, in 1966, the macrocyclic streptogramin antibiotic accommodating a 2,4-disubstituted oxazole ring, virginiamycin M2, was isolated from *Streptomyces virginiae* (Kondo *et al*., 1989). The 2,4-disubstituted oxazole ring is a common feature in several cyclic peptide natural

products including the antileukemic metabolites, orbiculamide and keramide E, which were both isolated from the marine sponge, *Theonella* sp., in 1991 (Fusetani *et al.,* 1991) and 1995 (Kobayashi *et al.,* 1995), respectively. Additionally, in 1998, Rastogi and colleagues reported the antimycobacterial activity of the oxazole alkaloid, texaline, isolated from the plants *Amyris texana* and *Amyris elemifera* (Rastogi *et al.,* 1998).

The stimulated interest in oxazoles (and thiazoles) can be attributed to their diverse biological properties and their ability to interact with a wide variety of intracellular bacterial targets, specifically proteins, DNA and RNA (Walsh, 2004; Milne *et al.,* 1998). Additionally, the oxazoline and thiazoline rings are important elements of bioactive natural products and well known antimicrobial products. The use of these rings as building blocks in pharmaceutical drug discovery is continually increasing and one example is the phenyloxazolines which are able to inhibit the growth of certain Gram-negative bacteria by impeding the production of lipid A, a molecule located in the outer membrane of most Gram-negative bacteria (Padmavathi *et al.,* 2009; Jackman *et al.,* 2000).

The biosynthesis of several oxazoles, thiazoles and their derivatives has been reported to be catalysed by NRPS and NRPS/PKS hybrid enzyme complexes. Examples include tallysomycin produced by *Streptoalloteichus hindustanus* ATCC 31158 (Tao *et al.,* 2007), virginiamycin produced by *S. virginiae* (Pulsawat *et al.,* 2007) and zorbamycin produced by *Streptomyces flavoviridis* ATCC 21892 (Wang *et al.,* 2007). More recently, novel antitrypanosomal antibiotics, the spoxazomicins, containing an oxazole moiety within their structure, were isolated from the endophytic actinomycete, *Streptosporangium oxazolinicum.* However, the biosynthetic machinery involved in the synthesis of the spoxazomicins is yet to be elucidated (Inahashi *et al.,* 2011).

1.4 NON-RIBOSOMAL PEPTIDE SYNTHETASES (NRPSs)

Non-ribosomally synthesized peptides (NRPs) represent a class of microbial natural products that exhibit unparalleled diversity in both structure and biological activity. These characteristics can be attributed to the incorporation of many unusual, nonproteinogenic residues and modifications (Konz & Marahiel, 1999; Marahiel, 1997). As the name implies, NRPs are normally synthesized from amino acids, but over 400 monomers are known to be used as substrates, which include non-proteinogenic amino acids such as ornithine and hydroxyphenylglycine, D-configured and N-methylated amino acids, as well as a variety of hydroxyl-amino acids, which can be further modified by acylation, glycosylation and heterocyclic ring formation mediated by associated tailoring enzymes (Bloudoff *et al.,* 2013; Konz & Marahiel, 1999). Despite the structural diversity, however, the majority of NRPs share a common biosynthetic origin, which has been linked to an enzymatic system known as the "thiotemplate multienzymic mechanism" (Marahiel, 1997; Kleinkauf & von Döhren, 1990). This enzymatic system is more commonly referred to as a non-ribosomal peptide synthetase (NRPS) and these large, multifunctional enzymes range in size from 100-2000 kDa and can catalyze up to several dozen reactions in a co-ordinated, assembly-line manner (Mootz *et al.,* 2002; Cane & Walsh, 1999; Marahiel, 1997).

Non-ribosomal peptide biosynthesis is widespread in a variety of bacterial and fungal species and plays a significant role in the production of several biologically relevant products (Schwarzer *et al*., 2003). These include antibiotics such as bacitracin A, surfactin, tyrocidine A, gramicidin S and the penicillin precursor, isopenicillin N; immunosuppressive agents such as cyclosporin (used in the aftercare of organ transplant patients); bleomycin A2 and epothilone (used in cancer therapy due to their cytostatic activity); and siderophores such as enterobactin and myxochelin A (Figure 1.4) (Schwarzer *et al.,* 2003: Marahiel, 1997). Additionally, the genes encoding bacterial NRPSs are organised into operons, which are usually 6-45 kilobases in length and encode protein templates that are often much smaller than their fungal counterparts (Marahiel, 1997).

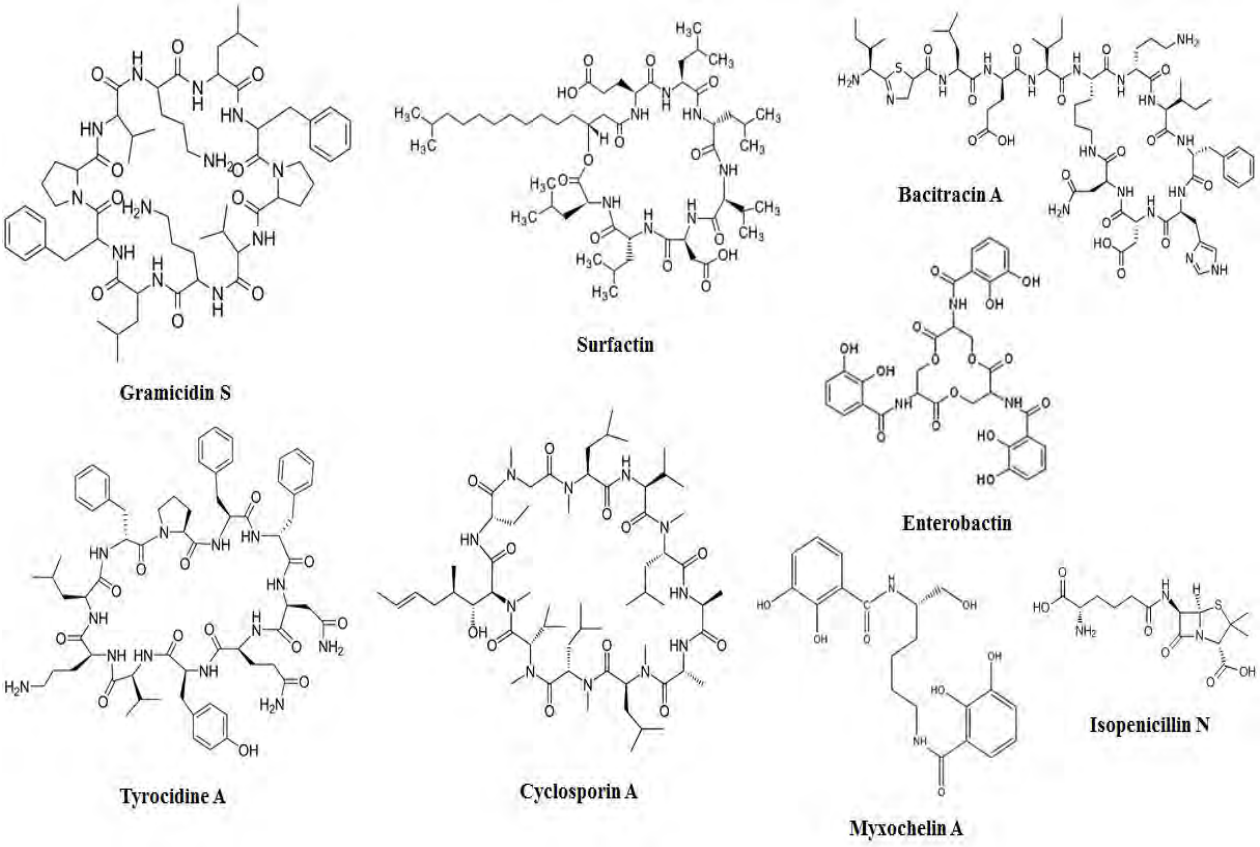

Figure 1.4 The chemical structures of some bacterial (gramicidin S, bacitracin A, tyrocidine A, surfactin and enterobactin) and fungal (cyclosporin A and isopenicillin) bioactive compounds whose peptide backbones are synthesized by the non-ribosomal thiotemplate mechanism (adapted from Marahiel, 1997).

More importantly, NRPSs are organised into repeated functional units known as modules, each of which is responsible for one stage of polypeptide chain elongation leading to the formation of the final non-ribosomal peptide product (Galm *et al.,* 2008; Cane & Walsh, 1999; Marahiel, 1997). The number of modules within an NRPS generally corresponds to the number of amino acids present in the structure of the peptide being synthesized and the modules are aligned in such a way as to be co-linear with the corresponding peptide product (Lautru & Challis, 2004; Marahiel, 1997). Additionally, each module consists of several structurally independent domains, with each one catalysing a single reaction step such as activation, substrate recognition, covalent bonding, optional modification of the incorporated amino acid monomer and condensation with the amino acid acyl or peptidyl group on the neighbouring module (Figure 1.5) (Mootz *et al.,* 2002).

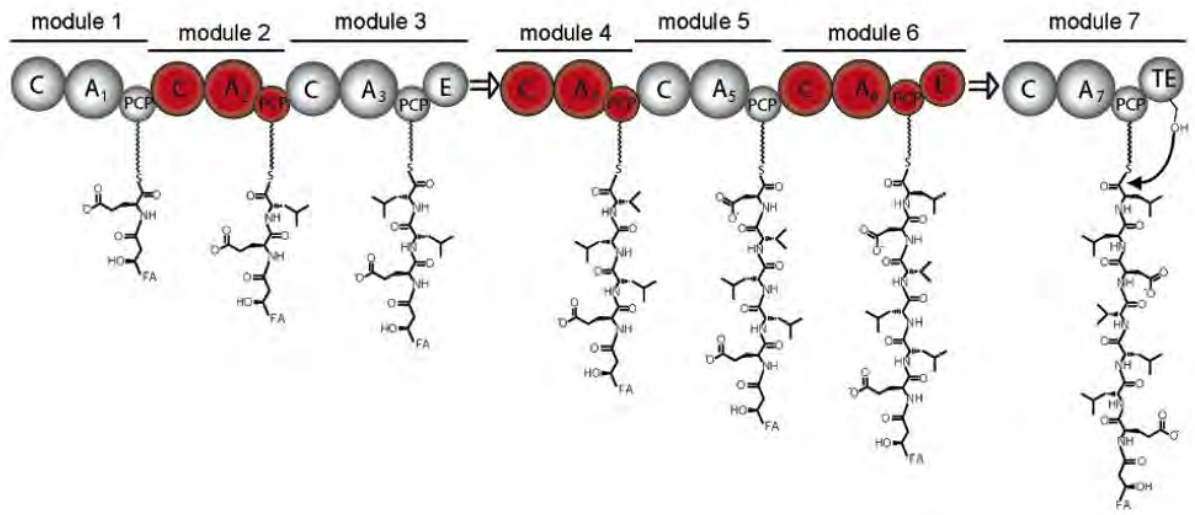

Figure 1.5 Example of a NRPS assembly line consisting of seven modules responsible for the incorporation of seven amino acids. Twenty-four chemical reactions are catalysed in the formation of the final peptide product by twenty-four domains of five different types (C, A, PCP, E and TE) (Sieber & Marahiel, 2005).

A basic elongation module consists of an adenylation (A) domain, a condensation (C) domain and a peptidyl carrier protein (PCP) domain. The core of each module is the A domain that is responsible for the activation of the cognate amino acid as an aminoacyl adenylate through ATP hydrolysis. The activated substrate is then transferred onto the thiol moiety of the PCP domain, which subsequently transports the substrate to the C domain. The C domain catalyzes peptide bond formation between this amino acid substrate and the peptide attached to the PCP domain of the preceding module, thus elongating the peptide chain. Following peptide bond formation, the elongated peptide chain is attached to the PCP domain of the downstream module, where it is passed off and further elongated in the next peptidyltransferase reaction. The thioesterase (TE) domain, in the termination module, catalyzes the release (and in some cases oligomerization and/or cyclization) of the mature NRPS-bound peptide product (Figure 1.6) (Bloudoff *et al.,* 2013).

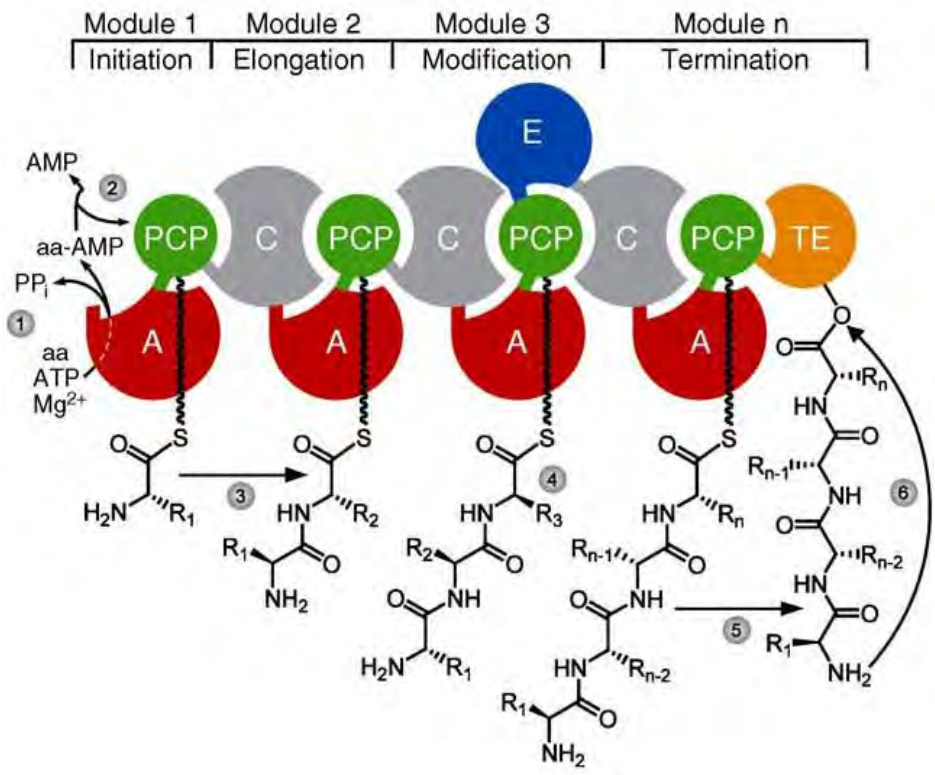

Figure 1.6 A simplified example of a typical NRPS assembly line. (**1**) The cognate amino acids are activated as aminoacyl-adenylates by the A domains. (**2**) The aminoacyl-adenylates are transferred onto the PCP domains. (**3**) Sequential condensation of the PCP-bound amino acids. (**4**) Possibility of modifications of the bound substrate. (**5**) The peptide chain is transferred from the terminal PCP domain onto the TE domain via a transesterification reaction. (**6**) Final product is released by hydrolysis or macrocyclization. The number of modules and modification domains varies with each specific multienzyme system (Strieker *et al.,* 2010).

1.4.1 Assembly logic of NRP synthesis

A typical NRPS consists of an initiation, extension and termination module (Figure 1.6). The minimal NRPS module consists of two catalytic domains and a carrier protein, which together carry out the selection and activation of the substrate and formation of the peptide bond. These core domains are the A, C and PCP (otherwise known as the thiolation (T) domain), which appear in the canonical order C-A-PCP(T) (Figure 1.5). The initiation module does not contain

a C domain (Finking & Marahiel, 2004; Mootz *et al.,* 2002). A fourth domain, a thioesterase (TE), is often found at the C-terminus of the NRPS termination module and catalyzes the release of the peptide from the NRPS (Felnagle *et al.,* 2008).

Furthermore, in addition to the required domains involved in constructing the peptide backbone, modules can contain a number of embedded editing/tailoring domains which are able to create structural diversity by modifying the incorporated amino acid. These auxiliary domains can act *in cis* or *in trans* during the biosynthetic process and examples include epimerization (E), heterocyclization (Cy), oxidation (Ox), reduction (RE), N-methyltransferase (MT), communication-mediating (COM) and aminotransferase (AMT) domains (Finking & Marahiel, 2004).

1.4.1.1 Activation by the Adenylation domain

The A domain is often referred to as the most important domain of each module and this is because it contains the substrate recognition and ATP-binding sites, which are necessary for the activation of the specific substrate amino acid and formation of its acyl adenylate through ATP hydrolysis (Marahiel, 1997).

The A domain selects the cognate amino acid from the available pool of substrates and tethers it to the PCP domain in a two-step process. Briefly, it first catalyzes the formation of an aminoacyl-adenylate intermediate through consumption of ATP, which is co-ordinated to Mg^{2+} and releases inorganic pyrophosphate (PPi) (Figure 1.7). Subsequently, the A domain allows the transfer of the activated acyl moiety onto the free thiol group on the phosphopantetheinyl arm of the active holo-PCP domain, thereby releasing AMP and producing a carboxy-thioester-bound intermediate that is both thermodynamically activated and kinetically labile (discussed in the following section) (Fischbach & Walsh, 2006; Sieber & Marahiel, 2005; Mootz *et al.,* 2002).

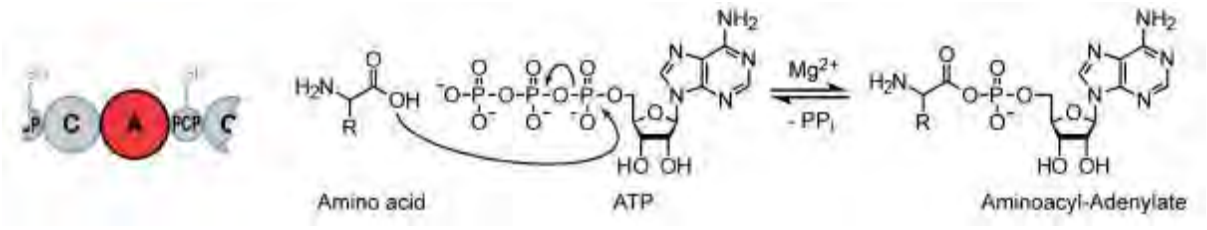

Figure 1.7 The formation of the aminoacyl-adenylate intermediate catalysed by the A domain, at the expense of ATP (adapted from Marahiel, 2009).

The A domain belongs to the large superfamily of adenylate-forming enzymes, which includes all A domains of modular peptide synthetases and distinct proteins such as acetyl-CoA synthetases, luciferases, oxidoreductases and 4-coumaryl CoA ligases. All of these enzymes possess the ability to activate their carboxylic acid or amino acid substrates to their acyl adenylates at the expense of ATP (Konz & Marahiel, 1999; Marahiel, 1997). They all share a homologous domain of approximately 550 amino acids that consists of a set of highly conserved signature sequences, which have been shown to be the major determinants of the substrate specificity exhibited by A domains (Marahiel, 1997).

The determination of the crystal structures of two members of this adenylate-forming enzyme superfamily, firefly luciferase of *Photinys pyralis* and the phenylalanine-specific A domain of the gramicidin S synthetase (GrsA) of *Bacillus brevis*, termed PheA, provided fundamental insight into the structural basis of substrate recognition and activation (Finking & Marahiel, 2004; Stachelhaus *et al.,* 1999; Conti *et al.,* 1997; Marahiel, 1997). Furthermore, it was observed that even though the protein sequence similarity between PheA and the firefly luciferase is only 16 %, the enzymes share a highly conserved three-dimensional structure. It can be expected that very similar three-dimensional structures would be observed for all A domains of peptide synthetase origin that show 30-60 % sequence identity to PheA (Marahiel, 1997).

The A domain of GrsA folds into a large amino-terminal and a smaller carboxy-terminal subdomain. An ordered layer of water molecules mediate interactions between the two sub-

domains. Ten highly conserved core motifs which surround the active site where the substrate binds are associated with the amino-terminal domain, however, a lysine residue located in the carboxy-terminal domain provides an essential interaction in co-ordinating the α-carboxyl group of the substrate amino acid (Sieber & Marahiel, 2005; Stachelhaus *et al.,* 1999).

The conserved motifs serve as functional anchors and it has been shown that many of the amino acid residues within these motifs perform a key role in the co-ordination of ATP binding and hydrolysis, as well as adenylation of the specific substrate carboxy moiety (Sieber & Marahiel, 2005; Stachelhaus *et al.,* 1999). It has, however, been observed that some A domains display "relaxed" substrate specificity and in these cases, chemically or sterically similar amino acids are also recognised and analogously processed (Keating & Walsh, 1999).

1.4.1.2 Transport of substrates and intermediates to the catalytic centres by the Peptidyl Carrier Protein

Located downstream of the A domain, the equally important PCP domain is approximately 80-100 amino acid residues in length and is highly homologous to the acyl carrier protein (ACP) involved in fatty acid synthases (FASs) and polyketide synthases (PKSs) (Sieber & Marahiel, 2005). Carrier proteins, also referred to as thiolation domains, may either be freestanding or embedded in enzymatic systems. A variant of the carrier protein commonly found in siderophore NRPS systems, is the aryl carrier protein (ArCP) (Qiao *et al*., 2007). The carrier proteins are tasked with keeping reaction intermediates bound to the enzymatic machinery and are responsible for transportation of substrates and elongation intermediates to the catalytic centres of the PKS, NRPS or FAS assembly lines (Qiao *et al.,* 2007; Stachelhaus *et al*., 1996).

In order for the PCP to become functional, each inactive apo-PCP domain must be post-translationally modified ("primed") by a phosphopantetheinyl transferase (PPTase), which involves the transfer of the cofactor, 4'-phosphopantetheine (4'-PP), from coenzyme A (CoA) to a conserved serine residue on the PCP domain. The PPTase stimulates the nucleophilic attack of the PCP serine hydroxyl group on the pyrophosphate bridge of CoA, resulting in the transfer of 4'-PP to the PCP domain and release of 3',5'-ADP (Lambalot *et al.,* 1996). This, in turn, results in an active holo-PCP domain with a flexible phosphopantetheinyl arm that is able to covalently bind both the amino acyl and peptidyl substrates as energy-rich thioesters (Figure

1.8) (Fischbach & Walsh, 2006; Schwarzer *et al.*, 2003). The PCP-domain is then able to act as the transport unit which enables the activated amino acids and elongation intermediates to move between all of the catalytically active entities, as the A and C domains associate closely to form a catalytic platform onto which the PCP domain is flexibly tethered between their active sites (Tanovic *et al.,* 2008).

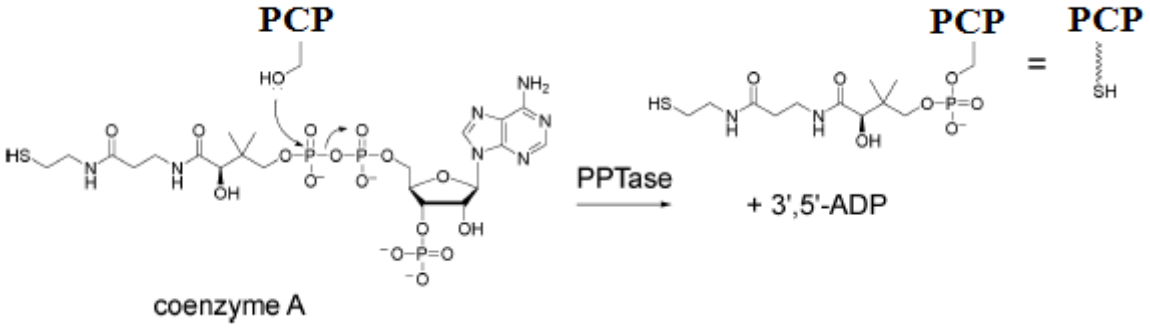

Figure 1.8 Modification of the PCP domain by a phosphopantetheinyltransferase (PPTase). The transfer of the 4'- phosphopantetheine cofactor from coenzyme A to a conserved serine residue in the PCP domain is catalysed by members of this enzyme class (adapted from Fischbach & Walsh, 2006).

1.4.1.3 Peptide elongation by the Condensation domain

The C domain is approximately 450 amino acids in length and is responsible for the elongation of the peptidyl chain. C domains are normally localized between each consecutive A-PCP domain and catalyse the condensation reaction between the peptidyl chain tethered to the phosphopantetheinyl arm of the upstream PCP domain and the amino acid bound to the downstream PCP domain (Lautru & Challis, 2004). However, in the first condensation reaction of an NRPS, both these reaction intermediates would typically be aminoacyl groups attached to their respective PCP domains. Furthermore, C domains are absent from modules involved in peptide initiation (Mootz *et al.,* 2002).

There is little information concerning the exact mechanism of peptide elongation and how the interaction of modules affects the direction of polymerization, but it is thought that peptide bond formation proceeds via the nucleophilic attack of the free α-amino group on the downstream PCP-bound acceptor amino acid on the activated carboxy-thioester of the upstream PCP-bound donor amino acid (Figure 1.9). This reaction facilitates the translocation of the growing peptide chain onto the next module for elongation and structural changes (Bloudoff *et al.,* 2013; Sieber & Marahiel, 2005; Mootz *et al.,* 2002; Marahiel, 1997). Moreover, this condensation reaction is strictly unidirectional, leading to a downstream-directed synthesis of the NRPS product (Samel *et al.,* 2007).

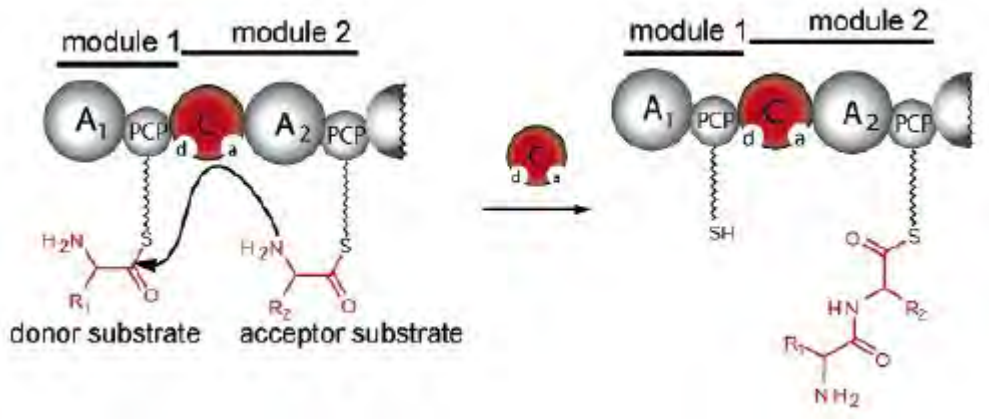

Figure 1.9 Peptide elongation catalysed by the C domain, which involves an attack of the nucleophilic amine of the acceptor substrate onto the electrophilic thioester of the donor substrate (Sieber & Marahiel, 2005).

In order to shed light on the catalytic mechanism of the C domain, three structures that contain NRPS C domains have been determined by X-ray crystallography, including a stand-alone C domain (Keating *et al.,* 2002), a C-PCP didomain complex (Samel *et al.,* 2007) and a C-A-PCP-TE termination module (Tanovic *et al.,* 2008). The overall architecture of the C domain revealed a pseudodimeric configuration consisting of both an N-terminal and C-terminal subdomain. The active site is located at the bottom of a "canyon" or "V-shape" formed by the two subdomains and is covered by a "latch" that crosses over from the C to N subdomain

(Figure 1.10) (Bloudoff *et al.,* 2013; Marahiel, 2009; Finking & Marahiel, 2004). The catalytic centre includes a conserved HHxxxDGxS core motif (where x is any amino acid and defined residues are represented by the single-letter amino acid code), that is also found in dihydrolipoyl transacetylases, chloramphenicol acetyltransferases, and NRPS epimerization and Cy domains (Sieber & Marahiel, 2005; Keating & Walsh, 1999; Marahiel *et al.*, 1997). A model, supported by mutational studies, suggests that the second histidine of this motif, which is located at the bottom of the canyon, may act as a catalytic base promoting the deprotonation of the NH_3^+ moiety of the thioester-bound nucleophile prior to peptide bond formation. However, recent pK value analysis of this active site residue suggests that peptide bond formation may depend mainly on electrostatic interactions rather than a general acid/base catalysis (Marahiel, 2009; Samel *et al.,* 2007; Konz & Marahiel, 1999).

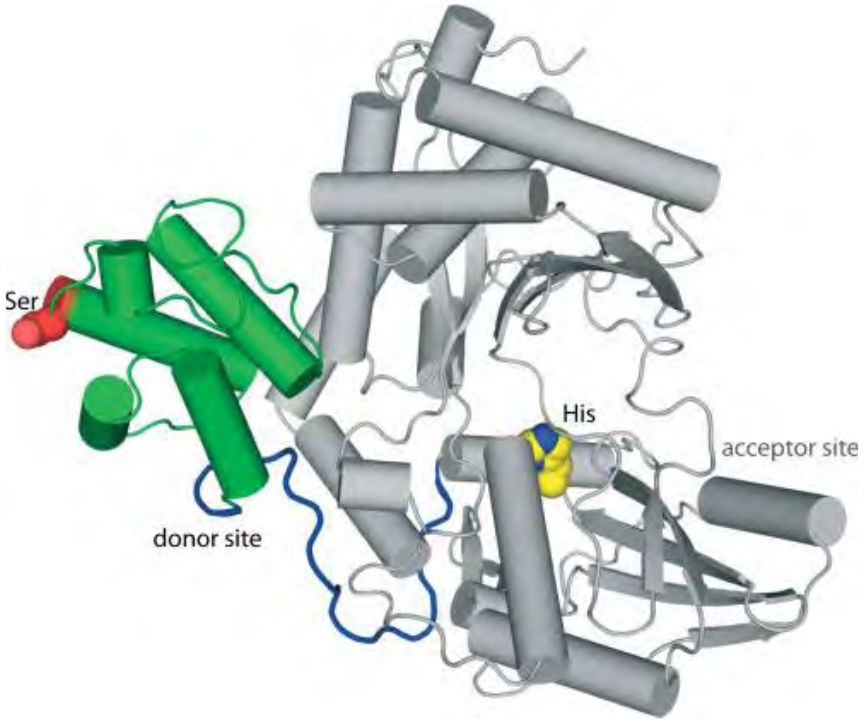

Figure 1.10 Structure of the PCP-C didomain from the surfactin synthetase illustrating the active site histidine (His) residue located on the floor of the C-domain "canyon" (Marahiel, 2009).

1.4.1.3.1 Heterocyclization domains

The C domain of a module, which catalyzes basic peptide bond formation only, can be replaced by a specialised heterocyclization (Cy) domain that shares striking structural and functional homology to the C domain. The Cy domain combines the condensation function of the C domain with additional heterocyclization and dehydration functions using the side chains of the amino acids cysteine, serine or threonine within the product peptide backbone to produce thiazoline (from cysteine), oxazoline (from serine) or 5-methyloxazoline (from threonine) heterocycles (Walsh *et al.,* 2001).

Cy domains were first identified in 1997 in the cyclic dodecylpeptide antibiotic bacitracin synthetase, produced by *Bacillus licheniformis* ATCC 10716 (Konz *et al.,* 1997) and were further validated in the biochemical characterisation of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis* (Gehring & Walsh, 1998). The catalytic core motif, HHxxxDGxS, of the C domain is modified to DxxxxDxxS in the Cy domain. The conserved aspartate residues are critical for both condensation and heterocyclization (Keating *et al*., 2002; Keating & Walsh, 1999). Examples of secondary metabolites produced using Cy domains are epothilone A, myxothiazol and mycobactin A produced by *Sorangium cellulosum* So ce90*, Stigmatella aurantiaca* DW4/3-1 and *M. tuberculosis*, respectively (Figure 1.11) (Duerfahrt *et al.,* 2004).

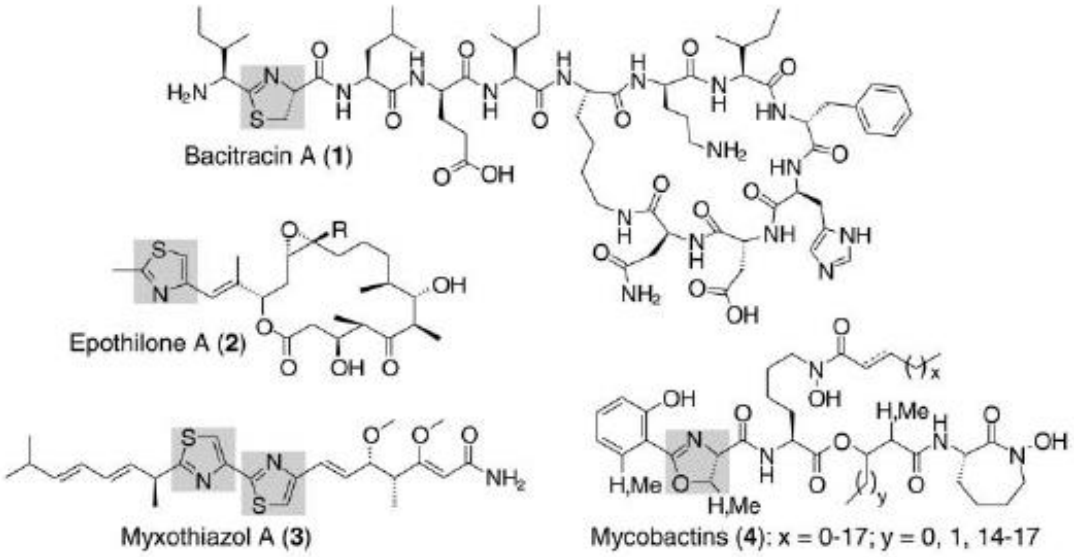

Figure 1.11 Heterocyclic ring-containing secondary metabolites from various organisms (the five-membered thiazole and oxazole rings are shaded grey) (Duerfahrt *et al.,* 2004).

The first reaction step catalysed by Cy domains is peptide bond condensation, carried out by the nucleophilic attack of a PCP-bound cysteine, serine or threonine acceptor substrate onto the thioester of the donor substrate. The next step involves the nucleophilic attack by the hydroxyl side chain of serine/threonine or thiol side chain of cysteine onto the carbonyl C atom of the newly-formed peptide bond to yield hemiaminal or thiohemiaminal intermediates and the five-membered heterocyclic ring. The intermediates are subsequently dehydrated to yield the C=N bond and the final oxazoline or thiazoline product (Figure 1.12) (Sieber & Marahiel, 2005; Walsh *et al.,* 2001). In non-ribosomal peptide products such as the glycopeptide antibiotic, bleomycin (produced by '*Streptomyces verticillatus*' ATCC 15003) or myxothiazol, additional oxidation (Ox) domains convert these heterocycles into more stable oxazole or thiazole rings. Heterocyclic rings are common structural features of NRPs and are important for the interaction with proteins, DNA and RNA, as well for chelating metal ions (Walsh, 2004).

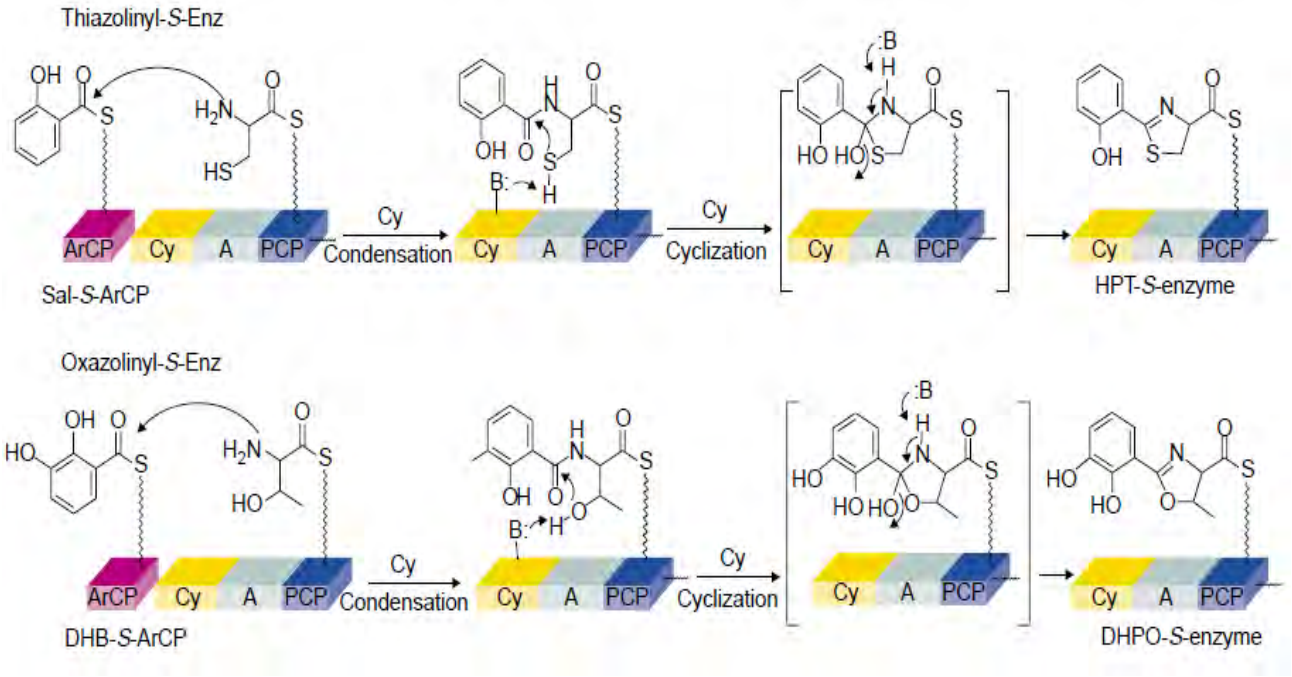

Figure 1.12 The formation of thiazoline (top of figure) and oxazoline heterocycles (bottom of figure) from cysteine and threonine precursors, respectively (Walsh *et al.,* 2001).

1.4.1.4 Peptide release by the Thioesterase domain

The termination of peptide synthesis on an NRPS assembly line is catalysed by the terminal enzyme of the last module. During synthesis, the growing peptide chain is transported from one module to the next until it reaches the final module's PCP domain. Product release is then achieved by the catalytic action of a C-terminal TE domain, which occurs when the nascent peptidyl chain is transferred from the terminal PCP to a highly conserved active site serine residue present in the TE domain to generate a covalent acyl-enzyme intermediate. This intermediate can then be released by the external nucleophile, water, in a hydrolysis reaction to form a linear acid or by the predominant mechanism of regio- and stereoselective intramolecular macrocyclization using an internal nucleophile to produce a cyclic peptide (Figure 1.13) (Sieber & Marahiel, 2005; Mootz *et al.,* 2002). Macrocyclization is believed to be favoured due to the fact that cyclic products are more resistant to proteolytic cleavage. In

addition, TE domains that catalyse a cyclization reaction are also referred to as peptide cyclases (Marahiel, 2009; Sieber & Marahiel, 2005; Finking & Marahiel, 2004).

The crystal structures of several dissected TE domains has revealed the presence of the common fold of α/β-hydrolases, such as lipases and esterases, as well as the fact that the active site consists of a catalytic triad composed of serine at position 80 (Ser80), histidine at position 207 (His207) and aspartate at position 107 (Asp107). Within this triad, the serine residue serves as the site of tetrahedral intermediate formation that is stabilized by an oxyanion hole en route to the acyl-enzyme intermediate. This intermediate is attacked by the action of an internal nucleophile of the peptide chain (which can be the N-terminal amino group or a functional side chain) or water, which results in a macrocyclic product or linear peptide, respectively. It remains uncertain as to how the active site is sufficiently sealed from water in order to catalyse cyclization rather than hydrolysis. However, the structures of the TE domains have also revealed a flexible "lid" region that may possibly adopt an open conformation for substrate entry and a closed conformation for excluding water from the active site (Condurso & Bruner, 2012; Marahiel, 2009; Schwarzer *et al.,* 2003).

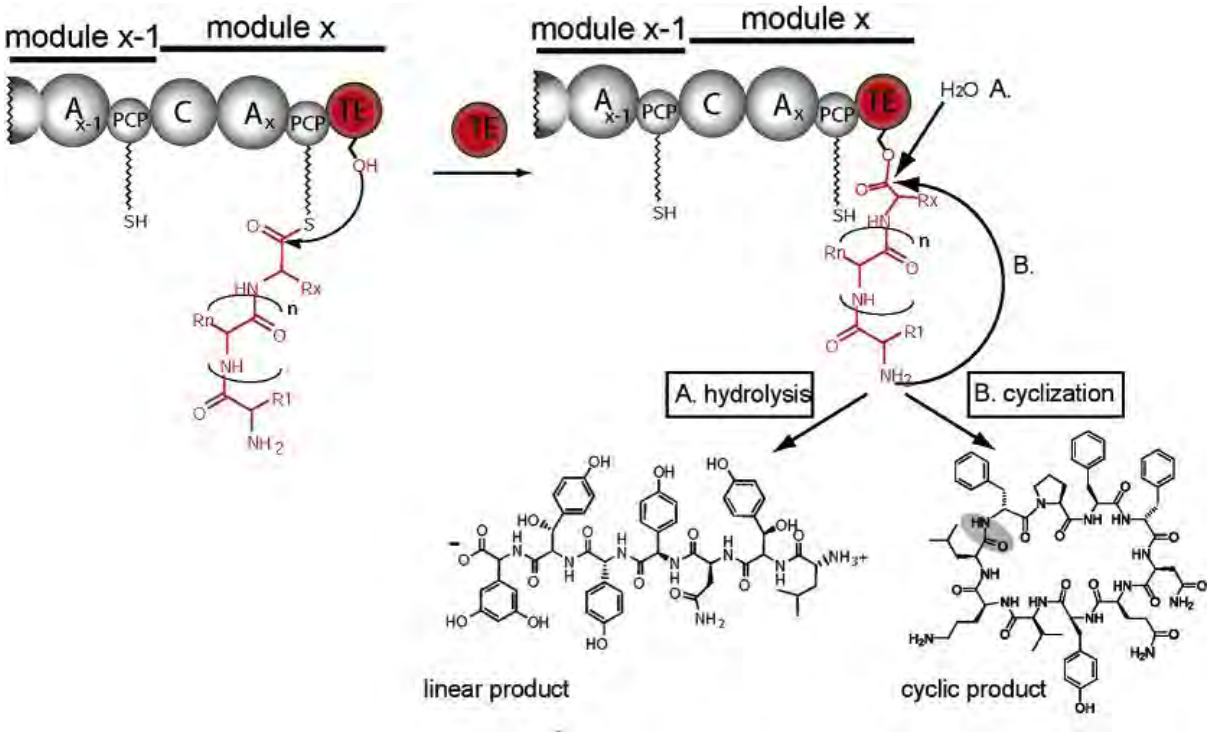

Figure 1.13 Illustration depicting peptide release by the TE domain. Product release can be achieved either by the external nucleophile, water, to produce a linear acid product as observed in (A) or by an internal nucleophile to produce a cyclic product as observed in (B) (Sieber & Marahiel, 2005).

1.4.1.5 Editing/tailoring domains

In addition to the essential domains found in every NRPS, further modifications to the final product can be achieved by the action of several editing/modifying domains that act to increase the diversity of NRPs. These so-called modifying domains are not present in every NRPS system, but are nevertheless vital for the proper processing of their designated substrate within their respective synthetase. Inactivation or deletion of these domains usually results in the synthesis of products with severely reduced or no bioactivity.

1.4.1.5.1 Methylation

Certain non-ribosomal peptides such as cyclosporin, enniatin, actinomycin and yersiniabactin possess *N*-methylated or *C*-methylated peptide bonds. These modifications are introduced by a methyltransferase (MT) domain, embedded within the canonical fold of an associated A domain, thus making the peptide less susceptible to proteolytic degradation. The MT domain, including both N- and C-methyltransferases, is able to catalyse the transfer of the *S*-methyl group from *S*-adenosylmethionine (SAM) to the α-amino group of the thioesterified aminoacyl intermediate in C-A-MT-PCP modules, thereby releasing *S*-adenosylhomocysteine as a reaction byproduct (Figure 1.14) (Fischbach & Walsh, 2006; Sieber & Marahiel, 2005; Finking & Marahiel, 2004).

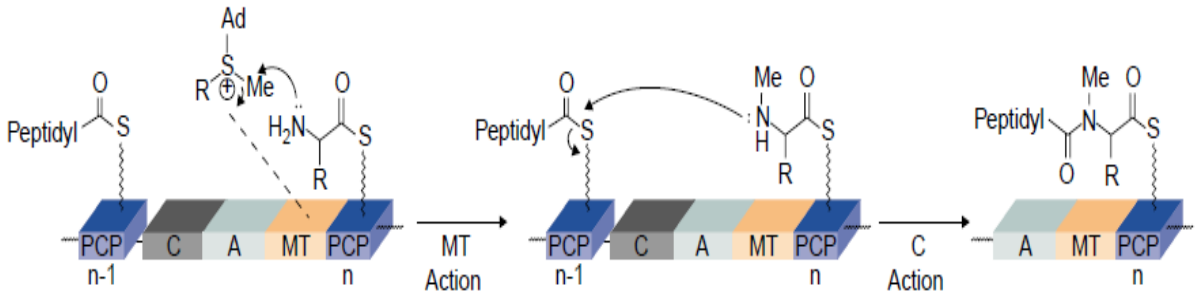

Figure 1.14 The MT domain catalyzes the transfer of the CH_3 (Me) group from SAM to the amino group of the aminoacyl-*S*-intermediate prior to peptide formation with the upstream peptidyl chain (Walsh *et al*., 2001).

1.4.1.5.2 Epimerization (control of stereochemistry)

A prevalent structural feature of NRPs is the incorporation of the less common D-amino acids. Two different strategies are utilized by NRPSs for their incorporation, which may be accomplished simply by an A domain with exclusive specificity towards a D-amino acid or by epimerization of L-amino acids by integrated epimerization (E) domains (Sieber & Marahiel, 2005). The E domains are found on the carboxy terminus of the respective module's PCP domain and catalyse the racemization of the PCP-bound amino acid or of the C-terminal amino acid of the growing peptidyl chain. Two different types of E domains catalyse these reactions and are known as either aminoacyl epimerases, which can only be part of initiation modules, or peptidyl epimerases, which are part of elongation modules. Due to the fact that E domains are only involved in catalyzing the racemization of their substrates, specific incorporation of the D-amino acid into the growing peptide chain is achieved by the enantiomer-selective donor site of the downstream C domain. The C domain therefore acts as a type of "filter", which selectively withdraws the correct enantiomer from the pool of L/D-amino acids (Schwarzer *et al.,* 2003). In some rare cases, such as in the arthrofactin synthetase found in *Pseudomonas* sp. MIS38, the C domain itself also exhibits epimerization activity in addition to its normal function and it is then known as a "dual C/E" domain (Balibar *et al.,* 2005).

1.4.1.5.3 Reduction domain

A reduction (RE) domain replaces the TE domain in a few NRPS systems, such as in the biosynthesis of gramicidin A in *B. brevis* and safracin in *Pseudomonas fluorescens*. RE domains, such as those found in the biosynthesis of myxochelin A in *Angiococcus disciformis,* catalyse an alternative mechanism of peptide release that involves the NADPH-dependent reduction of the PCP-bound peptidyl thioester to yield an unstable thioacetal intermediate that is converted into a linear aldehyde or alternatively into an alcohol via a second reduction of the thioacetal group (Figure 1.15) (Velasco *et al.,* 2005; Gaitatzis *et al.,* 2001). Another rare type of chain termination strategy is found in some NRPSs, where the RE domain triggers the reductive release, via nucleophilic attack by the free N-terminal amino group, of a highly reactive peptide aldehyde that results in the subsequent formation of a stable macrocyclic imine (Kopp & Marahiel, 2007). RE domains are also known to catalyse the NADPH-dependent

reduction of thiazoline into thiazolidine by the addition of two electrons, as witnessed in one of the rings in yersiniabactin (Sieber & Marahiel, 2005; Velasco *et al.,* 2005).

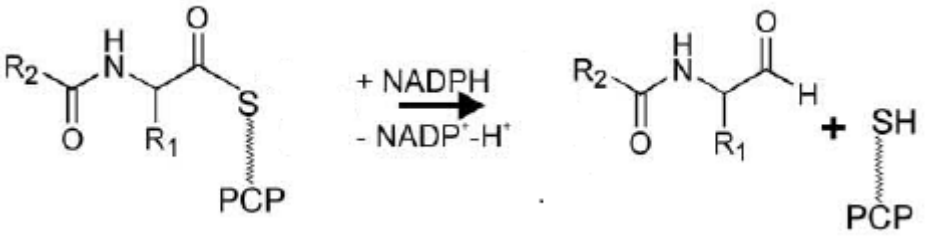

Figure 1.15 The RE domain reduces the C-terminal carboxy-group to an aldehyde or to the corresponding alcohol using NADPH as cofactor (Schwarzer *et al.,* 2003).

1.4.1.5.4 Oxidation domain

Oxidation (Ox) domains often appear in the same module as a Cy domain in order to catalyze the oxidation of the hydrolytically labile dihydroheterocyclic thiazolines and oxazolines by two-electron transfer to yield stable heteroaromatic thiazole or oxazole rings (Figure 1.16) (Duerfahrt *et al.,* 2004; Schwarzer *et al.,* 2003). Ox domains can be localized in the accompanying A domain, as is the case in the epothilone synthetase, or they can be C-terminally fused to the PCP domain of the respective module, as witnessed in the bleomycin synthetase (Schwarzer *et al.,* 2003; Du *et al.,* 2000). Furthermore, flavin mononucleotide (FMN) has been verified as a cofactor of Ox domains. FMN is reduced to $FMNH_2$ during the oxidation of the substrate and is subsequently re-oxidised by the reduction of molecular oxygen to superoxide. While the oxidation of thiazoline to thiazole in the epothilone synthetase is catalysed *in cis* by an FMN-containing Ox domain (a corresponding *in cis* flavoprotein domain exists in the bleomycin synthetase), a second *trans*-acting Ox domain has been proposed to be involved in the formation of the second thiazole ring found in the structure of bleomycin A2 (Finking & Marahiel, 2004; Du *et al.,* 2000). The fact that these enzymes are able to act *in trans*, where they are able to recognise the peptidyl chains on NRPS modules by protein-protein

interactions, is testament to the enormous structural variation created in these natural products (Walsh *et al.,* 2001).

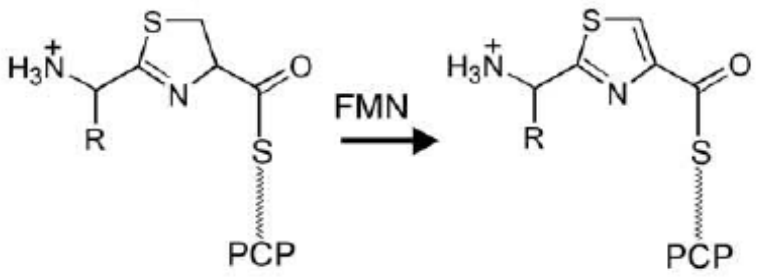

Figure 1.16 FMN-containing Ox domains catalyze the two electron oxidation of thiazolines and oxazolines to yield stable heteroaromatic thiazole or oxazole rings, respectively (Schwarzer *et al.,* 2003).

1.4.1.5.5 Further modifications

In the case of the linear gramicidin A, the N-terminus of the non-ribosomal peptide possesses a formyl group which is introduced by a formylation (F) domain, localized N-terminal to the A domain of the initiation module. The F domain is responsible for the *N*-formylation reaction by means of the cofactor *N*-formyltetrahydrofolate. Other formylated NRPs include coelichelin, whose ornithine residues are believed to be *N*-formylated *in trans* and anabaenopeptilide 90-A whose N-terminus exhibits an *N*-formylated glutamine residue (Lautru *et al.,* 2005; Sieber & Marahiel, 2005).

Surfactin and fengicin are lipopeptides, which represent a subgroup of NRPs whose peptide backbones are N-terminally bound to a fatty acid. An N-terminal C domain, in addition to the A and PCP domain, is usually present in the initiation module of these NRPSs and is thought to catalyse bond formation to the fatty acid.

An aminotransferase (AMT) domain exists within the mycosubtilin NRPS that converts a thioesterified β-ketoacyl intermediate to a covalently tethered β-aminoacyl thioester with

simultaneous removal of the α-amino group of glutamine to form α-keto-glutamine in a variation of the classic transamination reaction. It was therefore proposed to be responsible for the transfer of an amine group to the β-position of the growing acyl chain. The ability to incorporate amine functionality directly into NRPs creates a functional handle for macrocyclization strategies and acts as a potent hydrogen bond donor, which can significantly alter the biology of a given compound (Fischbach & Walsh, 2006).

All of the modifications that have been introduced thus far are catalysed by domains that are embedded into their respective NRPS modules and almost entirely affect only the peptide backbone. Additionally, certain enzymes such as glycosyl transferases, halogenases and hydroxylases are able to modify the peptide's side chains and are generally encoded in the same biosynthetic gene cluster.

1.4.2 NRPS substrate specificity prediction and the non-ribosomal code

The A domain has been described as the most important domain within each NRPS module due to the fact that it recognizes and activates only the cognate substrate amino acid as its acyl adenylate (using ATP to drive the reaction). Hence A domains act in a 'lock and key' process and are considered the primary determinant of substrate specificity, thereby also defining what molecules can be synthesized by the specific NRPS (Lautru & Challis, 2004; von Döhren *et al.*, 1999). Knowledge about how A domains are able to recognize their specific substrates is of fundamental importance to understanding and manipulating these complex enzymatic systems for use as scaffolds for combinatorial biosynthesis. Combinatorial biosynthesis (section 1.4.4) could allow for the generation of novel derivatives of the non-ribosomal peptide and ideally new, more potent antibiotics (Rausch *et al.,* 2005; Lautru & Challis, 2004; Challis *et al.,* 2000; Conti *et al.,* 1997).

The study of A domain specificity was greatly facilitated by the determination of the crystal structure of firefly luciferase and the phenylalanine activating domain, PheA, from the first module of gramicidin synthetase (GrsA). The co-crystalization of PheA in a ternary complex with L-phenylalanine (L-phe) and AMP allowed the discovery of ten highly conserved motifs, denoted A1 to A10, which extend over a region of 450 amino acids and mostly surround the active site where the substrate binds. The identification of this hydrophobic L-phe binding

pocket and the A domain residues making contact with L-phe provided a structural basis for understanding the specificity of peptide synthetases such as NRPSs (Rausch *et al.,* 2005; Lautru & Challis, 2004; Challis *et al.,* 2000; Conti *et al.,* 1997).

Expanding on the approach pursued by Conti and colleagues, Stachelhaus *et al.* (1999) and Challis *et al.* (2000) compared the L-phe binding pocket of GrsA with the corresponding sequences of over 150 aminoacyl and iminoacyl adenylate-forming domains of NRPSs to pinpoint the essential amino acid residues involved in substrate specificity and binding. Both studies examined a ~100 amino acid stretch between the core motifs A4 and A5 and identified 10 amino acid residues that were within approximately 5.5 Å of the active site-bound phenylalanine and in contact with the substrate (Rausch *et al.,* 2005; Lautru & Challis, 2004; Conti *et al.,* 1997).

These 10 amino acid residues were determined to be crucial for substrate binding and catalysis, as alignments of consensus A domain residues to the PheA domain revealed that all A domains display the same type of binding pockets, with different key residues that interact with different amino acid substrates. It was postulated that two residues, Asp235 and Lys517, stabilize the α-amino group of the amino acid substrate via two hydrogen bonds and are critical for the correct positioning of the substrate within the active site for ATP-dependent activation. These two residues are located in the conserved core motifs of A4 (Asp235) and A10 (Lys517). The other residues bordering the PheA specificity binding pocket were determined to be Ala236, Ile330 and Cys331 on the one side and Ala322, Ala301, Ile299 and Thr278 on the other side of the pocket. Both sides are separated by the indole ring of tryptophan at position 239 (Trp239), which is located at the bottom of the pocket (Stachelhaus *et al.,* 1999).

More importantly, it was determined that due to the high degree of sequence identity shared between NRPS A domains, the amino acid residues that correspond to those lining the PheA binding pocket could be used to reveal substrate specificity in other A domains. The consecutive order of the 10 residues was determined to constitute the signature sequence involved in the binding pockets of A domains and can be interpreted as the "specificity-conferring code" (also referred to as the non-ribosomal code), as it allows for the prediction of A domain selectivity on the basis of the A domain primary sequence (Figure 1.17) (Table 1.1) (Rausch *et al.,* 2005; von Döhren *et al.,* 1999).

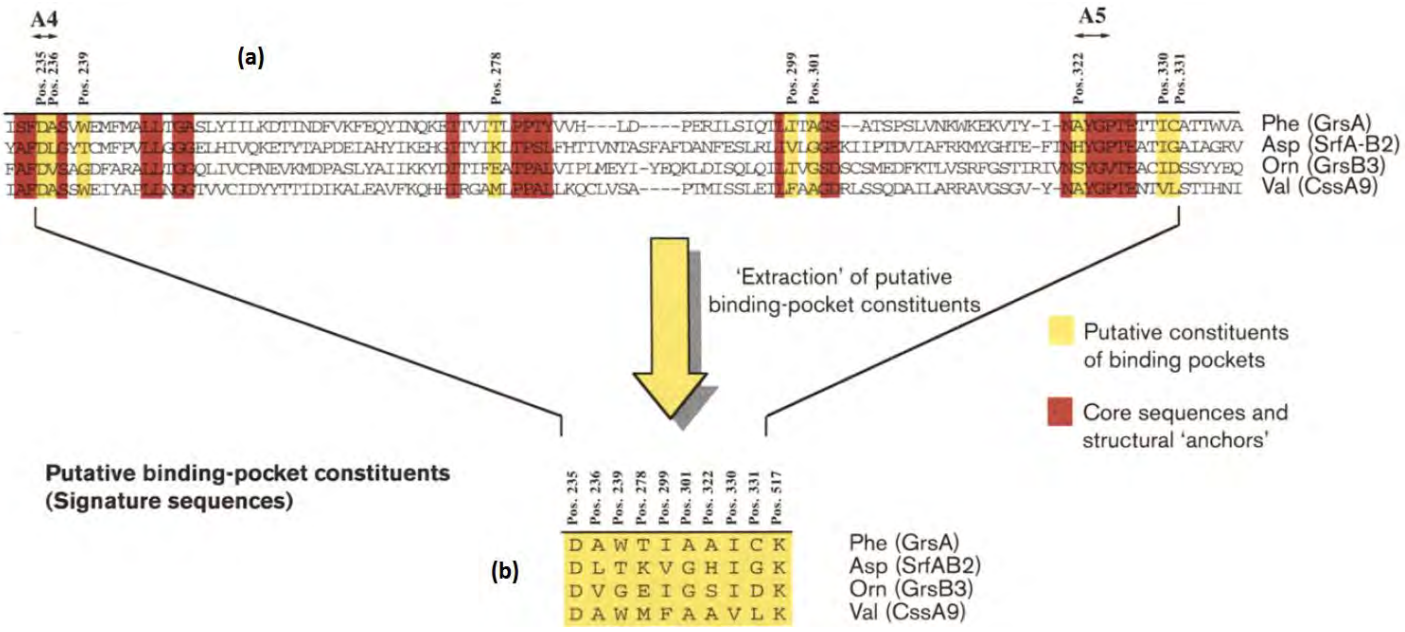

Figure 1.17 A multiple sequence alignment of the primary amino acid sequences from known A domains in order to determine the selectivity-conferring residues **(a)**. The sequence of ~100aa between the core motifs A4 and A5 from PheA from GrsA was aligned with the corresponding sequence of AspA from the surfactin synthetase, SrfA, OrnA from the gramicidin synthetase, GrsB3 and ValA from the cyclosporine synthetase. Yellow residues indicate those involved in the binding pocket positions and brown residues indicate conserved motifs which anchor the alignment. **(b)** Ten highly conserved residues were extracted from the sequence alignment and the consecutive order of the amino acids was determined to constitute the signature sequence involved in the binding pockets of the aligned A domains. The missing residue, Lys517, is highly conserved within motif A10, which is not shown in the protein sequence. The alignment was extended to 160 different A domains to confirm accuracy in determining the signature sequence (adapted from Stachelhaus *et al.*, 1999).

This discovery confirmed previous findings from site-directed mutagenesis and photoaffinity labelling experiments that indicated that these amino acid residues were involved in ATP binding and hydrolysis (Gocht & Marahiel, 1994; Pavela-Vrancic *et al.,* 1994). The non-ribosomal code was initially restricted to amino acid-activating A domains but was extended to carboxy acid-activating A domains when the crystal structure of the stand-alone 2,3-dihydroxybenzoic acid activating domain, DhbE, from *Bacillus subtilis* was solved (May *et al.,* 2002).

Table 1.1 Consensus specificity code for substrates from several 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	

Clusters of signature sequences extracted from A domains activating the same substrates were used to determine the consensus sequences for the recognition of several amino acid substrates. The biosynthetic template from which each A domain specificity code was derived is included, along with the overall similarity of each signature sequence. Variable constituents within each codon are represented by red residues and 'wobble'-like positions, which reveal a large degree of variability throughout all codons, are indicated in blue. 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 (Stachelhaus *et al.*, 1999).

The successful mutation of all the key residues of the PheA binding pocket, which resulted in relaxation or alteration of its substrate specificity, demonstrated the reliability of the non-ribosomal code. This has resulted in the development of web-based NRPS prediction services such as the NRPS-PKS knowledge base (Ansari *et al.,* 2004), NP.searcher (Li *et al.,* 2009), PKS/NRPS analysis (Bachmann & Ravel, 2009) and antiSMASH 3.0 (Weber *et al.,* 2015), which make use of the "specificity-conferring code" to predict putative A domain substrates in NRPS genes. Although these tools have been relatively successful at predicting the specificities of A domains in new NRPS genes, they do have a number of weaknesses (Challis *et al.,* 2000).

A clear shortcoming is that predictions of substrate specificities are based on known A domain sequences. Since not all A-domain sequences in nature are known and, in particular, since there are relatively few sequences for A domains that bind more unusual substrates, the accuracy of substrate specificity prediction is limited. The specificity of uncharacterised A domains must therefore be deduced from the available code for domains with known specificity. It has been observed that there are deviations from the code. For example, not all A domains specific for phenylalanine have the exact same specificity binding pocket sequence as GrsA, e.g. BarG of the barbamide synthesis gene cluster from *Lyngbya majuscula* (GenBank accession number: AAN32981) has DAWTVAAVCK instead of DAWTIAAICK. In addition, there are examples of codes where the predicted substrate specificity does not correspond to the actual activated amino acid, such as in the alanine-activating domain Sare0718 from the marine actinomycete *Salinispora arenicola* CNS-205 (which has the code for valine) and in the biosynthetic cluster for fusaricidin, a mixture of 12 depsipeptides from *Paenibacillus polymyxa* PKB1, which displays relaxed substrate specificity and allows for the incorporation of D-amino acids instead of their L-isomers (Xia *et al.,* 2012; Li & Jensen, 2008).

A further shortcoming has been observed in the analysis of the active sites within the A domains of certain types of NRPSs, particularly those belonging to fungi, as the GrsA crystal (from a bacterium) seems to be an inadequate model for fungal NRPSs or because the large number of sequence variants in the active site of fungal NRPSs does not allow for the identification of the key residues required for substrate-specificity prediction (Prieto *et al*., 2012). Certain positions are considered more variable or 'wobble'-like than others, particularly 239, 278, 299, 322 and 331, which are highly variant. Furthermore, positions 235 and 517 are considered invariant and positions 236, 301 and 330 are moderately variant (Table 1.1). The variability reflects each position's importance in contributing to substrate specificity, but also causes dissimilarity

between the signature sequences for A domains specific for identical substrate amino acids (Stachelhaus *et al.*, 1999).

It is due to these limitations that Rausch *et al.* (2005) expanded on the work of Stachelhaus *et al*. (1999) and Challis *et al.* (2000) by publishing a machine learning algorithm in which transductive support vector machines (TSVMs) were utilised to statistically propose NRPS A domain specificity using the physico-chemical fingerprint of the residues within 8 Å of the active site of the A domain (a total of 34 residues). The residues are encoded into TSVMs based on their physico-chemical properties such as the number of hydrogen bond donors, polarity, volume, secondary structure preferences, hydrophobicity and isoelectric point and a continuously updated dataset of A domains with known specificity are used in predicting substrate specificity. Due to the occurrence of relaxed/promiscuous specificity in certain A domains, such as in the NRPS responsible for xenematide biosynthesis in *Xenorhabdus nematophila*, the specificities for substrates with similar physico-chemical properties are clustered together (Crawford *et al.,* 2011; Rausch *et al.,* 2005).

An open source web-based predictor, NRPSpredictor, based on TSVMs, was built on the 34 active site residues in order to predict A domain specificity, which was later refined and replaced by the NRPSpredictor2, which possesses improved prediction performance, two new prediction levels and a larger database (Rottig *et al.,* 2011).

Although the TSVM-based method was based on a more recent database and was able to provide a substrate-specificity prediction in an additional 18% of cases, it is still beneficial to use it in combination with the empirical predictive method developed by Stachelhaus *et al.* (1999) and Challis *et al.* (2000) in order to create an even more accurate prediction tool (Rausch *et al.,* 2005). This is largely due to the fact that most of the weaknesses of the older method remain and by expanding the number of residues considered, it may have amplified the problems associated with including data with little or no influence on the specificity of the A domains. In addition, the clustering of specificities reduces the accuracy of the predictions and, although this is acceptable for A domains which possess relaxed specificity, other A domains are much more specific. There is currently no way to distinguish a highly specific A domain from a relaxed one (Rausch *et al.,* 2005; Mootz *et al.,* 2002; Challis *et al.,* 2000).

This situation prompted interest in developing new prediction methods supported by other approaches, such as the use of hidden Markov Models (HMM). Khurana *et al.* (2010) applied HMM to functionally classify the acyl-CoA synthetase superfamily members. The results of this work suggest that the application of HMM to classify this superfamily outperforms the predictions based on a restricted number of active site residues (Khurana *et al.,* 2010). Furthermore, a novel two-mode factor analysis model based on latent semantic indexing (LSI) has recently been published by Baranasic *et al.* (2013). This model is able to predict the specific amino acid that is activated by the A domain in contrast to a cluster of similar amino acids. The authors suggest that a detailed comparison of prediction quality against those of the NRPSpredictor, showed that the LSI model performed slightly better and is thus the most accurate method currently available for prediction of A domain substrate specificities (Baranasic *et al.*, 2013).

1.4.3 NRPS biosynthetic strategies

The order in which the modules of an NRPS are utilised to construct the final product and the assembly of the core domains within each module can vary considerably, thereby allowing these enzymatic systems to synthesize diverse chemical structures. The three most common biosynthetic strategies employed by NRPSs include the linear or type A NRPS, iterative or type B NRPS and the nonlinear or type C NRPS (Mootz *et al.,* 2002).

The simplest biosynthetic strategy is that of the linear or type A NRPSs, in which the three core domains are arranged in the order C-A-PCP in an elongation module, which is able to add one amino acid to the growing peptide chain (Figure 1.18). The initiation module is responsible for incorporating the first amino acid of the peptide chain and does not contain a C domain, whereas, the termination module contains a TE domain in order to catalyse the release of the final product. Therefore, a typical, linear NRPS template consists of *n* modules with a domain organization in the order of A-PCP-$(C\text{-}A\text{-}PCP)_{n-1}$-TE, for construction of a final peptide chain containing *n* amino acids. The sequence of the final peptide chain is determined exclusively by the number and order of the modules, which differs in the case of nonlinear and iterative NRPSs. Examples of linear NRPSs include those involved in the biosynthesis of cyclosporin, surfactin, actinomycin, peramine, ergovaline, tyrocidine and δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV), the isopenicillin and cephalosporin precursor (Mootz *et al.,* 2002).

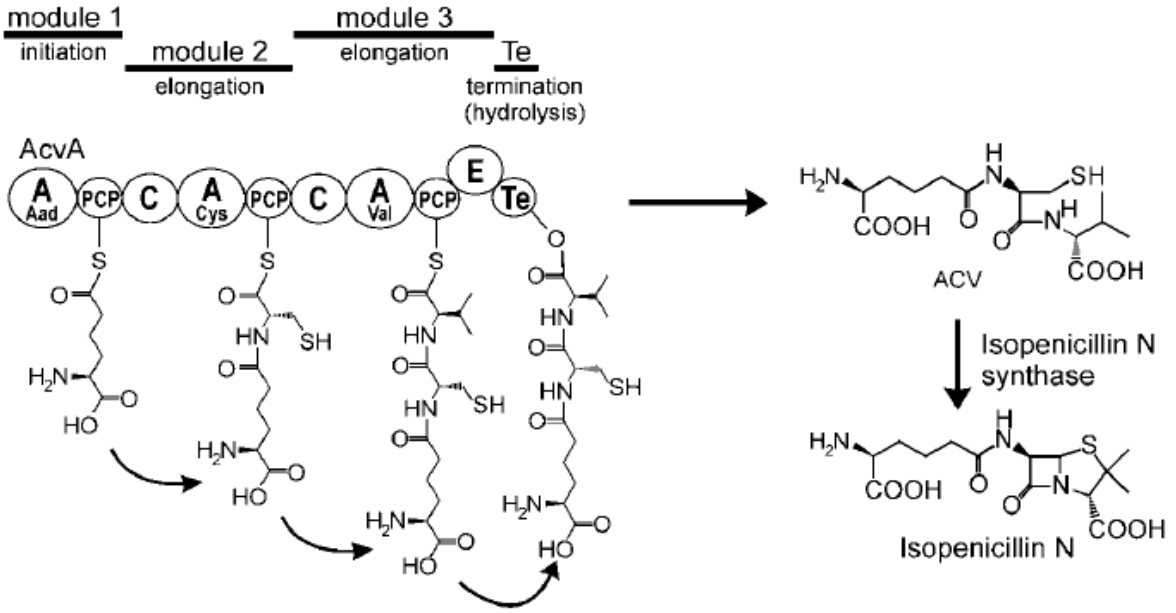

Figure 1.18 Example of the module and domain organization, A-PCP-(C-A-PCP)$_{n-1}$-TE, observed in a linear or type A NRPS, such as the ACV synthetase, which synthesizes the isopenicillin precursor ACV (Mootz *et al.,* 2002).

In the iterative or type B NRPSs, modules are used more than once in an efficient method of assembling multimeric peptide products. Instead of replicating the elongation modules, the entire NRPS is used repeatedly in an iterative manner to construct peptide chains that consist of recurring, short sequences. The product produced by the first cycle of synthesis is stalled on the C-terminal TE domain, thereby regenerating the NRPS for the assembly of the next product of the same amino acid sequence. Oligomerization of the completed product occurs on the TE domain and is released via hydrolysis or macrocyclization, with the latter being the preferred method. Examples of iterative NRPSs are those that synthesize enterobactin (Figure 1.19), gramicidin, enniatin and bacillibactin (Mootz *et al.,* 2002).

The final biosynthetic strategy is known as the nonlinear or type C NRPS, which can, in most cases, be identified from their primary sequence due to the fact that the identified module and

domain organization will deviate from the classical C-A-PCP domain organization. They are also characterized by their ability to incorporate small soluble molecules that are not covalently tethered to the NRPS template during synthesis, such as the involvement of amines in vibriobactin biosynthesis, instead of PCP-loaded amino acids (Finking & Marahiel, 2004; Keating *et al.,* 2002). Due to the fact that amines, such as norspermidine in vibriobactin biosynthesis, lack a carboxyl group necessary for their covalent attachment as a thioester, specialized C domains are employed to incorporate amines. The vibriobactin biosynthetic cluster also encodes an unusual tandem arrangement of Cy domains, in which one is responsible for heterocyclic ring formation, while the other's function remains unclear (Mootz *et al.,* 2002). Furthermore, unusual internal cyclizations are often associated with deviations from the standard domain organization observed in linear NRPSs, such as in the biosynthesis of bleomycin (Mootz *et al.,* 2002).

Other unusual biosynthetic strategies employed by type C NRPSs include utilizing a single domain to catalyse multiple different reactions, such as the cysteine-specific A domain in the yersiniabactin biosynthetic machinery, which is responsible for the loading of three different PCPs (Suo *et al.,* 2001), as well as the existence of a one-domain NRPS involved in the biosynthesis of the antibiotic novobiocin in *Streptomyces spheroides*. The enzyme, NovL, shares homology with acyl-adenylate forming enzymes of the same superfamily as A domains and catalyzes formation of the peptide bond between 3-dimethylallyl-4-hydroxybenzoic acid and 3-amino-4,7-dihydroxy-8-methyl coumarin in a reaction that is similar to those catalysed by acyl-CoA ligases. NovL is able to activate the carboxy acid of 3-dimethylallyl-4-hydroxybenzoic acid towards the acyl adenylate, but instead of transferring it onto a PCP domain, the acyl adenylate acts as the electrophile for the condensation with the amino group nucleophile of 3-amino-4,7-dihydroxy-8-methyl coumarin (Mootz *et al.,* 2002). It is also interesting to note that free-standing A domains are known to activate aromatic carboxylic acids and transfer (acylate) them to ArCPs, which are fused to isochorismate lyase, an enzyme involved in the synthesis of 2,3-dihydroxybenzoic acid, which is used as a starter unit in this type of synthetase (Crawford *et al.,* 2011; Schmoock *et al.,* 2005).

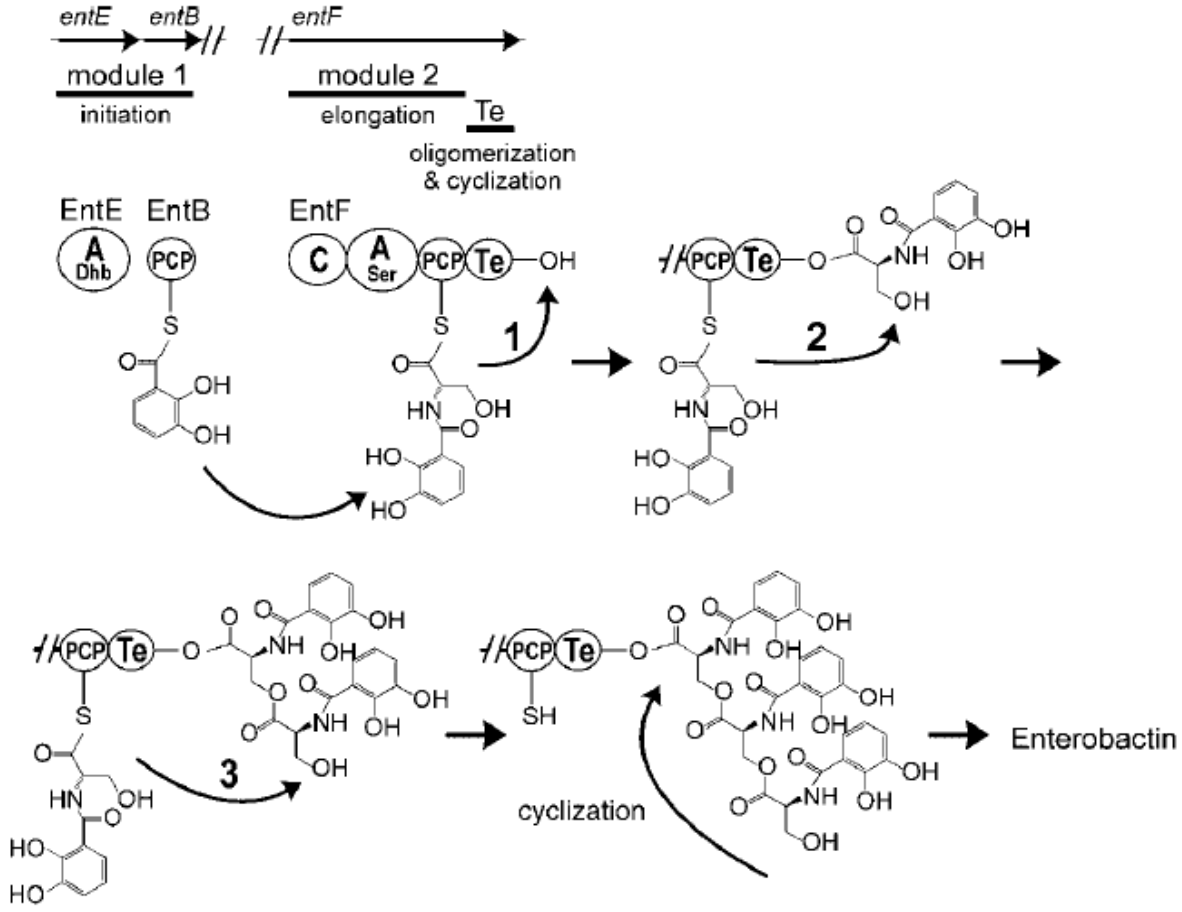

Figure 1.19 Example of the iterative or type B NRPS as observed in the biosynthesis of enterobactin. Three Dhb-Ser-*S*-ppant intermediates are produced on the two modules of the enterobactin NRPS and oligomerized and cyclized on the TE domain (Mootz *et al.,* 2002).

Additionally, fragments of NRPS assembly lines such as A-PCP didomains or separate, but adjacently encoded A and PCP domains are found in the absence of any other NRPS machinery (Fischbach & Walsh, 2006; Ullrich & Bender, 1994). It was discovered that these A-PCP didomains and freestanding A/PCP domains do in fact activate a specific L-amino acid as an aminoacyl adenylate, which then acts as a substrate for a partner enzyme to chemically modify the β-or γ-carbon of the thioesterified aminoacyl intermediate. Consequently, it has been inferred that the use of didomains or freestanding domains is a strategy to sequester a portion

of the pool of proteinogenic amino acids in order to modify them into a nonproteinogenic form, which can be used in the biosynthesis of extremely diverse secondary metabolites (Fischbach & Walsh, 2006).

The nonlinear NRPSs are impressive examples of how microbial producers are able to modify assembly-line organizations and operations in order to evolve novel combinations of the enzymatic components to generate new natural products (Fischbach & Walsh, 2006; Schwarzer *et al*., 2003).

1.4.4 Combinatorial biosynthesis and intermolecular communication

Combinatorial biosynthesis can be defined as ”the application of genetic engineering to modify biosynthetic pathways to natural products in order to produce new and altered structures using nature’s biosynthetic machinery” (Floss, 2006). In this approach, which was first demonstrated in work by David Hopwood and colleagues (Hopwood *et al.*, 1985), biosynthetic genes are cloned into an actinomycete host that does not naturally produce the antibiotic. Selected genes in the gene cluster are then deleted or replaced with carefully-chosen antibiotic biosynthetic genes from the producers of other actinomycete antibiotics, so that the host strain produces a chemical derivative of the original antibiotic (i.e. a hybrid antibiotic), dictated by the composition of the modified biosynthetic gene cluster (Weissman & Leadlay, 2005; Reynolds, 1998).

The order and domain composition of NRPS modules are the direct consequence of a meticulous selection process during evolution to synthesize a peptide molecule with particular bioactive properties. Once the logic and mechanisms of NRPS assembly had been investigated, interest developed in the rational manipulation of the NRPS template to synthesize novel peptide products (Sieber & Marahiel, 2005). Subsequently, over the past decade and a half, several strategies have been developed to redesign NRPS assembly lines in order to generate peptides with altered biological properties. Most of these exploits have focussed on the combinatorial approach of swapping or deleting domains or modules, as well as the mutation of individual domains to alter their specificity (Hahn & Stachelhaus, 2006). Although some of the domain swapping experiments have been successful in the production of altered products, a major stumbling block has been the fact that most of the chimeric NRPSs were functionally

impaired and produced small amounts of the product (Giessen & Marahiel, 2012). However, the identification of linker regions that bridge the individual modules within an NRPS and a better understanding of the domain borders has allowed these approaches to remain viable. The linker regions are about 15 amino acids in length and do not contain conserved residues. The fact that these linker regions are able to connect individual modules, in the absence of any conserved residues, has made them the ideal candidate for artificial module fusions. Indeed, *in vivo* biochemical studies on the tyrocidin synthetase showed that the fusion of modules within the linker regions resulted in the production of satisfying amounts of the altered peptide product (Schwarzer *et al.,* 2003).

Furthermore, the elucidation of the molecular basis for the selective interaction between NRPS modules via communication-mediating (COM) domains has also widened the toolbox for the combinatorial manipulation of NRPSs (Hahn & Stachelhaus, 2006). According to a study by Hahn & Stachelhaus (2004), a donor COM domain located at the C terminus of an aminoacyl- or peptidyl-donating NRPS and an acceptor COM domain located at the N terminus of the accepting cognate NRPS form a compatible set, which facilitates the intermolecular communication between adjacent modules. Recognition between the donor and acceptor domains are mediated by electrostatic and/or polar interactions between five pairs of residues located in both of the compatible COM domains. Additional studies have revealed that the swapping of COM domains can be used to manipulate the interaction of enzymes within the NRPS complex leading to the simultaneous production of different peptide products (Hahn & Stachelhaus, 2006).

The recombination of whole modules does represent a rather drastic intervention in the biosynthesis of NRPS-produced compounds, but eventually these approaches may lead to the combinatorial biosynthesis of innovative peptide antibiotics. Combinatorial biosynthesis has the potential to be a powerful way to generate antibiotic derivatives with greater antibacterial activity and/or improved pharmacokinetic properties. It has advantages over the chemical derivatisation of antibiotic molecules, since the bacterial biosynthetic enzymes carry out site-specific and enantiomer-specific reactions, by-passing the need to protect reactive functional groups in each reaction step in the chemical approach. This ensures that only the desired product is produced, which enhances the product yield.

1.5 RESEARCH AIMS

The actinomycete, *Streptomyces polyantibioticus* SPR^T, was isolated from soil collected from the banks of the Umgeni River, KwaZulu-Natal Province, South Africa, as part of an antibiotic-screening programme (Le Roes-Hill & Meyers, 2009). It exhibits antibiosis against various Gram-positive and Gram-negative bacteria, including *Enterococcus faecium* VanA (a vancomycin-resistant strain), *Mycobacterium aurum* A+ (which has a similar antibiotic susceptibility to *M. tuberculosis*, but is not pathogenic; Chung *et al.*, 1995) and *E. coli* ATCC 25922 (Le Roes, 2005). Of great interest, however, was its antibiotic activity against *M. tuberculosis* H37Rv^T, which prompted interest in its antibiotic production.

An antibacterial compound produced by *S. polyantibioticus* SPR^T was isolated and its structure was determined by X-ray crystallography and nuclear magnetic resonance (NMR) to be 2,5-diphenyloxazole (DPO; Figure 1.20) (Le Roes, 2005). An independent report by Giddens *et al.* (2005) confirmed that DPO (produced by chemical synthesis) showed activity against non-replicating persistent cells of *M. tuberculosis,* which are difficult to eradicate using traditional anti-TB drugs. The authors suggested that simple oxazole derivatives such as DPO may therefore be feasible options in the search for new antitubercular agents. DPO is currently only known to be synthesised chemically (Adrova *et al*., 1956), therefore its discovery from a biological origin is of great interest. DPO is unusual in that it is a 2,5-disubstituted oxazole, whereas most other disubstituted oxazoles from biological sources are 2,4-substituted.

Based on the structure of DPO, a biosynthetic scheme for the synthesis of this molecule was proposed, whereby a NRPS condenses a molecule of benzoic acid with 3-hydroxyphenylalanine. The resulting benzoyl-β-hydroxyphenylalanine is converted to DPO by heterocyclisation, oxidation and decarboxylation.

It seems likely that DPO is synthesized non-ribosomally by *S. polyantibioticus* SPR^T due to the presence of an oxazole. The NRPS responsible for the biosynthesis of DPO is proposed to have an A domain specific for the activation of phenylalanine or 3-hydroxyphenylalanine, plus ArCP, Cy, Ox, PCP and TE domains.

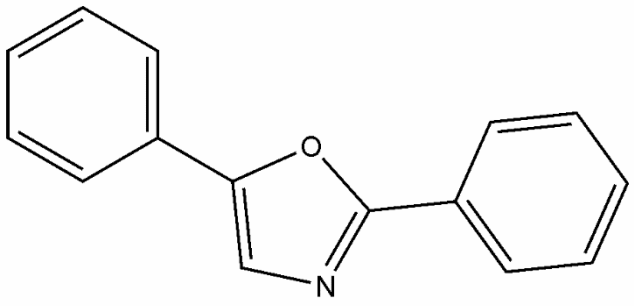

Figure 1.20 The structure of 2,5-diphenyloxazole (DPO).

The overall aim of this study was to confirm the production of DPO by *S. polyantibioticus* SPR^T and then to focus on the elucidation of the biosynthetic pathway involved in its synthesis. This consisted of the identification and partial characterization of the genes involved in the production of DPO to establish whether an NRPS is involved.

In order to determine whether the proposed biosynthetic pathway is correct, the genes coding for benzoic acid synthesis and the DPO NRPS had to be identified in the *S. polyantibioticus* SPR^T genome. The aims of the study were thus:

(i) To identify the genes involved in joining benzoic acid and phenylalanine/3-hydroxyphenylalanine in *S. polyantibioticus* SPR^T to create a 2,5-disubstituted oxazole.

(ii) To identify the genes responsible for the biosynthesis of benzoic acid in *S. polyantibioticus* SPR^T.

(iii) To propose a pathway for the biosynthesis of DPO in *S. polyantibioticus* SPR^T.

(iv) To develop a method to introduce recombinant plasmids into *S. polyantibioticus* SPR^T.

(v) To prove the involvement of genes in DPO biosynthesis through the creation of selected gene knockouts in *S. polyantibioticus* SPR^{T}.

Lastly, the genetic information from *S. polyantibioticus* SPR^{T} will help to lay the foundation for future combinatorial biosynthetic studies, using the genes responsible for DPO production to develop a range of oxazole derivatives that could be tested for enhanced antitubercular activity and therefore could become candidates for development as novel drugs to treat drug-resistant tuberculosis.

1.6 REFERENCE LIST

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*, **63(3):** 394-395.

Ansari, M.Z., Yadav, G., Gokhale, R.S. & Mohanty, D. (2004). NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Research*, **32:** 405-413.

Applebaum, P.C. (2006). The emergence of vancomycin-intermediate and vancomycin-resistant *Staphylococcus aureus*. *Clinical Microbiology and Infection,* **12(1):**16-23.

Ashforth, E. J., Fu, C., Liu, X., Dai, H., Song, F., Guo, H. & Zhang, L. (2010). Bioprospecting for antituberculosis leads from microbial metabolites, *Natural Product Reports*. **27**: 1709–1719.

Bachmann, B.O. & Ravel, J. (2009). Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Methods in Enzymology*, **458:** 181-217.

Balibar, C.J., Vaillancourt, F.H. & Walsh, C.T. (2005). Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chemistry & Biology*, **12(11):** 1189-1200

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

Baltz, R.H. (2014). MbtH homology codes to identify gifted microbes for genome mining. *Journal of Industrial Microbiology and Biotechnology,* **41(2):** 357-69.

Baranasic, D., Zucko, J., Diminic, J., Gacesa, R., Long, P.F., Cullum, J., Hranueli, D. & Starcevic, A. (2013). Predicting substrate specificity of adenylation domains of nonribosomal peptide synthetases and other protein properties by latent semantic indexing. *Journal of industrial microbiology & biotechnology*, **41(2):** 461-467.

Barry, C. E. 3rd, Boshoff, H. I., Dartois, V., Dick, T., Ehrt, S., Flynn, J., Schnappinger, D., Wilkinson, R. J. & Young, D. (2009). The spectrum of latent tuberculosis: rethinking the biology and intervention strategies *Nature Reviews Microbiology*, **7**: 845-855.

Bentley, S.D., Chater, K.F., Cerdeño-Tárraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, E., Quail, M., A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J. & Hopwood, D.A. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature,* **417**: 141-147.

Bérdy, J. (2005). Bioactive microbial metabolites. *The Journal of Antibiotics*, **58:** 1-26.

Bhate, D. S., Hulyalkar, R. K. & Menon, S. K. (1960). Isolation of iso-butyropyrrothine along with thiolutin and aureothricin from a *Streptomyces* sp. *Experientia,* **16:** 504.

Bloom, B.R. & Murray, C.J.L. (1992). Tuberculosis: commentary on a re-emergent killer. *Science*, **257**: 1055-1064.

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

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

Cane, D.E. & Walsh, C.T. (1999). The parallel and convergent universe of polyketide synthases and nonribosomal peptide synthetases. *Chemistry & Biology*, **6(12):** 319-325.

Cape Gateway, Department of Health (Provincial Government of the Western Cape). (2006). Last accessed January 2015. http://www.capegateway.gov.za/eng/pubs/public_info/M/115549.

Centers for Disease Control and Prevention (CDC). (2013). Last accessed January 2015. http://www.cdc.gov/drugresistance/threat-report-2013/.

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.

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(3):** 211-224.

Challis, G.L. (2014). Exploitation of the *Streptomyces coelicolor* A3(2) genome sequence for discovery of new natural products and biosynthetic pathways. *Journal of Industrial Microbiology and Biotechnology,* **41:** 219–232.

Chater, K.F. (1990). The Improving Prospects for Yield Increase by Genetic Engineering in Antibiotic-Producing Streptomycetes. *Nature Biotechnology.* **8**: 115 – 121.

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.

Churchyard, G.J., Mametja, L.D., Mvusi, L., Ndjeka, N., Hesseling, A.C., Reid, A., Babatunde, S. & Pillay, Y. (2014). Tuberculosis control in South Africa: Successes, challenges and recommendations. *South African Medical Journal*, **104(3):** 244-248.

Condurso, H.L. & Bruner, S.D. (2012). Structure and noncanonical chemistry of nonribosomal peptide biosynthetic machinery. *Natural Product Reports*, **29(10):** 1099-1110.

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(14):** 4174-4183.

Cragg, G.M. & Newman, D.J. (2013). Natural products: A continuing source of novel drug leads. *Biochimica et Biophysica Acta*, **1830(6):** 3670–3695.

Crawford, J.M., Portmann, C., Kontnik, R., Walsh, C.T. & Clardy, J. (2011). NRPS substrate promiscuity diversifies the Xenematides. *Organic Letters*, **13(19):** 5145-5147.

Dantas, G. & Sommer, M.O.A. (2014). How to Fight Back Against Antibiotic Resistance. *American Scientist*, **102:** 42-51

Davies, J. (2007). Microbes have the last word. *European Molecular Biology Organization Reports*, **8(7):** 616-621.

Davies, J. (2011). How to discover new antibiotics: harvesting the parvome. *Current Opinion in Chemical Biology*, **15:** 5–10.

Davies, J. (2012). Antibiotic discovery: then and now. *Microbiology Today*, **39(4):** 200-203.

Davies, J. & Ryan, K.S. (2012). Introducing the Parvome: Bioactive Compounds in the Microbial World. *ACS Chemical Biology*, **7:** 252−259

Demain, A.L. & Sanchez, S. (2009). Microbial drug discovery: 80 years of progress. *The Journal of Antibiotics,* **62:** 5–16.

Du, L., Chen, M., Sanchez, C. & Shen, B. (2000). An oxidation domain in the BlmIII non-ribosomal peptide synthetase probably catlyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in *Streptomyces verticillus* ATCC15003. *FEMS Microbiology Letters*, **189:** 171-175.

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.

Felnagle, E.A., Jackson, E.E., Chan, Y.A., Podevels, A.M., Berti, A.D., McMahon, M.D. & Thomas, M.G. (2008). Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Molecular Pharmaceutics*, **5(2):** 191-211.

Finking, R. & Marahiel, M.A. (2004). Biosynthesis of nonribosomal peptides. *Annual Review of Microbiology*, **58:** 453-488.

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

Fischbach, M.A. & Walsh, C.T. (2009). Antibiotics For Emerging Pathogens. *Science*, **325:** 1089-1093.

Floss, H.G. (2006). Combinatorial Biosynthesis – Potential and Problems. *Journal of Biotechnology,* **124(1):** 242–257.

Fusetani, N., Sugawara, T. & Matsunaga, S. (1991). Orbiculamide A: A novel cytotoxic cyclic peptide from a marine sponge Theonella sp. *Journal of The American Chemical Society,* **113:** 7812-7813.

Gaitatzis, N., Kunze, B. & Muller, R. (2001). In *vitro* reconstitution of the myxochelin biosynthetic machinery of the *Stigmatella aurantiaca* Sg a15: Biochemical characterization of a reductive release meachansim from nonribosomal peptide synthetases. *Proceedings of the National Academy of Sciences of the USA,* **98(20):** 11136-11141.

Galm, U., Wang, W., Wendt-Pienkowski, E., Yang, R., Liu, W., Tao, M., Coughlin, J.,M., Shen, B. (2008). *In Vivo* Manipulation of the Bleomycin Biosynthetic Gene Cluster in *Streptomyces verticillus* ATCC15003 Revealing New Insights into Its Biosynthetic Pathway. *The Journal of Biological Chemistry*, **283(42):** 28236-28245.

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.

GenBank. (2015). Last accessed 13 July 2015. http://www.ncbi.nlm.nih.gov/genome/browse/.

Giddens, A.C., Boshoff, H.I.M., Franzblau, S.G., Barry III, C.E. & Copp, B.R. (2005). Antimycobacterial natural products: synthesis and preliminary biological evaluation of the oxazole-containing alkaloid texaline. *Tetrahedron letters,* **46(43):** 7355-7357

Giessen, T.W. & Marahiel, M.A. (2012). Ribosome-independent biosynthesis of biologically active peptides: Application of synthetic biology to generate structural diversity. *FEBS Letters*, **586:** 2065-2075.

Gocht, M. & Marahiel, M.A. (1994). Analysis of core sequences in the D-Phe activating domain of the multifunctional peptide synthetase TycA by site directed mutagenesis. *Journal of Bacteriology*, **176:** 2654-2662.

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 USA*, **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 USA*, **103:** 275-280.

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., Goodman, R.M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, **5(10):** 245-249.

Harvey, A. (2000). Strategies for discovering drugs from previously unexplored natural products. *Drug Discovery Today*, **5(7):** 294-300.

Hopwood D.A., Malpartida, F., Kieser, H.M., Ikeda, H., Duncan, J., Fujii, I., Rudd, B.A.M., Floss, H.G. & Omura, S. (1985). Production of "hybrid" antibiotics by genetic engineering. *Nature*, **314:** 642–644.

Howden, B.P., Davies, J.K., Johnson, P.D.R., Stinear, T.P. & Grayson, M.L. (2010). Reduced Vancomycin Susceptibility in *Staphylococcus aureus*, including Vancomycin-Intermediate and Heterogeneous Vancomycin-Intermediate Strains: Resistance Mechanisms, Laboratory Detection, and Clinical Implications. *Clinical Microbiology Reviews*, **23(1):** 99–139.

Hugenholtz, P., Goebel, B.M. & Pace, N.R. (1998). Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity. *Journal of Bacteriology*, **180(18):** 4765-4774.

Inahashi, Y., Iwatsuki, M., Ishiyama, A., Namatame, M., Nishihara-Tsukashima, A., Matsumoto, A., Hirose, T., Sunazuka, T., Yamada, H., Otoguro, K., Takahashi, Y., Omura, S. & Shiomi, K. (2011). Spoxazomicins A-C, novel antitrypanosomal alkaloids produced by an endophytic actinomycete, *Streptosporangium oxazolinicum* K07-0460(T). *Journal of Antibiotics (Tokyo),* **64(4):** 303-307.

Jackman, J.E., Fierke, C.A., Tumey, L.N., Pirrung, M., Uchiyama, T., Tahir, S.H., Hindsgaul, O. & Raetz, C.R. (2000). Antibacterial agents that target lipid A biosynthesis in gram-negative bacteria. Inhibition of diverse UDP-3-O-(r-3-hydroxymyristoyl)-n-acetylglucosamine deacetylases by substrate analogs containing zinc binding motifs. *Journal of Biological Chemistry*, **275(15):** 11002-11009.

John, J.E. (2009). Natural products-based drug discovery: some bottlenecks and considerations. *Current Science*, **96(6):** 753.

Keating, T.A. & Walsh, C.T. (1999). Initiation, elongation and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Current Opinion in Chemical Biology*, **3:** 598-606.

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.

Khurana, P., Gokhale, R.S. & Mohanty, D. (2010). Genome scale prediction of substrate specificity for acyl adenylate superfamily of enzymes based on active site residue profiles. *BMC Bioinformatics*, **11:** 57.

Kleinkauf, H. & von Döhren, H. (1990). Nonribosomal biosynthesis of peptide antibiotics. *European Journal of Biochemistry*, **192(1):** 1–15.

Kobayashi, J., Itagaki, F., Shigemori, I., Takao, T. & Shimonishi, Y. (1995). Keramamides E, G, H, and J, new cyclic peptides containing an oxazole or a thiazole ring from a *Theonella* sponge. *Tetrahedron*, **51(9):** 2525-2532.

Kondo, K., Higuchi, Y., Sakuda, S., Nihira, T. & Yamada, Y. (1989). New virginiae butanolides from *Streptomyces virginiae*. *The Journal of antibiotics*, **42(12):** 1873-1876.

Konz, D., Klens, A., Schbrgendorfer, K., Marahiel, M.A. (1997). The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chemistry & Biology*, **4:** 927-937.

Konz, D. & Marahiel, M.A., (1999). How do peptide synthetases generate structural diversity? *Chemistry & Biology*, **6(2):** 39-48.

Kopp, F. & Marahiel, M.A. (2007). Macrocyclization Strategies in Polyketide and Nonribosomal Peptide Biosynthesis. *Natural Product Reports*, **24(4):** 735-749.

Koul, A., Arnoult, E., Lounis, N., Guillemont, J. & Andries, K. (2011). The challenge of new drug discovery for tuberculosis. *Nature*, **469**: 483-490.

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

Lamichhane, G. (2011). Novel targets in *M. tuberculosis*: search for new drugs. *Trends in Molecular Medicine*, **17(1):** 25-33.

Land, M., Hauser, L., Jun, S., Nookaew, I., Leuze, M.R., Ahn, T., Karpinets, T., Lund, O., Kora, G., Waasenaar, T., Poudel, S. & Ussery, D.W. (2015). Insights from 20 years of bacterial genome sequencing. *Functional & Integrative Genomics*, DOI 10.1007/s10142-015-0433-4.

Lautru, S. & Challis, G.L. (2004). Substrate recognition by nonribosomal peptide synthetase multi-enzymes. *Microbiology*, **150:** 1629-1636.

Lautru, S., Deeth, R.J., Bailey, L.M., Challis, G.L. (2005). Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nature Chemical Biology*, **1:** 265-269.

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.

Li, J. & Jensen, S.E. (2008). Nonribosomal biosynthesis of fusaricidins by Paenibacillis polymyxa PKB1 involves direct activation of a D-amino acid. *Chemistry & Biology*, **15(2):** 118-127.

Li, J.W. & Vederas, J.C. (2009). Drug discovery and natural products: end of an era or an endless frontier? *Science*, **325(5937):** 161-165.

Li, M.H.T., Ung, P.M.U., Zajkowski, J., Garneau-Tsodikova, S. & Sherman, D.H. (2009). Automated genome mining for natural products. *BMC Bioinformatics*, **10:** 185.

Liu, X., Chen, C., He, W., Huang, P., Liu, M., Wang, Q., Guo, H., Bolla, K., Lu, Y., Song, F., Dai, H., Liu, M. & Zhang, L. (2012). Exploring anti-TB leads from natural products library originated from marine microbes and medicinal plants. *Antonie van Leeuwenhoek,* **102:** 447-461.

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

Marahiel, M.A. (1997). Protein templates for the biosynthesis of peptide antibiotics. *Chemistry & Biology*, **4(8):** 561-567.

Marahiel, M.A. (2009). Working outside the protein-synthesis rules: insights into non-ribosomal protein synthesis. *Journal of Peptide Science*, **15:** 799-807.

May, J.J., Kessler, N., Marahiel, M.A. & Stubbs, M.T. (2002). Crystal structure of DhbE, an archetype for aryl acid activating domains of nonribosomal peptide synthetases. *Proceedings of the National Academy of Science of the USA*, **99(19):** 12120-12125.

Miao, V. & Davies, J. (2010). Actinobacteria: the good, the bad, and the ugly. *Antonie van Leeuwenhoek*, **98:** 143–150.

Michels, P.C., Khmelnitsky, Y.L., Dordick, J.S. & Clark, D.S. (1998). Combinatorial biocatalysis: a natural approach to drug discovery. *Trends in Biotechnology*, **16(5):** 210-215.

Milne, J.C., Eliot, A.C., Kellher, N.L. & Walsh, C.T. (1998). ATP/GTP hydrolysis is required for oxazole and thiazole biosynthesis in the peptide antibiotic microcin B17. *Biochemistry*, **37(38):** 13250-13261.

Mootz, H.D., Schwarzer, D. & Marahiel, M.A. (2002). Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem,* **3**: 490-504.

Nett, M., Ikedab, H. & Moore, B., S. (2009). Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Natural Product Reports*, **26:** 1362–1384.

Padmavathi, V., Mahesh, K., Rangayapalle, D., Venkata, C., Deepti, D. & Reddy, G.S. (2009). Synthesis and biological activity of a new class of sulfone linked bis(heterocycles). *Arkivoc*, **2:** 195-208.

Pavela-Vrancic, M., van Liempt, H., Pfeifer, E., Freist, W. & von Dohren, H. (1994). Nucleotide binding by multienzyme peptide synthetases. *European Journal of Biochemistry*, **220(2):** 535-542.

Petrini, B. & Hoffner, S. (1999). Drug-resistant and multidrug-resistant tubercle bacilli. *International Journal of Antimicrobial Agents*, **13**: 93-97.

Prieto, C., Garcia-Estrada, C., Lorenzana, D. & Martin, J.F. (2012). NRPSsp: non-ribosomal peptide synthase substrate predictor. *Bioinformatics*, **28(3):** 426-427.

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.

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.

Rastogi, N., Abaul, J., Goh, K.S., Devallois, A., Philogene, E. & Bourgeois, P. (1998). Antimycobacterial activity of chemically defined natural substances from the Caribbean flora in Guadeloupe. *FEMS Immunology and Medical Microbiology*, **20(4):** 267-273.

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.

Reynolds, K. A. (1998). Combinatorial biosynthesis: lesson learned from nature. *Proceedings of the National Academy of Sciences of the USA*, **95**: 12744-12746.

Röttig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C. & Kohlbacher, O. (2011). NRPSpredictor2 – a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Research*, 1-6.

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.

Rivers, E. C. & Mancera, R. L. (2008). New anti-tuberculosis drugs in clinical trials with novel mechanisms of action. *Drug Discovery Today*, **13**: 1090-1098.

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

Schmoock, G., Pfennig, F., Jewiarz, J., Schlumbohm, W., Laubinger, W., Schauwecker, F. & Keller, U. (2005). Functional Cross-talk between Fatty Acid Synthesis and Nonribosomal Peptide Synthesis in Quinoxaline Antibiotic-producing Streptomycetes. *The Journal of Biological Chemistry*, **280(6):** 4339-4349.

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

Seow, K.T., Meurer, G., Gerlitz, M., Wendt-Pienkowski, E., Hutchinson, C.R. & Davies, J. (1997). A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. *Journal of Bacteriology*, **179(23):** 7360-7368.

Sieber, S.A. & Marahiel, M.A. (2005). Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chemical Reviews*, **105:** 715-738.

Spratt, B.G. (1994). Resistance to antibiotics mediated by target alterations. *Science*, **264:** 388.

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.

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

Subramani, R. & Aalbersberg, W. (2013). Culturable rare Actinomycetes: diversity, isolation and marine natural product discovery. *Applied Microbiology & Biotechnology,* **97:** 9291-9321.

Suo, Z., Tseng, S.A. & Walsh, C.T. (2001). Purification, priming, and catalytic acylation of carrier protein domains in the polyketide synthase and nonribosomal peptidyl synthetase modules of the HMWP1 subunit of yersiniabactin synthetase. *Proceedings of the National Academy of Science of the USA*, **95(1):** 99-104.

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

Tao, M., Wang, L., Wendt-Pienkowski, E., George, N.P., Galm, U., Zhang, G., Coughlin, J.M. & Shen, B. (2007). The tallysomycin biosynthetic gene cluster from *Streptoalloteichus hindustanus* E465-94 ATCC 31158 unveiling new insights into the biosynthesis of the bleomycin family of antitumor antibiotics. *Molecular Biosystems*, **3(1):** 60-74.

Udwadia, Z., F. (2012). MDR, XDR, TDR tuberculosis: ominous progression. *Thorax,* **67(4):** 286-288.

Udwadia Z., F., Vendoti, D. (2013). Totally drug-resistant tuberculosis (TDR-TB) in India: every dark cloud has a silver lining. *Journal of Epidemiology and Community Health.* 67(6): 471-472.

Ullrich, M. & Bender, C.L. (1994). The biosynthetic gene cluster for coronamic acid, an ethylcyclopropyl amino acid, contains genes homologous to amino acid-activating enzymes and thioesterases. *Journal of Bacteriology*, **176:** 7574-7586.

Van Lanen, S. & Shen, B. (2006). Microbial genomics for the improvement of natural product discovery. *Current Opinion in Microbiology*, **9**: 252-262.

Velasco, A., Acebo, P., Gomez, A., Schleissner, C., Rodriguez, P., Aparicio, T., Conde, S., Munoz, R., de la Calle, F., Garcia, J.L. & Sanchez-Puelles, J.M. (2005). Molecular characterization of the safracin biosynthetic pathway from *Pseudomonas fluorescens* A2-2: designing new cytotoxic compounds. *Molecular Microbiology*, **56(1):** 144-154.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G., F., Chater, K., F. & van Sinderen, D. (2007). Genomics of *Actinobacteria*: Tracing the Evolutionary History of an Ancient Phylum. *Microbiology and Molecular Biology Reviews*. **71(3):** 495–548.

Verma, M., Lal, D., Kaur, J., Saxena, A., Kaur, J., Anand, S. & Lal, R. (2013). Phylogenetic analyses of phylum Actinobacteria based on whole genome sequences. *Research in Microbiology*. **164:** 718-728.

Von Döhren, H., Dieckmann, R. & Pavela-Vrancic, M. (1999). The nonribosomal code. *Chemistry & Review*, **6(10):** 273-279.

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

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. (2004). Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science*, **303:** 1805-1810.

Walsh, C.T. & Fischbach, M.A. (2010). Natural Products Version 2.0: Connecting Genes to Molecules. *Journal of the American Chemical Society*, **132(8):** 2469-2493.

Wang, L., Yun, B., George, N.P., Wendt-Pienkowski, E., Galm, U., Oh, T., Coughlin, J.M., Zhang, G., Tao, M. & Shen, B. (2007). The Glycopeptide Antitumor Antibiotic Zorbamycin from *Streptomyces flavoviridis* ATCC 21892: Strain Improvement and Structure Elucidation. *Journal of Natural Products*, **70(3):** 402-406.

Watve, M.G., Tickoo, R., Jog, M.M., Bhole, B.D. (2001). How many antibiotics are produced by the genus Streptomyces? *Archives of microbiology*, **176(5):** 386-90.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbacj, M.A., Muller, R., Wohlleben, W., Breitling, R., Takano, E., & Medema, M.H. (2015). antiSMASH 3.0 — a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*, doi: 10.1093/nar/gkv437

Weissman, K. J. & Leadlay, P. F. (2005). Combinatorial biosynthesis of reduced polyketides. *Nature Reviews: Microbiology*; **3**: 925-936.

World Health Organisation (WHO). (2014). Last accessed 12 January 2015. http://www.who.int/mediacentre/factsheets/fs104/en/.

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.

Xia, S., Ma, Y., Zhang, W., Yang, Y., Wu, S., Zhu, M., Deng, L., Li, B., Liu, Z. & Qi, C. (2012). Identification of Sare0718 as an alanine-activating adenylation domain in marine actinomycete *Salinispora arenicola* CNS-205. *PLoS One*, **7(5):** 1-12.

Yeh, V.S.C. (2004). Recent advances in the total synthesis of oxazole-containing natural products. *Tetrahedron* **60**: 11995 – 12042.

Yoshioka, T., Mohri, K., Oikawa, Y. & Yonemitsu, 0. (1981). Metabolism of *Streptoverticilium olivoreticuli* and Streptomyces *cinnamomeus*. *Journal of Chemical Research*, 194.

Young, D. B., Gideon, H. P. & Wilkinson, R. J. (2009). Eliminating latent tuberculosis. *Trends in Microbiology*, **17**: 183-188.

CHAPTER 2

EARLY ATTEMPTS AT ISOLATING THE GENE CLUSTER RESPONSIBLE FOR DPO BIOSYNTHESIS IN *STREPTOMYCES POLYANTIBIOTICUS* SPR^{T}

CHAPTER 2

EARLY ATTEMPTS AT ISOLATING THE GENE CLUSTER RESPONSIBLE FOR DPO BIOSYNTHESIS IN *STREPTOMYCES POLYANTIBIOTICUS* SPR^{T}

CHAPTER 2

EARLY ATTEMPTS AT ISOLATING THE GENE CLUSTER RESPONSIBLE FOR DPO BIOSYNTHESIS IN *STREPTOMYCES POLYANTIBIOTICUS* SPR^T

2.1 ABSTRACT

To determine whether the hypothesis pertaining to the DPO biosynthetic pathway is correct, the genes coding for benzoic acid synthesis and the DPO NRPS in the *S. polyantibioticus* SPR^T genome need to be identified. As benzoic acid synthesis is unusual in bacteria, the initial focus was on detecting an NRPS gene in *S. polyantibioticus* SPR^T with an A domain exhibiting a binding pocket substrate specificity for phenylalanine or 3-hydroxyphenylalanine, as this would most likely indicate an NRPS involved in the biosynthesis of DPO. This first step was achieved by performing PCR amplification of NRPS A domain sequences, which led to the identification of 12 unique NRPS A domains. One of the A domains was specific for phenylalanine. This phenylalanine–specific domain (designated A-18) was used as a probe in Southern hybridization experiments in an effort to identify larger DNA fragments surrounding the A-18 NRPS gene cluster. However, no further sequence information could be obtained. The identification of Cy domains in the genome of *S. polyantibioticus* SPR^T was analysed in a similar fashion to the A domains, as the presence of a Cy domain would be an indicator of heterocyclization across the amide bond, a reaction required for the formation of an oxazole ring. PCR primers specific for oxazole- and thiazole-producing Cy domains were designed in an effort to identify Cy-domain-containing NRPS genes in *S. polyantibioticus* SPR^T. The Cy-domain primers were unsuccessful. Similar attempts were made at identifying genes involved in the synthesis of benzoic acid in *S. polyantibioticus* SPR^T, as the presence of a phenylalanine ammonia-lyase (PAL) encoding gene would allow the synthesis of benzoyl-CoA for DPO

production. PCR primers were utilised to amplify the *encP* gene (encoding PAL) from '*Streptomyces maritimus*' DSM 41777^T and Southern hybridization confirmed the presence of *encP* in this strain. In contrast, an *encP* homologue could not be detected in the genome of *S. polyantibioticus* SPR^T. As *S. polyantibioticus* SPR^T appears not to possess a PAL gene, it was suggested that this organism may produce benzoyl-CoA via the phenylacetate (PA) pathway involved in the degradation of phenylalanine. The initial step of the PA pathway is catalyzed by a PA-CoA ligase and degenerate primers were used to amplify a homologue of the gene, *paaK*, encoding this enzyme in *S. polyantibioticus* SPR^T. The presence of a *paaK* gene in *S. polyantibioticus* SPR^T was confirmed, suggesting that the PA pathway is present.

2.2 INTRODUCTION

2.2.1 BIOSYNTHESIS OF DPO

DPO is best known for its properties as a scintillator and as a high efficiency luminophore (Ionescu *et al.,* 2005; Semenova *et al.,* 2004). It is produced for industry by chemical synthesis, whereby it is used extensively in laser dyes and as a light activator for liquids (Semenova *et al.,* 2004; Adrova *et al*., 1957). However, the biological synthesis of DPO has never been reported before and therefore its production by *S. polyantibioticus* SPR^T is of great interest.

A biochemical pathway for the production of DPO by *S. polyantibioticus* SPR^T has been hypothesized whereby the compound is synthesized from starter units of benzoic acid (Figure 2.1A) and phenylalanine or 3-hydroxyphenylalanine (Figure 2.1B). Nucleophilic attack by the α-amino group of 3-hydroxyphenylalanine on the electrophilic carboxyl group of benzoic acid would result in the formation of an amide bond and the reaction intermediate benzoyl-β-hydroxyphenylalanine (Figure 2.1C). This molecule is postulated to undergo a heterocyclization or cyclodehydration reaction to create an oxazoline intermediate, which would subsequently be oxidized to form an oxazole, 4-carboxy 2,5-diphenyloxazole (Figure 2.1D), by the nucleophilic attack of the phenylalanine β-hydroxy group on the carbonyl group of the amide. The decarboxylation of the oxazole moiety would result in the formation of DPO as the final product (Figure 2.1E).

The formation of the amide bond and subsequent heterocyclization reaction are assumed to be catalyzed by an NRPS. The NRPS responsible for the biosynthesis of DPO is proposed to have an A domain, plus ArCP, Cy, Ox, PCP and TE domains (Figure 2.2). The A domain would recognise either 3-hydroxyphenylalanine or phenylalanine as the substrate and tether it to the PCP domain. Although it is unclear whether the A domain would bind phenylalanine or β-hydroxyphenylalanine, it seems more likely that phenylalanine would be the substrate, with subsequent hydroxylation forming 3-hydroxyphenylalanine. If this is correct, then the DPO NRPS would be expected to have an additional domain for a P450 monooxygenase to allow for the β-hydroxylation of phenylalanine before the heterocyclization reaction (this P450 monooxygenase domain is not shown in Figure 2.2). The hydroxylation activity could also be provided *in trans*. The ArCP domain would be responsible for loading benzoic acid.

The PCP domain would be responsible for keeping the 3-hydroxyphenylalanine intermediate bound to the enzymatic machinery, while the ArCP domain would perform the identical function in keeping the benzoyl intermediate bound. The Cy domain is proposed to perform the dual function of facilitating the condensation between benzoic acid and 3-hydroxyphenylalanine and carrying out the heterocyclization of benzoyl-β-hydroxyphenylalanine to 4-carboxy 2,5-diphenyloxazole.

Furthermore, the Ox domain, which may exist as an external tailoring enzyme, would be necessary for the oxidation of the hydrolytically labile oxazoline to form the oxazole moiety in the heterocyclization reaction (the oxidation of the oxazolidine intermediate is not shown in Figure 2.1). A decarboxylase reaction would be required to convert 4-carboxy 2,5-diphenyloxazole to DPO. This decarboxylating activity could also be supplied *in trans*, as occurs in curacin A biosynthesis in *L. majuscula* (Gu *et al*., 2006). Finally, the TE domain would be involved in the release of DPO by breaking the covalent linkage between DPO and the 4'-phosphopantetheine (4'-PP) thiol arm.

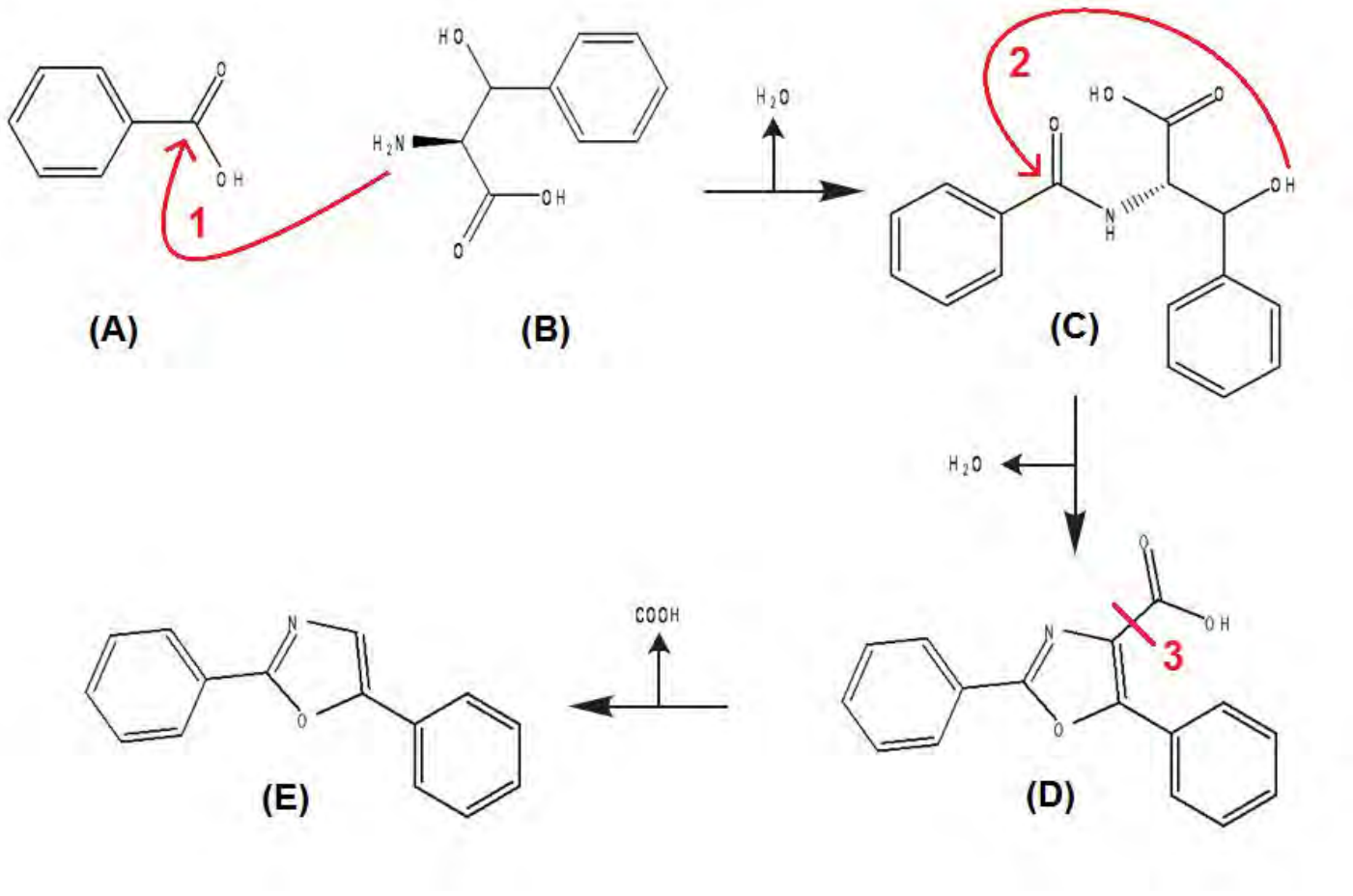

Figure 2.1 Proposed reaction scheme for the synthesis of DPO in *S. polyantibioticus* SPR^{T}. **(A)** Benzoic Acid, **(B)** 3-Hydroxyphenylalanine, **(C)** Benzoyl-β-Hydroxyphenylalanine, **(D)** 4-Carboxy 2,5-Diphenyloxazole, **(E)** DPO. Reactions are shown by red arrows or lines, **1**- Amide bond formation, **2**- Heterocyclization, **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)** (Stegmann, 2011). Note that the oxazoline intermediate formed by the heterocyclization reaction is not shown.

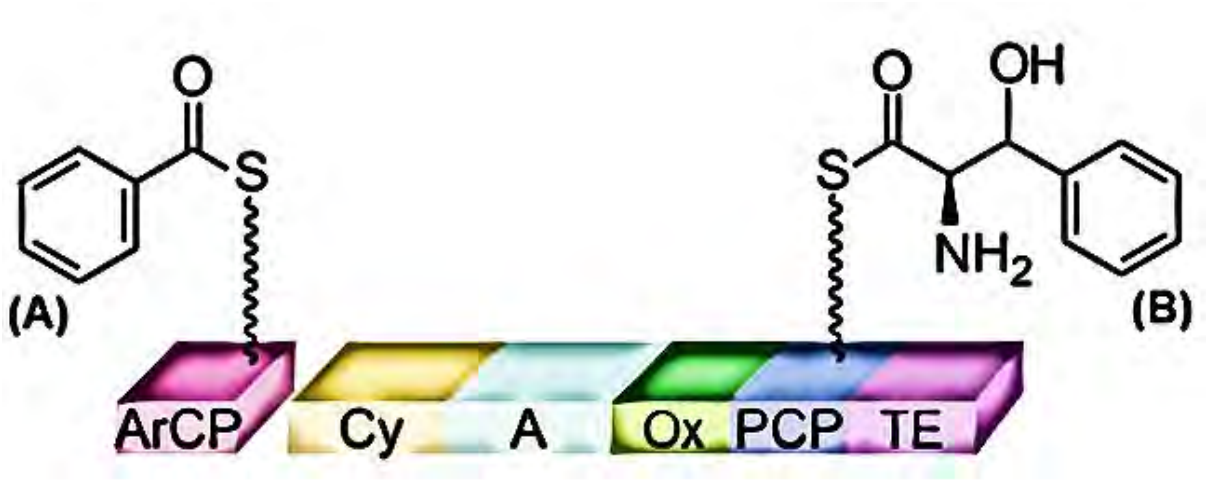

Figure 2.2 The postulated NRPS module arrangement involved in the biosynthesis of DPO. The ArCP and PCP domains are activated by the transfer of 4'-PP to the conserved serine residue on each domain, which allows the covalent binding of benzoate **(A)** and 3-hydroxyphenylalanine **(B)**, as activated thioester derivatives. Domains: **ArCP** - aryl carrier protein, **Cy** – heterocyclization domain, **Ox** – oxidation domain, **PCP** - peptidyl carrier protein domain, **TE** – thioesterase domain. (Stegmann, 2011).

2.2.2 ISOLATION OF GENES INVOLVED IN DPO BIOSYNTHESIS

To assess whether the proposed biosynthetic pathway is correct, the genes required for the synthesis of benzoic acid and the DPO NRPS, which are expected to be clustered on the *S. polyantibioticus* SPR^T genome, had to be identified.

The identification of an A domain in the genome of *S. polyantibioticus* SPR^T, with a binding pocket substrate specificity for phenylalanine, would indicate an NRPS that might be responsible for the biosynthesis of DPO. PCR amplification, using suitable degenerate primers, would allow for the detection of phenylalanine-specific A domains in *S. polyantibioticus* SPR^T. The substrate specificities of these PCR-amplified A domains could be determined by comparison with the binding-pocket specificities of A domains for which the amino-acid substrates are known. Larger fragments of the DPO biosynthetic gene cluster could then be

detected by Southern hybridization and further sequencing would reveal the remainder of the genes constituting the cluster. Identification of the DPO biosynthetic gene cluster would add to the handful of oxazole biosynthetic gene clusters characterised from actinomycetes (Pulsawat *et al.,* 2007; Zhao *et al.,* 2006; Onaka *et al.,* 2005).

The identification of a Cy domain in *S. polyantibioticus* SPR^T may be analysed in a similar fashion to the A domains and their presence would indicate NRPS enzymes involved in making oxazoles or thiazoles. PCR primers specific for oxazole- and thiazole-producing Cy domains may be used to amplify such domains.

In addition to searching for NRPS genes, the genes for benzoic acid biosynthesis need to be identified. The presence of an orthologue of the *encP* gene, coding for phenylalanine ammonia-lyase (PAL), in the genome of *S. polyantibioticus* SPR^T would be of great interest, as PAL is a key enzyme in the generation of benzoyl coenzyme A (benzoyl CoA) via *trans*-cinnamic acid in a plant-like manner from phenylalanine in the biosynthesis of enterocin in '*Streptomyces maritimus*'strain DSM 41777^T (Figure 2.3) (Moore *et al*., 2002; Hertweck & Moore, 2000). In spite of its wide occurrence in fungi and plants, benzoic acid is an extremely rare bacterial metabolite that has only been described in a few bacterial biosynthetic systems and as a central intermediate of anaerobic aromatic molecule metabolism (Xiang & Moore, 2003). The existence of an *encP* orthologue in the genome of *S. polyantibioticus* SPR^T could be elucidated using suitable PCR primers. Such an *encP* orthologue would be a strong candidate for involvement in the biosynthesis of DPO.

If benzoic acid is not synthesized via a PAL-mediated pathway in *S. polyantibioticus* SPR^T, it could be synthesized via a novel variation on the phenylacetate pathway for the degradation of phenylalanine. Besides the aerobic process catalysed by PAL in '*S. maritimus*' strain DSM 41777^T, other aerobic and anaerobic pathways for the production of benzoyl-CoA and derivatives do exist, such as in the β-proteobacterium *Azoarcus evansii*, in which phenylacetic acid is degraded via an anaerobic mechanism to benzoyl-CoA (Gescher *et al*., 2005). It has been reported that phenylacetate-coenzyme A ligase (PA-CoA ligase) catalyses the initial reaction in this pathway, which involves the activation of PA to PA-CoA. The gene coding for PA-CoA ligase, *paaK*, has also been identified in *Streptomyces* species (Pometto and Crawford, 1985). Identification of a similar gene to *paaK* in *S. polyantibioticus* SPR^T could indicate its

ability to synthesise benzoyl-CoA aerobically from PA (Figure 2.3B). Amino acid catabolism has been linked to antibiotic synthesis in the production of macrolides in *Streptomyces ambofaciens* and *Steptomyces fradiae* and therefore it is possible that an aromatic amino acid degradation pathway could be involved in the production of a benzoic acid intermediate for DPO biosynthesis (Tang *et al.,* 1994). Alternatively, since the chorismate pathway is a common pathway for generating molecules with benzene rings, such as shikimic acid, in bacteria (and other organisms), *S. polyantibioticus* SPR^T could perhaps also use a novel variant of one of the aromatic biosynthetic pathways to generate benzoic acid (Figure 2.3D).

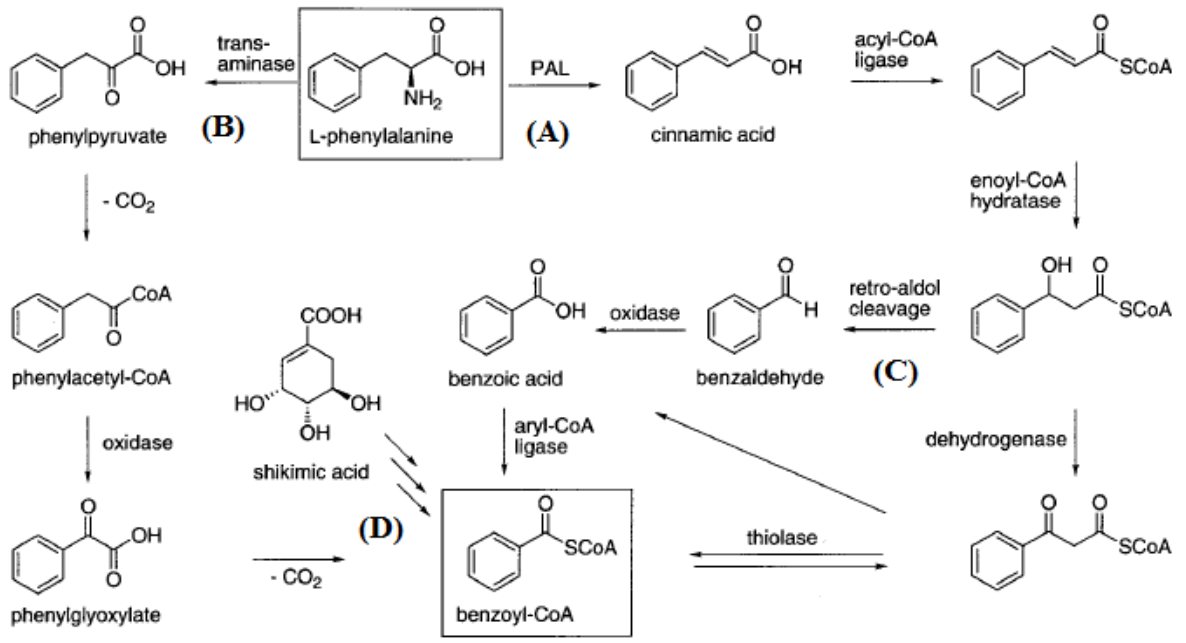

Figure 2.3 Biosynthetic pathways involved in the synthesis of benzoyl-CoA in '*S. maritimus*' and plants (route A), in anaerobic bacteria (route B), in fungi (route C) and from shikimic acid (route D) (Hertweck & Moore, 2000).

2.3 MATERIALS AND METHODS

2.3.1 STRAINS, MEDIA AND GROWTH CONDITIONS

'*S. maritimus*' DSM 41777^T, *Streptomyces virginiae* NRRL B-1446^T, *Streptomyces avermitilis* MA-4680^T, *Streptomyces coelicolor* A3(2) and *Streptosporangium oxazolinicum* JCM 17388^T were grown in yeast extract-malt extract broth (YEME; International *Streptomyces* Project medium 2) (Shirling & Gottlieb, 1966) and *S. polyantibioticus* SPR^T was grown in *Streptosporangium* Medium (SM) (Pfefferle *et al.*, 2000) for 7 days at 30^{o}C 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 (gDNA) was extracted from bacterial cell mass using the gDNA extraction method of Wang *et al.* (1996), with the following modifications: 25 mg lysozyme/ml was used instead of 5 mg/ml and the cells were incubated in the lysozyme buffer at 37 ^{o}C overnight instead of for 30 min; 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 3 M sodium acetate (pH 5.2) and one volume of room-temperature isopropanol. The concentration of each gDNA sample was quantitated spectrophotometrically using the Nanodrop® ND-1000 Spectrophotometer (Coleman Technologies Inc., USA) and analysed by agarose gel electrophoresis. Visualization was performed with a Gel Doc XR (Bio-Rad Laboratories Inc., USA) system and Quantity One Version 4.5.2 Software in order to assess the quantity and integrity of the DNA samples. To confirm that the extracted gDNA was of good quality, the 16S rRNA gene was PCR amplified as detailed in Cook & Meyers (2003). gDNA samples were stored at 4 °C.

2.3.3 PRIMER DESIGN

All of the degenerate primers used in the early attempts to isolate the NRPS gene cluster are listed in Table 2.1, accompanied by their sequences and binding positions. The A3F and A7R primer set was designed to bind to the conserved motifs of actinomycete A domains, namely

A3 and A7, respectively (Ayuso-Sacido & Genilloud, 2005). This primer set was used to amplify NRPS A domain fragments of approximately 700 bp in size.

A multiple amino acid sequence alignment of the Cy-domain regions of one oxazole and four thiazole producing NRPSs from various bacterial strains (Table 2.2) obtained from the GenBank database of the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) was performed using DNAMAN version 4.13 (Lynnon Biosoft, USA) in order to design the degenerate primer set CyF and CyR. The Cy domain primer set amplifies a product of 1050 bp from *S. polyantibioticus* SPR^T.

Another two Cy-domain forward primers, VVFTS CycF and QTPQV CycF, were designed based on a multiple sequence alignment of the amino acid sequences used in the initial primer design (Table 2.2) plus three new Cy domain sequences from *S. coelicolor* strain A3(2) (GenBank accession number: NP_631722), *Streptomyces roseosporus* strain NRRL 111379 (GenBank accession number: ZP_04712035) and *Streptomyces griseus* subsp. *griseus strain* NBRC 11350 (GenBank accession number: WP_012379765). These forward primers were used in conjunction with the A domain reverse primer, A7R, and were expected to amplify a PCR product of 1449 bp.

The primer set, PALHAL_F and PALHAL_R, was designed using DNAMAN from the consensus sequence of a multiple sequence alignment of the PAL amino acid sequence from '*S. maritimus*' DSM 41777^T (GenBank accession number: AF254925) and the histidine ammonia lyase (HAL) amino acid sequences from eleven *Streptomyces* species obtained from the GenBank database (http://www.ncbi.nlm.nih.gov/) (Table 2.3). An amplification product of approximately 350 bp was expected to be amplified from *S. polyantibioticus* SPR^T. The EncP-F and EncP-R primer set was designed to amplify a 721 bp *encP* fragment from the biosynthetic gene cluster for benzoyl-CoA-derived enterocin in '*S. maritimus*' DSM 41777^T (AF254925) (Stegmann, 2011).

The genes involved in the aerobic PA pathway are organized differently in different *Streptomyces* strains and therefore two different sets of primers for the amplification of *paaK* were designed. The primer set, Paak_AveF and Paak_AveR, was designed based on the multiple sequence alignment of *paaK* nucleotide sequences from eight different *Streptomyces*

strains obtained from the GenBank database (http://www.ncbi.nlm.nih.gov/), using DNAMAN (Table 2.4). An amplification product of 490 bp was expected to be amplified from *S. polyantibioticus* SPR^T. In the same manner, the primer set, Paak_CoeF and Paak_CoeR was designed based on the multiple sequence alignment of *paaK* nucleotide sequences from six different *Streptomyces* strains using DNAMAN (Table 2.5). An amplification product of 709 bp was expected to be amplified from *S. polyantibioticus* SPR^T.

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 [c]	AGGCAGGTCGGAGGTGAAGACGAC
VVFTS CycF	1274-1293 [d]	TSGTSTTCACSWSSIHSYTG
QTPQV CycF	1380-1398 [d]	TSWSSCAGACSCCSCAGGT
PALHAL_F	571-593 [e]	TACGGIGTSWVCACCRGBWTSGG
PALHAL_R	879-902 [e]	GIRCASCGBABSGARTASGCGTCC
Paak_AveF	559-583 [f]	CCBTCSTACMTGCTSACSCTSCTSGACG
Paak_AveR	1049-1022 [f]	GSAGSASGATCTCCTCGARCTGSSTGG
Paak_CoeF	315-340 [g]	CCRCGCAGGATSAYCATGTCGTCGC
Paak_CoeR	1024-1001 [g]	GCCTCCAGCGGNACSACSGGBCG
EncP-F	766-787 [e]	GACTCGCACCTGGCGGTCAAC
EncP-R	1486-1465 [e]	GTAGTCGGTGATGGTCTCGTC
M13F	2949-2972	CGCCAGGGTTTTCCCAGTCACGAC
M13R	197-176	TCACACAGGAAACAGCTATGAC

[a] Based on the *Brevibacillus brevis grsA1* sequence (GenBank accession number: D00519).

[b] Based on the *S. virginiae* NRRL B-1446^T *virH* sequence (Table 2.2).

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

[d] Based on the *S. coelicolor* A3(2) non-ribosomal peptide synthetase sequence (GenBank accession number: NP631722)

[e] Based on the *'S. maritimus'* DSM 41777^T *encP* gene (Table 2.2)

[f] Based on the *S. avermitilis* MA-4680^T *paaK* gene (Table 2.2)

[g] Based on the *S. coelicolor* A3(2) *paaK* gene (Table 2.2)

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

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

Table 2.2. Bacterial species and genes used to design the CyF/CyR primers

Species name	Accession Number	Gene	Aromatic Compound
Streptoalloteichus hindustanus ATCC 31158	EF032505	*tlmIV*	Tallysomycin
Streptomyces atroolivaceus (no strain number given)	AF484556	*lnmI*	Leinamycin
Streptomyces flavoviridis ATCC 21892	EU670723	*zbmIV*	Zorbamycin
Streptomyces verticillatus strain ATCC 15003	AF210249	*blmIV*	Bleomycin
Streptomyces virginiae	AB283030	*virH*	Virginiamycin M_1

Table 2.3. Bacterial species and genes used to design the PALHAL primers

Species name	Accession Number	Protein
'*Streptomyces maritimus*'	AF254925	PAL
Streptomyces sp. SirexAA-E	AEN12031.1	HAL
Streptomyces bingchenggensis BCW-1	ADI07714.1	HAL
Streptomyces violaceusniger Tü 4113	AEM87830.1	HAL
Streptomyces avermitilis MA-4680^T	WP_037644951.1	HAL
'*Streptomyces cattleya*' NRRL 8057	CCB76521.1	HAL
Streptomyces coelicolor A3(2)	NP_629085.1	HAL
Streptomyces pratensis ATCC 33331	ADW03754.1	HAL
Streptomyces fradiae NRRL 18158	KDS89354.1	HAL
Streptomyces griseus NBRC 13350	BAG19440.1	HAL
Streptomyces scabiei 87.22	CBG72492.1	HAL
Streptomyces tsukubaensis NRRL 18488	EIF92105.1	HAL

Table 2.4. Bacterial species and genes used to design the Paak_Ave primers

Species name	Accession Number	Gene
Streptomyces avermitilis MA-4680^T	BA000030	*paaK*
'*Streptomyces lividans*' TK24	ACEY01000018	*paaK*
Streptomyces ambofaciens ATCC 23877	AM238663	*paaK*
Streptomyces viridochromogenes DSM 40736	GG657757	*paaK*
Streptomyces bingchenggensis BCW-1	CP002047	*paaK*
Streptomyces albus J1074	DS999645,	*paaK*
Streptomyces clavuligerus ATCC 27064	CM000913	*paaK*
Streptomyces sviceus ATCC 27064	CM000951	*paaK*

Table 2.5. Bacterial species and genes used to design the Paak_Coe primers

Species name	Accession Number	Gene
Steptomyces coelicolor A3(2)	AL645882.2	*paaK*
Streptomyces hygroscopicus ATCC 53653	GG657754	*paaK*
Streptomyces pratensis ATCC 33331	CP002475.1	*paaK*
Streptomyces pristinaespiralis ATCC 25486	CM000950	*paaK*
Streptomyces ghanaensis ATCC 14672	DS999641	*paaK*
Streptomyces sp. SPB78	GG657742	*paaK*

2.3.4 PCR PROTOCOLS

All PCR amplifications were performed using a Techne TC512 Thermal Cycler fitted with a heated lid and gradient sample block.

2.3.4.1 A DOMAIN AMPLIFICATION

The cycling conditions used for the amplification of the A domains from *S. polyantibioticus* SPR^T and *Sts. oxazolinicum* gDNA using the A3F/A7R primer set was as follows: initial denaturation at 95 ºC for 5 min, followed by 35 cycles of denaturation at 95 ºC for 30 s, annealing at 64 ºC for 90 s and elongation at 72 ºC for 4 min, with a final elongation at 72 ºC for 10 min. PCR reactions consisted of: 500 ng of DNA, 2 U SuperTherm *Taq* polymerase (JMR Holdings, USA), 1.5 μM of each primer, 0.2 mM of each dNTP, 4 mM $MgCl_2$ and 3 % (v/v) glycerol in a total volume of 50 μl.

2.3.4.2 CY DOMAIN AMPLIFICATION

Cycling conditions for the amplification of the Cy domain from *S. polyantibioticus* SPR^T gDNA using the CyF/CyR primer set were similar to those used for amplification of the A domain fragment (section 2.3.4.1), with the exception of the annealing temperature, which was performed as a gradient ranging from 52 ºC to 65 ºC. PCR reactions were set up as described in section 2.3.4.1, with the exception of the $MgCl_2$ concentration, which ranged from 2 mM to 4 mM, the primer concentration, which ranged from 0.5 μM to 2.5 μM for each primer and the

glycerol concentration, which ranged from 2 % to 8 % (v/v). Template concentration was also varied from 500 ng to 1500 ng. *S. coelicolor* A3(2) gDNA was used as a positive control.

Amplification using the VVFTS CycF/A7R and QTPQV CycF/A7R primer sets was performed using the same cycling conditions and PCR reaction set up mentioned above.

2.3.4.3 AMMONIA-LYASE GENE AMPLIFICATION

Cycling conditions for the amplification of *encP* from *S. polyantibioticus* SPR^T gDNA, using the EncP-F/EncP-R primer set, consisted of the following: initial denaturation at 95 ºC for 2 min, followed by 30 cycles of denaturation at 95 ºC for 45 s, annealing at 60 ºC for 30 s and elongation at 72 ºC for 70 s, with a final elongation at 72 ºC for 5 min. PCR reactions contained: 500 ng of DNA, 1 U SuperTherm *Taq* polymerase (JMR Holdings, USA), 0.5 µM of each primer, 0.2 mM of each dNTP and 4 mM $MgCl_2$, in a total volume of 50 µl. '*S. maritimus*' DSM 41777^T gDNA was used as a positive control for the amplification of the *encP* fragment.

Cycling conditions using the primer set PALHALF/PALHALR were as follows: initial denaturation at 95 ºC for 5 minutes, followed by 35 cycles of denaturation at 95 ºC for 30 seconds, annealing at 56 ºC for 30 seconds and elongation at 72 ºC for 60 seconds, with a final elongation at 72 ºC for 5 minutes. PCR reactions consisted of: 200 ng of DNA, 2 U SuperThem *Taq* polymerase (JMR Holdings, USA), 0.5 µM of each primer, 0.2 mM of each dNTP and 3 mM $MgCl_2$ in a total volume of 20 µl.

2.3.4.4 *paaK* GENE AMPLIFICATION

Cycling conditions for the amplification of *paaK* using the primer sets Paak_AveF/Paak_AveR and Paak_CoeF/Paak_CoeR, from *S. polyantibioticus* SPR^T were as follows: initial denaturation at 95 ºC for 5 min, followed by 35 cycles of denaturation at 95 ºC for 30 s, annealing at 60 ºC for 45 s and elongation at 72 ºC for 60 s, with a final elongation at 72 ºC for 5 min. PCR reactions consisted of: 400 ng of DNA, 2 U SuperThem *Taq* polymerase (JMR Holdings, USA), 0.5 µM of each primer, 0.2 mM of each dNTP and 2 mM $MgCl_2$ in a total volume of 50 µl. *S. coelicolor* A3(2) and *S. avermitilis* MA-4680^T gDNA were used as a

positive control for the amplification of the *paaK* fragment using the Paak_CoeF/Paak_CoeR and Paak_AveF/Paal_AveR primer sets, respectively.

2.3.4.5 LONG RANGE PCR

In order to efficiently amplify large gDNA fragments of 5-25 kb in size, the Expand Long Range dNTPack (Roche Diagnostics, Germany) was used. The cycling conditions for the amplification of the DPO gene cluster using the A3F/Paak_AveR, Paak_AveF/A7R, VVFTS CycF/A7R and QTPQV CycF/A7R primer sets were performed according to the manufacturer's instructions. Each PCR reaction consisted of 500 ng gDNA, 0.5 mM of each dNTP, 0.5 µM of each primer, 8 % (v/v) DMSO, 2.5 mM $MgCl_2$ and 3.5 U Expand Long Range Enzyme mix in a total volume of 50 µl.

2.3.4.6 COLONY PCR

A colony PCR protocol was used to confirm the presence of cloned inserts in *E. coli* transformants harbouring recombinant constructs. The PCR cycling conditions for the amplification of these inserts was as follows: initial denaturation at 95 ^oC for 5 min, followed by 35 cycles of denaturation at 95 ^oC for 30 s, annealing at 60 ^oC for 90 s and elongation at 72 ^oC for 60 s, with a final elongation at 72 ^oC for 10 min. PCR reactions consisted of: toothpick-tip size amount of cell mass from an *E. coli* transformant colony, 2 U SuperThem *Taq* polymerase (JMR Holdings, USA), 0.5 µM of each primer, M13F and M13R (Table 2.1), 0.8 mM of each dNTP and 3 mM $MgCl_2$ in a total volume of 20 µl.

All PCR amplification products, including probes used for Southern hybridizations, were resolved by electrophoresis alongside a λ-*Pst*I molecular marker on 0.8 % agarose gels containing 0.8 µg/ml ethidium bromide in order to analyse amplicon size and assess primer specificity. The products were visualized using a long wavelength UV light box (Bio-Rad Gel Doc EQ-system™, Bio-Rad Laboratories Inc., USA). For cloning, the fragments of interest were excised from the gel and purified using the FavorPrep Gel/PCR Purification kit (FavorGen^{TM}, Germany) according to the manufacturer's instructions. PCR products for sequencing were purified using the MSB® Spin PCRapace kit (STRATEC Molecular, Germany).

2.3.5 SPLINKERETTE METHOD

The splinkerette method, originally described by Devon *et al*. (1995), was used in an attempt to obtain additional sequence information both upstream and downstream of the putative phenylalanine-specific A domain sequence identified by amplification with the A3F/A7R primer set. The protocol was modified by using an oligonucleotide adaptor sequence that recognizes a *Xho*I overhang and *Xho*I-digested gDNA in the ligation reaction. A standard three step PCR reaction using an annealing temperature of 55 ^oC was performed, as described in the original method, using the splk0/PheAd GSP3 primer set. A second round of PCR was performed using the same cycling conditions and PCR reagent concentrations, but the splk1/PheAd GSP2 primer set was used instead and 10 µl of the initial PCR reaction was used as the template. The primers are listed in Table 2.6.

The PCR amplification products were resolved by electrophoresis alongside a λ-*Pst*I molecular marker on a 1 % agarose gel, containing 0.8 µg/ml ethidium bromide, in order to analyse amplicon size and primer specificity. The products were visualized with a long wavelength UV light box (Bio-Rad Gel Doc EQ-system™, Bio-Rad Laboratories Inc., USA). The fragments of interest were purified for sequencing as described in section 2.3.4.6.

Table 2.6 Oligonucleotide primers used in the splinkerette method

Oligonucleotide name	Primer Sequence (5'→3')	Reference
Splnktop	cgaatcgtaaccgttcgtacgagaattcgtacgagaatcgctgtcctctccaacgagccaaga	Devon *et al*. (1995)
Splk0	Cgaatcgtaaccgttcgtacgagaa	Devon *et al*. (1995)
Splk1	tcgtacgagaatcgctgtcctctcc	Devon *et al*. (1995)
PheAd GSP2	tgaacgaggcgaactggagcacc	This study
PheAd GSP3	ctgcggcgggatcgtcacatgg	This study
Splnkxho	tcgagcttggctcgttttttttgcaaaaa	This study

2.3.6 CLONING AND SEQUENCING

The amplification products obtained from the primer sets Paak_AveF/A7R (2.3.4.4) and splk1/PheAD GSP2 (section 2.3.5) were purified using the MSB® Spin PCRapace kit (STRATEC Molecular, Germany) and sent directly for sequencing, whereas the amplification products obtained from the primer sets A3F/A7R, VVFTS/A7R, Paak_AveF/Paak_AveR, PALHAL_F/PALHAL_R were ligated individually into the pGEM-T plasmid as described in the pGEM®-T Easy Vector System kit (Promega, USA). The ligation reaction was incubated at 22 °C for 14 h after which 10 ng of the reaction was transformed into *E. coli* α-Select Bronze Efficiency Competent Cells (Bioline, UK) according to the manufacturer's instructions. The transformants were inoculated onto Luria-Bertani agar (LA) (Sambrook *et al.,* 1989) supplemented with 100 µg/ml ampicillin, 40 µg/ml X-gal and 0.2 mM IPTG and incubated at 37 °C for 18 h. Transformants harbouring recombinant pGEM-T constructs were identified by blue/white selection and white colonies were subcultured onto fresh LA plates supplemented with 100 µg/ ml ampicillin, 40 µg/ml X-gal and 0.2 mM IPTG and incubated at 37 °C for 18 h to allow confirmation of the white phenotype. The presence of the desired inserts was determined by performing colony PCR (section 2.3.4.6) using the M13F and M13R primer set (Table 2.1).

Transformants identified as harbouring the correct recombinant constructs were inoculated into 5 ml Luria-Bertani broth (LB) containing 100 µg/ml ampicillin and grown overnight with shaking at 37 °C. Thereafter, plasmid DNA was isolated using the NucleoSpin Plasmid Isolation kit (Machery Nagel, Germany) according to the manufacturer's instructions and quantitated spectrophotometrically as described before (section 2.3.2). The cloned DNA fragments were sequenced in both the forward and reverse direction using the M13F and M13R primers (Table 2.1), by the dideoxy chain-termination method on an Applied Biosystems Big Dye terminator v3.1 DNA sequencer using BIOLINE Half Dye Mix (Sanger *et al.,* 1977) (Macrogen Inc., South Korea).

S. polyantibioticus SPR^T gDNA, digested with various restriction endonucleases (section 2.3.8.1), was excised from agarose gels and purified using the Favorgen Gel/PCR Purification Kit (Favorgen™, Germany), before ligating into the blue/white selection vector, pSK. One microgram of pSK vector was digested individually with *Not*I, *Pst*I or *Sac*II overnight at 37 ºC using 1.5 U of each restriction endonuclease along with the appropriate restriction buffer. The

digested pSK vector was dephosphorylated using rAPid Alkaline Phosphatase (Roche, Switzerland) according to the manufacturer's instructions, but the incubation period at 37 ºC was extended to 24 h instead of the recommended 1 h. Subsequently, ligation of the digested gDNA into the dephosphorylated pSK vector was performed according to the sticky-end protocol for the Rapid DNA Ligation Kit (Thermo Fisher Scientific, USA) for 14 h at 4 ºC, after which 10 ng of the reaction was transformed into chemically competent *E. coli* cells. Transformants were screened for inserts and sequenced as described above.

2.3.7 SEQUENCE ANALYSIS

All sequence chromatograms were viewed and edited using Chromas v2.01, (Technelysium, Australia). DNA alignments, assembly of nucleotide sequences obtained by PCR amplification, translations, *in silico* digestions and other analyses were performed using DNAMAN v4.13 (Lynnon Biosoft, USA.).

The isolated DNA sequences were used to search for similar sequences in the GenBank database (http://www.ncbi.blast.nlm.nih.gov/BLAST/) using the BLASTX algorithm (Altschul *et al.*, 1997). The NRPSpredictor2 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 A domains. The nucleotide sequence of each A domain was translated and provided to the program, which identified the key binding-pocket residues by an alignment with the PheA domain of GrsA and determined the probable specificity of the A domain based on the biochemical and biophysical properties of the binding-pocket residues.

2.3.8 SOUTHERN HYBRIDIZATION

2.3.8.1 RESTRICTION ENDONUCLEASE DIGESTION

For Southern hybridizations using the *paaK* probe, *S. polyantibioticus* SPR^T gDNA was digested with the restriction endonucleases *Not*I & *Pst*I and *Pst*I & *Sac*II, as well as with the single restriction endonucleases *Not*I, *Pst*I and *Sac*II.

For Southern hybridization using the A domain probe, *S. polyantibioticus* SPR^T gDNA was digested with the single restriction endonucleases *Not*I, *Pst*I and *Sac*II.

For Southern hybridization using the *encP* probe, gDNA of *S. polyantibioticus* SPR^T was digested with the following restriction endonucleases: *Sph*I & *Stu*I, *Pvu*II & *Sph*I, *Avr*II & *Sph*I and *Avr*II & *Stu*I. '*S. maritimus*' DSM 41777^T genomic DNA was digested with *Sph*I & *Stu*I.

Each reaction consisted of 50 µg of gDNA, 1.5 U of each restriction endonuclease and the appropriate restriction buffer in a total volume of 50 µl. Digestions were performed overnight at 37 ^oC.

2.3.8.2 PROBE PREPARATION

The recombinant constructs harbouring the phenylalanine-specific A domain insert (pGEMA-18) and *paaK* gene insert (pGEM-PAAK) served as the templates for probe preparation using the PCR DIG Probe Synthesis Kit (Roche). Additionally, PCR-amplified *encP* from '*S. maritimus*' DSM 41777^T served as a template for probe preparation and was also used in a Southern hybridization.

Probes for both the A-18 A domain and *paaK* gene were synthesized using the A domain (section 2.3.4.1) and *paaK* (section 2.3.4.4) PCR protocols, respectively, and the plasmid specific primers, M13F and M13R. The amplified *encP* gene probe was synthesized using the DIG Probe Synthesis Kit (Roche, Switzerland), the *encP* PCR protocol (section 2.3.4.3) and the EncP-F/EncP-R primer set.

2.3.8.3 SOUTHERN BLOT HYBRIDIZATION

Restriction endonuclease digested *S. polyantibioticus* SPR^T and '*S. maritimus*' DSM 41777^T gDNAs were electrophoresed on 0.7 % agarose gels containing 0.8 µg/ml ethidium bromide and individually blotted onto Hybond N$^+$ membranes (Amersham Biosciences, UK) using a Trans-blot® SD Semi-dry Transfer Cell (Bio-Rad Laboratories Inc., USA) for 1 h at 15 V and a current of 355 mA.

Each membrane was air dried and placed into a sealable plastic bag with DIG Easy Hybridization Buffer (Roche, Switzerland) (approximately 10 ml/100 cm^2 of membrane). This was followed by incubation with gentle shaking for 2 h at 56.3 ºC for the A domain probe, 53.8 ºC for the *encP* probe and 48 ºC for the *paaK* probe (these being the calculated optimum hybridization temperatures for each probe). The hybridization solution was prepared by adding 60 µl denatured probe to 20 ml hybridization buffer (Roche, Switzerland) and hybridization was performed at the hybridization temperature overnight with gentle shaking. The temperature of hybridization was calculated using 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 pairs

After hybridization, the membrane was washed twice in low stringency buffer, containing 2 x standard saline citrate (SSC; 0.3 M NaCl, 30 mM $Na_3C_6H_5O_7$) and 0.1 % sodium dodecyl sulphate (SDS), at room temperature for 5 min with gentle shaking, followed by washing twice in high stringency buffer (0.1 x SSC; 0.1 % SDS) at 68 ºC for 15 min with gentle shaking. This was followed by washing with washing buffer for 2 min at room temperature (0.5 M maleic acid, pH 7.5; 0.3 % Tween 20) and subsequent blocking with 2 % skim milk for 2 h at room temperature.

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

Before detection, the membrane was equilibrated for 3 min in detection buffer (0.1 M Tris-HCl, 0.1 M NaCl, pH 9.5) at room temperature. Detection was achieved by incubating the membrane in 20 ml detection buffer containing 0.175 mg/ml 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) and 0.25 mg/ml 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) in the dark at room temperature for 0.5-12 h. The reaction was stopped by rinsing the membrane with Tris-EDTA buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.6) for 5 min, after which the result was digitally captured.

2.4 RESULTS AND DISCUSSION

2.4.1 ISOLATION OF THE NRPS GENE CLUSTER

2.4.1.1 A DOMAIN AMPLIFICATION FROM *S. POLYANTIBIOTICUS* SPR^T

The identification of an A domain in the genome of strain SPR with a binding pocket substrate specificity for phenylalanine or 3-hydroxyphenylalanine could indicate that the A domain is part of the NRPS responsible for the biosynthesis of DPO.

NRPS A domains were amplified from *S. polyantibioticus* SPR^T by primers specific for A domain conserved motifs (Ayuso-Sacido & Genilloud, 2005) (Figure 2.4). The amplified products were individually cloned into the pGEM-T Easy vector, transformed into competent *E. coli* cells and screened to confirm the presence of the correct size insert using colony PCR. The inserts were sequenced before being subjected to protein BLAST (BLASTP) analysis in order to confirm that they were A domains. Thirty-one recombinant constructs were identified as carrying an insert (ranging in size from 688 to 729 nucleotide bases) with homology to known A domains in the GenBank database (Table 2.7).

The first 9 amino acid residues of the binding-specificity code of the A domains were identified for each unique sequence by aligning the protein sequence of each A domain against that of the PheA domain of GrsA, in a similar manner to that shown in Figure 1.17. This resulted in the identification of twelve unique amino acid binding pocket codes. Additionally, the A domain protein sequences were inserted individually into the NRPSpredictor2 program, which performs an alignment against the PheA domain of GrsA, but also analyses the physico-chemical fingerprint of the residues lining the amino acid binding pocket to give an indication of the most probable amino acid specificity of the binding pocket. Due to the occurrence of relaxed substrate specificity in certain A domains, more than one amino acid substrate can be predicted for any given A domain (e.g. substrates with similar physico-chemical properties) (Rausch *et al.,* 2005). The specificity binding code and likely amino acid substrate specificity for the twelve unique A domains obtained from *S. polyantibioticus* SPR^T are listed in Table 2.7.

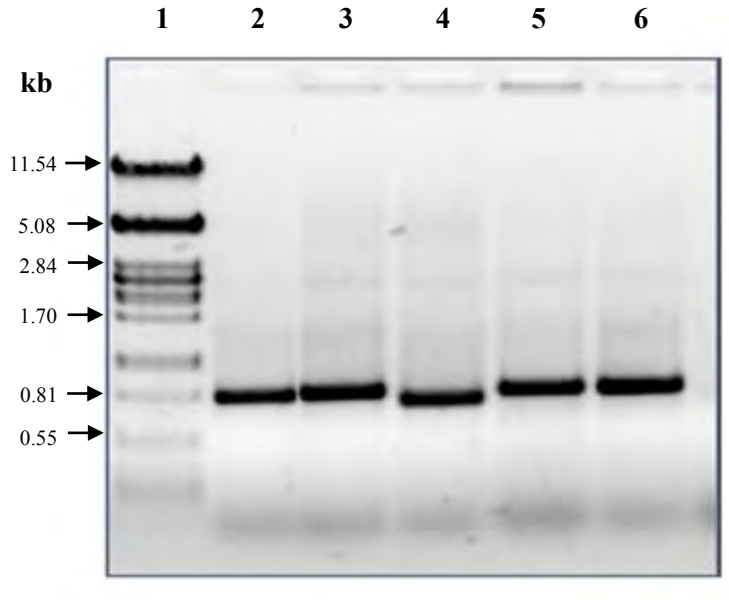

Figure 2.4 Gel electrophoresis of A domain amplicons from *S. polyantibioticus* SPR^T amplified by PCR using the degenerate primer set A3F/A7R. Lanes: **1**- λ-*Pst*I molecular marker, **2-6** – different A domains amplified from *S. polyantibioticus* SPR^T gDNA.

The sequences of the inserts in clones pGEMA-18 and pGEMA-66 were identical and were determined by the NRPSpredictor2 to be specific for phenylalanine. Protein BLAST analysis of the amino acid sequence of both inserts showed a high similarity to an A domain in *Granuliella mallensis* MP5ACIX8 (GenBank accession number: YP_00507340.1) with an amino acid similarity of 57 %, an A domain in *Streptomyces griseus* XYLEBKG1 (GenBank accession number: YP_08236938.1) with an amino acid similarity of 60 % and an A domain in *Streptomyces netropsis* (GenBank accession number: BAH_68437.1) with an amino acid similarity of 54 %. Additionally, all of these BLASTP hits had an amino acid substrate specificity for phenylalanine, as predicted by the NRPSpredictor2 program.

When comparing the amino acid specificity code of the pGEMA-18/pGEMA-66 clones to the well-known phenylalanine-specific A domains of GrsA and BarG, there are differences observed at the 236, 239, 299, 330 and 331 positions (Table 2.7). At position 236, pGEMA-

18/A-66 differs from GrsA and BarG in having a cysteine residue instead of an alanine residue. However, according to the gross clustering of NRPS substrates by chemical similarity, both alanine and cysteine are classified as aliphatic, hydrophobic residues and thus could perform the same function in the binding pocket interacting with non-polar, hydrophobic phenylalanine (Rausch *et al.,* 2005). At position 299, pGEMA-18/A-66 displays an alanine residue instead of isoleucine in the GrsA sequence and valine in the BarG sequence. Isoleucine, alanine and valine are, however, all non-polar amino acids, which may make them interchangeable, and therefore they could perform the same function in the binding pocket in interacting with phenylalanine. The same argument can also be used to explain the differences at position 330, where GrsA displays an isoleucine residue in comparison to valine in both GrsA and pGEMA-18/A-66.

The key differences are at position 239 where GrsA and BarG both display a tryptophan residue in comparison to a glycine residue in pGEMA-18/A-66 and at position 331 where GrsA and BarG both present a cysteine residue in contrast to aspartic acid in pGEMA-18/A-66. However, both of these key differences occur at positions of medium to high variability and thus should not have a large impact on the substrate affinity, as these residues may have minimal interaction with the substrate amino acid.

From the specificity-conferring code and NRPS prediction software, it is possible to infer that the pGEMA-18/pGEMA-66 A domain may be involved in the recognition and activation of phenylalanine as a starter molecule in the biosynthesis of DPO. However, it is also possible that the pGEMA-18/pGEMA-66 A domain may have a substrate specificity for amino acids other than phenylalanine and this is due to the fact that the NRPSpredictor2 predicts aromatic substrates less reliably due to the observed promiscuity of the A domains utilizing these substrates (Rausch *et al.,* 2005).

Furthermore, as serine residues are common amino acids for the biosynthesis of oxazoles, it is feasible that any of the A domains specific for serine, such as the inserts in clones pGEMA-5, A-7, A-12, A-15, A-22, A-36, A-51 and A-63 may activate serine as the starter unit instead of phenylalanine, thereby utilising it in the biosynthesis of DPO (Roy *et al.,* 1999). However, if serine is recognised by the A domain involved in DPO biosynthesis, then it would require the

addition of a phenyl group to carbon-3 of the serine residue via an unusual β-phenylation reaction.

The identification of the genes surrounding the phenylalanine A domain, as well as the A domains specific for serine, would help to elucidate the DPO biosynthetic strategy. However, attempts at isolating additional sequence information upstream and downstream of the A domain specific for phenylalanine using the splinkerette method failed, as non-specific product amplification was observed (data not shown).

Table 2.7 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	
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
Clone pGEMAD-2	D	V	Q	F	N	A	H	M	V	-	Pro
Clone pGEMAD-16	D	A	F	F	L	G	V	T	F	-	Ile, Leu, Val
Clone pGEMA-1	D	M	V	Q	F	G	L	V	Y	-	Gly, Ala, Val
Clone pGEMA-5	D	F	W	N	V	G	M	V	H	-	Ser, Thr, Dht
Clone pGEMA-7	D	F	W	N	V	G	M	V	H	-	Ser, Thr, Dht
Clone pGEMA-8	D	M	V	Q	F	G	L	V	Y	-	Gly, Ala, Val
Clone pGEMA-11	D	M	V	Q	F	G	L	V	Y	-	Gly, Ala, Val
Clone pGEMA-12	D	F	W	N	V	G	M	V	H	-	Ser, Thr, Dht
Clone pGEMA-14	D	M	V	Q	F	G	L	V	H	-	Gly, Ala, Val
Clone pGEMA-15	D	V	W	H	F	S	L	I	D	-	Ser, Thr
Clone pGEMA-16	D	A	F	F	L	G	A	T	F	-	Ile, Leu, Val
Clone pGEMA-18	D	C	G	T	A	A	A	V	D	-	Phe, Trp, Phg, Tyr Bht
Clone pGEMA-19	D	M	V	Q	F	G	L	V	Y	-	Gly, Ala, Val
Clone pGEMA-22	D	F	W	N	V	G	M	V	H	-	Ser, Thr, Dht
Clone pGEMA-28	D	M	E	N	L	G	L	I	N	-	Orn, Lys, Arg
Clone pGEMA-31	D	A	F	F	L	G	A	T	F	-	Ile, Leu, Val
Clone pGEMA-32	G	I	Y	H	L	G	L	L	C	-	Dhpg, hpg
Clone pGEMA-33	G	I	Y	H	L	G	L	L	C	-	Dhpg, hpg
Clone pGEMA-36	D	V	W	H	F	S	L	I	D	-	Ser, Thr

Clone pGEMA-40	D	M	V	Q	F	G	L	V	Y	-	Gly, Ala, Val
Clone pGEMA-48	-	-	-	D	V	G	F	V	-	-	NO HIT
Clone pGEMA-51	D	V	W	H	F	S	L	I	D	-	Ser, Thr
Clone pGEMA-53	D	M	V	Q	F	G	L	V	Y	-	Gly, Ala, Val
Clone pGEMA-58	D	A	F	F	L	G	V	T	F	-	Val, Leu, Ile
Clone pGEMA-60	D	A	F	F	L	G	A	A	T	-	Val, Leu, Ile
Clone pGEMA-63	D	V	W	H	F	S	L	I	D		Ser, Thr
Clone pGEMA-66	D	C	G	T	A	A	A	V	D	-	Phe, Trp, Phg, Tyr, Bht
Clone pGEMA-68	D	M	V	Q	F	G	L	V	Y	-	Gly, Ala, Val
Clone pGEMA-69	D	M	V	Q	F	G	L	V	Y	-	Gly, Ala, Val
Clone pGEMA-70	D	M	V	Q	F	G	L	V	Y	-	Gly, Ala, Val
Clone pGEMA-74	D	A	A	D	V	G	F	V	D	-	Glu, Gln, Asp, Asn
Variability[d] %	3	16	16	39	52	13	26	23	26	0	

[a] According to GrsA numbering

[b] Predictions in order of decreasing preference, where A domain specificities have been clustered according to the physico-chemical fingerprint of the residues lining the amino acid binding pocket to give an indication of the most probable amino acid specificity of the binding pocket.

[c] Part of the barbamide gene cluster in *Lyngbya majuscule* (GenBank accession number: AAN32981), a cyanobacterium.

[d] The amount of variation observed at each amino acid position in the A domain substrate-determining signature sequence. The most variable positions (278 and 299) are considered to be 'wobble' positions (similar to the third position in DNA codons (Stachelhaus *et al.*, 1999).

Residue position 517 was not obtained for *S. polyantibioticus* SPR^T A 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, Hpg – hydroxyphenylglycine, Dhpg – dihydroxyphenylglycine. Highlighted rows indicate the clones carrying an A domain insert specific for phenylalanine.

2.4.1.2 SOUTHERN HYBRIDIZATION USING THE PHENYLALANINE-SPECIFIC A DOMAIN PROBE

A Southern hybridization experiment was performed using a collection of single restriction endonuclease digestions (*Pst*I, *Not*I and *Sac*II) of *S. polyantibioticus* SPR^T gDNA in an attempt to isolate DNA fragments large enough to provide additional sequence data on the NRPS gene cluster involved in DPO biosynthesis. The phenylalanine-specific A domain probe detected numerous bands of varying size, ranging from 0.7 kb to 4.5 kb, for each restriction endonuclease in the Southern hybridization experiment (Figure 2.5). The presence of multiple DNA fragments generated by the *Sac*II and *Not*I digestions may be explained as a result of the target A domain sequence being cleaved by these enzymes. An *in silico* digest of the *S. polyantibioticus* SPR^T phenylalanine (Phe)-specific A domain probe revealed that it is digested by *Sac*II at position 49 and 577 and by *Not*I at position 518. However, *Pst*I does not cleave within the Phe-specific A domain and thus it is proposed that the multiple hybridization bands generated during this digestion may indicate the presence of other Phe-specific A domains in the genome.

Due to the fact that the sizes of these bands were known, gel electrophoresis of gDNA digested with *Pst*I, *Sac*II and *Not*I was performed and DNA of the appropriate size was purified and cloned. Unfortunately, sequencing of the inserts did not yield any A-domain sequences, other NRPS domains or their flanking regions.

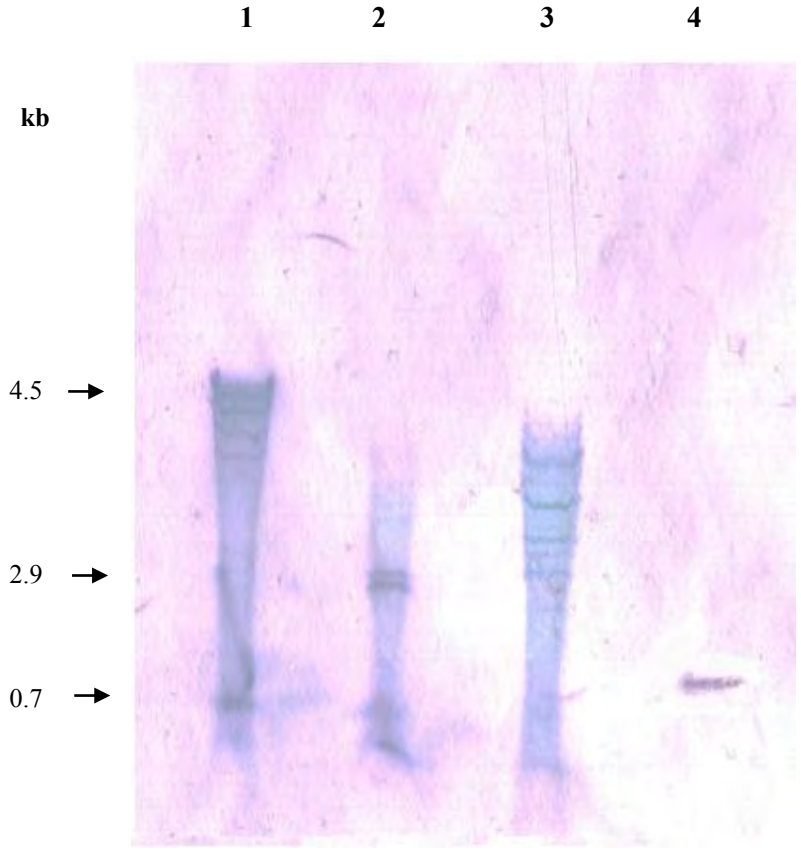

Figure 2.5 Southern hybridization of restriction endonuclease digested *S. polyantibioticus* SPR^T gDNA using the Phe-specific A domain as a probe. Lanes **1-4** each contain 50 µg of *S. polyantibioticus* SPR^T gDNA digested singly with the following enzymes; Lanes: **1**- *Pst*I, **2**- *Sac*II **3**- *Not*I, **4**- Unlabelled Probe (positive control).

2.4.1.3 CY DOMAIN AMPLIFICATION

Initially, the degenerate Cy domain PCR primer set, CyF/CyR, was designed based on sequences from characterised *Streptomyces* and *Streptoalloteichus* thiazole and oxazole producers and was expected to amplify a fragment of approximately 1050 bp from *S. polyantibioticus* SPR^T gDNA. However, there was no amplification of the correct-sized band from either *S. polyantibioticus* SPR^T gDNA or *S. virginiae* NRRL B-1446^T (Figure 2.6A). It should be noted that *S. virginiae* NRRL B-1446^T may not be the same strain as that used for

the design of the CyF/CyR primers (the papers describing virginiamycin M biosynthesis do not specify which *S. virginiae* strain was used). As it is not known whether strain NRRL B-1446^T has the biosynthetic genes for virginiamycin M_1, it may not have been a suitable positive control.

Consequently, two new degenerate Cy domain PCR forward primers, VVFTS CycF and QTPQV CycF, were designed based on a more comprehensive multiple sequence alignment consisting of the original Cy domain sequences and the addition of three new *Streptomyces* Cy domains. The new primers were used in conjunction with the A7R primer and were expected to amplify a product of approximately 1449 bp. The reason for using a Cy-domain forward primer with an A-domain reverse primer is that a typical linear NRPS elongation module consists of the core domains arranged in the canonical order of C/Cy-A-PCP. An amplification product of the correct size was observed using the VVFTS CycF/A7R primer set, but sequencing revealed that a non-target product had been amplified. This was explained as being due to the single VVFTS CycF primer on its own amplifying non-target sequences (Figure 2.6B, lane 9). The primer set QTPQV CycF/A7R failed to amplify the correct size product (Figure 2.6C).

The lack of amplification of a Cy domain fragment may have been due to the fact that the target sequence is not present in the genome of *S. polyantibioticus* SPR^T, but a more likely scenario is that the primers did not bind to the target sequences in *S. polyantibioticus* SPR^T. A lack of primer binding is plausible given the low degree of homology observed in the multiple sequence alignments of Cy domains from the various thiazole and oxazole producers. The lack of amplification in the case of the VVFTS CycF/A7R and QTPQV CycF/A7R primer sets may be due to the fact that the NRPS biosynthetic strategy for DPO production may not follow the conventional linear NRPS arrangement, but may in fact consist of a non-linear organization or even contain unusual freestanding A or C/Cy domains. Investigation of this hypothesis is described in the following chapter on the *S. polyantibioticus* SPR^T genome sequence.

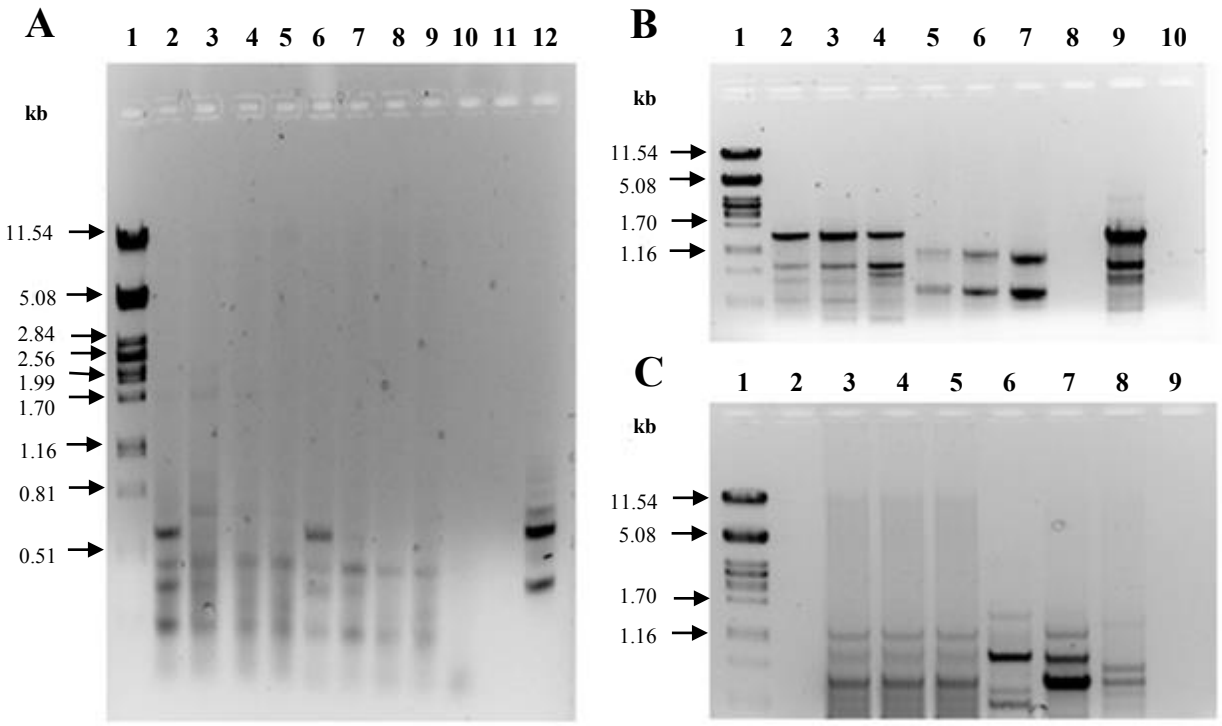

Figure 2.6 Gel electrophoresis of PCR amplified DNA from *S. polyantibioticus* SPR^T using the degenerate Cy domain primer sets **(A)** CyF/CyR, **(B)** VVFTS CycF/A7R and **(C)** QTPQV CycF/A7R. **(A)** Lane 1: λ-*Pst*I molecular marker, Lane 2-4: Amplification at annealing temperature of 52 ºC and increasing $MgCl_2$ concentrations of 2, 3 and 4 mM respectively, Lane 5-7: Amplification at annealing temperature of 58 ºC and increasing $MgCl_2$ concentrations of 2, 3 and 4 mM respectively, Lane 8-10: Amplification at annealing temperature of 65 ºC and increasing $MgCl_2$ concentrations of 2, 3 and 4 mM respectively, Lane 11: No template control, Lane 12: Amplification of *Streptomyces virginiae* NRRL B-1446^T gDNA at annealing temperature of 52 ºC and $MgCl_2$ concentration of 3 mM . **(B)** Lane 1: λ-*Pst*I molecular marker, Lane 2-4: Amplification at annealing temperature of 55 ºC and increasing $MgCl_2$ concentrations of 2, 3 and 4 mM, respectively, Lane 5-7: Amplification of *S. coelicolor* (A3)2 gDNA at annealing temperature of 55 ºC and increasing $MgCl_2$ concentrations of 2, 3 and 4 mM respectively, Lane 8: No template control, Lane 9: VVFTS CycF primer only, Lane 10: A7R primer only. **(C)** Lane 1: λ-*Pst*I molecular marker, Lane 2: No template control, Lane 3-5: Amplification at annealing temperature of 55 ºC and increasing $MgCl_2$ concentrations of 2, 3 and 4 mM respectively, Lanes 6-8: Amplification of *S. coelicolor* (A3)2 gDNA at annealing temperature of 55 ºC and increasing $MgCl_2$ concentrations of 2, 3 and 4 mM respectively, Lane 9: QTPQV CycF primer only.

2.4.1.4 NRPS DOMAIN AMPLIFICATION FROM *STS. OXAZOLINICUM*

Sts. oxazolinicum was recently characterised as producing a novel group of antibiotics, the spoxazomicins, which contain an oxazole moiety and are therefore likely to be produced non-ribosomally (Inahashi *et al.,* 2011). In light of this, it was decided that amplification of an A domain and/or Cy domain from *Sts. oxazolinicum* would aid in the identification of the correct A domain for DPO biosynthesis in *S. polyantibioticus* SPR^T.

Subsequently, two unique A domains, bearing substrate specificity for serine and glycine respectively, were identified in *Sts. oxazolinicum* following their amplification using the A3F/A7R primer set. The Cy domain amplification using the VVFTS CycF/A7R and QTPQV CycF/A7R primer sets was unsuccessful. Due to the fact that Cy domains are absolutely necessary for the heterocyclization of serine or cysteine in the formation of an oxazole or thiazole heterocyclic ring, respectively, it was decided that the A domains in *S. polyantibioticus* SPR^T that are specific for serine would remain of significant interest in the quest to uncover the biosynthetic pathway involved in DPO production. Investigation into this hypothesis is described in the following chapters.

2.4.2 ISOLATION OF GENES INVOLVED IN BENZOIC ACID BIOSYNTHESIS

2.4.2.1 AMPLIFICATION OF PAL/HAL

It has been shown that the presence of the PAL-encoding gene, *encP*, is absolutely required for benzoyl-CoA formation in '*S. maritimus*' DSM 41777^T and therefore *encP* is a prime candidate to screen for when searching for benzoyl-CoA biosynthetic potential (Xiang & Moore, 2003). PCR amplification was performed using the designed *encP* primers on gDNA extracted from *S. polyantibioticus* SPR^T and '*S. maritimus*' DSM 41777^T. Amplification was observed for '*S. maritimus*' DSM 41777^T gDNA resulting in a clear, single band of about 0.7 kb (Figure 2.7). There was no similar amplification observed for *encP* amplification from *S. polyantibioticus* SPR^T. The lack of amplification from *S. polyantibioticus* SPR^T DNA could be because the organism does not contain an *encP* homologue. However, it could also be because the design of the EncP-F/EncP-R primers was based on a single sequence (from '*S. maritimus*' DSM 41777^T), containing no variable nucleotide positions and therefore may not have bound efficiently to the *S. polyantibioticus* SPR^T *encP* homologue.

The PAL amino acid sequence from '*S. maritimus*' DSM 41777^T was used in a multiple sequence alignment together with the HAL amino acid sequences from 11 *Streptomyces* strains in order to design new degenerate primers for the amplification of a PAL/HAL from *S. polyantibioticus* SPR^T. The HAL sequences from various *Streptomyces* strains share a high degree of homology with each other and to the PAL sequence of '*S. maritimus*' DSM 41777^T. A 350 bp fragment was amplified from *S. polyantibioticus* SPR^T gDNA, which was sequenced and identified as a HAL after analysis using BLASTX. It was concluded that the existence of a PAL in the genome of *S. polyantibioticus* SPR^T is unlikely and therefore that the synthesis of benzoic acid for incorporation into DPO does not proceed via a PAL-catalysed conversion of phenylalanine to *trans*-cinnamic acid. Benzoic acid could be produced in a novel manner in *S. polyantibioticus* SPR^T, perhaps via the phenylalanine degradation pathway mentioned earlier (section 2.2.2).

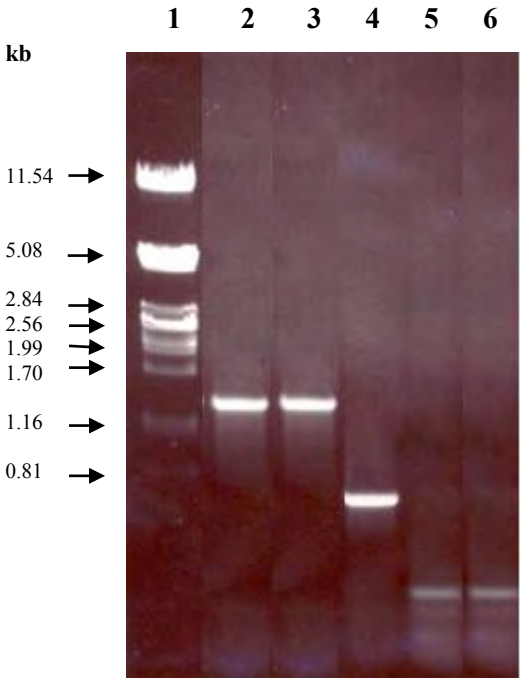

Figure 2.7 Gel electrophoresis of PCR amplified *encP* and 16S rRNA genes from *S. polyantibioticus* SPR^T and '*S. maritimus*' DSM 41777^T. Lanes: **1**- λ-*Pst*I molecular marker, **2**- '*S. maritimus*' DSM 41777^T amplified 16S rRNA gene, **3**- *S. polyantibioticus* SPR^T amplified 16S rRNA gene, **4**- '*S. maritimus*' DSM 41777^T amplified *encP*, **5-6** *S. polyantibioticus* SPR^T with no amplified *encP*.

2.4.2.2 SOUTHERN HYBRIDIZATION USING THE *ENCP* GENE PROBE

Southern hybridization was performed using various pairwise restriction endonuclease digestions of *S. polyantibioticus* SPR^T gDNA along with a pairwise restriction endonuclease digestion of '*S. maritimus*' DSM 41777^T gDNA. There was no hybridization observed for any of the *S. polyantibioticus* SPR^T genomic digests, while a weak hybridization band was observed for the '*S. maritimus*' DSM 41777^T positive control (data not shown). The fact that no hybridization signal was detected, corroborates the result of the PCR amplification experiment (section 2.4.2.1) and suggests that there is no *encP* homologue in *S. polyantibioticus* SPR^T. However, neither the PCR nor the Southern hybridization result rules out the possibility that there is an *encP* homologue in *S. polyantibioticus* SPR^T involved in the synthesis of benzoic

acid, but which has a very dissimilar nucleotide sequence to the '*S. maritimus*' DSM 41777^T *encP* gene.

2.4.2.3 AMPLIFICATION OF *pAAK*

Despite *S. polyantibioticus* SPR^T appearing not to possess a PAL-like gene similar to *encP*, it may produce benzoyl-CoA via the phenylacetate pathway used for the degradation of phenylalanine. The initial step of the phenylacetate pathway is catalyzed by a PA-CoA ligase and degenerate primers, paaK_AveF and paak_AveR, were used to amplify a 490 bp fragment of the homologue of the gene, *paaK*, encoding this enzyme in *S. polyantibioticus* SPR^T (Figure 2.8). The degenerate primer set, paaK_CoeF and paaK_CoeR did not amplify the expected size product from *S. polyantibioticus* SPR^T gDNA (data not shown). Additionally, based on the fact that genes encoding functions for the biosynthesis of many antibiotics are clustered, the primer sets A3F/paak_AveR and paak_AveF/A7R, were used in an effort to amplify larger gDNA fragments containing both an NRPS A domain and the phenylacetate pathway gene cluster (Ikeda *et al.,* 1999). These attempts proved unsuccessful (data not shown).

The presence of *paaK* within the genome of *S. polyantibioticus* SPR^T suggests that the phenylacetate pathway is present, which was expected, as the prototrophic *S. polyantibioticus* SPR^T should have all the pathways for amino acid biosynthesis and degradation. However, whether a derivative of this pathway is used by *S. polyantibioticus* SPR^T to produce benzoyl-CoA remains to be determined. It is possible that the genes involved in benzoic acid synthesis are clustered together with the NRPS involved in DPO biosynthesis and therefore sequencing further upstream and downstream of the amplified *paaK* fragment may help to uncover the DPO NRPS.

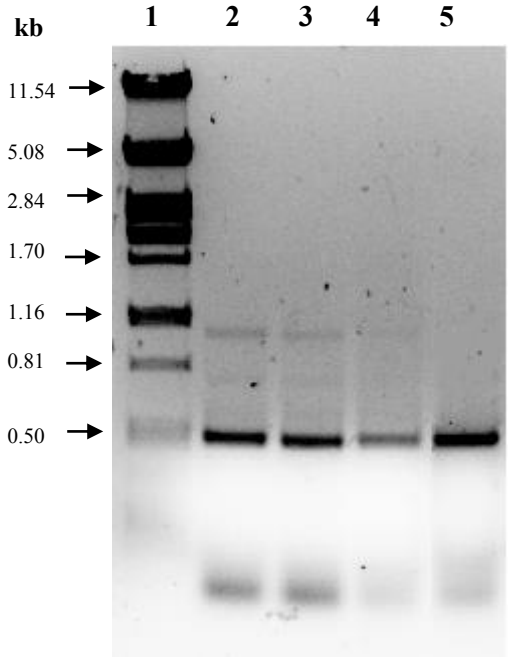

Figure 2.8 Gel electrophoresis of PCR amplified *paaK* from *S. polyantibioticus* SPR^T and *S. avermitilis* MA-4680^T. Lanes: **1**- λ-*Pst*I molecular marker, **2 -** *S. polyantibioticus* SPR^T amplified *paaK* using 2 mM $MgCl_2$ in PCR reaction, **3 -** *S. polyantibioticus* SPR^T amplified *paaK* using 3 mM $MgCl_2$ in PCR reaction, **4 -** *S. polyantibioticus* SPR^T amplified *paaK* using 4 mM $MgCl_2$ in PCR reaction **5**- *S. avermitilis* MA-4680^T amplified *paaK*.

2.4.2.4 SOUTHERN HYBRIDIZATION USING THE *PAAK* GENE PROBE

Southern hybridization using the *paaK* probe resulted in the detection of single bands of approximately 1.8 kb for *S. polyantibioticus* SPR^T gDNA digested with *Sac*II, 2.5 kb for gDNA digested with *Not*I and 4.5 kb for gDNA digested with *Pst*I (Figure 2.9). *S. polyantibioticus* SPR^T digested gDNA fragments of about 1.8 kb and 2.5 kb were gel purified and cloned into the plasmid vector pSK, before using colony PCR to identify clones carrying inserts. Sequencing of the 1.8 kb insert revealed the presence of a monooxygenase gene and sequencing of the 2.5 kb insert identified an A domain specific for proline. No clones containing the *paaK* gene were identified. Thus, these results did not establish that the monooxygenase gene and the proline A domain are near the *paaK* gene, nor that the *paaK* gene is near an NRPS containing an A domain that activates phenylalanine.

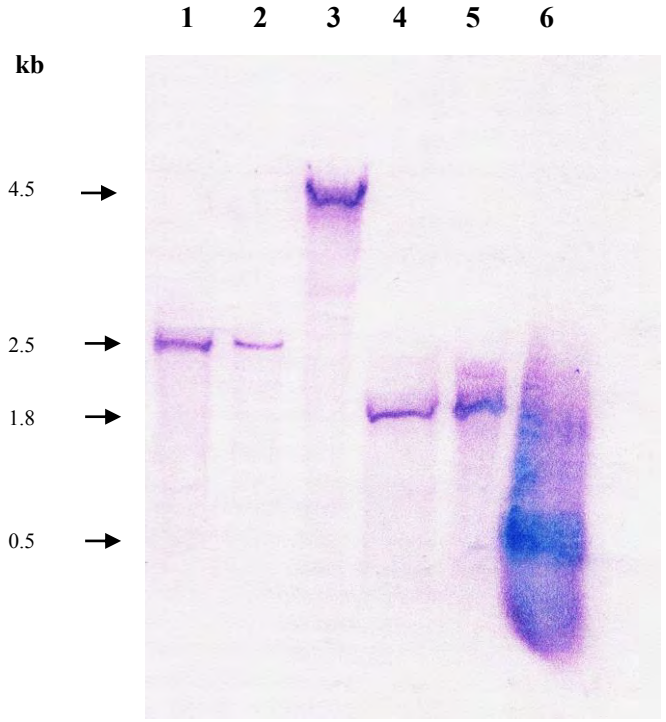

Figure 2.9 Southern hybridization of restriction endonuclease digested *S. polyantibioticus* SPR^T genomic DNA using the amplified *paaK* gene as a probe. Lanes 1-6 contain single and double endonuclease digestions of 50 µg of *S. polyantibioticus* SPR^T genomic DNA in each Southern hybridization. Lanes: **1**- *Not*I/*Pst*I, **2**- *Not*I, **3**- *Pst*I, **4**- *Pst*I/*Sac*II, **5**– *Sac*II, **6**- Unlabelled Probe (positive control).

2.5 CONCLUSION

In the efforts to isolate and identify the biosynthetic gene cluster for the production of DPO in *S. polyantibioticus* SPR^T, twelve unique A domains were identified. Importantly, one A domain was predicted to be specific for the activation of phenylalanine, while two A domains were predicted to be specific for serine. Serine could be involved in DPO biosynthesis due to its known involvement in heterocyclic ring formation.

This chapter describes the extensive efforts to isolate additional sequence information upstream and downstream of the NRPS A domains of interest, using the techniques of Southern hybridization, the splinkerette method and long-range PCR. These efforts were all unsuccessful. Furthermore, the inability to identify an NRPS Cy domain or any genes involved in benzoic acid biosynthesis, led to the conclusion that a completely different approach was required, namely, sequencing of the full *S. polyantibioticus* SPR^T genome. The sequencing of the genome and its annotation are described in the following chapter.

2.6 REFERENCE LIST

Adrova, N. A., Koton, M. M., & Florinsky, F. S. (1957). Preparation of 2, 5-diphenyloxazole and its scintillation efficiency in plastics. *Russian Chemical Bulletin*, **6(3):** 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.

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.

Devon, R.S., Porteous, D.J. & Brookes, A.J. (1995). Splinkerettes--improved vectorettes for greater efficiency in PCR walking. *Nucleic Acids Research*, **23(9):** 1644–1645.

Gescher, J., Eisenreich, W., Worth, J., Bacher, A. & Fuchs, G. (2005). Aerobic benzoyl-CoA catabolic pathway in *Azoarcus evansii*: studies on the non-oxygenolytic ring cleavage enzyme. *Molecular Microbiology*, **56(6):** 1586-1600.

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.

Hertweck, C. & Moore, B.S. (2000). A plant-like biosynthesis of Benzoyl-CoA in the marine bacterium '*Streptomyces maritimus'*. *Tetrahedron*, **56:** 9115-9120.

Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. & Mura, S.O. (1999). Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. *Proceedings of the National Academy of Sciences of the USA*, **96:** 9509-9514.

Inahashi, Y., Iwatsuki M., Ishiyama A., Namatame M., Nishihara-Tsukashima A., Matsumoto A., Hirose T., Sunazuka T., Yamada H., Otoguro K., Takahashi Y., Omura S. & Shiomi K. (2011). Spoxazomicins A-C, novel antitrypanosomal alkaloids produced by an endophytic actinomycete, *Streptosporangium oxazolinicum* K07-0460(T). The *Journal of Antibiotics (Tokyo),* **64(4):** 303-307.

Ionescu, S., Popovici, D., Balaban, A.T. & Hillebrand, M. (2005). Experimental and theoretical study of 2,5-diaryloxazoles whose aryl are *para*-substituted phenyl groups. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **62(1-3):** 252-60.

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. & Noel, J.P. (2002). Plant-like biosynthetic pathways in bacteria: from benzoic acid to chalcone. *Journal of Natural Products*, **65:** 1956-1962.

Onaka, H., Nakaho, M., Hayashi, K., Igarashi, Y. & Furumai, T. (2005). Cloning and characterization of the goadsporin biosynthetic gene cluster from *Streptomyces* sp. TP-A0584. *Microbiology*, **151**: 3923-3933.

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.

Pometto III, A.L., & Crawford, D.L. (1985). L-Phenylalanine and L-tyrosine catabolism by selected Streptomyces species. *Applied Environmental Microbiology*, **49(3):** 727-729.

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 (TSVMs). *Nucleic Acids Research*, **33:** 5799-808.

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.

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

Sanger, F., Nicklen, S. & Coulson, A., R. (1977). DNA Sequencing with chain-terminating inhibitors. *Biotechnology*, **24:** 104-108.

Semenova, O.N., Galkina, O.S., Patsenker, L.D., Yermolenko, I.G. & Fedyunyayeva, I.A. (2004). Experimental and theoretical investigation of the reaction of 2,5-diphenyl-1,3-oxazole and 2,5-diphenyl-1,3,4-oxadiazole dimethylamino derivatives with the Vilsmeier reagent. *Functional Materials,* **11**: 67–75.

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

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.

Stegmann, D.E. (2011). The Investigation into the synthesis of 2, 5-diphenyloxazole in *Streptomyces polyantibioticus* SPR^T. M.Sc. Thesis, Department of Molecular and Cell Biology, University of Cape Town.

Tang, L., Zhang, Y. & Hutchinson, C.R. (1994). Amino Acid Catabolism and Antibiotic Synthesis: Valine Is a Source of Precursors for Macrolide Biosynthesis in *Streptomyces ambofaciens* and *Streptomyces fradiae*. *Journal of Bacteriology*, **176(19):** 6107-6119.

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.

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.

Zhao, C., Ju, J., Christenson, S. D., Smith, W. C., Song, D., Zhou, X. Shen, B. & Deng, Z. (2006). Utilization of the methoxymalonyl-acyl carrier protein biosynthesis locus for cloning the oxazolomycin biosynthetic gene cluster from *Streptomyces albus* JA3453. *Journal of Bacteriology*, **188**: 4142–4147.

CHAPTER 3

STREPTOMYCES POLYANTIBIOTICUS SPR^{T} GENOME EXPLORATION

CHAPTER 3

STREPTOMYCES POLYANTIBIOTICUS SPR^T GENOME EXPLORATION

CHAPTER 3

STREPTOMYCES POLYANTIBIOTICUS SPR^T GENOME EXPLORATION

3.1 ABSTRACT

Initial efforts to sequence the *S. polyantibioticus* SPR^T genome using the 454 pyrosequencing approach proved unsuccessful. However, a hybrid approach using the "third generation" NGS technology, PacBio, in combination with the Illumina MiSeq approach provided a draft *S. polyantibioticus* SPR^T genome sequence and with it the framework to identify the putative DPO biosynthetic cluster. The draft *S. polyantibioticus* SPR^T genome sequence was annotated using antiSMASH 3.0, which identified 43 secondary metabolite gene clusters, thereby revealing the potential of this strain to produce a range of biotechnologically pertinent compounds.

A gene cluster within the draft *S. polyantibioticus* SPR^T genome was identified as the putative DPO biosynthetic gene cluster due to the fact that the NRPS genes in this cluster more closely resembled the enzyme architecture hypothesized to be involved in DPO biosynthesis than any other secondary metabolite cluster identified in the genome. The gene content and the arrangement of the genes in the proposed DPO gene cluster was identical to the congocidine biosynthetic gene cluster identified in *Streptomyces ambofaciens* ATCC 23877^T.

Analysis of the NRPS domains within the putative DPO biosynthetic gene cluster revealed a nonlinear NRPS initiation module, two lone C domains, a lone PCP domain and a lone A domain encoded by an acyl-CoA synthetase. The *in silico* prediction of the substrate specificity of the A domains within both the initiation module and the acyl-CoA synthetase proved inconclusive due to the different predictions obtained from a range of NRPS prediction programmes. However, it was hypothesized that both A domains could display a degree of

relaxed amino acid substrate specificity, commonly observed in A domains displaying specificity for aromatic amino acids, thereby allowing one A domain to activate phenylalanine and the other to activate benzoic acid, in line with the hypothetical DPO biosynthetic scheme presented in Chapter 2 (Figure 2.1). Furthermore, a phylogenetic analysis in combination with an analysis of the conserved signature motifs frequently observed in NRPS domains revealed the presence of *S. polyantibioticus* SPR^T genes encoding putative Cy and TE domains within the proposed DPO biosynthetic cluster. In addition, one PCP domain within the putative DPO biosynthetic cluster was identified as likely to be active from the conserved signature motif analysis, while a second PCP domain was classified as inactive due to the absence of a key catalytic residue. Lastly, a biosynthetic pathway for benzoyl-CoA production in *S. polyantibioticus* SPR^T is proposed, based on the genome analysis.

3.2 INTRODUCTION

The explosion in microbial genome sequencing over the past decade and a half, which has appropriately been named the genomics era, has boosted the field of drug discovery from microbial natural products. The chain-termination DNA sequencing method, also referred to as dideoxy sequencing, developed by Sanger and colleagues in the 1970s has remained the most commonly employed DNA sequencing technique to date. Recently, however, this method has been partially superseded by several next-generation sequencing (NGS) technologies that offer attractive increases in cost-effective sequence throughput (Morozova & Marra, 2008). The automated chain-termination (Sanger) method is considered as a 'first-generation' technology, and newer methods are referred to as next-generation sequencing. These newer technologies constitute various strategies that rely on a combination of template preparation, sequencing and imaging, and genome alignment and assembly methods (Metzker, 2011; Niedringhaus *et al.,* 2011).

A major limitation of the chain-termination method is the requirement of *in vivo* PCR amplification of DNA fragments that need to be sequenced, which is usually accomplished by the laborious and labour intensive practice of cloning into bacterial hosts (Morozova & Marra, 2008). The 454 sequencing technology (Figure 3.1) developed by Roche Applied Science (Penzberg, Germany) was the first NGS technology released onto the market and bypasses the cloning requirement by using a highly efficient *in vitro* DNA amplification method known as emulsion PCR (Tawflik & Griffiths, 1998). During emulsion PCR, individual DNA fragments are linked to specific 454-adaptors bound to single streptavidin-coated beads which are suspended in a water-in-oil emulsion, so that each bead with a single DNA fragment resides in an individual aqueous droplet. These discrete emulsion droplets act as distinct amplification reactors producing approximately 1 million clonal copies of a unique DNA template per bead (Margulies *et al.,* 2005). Each template-containing bead is then transferred into the well of a picotiter plate and the templates are analysed using pyrosequencing. Instead of using dideoxynucleotides to terminate the chain amplification, pyrosequencing technology relies on the detection of pyrophosphate released during nucleotide incorporation and can be described as a sequencing-by-synthesis technique that measures the release of inorganic pyrophosphate (PPi) by chemiluminescence (Liu *et al.,* 2012; Morozova & Marra, 2008). Briefly, the template DNA is immobilized, and solutions of dNTPs are added sequentially; the release of PPi,

whenever the complementary nucleotide is incorporated, is detectable by light produced by a chemiluminescent enzyme, such as luciferase, present in the reaction mix. The sequence of the DNA template is determined from a "pyrogram," which corresponds to the order of correct nucleotides incorporated. Since chemiluminescent signal intensity is proportional to the amount of pyrophosphate released and hence the number of bases incorporated, the pyrosequencing approach is prone to errors that result from incorrectly estimating the length of homopolymeric sequence stretches (i.e. indels). The current state-of-the-art 454 platform marketed by Roche Applied Science (Penzberg, Germany) is capable of generating 80–120 Mb of sequence in 200- to 300-bp reads in a 4 h run (Morozova & Marra, 2008).

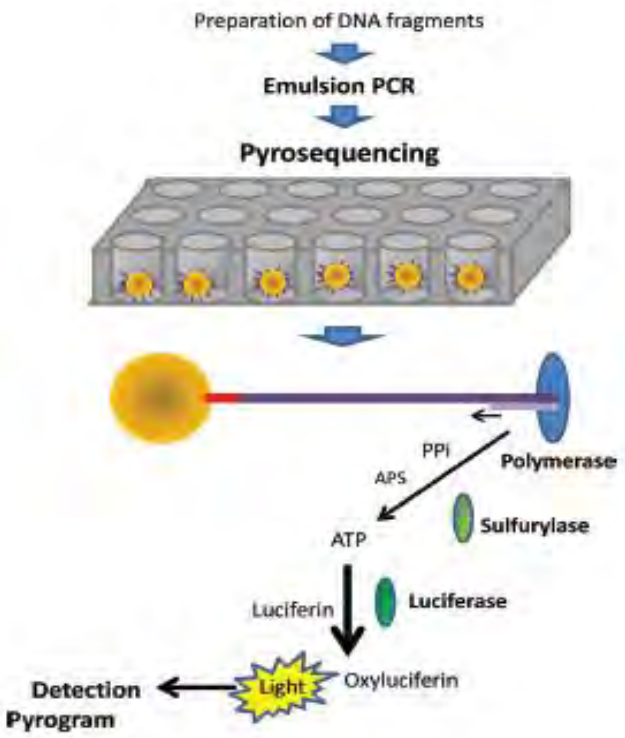

Figure 3.1 The 454 pyrosequencing approach. In this sequencing approach, DNA is isolated, sheared and ligated to special adaptors. Emulsion PCR is then performed for clonal amplification after which the pyrosequencing reaction takes place using a mixture of the single-stranded DNA template, the sequencing primer, and the enzymes DNA polymerase, ATP sulfurylase, luciferase, and apyrase (Siqueira *et al.,* 2012).

Other "second generation" NGS technologies include the Illumina MiSeq/Solexa approach, which involves sequencing-by-synthesis of single-molecule arrays with reversible terminators. A solid glass surface, known as a flowcell, which consists of an acrylamide coating, is used to immobilize individual DNA molecules and bridge PCR amplification (PCR amplification that occurs between primers bound to a surface) is used to amplify DNA into clusters of identical molecules. This sequencing method is similar to the chain-termination method, in that it also relies on dye terminator nucleotides incorporated in the sequence by a DNA polymerase. However, Illumina/Solexa terminators are reversible, permitting polymerization to proceed even after fluorophore detection. DNA sequencing commences with addition of the sequencing primer, DNA polymerase, and four reversible dye terminators. Fluorescence is recorded after incorporation by a four-channel fluorescent scanner (Quail *et al.,* 2012; Siqueira *et al.,* 2012; Glenn, 2011). Furthermore, massively parallel sequencing by hybridization–ligation, which was implemented in the supported oligonucleotide ligation and detection system (SOLiD), was developed by Applied Biosystems (Waltham, USA) (Morozova & Maraa, 2008). This approach involves fluorescent probes that undergo repetitive steps of hybridization and ligation to complementary positions in the template strand, followed by fluorescence imaging to identify the ligated probe (Loman *et al.,* 2012). The ligation chemistry used in SOLiD is based on the polony sequencing technique that was published in the same year as the 454 method (Niedringhaus *et al.,* 2011; Shendure *et al.,* 2005). Meanwhile, "third generation" NGS technologies include the PacBio Single Molecule, Real-Time (SMRT®) DNA Sequencing System developed by Pacific Biosystems (Menlo Park, USA) that provides the longest read lengths of any available sequencing technology (Roberts *et al.,* 2013; Niedringhaus *et al.,* 2011). "Third-generation sequencing" is comprised of two main characteristics; 1) PCR is not required before sequencing, which reduces DNA preparation time for sequencing and 2) the signal is captured in real time, which means that the signal, no matter whether it is fluorescent (Pacbio) or an electric current (Nanopore), is monitored during the enzymatic reaction of adding a nucleotide to the growing complementary strand (Liu *et al*., 2012).

SMRT® sequencing is a sequencing-by-synthesis technology based on real-time imaging of fluorescently tagged nucleotides as they are incorporated along individual DNA template molecules. Due to the fact that the technology uses a DNA polymerase to drive the reaction, and because it images single molecules, there is no degradation of signal over time (Roberts *et al.,* 2013). SMRT® is based on a chip pioneered by Levene *et al.* (2003) that contains an array of zero-mode waveguides (ZMWs). Due to the fact that biological systems characteristically

pack their molecules at spacing much smaller than the wavelength of visible light, fluorescent microscopy techniques that target single-molecule studies must overcome the diffraction-limited resolution of conventional optical microscopy. Fluorescence microscopy techniques such as ZMWs have been developed for this purpose and consist of simple nanostructures that allow real-time observation of individual molecules at high concentrations (Zhu & Craighead, 2012).

During SMRT® sequencing, a single DNA polymerase is attached to the bottom of a well on a SMRT® Cell, and the millions of ZMWs create an illuminated volume that is small enough to observe the incorporation of a single nucleotide. Each time a nucleotide is added to the DNA at the bottom of the well, the dye is detected before being cleaved off and diffusing away (Liu *et al.,* 2012; Levene *et al.,* 2003). The camera inside the instrument computer captures the signal in real-time observation (Mardis, 2013; Liu *et al.,* 2012).

The Ion Torrent platform, marketed by Life Technologies (Carlsbad, USA), uses a sequencing strategy similar to the 454 platform, except that hydrogen ions (H^+) are detected instead of a pyrophosphatase cascade. The use of H^+ means that no lasers, cameras or fluorescent dyes are needed (Glenn, 2011). Furthermore, this platform utilizes semiconductor technology that detects the release of protons as nucleotides are incorporated during DNA synthesis. DNA fragments with specific adapter sequences are linked to and then clonally amplified by emulsion PCR on the surface of 3-micron diameter beads, known as Ion Sphere Particles. The templated beads are loaded into proton-sensing wells that are fabricated on a silicon wafer and sequencing is primed from a specific location in the adapter sequence. As sequencing proceeds, each of the four bases is introduced sequentially and if bases of that type are incorporated, protons are released and a signal is detected proportional to the number of bases incorporated (Quail *et al.,* 2012).

Nanopore sequencing is another "third generation sequencing" method that relies on the transit of a DNA molecule or its component bases through a biological nanopore. The bases are detected by their effect on an electrical current or optical signal (Steinbock & Radenovic, 2015; Schadt *et al.,* 2010).

The increase in speed and decrease in the cost of genome sequencing, as well as the rise of metagenomic sequencing projects, have been the key driving factors for the revolution in the field of microbial drug discovery. Particularly, it has been revealed that streptomycetes still represent a major source for the discovery of novel natural products due to the wealth of NRPS and PKS structural diversity that they encode (Boddy, 2014). This was emphasized by the complete genome sequence of *S. coelicolor* strain A3(2), which revealed the unexpected potential of this microorganism to produce natural products that were undetectable by classical screening methods, such as bioactivity-guided fractionation of crude fermentation broth extracts and chemical screening methods (Aigle *et al.,* 2014; Bachmann *et al.,* 2014).

Indeed, whereas just five secondary metabolites were identified over 50 years using classical screening approaches, the advancement in genome sequencing technologies and bioinformatics has revealed that *S. coelicolor* A3(2) has the capability to produce almost 20 additional natural products (Bentley *et al.,* 2002).

Furthermore, the mining of the genome of *S. ambofaciens* ATCC 23877^T which, for many decades, was only known to produce the macrolide spiramycin and the pyrrolamide congocidine, has allowed for the identification of an additional 23 gene clusters putatively involved in the production of a variety of other secondary metabolites.

Importantly, it is critical to note that this situation is not specific to *S. coelicolor* A3(2) and *S. ambofaciens* ATCC 23877^T, as both complete and partial genome sequencing of numerous streptomycetes has shown that they all contain a large number of cryptic secondary metabolite gene clusters. Thus, these strains contain genes that are not expressed under standard laboratory conditions or the products are formed at a level too low to be detected in laboratory growth conditions (Aigle *et al.,* 2014; Ikeda *et al.,* 2014, Jensen *et al.,* 2014, Nett *et al.*, 2009; Baltz, 2008).

The annotation of sequenced genomes from several *Streptomyces* species has shown that a single strain can carry more than 30 secondary metabolite gene clusters, thereby supporting the idea that the biosynthetic potential of this bacterial genus is far from being fully exploited (Aigle *et al.,* 2014).

The recent advances in genome sequencing are rapidly changing the field of bacterial natural products research by providing opportunities to assess the biosynthetic potential of strains prior to chemical analysis or biological testing. Furthermore, the easy access to sequence data is driving the development of new bioinformatic tools and methods to identify the products of silent or cryptic pathways (Jensen *et al.,* 2014; Bérdy, 2012).

Indeed, automatic bioinformatics platforms, such as antiSMASH 3.0 (Weber *et al.,* 2015), allow the efficient detection of secondary metabolite gene clusters belonging to a range of different classes of natural products and also facilitate the semi-automated prediction of the structure of the natural products encoded by these secondary metabolite blueprints (Bachmann *et al.,* 2014). It has been reported that antiSMASH analysis, which has become the gold standard for mining secondary metabolite gene clusters in genome sequences, compares favourably with high-quality, detailed manual annotation of bacterial whole genomes and provides a more detailed description of individual gene clusters identified than NaPDOS (Ziemert *et al.,* 2012) and NP.searcher (Li *et al.,* 2009) (Boddy, 2014; Harrison & Studholme, 2014). Furthermore, the ClusterBlast and SubClusterBlast tools available within antiSMASH enable rapid identification of unique gene clusters from sequence data sets (Boddy, 2014).

However, a concern with automated genome analysis is that gene clusters located near to each other, may be merged into superclusters, as occurs with the salinilactam gene cluster (Udwary *et al.,* 2007), as antiSMASH defines clusters as groups of signature genes within 10 kb of each other and extends the cluster 20 kb on each side of the last signature gene to delineate the boundaries of PKS and NRPS biosynthetic gene clusters (Boddy, 2014; Medema *et al.,* 2011).

Nevertheless, the recent innovation in genome sequencing has disclosed the astonishing wealth of new polyketide and non-ribosomal peptide natural product diversity to be mined from genetic data and these web-based bioinformatics tools can play a key role in guiding the discovery of novel natural products in genome mining projects, enabling the focus on biosynthetic gene clusters likely to encode new natural product diversity (Boddy, 2014).

3.3 MATERIALS AND METHODS

3.3.1 GENOMIC DNA EXTRACTION

The gDNA for both 454 sequencing and sequencing using the hybrid approach was extracted from *S. polyantibioticus* SPR^T bacterial cell mass using the gDNA extraction method described in section 2.3.2.

3.3.2 ETHANOL PRECIPITATION OF GENOMIC DNA

Total gDNA isolated from *S. polyantibioticus* SPR^T was subjected to ethanol precipitation so that it would be stable against any fluctuations in temperature during transportation to ChunLab (Seoul, South Korea) for hybrid genome sequencing. Briefly, 1 µl of 10 mg/ml glycogen was added to 100 µl of 1 µg/µl gDNA sample, followed by the addition of 10 µl of 3 M sodium acetate (pH 5.2), after which the sample was briefly vortexed. Three hundred microliters of 100 % ethanol was then added to the sample, which was vortexed briefly again. The sample was placed on ice for 30 min and thereafter transported to the sequencing facility.

3.3.3 GENOME SEQUENCING AND ASSEMBLY USING THE 454 STAND-ALONE PLATFORM

The first attempt at sequencing the *S. polyantibioticus* SPR^T genome was performed at the Next Generation Sequencing Facility at the University of the Western Cape (Cape Town, South Africa) using a 454 GS FLX Titanium system (Roche, Switzerland). The Newbler Assembler 2.6 (Roche, Switzerland) was used in an attempt to assemble the reads.

3.3.3.1 GENOME ANNOTATION

Initially, the genes in each of the individual contigs obtained from the 454 genome sequencing were predicted using the Rapid Annotation using Subsystem Technology (RAST) 2.0 server database (Overbeek *et al.,* 2014). Additionally, the sequences of the contigs were aligned and edited using Sequencher version 5.1 (Genecodes Corporation, USA) in order to achieve assembly.

3.3.4 GENOME SEQUENCING AND ASSEMBLY USING A HYBRID APPROACH

The genome of *S. polyantibioticus* SPR^T was sequenced at ChunLab (Seoul, South Korea) using a combination of an Illumina Miseq PE-300 system (Illumina, USA) with 2 × 300 paired-end reads and a PacBio RSII system (Pacific Biosciences, USA). The Illumina and PacBio reads were assembled using CLC Genomics Workbench 7.0.4 (CLCbio, Denmark) and PacBio SMRT Analysis 2.2.0, respectively, and used to construct the draft *S. polyantibioticus* SPR^T genome sequence.

3.3.4.1 GENOME ANNOTATION

The genes in the assembled genome were predicted using the Rapid Annotation using Subsystem Technology (RAST) 2.0 server database (Overbeek *et al.,* 2014) and the gene-caller GLIMMER 3.02 (Delcher *et al.,* 2007). The predicted ORFs were annotated by searching clusters of orthologous groups (COGs) using the SEED database (Disz *et al.,* 2010; Tatusov *et al.,* 1997). CLgenomics™ 1.06 (ChunLab) was used to visualize the genomic features, while protein homologies and conserved domains were analysed by application of BLASTp and the Conserved Domain Database (CDD; Marchler-Bauer *et al*., 2011). AntiSMASH 3.0 (Weber *et al.,* 2015) was used to identify secondary metabolite clusters within the draft genome, while the NRPSpredictor2 program (Röttig *et al.,* 2011) in combination with the latent semantic indexing (LSI) model (developed by Baranasic *et al.*, 2013), the NRPS/PKS prediction server (Bachmann & Ravel, 2009), the Non-Ribosomal Peptide Synthase Substrate Predictor (NRPSsp) (Preito *et al.,* 2012) and the Stachelhaus code (Stachelhaus *et al.,* 1999) were used to predict the substrate specificity of all contigs containing NRPS A domains. The NaPDos tool (available via http://napdos.ucsd.edu/) was utilized in the analysis of all NRPS C domains (Ziemert *et al.,* 2012). Other sequence features were identified using RNAmmer (Lagesen *et al.,* 2007) and tRNAscan-SE (Lowe & Eddy, 1997).

3.3.4.2 SEQUENCE ANALYSIS

For the phylogenetic analysis of the putative domains identified as members of the DPO biosynthetic cluster, amino acid sequences of each separate domain, A, C, PCP, Ox and TE, were obtained from the detailed antiSMASH 3.0 annotation and aligned against homologous amino acid sequences that were selected from the GenBank database (http://www.ncbi.nlm.nih.gov/). The sequences were aligned using the Multiple Sequence

Comparison by Log-Expectation (MUSCLE) algorithm (Edgar, 2004) in Molecular Evolutionary Genetics Analysis (MEGA) version 6.0 (Tamura *et al.,* 2013).

Phylogenetic analyses were performed by construction of unrooted phylogenetic trees using the neighbour joining (Saitou & Nei, 1987), maximum likelihood (Jones *et al.,* 1992) and maximum parsimony (Takahashi & Nei, 2000) methods. A bootstrap test was employed, based on 1000 replicates, to assess the reliability of the topology of each phylogenetic tree. All columns in the alignment containing gaps and missing data were eliminated from the dataset.

Additionally, the conserved NRPS signature motifs, as described by Schwarzer *et al.* (2003), were examined in the multiple sequence alignment of the putative *S. polyantibioticus* SPR^T A, C, PCP, Ox and TE domains in comparison to homologous reference sequences obtained from the GenBank database (http://www.ncbi.nlm.nih.gov/).

3.4 RESULTS AND DISCUSSION

3.4.1 INITIAL GENOME SEQUENCING USING 454 TECHNOLOGY

The 454 pyrosequencing platform generated 7 178 853 bp of *S. polyantibioticus* SPR^T genome sequence consisting of 3553 contigs, however, the *de novo* assembly of the contigs, ranging in size from 100 bp to 78 914 bp, proved unsuccessful. This was attributed to the inability of the sequence assembly program to deal with the highly repetitive strings of G + C residues.

In an attempt to manually assemble the contigs, the sequence analysis program Sequencher version 5.1 (Genecodes Corporation, USA) was employed, but this strategy was unsuccessful. After trying to work with the 454 sequence data for over a year, it was concluded that the reason the genome-sequence contigs could not be assembled was due to there being pieces of the genome missing. Indeed, *S. polyantibioticus* SPR^T has been characterized to use salicin and *myo*-inositol as sole carbon sources (Le Roes-Hill & Meyers, 2009) yet the RAST analysis showed that critical enzymes, such as the protein-N$^{\pi}$-phosphohistidine sugar phosphotransferase (EC 2.7.1.69) involved in the catabolism of salicin to salicin-6-phosphate and all of the enzymes involved in *myo*-inositol metabolism were absent. Additionally, *S. polyantibioticus* SPR^T has been characterized to use L-phenylalanine as a sole nitrogen source (Le Roes-Hill & Meyers, 2009), yet the deaminating phenylalanine dehydrogenase (EC 1.4.1.20), critical for its catabolism, was also absent from the sequencing data. The absence of these key enzymes, in addition to critical enzymes involved in fatty acid metabolism, menaquinone biosynthesis and glycolysis/gluconeogenesis, prompted the decision to re-sequence the *S. polyantibioticus* SPR^T genome using a different platform.

3.4.2 *S. POLYANTIBIOTICUS* SPR^T GENOME PROPERTIES

The Illumina platform provided 482.63-fold coverage (for a total of 15 763 940 sequencing reads) of the genome, while the PacBio platform generated a 38.13-fold coverage of the genome (for a total of 65 082 sequencing reads). The assembled sequences from the two approaches resulted in 1 scaffold that consisted of 12 large contigs, while the assembled draft genome comprised a chromosome (Appendix A) with a length of 8 821 522 bp and 71.53 % G + C content (520.76-fold coverage), which is lower than the 74.4 ± 0.2 mol% determined by the thermal denaturation method (Le Roes-Hill & Meyers, 2009). The draft genome sequence

contained 8148 ORFs, of which the majority (5360, 66%) were assigned putative functions according to the COG functional categories (Table 3.1). In addition, the draft genome sequence consisted of 69 tRNA genes and 24 rRNA genes. Le Roes-Hill and Meyers (2009) reported 7 rRNA operons in the description of *Streptomyces polyantibioticus* sp. nov. The existence of 69 tRNA genes is somewhat surprising considering the fact that there are 64 possible codons (of which three are stop codons and would not be involved in the activation of amino acids). It may be noted that the *Streptomyces davawensis* JCM 4913 genome sequence also revealed the presence of 69 tRNA genes, which suggests that the occurrence of more than the usual 30-40 tRNA encoding genes in bacterial cells is not uncommon (Jankowitsch *et al.,* 2012; Lodish *et al.,* 2000).

Table 3.1. Number of genes associated with the general COG functional categories

Description	Value	% of total[a]
Translation, ribosomal structure and biogenesis	215	4.01
Transcription	633	11.81
Replication, recombination and repair	169	3.15
Cell cycle control, cell division, chromosome partitioning	41	0.76
Posttranslational modification, protein turnover, chaperones	147	2.74
Cell wall/membrane/envelope biogenesis	302	5.63
Cell motility	8	0.15
Inorganic ion transport and metabolism	177	3.30
Signal transduction mechanisms	356	6.64
Energy production and conversion	304	5.67
Carbohydrate transport and metabolism	461	8.60
Amino acid transport and metabolism	490	9.14
Nucleotide transport and metabolism	125	2.33
Coenzyme transport and metabolism	226	4.22
Lipid transport and metabolism	321	5.99
Secondary metabolite biosynthesis, transport and catabolism	159	2.97
General function prediction only	766	14.29
Function unknown	460	8.58

[a] The total is based on the 8148 ORFs identified in the annotated genome.

3.4.3 IDENTIFICATION OF THE PUTATIVE DPO BIOSYNTHETIC GENE CLUSTER

AntiSMASH 3.0 identified 43 gene clusters encoding secondary metabolites in the draft genome of *S. polyantibioticus* SPR^T, revealing the potential of this strain to produce a wide range of biotechnologically relevant compounds: 4 clusters for lantipeptides, 6 clusters for NRPSs, 1 cluster for a siderophore, 10 clusters for PKSs, 2 hybrid NRPS-PKS clusters, 6 clusters encoding terpenes, 2 clusters encoding bacteriocins, 2 clusters encoding melanin, 1 cluster encoding a phenazine/melanin hybrid, 1 cluster encoding a β-lactam, 1 cluster encoding a terpene/lantipeptide hybrid, 1 cluster encoding a lantipeptide/lassopeptide/terpene hybrid, 1 cluster encoding an aminoglycoside/aminocyclitol, 1 cluster encoding a butyrolactone, 1 cluster encoding an ectoine and 3 unidentified clusters.

The putative biosynthetic scheme for the production of DPO proposes that a single NRPS condenses a molecule of benzoic acid with 3-hydroxyphenylalanine. This NRPS is proposed to possess a single A domain, plus ArCP, Cy, Ox, PCP and TE domains (Figure 2.2). In light of this hypothesis, the antiSMASH results were analysed to identify the genes involved in the biosynthesis of DPO amongst the NRPS gene clusters.

The secondary metabolite gene cluster spanning the region located from nucleotide 637827-688770 (Appendix A) of the draft *S. polyantibioticus* SPR^T genome consisted of a linear NRPS encoding 3 A domains with two of them predicted to be specific for the activation of ornithine and one specific for the activation of threonine. Additionally, all of the genes comprising this cluster shared homology to the genes comprising the coelichelin biosynthetic cluster identified in *S. coelicolor* A3(2).

An NRPS gene cluster spanning the region located from nucleotide 669778-735243 (Appendix A) overlapped the NRPS cluster mentioned above, but consisted of 2 A-PCP di-domains and freestanding, adjacently encoded A, C and PCP domains. The three A domains were predicted to be specific for the activation of threonine, proline or valine and leucine, isoleucine, isovaline or 2-amino-butyric acid, respectively. The composition of this gene cluster was most similar in gene content to the tobramycin biosynthetic gene cluster identified in *Streptoalloteichus tenebrarius,* as 24 % of the *S. polyantibioticus* SPR^T genes shared homology to genes within this cluster according to the antiSMASH analysis.

Due to the fact that both of the NRPS gene clusters mentioned above consisted of 3 A domains each, of which none were predicted to be specific for the activation of aromatic amino acids such as phenylalanine or 3-hydroxyphenylalanine, and because the homologous gene clusters encode coelichelin and tobramycin, neither of which contains an aromatic heterocycle, they were dismissed as putative DPO biosynthetic clusters.

The secondary metabolite gene cluster spanning the region located from nucleotide 3093585 – 3154158 (Appendix A) consisted of 4 linear NRPS modules in close proximity and also overlapped an upstream secondary metabolite cluster encoding a terpene, which was predicted to encode a compound of complex molecular structure. Indeed, the gene content of this gene cluster shared the highest homology (12 %) with the biosynthetic gene cluster encoding the cyclic depsipeptide, skyllamycin A, isolated from *Streptomyce*s sp. Acta 2897. The skyllamycin A structure consists of a cinnamoyl side chain and incorporates a large number of β-hydroxylated amino acids as well as an unusual α-hydroxyglycine moiety as a rare structural modification (Pohle *et al.,* 2011). This structure does not bear any resemblance to that of DPO and this gene cluster was therefore dismissed as a putative DPO biosynthetic cluster.

Similarly, the secondary metabolite cluster spanning the region from nucleotide 3320540 – 3386004 (Appendix A) consisted of 4 linear NRPS modules, in addition to an A-PCP di-domain, as well as lone A and PCP domains, thus also predicted to encode a molecule of complex structure. The gene content of this gene cluster shared the highest overall homology (17 %) to the calcium-dependent cyclic lipopetide antibiotic gene cluster identified in *S. coelicolor* (A3)2. Additionally, the composition and arrangement of the genes shared 100 % sub-cluster homology with the clinically important glycopeptide antibiotic teicoplanin isolated from *Actinoplanes teichomyceticus.* Due to the structural dissimilarities between DPO and teicoplanin, this gene cluster was ruled out as the putative DPO biosynthetic cluster.

The secondary metabolite gene cluster spanning the region located from nucleotide 5343994-5398013 (Appendix A) consisted of 2 A-PCP di-domains, 1 NRPS elongation module, as well as 2 freestanding C domains and a freestanding PCP domain. The A domains within this gene cluster were predicted to be specific for the activation of valine, aspartate and alanine, therefore ruling out the formation of a heterocycle such as DPO.

The secondary metabolite gene cluster spanning the region from nucleotide 5634984-5665006 (Appendix A) consisted of a nonlinear NRPS module which deviated from the classical C-A-PCP domain organization and consisted of a lone A domain encoded by an acyl-CoA synthetase, two lone C domains, a lone PCP domain and a lone TE domain. All of the genes within this cluster shared homology to the congocidine biosynthetic gene cluster identified in *S. ambofaciens* ATCC 23877^T. Due to the fact that this cluster was significantly smaller than any of the other secondary metabolite clusters encoding NRPSs and because the A domains were predicted to be specific for the activation of aromatic amino acids, it was explored in greater depth.

3.4.4 INVESTIGATION INTO THE PUTATIVE DPO GENE CLUSTER

The gene cluster spanning the region on the draft *S. polyantibioticus* SPR^T genome from positions 5634984-5665006 was preliminarily identified as the DPO biosynthetic gene cluster based on the antiSMASH 3.0 analysis and due to the fact that the NRPS genes comprising this cluster most closely resembled the gene architecture hypothesized to be involved in DPO biosynthesis, in comparison to the remainder of the secondary metabolite clusters identified in the genome. The genes in this cluster were further analysed with BLASTp to determine protein homologies and conserved domains (Table 3.2).

The *in silico* analysis identified an atypical NRPS, consisting of a freestanding nonlinear module and several single-domain proteins encoded by four genes, as the components of the putative DPO biosynthetic cluster. Interestingly, the composition of the gene cluster and the arrangement of the genes were most similar to the pyrrole-amide congocidine (also known as netropsin) (Figure 3.2) biosynthetic gene cluster identified in *S. ambofaciens* ATCC 23877^T, as 100 % of the *S. polyantibioticus* SPR^T genes shared homology to the genes within the congocidine cluster. Congocidine biosynthesis is characterised by an NRPS that displays some unusual features, namely that the A domain acts iteratively and the C domain uses a CoA-activated substrate rather than a PCP-phosphopantetheinyl-activated substrate. Additionally, the organization of the only complete NRPS module, A-PCP-C, differs from the canonical C-A-PCP organization and the remaining NRPS-associated domains are freestanding (Juguet *et al.,* 2009).

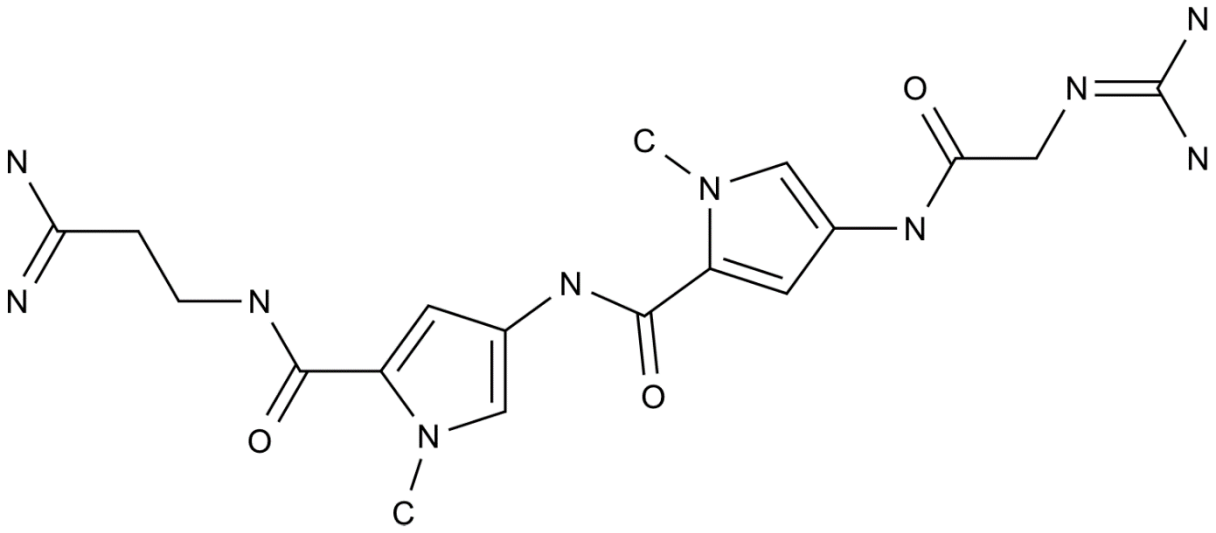

Figure 3.2 Outline of the chemical structure of congocidine.

Freestanding A and PCP domains, such as the A domain encoded by gene SPR_52860 and the PCP encoded by SPR_53070, are normally found when a particular amino acid is modified before its incorporation into the secondary metabolite, such as found in the biosynthesis of the aminocoumarins and vancomycin (Juguet *et al.,* 2009; Fischbach & Walsh, 2006). Freestanding C domains, such as those encoded by SPR_52900 and SPR_53040, as well as those involved in the biosynthesis of vibriobactin, are, however, far less frequent. The condensation reactions catalysed by freestanding C domains generally necessitate a PCP-tethered biosynthetic intermediate and a nucleophilic substrate free in solution (Juguet *et al.,* 2009; Keating *et al.,* 2002), such as the ArCP-bound benzoyl intermediate proposed during DPO biosynthesis.

There are 24 genes constituting the putative DPO biosynthetic cluster starting with SPR_52860 and ending at SPR_53090 (Figure 3.3). Table 3.2 presents an overview of the predicted function of each of the products of the genes, as well as their proposed role in DPO biosynthesis. Two of the genes, SPR_52870 and SPR_52880, are predicted to confer resistance to DPO and one of the genes, SPR_52890, is predicted to encode a regulator for DPO biosynthesis. Out of the remaining 21 genes, 5 genes, SPR_52900, SPR_53040, SPR_53060,

SPR_53070 and SPR_53090, are predicted to encode NRPS-associated domains, while another, SPR_52860, encodes an enzyme belonging to the AMP-binding superfamily of proteins and highly resembles acyl-CoA synthetases. The role of the proteins encoded by the remaining 15 genes cannot be easily inferred from the *in silico* analysis. Indeed, the presence of 5 genes, SPR_52960 to SPR_53000, encoding proteins involved in sugar biosynthesis is perplexing, as DPO does not contain a sugar moiety. It is possible that a glycosylated form of DPO is synthesised by *S. polyantibioticus* SPR^T, however, this putative glycosylated derivative has not been isolated. It is also possible that the putative glycosyltransferase encoded by SPR_52980 and the putative mannanase encoded by SPR_53010 are involved in a self-resistance mechanism involving the inactivation of DPO by intracellular glycosylation and extracellular reactivation by hydrolysis of the added sugar, similar to the type of mechanism demonstrated in *Streptomyces antibioticus* (Juguet *et al.,* 2009; Quirós *et al.,* 1998).

The genes, *cgc4*, *cgc5*, *cgc6* and *cgc7*, present in the congocidine biosynthetic pathway and homologous to SPR_52920, SPR_52930, SPR_52940 and SPR_52950, respectively, were characterized as having no involvement in the biosynthesis of congocidine (Lautru *et al.,* 2012). Similarly, it is possible that these genes play no part in DPO biosynthesis.

Although the putative DPO gene cluster closely resembles both the type and order of the genes present in the congocidine biosynthetic cluster, there are a number of differences, namely: 1) the mannanase encoded by SPR_53010 is absent in congocidine biosynthesis, 2) the TE domain encoded by SPR_53090 is absent in congocidine biosynthesis, and 3) the A domains encoded by SPR_52860 and SPR_53060 confer specificity towards different substrates than the A domains involved in congocidine biosynthesis.

In light of this and due to the existence of transposases encoded by SPR_49500 and SPR_57250, it is possible that *S. polyantibioticus* SPR^T acquired the region between the genes encoding these enzymes in the form of a transposable genetic element, via horizontal gene transfer facilitated by the transposases. The G + C content of this region is 72.5 %, which is slightly higher than the G + C content of the *S. polyantibioticus* SPR^T draft genome (71.53 %), thereby suggesting that the transposable genetic element may have been acquired from another actinomycete, such as *Streptomyces netropsis* DSM 40093 or *S. ambofaciens* ATCC 23877^T. It also suggests that *S. polyantibioticus* SPR^T has adapted the pathway for pyrrolamide biosynthesis for the production of DPO.

Table 3.2 Components of the putative DPO biosynthetic cluster and their proposed function

Gene ID	Size (aa)	GenBank accession number	Closest species match	Identity (%)	Putative function	Proposed role in DPO biosynthesis
SPR_52860	518	AHF81557	*Streptomyces netropsis* DSM 40093	65	Acyl-CoA synthetase	DPO assembly
SPR_52870	626	AHF81556	*Streptomyces netropsis* DSM 40093	75	ABC transporter, ATP-binding protein	Resistance
SPR_52880	616	AHF81555	*Streptomyces netropsis* DSM 40093	74	ABC transporter, transmembrane protein	Resistance
SPR_52890	268	AHF81554	*Streptomyces netropsis* DSM 40093	57	Two-component response regulator	Regulation
SPR_52900	478	AHF81553	*Streptomyces netropsis* DSM 40093	62	NRPS, C domain	Not involved
SPR_52910	479	CAJ88625	*Streptomyces ambofaciens* ATCC 23877	71	Aldehyde dehydrogenase	Unknown
SPR_52920	182	AHF81552	*Streptomyces netropsis* DSM 40093	65	Nucleoside 2-deoxyribosyltransferase	Unknown
SPR_52930	319	AHF81551	*Streptomyces netropsis* DSM 40093	73	Dihydroorotate dehydrogenase	Unknown
SPR_52940	266	AHF81550	*Streptomyces netropsis* DSM 40093	79	Creatinine amidohydrolase	Unknown
SPR_52950	377	AHF81549	*Streptomyces netropsis* DSM 40093	79	Hypothetical protein - hydrolase	Unknown
SPR_52960	437	AHF81548	*Streptomyces netropsis* DSM 40093	75	Nucleotide sugar dehydrogenase	Unknown
SPR_52970	330	AHF81547	*Streptomyces netropsis* DSM 40093	78	Nucleoside-diphosphate sugar epimerase	Unknown
SPR_52980	354	AHF81546	*Streptomyces netropsis* DSM 40093	72	Glycosyltransferase	Unknown
SPR_52990	254	CAJ88617	*Streptomyces ambofaciens* ATCC 23877	73	Sugar nucleotidyl transferase	Unknown
SPR_53000	383	AHF81544	*Streptomyces netropsis* DSM 40093	77	Nucleotide sugar aminotransferase	Unknown
SPR_53010	641	AHF81543	*Streptomyces netropsis* DSM 40093	71	Mannanase	Unknown

SPR_53020	287	AHF81561	*Streptomyces netropsis* DSM 40093	76	Amidohydrolase	Unknown
SPR_53030	271	AHF81542	*Streptomyces netropsis* DSM 40093	70	Methyltransferase	Unknown
SPR_53040	450	AIS24860	*Streptomyces netropsis* DSM 40093	60	NRPS, C domain	DPO assembly
SPR_53050	348	AHF81540	*Streptomyces netropsis* DSM 40093	78	Alcohol dehydrogenase	Unknown
SPR_53060	1084	AHF81539	*Streptomyces netropsis* DSM40093	60	NRPS, A-PCP-C	DPO assembly
SPR_53070	124	ASI24864	*Streptomyces netropsis* DSM40093	66	NRPS, PCP domain	Not involved
SPR_53080	221	WP_037790144	*Streptomyces* sp. Mg1	72	Hypothetical protein	Not involved
SPR_53090	259	WP_030674398	*Streptomyces sp.* NRRL B-1347	79	TE domain	DPO assembly

NRPS, nonribosomal peptide synthetase; C, condensation; A, adenylation; PCP, peptidyl carrier protein; TE, thioesterase. Percentages of identity refer to deduced amino acid sequence comparisons.

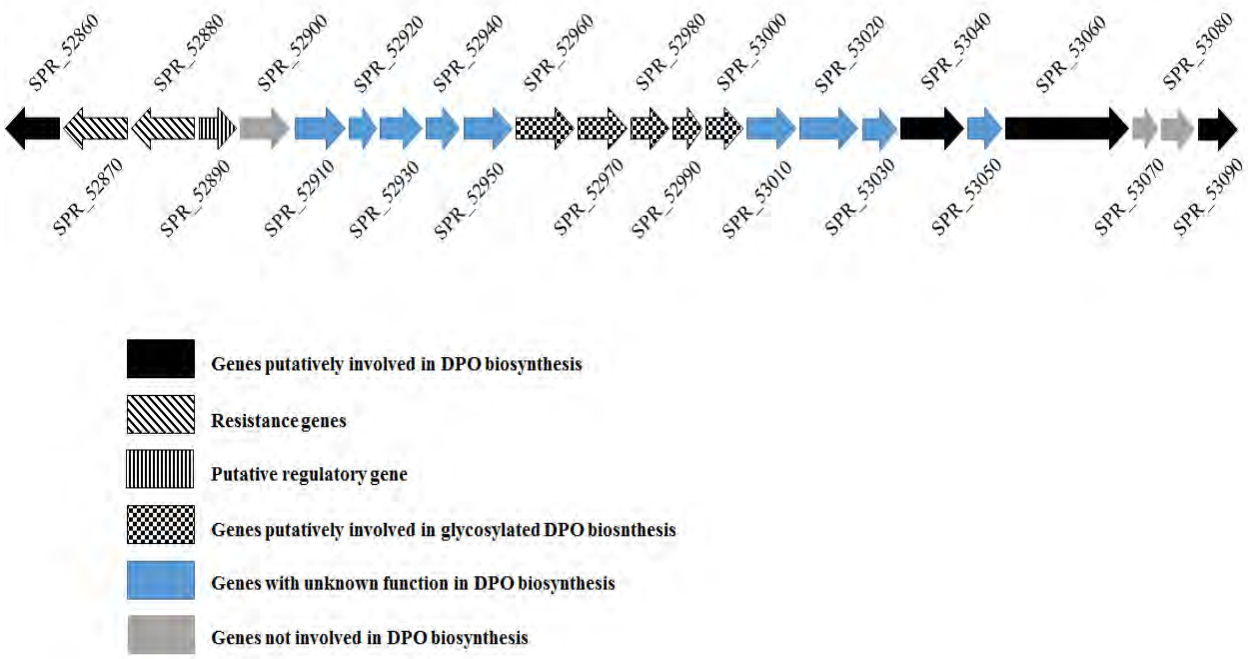

Figure 3.3 Genetic organization of the putative DPO biosynthetic gene cluster.

Furthermore, in order to further characterize the individual NRPS domains contained within the putative DPO biosynthetic cluster, known signature motifs and catalytic residues were identified and assessed for integrity against a selection of reference proteins derived from the CDD (Marchler-Bauer *et al.,* 2011). These analyses are described in the following sections.

3.4.4.1 A DOMAIN SPECIFICITY

Prediction of the substrate specificity of the A domains encoded by SPR_52860 and SPR_53060 was carried out using the NRPSpredictor2 (Röttig *et al.,* 2011), the LSI algorithm (Baranasic *et al.*, 2013), NRPS/PKS predictor (Bachmann & Ravel, 2009), the Non-Ribosomal Peptide Synthase Substrate Predictor (NRPSsp) (Prieto *et al.,* 2012) and the Stachelhaus specificity-conferring code (Stachelhaus *et al.,* 1999) (Table 3.3). The A domain specificity predictions were not conclusive as the web-based NRPS prediction services produced different

results for the A domains encoded by SPR_52860 and SPR_53060. However, the A domain specificity results for the domains encoded by SPR_52860 and SPR_53060 were in agreement with the prediction that the substrate would be aromatic i.e. tryptophan, phenylalanine and/or 2,3-dihydroxybenzoic acid.

The NRPSpredictor2 result for the query domain encoded by SPR_52860, based on the gross physico-chemical properties, was that the substrate is hydrophobic and aliphatic, but with a very low score of 0.190785. The result based on the gross physico-chemical properties for the query domain encoded by SPR_53060, predicts that the substrate is hydrophilic with a slightly higher, but still inconclusive score of 0.375341. On the basis of these highly questionable predictions, it was not possible to infer the A domain specificity of the query domains encoded by SPR_52860 and SPR_53060 using the NRPSpredictor2 output. Furthermore, Rausch *et al.* (2005) commented that the NRPSpredictor2 server predicts aromatic substrates less reliably due to the observed promiscuity of the A domains utilizing these substrates (Rausch *et al.,* 2005). Indeed, substrate promiscuity has been observed in *Xenorhabdus nematophila* where two A domains that exhibit specificity for tryptophan in xenematide A biosynthesis can accept either tryptophan or phenylalanine as substrates to produce a diverse family of products. Markedly, this tryptophan/phenylalanine promiscuity is accommodated in the downstream domains and results in a small family of cyclic depsipeptides (Crawford *et al.* 2011).

Instances of relaxed substrate specificity have also been observed in cyanobacterial NRPS A domains, where structurally related amino acids such as valine, isoleucine and leucine are located in equivalent positions of the final peptide products. These NRPS pathways produce entire families of structurally related compounds all co-occurring in one specific isolate (Christiansen *et al.,* 2011).

Both of the A-domain specificity prediction results obtained from the LSI algorithm were based on a LSI score of below 0.66, which is classified as an unreliable prediction (Baranasic *et al.*, 2013).

The A domain specificity prediction based on the profile Hidden Markov Model (pHMM) approach developed by Prieto *et al.* (2012) and known as NRPSsp, resulted in both query domains predicted to be specific for the activation of phenylalanine. However, the authenticity of this prediction server was questioned by Khayatt *et al*. (2013) who claimed that the dataset

that the NRPSsp server is based on contains erroneous annotations and several sequences not related to NRPSs.

The NRPS/PKS prediction server developed by Bachmann & Ravel (2009) could not identify the A domain specificity for either query domain. The Stachelhaus specificity-conferring code predicted the substrates 2,3-dihydroxybenzoic acid and tryptophan for the query domains encoded by SPR_52860 and SPR_53060, respectively, with a score of 60 % in each case. In comparison to the scores achieved by the other prediction servers/algorithms, these are the most reliable. However, Rausch *et al.* (2005) noted that Stachelhaus predictions at less than a 70% threshold are less reliable and lead to inconsistent predictions. In light of this statement and the fact that tryptophan-and phenylalanine-specific domains are known to be interchangeable, it is possible that the query domain, SPR_53060, displays a relaxed specificity and activates phenylalanine (or 3-hydroxyphenylalanine) instead of tryptophan. Furthermore, it is possible that the query domain, SPR_52860, also displays relaxed substrate specificity, allowing it to activate benzoic acid rather than 2,3-dihydroxybenzoic acid.

Table 3.3 A domain specificity predictions

	NRPSpredictor 2 (Röttig *et al.,* 2011)	**LSI** (Baranasic *et al.*, 2013)	**NRPS/PKS** (Bachmann & Ravel, 2009)	**NRPSsp** (Prieto *et al.,* 2012)	**Stachelhaus code** (Stachelhaus *et al.,* 1999)
SPR_52860	Hydrophobic-aliphatic	tryptophan	NO HIT	phenylalanine	2,3-dihydroxy-benzoic acid
SPR_53060	Hydrophilic	tryptophan	NO HIT	phenylalanine	tryptophan

3.4.4.2 CY DOMAINS

The NaPDos online tool did not identify any obvious Cy domains within the *S. polyantibioticus* SPR^T genome sequence. In light of this, a phylogenetic analysis was performed in order to determine the evolutionary relatedness of all the annotated C domains within the *S. polyantibioticus* SPR^T genome in comparison to reference Cy domain sequences from the *Streptomyces virginiae* NRRL B-1446^T VirH sequence (GenBank accession number:

AB283030), *S. coelicolor* A3(2) putative non-ribosomal peptide synthetase sequence (CAC17500), *Streptomyces verticillatus* strain ATCC 15003 BlmIV sequence (AAG02364), *Pseudomonas aeurginosa* PA14 PchE sequence (AAD55800), *Bacillus licheniformis* ATCC 10716 BacA sequence (O68006), *Sorangium cellulosum* So ce90 EposP sequence (AAF26925) and the *Streptomyces flavoviridis* ATCC 21892 Zbm sequence (EU670723).

The phylogenetic analysis, using the neighbour joining (NJ) (Figure 3.4) and maximum likelihood (ML) (Figure 3.5) algorithms, resolved the putative C domains from *S. polyantibioticus* SPR^T and the reference Cy domain sequences into two distinct groups. The first group consisted of 19 of the putative C domains found within the *S. polyantibioticus* SPR^T genome sequence, while the second group consisted of all of the reference Cy domains plus the *S. polyantibioticus* SPR^T C domains encoded by SPR_52900, SPR_53040, SPR_53060, SPR_6230, SPR_6240 and SPR_6360. From the neighbour-joining phylogenetic analysis, it can be inferred that the *S. polyantibioticus* SPR^T genes in the second group are further sub-divided into three clusters, wherein SPR_6230 was more closely related to the reference Cy domains than SPR_6360, SPR_6240, SPR_53040, SPR_53060, and SPR_52900. Moreover, SPR_6240 was more closely related to SPR_53040 and SPR_53060 than was SPR_6230, with SPR_52900 being the least related to the reference Cy domains. However, the bootstrap support for the association of SPR_6230, SPR_6240, SPR_6360, SPR_53040 and SPR_53060 with known Cy domain sequences was very weak (<50%).

In contrast, the phylogenetic analysis using the ML algorithm resolved the *S. polyantibioticus* SPR^T genes in the second group into five clusters, wherein SPR_6240 was the most closely related to the reference Cy domains, followed by SPR_53040, SPR_53060, SPR_6360, SPR_52900 and SPR_6230. Similarly to the NJ tree, the bootstrap support for the association of SPR_6230, SPR_6240, SPR_6360, SPR_52040 and SPR_53060 with known Cy domain sequences was very weak (<50%).

Similarly, the phylogenetic analysis using the maximum parsimony (MP) (Figure 3.6) algorithm, resolved the C domains from *S. polyantibioticus* SPR^T and the reference Cy domain sequences into two distinct groups. However, while the first, most distantly-related group consisted of only *S. polyantibioticus* SPR^T encoded C domains, the second group consisted of the Cy reference domains, in addition to the *S. polyantibioticus* SPR^T genes encoded by SPR_6230, SPR_6240, SPR_6360, SPR_53060 and SPR_53040. This algorithm established

that SPR_6360 was most closely related to the reference Cy domains, followed by SPR_6240, SPR_53060, SPR_53040 and SPR_6230. In contrast to the NJ and ML algorithms, SPR_52900 was not associated with the Cy domain reference sequences in this instance.

It is clear from the phylogenetic analyses that the *S. polyantibioticus* SPR^T amino acid sequences from SPR_6230, SPR_6240, SPR_6360, SPR_53060 and SPR_53040 were consistently associated with the reference Cy domains, suggesting that SPR_6230, SPR_6240, SPR_6360, SPR_53060 and SPR_53040 encode Cy domains, despite being annotated as C domains. It has been reported that despite secondary structure predictions indicating a similar overall fold for both Cy and C domains, there is weak primary sequence homology between the two (Kelly *et al.,* 2005; Keating *et al.,* 2002). This has contributed to the lack of knowledge regarding the catalytic mechanism of the Cy domains and only a detailed structural and mutational analysis of the catalytic residues of the domains will help to elucidate the mechanisms involved (Hur *et al.,* 2012). Despite primary sequence differences, Figures 3.3-3.5 show that C and Cy domains can be distinguished by phylogenetic analyses.

The phylogenetic analysis of the putative C and Cy domains prompted a comprehensive investigation of the signature motifs and catalytic residues contained within the domains encoded by SPR_53060, SPR_52900, SPR_53040, SPR_6230, SPR_6240 and SPR_6360, as these domains shared the highest degree of evolutionary relatedness to the reference Cy domains (Figure 3.7).

Heterocyclization domains replace C domains in NRPS modules that produce thiazoline or oxazoline rings. In addition, the catalytic core motif, H-H-x-x-x-D-G, of regular C domains is replaced by a D-x-x-x-x-D-x-x-S motif found in Cy domains; the two aspartate residues play an essential part in both condensation and heterocyclization reactions (Keating *et al.,* 2002). An additional seven Cy domain signature motifs were described by Schwarzer *et al.* (2003), all of which were analysed using a multiple sequence alignment with reference Cy domain sequences from the *S. virginiae* NRRL B-1446^T VirH sequence (GenBank accession number: AB283030), *S. coelicolor* A3(2) putative non-ribosomal peptide synthetase sequence (CAC17500), *S. verticillatus* strain ATCC 15003 BlmIV sequence (AAG02364), *Pseudomonas aeurginosa* PA14 PchE sequence (AAD55800), *Bacillus licheniformis* ATCC 10716 BacA sequence (O68006), *Sorangium cellulosum* So ce90 EposP sequence (AAF26925) and the *Streptomyces flavoviridis* ATCC21892 Zbm sequence (EU670723) (Figure 3.7).

The putative Cy domains encoded by SPR_53040, SPR_53060 and SPR_52900 shared a 35 %, 41 % and 24 % homology, respectively, to the eight NRPS Cy domain signature motifs. By comparison, the two Cy domains encoded by the *angN* gene found in the NRPS module of *Vibrio anguillarum* 775, revealed only 23 % and 19 % homology with 7 other characterised Cy domains in a multiple sequence alignment (Di Lorenzo *et al.,* 2008). Within the core motif (D-x-x-x-x-D-x-x-S), SPR_52900 displayed neither of the essential catalytic aspartate residues and was therefore concluded to be inactive. In contrast, SPR_6230 displayed both of the critical catalytic aspartate residues, while SPR_6240 and SPR_6360 each displayed only the second catalytic aspartate residue. Similarly, SPR_53040 and SPR_53060 each displayed only one of the catalytic aspartate residues. However, SPR_53040 displayed the first catalytic aspartate residue, while SPR_53060 displayed the second aspartate residue, suggesting the possible complementation of the catalytic activity required in the formation of DPO.

Indeed, it has been demonstrated that the two different Cy domains found in the VibF NRPS of the vibriobactin biosynthetic pathway perform either the role of the condensation reaction or the role of heterocyclization. A mutation of the catalytic aspartate residue in one domain resulted in lower heterocyclic product formation, whilst maintaining condensation activity. A mutation in the second domain caused the opposite effect (Keating *et al.,* 2002). However, from the multiple sequence alignment alone, it cannot be determined whether the two domains, SPR_53040 and SPR_53060, together provide the catalytic activity required for the formation of DPO by *S. polyantibioticus* SPR^T.

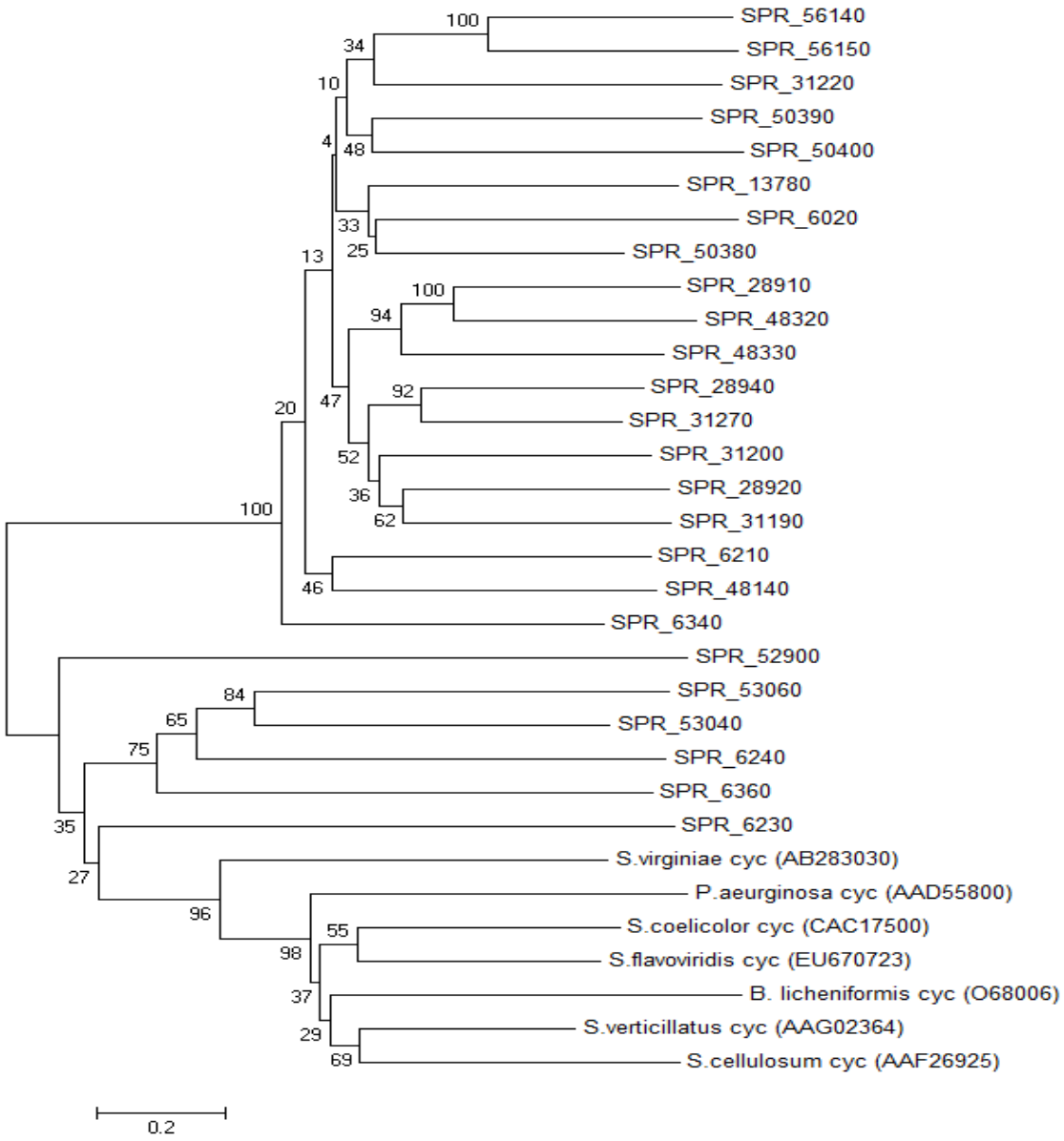

Figure 3.4 Unrooted phylogenetic tree obtained by the neighbour-joining method (Saitou & Nei, 1987) from an alignment of the amino acid sequences of *S. polyantibioticus* SPR^T C domains and heterocyclization domain sequences from the *S. virginiae* NRRL B-1446^T Vir sequence (S.virginiae cyc), *S. coelicolor* A3(2) putative non-ribosomal peptide synthetase sequence (S. coelicolor cyc), *S. verticillatus* strain ATCC 15003 BlmIV sequence (S.verticillus cyc), *Pseudomonas aeurginosa* PA14 PchE sequence (P.aeurginosa cyc), *Bacillus licheniformis* ATCC 10716 BacA sequence (B. licheniformis cyc), *Sorangium cellulosum* So ce90 EposP sequence (S. cellulosum cyc) and the *Streptomyces flavoviridis* ATCC21892 Zbm sequence (S.flavoridis cyc). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches and the GenBank accession numbers are displayed in parenthesis. The analysis involved 32 amino acid sequences and all positions containing gaps and missing data were eliminated. Scale bar, 20 amino acid substitutions per 100 amino acids.

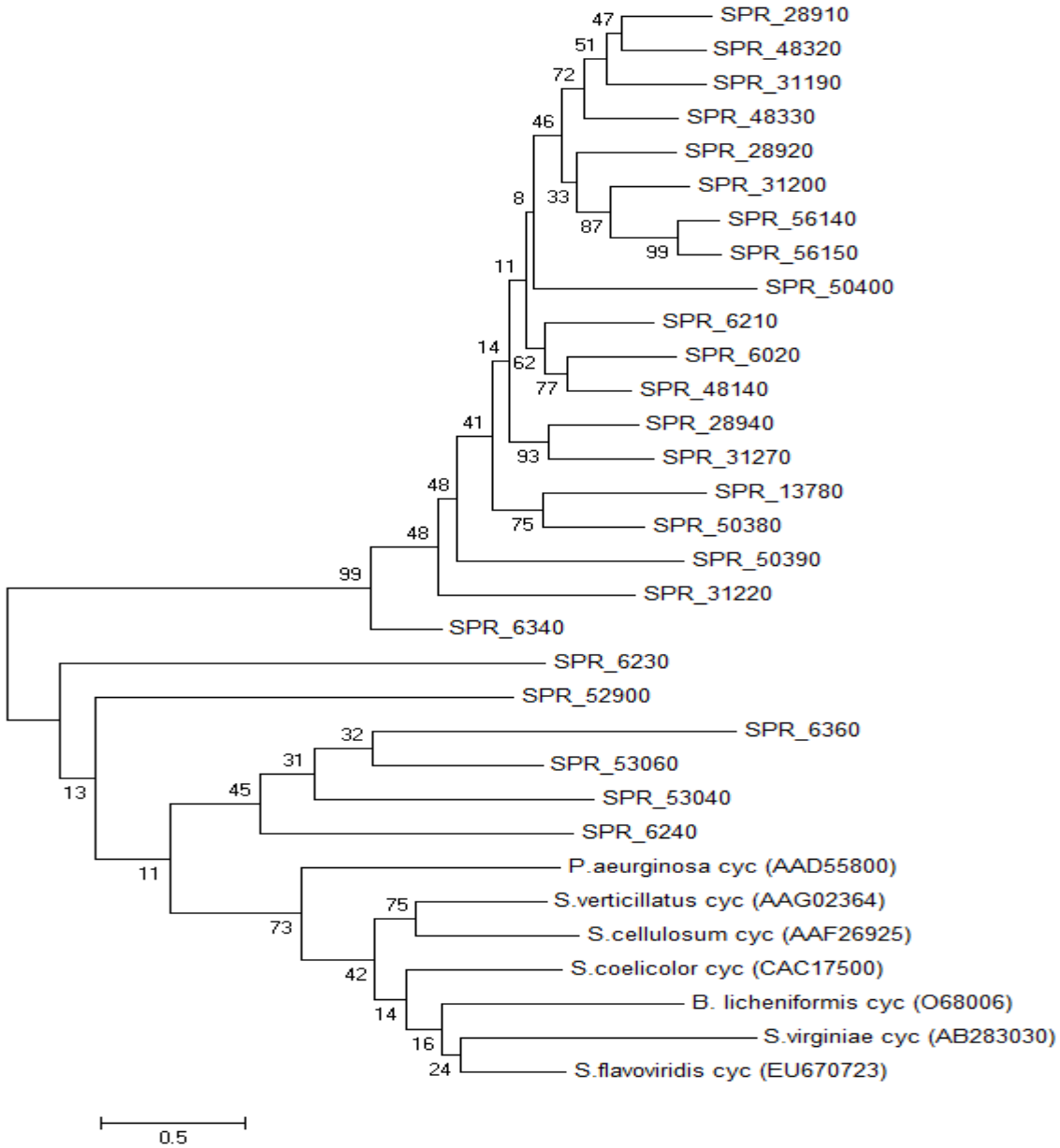

Figure 3.5 Unrooted phylogenetic tree obtained by the Maximum Likelihood method (Jones *et al.,* 1992) based on the JTT matrix-based model for the amino acid sequences of *S. polyantibioticus* SPR^T C domains and heterocyclization domain sequences from the *S. virginiae* NRRL B-1446^T VirH sequence (S.virginiae cyc), *S. coelicolor* A3(2) putative non-ribosomal peptide synthetase sequence (S. coelicolor cyc), *S. verticillatus* strain ATCC 15003 BlmIV sequence (S.verticillus cyc), *Pseudomonas aeurginosa* PA14 PchE sequence (P.aeurginosa cyc), *Bacillus licheniformis* ATCC 10716 BacA sequence (B. licheniformis cyc), *Sorangium cellulosum* So ce90 EposP sequence (S. cellulosum cyc) and the *Streptomyces flavoviridis* ATCC21892 Zbm sequence (S.flavoridis cyc). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches and the GenBank accession numbers are displayed in parenthesis. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value (Tamura *et al.,* 2013). The analysis involved 32 amino acid sequences. All positions containing gaps and missing data were eliminated. Scale bar, 50 amino acid substitutions per 100 amino acids.

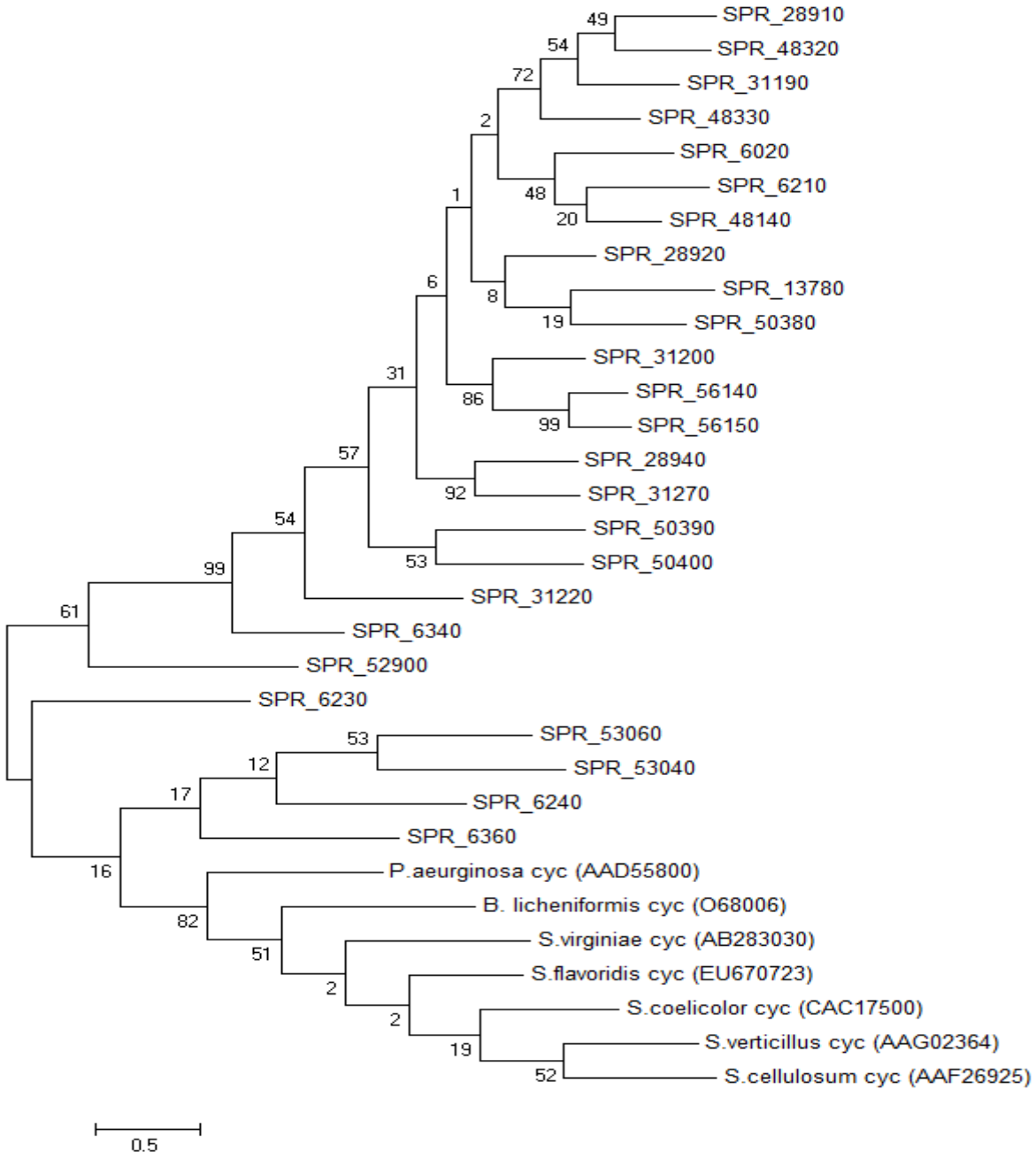

Figure 3.6 Unrooted phylogenetic tree obtained by the Maximum Parsimony (MP) method (Felsenstein, 1985) for the amino acid sequences of *S. polyantibioticus* SPR^T C domains and heterocyclization domain sequences from the *S. virginiae* NRRL B-1446^T VirH sequence (S.virginiae cyc), *S. coelicolor* A3(2) putative non-ribosomal peptide synthetase sequence (S. coelicolor cyc), *S. verticillatus* strain ATCC 15003 BlmIV sequence (S.verticillus cyc), *Pseudomonas aeurginosa* PA14 PchE sequence (P.aeurginosa cyc), *Bacillus licheniformis* ATCC 10716 BacA sequence (B. licheniformis cyc), *Sorangium cellulosum* So ce90 EposP sequence (S. cellulosum cyc) and the *Streptomyces flavoviridis* ATCC21892 Zbm sequence (S.flavoridis cyc). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches and the GenBank accession numbers are displayed in parenthesis. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates) (Nei & Kumar, 2000). The analysis involved 32 amino acid sequences. All positions containing gaps and missing data were eliminated.

Cyc motif 1	**F**	**P**	**L**	**T/S**	**2x**	**Q**	**x**	**A**	**Y**	**2x**	**G**	**R**
VirH-cyc	L	P	L	T	(2x)	Q	S	A	Y	(2x)	G	R
NRPS-cyc	F	P	L	T	(2x)	Q	A	A	Y	(2x)	G	R
blmIV-cyc	F	P	L	T	(2x)	Q	R	A	Y	(2x)	G	R
BacA-cyc	F	P	L	T	(2x)	Q	L	A	Y	(2x)	G	R
EposP-cyc	F	P	L	T	(2x)	Q	E	S	Y	(2x)	G	R
PchE-cyc	F	E	L	S	(2x)	Q	Q	A	Y	(2x)	G	R
Zbm-cyc	F	P	L	T	(2x)	Q	Q	A	Y	(2x)	G	R
SPR_6230	A	P	P	S	(2x)	Q	E	E	H	(2x)	Q	A
SPR_6240	A	P	L	S	(2x)	Q	E	E	M	(2x)	N	I
SPR_6360	G	P	L	S	(2x)	Q	R	R	F	(2x)	A	E
SPR_53040	G	P	L	S	(2x)	Q	L	G	M	(2x)	L	E
SPR_53060	-	P	L	T	(2x)	Q	W	R	M	(2x)	H	H
SPR_52900	A	S	A	T	(2x)	Q	R	Q	M	(2x)	L	I

Cyc motif 2	**R**	**H**	**P/A/L**	**G**	**x**	**Q**
VirH-cyc	H	H	R	V	H	Q
NRPS-cyc	R	H	R	I	H	Q
blmIV-cyc	R	H	R	V	L	P
BacA-cyc	R	H	R	V	F	E
EposP-cyc	R	H	R	T	L	P
PchE-cyc	R	H	R	F	F	D
Zbm-cyc	R	H	R	I	D	A
SPR_6230	R	H	R	F	R	T
SPR_6240	R	H	R	I	A	Y
SPR_6360	R	H	R	Y	P	W
SPR_53040	R	H	R	F	T	V
SPR_53060	R	H	R	L	A	T
SPR_52900	R	H	E	S	L	R

Cy motif 3	**D**	**4x**	**D**	**2x**	**S**
VirH-cyc	D	(4x)	D	(2x)	S
NRPS-cyc	D	(4x)	D	(2x)	S
blmIV-cyc	D	(4x)	D	(2x)	S
BacA-cyc	D	(4x)	D	(2x)	S
EposP-cyc	D	(4x)	D	(2x)	S
PchE-cyc	D	(4x)	D	(2x)	S
Zbm-cyc	D	(4x)	D	(2x)	S
SPR_6230	D	(4x)	D	(2x)	S
SPR_6240	H	(4x)	D	(2x)	S
SPR_6360	H	(4x)	D	(2x)	S
SPR_53040	D	(4x)	W	(2x)	V
SPR_53060	H	(4x)	D	(2x)	S
SPR_52900	S	(4x)	A	(2x)	S

Cy motif 4	**L**	**P**	**2x**	**P**	**x**	**L**	**P**	**L**	**3x**	**P**
VirH-cyc	L	P	(2x)	P	P	L	P	D	(3x)	Q
NRPS-cyc	L	P	(2x)	P	E	L	P	L	(3x)	A
blmIV-cyc	L	P	(2x)	P	G	L	P	L	(3x)	P
BacA-cyc	F	P	(2x)	P	E	L	P	L	(3x)	P
EposP-cyc	L	P	(2x)	P	T	L	P	M	(3x)	P
PchE-cyc	L	P	(2x)	P	A	L	P	L	(3x)	P
Zbm-cyc	L	P	(2x)	P	E	L	P	F	(3x)	P
SPR_6230	L	S	(2x)	L	L	P	Q	A	(3x)	P
SPR_6240	L	A	(2x)	T	G	E	V	A	(3x)	P
SPR_6360	A	A	(2x)	P	V	L	P	T	(3x)	N
SPR_53040	V	A	(2x)	P	E	L	T	L	(3x)	T
SPR_53060	A	P	(2x)	T	R	L	P	T	(3x)	P
SPR_52900	A	P	(2x)	M	F	P	G	R	(3x)	G

Cy motif 5	**T/S**	**P/A**	**3x**	**L/A/F**	**6x**	**I/V/T**	**L**	**2x**	**W**
VirH-cyc	P	A	3x	L	6x	E	L	2x	T
NRPS-cyc	T	P	3x	L	6x	T	V	2x	W
blmIV-cyc	T	P	3x	I	6x	V	L	2x	W
BacA-cyc	T	P	3x	L	6x	I	L	2x	W
EposP-cyc	T	P	3x	I	6x	V	I	2x	W
PchE-cyc	T	L	3x	F	6x	V	L	2x	W
Zbm-cyc	T	P	3x	L	6x	I	V	2x	W
SPR_6230	T	L	3x	L	6x	A	S	2x	I
SPR_6240	S	P	3x	L	6x	A	I	2x	V
SPR_6360	A	T	3x	C	6x	G	R	2x	P
SPR_53040	T	L	3x	L	6x	V	L	2x	H
SPR_53060	T	P	3x	I	6x	V	A	2x	S
SPR_52900	S	T	3x	L	6x	L	L	2x	R

Cy motif 6	**G/A**	**D**	**F**	**T**	**x**	**L**	**x**	**L**	**L**
VirH-cyc	A	D	F	T	Q	L	A	W	V
NRPS-cyc	G	D	F	T	S	L	E	L	L
blmIV-cyc	G	D	F	T	T	T	T	L	L
BacA-cyc	G	D	F	T	S	L	M	L	L
EposP-cyc	G	D	F	T	S	M	V	L	L
PchE-cyc	A	D	F	T	T	L	L	L	L
Zbm-cyc	G	D	F	T	S	L	I	P	L
SPR_6230	G	W	Y	I	N	A	A	P	I
SPR_6240	G	H	L	L	N	T	R	L	T
SPR_6360	V	V	D	T	A	A	D	Q	V
SPR_53040	G	L	F	V	N	M	L	P	I
SPR_53060	G	L	F	T	H	T	V	P	L
SPR_52900	A	N	L	H	Q	E	V	L	T

Cy motif 7	**P**	**V**	**V**	**F**	**T**	**S**	**x**	**L**
VirH-cyc	P	V	V	F	T	R	V	P
NRPS-cyc	P	V	V	F	T	S	A	L
blmIV-cyc	P	V	V	F	T	S	T	L
BacA-cyc	P	I	V	F	T	S	V	L
EposP-cyc	P	V	V	L	T	S	A	L
PchE-cyc	P	V	V	F	A	S	N	L
Zbm-cyc	P	V	V	F	T	S	N	L
SPR_6230	S	Y	L	D	F	R	P	L
SPR_6240	L	S	P	E	W	E	A	Q
SPR_6360	S	S	V	R	V	R	P	P
SPR_53040	I	N	V	C	V	S	F	Q
SPR_53060	Q	M	V	F	C	H	G	A
SPR_52900	T	V	V	A	S	S	D	F

Cy motif 8	S/T	Q/R	T	P	Q	V	x	L/I	D	13x	W	D
VirH-cyc	S	Q	T	A	Q	V	A	L	D	13x	W	D
NRPS-cyc	T	R	T	P	Q	V	W	L	D	13x	W	D
blmIV-cyc	S	Q	T	P	Q	V	L	L	D	13x	W	D
BacA-cyc	T	R	T	S	Q	V	Y	I	D	13x	W	D
EposP-cyc	T	Q	T	P	Q	L	L	L	D	13x	W	D
PchE-cyc	S	Q	T	P	Q	V	W	L	D	13x	W	D
Zbm-cyc	-	-	-	-	-	-	-	-	-	-	-	-
SPR_6230	S	G	T	G	I	D	L	F	L	13x	P	D
SPR_6240	S	W	R	S	F	A	V	L	W	13x	R	P
SPR_6360	Q	W	R	E	D	V	M	D	R	13x	A	E
SPR_53040	S	P	F	D	L	D	L	G	F	13x	N	P
SPR_53060	A	K	F	D	V	T	V	M	V	13x	Y	D
SPR_52900	C	A	E	T	A	A	L	S	P	13x	L	V

Figure 3.7 A multiple sequence alignment of the conserved signature motifs of the putative *S. polyantibioticus* SPR^T Cy domains, encoded by SPR_6230, SPR_6240, SPR_6360, SPR_53040, SPR_53060 and SPR_52900, and reference Cy domain amino acid sequences taken from the *S. virginiae* NRRL B-1446^T VirH sequence (VirH-cyc) (AB283030), *S. coelicolor* A3(2) putative non-ribosomal peptide synthetase sequence (NRPS-cyc) (CAC17500), *S. verticillatus* strain ATCC 15003 BlmIV sequence (blmIV-cyc) (AAG02364), *Pseudomonas aeurginosa* PA14 PchE sequence (PchE-cyc) (AAD55800), *Bacillus licheniformis* ATCC 10716 BacA sequence (BacA-cyc) (O68006), *Sorangium cellulosum* So ce90 EposP sequence (EposP-cyc) (AAF26925) and the *Streptomyces flavoviridis* ATCC21892 Zbm sequence (Zbm-cyc) (EU670723). Residues critical for the domain function are coloured green while other residues corresponding to the consensus motif are coloured yellow.

3.4.4.3 PCP DOMAINS

The peptidyl carrier protein enables the transfer of activated amino acids and elongation intermediates between catalytic centres. The PCP domain is characterised by a conserved serine residue, which accepts a phosphopantetheine moiety as a prosthetic group, thereby converting the inactive *apo*-PCP to its active *holo*-form. A variant of the PCP has been identified in the synthetases of siderophores and other aryl-capped non-ribosomal peptides for the incorporation of aromatic carboxy acids, where aryl acids such as salicylate and 2,3-dihydroxybenzoic acid are tethered on ArCPs (Crosa & Walsh, 2002). In particular, freestanding A domains are known to activate aromatic carboxylic acids and transfer them to aryl-carrier proteins (ArCP), which are used as starter substrates in, for example, quinoxaline antibiotics (Crawford *et al.,* 2011).

Marahiel *et al.* (1997) identified the signature motif, L-G-G-(D/H)-S-L, in all PCP and ArCP domains. The antiSMASH analysis of the draft *S. polyantibioticus* SPR^T genome identified two PCP domains in the putative DPO gene cluster (Table 3.2), which were examined in a multiple sequence alignment with reference ArCP domains from *Yersinia enterocolitica* subsp. *enterocolitica* 8081 (GenBank accession number: YE2617), *Brucella melitensis* bv. 1 str. 16M (BMEII0079), *Bacillus subtilis* (BSU31970), *Escherichia coli* (P0ADI4), *Pseudomonas aeruginosa* PA14 (AAD55800), *Salmonella enterica* subsp. *enterica* (STM0597) and *Streptomyces pyridomyceticus* (AEF33099). The alignment revealed the presence of the critical catalytic serine residue in the putative ArCP domain encoded by SPR_53060, and the absence of this serine residue in the putative PCP domain encoded by SPR_53070 (Figure 3.8). The signature motif is well conserved in the SPR_53060 domain and very poorly conserved in the SPR_53070 domain, with the exception of the final leucine residue, suggesting the possible inactivity/redundancy of the domain encoded by SPR_53070. Indeed, the essential serine residue is replaced by an arginine residue, which would prevent the attachment of the phosphopantetheine prosthetic group.

PCP motif	**L**	**G**	**G**	**D/H**	**S**	**L**
Y.enterocolitica-ArCP	A	G	L	D	S	I
B.melitensis-ArCP	Y	G	L	D	S	L
B.subtilis-ArCp	R	G	L	D	S	V
E.coli-ArCp	Y	G	L	D	S	V
P.aeruginosa-ArCP	C	G	L	D	S	I
S.enteric-ArCP	Y	G	L	D	S	V
S.pyridomyceticus-ArCP	L	G	L	D	S	I
SPR_53060	R	G	G	D	S	L
SPR_53070	Q	A	I	G	R	L

Figure 3.8 A multiple sequence alignment of the conserved signature motifs of the putative *S. polyantibioticus* SPR^T ArCP domain encoded by SPR_53060 and the PCP domain encoded by SPR_53070 with reference ArCP domains from *Yersinia enterocolitica* subsp. *enterocolitica* 8081 (*Y.enterocolitica*-ArCP), *Brucella melitensis* bv. 1 str. 16M (*B.melitensis*-ArCP), *Bacillus subtilis* (*B.subtilis*-ArCP), *Escherichia Coli* (*E.coli*-ArCP), *Pseudomonas aeruginosa* PA14 (*P.aeruginosa*-ArCP), *Salmonella enterica* subsp. *enterica* (*S.enteric*-ArCP)) and *Streptomyces pyridomyceticus* (*S.pyridomyceticus*-ArCP). Residues critical for the domain function are coloured green while other residues corresponding to the consensus motif are coloured yellow.

Due to the fact that both PCP and ArCP domains share the same conserved signature motif, a phylogenetic analysis was performed (Figure 3.9) in order to determine the evolutionary relatedness of the putative ArCP domain encoded by SPR_53060 and the PCP domain encoded by SPR_53070 to the reference ArCP domains used in the analysis of the conserved signature motifs (Figure 3.8). The PCP domains from *Streptomyces netropsis* DSM 40846 (GenBank accession number: AHF81539), *Streptomyces* sp. CNS615 (WP_037734274), *Streptomyces himastatinicus* ATCC 53653 (WP_009720528), *S.coelicolor* A3(2) (NP_631722), *Streptomyces lividans* 1326 (EOY52612) were added to the phylogenetic analysis.

The phylogenetic analysis, using the NJ, ML and MP algorithms, resolved the putative ArCP and PCP domains from *S. polyantibioticus* SPR^T and the reference sequences into two distinct groups. The reference ArCP domains clustered together in all three trees (with very high bootstrap support (98 %) in the NJ tree) and the reference PCP domains clustered together with the amino acid sequences of SPR_53060 and SPR_53070. There was strong bootstrap support, 93 % and 67 %, respectively, for the association of SPR_53060 and SPR_53070 with known PCP domain sequences.

This analysis indicated that the domain encoded by SPR_53060 is in fact a PCP domain and not an ArCP, as originally suggested by the conserved signature motif analysis.

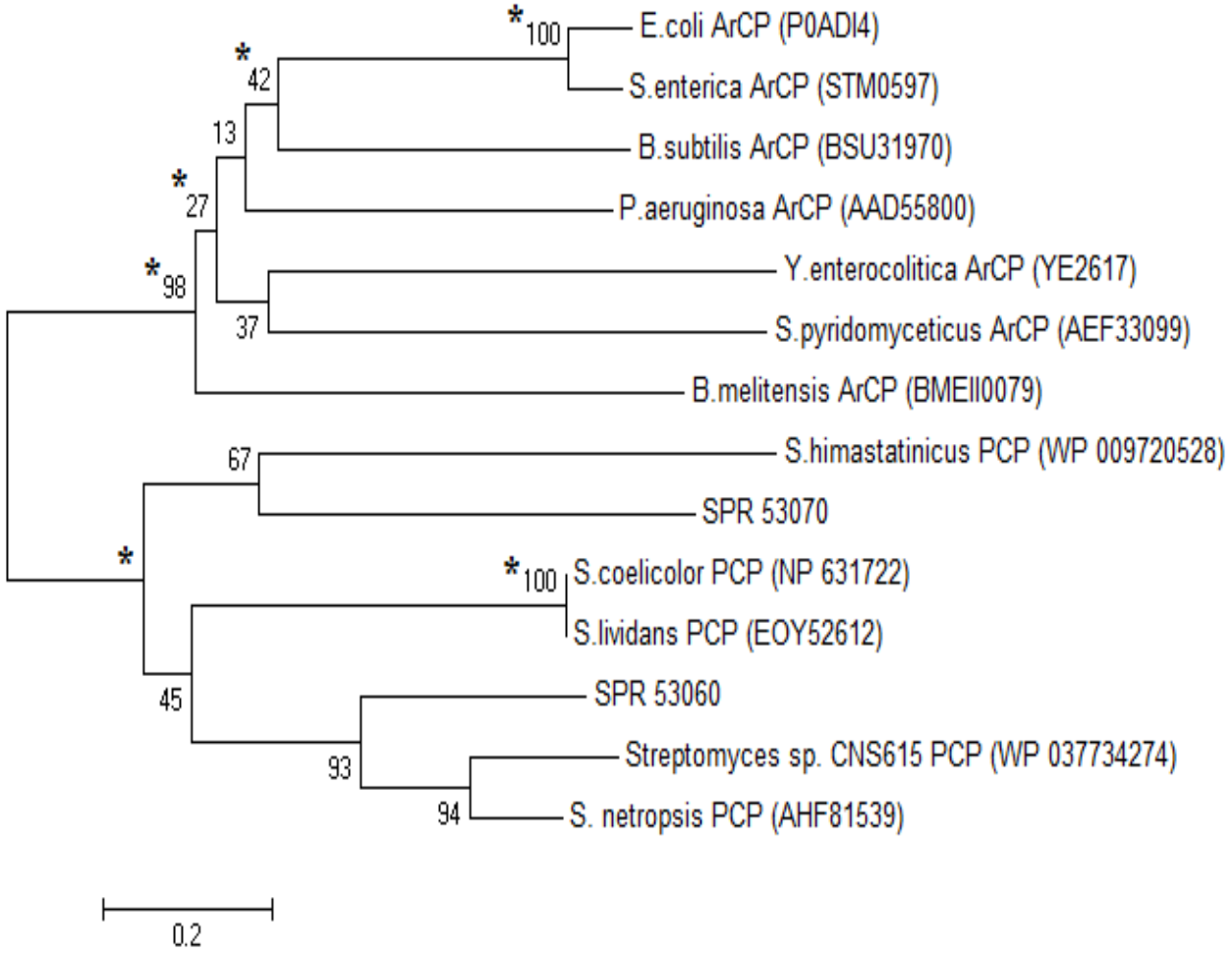

Figure 3.9 Unrooted, neighbour joining tree (Saitou & Nei, 1987) obtained from an alignment of the amino acid sequences of the putative *S. polyantibioticus* SPR^T ArCP domain encoded by SPR_53060 and the PCP domain encoded by SPR_53070 with reference ArCP domains from *Yersinia enterocolitica* subsp. *enterocolitica* 8081 (Y.enterocolitica-ArCP), *Brucella melitensis* bv. 1 str. 16M (B.melitensis-ArCP), *Bacillus subtilis* (B.subtilis-ArCP), *Escherichia Coli* (E.coli-ArCP), *Pseudomonas aeruginosa* PA14 (P.aeruginosa-ArCP), *Salmonella enterica* subsp. *enterica* (S.enterica-ArCP)) and *Streptomyces pyridomyceticus* (S.pyridomyceticus-ArCP), in addition to PCP domains from *Streptomyces netropsis* DSM 40846 (S.netropsis PCP), *Streptomyces* sp. CNS615 (Streptomyces sp. CNS615 PCP), *Streptomyces himastatinicus* ATCC 53653 (S.himastatinicus PCP), *S.coelicolor* A3(2) (S.coelicolor PCP) and *Streptomyces lividans* 1326 (S.lividans PCP). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches and the GenBank accession numbers are displayed in parenthesis. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site (Zuckerkandl & Pauling, 1965). The analysis involved 14 amino acid sequences and all positions containing gaps and missing data were eliminated. Asterisks indicate clades that were conserved in neighbour-joining, maximum-likelihood and maximum-parsimony trees. Scale bar, 20 amino acid substitutions per 100 amino acids.

3.4.4.4 OX DOMAINS

An Ox domain is required during DPO biosynthesis for the oxidation of the initial unstable oxazoline ring to an oxazole ring, as postulated in Chapter 2 (section 2.2.1). Ox domains have been identified as members of the McbC-like oxidoreductase superfamily and are normally associated with heterocyclization modules (Marchler-Bauer *et al.,* 2011).

A BLASTP search against the *S. polyantibioticus* SPR^T draft genome, using well-characterised Ox domain amino acid sequences from: *Streptomyces hygroscopicus* McbC-like oxidoreductase (GenBank accession number: ACS50132), *Streptomyces* sp. NRRL WC-3773 NADH oxidase (WP_031005407), *Streptomyces verticillus* ATCC15003 peptide synthetase (AAG02365), *S. cellulosum* epothilone biosynthetic cluster (AAF62881) and *Angiococcus disciformis* tubulysin biosynthetic cluster (CAF05649), revealed the existence of putative Ox domains within the genome encoded by the genes SPR_01300, SPR_37410, SPR_39290, SPR_39400, SPR_44390, SPR_49120 and SPR_53180 amongst others. However, no Ox domains were found within the putative DPO biosynthetic cluster (section 3.4.4). This suggests that the oxidation activity required in DPO biosynthesis may be provided by an external tailoring enzyme. Indeed, the *in trans* thiazoline to thiazole oxidase activity of the *S. cellulosum* EpoB Ox domain was demonstrated by Schneider *et al.* (2003) who heterologously expressed this domain in *E. coli.* Additionally, the mechanism of the oxidation process remains unclear, but it does share a clear analogy to the oxidation of dihydroorotate to orotate observed in pyrimidine biosynthesis, in which an FMN co-factor-mediated redox step is involved (Schenider *et al.,* 2003). In light of this, it is interesting to note that a homologue of the key enzyme involved in this reaction, dihydroorotate dehydrogenase, is encoded by the gene SPR_52930, identified as a member of the putative DPO biosynthetic pathway. Although DPO is not a pyrimidine, both compounds are aromatic heterocycles and therefore the presence of a homologue of dihydroorotate dehydrogenase within the putative DPO biosynthetic cluster may indicate its involvement in the oxidation process required for DPO biosynthesis.

Lastly, in contrast to all other NRPS domains, whose relative position within a given module is generally highly conserved, the location of the EpoB and BlmIII Ox domains from *S. cellulosum* and *Streptomyces verticillatus* ATCC 15003, respectively, suggest that Ox domains can be located either upstream or downstream of the PCP domains presenting thiazolinyl-*S*-pantotheinyl chains for oxidation. This may be indicative of the fact that Ox-PCP

and PCP-Ox pairs may be transferable to other Cy-containing modules in NRPS assembly lines, thereby preserving the ability to direct the catalysis of thiazolinyl and oxazolinyl-*S*-enzyme intermediates to heterocycle oxidation states (Schneider *et al.,* 2003). A similar atypical positioning of the Ox domain may also occur in the genes for DPO biosynthesis in *S. polyantibioticus* SPR^T.

3.4.4.5 TE DOMAIN

The gene SPR_53090 was identified by antiSMASH to encode a TE domain. These domains share the same G-x-S-x-G motif, described by Schwarzer *et al.* (2003), as the acyltransferase enzymes found in PKS systems that catalyse the transfer of acyl moieties between CoA and ACP using a serine-histidine catalytic diad (Yadav *et al*., 2003).

A multiple sequence alignment of SPR_53090 with known TE domains from *Bacillus* and *Streptomyces* species revealed the highly conserved G-x-S-x-G signature motif in SPR_53090 (Figure 3.10), thereby indicating that it is likely responsible for the thioesterase activity in the proposed DPO biosynthetic gene cluster.

Te motif	**G**	**x**	**S**	**x**	**G**
B.cereus-Te	G	H	S	M	G
B.licheniformis-Te	G	H	S	M	G
S.coelicolor-Te	G	H	S	M	G
S.aureofaciens-Te	G	H	S	M	G
SPR_53090	G	H	S	L	G

Figure 3.10 A multiple sequence alignment of the conserved signature motif of the putative DPO TE domain from *S. polyantibioticus* SPR^T and various TE reference sequences taken from *S. coelicolor* A3(2) (*S.coelicolor*-Te) (GenBank ccession number: NP_631726), *S. aureofaciens* (*S.aureofaciens*-Te) (YP_009060793), *Bacillus cereus* AH1134 (*B.cereus*-Te) (EDZ49042) and *B. licheniformis* IBL200 (*B.licheniformis*-Te) (AAU22005). Residues critical for the domain function are coloured green while other residues corresponding to the consensus motif are coloured yellow.

3.4.4.6 GENES INVOLVED IN BENZOIC ACID BIOSYNTHESIS

In addition to identifying the NRPS-associated genes involved in DPO biosynthesis, the genes for benzoic acid biosynthesis also need to be identified. Despite its common occurrence in plant metabolites, benzoic acid biosynthesis is not common in bacteria (Moore *et al*., 2002). However, the biosynthesis of benzoate as a starter unit in the biosynthesis of enterocin in '*Streptomyces maritimus*' strain DSM 41777^T prompted a search of the draft *S. polyantibioticus* SPR^T genome for an orthologue of the *encP* gene, encoding the key enzyme, phenylalanine ammonia-lyase. An *encP* homologue is not present in the putative DPO biosynthetic cluster, or anywhere else in the *S. polyantibioticus* SPR^T genome.

The annotation of the draft genome of *S. polyantibioticus* SPR^T revealed the presence of an enoyl-CoA hydratase, cinnamate-CoA ligase, and an acyl-CoA dehydrogenase encoded by genes SPR_60140, SPR_60150 and SPR_60160, respectively. The enoyl-CoA hydratase and cinnamate CoA ligase shared 25 % and 27 % homology, respectively, to the homologous proteins found in the benzoate biosynthetic pathway identified in '*S. maritimus*' strain DSM 41777^T.

Despite the fact that a PAL homologue was not identified in the *S. polyantibioticus* SPR^T draft genome, an alternative route to cinnamic acid was proposed whereby phenylalanine would be catabolized to cinnamic acid via phenylpyruvate and phenyllactate (Figure 3.9). Indeed, several catabolic pathways for L-phenylalanine have been reported in various organisms, including *Streptomyces setonii* 75Vi2, which metabolises L-phenylalanine via phenylpyruvate and phenyllactate (Pometto III & Crawford, 1985). Moreover, it has been reported that lactic acid bacteria catabolize phenylalanine to phenylpyruvate via an aminotransferase, after which phenylpyruvate is catabolized to either D- or L-phenyllactate via dehydrogenase activity (Mu *et al*., 2012). It has also been demonstrated that a D-lactate dehydrogenase from *Leuconostoc mesenteroides* can reduce phenylpyruvate to phenyllactate (Simon *et al.,* 1989).

In light of these reports and the fact that a putative aminotransferase encoded by SPR_60250 and a putative D-lactate dehydrogenase encoded by SPR_60260 were clustered together with the genes encoding the enoyl-CoA hydratase, cinnamate-CoA ligase, and an acyl-CoA dehydrogenase mentioned above, it was decided to explore the possibility of these genes being involved in the biosynthesis of benzoate and therefore DPO production, as functionally related

genes are usually clustered in bacterial genomes (Kaneko *et al.,* 2003). Therefore, the genes encoding the cinnamate-CoA ligase and D-lactate dehydrogenase were identified as putative external constituents of the DPO biosynthetic pathway. The investigation into their involvement in the DPO biosynthetic pathway is explored in Chapter 4.

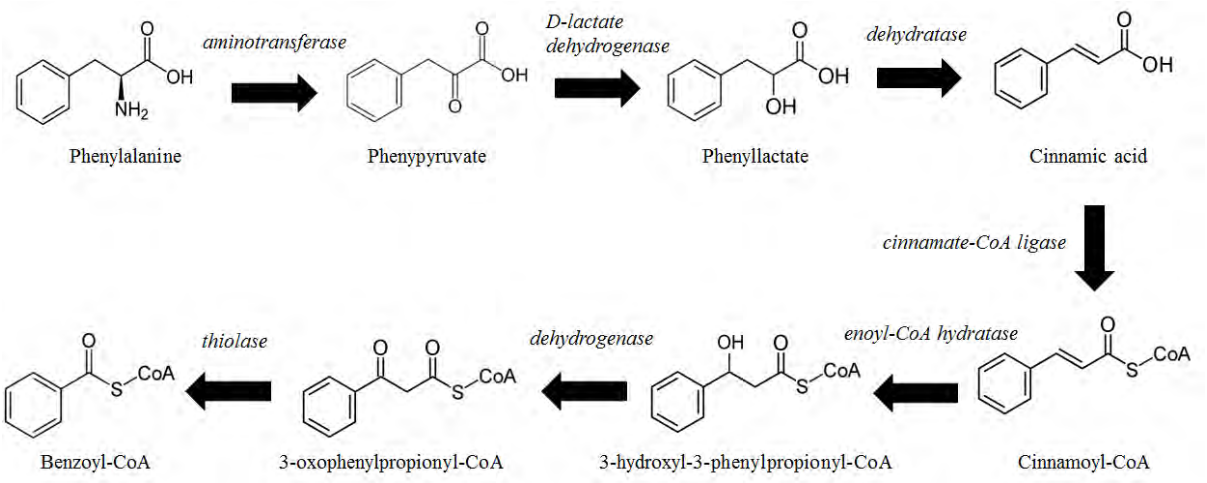

Figure 3.9 Proposed biosynthetic pathway for benzoyl-CoA production in *S. polyantibioticus* SPR^T.

Lastly, the gene coding for the phenylacetate-coenzyme A ligase, *paaK*, which was identified in the *S. polyantibioticus* SPR^T genome via PCR amplification and sequencing (Chapter 2), was identical to the sequence of the gene SPR_46390 obtained from the annotation of the draft *S. polyantibioticus* SPR^T genome. The antiSMASH 3.0 (Weber *et al.,* 2015) analysis identified the cluster that the *paaK* gene was located in as a PKS/NRPS hybrid encoding an antimycin-type antibiotic. The composition of the gene cluster was most similar in gene content and gene order to the antimycin biosynthetic gene cluster from *Streptomyces blastmyceticus,* as 100 % of the *S. polyantibioticus* SPR^T genes shared homology to genes within this cluster but, due to the complicated structure of antimycin (PubChem ID: 14957) (Figure 3.10), it was deemed unlikely that this cluster would be involved in DPO biosynthesis. However, further efforts to characterize the involvement of *paaK* in the synthesis of DPO are described in Chapter 4.

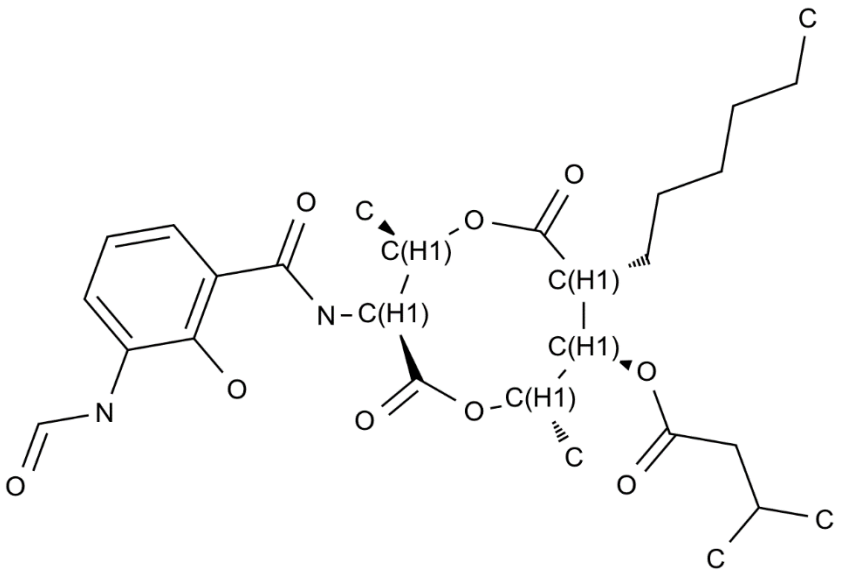

Figure 3.10 Outline of the *S. polyantibioticus* SPR^{T} antimycin structure predicted by antiSMASH 3.0.

3.5 CONCLUSION

The draft *S. polyantibioticus* SPR^T genome sequence was obtained using a hybrid sequencing approach, after which the sequence was annotated and secondary metabolite gene clusters were identified using antiSMASH 3.0 analysis (Weber *et al.,* 2015). The putative DPO biosynthetic gene cluster was identified using a genome mining approach in which each of the six NRPS-containing gene clusters was assessed for the likelihood of it being responsible for producing DPO, based on the structure of DPO and the gene content of each gene cluster.

The putative DPO biosynthetic cluster shared a high degree of homology to the biosynthetic gene cluster for the pyrrole-amide, congocidine, identified in *S. ambofaciens* ATCC 23877^T. The putative acyl-CoA synthetase encoded by SPR_52860, the A domain encoded by SPR_53060, the putative Cy domain encoded by SPR_53040, the TE domain encoded by SPR_53090, the cinnamate-CoA ligase encoded by SPR_60150 and the putative D-lactate dehydrogenase encoded by SPR_60260 were identified as genes that would be disrupted in order to determine their involvement in DPO biosynthesis. The development of a DNA transformation method for the introduction of exogenous DNA into *S. polyantibioticus* SPR^T and subsequent gene disruption experiments are described in the following chapter.

3.6 REFERENCE LIST

Aigle, B., Lautru, S., Spiteller, D., Dickschat, J. S., Challis, G. L., Leblond, P. & Pernodet, J. L. (2014). Genome mining of *Streptomyces ambofaciens*. *Journal of industrial microbiology & biotechnology*, **41(2):** 251-263.

Bachmann, B.O. & Ravel, J. (2009). *In Silico* Prediction of Microbial Secondary Metabolic Pathways from DNA Sequence Data. *Methods in Enzymology*, **458:** 181-217.

Bachmann, B.O., Van Lanen, S.G. & Baltz, R.H. (2014). Journal of Industrial Microbiology and Biotechnology, **41:** 175-184.

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

Baranasic, D., Zucko, J., Diminic, J., Gacesa, R., Long, P.F., Cullum, J., Hranueli, D. & Starcevic, A. (2013). Predicting substrate specificity of adenylation domains of nonribosomal peptide synthetases and other protein properties by latent semantic indexing. *Journal of Industrial Microbiology and Biotechnology*, DOI 10.1007/s10295-013-1322-2.

Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J. & Hopwood DA. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, **417(6885):**141–147.

Bérdy, J. (2012). Thoughts and facts about antibiotics: Where we are now and where we are heading. *The Journal of Antibiotics*, **65:** 385-395.

Boddy, C.N. (2014). Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. *Journal of Industrial Microbiology and Biotechnology*, **41:** 443-450.

Christiansen, G., Philmus, B., Hemscheidt, T. & Kurmayer, R. (2011). Genetic Variation of Adenylation Domains of the Anabaenopeptin Synthesis Operon and Evolution of Substrate Promiscuity. *Journal of Bacteriology*, **193(15):** 3822–3831.

Crawford, J.M., Portmann, C., Kontnik, R., Walsh, C.T. & Clardy, J. (2011). NRPS Substrate Promiscuity Diversifies the Xenematides. *Organic Letters*, **13(19):** 5144-5147.

Crosa, J.G. & Walsh, C.T. (2002). Genetics and Assembly Line Enzymology of Siderophore Biosynthesis in Bacteria. *Microbiology and Molecular Biology Reviews*, **66 (2):** 223-249.

Delcher, A.L, Bratke, K.A., Powers, E.C. & Salzberg, S.L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics,* **23:** 673–679.

Di Lorenzo, M., Stork, M., Naka, H., Tolmasky, M. E. & Crosa, J. H. (2008). Tandem heterocyclization domains in a nonribosomal peptide synthetase essential for siderophore biosynthesis in Vibrio anguillarum. *Biometals*, **21(6):** 635-648.

Disz, T., Akhter, S., Cuevas, D., Olson, R., Overbeek, R., Vonstein, V., Stevens, R. & Edwards, R.A. (2010). Accessing the SEED genome databases via Web ser vices API: tools for programmers. *BMC Bioinformatics,* **11 (31):** 1-11

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

Felsenstein J. (**1985**). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution,* **39**:783-791.

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

Glenn, T.C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, **11:** 759–769.

Harrison, J., & Studholme, D. J. (2014). Recently published Streptomyces genome sequences. *Microbial biotechnology*, **7(5):** 373-380.

Hur, G.H., Vickery, C.R. & Burkart, M.D. (2012). Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Natural Products Reports*, **29:** 1074-1098.

Ikeda, H., Shin-ya, K. & Omura, S. (2014). Genome mining of the *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. *Journal of industrial microbiology & biotechnology*, **41(2):** 233-250.

Jankowitsch, F., Schwarz, J., Rückert, C., Gust, B., Szczepanowski, R., Blom, J., Pelzer, S., Kalinowski, J. & Mack, M. (2012). Genome Sequence of the Bacterium *Streptomyces davawensis* JCM 4913 and Heterologous Production of the Unique Antibiotic Roseoflavin. *Journal of Bacteriology*, **192(24):** 6818-6827.

Jensen, PR, Chavarria KL, Fenical W, Moore BS, Ziemert N. (2014). Challenges and triumphs to genomics-based natural product discovery. *Journal of Industrial Microbiology & Biotechnology*. **41:** 203-209.

Jones, D.T., Taylor, W.R., Thornton, J.M. (**1992**). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences,* **8**: 275-282.

Juguet, M., Lautru, S., Francou, F. X., Nezbedová, Š., Leblond, P., Gondry, M., Pernodet, J. L. (2009). An iterative nonribosomal peptide synthetase assembles the pyrrole-amide antibiotic congocidine in *Streptomyces ambofaciens*. *Chemistry & biology*, **16(4):** 421-431.

Kaneko, M., Hwang, E. I., Ohnishi, Y., Horinouchi, S. (2003). Heterologous production of flavanones in *Escherichia coli*: potential for combinatorial biosynthesis of flavonoids in bacteria. *Journal of Industrial Microbiology and Biotechnology*, **30(8):** 456-461.

Keating, T.A., Marshall, C.G., Walsh, C.T. & Keating, A.E. (2002). Excision of the Epothilone Synthetase B Cyclization Domain and Demonstration of in Trans Condensation/Cyclodehydration Activity. *Nature Structural Biology*, **9(7):** 522-526.

Kelly, W.L., Hillson, N.J. & Walsh, C.T. (2005). Excision of the Epothilone Synthetase B Cyclization Domain and Demonstration of in Trans Condensation/Cyclodehydration Activity. *Biochemistry*, **44:** 13385-13393.

Khayatt, B. I., Overmars, L., Siezen, R. J. & Francke, C. (2013). Classification of the adenylation and acyl-transferase activity of NRPS and PKS systems using ensembles of substrate specific hidden Markov models. *PloS one*, **8(4):** 2136-2140.

Lagesen K, Hallin, P., Rødland, E.A., Staerfeldt, H.H., Rognes, T. & Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, **35:** 3100–3108.

Lautru, S., Song, L., Demange, L., Lombès, T., Galons, H., Challis, G. L. & Pernodet, J. L. (2012). A Sweet Origin for the Key Congocidine Precursor 4-Acetamidopyrrole-2-carboxylate. *Angewandte Chemie International Edition*, **51(30):** 7454-7458.

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.

Levene, M. J., Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G. & Webb, W. W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science*, **299(5607):** 682-686.

Li, M. H., Ung, P. M., Zajkowski, J., Garneau-Tsodikova, S. & Sherman, D. H. (2009). Automated genome mining for natural products. *BMC bioinformatics*, **10(1),**:185.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. & Law, M. (2012). Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*, **1:** 1-11.

Lodish, H., Berk, A., Zipursky, S.,L., Matsudaira, P., Baltimore, D. & Darnell, J. (2000). Molecular Cell Biology, 4th edition. New York: W. H. Freeman.

Loman, N.J., Constantinidou, C., Chan, J.Z., Halachev, M., Sergeant, M., Penn, C.W., Robinson, E.R. & Pallen, M.J. (2012). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature reviews*, **10:** 599-606.

Lowe, T.M. & Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*. **25:** 955–964

Marahiel, M.A. (1997). Protein templates for the biosynthesis of peptide antibiotics. *Chemistry & Biology*, **4(8):** 561-567.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, Lu, F., Marchler, G.H., Mullokandov, M., Omelchencko, M.V., Robertson, C.L., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, N., Zheng, C. & Bryant, S.H. (2011) . CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research,* **39**: 225–229.

Mardis, E.R. (2013). Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry*, **6:** 287–303.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C. & He, W. (2005). Genome sequencing in microfabricated high-density picolitre reactors, *Nature*, **437:** 376–380.

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A. & Breitling, R. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic acids research*, **39(2):** W339-W346.

Metzker, M.L. (2011). Sequencing technologies – the next generation. *Nature Reviews*, **11:** 31-46.

Moore, B. S., Hertweck, C., Hopke, J. N., Izumikawa, M., Kalaitzis, J. A., Nilsen, G. & Noel, J. P. (2002). Plant-like biosynthetic pathways in bacteria: from benzoic acid to chalcone 1. *Journal of natural products*, **65(12):** 1956-1962.

Morozova, M.A., & Marra, M.A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*, **92:** 255-264.

Mu, W., Yu, S., Zhu, L., Zhang, T. & Jiang, B. (2012). Recent research on 3-phenyllactic acid, a broad-spectrum antimicrobial compound. *Applied microbiology and biotechnology*, **95(5):** 1155-1163.

Nei M., & Kumar S. (**2000**). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Nett, M., Ikeda, H. & Moore, B., S. (2009). Genomic basis for natural product biosynthetic diversity in the actinomycetes. Natural Product Reports. **26:** 1362–1384.

Niedringhaus, T.P., Milanova, D., Kerby, M.B., Snyder, M.P. & Barron, A.E. (2011). Landscape of Next-Generation Sequencing Technologies. *Analytical Chemistry*, **15:** 4327-4341.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F. & Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, **42:** D206–D214.

Pohle, S., Appelt, C., Roux, M., Fiedler, H.P. & Süssmuth, R.D. (2011). Biosynthetic gene cluster of the non-ribosomally synthesized cyclodepsipeptide skyllamycin: deciphering unprecedented ways of unusual hydroxylation reactions. *Journal of the American Chemical Society,* **133(16):** 6194-205.

Pometto III, A.L., & Crawford, D.L. (1985). L-Phenylalanine and L-tyrosine catabolism by selected Streptomyces species. *Applied Environmental Microbiology*, **49(3):** 727-729.

Prieto, C., García-Estrada, C., Lorenzana, D., Martín, J. F. (2012). NRPSsp: non-ribosomal peptide synthase substrate predictor. *Bioinformatics*, **28(3):** 426-427.

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., & Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics,* **13**: 341-346.

Quirós, L. M., Aguirrezabalaga, I., Olano, C., Méndez, C. & Salas, J. A. (1998). Two glycosyltransferases and a glycosidase are involved in oleandomycin modification during its biosynthesis by Streptomyces antibioticus. *Molecular microbiology*, **28(6):** 1177-1185.

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.

Roberts, R.J., Carneiro, M.O. & Schatz, M.C. (2013). The advantages of SMRT sequencing. *Genome Biology*, **14:** 405-410.

Röttig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C. & Kohlbacher, O. (2011). NRPSpredictor2 – a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Research*, 1-6.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406 – 425.

Schadt, E.E., Turner, S. & Kasarskis, A. (2010). A window into third-generation sequencing. *Human Molecular Genetics*, **19(2):** R227–R240

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

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

Shendure, J., Porreca, G.J, Reppas, N.B, Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D. & Church, G.M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome, *Science*, **309:** 1728–1732.

Simon, E. S., Plante, R. & Whitesides, G. M. (1989). D-lactate dehydrogenase. *Applied biochemistry and biotechnology*, **22(2):** 169-179.

Siqueira, J.F., Fouad, A.F. & Rocas, I.N. (2012). Pyrosequencing as a tool for better understanding of human microbiomes. *Journal of Oral Microbiology*, **4:** 1-15

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.

Steinbock, L.J., & Radenovic, A. (2015). The emergence of nanopores in next-generation sequencing. *Nanotechnology*, **26(7):** 074003.

Takahashi, K., & Nei, M. (2000). Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Molecular Biology and Evolution*, **17(8):** 1251-1258.

Tamura K., Stecher G., Peterson D., Filipski A. & Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution,* **30:** 2725-2729.

Tatusov, R.L., Koonin, E.V., Lipman, D.J. (1997). A genomic perspective on protein families. *Science,* **278:** 631–637.

Tawflik, D.S. & Griffiths, A.D. (1998). Man-made cell-like compartments for molecular evolution, *Nature Biotechnology*, **16:** 652–656.

Udwary, D.W., Zeigler, L., Asolkar, R.N., Singan, V., Lapidus, A., Fenical, W., Jensen, P.R. & Moore, B.S. (2007). Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica. Proceedings of the National Academy of Sciences USA, **104(25):** 10376–10381

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbacj, M.A., Muller, R., Wohlleben, W., Breitling, R., Takano, E., & Medema, M.H. (2015). antiSMASH 3.0 — a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*, doi: 10.1093/nar/gkv437.

Yadav, G., Gokhale, R. S. & Mohanty, D. (2003). Computational Approach for Prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. *Journal of Molecular Biology,* **328**: 335–363.

Zhu, P. & Craighead, H.G. (2012). Zero-Mode Waveguides for Single-Molecule Analysis. *Annual Review of Biophysics*, **41:** 269–93

Ziemert, N., Podell, S., Penn, K., Badger, J.H., Allen, E. & Jensen, P.R. (2012). The Natural Product Domain Seeker NaPDoS: A Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. *PLoS One*, **7(3):** 34064-34068.

Zuckerkandl, E. & Pauling L. (**1965**). Evolutionary divergence and convergence in proteins. Edited in *Evolving Genes and Proteins* by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York.

CHAPTER 4

DEVELOPMENT OF A TRANSFORMATION PROTOCOL FOR *STREPTOMYCES POLYANTIBIOTICUS* SPR^{T} AND GENE DISRUPTION EXPERIMENTS

CHAPTER 4

DEVELOPMENT OF A TRANSFORMATION PROTOCOL FOR *STREPTOMYCES POLYANTIBIOTICUS* SPR^T AND GENE DISRUPTION EXPERIMENTS

CHAPTER 4

DEVELOPMENT OF A TRANSFORMATION PROTOCOL FOR *STREPTOMYCES POLYANTIBIOTICUS* SPR^T AND GENE DISRUPTION EXPERIMENTS

4.1 ABSTRACT

In order to identify the genes involved in DPO biosynthesis, a DNA transformation protocol was required for the introduction of plasmid DNA into *S. polyantibioticus* SPR^T, so that specific genes could be disrupted to establish whether their products are involved in the biosynthesis of DPO. Although there are several published transformation protocols for the introduction of DNA into streptomycetes, there is no method that is generally applicable to all species. Due to the fact that *S. polyantibioticus* SPR^T is a novel antibiotic-producing actinobacterium, it was necessary to develop a unique DNA transformation method.

In this study, the classical methods of electroporation and protoplast transformation of plasmid DNA proved unsuccessful in the transformation of *S. polyantibioticus* SPR^T. However, an optimised method of intergeneric conjugation using the methylation-deficient *E. coli* ET12567/pUZ8002 strain allowed the transfer plasmid DNA into *S. polyantibioticus* SPR^T with a conjugation frequency of 8.3 x 10^{-6} exconjugants per 1 x 10^7 donor *E. coli* cells, thereby providing a platform for the gene disruption experiments.

Subsequently, the genes identified from the *S. polyantibioticus* SPR^T genome sequence (Chapter 3) as being putatively involved in the biosynthesis of DPO were individually cloned into the suicide vector, pOJ260, transformed into *S. polyantibioticus* SPR^T by means of mating with *E. coli* ET12567/pUZ8002 and insertionally activated via homologous recombination in the recipient strain. The genes that were disrupted were: the A domain encoded by gene

SPR_53060, the putative Cy domain encoded by gene SPR_53040, the acyl-coA synthetase encoded by gene SPR_52860, the cinnamate-CoA ligase encoded by gene SPR_60150, the putative D-lactate dehydrogenase encoded by gene SPR_60250, the thioesterase encoded by gene SPR_53090 and a selection of the A domains identified in Chapter 2.

The method for isolating DPO was carried out on each of the mutant strains, *S. polyantibioticus* ΔAD2, *S. polyantibioticus* ΔA18, *S. polyantibioticus* ΔA16, *S. polyantibioticus* ΔA28, *S. polyantibioticus* ΔA7, *S. polyantibioticus* ΔA99, *S. polyantibioticus* ΔPAAK, *S. polyantibioticus* ΔCYC, *S. polyantibioticus* ΔLAC, *S. polyantibioticus* ΔACY, *S. polyantibioticus* ΔCIN and *S. polyantibioticus* ΔTHI, and the extracts were assayed for activity against *M. aurum* A+. The absence of activity against *M. aurum* A+ in the extracts from strains *S. polyantibioticus* ΔA99, *S. polyantibioticus* ΔCYC and *S. polyantibioticus* ΔACY suggested the involvement of these genes in the biosynthesis of DPO.

4.2 INTRODUCTION

Streptomycetes are well known for their ability to produce commercially useful secondary metabolites with a variety of biological activities (Hopwood, 1989). In order to discover novel biochemical pathways used in the production of these important compounds, or to enhance or modify their production, it is necessary to manipulate the producing strains by genetic engineering techniques. This requires transformation of exogenous DNA into the producing strains and subsequent gene disruption experiments (Mazy-Servais *et al.,* 1997).

Due to the fact that streptomycetes are not naturally competent for the uptake of exogenous plasmid DNA and because a general system for competence induced by cold shock and calcium treatment, such as that developed for *E. coli,* has not yet been developed for *Streptomyces*, it has hampered progress in *Streptomyces* strain manipulation (Marcone *et al.,* 2010; Kieser *et al.,* 2000). The work by Okanishi *et al.* (1974) on the preparation, regeneration and transfection of protoplasts in *Streptomyces* spp. led to the discovery that plasmid DNA could be transformed into protoplasts at a very high frequency in the presence of polyethylene glycol (PEG; Bibb *et al.,* 1978). Even though this method, which was subsequently modified and adapted, is still widely used in the transformation of *Streptomyces* species, it became apparent after limited success in species such as *Streptomyces fradiae* (Matshushima and Baltz, 1985), that no method exists which is equally efficient for all *Streptomyces* species and strains. Each transformation method therefore requires the optimization of each step for each individual strain (Mazy-Servais et al., 1997).

The reason for the different response to protoplast transformation from taxonomically related species may be linked to the variation in the composition and density of the bacterial cell wall. Indeed, variation in the peptidoglycan structure in response to environmental conditions, aging, and maturation in order to adapt to changing conditions may explain the intraspecific variability in the susceptibility to cell wall digestion (Marcone *et al*., 2010; Vollmer, 2008). Another serious limitation to this method is the occurrence of restriction-modification (R-M) systems, which are widespread in streptomycetes. R-M systems are vital components of prokaryotic defence mechanisms against invading foreign DNA, such as phage genomes, that can decrease transformation frequencies or make transformants undetectable, depending on the origin of the DNA being introduced (Vasu & Nagaraja, 2013; MacNeil, 1988). The R-M systems generally

consist of two different enzymatic activities: a restriction endonuclease (REase) and a methyltransferase (MTase). The REase is able to recognize and cleave foreign DNA sequences at specific sites, whereas the MTase activity ensures differentiation between self and non-self DNA by transfer of methyl groups to specific DNA sequences within the host's genome (Figure 4.1) (Vasu & Nagaraja, 2013). However, methyl-specific restriction systems have also been described (Raleigh & Wilson, 1986; Lacks & Greenberg, 1977), whereby foreign methyl-modified DNA is restricted and the host strain does not modify its own DNA. Indeed, *Streptomyces avermitilis* NRRL 8165 contains a unique restriction system that restricts plasmid DNA containing N^6-methyladenine or 5-methylcytosine. Additionally, *S. coelicolor* A3(2) is known to strongly restrict DNA from modification-proficient *E. coli* strains such as *E. coli* K12, but readily accepts it when it originates from methylation-deficient strains (Kieser *et al.,* 2000; Flett et al., 1997; Kieser & Hopwood, 1991). Consequently, for the purpose of transforming exogenous DNA into *S. coelicolor* A3(2), it is possible to thwart the R-M system by isolating plasmid DNA from the methylation-deficient *E. coli* ET12567 strain (Kieser *et al.*, 2000). Generally, it is also possible to limit the restriction barrier by subjecting protoplasts to heat treatment prior to transformation which serves to inactivate the restriction enzymes (Hussain & Ritchie, 1991; Bailey & Winstanley, 1986).

Even though PEG-mediated transformation of protoplasts has allowed for the rapid development of gene cloning systems in various *Streptomyces* species such as *S. lividans* 66, *S. rimosus* R6 and *S. coelicolor* A3(2) amongst others, transformation of fragile protoplasts is tedious and frequently not reproducible, and therefore many strains remain recalcitrant to transformation (Pigac & Schrempf, 1995).

An alternative method to the use of chemicals in assisting the uptake of DNA by bacterial cells is electroporation, which involves the application of a brief, high voltage electrical pulse to a suspension of cells and DNA that results in the formation of transient membrane pores and subsequent uptake of DNA by the cells (Pigac & Schrempf, 1995; Shigekawa & Dower, 1988). Several *Streptomyces* species, such as *Streptomyces parvulus* IMET 41380 (Mazy-Servais *et al.,* 1997), *Streptomyces vinaceus* NCIB 8852 (Macy-Servais *et al.,* 1997), *S. lividans* ATCC 1326 (Tyurin *et al.,* 1995) and *S. virginiae* ATCC 13161 (Tyurin *et al.,* 1995), have been successfully transformed using electroporation. Due to the fact that electroporation avoids the need to optimise conditions for protoplast preparation and regeneration, it is less tedious and

time consuming than protoplast transformation and has been reported to be particularly useful in transforming previously "untransformable" strains (Pigac & Schempf, 1995). However, conditions for electroporation have also been described as being strain-specific (Kieser *et al.,* 2000).

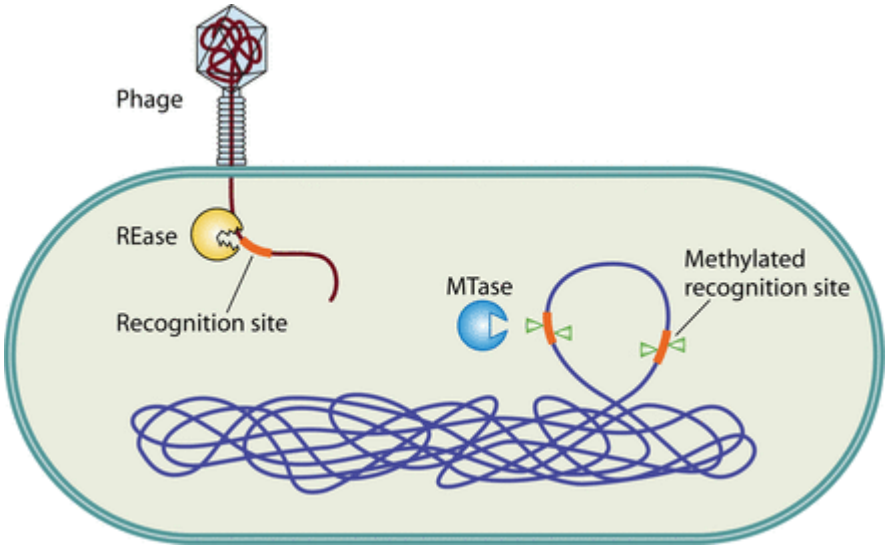

Figure 4.1 Illustration depicting R-M systems in their role as a defence mechanism in prokaryotes. R-M systems are able to identify the methylation status of invading foreign DNA, as methylated sequences are recognised as "self", while any recognition sequences that lack methylation are identified as "non-self" and cleaved by a REase. This is due to the fact that "self" DNA is methylated at its appropriate recognition sites by a cognate MTase belonging to the specific R-M system (Vasu & Nagaraja, 2013).

In recent years, there has been considerable interest in the use of intergeneric conjugation, which evades restriction barriers, as a means of transferring plasmid DNA into actinomycetes. This method allows for the construction and manipulation of recombinant plasmids in a host such as *E. coli*, which is used to transfer the genetic material into the required recipient via horizontal gene transmission (Giebelhaus *et al.,* 1996). Conjugal DNA transfer is a highly conserved mechanism that is generally mediated by plasmids which encode many critical transfer-related functions (Mazodier & Davies, 1991). The first plasmid-mediated transfer to

be described involved the *E. coli* F plasmid, which has served as the model for nearly all subsequently described conjugative plasmids from both Gram-negative and Gram-positive bacteria (Lederberg and Tatum, 1946). Briefly, the plasmid DNA molecule is endonucleolytically cleaved on a single strand by a plasmid-encoded relaxase protein, which leads to the formation of a DNA-protein complex known as a relaxosome (Lanka & Wilkins, 1995). The cleaved strand is then transferred from the donor to the recipient cell via a bridge-like structure known as a pilus (which connects the two cells). After conjugation, the donor and recipient both contain the plasmid, which allows the recipient to serve as a donor in future matings (Willetts & Wilkins, 1984).

In contrast to this conserved mechanism of conjugation witnessed throughout most of the bacterial domain, conjugative plasmids of *Streptomyces* origin employ a different mode of DNA transfer (Grohmann *et al.,* 2003). The conjugative elements and transfer functions present in *Streptomyces* plasmids are unique and encode fewer transfer functions compared to plasmids from other genera. Indeed, in comparison to the approximately 25 functions required by the *E. coli* F plasmid to mobilize DNA, the *S. lividans* plasmid pIJ101 only requires a 70 kDa membrane-associated product of the *tra* gene and the *cis*-acting locus of transfer (*clt*) to mediate plasmid transfer (Grohmann *et al.,* 2003). Moreover, the major transfer proteins of conjugative *Streptomyces* plasmids are homologous to proteins which mobilize double-stranded DNA in processes such as sporulation and cell division, instead of sharing similarity with relaxases (Begg *et al*., 1995; Wu *et al*., 1995; Hopwood & Kieser, 1993). Additionally, the conjugative transfer of *Streptomyces* plasmids has been shown to occur via a double-stranded intermediate. Due to the fact that the *Streptomyces* conjugative system is unique, there has been considerable interest in elucidating the manner in which streptomycetes mediate plasmid transfer via conjugation (Grohmann *et al.,* 2003; Possoz *et al.,* 2001).

It was initially assumed that the transfer of plasmids by conjugation was confined to closely related bacterial species or genera until Trieu-Cuot *et al.* (1987) dispelled this misconception by demonstrating conjugation between Gram-negative *E. coli* and various Gram-positive bacteria, which included *Streptococcus lactis*, *Bacillus thuringiensis*, *S. aureus* and *Enterococcus faecalis*, amongst others. The shuttle plasmid used in the study, pAT187, contained the origin of replication for *E. coli* and a broad-host range origin of replication for Gram-positive bacteria. The transfer of the plasmid was contingent on the origin of transfer

(*oriT*) of the incompatability group P (IncP) broad-host range *E. coli* plasmid RK2 acting in *cis* and the transfer gene (*tra*) functions of the IncPα *E. coli* plasmid RP4 supplied *in trans* (Gormley & Davies, 1991; Mazodier *et al.,* 1989). Since then, additional studies with similarly constructed shuttle plasmids have shown that conjugative transfer between phylogenetically unrelated microorganisms is possible (Gormley & Davies, 1991).

The first protocol for the intergeneric transfer of plasmids between *E. coli* and *Streptomyces* spp. was reported by Mazodier *et al.* (1989). The success of this transfer was dependent on the presence of a 760 bp, *cis*-acting, *oriT*-containing fragment from the plasmid RK2 and on the conjugative functions, such as the *Tra2* core of plasmid RP4, which encodes the DNA transport apparatus important for pilus formation, supplied *in trans* by the *E. coli* donor strain. The same method has been used successfully with several *Streptomyces* species such as *S. fradiae*, *S. ambofaciens*, *S. lividans* and *S. coelicolor* (Bierman *et al.,* 1992) and other actinomycetes such as *Amycolatopsis* (Stegmann *et al.,* 2001), *Actinoplanes* (Heinzelmann *et al.,* 2003) and *Saccharopolyspora* (Matshushima *et al.,* 1994). Importantly, the conjugative functions of plasmid RP4 and the RK2 derivative, pUZ8002, were used to mobilize the resident plasmids in these studies (Paranthaman & Dharmalingam, 2003). It was also demonstrated that the major biochemical events during the intergeneric conjugal transfer occurred at the *oriT*, which included the formation of a relaxosome, nicking of the closed circular dsDNA molecule and transfer of the ssDNA intermediate from donor to recipient (Grohmann *et al.,* 2003; Frost *et al.,* 1994).

Numerous cloning vectors for the conjugative transfer from *E. coli* to *Streptomyces* spp. have been constructed, most notably by Bierman *et al.* (1992), who reported the construction of plasmid and cosmid vectors that function in *Streptomyces* spp. due to features that allow for: a) integration via homologous recombination between the cloned DNA fragment within the vector and the *Streptomyces* chromosome (e.g. pOJ260 and pKC1132), b) autonomous replication (e.g. pOJ446 and pKC1139) and c) site-specific integration via the bacteriophage ΦC31 attachment site (e.g. pSET152 and pKC1163) (Kieser *et al*., 2000). All of these vectors contain the 760 bp *oriT* fragment from plasmid RK2 and *E. coli* replication functions from pUC, P15A or P1, which allow the plasmids to serve as cloning vectors for construction to occur in an *E. coli* host. The recombinant plasmids can then be transferred to the desired *Streptomyces* strain. Plasmids such as pOJ260 and pKC1132 were designed to be unable to

replicate in *Streptomyces* spp. and are therefore most useful for gene disruption experiments, whereas plasmids such as pOJ446 and the recently designed pJN100 (Nikodinovic & Priestley, 2006), contain replication functions for *Streptomyces* spp. and therefore can exist as autonomous, multicopy plasmids that are useful for complementation studies (Bierman *et al*., 1992).

The use of intergeneric conjugation as a means of transforming DNA into streptomycetes has been extensively utilized due to the simplicity of performing the method in comparison to developing separate procedures for protoplast preparation and regeneration. Additionally, restriction barriers such as R-M systems are by-passed or have a drastically reduced effect when ssDNA is transferred into the recipient (Flett *et al.,* 1997; Matshushima et al., 1994). This is due to the fact that the majority of REases require dsDNA for the recognition and cleavage of foreign DNA (Vasu & Nagaraja, 2013). In addition, the use of *E. coli* shuttle vectors that contain *oriT* allow site-specific or insert-directed chromosomal integration that is extremely useful for targeted gene disruption (Kieser *et al.,* 2000).

In summary, electroporation and protoplast transformation methods are widely used for introducing plasmid DNA into streptomycetes, but due to their low efficiency, intergeneric conjugation has provided an alternative means. It has been reported that optimal conditions for different strains may vary and therefore a defined procedure for each strain must be established to enable a high conjugation efficiency. Furthermore, conjugation with streptomycete mycelial fragments has proven to provide higher numbers of recombinants than conjugation performed with freshly germinated spores (Kieser *et al.,* 2000; Matshushima et al., 1994).

The development of an efficient transformation method for the introduction of plasmid DNA into a specific strain of *Streptomyces* allows for the targeted manipulation of biosynthetic pathways in order to characterize individual genes and their functions. This is performed experimentally via the disruption of target genes using the method of homologous recombination. In this method, a DNA fragment internal to the target gene is inserted into a plasmid vector, which integrates into the chromosome of the recipient *Streptomyces* strain by a single or double homologous crossover, depending on whether a selectable marker is contained within the DNA fragment carrying all or part of the target gene. A single crossover

results in an integrated copy of the plasmid vector, which is flanked on either side by two mutant alleles of the target gene i.e. one truncated at the 5' end and the other truncated at the 3' end. A double crossover results in a mutant allele that replaces the chromosomal copy of the target gene via two crossovers that cannot revert and is therefore more stable than a single crossover mutant (Keiser *et al.,* 2000).

4.3 MATERIALS AND METHODS

4.3.1 BACTERIAL STRAINS AND PLASMIDS

All *Streptomyces* and *E. coli* strains and plasmids that were used and constructed in this study are listed in Table 4.1. The common conjugative donor strain *E. coli* S17-1, which carries an integrated derivative of IncP-group plasmid RP4 encoding all of the required plasmid mobilization functions, was used as a donor strain for intergeneric conjugation. *E. coli* DH5α was used as a general host for all standard cloning procedures. *E. coli* JM109 was used as a recipient strain in conjugation with the methylation deficient *E. coli* ET12567 (provided by Dr P. Whitney Swain III, Promiliad Biopharma Inc., USA). *E. coli* ET12567 contains the 'driver' plasmid pUZ8002 (derived from RK2) and was also used as the donor strain for intergeneric conjugation. Plasmid pUZ8002 carries the plasmid RK2 transfer functions necessary for intergeneric conjugation. Plasmid pJN100 (Figure 4.2) (provided by Dr. P. Whitney Swain III, Promiliad Biopharma Inc., USA) (GenBank accession number: DQ309424) (Nikodinovic & Priestley, 2006) is a high copy number *E. coli*-*Streptomyces* shuttle vector derived from plasmid RK2 and was used for the optimization of the intergeneric conjugation method between the different *E. coli* donor strains and *S. polyantibioticus* SPR^T, as well as the complementation experiments in *S. polyantibioticus* SPR^T.

Plasmid pOJ260 (Figure 4.3) (provided by Dr Bohdan Ostash, Department of Genetics and Biotechnology, Ivan Franko National University of L'viv, L'viv, Ukraine) (GenBank accession number: GU270843) is a non-replicating, integrative *E. coli*-*Streptomyces* shuttle vector derived from the conjugative broad host range plasmid RK2, which was used to perform gene disruption experiments by chromosomal integration via single crossover homologous recombination in *S. polyantibioticus* SPR^T. Plasmid pOJ260 does not replicate in *Streptomyces* and can therefore be used as a suicide vector for gene disruption and replacement (Kieser *et al.,* 2000).

Plasmid pJNHYG is a derivative of pJN100 carrying a hygromycin resistance cassette. This derivative was created by first performing a restriction endonuclease digestion of pORI101 (Figure 4.4) (GenBank accession number: EF216315) (Bourn *et al.,* 2007) using *Not*I to release a 2.44 kb fragment containing the entire hygromycin resistance cassette. Concomitantly, the plasmid pJN100 was subjected to a restriction endonuclease digestion with *Not*I. The 2.44 kb

hygromycin resistance cassette was ligated into the *Not*I site of the restriction endonuclease digested pJN100 using the Rapid DNA Ligation Kit (Thermo Scientific) (14 h at 4 °C) according to the manufacturer's instructions to generate the 8.9 kb pJNHYG plasmid. The T4 DNA ligase was inactivated at 80 °C for 20 min before transformation into *E. coli* DH5α. The plasmid was isolated using a NucleoSpin Plasmid Isolation Kit (Machery Nagel, Germany) according to the manufacturer's instructions.

All of the other plasmids utilized in this study are derivatives of pOJ260 and pJN100 carrying the PCR-amplified target genes used in the gene disruption experiments (section 4.3.7). *S. polyantibioticus* SPR^T was isolated by Dr Paul Meyers. *S. coelicolor* A3(2) (NRRL B-16638) was obtained from the culture collection of the Agricultural Research Service, United States Department of Agriculture (NRRL) (Peoria, Illinois, USA).

4.3.2 MEDIA AND CULTURE CONDITIONS

All *Streptomyces* strains used in this study were grown on ISP4 agar (10 g soluble starch, 1 g K_2HPO_4, 1 g $MgSO_4$, 2 g $(NH4)_2SO4$, 1 g $CaCO_3$, 1 mg $FeSO_4$, 1 mg $MgCl_2$, 1 mg $ZnSO_4$, 18 g bacteriological agar per litre of dH_20, pH 7.2), (Shirling & Gottlieb, 1966) at 30 °C for 7 days in order for sporulation to occur, after which the spores were scraped off the surface of the agar using a sterile inoculating loop and stored in 20 % glycerol at -20 °C until needed. Liquid SMC medium (10 g glucose, 4 g yeast extract, 4 g peptone, 4 g K_2HPO_4, 2 g KH_2PO_4, 0.5 g $MgSO_4$ per litre of dH_20, pH 7.0) (Zhang *et al*., 1992) containing 10.3 % sucrose, ISP2 (YEME) medium (10 g malt extract, 4 g glucose, 4 g yeast extract per litre of dH_20, pH 7.3) (Shirling & Gottlieb, 1966), Luria-Bertani broth (LB) (10 g tryptone, 5 g yeast extract, 5 g NaCl, per litre of dH_20) (Sambrook *et al*., 1989) and V medium (2.4 g soluble starch, 10 g glucose, 3 g meat extract, 5 g yeast extract, 5 g tryptose per litre of dH_20, pH 7.2) (Marcone *et al*., 2010) were used to culture *S. polyantibioticus* SPR^T for the collection of mycelium for intergeneric conjugation experiments. CRM (10 g glucose, 100.3 g sucrose, 10 g $MgCl_2.6H_20$, 15 g tryptic soy broth (TSB), 5 g yeast extract per litre of dH_20, pH 7.2), tryptic soy broth (TSB) supplemented with 0.24 mM threonine and 1 % glycine and YEME supplemented with 34 % sucrose and 0.5 % glycine were used to culture *S. polyantibioticus* SPR^T for electroporation of mycelia. Solid agar media used for the cultivation of *S. polyantibioticus* SPR^T and *S. coelicolor* A3(2) exconjugants were mannitol soya (MS) medium (20 g mannitol, 20 g soya flour, 20 g

bacteriological agar per litre of tap water, pH 7.0), Middlebrook 7H9, YEME and ISP4 all containing 10 or 20 mM $MgCl_2$. Electroporated mycelia were plated on TSB agar or R2YE agar (103 g sucrose, 0.25 g K_2SO_4, 10.12 g $MgCl_2.6H_2O$, 10 g glucose, 0.1 g casamino acids, KH_2PO_4, 10 ml 0.5 % KH_2PO_4, 80 ml 3.68 % $CaCl_2.2H_2O$, 15 ml 20 % L-proline, 100 ml 5.73 % TES buffer, 2 ml trace elements solution (40 mg $ZnCl_2$, 200 mg $FeCl_3.6H_2O$, 10 mg $CuCl_2.2H_2O$, 10 mg $MnCl_2.4H_2O$, 10 mg $Na_2B_4O_7.10H_2O$, 10 mg $(NH)_4Mo_7O_{24}.4H_2O$ per litre of dH_2O), 0.5 ml 1 N NaOH, 40 ml 10 % yeast extract per litre of dH_20, pH 7.2). Finally, hypertonic V medium was used to cultivate *S. polyantibioticus* SPR^T for preparing protoplasts, while R2YE or VMSO.1 (2.4 g soluble starch, 0.1 g glucose, 0.3 g meat extract, 0.5 g yeast extract, 0.5 g tryptose, 3.5 g L-proline, 4 g bacteriological agar per litre of dH_20, pH 7.2) media were used for the regeneration of protoplasts. LB broth supplemented with antibiotics (50 µg/ml apramycin, 34 µg/ml chloramphenicol and 50 µg/ml kanamycin, as required) was used for the cultivation of all *E. coli* strains.

Table 4.1 Strains and plasmids used in this study

Strain/plasmid	Genotype/characteristics	Reference
Strains		
S. polyantibioticus		
SPR^T	Wild type	This study
ΔAD2	SPR^T derivative carrying an integrated copy of pOJAD2; apr^R	This study
ΔA16	SPR^T derivative carrying an integrated copy of pOJA16; apr^R	This study
ΔA18	SPR^T derivative carrying an integrated copy of pOJA18; apr^R	This study
ΔA28	SPR^T derivative carrying an integrated copy of pOJA28; apr^R	This study
ΔA7	SPR^T derivative carrying an integrated copy of pOJA7; apr^R	This study
ΔPAAK	SPR^T derivative carrying an integrated copy of pOJPAAK; apr^R	This study
ΔA99	SPR^T derivative carrying an integrated copy of pOJA99; apr^R	This study
ΔCYC	SPR^T derivative carrying an integrated copy of pOJCYC; apr^R	This study
ΔACY	SPR^T derivative carrying an integrated copy of pOJACY; apr^R	This study
ΔLAC	SPR^T derivative carrying an integrated copy of pOJLAC; apr^R	This study
ΔCIN	SPR^T derivative carrying an integrated copy of pOJCIN; apr^R	This study
ΔTHI	SPR^T derivative carrying an integrated copy of pOJTHI; apr^R	This study
PJNACY	ΔACY derivative complemented with pJNACY; apr^R	This study

S.coelicolor		
A3(2)	Wild type	NRRL
E. coli		
ET12567	F$^-$ *dam*13*::Tn*9 *dcm*6 *hsd*M *hsd*R *zjj*202*::Tn*10 *rec*F143 *gal*K2 *gal*T22 *ara*14 *lac*Y1 *xyl*5 *leu*B6 *thi*1 *ton*A31 *rps*L136 *his*G4 *tsx*78 *mtl*1 *gln*V44	MacNeil *et al.* (1992)
DH5α	F$^-$ *deo*R *end*A1 *rec*A1 *rel*A1 *gyr*A96 *hsd*R17(rk-, mk+) *sup*E44 thi-1 *pho*A Δ(*lac*ZYA-argF)U169 Φ80*lac*ZΔM15 λ-	Bioline (UK)
S17-1	*recA pro hsdR RP4-2-Tc::Mu-Km::Tn7*	Simon *et al.* (1983)
dam$^-$ dcm$^-$	F$^-$ *dam*-13:*Tn*9(Cam^R) *dcm*-6 *ara*-14 *his*G4 *leu*B6 *thi*-1 *lac*Y1 *gal*K2 *gal*T22 *gln*V44 *hsd*R2 *xyl*A5 *mtl*-1 *rps*L 136(Str^R) *rtb*D1 *ton*A31 *tsx*78 *mcr*A *mcr*B1	Bioline (UK)
JM109	*end*A1, *rec*A1, *gyr*A96, *thi, hsd*R17 (r_k^-, m_k^+), *rel*A1, *sup*E44, Δ(*lac-pro*AB)	Yanisch-Perron *et al.* (1985)
Plasmids		
pGEM®-T Easy	3015 bp vector with 3' - T overhangs, designed for cloning PCR products; Amp^R *oriT lacZ*	Promega (USA)
pGEMAD-2	pGEM®-T Easy vector harbouring the AD2 PCR product	This study
pGEMA-7	pGEM®-T Easy vector harbouring the A7 PCR product	This study
pGEMA-18	pGEM®-T Easy vector harbouring the A18 PCR product	This study
pGEMA-16	pGEM®-T Easy vector harbouring the A16 PCR product	This study
pGEMA-28	pGEM®-T Easy vector harbouring the A28 PCR product	This study
pGEM-PAAK	pGEM®-T Easy vector harbouring the *paaK* PCR product	This study
pGEMA-99	pGEM®-T Easy vector harbouring the A99 PCR product	This study
pGEM-CYC	pGEM®-T Easy vector harbouring the CYC PCR product	This study
pGEM-ACY	pGEM®-T Easy vector harbouring the ACY PCR product	This study
pGEM-LAC	pGEM®-T Easy vector harbouring the LAC PCR product	This study
pGEM-CIN	pGEM®-T Easy vector harbouring the CIN PCR product	This study
pGEM-THI	pGEM®-T Easy vector harbouring the THI PCR product	This study
pJN100	6460 bp *E. coli-Streptomyces* shuttle vector derived from RK2; apr^R oriT *snp*A *rep^{ColE1}* rep^{pIJ101}	Nikodinovic & Priestley, (2006)
pOJ260	3469 bp non-replicating integrative *E. coli-Streptomyces* shuttle vector derived from pKC787; apr^R rep^{puc} oriT *lacZ*	Bierman *et al.* (1992)

pOJAD2	720 bp *Pst*I/*Sac*II fragment of pGEMAD-2 inserted into the *Pst*I/*Sac*II sites of pOJ260	This study
pOJA18	670 bp *Pst*I/*Sac*II fragment of pGEMA-18 inserted into the *Pst*I/*Sac*II sites of pOJ260	This study
pOJA16	700 bp *Pst*I/*Sac*II fragment of pGEMA-16 inserted into the *Pst*I/*Sac*II sites of pOJ260	This study
pOJA28	708 bp *Pst*I/*Sac*II fragment of pGEMA-28 inserted into the *Pst*I/*Sac*II sites of pOJ260	This study
pOJA7	717 bp *Pst*I/*Sac*II fragment of pGEMA-7 inserted into the *Pst*I/*Sac*II sites of pOJ260	This study
pOJPAAK	700 bp *Pst*I/*Sac*II fragment of pGEM-PAAK inserted into the *Pst*I/*Sac*II sites of pOJ260	This study
pOJA99	719 bp *Eco*RI fragment of pGEMA-99 inserted into the *Eco*RI site of pOJ260	This study
pOJCYC	980 *Eco*RI fragment of pGEM-CYC inserted into the *Eco*RI site of pOJ260	This study
pOJACY	921 bp *Eco*RI fragment of pGEM-ACY inserted into the *Eco*RI site of pOJ260	This study
pOJLAC	608 bp *Pst*I/*Sac*II fragment of pGEM-LAC inserted into the *Pst*I/*Sac*II sites of pOJ260	This study
pOJCIN	649 bp *Pst*I/*Sac*II fragment of pGEM-CIN inserted into the *Pst*I/*Sac*II sites of pOJ260	This study
pOJTHI	473 bp *Pst*I/*Sac*II fragment of pGEM-THI inserted into the *Pst*I/*Sac*II sites of pOJ260	This study
pJNACY	1557 bp *Eco*RI fragment of pGEM-ACY inserted into the *Eco*RI site of pOJ260	This study
pUZ8002	Non-transmissable *oriT*-mobilizing RK2 derivative; *dam dcm hsd*S Str^{r} Tel^{r} Clm^{r} Km^{r}	Paget *et al.* (1999)
pJNHYG	Derivative of pJN100 carrying a 2.44 kb hygromycin resistance casette; apr^{R} hyg^{R} oriT *snp*A *rep*^{ColE1} rep^{pIJ101}	This study
pORI101	5316 bp *Mycobacterium-E. coli* shuttle vector derived from pAL5000; hyg^{R} Ori2 OriM rep^{pMB1} rep^{pAL5000} *lacZ*	Bourn *et al.* (2007)

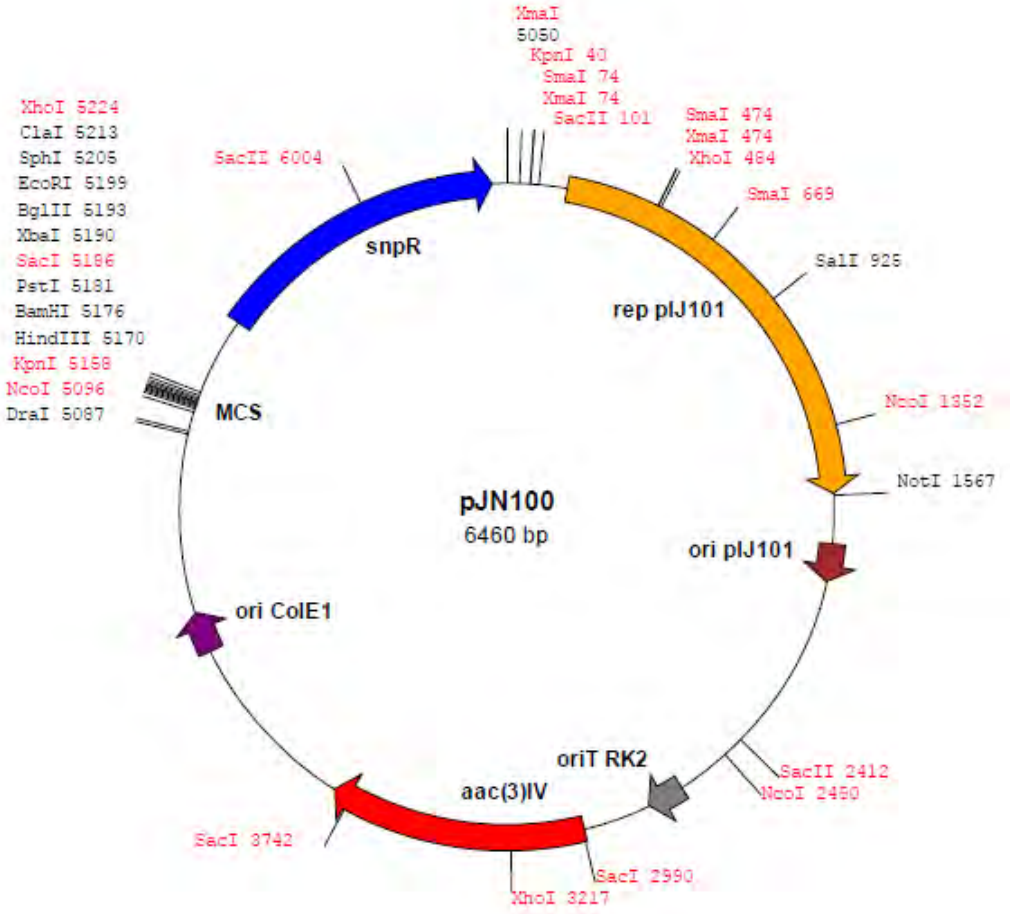

Figure 4.2 Plasmid map of the *E. coli-Streptomyces* shuttle expression vector, pJN100, indicating the position of selected restriction endonuclease sites (multiple sites for the same enzyme are shown in red).. The red arrow represents an apramycin resistance cassette, a selectable marker in both hosts. The grey arrow represents an $oriT_{RK2}$ insertion that allows intergeneric conjugation between *E. coli* and *Streptomyces*. The orange arrow represents the origin of replication of plasmid pIJ101, which allows for replication to take place in the *Streptomyces* host, while the purple arrow represents the origin of replication of plasmid ColE1, which allows for replication to take place in the *E. coli* host. The blue arrow indicates the presence of the *snpR* gene, the product of which activates the *snpA* promoter, for use in protein expression studies. The position of a multiple cloning region (showing only the restriction sites relevant to this study) is indicated as MCS (Nikodinovic & Priestley, 2006).

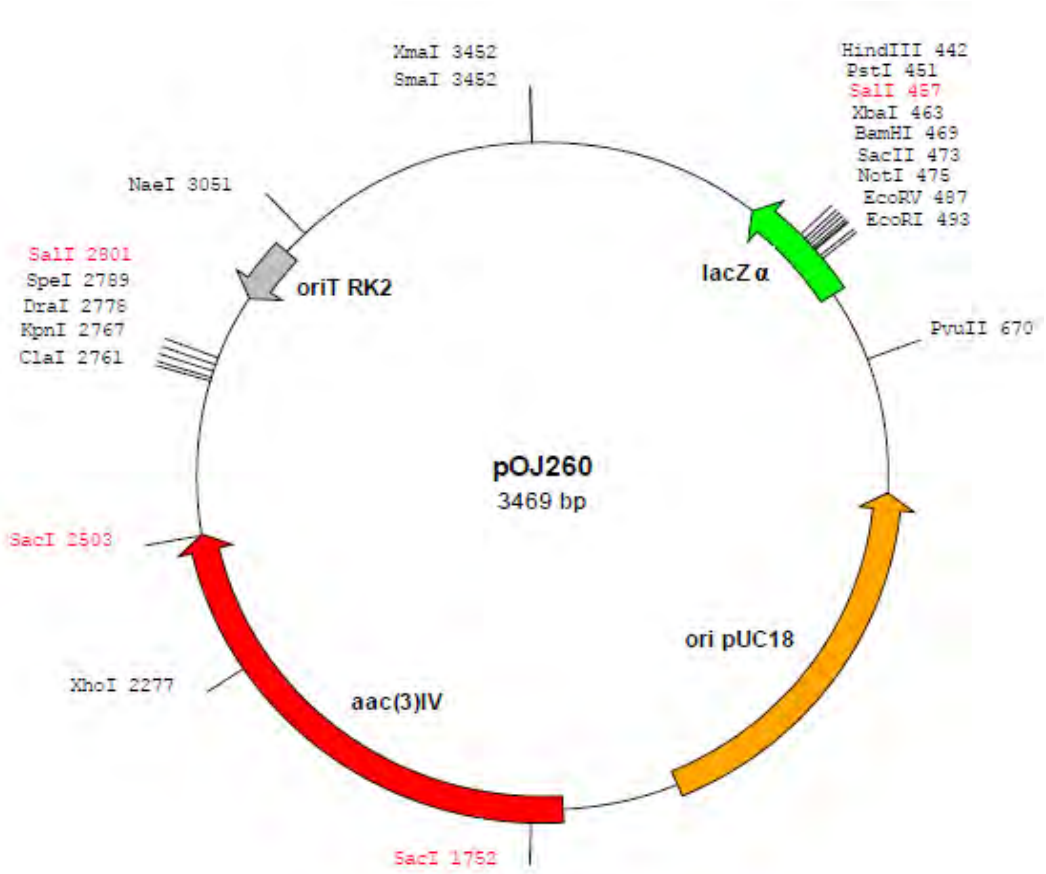

Figure 4.3 Plasmid map of the non-replicating integrative *E. coli-Streptomyces* shuttle vector, pOJ260, indicating the position of selected restriction endonuclease sites (multiple sites for the same enzyme represented by red text). The red arrow represents an apramycin resistance cassette, a selectable marker in both hosts. The grey arrow represents an *oriT*$_{RK2}$ insertion that allows intergeneric conjugation between *E. coli* and *Streptomyces*, while the orange arrow represents the origin of replication of plasmid pUC18, which allows for replication to take place in the *E. coli* host. The green arrow represents the α-peptide coding region of the *E. coli lacZ* gene, encoding the enzyme β-galactosidase, which also includes restriction sites conveniently arranged as a multiple cloning region (Bierman *et al.,* 1992).

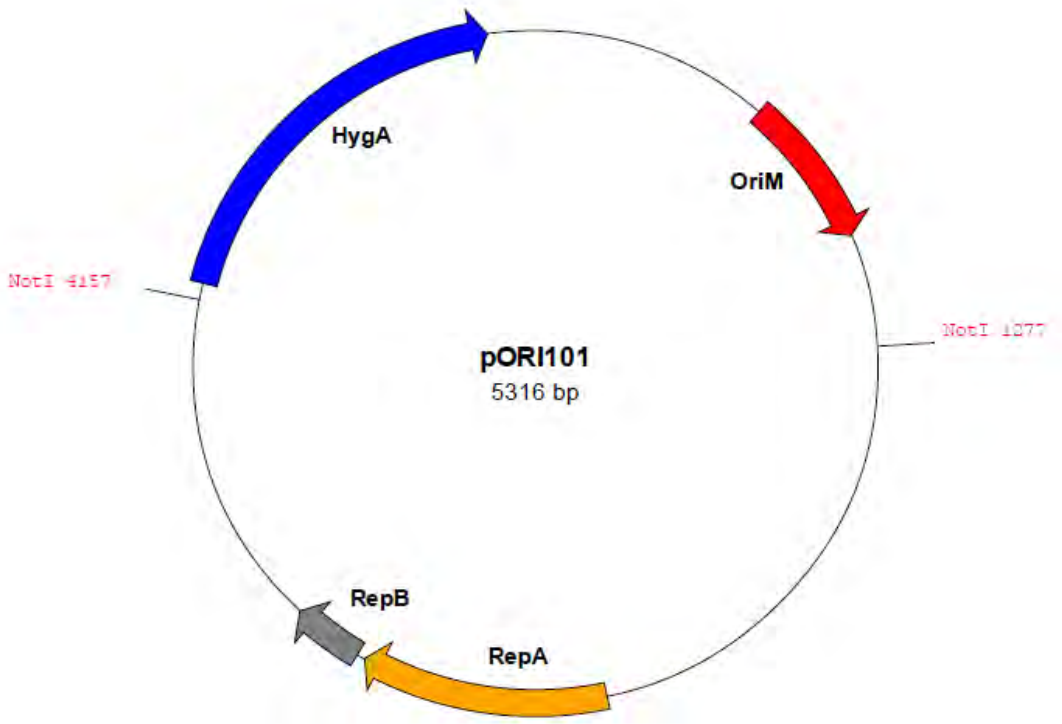

Figure 4.4 Plasmid map of the mycobacterial shuttle vector, pORI101. The *Not*I restriction sites used in a restriction endonuclease digestion to generate a 2.44 kb fragment containing the hygromycin resistance gene, *hygA*, are depicted in red text. The mycobacterial replicon genes are represented by RepA and RepB. The mycobacterial origin of replication is represented by OriM (Bourn *et al.,* 2007).

4.3.3 PRIMER DESIGN

PCR primers were designed for the amplification of the following genes based on the *S. polyantibioticus* SPR^T genome sequence (Chapter 3): the putative Cy domain (CYC) encoded by gene SPR_53040, the A domain (AD99) encoded by gene SPR_53060, the acyl-CoA synthetase (ACY) encoded by gene SPR_52860, the cinnamate CoA ligase (CIN) encoded by

gene SPR_60150, the putative D-lactate dehydrogenase (LAC) encoded by gene SPR_60250 and the thioesterase (THI) encoded by gene SPR_53090. The primer sequences and their respective amplification product sizes are shown in Table 4.2.

Additionally, the primer set, ACYORF_F/ACYORF_R, was designed to amplify the full open reading frame of the acyl-CoA synthetase (gene SPR_52860) and incorporated the *Cla*I restriction enzyme recognition sequence at the 5' end of the ACYORF_F primer sequence and the *Hind*III restriction enzyme recognition sequence at the 5' end of the ACYORF_R primer sequence, in order to achieve the in-frame cloning of this gene into plasmid pJN100.

Table 4.2 Oligonucleotide primers designed in this study

Primer Name	Primer Sequence (5'→3')	Expected product size for primer pair
ADFWD	GTACACCTCGGGATCCACC	719 bp
ADREV	TCGCCCGTGCGGTACATGC	
CYCFWD	CGTGGTCGACGGAGCTCC	980 bp
CYCREV	GGTGTAGAAGCCGAGATCG	
THIF	TATGCGTGCGTTCTTCACCG	473 bp
THIR	AAGGTGCATTCGAGATAGGG	
CINF	CCCAGTACGACCTCTCCTCC	649 bp
CINR	CGGCGGATCTTCTTGTACGG	
LACF	ACTTCGACCTGCTGAGTACG	608 bp
LACR	CAGGAACTCCGCCTTGTACG	
ACYF	CTGTGGGAGCTCATCGACC	921 bp
ACYR	CTCGGTGCTTCCGTAGACC	
ACYORF_F	ATTATAATCGATGCCCTCGACCACTGAATCCG	1557 bp
ACYORF_R	GATAATAAGCTTTCAGGCCGTGTCGTCCCGGTGC	
POJFWD	GCTGCAAGGCGATTAAGTTGG	Variable
POJREV	CAGGCTTTACACTTTATGCTTCC	

"Variable" denotes that the expected product size changes with each different target gene being amplified due to the fact that a different target gene specific primer is used in conjunction with either POJFWD or POJREV (section 4.3.7.1.4)

4.3.4 PCR PROTOCOLS

All PCR amplifications were performed using a Techne TC512 Thermal Cycler fitted with a heated lid and gradient sample block.

4.3.4.1 AMPLIFICATION OF GENES FOR CLONING INTO PLASMID VECTORS

Amplification of the A domain in gene SPR_53060 and the putative Cy domain in gene SPR_53040 were performed using the ADFWD/ADREV and CYCFWD/CYCREV primer sets, respectively. The cycling conditions used for the amplification of both domains were as follows: initial denaturation at 98 ºC for 1 min, followed by 35 cycles of denaturation at 98 ºC for 10 s, annealing at 55 ºC for 30 s and elongation at 72 ºC for 2 min, with a final elongation at 72 ºC for 10 min. PCR reactions consisted of: 100 ng of DNA, 1 U Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, USA), 0.5 µM of each primer, 0.2 mM of each dNTP, 1 x Phusion HF buffer and 4 % (v/v) glycerol in a total volume of 50 µl.

Amplification of sections of the cinnamate CoA ligase (gene SPR_60150), the putative D-lactate dehydrogenase (gene SPR_60250), the thioesterase (gene SPR_53090) and the acyl-CoA synthetase (gene SPR_52860) were performed using the CINF/CINR, LACF/LACR, THIF/THIR, ACYF/ACYR and ACYORF_F/ACYORF_R primer sets, respectively. The cycling conditions used for the amplification of these genes were similar to those used for the amplification with the ADFWD/ADREV and CYCFWD/CYCREV primer sets, except that the annealing temperature was 51 ºC for the THIF/THIR primer set, 53 ºC for the ACYF/ACYR primer set, 56 ºC for the CINF/CINR and LACF/LACR primer sets and 60 ºC for the ACYORF_F/ACYORF_R primer set. PCR reactions were set up as described for the amplification of the A and Cy domains mentioned above, with the exception of the glycerol concentration, which was 2 % (v/v) for amplification using the CINF/CINR, LACF/LACR, THIF/THIR and ACYF/ACYR primer sets and 8 % (v/v) for the ACYORF_F/ACYORF_R primer set.

All PCR amplification products were resolved by electrophoresis alongside a λ-*Pst*I molecular marker on 0.8 % agarose gels containing 0.8 μg/ml ethidium bromide in order to analyse amplicon size and assess primer specificity. The products were visualized using a long wavelength UV light box (Bio-Rad Gel Doc EQ-system™, Bio-Rad Laboratories Inc., USA).

4.3.4.2 COLONY PCR

A colony PCR protocol was used to confirm the presence and correct size of the amplification product in transformants harbouring recombinant constructs. The PCR cycling conditions for the amplification of these inserts was as follows: initial denaturation at 95 ^oC for 5 min, followed by 35 cycles of denaturation at 95 ^oC for 30 s, annealing at 60 ^oC for 90 s and elongation at 72 ^oC for 60 s, with a final elongation at 72 ^oC for 10 min. PCR reactions consisted of: a toothpick-tip size amount of cell mass from a transformant colony, 2 U SuperThem *Taq* polymerase (JMR Holdings, USA), 0.5 μM of each primer, M13F and M13R (Table 2.1), 0.8 mM of each dNTP and 3 mM $MgCl_2$ in a total volume of 20 μl.

All PCR amplification products were resolved by agarose gel electrophoresis alongside a λ-*Pst*I molecular marker on 0.8 % agarose gels, containing 0.8 μg/ml ethidium bromide, in order to analyse amplicon size and primer specificity. The products were visualized using a long wavelength UV light box (Bio-Rad Gel Doc EQ-system™, Bio-Rad Laboratories Inc., USA). The fragments of interest were excised from the gel and purified using the FavorPrep Gel/PCR Purification kit (FavorGen^{TM}, Germany), according to the manufacturer's instructions, if they were to be cloned into pGEM®-T Easy for DNA sequencing (section 2.3.6), or the amplified products were purified using the MSB® Spin PCRapace kit (STRATEC Molecular, Germany) if they were to be sent directly for sequencing as PCR products.

4.3.5 CLONING AND RESTRICTION ENDONUCLEASE DIGESTIONS

For cloning, the fragments of interest were excised from the gel (section 4.3.4.2) and purified using the FavorPrep Gel/PCR Purification kit (FavorGen^{TM}, Germany) according to the manufacturer's instructions. The amplification products obtained from the primer sets ADFWD/ADREV, CYCFWD/CYCREV, THIF/THIR, CINF/CINR, LACF/LACR and ACYF/ACYR were ligated individually into the pGEM®-T plasmid as described in the

pGEM®-T Easy Vector System kit (Promega, USA). The ligation reaction was incubated at 22 °C for 14 h after which 10 ng of the reaction was transformed into *E. coli* α-Select Bronze Efficiency Competent Cells (Bioline, UK) according to the manufacturer's instructions.

The transformants were inoculated onto LB agar (Sambrook *et al.,* 1989) supplemented with 100 µg/ml ampicillin, 40 µg/ml X-gal and 0.2 mM IPTG and incubated at 37 °C for 18 h. Transformants harbouring recombinant pGEM®-T constructs were identified by blue/white selection and white colonies were subcultured onto fresh LB agar plates supplemented with 100 µg/ml ampicillin, 40 µg/ml X-gal and 0.2 mM IPTG and incubated at 37 °C for 18 h to confirm the white phenotype. The presence of the desired inserts was determined by performing colony PCR using the M13F and M13R primer set (Table 2.1) in order to ensure the correct size of the cloned fragment (Section 4.3.4.2).

Transformants identified as harbouring the correct recombinant constructs were inoculated into 5 ml LB broth containing 100 µg/ml ampicillin and cultured overnight with gentle shaking at 37 °C. Thereafter, plasmid DNA was isolated using the NucleoSpin Plasmid Isolation kit (Machery Nagel, Germany) according to the manufacturer's instructions and quantitated spectrophotometrically as described before (section 2.3.2). One microgram of each of the recombinant pGEM®-T vectors, individually carrying the THIF/THIR, LACF/LACR and CINF/CINR amplification products, was subjected to a double restriction endonuclease digestion with *Pst*I and *Sac*II (1.5 U of each enzyme) in the appropriate restriction buffer overnight at 37 °C. The recombinant pGEM®-T vectors, individually carrying the ADFWD/ADREV, CYCFWD/CYCREV and ACYF/ACYR amplification products, were subjected to a single restriction endonuclease digestion with 1.5 U *Eco*RI in the appropriate restriction buffer overnight at 37 °C. Each reaction was subjected to agarose gel electrophoresis as described earlier and the desired inserts obtained from the restriction endonuclease digestions were excised from the gel and purified using the Favorgen Gel/PCR Purification Kit (Favorgen™, Germany). The purified inserts were ligated into the plasmid vector, pOJ260, which had been digested overnight at 37 ºC with 1.5 U of *Eco*RI or a double digestion consisting of *Sac*II and *Pst*I and thereafter dephosphorylated using rAPid Alkaline Phosphatase (Roche, Switzerland) according to the manufacturer's instructions. The alkaline phosphatase incubation period was 24 h at 37 ºC instead of the recommended 1 h.

Subsequently, ligation of each of the digested purified inserts into the dephosphorylated pOJ260 vector was performed according to either the sticky-end or blunt-end protocol for the Rapid DNA Ligation Kit (Thermo Fisher Scientific, USA) for 14 h at 4 ^oC, after which 10 ng of the reaction was transformed into chemically competent *E. coli* α-Select Bronze Efficiency Competent Cells (Bioline, UK) according to the manufacturer's instructions. The transformants were inoculated onto LB agar supplemented with 50 µg/ml apramycin, 40 µg/ml X-gal and 0.2 mM IPTG and incubated at 37 °C for 18 h. Transformants harbouring recombinant pOJ260 constructs were identified by blue/white selection and white colonies were subcultured onto fresh LB plates supplemented with 50 µg/ml apramycin, 40 µg/ml X-gal and 0.2 mM IPTG and incubated at 37 °C for 18 h to confirm the white phenotype.

Transformants with a white phenotype were identified as harbouring the correct recombinant constructs and were inoculated into 5 ml LB broth containing 50 µg/ml apramycin and cultured overnight with shaking at 37 °C. Thereafter, plasmid DNA was isolated using the NucleoSpin Plasmid Isolation kit (Machery Nagel, Germany) according to the manufacturer's instructions and quantitated spectrophotometrically as described before (section 2.3.2). The recombinant constructs carrying the amplification products obtained from the THIF/THIR, CINF/CINR, LACF/LACR, ACYF/ACYR, CYCFWD/CYCREV and ADFWD/ADREV primer sets were designated pOJTHI, pOJCIN, pOJLAC, pOJACY, pOJCYC and pOJA99, respectively (Table 4.1) and sequenced by the dideoxy chain-termination method (Sanger *et al.,* 1977) on an Applied Biosystems Big Dye terminator v3.1 DNA sequencer using BIOLINE Half Dye Mix (Macrogen Inc., South Korea) using the appropriate gene specific primers in order to confirm the presence of the desired inserts.

Additionally, the pGEMA-16, pGEMA-18, pGEMA-28, pGEMA-7 and pGEMAD-2 A domain clones and the pGEM-PAAK clone described in Chapter 2 (section 2.3.8.2) were subjected to restriction endonuclease digestion with 1.5 U *Eco*RI and the appropriate restriction buffer at 37 °C overnight. The DNA samples were subjected to agarose gel electrophoresis, excised from the gel and purified as described above. Each of the inserts was ligated into dephosphorylated pOJ260, which had been subjected to restriction endonuclease digestion with *Eco*RI, and transformed into chemically competent *E. coli* α-Select Bronze Efficiency Competent Cells (Bioline, UK) according to the manufacturer's instructions. Selection of positive transformants carrying the desired insert and isolation of plasmid DNA was performed

as described above and the recombinant constructs carrying the A16, A18, A28, A7, AD2 and *paaK* insert sequences were designated pOJA16, pOJA18, pOJA28, pOJA7, pOJAD2 and pOJPAAK.

Furthermore, the amplification product obtained from the ACYORF_F/ACYORF_R primer set was ligated directly into the pJN100 plasmid, which had been digested with 1.5 U of each of *Cla*I and *Hind*III in the appropriate restriction buffer overnight at 37 °C. The ligation was performed at 4 °C for 18 h and thereafter 10 ng of the ligation product was transformed into chemically competent *E. coli* α-Select Bronze Efficiency Competent Cells (Bioline, UK) according to the manufacturer's instructions. The transformants were inoculated onto LB agar supplemented with 50 µg/ml apramycin, 40 µg/ml X-gal and 0.2 mM IPTG and incubated at 37 °C for 18 h. Transformants harbouring recombinant pJN100 constructs were identified by colony PCR using the ACYF/ACYR primer set (section 4.3.4.2).

4.3.6 DETERMINATION OF THE ANTIBIOTIC SUSCEPTIBILITY OF *S. POLYANTIBIOTICUS* SPR^T

The antibiotic susceptibility of *S. polyantibioticus* SPR^T was determined by inoculation onto YEME agar containing 0, 10, 15, 20, 25, 30, 35, 40, 45 and 50 µg/ml of apramycin, hygromycin B or kanamycin and incubating for 72 h at 30 °C. Subsequently, growth of *S. polyantibioticus* SPR^T was assessed to determine the susceptibility to each antibiotic.

4.3.7 KNOCKOUT CONSTRUCTION BY GENE DISRUPTION EXPERIMENTS

4.3.7.1 CONJUGATION

4.3.7.1.1 TRANSFORMATION OF PLASMID DNA INTO *E. COLI* ET12567/pUZ8002

Chemically competent *E. coli* ET12567/pUZ8002 cells were prepared according to the method described by Dagert and Ehrlich (1979). The plasmids designated pOJA16, pOJA18, pOJA28, pOJA7, pOJAD2, pOJPAAK, pOJTHI, pOJCIN, pOJLAC, pOJACY, pOJCYC, pOJA99 and

pJNACY were transformed into chemically competent *E. coli* ET12567/pUZ8002 cells for subsequent intergeneric conjugation.

4.3.7.1.2 CONJUGATION BETWEEN *E. COLI* ET12567/pUZ8002 AND *E. COLI* JM109

In order to test the conjugal ability of *E. coli* ET12567/pUZ8002, conjugation between *E. coli* ET12567/pUZ8002 and *E. coli* JM109 was performed according to the method described by Zhang *et al.* (2013) with some modifications. Briefly, the donor strain, *E. coli* ET12567/pUZ8002 harbouring the plasmid pOJAD2, was cultured in LB broth containing 50 µg/ml apramycin, 100 µg/ml chloramphenicol and 25 µg/ml kanamycin, while the recipient strain, *E. coli* JM109, was cultured in LB broth without antibiotics at 37 °C with shaking at 200 rpm until an OD_{600nm} = 0.5 was reached (mid-exponential phase). The cells were collected by centrifugation at 5000 rpm for 5 min, washed once with an equal volume of LB broth (without antibiotics) and resuspended in 1 ml LB medium before 50 µl of the donor strain was mixed with 50 µl of the recipient in a 1.5 ml microcentrifuge tube and supplemented with 900 µl LB broth which had been pre-warmed to 37 °C. After mixing, the cells were incubated at 37 °C without shaking for 24 h. The conjugation process was interrupted by vortexing for 10 sec and the mixture was immediately serially diluted 10-fold and plated on LB agar containing 50 µg/ml nalidixic acid and 50 µg/ml apramycin in order to select for the recipient strain (strain JM109 has a DNA gyrase mutation which makes it resistant to nalidixic acid).

4.3.7.1.3 INTERGENERIC CONJUGATION BETWEEN *E. COLI* AND *S. POLYANTIBIOTICUS* SPR^T

Classical intergeneric conjugation was performed between *E. coli* ET12567/pUZ8002 and *S. polyantibioticus* SPR^T and between *E. coli* ET12567/pUZ8002 and *S. coelicolor* A3(2), according to the method described by Flett *et al.* (1997). Conjugation was also performed between *E. coli* S17-1 and both *Streptomyces* strains according to the method described by Mazodier *et al.* (1989). Briefly for both methods, the donor *E. coli* strain (ET12567/pUZ8002 or S17-1 carrying pJN100) was cultured in 25 ml LB broth supplemented with 50 µg/ml apramycin, 34 µg/ml chloramphenicol and 50 µg/ml kanamycin to an OD_{600nm} = 0.5, pelleted by centrifugation at 5000 rpm for 5 min, washed twice with LB broth and resuspended in 1 ml

LB. Aliquots of *S. polyantibioticus* SPR^T and *S. coelicolor* A3(2) spore suspensions, which had been stored at -20 °C, were used as recipients. Spores were induced to germinate by heat-shock at 50 °C for 10, 20, 30 or 60 min, after which the donor *E. coli* cells were added to the prepared spores and the mixture was inoculated onto MS + 20 mM $MgCl_2$ plates. The conjugation plates were incubated for 16 h at 30 °C, after which the surface of each plate was overlaid with 1 ml of sterile dH_20 containing 500 µg nalidixic acid and 1 mg apramycin. The plates were incubated for a further 96 h at 30 °C and the exconjugant colonies were counted.

The classical conjugation method was modified in order to obtain *S. polyantibioticus* SPR^T exconjugants by growing the donor *E. coli* culture to an OD_{600nm} of 0.3, 0.4 or 0.6 (instead of 0.5, by decreasing the centrifugation speed to 3500 rpm (instead of 5000 rpm), by using a culture of *S. polyantibioticus* SPR^T which had been grown for 5 h, 12 h, 18 h, 24 h, 48 h and 72 h (instead of using a spore preparation), by cultivating *S. polyantibioticus* SPR^T mycelia in YEME, V or SMC media, and by plating the conjugation mixture on 7H9, YEME or ISP4 supplemented with 10 mM or 20 mM $MgCl_2$ (instead of MS supplemented with 10 mM or 20 mM $MgCl_2$). All of these modifications were tested as individual changes to the original method and in different combinations in order to optimise the method for transformation of *S. polyantibioticus* SPR^T.

A reproducible method for intergeneric conjugation between *E. coli* ET12567/pUZ8002 and *S. polyantibioticus* SPR^T was finally established based on the methods described by Marcone *et al.* (2010) and Du *et al.* (2012). Briefly, a culture of the donor strain, *E. coli* ET12567/pUZ8002 containing the selected plasmid was cultivated in 25 ml LB broth supplemented with 50 µg/ml apramycin, 34 µg/ml chloramphenicol and 50 µg/ml kanamycin to an OD_{600nm} = 0.4 (approximately 1 x 10^7 cells/ml). Cells were collected by centrifugation at 3500 rpm for 7 min, washed twice with an equal volume of LB broth and resuspended in 1 ml LB broth. *S. polyantibioticus* SPR^T mycelia were prepared as follows: fresh spores were scraped off the surface of an ISP4 plate and used to inoculate 25 ml SMC medium, containing 10.3 % sucrose and incubated at 30 °C on a rotary shaker at 200 rpm for 18 h (exponential phase). Mycelia were collected by centrifugation at 10 000 rpm for 10 min at 4 °C and resuspended in 2 ml ice cold LB. For conjugation, 0.5 ml donor *E. coli* ET12567/pUZ8002 cells carrying the selected plasmid (OD_{600nm} = 0.4) were added to 0.5 ml *S. polyantibioticus* mycelium and incubated in a 2 ml microcentrifuge tube at 30 °C for 1 h before serially diluting 1000 fold in LB broth and

plating on YEME agar supplemented with 10 mM or 20 mM $MgCl_2$. The conjugation plates were incubated at 30 °C for 18 h and then overlaid with 25 µg/ml nalidixic acid and 25 µg/ml apramycin. The plates were incubated for a further 72-96 h at 30 °C before single *Streptomyces* colonies were sub-inoculated onto fresh YEME plates containing 50 µg/ml nalidixic acid and 30 µg/ml apramycin to confirm the single crossover exconjugants. The conjugation frequency was calculated as the number of exconjugants per total number of donor *E. coli* cells.

To confirm the chromosomal integration of the pOJ260 and pJN100 derivative plasmids, gDNA was extracted from the exconjugants using the method described in section 2.3.2 and PCR amplification using the pOJFWD/pOJREV primers was performed (section 4.3.7.1.4).

4.3.7.1.4 CONFIRMATION OF SINGLE CROSSOVER EXCONJUGANTS BY PCR AMPLIFICATION

Amplification of gDNA isolated from *S. polyantibioticus* mutant strains ΔAD2, ΔA16, ΔA18, ΔA28, ΔA7, ΔPAAK, ΔTHI, ΔCIN, ΔLAC, ΔACY, ΔCYC and ΔA99 was performed using the POJFWD primer in combination with the respective reverse gene specific primer of interest: A7R for AD2, A16, A18, A28 and A7, PAAKREV for PAAK, THIR for THI, CINR for CIN, LACR for LAC, ACYR for ACY, CYCREV for CYC and ADREV for A99, in order to confirm the integration of the target gene within the mutant *S. polyantibioticus* genomes. Additionally, the POJREV primer was used in combination with the respective forward gene specific primer of interest: A3F for AD2, A16, A18, A28 and A7, PAAKFWD for PAAK, THIF for THI, CINF for CIN, LACF for LAC, ACYF for ACY, CYCFWD for CYC and ADFWD for A99, in an amplification reaction to confirm the integration of the target within the mutant *S. polyantibioticus* genomes.

The cycling conditions used for the amplification reactions were performed as follows: initial denaturation at 95 ^oC for 5 min, followed by 35 cycles of denaturation at 95 ^oC for 30 s, annealing at 56 ^oC for 30 s and elongation at 72 ^oC for 60 s, with a final elongation at 72 ^oC for 5 min. PCR reactions consisted of: 200 ng of DNA, 2 U SuperThem *Taq* polymerase (JMR Holdings, USA), 1.5 µM of each primer, 0.2 mM of each dNTP and 3 mM $MgCl_2$ in a total volume of 50 µl.

All PCR amplification products were resolved by electrophoresis alongside a λ-*Pst*I molecular marker on 0.8 % agarose gels containing 0.8 µg/ml ethidium bromide in order to analyse amplicon size. The products were visualized using a long wavelength UV light box (Bio-Rad Gel Doc EQ-system™, Bio-Rad Laboratories Inc, USA) and excised from the agarose gel, purified using the Favorgen Gel/PCR Purification Kit (Favorgen^{TM}, Germany) and sequenced by the dideoxy chain-termination method (Sanger *et al.,* 1977) on an Applied Biosystems Big Dye terminator v3.1 DNA sequencer using BIOLINE Half Dye Mix (Macrogen Inc., South Korea).

4.3.7.2 ELECTROPORATION

4.3.7.2.1 PREPARATION OF PLASMID DNA FOR ELECTROPORATION

The plasmids, pJN100 and pJNHYG, were transformed into *E. coli* dam$^-$/dcm$^-$ chemically competent cells, according to the method described by Dagert and Ehrlich, (1979), in order to generate unmethylated plasmid DNA, which was deemed necessary for efficient electroporation (Spath *et al.,* 2012). Thereafter, plasmid DNA was isolated using the NucleoSpin Plasmid Isolation kit (Machery Nagel, Germany) according to the manufacturer's instructions and quantitated spectrophotometrically as described before (section 2.3.2).

4.3.7.2.2 ELECTROPORATION

Electroporation of plasmid DNA into *S. polyantibioticus* SPR^T and *S. coelicolor* A3(2) was initially performed based on the method described by Pigac & Schrempf (1995). Briefly, *S. polyantibioticus* SPR^T and *S. coelicolor* A3(2) mycelia was individually cultured in 25 ml CRM at 30 °C on a rotary shaker at 220 rpm for 24-72 h. Each culture was harvested by centrifugation at 10 000 rpm at 4 °C, washed once with 25 ml ice cold 10 % sucrose, centrifuged at 10 000 rpm at 4 °C, re-suspended in 12.5 ml 15 % glycerol, centrifuged at 10 000 rpm at 4 °C and finally re-suspended in 2 ml ice cold 15 % glycerol containing 100 µg/ml lysozyme. The mycelial suspension was incubated at 37 °C for 1 h and then washed twice with ice-cold 15 % glycerol by centrifugation at 10 000 rpm at 4 °C, before re-suspension in 1 ml of the electroporation buffer, which consisted of sterile de-ionized dH_20 containing 30 % PEG-1000,

10 % glycerol and 6.5 % sucrose. The mycelial suspension was dispensed into 50 µl aliquots in 1.5 ml microcentrifuge tubes and immediately stored at -80 °C overnight. Subsequently, a 50 µl aliquot was thawed on ice and 10 ng to 1 µg of plasmid DNA (pJN100 or pJNHYG) was added to it, before the mixture was transferred into a 2 mm-gapped electrocuvette (Bio-Rad Laboratories Inc., USA) and subjected to an electrical pulse ranging from 0 to 12.5 kV/cm using a Gene Pulser (Bio-Rad Laboratories Inc., USA), which was connected to a pulse controller (25 µF capacitor) with a parallel resistor setting of 200, 400, 600 or 800 Ω. The pulsed mycelium was immediately diluted with ice cold 0.75 ml CRM and incubated on a rotary shaker at 220 rpm for 3 h at 30 °C. Before plating, CRM was added to a final volume of 1 ml and dilutions were inoculated onto TSB plates containing 30 µg/ml apramycin. Control reactions were performed by omitting either the plasmid DNA or the electrical pulse.

An alternative method described by Tyurin *et al.* (1995) for electroporation of freshly germinated mycelial fragments was also performed. Briefly, freshly harvested *S. polyantibioticus* SPR^T and *S. coelicolor* A3(2) spores were inoculated individually into 25 ml TSB and incubated on a rotary shaker (220 rpm) at 30 °C for 3-48 h. The cells were harvested by centrifugation at 10 000 rpm (4 °C) for 10 min, followed by washing twice with 0.15 M sucrose (centrifugation at 10 000 rpm for 10 min each time) and resuspension of the pellet in the electroporation buffer consisting of sterile de-ionized dH_20 containing 7 mM HEPES, 75 mM sucrose and 1 mM $MgCl_2$. The mycelial suspension was dispensed into 50 µl aliquots and stored at -80°C overnight. Subsequently, 10 ng to 1 µg of plasmid DNA was added to a thawed 50 µl mycelial aliquot and pulsed as described above. The pulsed mycelia were immediately diluted with 0.5 ml TSB, left on ice for 10 min and incubated at 30 °C for 1.5 h. Each electroporation reaction was inoculated onto TSB plates containing 30 µg/ml apramycin. Controls were performed by omitting either the plasmid DNA or the electrical pulse.

Another two methods described by Mazy-Servais *et al.* (1997) and Ma *et al.* (2013) were also tested. For the method described by Mazy-Servais *et al.* (1997), freshly harvested *S. polyantibioticus* SPR^T spores were inoculated into 25 ml YEME medium supplemented with 34 % sucrose and 0.5 % glycine and incubated on a rotary shaker (220 rpm) at 30 °C for 12-72 h. The mycelia were harvested by centrifugation at 10 000 rpm (4 °C) for 15 min, followed by washing three times with an equal volume of sterile dH_20 (centrifugation at 10 000 rpm for 10

min each time). Two different treatments were then applied to the cells: lysozyme digestion or suspension in PEG-supplemented electroporation buffer. In the case of the lysozyme treatment, cells were resuspended in 25 ml electroporation buffer (10 % sucrose, 15 % glycerol and 3 mM NaH_2PO_4-Na_2HPO_4 buffer, pH 7.4) containing 0.5 mg/ml lysozyme and incubated at 37 °C for 1 h, before the cells were harvested by centrifugation at 10 000 rpm (4 °C) for 15 min. The cells were resuspended in 1 ml ice cold electroporation buffer and distributed into 50 µl aliquots before electroporation. In the case of PEG treatment, cells were resuspended in 25 ml ice cold electroporation buffer, subjected to the same centrifugation conditions mentioned above and resuspended in 1 ml electroporation buffer containing 28.5 % PEG-1000, before distribution into 50 µl aliquots. Ten nanograms (10 ng) to one microgram (1 µg) of plasmid DNA was then added to both sets of treated aliquots, which were kept on ice for 1 min before the mixture was transferred to an electrocuvette and a single electrical pulse was applied according to the conditions shown in Table 4.3. The pulsed mycelia were diluted with ice cold 0.4 ml SOC solution, kept on ice for 5 min and plated onto R2YE agar plates containing 30 µg/ml apramycin.

Table 4.3 Electroporation conditions used in this study

Parallel resistance (Ω)	Pulse (kV/cm)	Capacitance (µF)
200	7.5	25
200	10	25
200	12.5	25
400	5	25
400	7.5	25
400	10	25
400	12.5	25
600	5	25
600	7.5	25
600	10	25
600	12.5	25
800	2.5	25
800	5	25
800	7.5	25
800	10	25
800	12.5	25

The method described by Ma *et al.* (2013) was performed in a similar manner to the method described by Tyurin *et al.* (1995) with some modifications. Briefly, the resuspension step consisted of resuspending the cell pellet in 15 % glycerol containing 25 mg/ml lysozyme and 10 µg/ml penicillin G, incubating at 37 °C overnight and then for a further 2 h at room temperature. Furthermore, the electroporation buffer consisted of 10 % PEG-1000, 10 % glycerol and 6.5 % sucrose, each electroporated sample was diluted in 1 ml CP medium (1 % glucose, 0.2 % tryptone, 0.4 % yeast extract, 0.05 % $MgSO_4.7H_2O$, 0.05 % K_2HPO_4, 0.05 % NaCl per litre of dH_2O, pH 7.2) instead of CRM and was plated on R2YE and YEME solid media.

In addition, modifications to the above mentioned methods were performed by replacing the wash buffers with 1 M sorbitol supplemented with 2, 4, 6 or 8 % DMSO, as well as with 1 mM citrate supplemented with 16 % raftilose and replacing the initial culture medium with TSB supplemented with 0.24 mM threonine and 1 % glycine. Other modifications to the electroporation conditions included double pulsing each sample and the addition of 5 µg of TypeOne Restriction Inhibitor (Epicentre, USA) to each electroporation reaction before pulsing in order to circumvent potential R-M systems.

4.3.7.3 PROTOPLAST TRANSFORMATION

4.3.7.3.1 PREPARATION OF PROTOPLASTS

S. polyantibioticus SPR^T and *S. coelicolor* A3(2) protoplasts were prepared according to the method described by Marcone *et al.* (2010) with some modifications. Briefly, a glycerol spore preparation (Kieser *et al.,* 2000) of each strain was inoculated into 25 ml V medium and cultured for 36-48 h at 30 °C on a rotary shaker (220 rpm). The culture was centrifuged at 4500 rpm for 10 min at 4 °C and washed once (centrifugation at 3250 rpm for 10 min) with 25 ml P medium (103 g sucrose, 0.25 g K_2SO_4, 2.02 g $MgCl_2.6H_2O$, 2 ml trace element solution, 1 ml 0.5 % KH_2PO_4, 3.68 % $CaCl_2.2H_2O$, 10 ml 5.73 % TES buffer per litre of dH_2O, pH 7.2) (Okanishi *et al.,* 1974). In order to achieve cell-wall digestion, 12.5 ml P medium (consisting of 20 mg/ml lysozyme, 0.018 mg/ml mutanolysin and 100 mg/l pluronic) was added to the cell pellet and incubated at 30 °C on a rotary shaker (50 rpm) for 48 h. Protoplast formation was monitored by using an Olympus CH20 microscope at 1000 x magnification. Protoplasts were

detached from residual mycelium clumps by gently drawing through a 5 ml glass pipette tip and then filtering through sterile cotton wool. Protoplasts were harvested by centrifugation at 4500 rpm for 10 min at 4 °C and finally resuspended in 1 ml P medium. Fifty microlitre (50 µl) aliquots of the protoplast suspension were added to microcentrifuge tubes, which were placed in a beaker of ice and then stored at -80 °C until needed.

4.3.7.3.2 PROTOPLAST TRANSFORMATION

One microgram of plasmid DNA (approximately 5 µl) was added to each protoplast suspension (section 4.3.7.3.1) in a microcentrifuge tube, followed by the addition of 200 µl T buffer and gentle mixing by pipetting up and down. After incubation at room temperature for 2 min, the transformation mixture was inoculated onto R2YE medium (Kieser *et al.,* 2000) or VMSO.1 agar medium (Marcone *et al.,* 2010). The plates were incubated at 30 °C for 20 h and then overlaid with 30 µg/ml apramycin.

4.3.8 FERMENTATION AND ISOLATION OF DPO

Spores and mycelial mass were collected from *S. polyantibioticus* strains SPR^T, ΔAD2, ΔA16, ΔA18, ΔA28, ΔA7, ΔPAAK, ΔCIN, ΔLAC, ΔCYC, ΔA99, ΔACY, ΔTHI and PJNACY from 10 day old ISP4 agar plates and heavy suspensions were made in 2 ml microcentrifuge tubes containing 1.5 ml sterile distilled water. The spore suspension of each strain was used to inoculate a seed culture of 25 ml Hacène's medium (HM; 5 g glucose, 4 g yeast extract powder, 10 g malt extract powder and 1 g NaCl per litre of dH_2O, pH 7.0) (Hacène & Lefebvre, 1995) in a 250 ml Erlenmeyer flask. The flask was incubated at 30 °C for 48 h on a rotary shaker, before inoculation into a 5 L Erlenmeyer flask containing 1 litre of HM. The flask was incubated at 30 °C on a rotary shaker for an additional 192 h (8 days) for a total fermentation time of 10 days.

The purity of each culture was confirmed by performing a standard Gram stain, after which DPO was isolated and purified from the cultures according to the method depicted in Figure 4.5 (method adapted from le Roes, 2005). Briefly, the culture was filtered through a 1 x 4 sized coffee filter (House of Coffees, RSA), after which DPO was extracted from the mycelial mass with methanol and from the culture filtrate with ethyl acetate by stirring at room temperature

on a magnetic stirrer overnight. After concentration to 25 ml, by evaporation in a fumehood, the methanol and ethyl acetate extracts were combined, the pH was adjusted to 7.0 (monitored with pH strips, Merck, Germany) and the mixture was re-extracted with 500 ml toluene by stirring at room temperature on a magnetic stirrer overnight. The toluene extract was concentrated to 50 ml by evaporation in a fumehood and extracted with 0.1 M sodium acetate buffer (pH 3.5) by stirring at room temperature on a magnetic stirrer overnight. The toluene layer was concentrated to 5 ml by evaporation in a fumehood and used directly in subsequent HPLC and TLC bioautography analysis experiments.

In addition, following the isolation of pure DPO from strains *S. polyantibioticus* SPR^T, *S. polyantibioticus* ΔAD2, *S. polyantibioticus* ΔA18, *S. polyantibioticus* ΔA16, *S. polyantibioticus* ΔA28, *S. polyantibioticus* ΔA7, *S. polyantibioticus* ΔPAAK, *S. polyantibioticus* ΔCYC, *S. polyantibioticus* ΔA99, *S. polyantibioticus* ΔACY, *S. polyantibioticus* ΔLAC, *S. polyantibioticus* ΔCIN and *S. polyantibioticus* ΔTHI, each DPO sample was subjected to UV spectrophotometry at a wavelength of 280 nm using the Nanodrop® ND-1000 Spectrophotometer (Coleman Technologies Inc., USA).

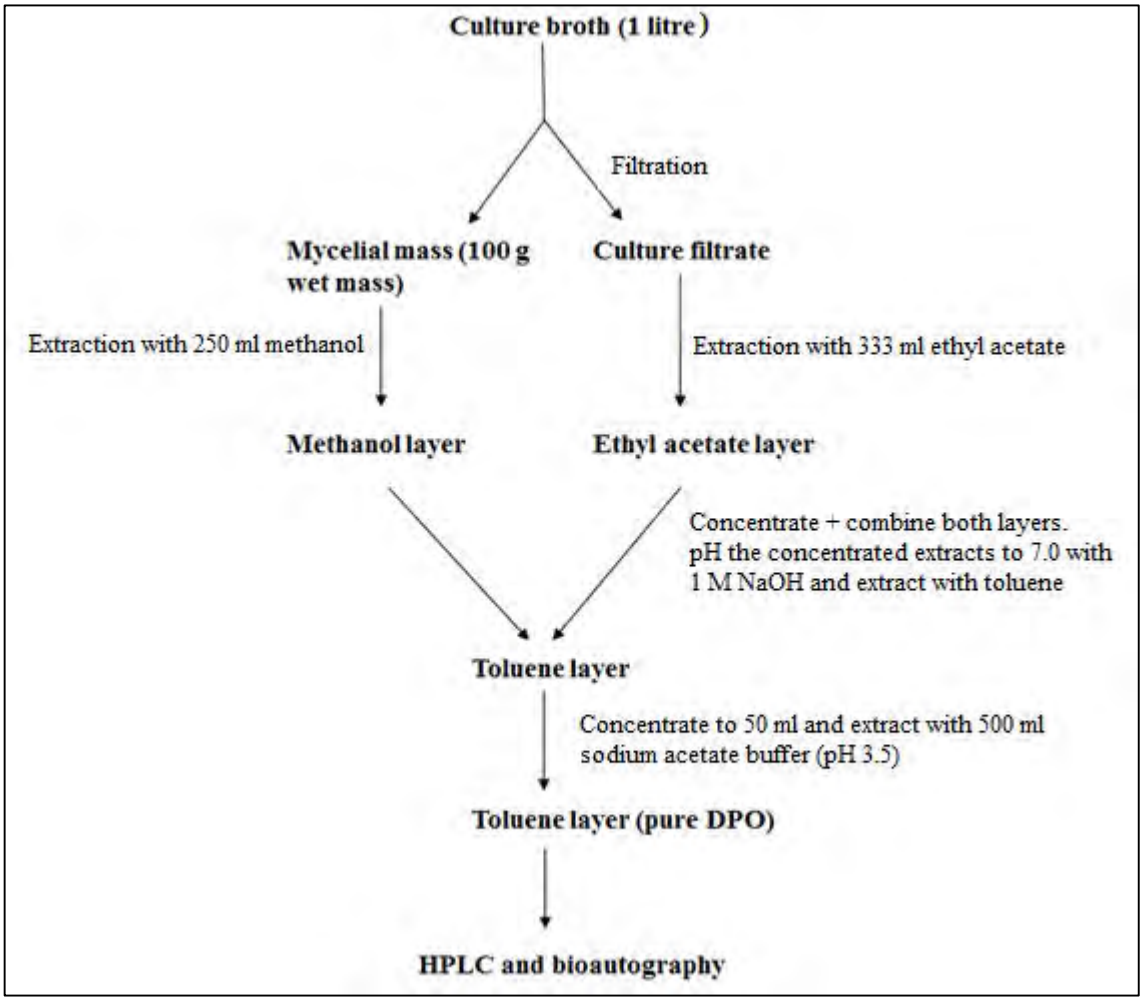

Figure 4.5 Schematic diagram depicting the isolation and purification of DPO.

4.3.9 HIGH PERFORMANCE LIQUID CHROMATOGRAPHY (HPLC)

Analytical reverse phase-HPLC analysis of a pure DPO sample (approximately 1 µg/ml) isolated from *S. polyantibioticus* SPR^T (section 4.3.8) and a commercial sample of DPO (Sigma-Aldrich, USA, catalogue number: D210404) (1 mg/ml dissolved in toluene) was carried out on an Agilent system (Agilent Technologies, USA) using a Vydac-C^{18} column (5 µm particle size, 4.66 x 250 mm) with 50 % acetonitrile as the mobile phase. The flow rate was 1 ml/min and DPO detection was at 280 nm.

4.3.10 THIN LAYER CHROMATOGRAPHY (TLC) BIOAUTOGRAPHY ANALYSIS

Each pure DPO sample (approximately 1 mg/ml) (section 4.3.8), a commercial DPO sample (Sigma-Aldrich, USA) (1 mg/ml dissolved in toluene) and the solvent, toluene, to serve as a negative control, was applied to the surface of a silica gel 60 F_{254} TLC plate (Merck, Germany, catalogue number: 1055530001) and allowed to dry in a fumehood for 15 min at room temperature. The plates were developed in a pre-saturated solvent chamber using ethyl acetate as the mobile phase until the solvent front reached approximately 1 cm from the top of the plates. The developed TLC plates were removed from the chamber and allowed to air-dry for 30 min, before being subjected to bioautography analysis against *M. aurum* A+.

M. aurum A+ was inoculated in LB broth and incubated at 37 °C with shaking for two days. After performing a standard Gram stain to confirm the purity of the culture, the optical density of the culture was determined on a Novaspec II spectrophotometer (Pharmacia Biotech, Sweden) at a wavelength of 600 nm. The culture was diluted to OD_{600nm} = 0.5 using sterile LB broth. The test bacterium was applied to the surface of the TLC plate (containing the DPO test sample, commercial DPO sample and toluene) with sterile non-absorbent cotton wool, the TLC plate was placed in a plastic sealable container lined with moist paper towel and incubated overnight at 37 °C (about 18 h).

Thiazolyl blue (MTT; Sigma-Aldrich, USA, catalogue number: M2128), dissolved in phosphate buffered saline (4.26 g $Na_2HPO_4.7H_2O$, 2.27 g KH_2PO_4, 8 g NaCl per litre of dH_2O, pH 7.0) at a final concentration of 0.25 % w/v, was applied to the plates using sterile non-absorbent cotton wool and the plates were incubated at 37 °C for a further 2 h. Colour changes were noted: MTT is yellow and turns purple when reduced, thereby indicating the presence of living cells, whereas white areas on the plate indicate zones where the test bacteria have been killed.

4.4 RESULTS AND DISCUSSION

4.4.1 DETERMINATION OF THE ANTIBIOTIC SUSCEPTIBILITY OF *S. POLYANTIBIOTICUS* SPR^T

The antibiotic susceptibility of *S. polyantibioticus* SPR^T was determined against apramycin, hygromycin B and kanamycin in order to establish an appropriate marker to select against untransformed cells in each transformation experiment (Table 4.4).

Table 4.4 The antibiotic susceptibility of *S. polyantibioticus* SPR^T against various antibiotics

Antibiotic	Concentration (µg/ml)	Growth
Apramycin	20	-
Kanamycin	50	-
Hygromycin B	100	-

S. polyantibioticus SPR^T exhibited sensitivity to apramycin, kanamycin and hygromycin B at concentrations of 20 µg/ml, 50 µg/ml and 100 µg/ml, respectively. Apramycin was chosen as the selectable marker for all subsequent gene disruption experiments due to the extensive availability of *E. coli-Streptomyces* shuttle vectors carrying the gene for apramycin resistance.

4.4.2 THE DEVELOPMENT OF AN OPTIMISED TRANSFORMATION PROTOCOL FOR *S. POLYANTIBIOTICUS* SPR^T

4.4.2.1 ELECTROPRATION AND PROTOPLAST TRANSFORMATION

In order to identify the genes involved in DPO biosynthesis, an efficient DNA transformation protocol was required for the introduction of foreign DNA into *S. polyantibioticus* SPR^T, to be able to disrupt genes to prove their involvement in the DPO biosynthetic pathway. Although published transformation protocols exist for the introduction of DNA into streptomycetes, there

is no method that is generally applicable to all species. Due to the fact that *S. polyantibioticus* SPR^T is a novel antibiotic-producing actinomycete, the establishment of a transformation protocol was required to make this strain amenable to genetic modifications.

Various methods for the delivery of pJN100 DNA were applied to *S. polyantibioticus* SPR^T cells in this study. Electroporation was the first method to be employed for transformation of DNA into *S. polyantibioticus* SPR^T, as it has been used in numerous *Streptomyces* species, including *S. coelicolor* A3(2), which was used as a positive control in the present study. The optimization of the classical electroporation methods described by Pigac & Schrempf (1995) and Tyurin *et al*. (1995) proved successful for transforming plasmid DNA into *S. coelicolor* A3(2). The highest number of *S. coelicolor* A3(2) transformants per µg of plasmid DNA was obtained using a field strength of 7.5 kV/cm, a parallel resistor setting of 800 Ω and a time constant of 18 ms (Figure 4.6). These optimized conditions were applied to the electroporation of *S. polyantibioticus* SPR^T cells without success.

Subsequently, an extensive optimization of four electroporation protocols was conducted. Addition of various agents that affect cell wall structure to the growth medium has been shown to enhance the electrotransformation of Gram-positive bacteria. High concentrations of glycine (1-4 % w/v) have been used to successfully inhibit cell wall synthesis and it has been reported that the addition of glycine and the osmotic stabilizer, sucrose, resulted in an increased transformation efficiency of *Lactococcus lactis* (Papagianni *et al.,* 2007; Holo and Nes, 1989). Furthermore, supplementation of the growth medium with glycine and L-threonine produced *B. subtilis* cells that could be electrotransformed much more efficiently at frequencies up to 2.5 x 10^3 transformants/µg of plasmid DNA (McDonald *et al*., 1995). Additionally, the supplementation of growth media with penicillin G, which affects the cross-linking of the cell wall in Gram-positive bacteria, was shown to enhance the transformation frequency in *Rhodococcus rhodochrous* CF222 (Sunairi *et al.,* 1996).

Initially, the filamentous growth of *S. polyantibioticus* SPR^T proved an obstacle for all of the methods, as the filaments aggregated. However, the addition of PEG-1000, sucrose and/or glycine was sufficient to impair the formation of the filaments. It was therefore noted that the most appropriate time span for electroporation was within the first 24 h of growth due to the

fact that a longer incubation time resulted in a high cell density culture containing aggregates of intertwined hyphae, which caused arcing during electroporation.

In order to circumvent the potent restriction-modification systems present in many *Streptomyces* species, all plasmid DNA used for electroporation into *S. polyantibioticus* SPR^T was isolated from *E. coli* dam$^-$ dcm$^-$ cells. Additionally, a restriction inhibitor, TypeOne Restriction Inhibitor (Epicentre, USA), was incorporated into each electroporation reaction in order to block the DNA binding site of type 1 R-M systems and inhibit cleavage of unmodified DNA.

However, despite efforts to optimize an efficient electroporation protocol for *S. polyantibioticus* SPR^T, none proved successful. Consequently, in an effort to avoid the problems caused by cell filaments and/or cell wall structures, protoplasts were used for the introduction of plasmid DNA into *S. polyantibioticus* SPR^T. Protoplast formation was achieved after a period of 48 h in *S. polyantibioticus* SPR^T, which was microscopically observed every 12 h up until the minimum required time for protoplast formation of 48 h, in contrast to the rapid protoplast formation, approximately 25 min, observed in *S. coelicolor* A3(2). Despite the success in preparing *S. polyantibioticus* SPR^T protoplasts using standard procedures, they could not be regenerated on the standard regeneration medium, R2YE (Kieser *et al*., 2000) or the alternative medium, VMSO.1 (Marcone *et al.,* 2010) under conditions where protoplasts of other *Streptomyces* spp. were regenerated. Thus, it was never possible to test protoplast transformation of *S. polyantibioticus* SPR^T cells.

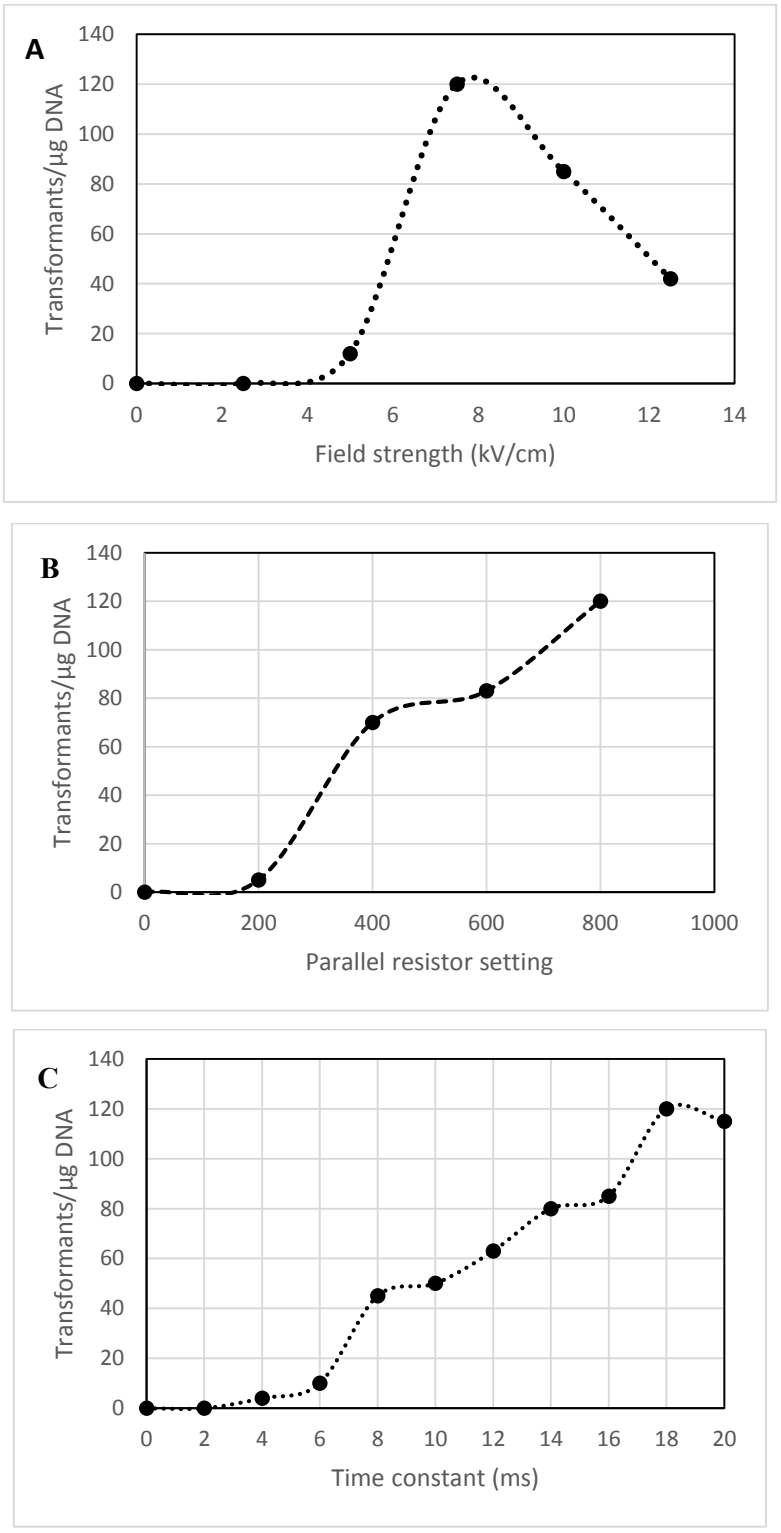

Figure 4.6 Effects of various electroporation parameters on the transformation efficiency of *S. coelicolor* A3(2). (A) Effect of the initial voltage of the applied electrical pulse on transformation efficiency of the electroporated cells at 25 µF and 800 Ω. (B) Effect of the external parallel resistance on transformation efficiency of the electroporated cells at 25 µF and 7.5 kV/cm. (C) Effect of the electrical pulse duration (time constant) on transformation efficiency of electroporated cells at 25 µF and 7.5 kV/cm.

It is possible that the inability to transform DNA into *S. polyantibioticus* SPR^T, using electroporation, could be attributed to the inefficient expression of the heterologous apramycin-resistance gene promoter in plasmid pJN100 when transformed into *S. polyantibioticus* SPR^T. The apramycin resistance gene was originally isolated from *Klebsiella pneumoniae* and it has previously been reported that when the promoter being used is heterologous to the host strain, expression may be feeble or non-existent due to unknown reasons (Wilkinson *et al.,* 2002). The apramycin resistance gene cassette in pJN100 was therefore replaced with the hygromycin resistance cassette (originally isolated from *Streptomyces hygroscopicus*) to generate the modified vector, pJNHYG. It was hoped that the hygromycin resistance gene would be more efficiently expressed in *S. polyantibioticus* SPR^T, due to the fact that it was derived from the same genus, thereby allowing transformation and integration of pJNHYG in *S. polyantibioticus* SPR^T. However, no exconjugants were observed after transformation with pJNHYG, suggesting that the problem was one of getting plasmid DNA into *S. polyantibioticus* SPR^T rather than one of antibiotic selection.

4.4.2.2 INTERGENERIC CONJUGATION

Intergeneric conjugation has been used to circumvent plasmid DNA transformation problems in *Streptomyces* strains that are recalcitrant to protoplast transformation, most likely due to the fact that ssDNA is transferred thereby bypassing any R-M systems (Voeikova, 1999). Additionally, intergeneric conjugation allows for the utilization of *E. coli* shuttle vectors containing the *oriT* function that can be used for targeted gene disruption (Kieser *et al.,* 2000).

Two *E. coli* strains were used as donors in intergeneric conjugations, S17-1 and ET12567/pUZ8002, with *S. coelicolor* A3(2) and *S. polyantibioticus* SPR^T as the recipient strains. Additionally, the conjugal ability of *E. coli* ET12567/pUZ8002 harbouring pOJAD2 was determined by successful mating with *E. coli* JM109.

Donor cultures carrying pJN100 were mated individually with *S. coelicolor* A3(2) and *S. polyantibioticus* SPR^T. The *E. coli* S17-1 and ET12567/pUZ8002 strains were efficient in transferring the pJN100 plasmid to *S. coelicolor* A3(2) using classical conjugation methods with conjugation frequencies of 2.1 x 10^{-5} and 1.83 x 10^{-5} exconjugants per 1 x 10^7 donor *E. coli* cells, respectively, which is comparable to the 1 x 10^{-5} exconjugants obtained by Flett *et*

al. (1997) (Figure 4.7). However, the transfer of pJN100 into *S. polyantibioticus* SPR^T was only successful using *E. coli* ET12567/pUZ8002 as the donor strain with a conjugation frequency of 8.3 x 10^{-6} exconjugants per 1 x 10^7 donor *E. coli* cells (Figure 4.7). The unsuccessful mating between *E. coli* S17-1 and *S. polyantibioticus* SPR^T may have been due to the fact that the vigorous growth of *E. coli* S17-1 caused the donor strain to rapidly reach stationary phase and thereby overgrow relative to *S. polyantibioticus* SPR^T, before *E. coli*-streptomycete mating junctions had been established (Flett *et al.,* 1997).

In contrast to the classical conjugation methods described by Mazodier *et al.* (1989) and Flett *et al.* (1997), whereby a glycerol spore preparation of the recipient strain was subjected to heat treatment to induce germination and thereafter mated directly with the donor *E. coli* strain, freshly germinated *S. polyantibioticus* SPR^T spores were inoculated into liquid growth medium for cultivation for 18 h before mating took place. Furthermore, only liquid SMC medium containing 10.3 % sucrose produced a dispersed culture of *S. polyantibioticus* SPR^T mycelia that promoted the transfer of plasmid DNA across the cell wall. No exconjugants were observed when the *S. polyantibioticus* SPR^T cells were grown in any other medium.
In addition, it has been reported that the mycelial age of the recipient strain is a critically important factor for the uptake of plasmid DNA in terms of competence of the recipient cells. Luo *et al.* (2009) reported that *Nonomuraea* mycelia collected during the exponential phase were optimal for intergeneric conjugation, whereas mycelia collected from *Streptomyces peucetius* after 34 h growth were most favourable for intergeneric conjugation (Paranthaman *et al.,* 2003). Indeed, it was observed that *S. polyantibioticus* SPR^T mycelia collected after 18 h were optimal for intergeneric conjugation with *E. coli* ET12567/pUZ8002/pJN100 with a conjugation frequency of 8.3 x 10^{-6} exconjugants per 1 x 10^7 donor *E. coli* cells (Figure 4.8). The conjugation frequency decreased 3.3-fold to 2.5 x 10^{-6} after 24 h of cultivation. Mycelia collected earlier than 18 h or later than 24 h were not suitable for conjugation as no exconjugants were observed.

Lastly, it has been observed that the solid medium on which the intergeneric conjugation takes place has a significant influence on the conjugation frequency in various streptomycetes (Du *et al*., 2012; Guan & Pettis, 2009; Kim *et al.,* 2008; Choi *et al.,* 2004; Kitani *et al.,* 2000; Flett *et al*., 1997). In this study, four solid media, MS, YEME, 7H9 and ISP Medium 4, each supplemented with either 10 mM $MgCl_2$ or 20 mM $MgCl_2$, were tested in order to determine

which one promoted intergeneric conjugation between *E. coli* ET12567/pUZ8002/pJN100 and *S. polyantibioticus* SPR^T the most. Exconjugants were only obtained on YEME and 7H9 media, while no exconjugants were observed on MS or ISP4 media (Figure 4.9). YEME media supplemented with 20 mM $MgCl_2$ proved to be optimal for the conjugation of *E. coli* ET12567/pUZ8002/pJN100 and *S. polyantibioticus* SPR^T with a conjugation frequency of 8.2 x 10^{-6} exconjugants per 1 x 10^7 donor *E. coli* cells, 1.1-fold higher than that of YEME supplemented with 10 mM $MgCl_2$ and 2.3-fold higher than 7H9 medium supplemented with 10 mM $MgCl_2$ (conjugation frequency of 3.5 x 10^{-6} exconjugants per 1 x 10^7 donor *E. coli* cells). No significant differences were observed between media supplemented with 10 mM $MgCl_2$ or 20 mM $MgCl_2$.

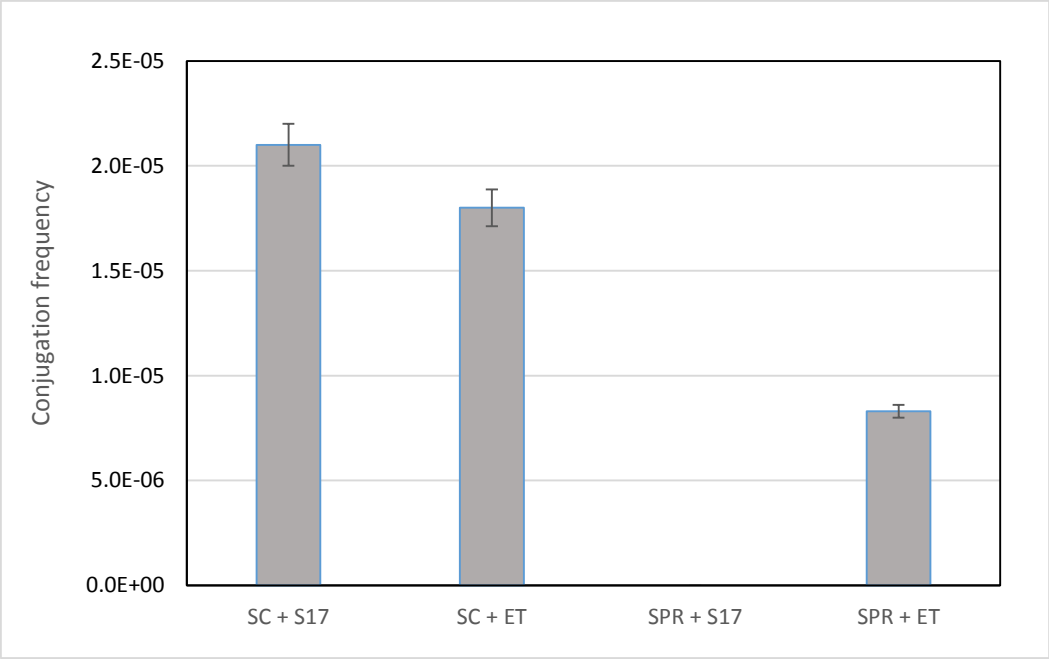

Figure 4.7 Effect of *E. coli* donor on intergeneric conjugation with *S. coelicolor* A3(2) and *S. polyantibioticus* SPR^T. The data are presented as the mean ± SEM of three independent experiments of the number of *S. polyantibioticus* SPR^T exconjugants observed per 1 x 10^7 donor *E. coli* cells. SC denotes *S.coelicolor* A3(2), SPR denotes *S. polyantibioticus* SPR^T, S17 denotes *E. coli* S17-1/pJN100 and ET denotes *E. coli* ET12567/pUZ8002/pJN100.

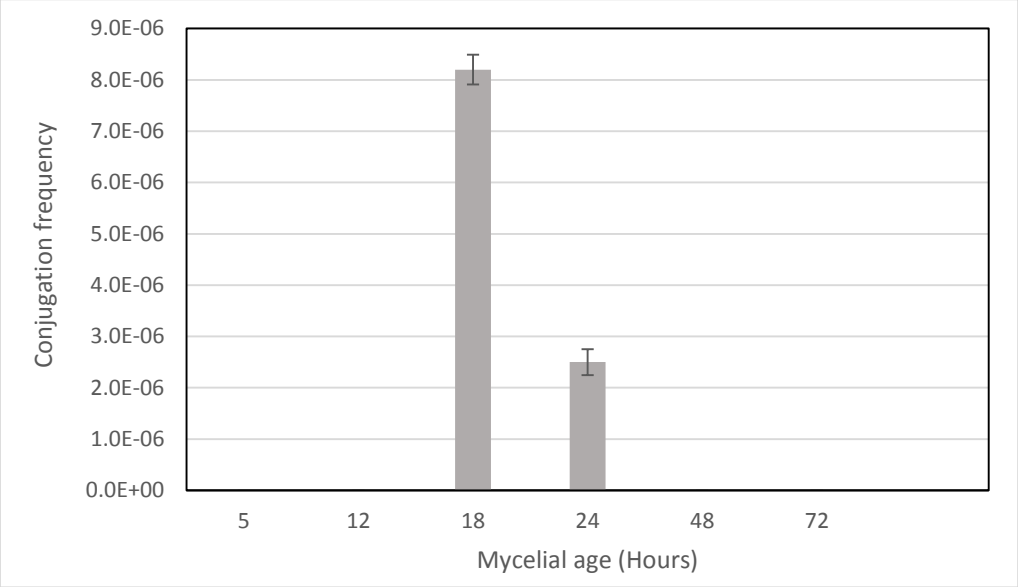

Figure 4.8 Effect of the mycelial age of the recipient strain, *S. polyantibioticus* SPR^T, on intergeneric conjugation with *E. coli* ET12567/pUZ8002/pJN100. The data are presented as the mean ± SEM of three independent experiments of the number of *S. polyantibioticus* SPR^T exconjugants observed per 1 x 10^7 donor *E. coli* cells.

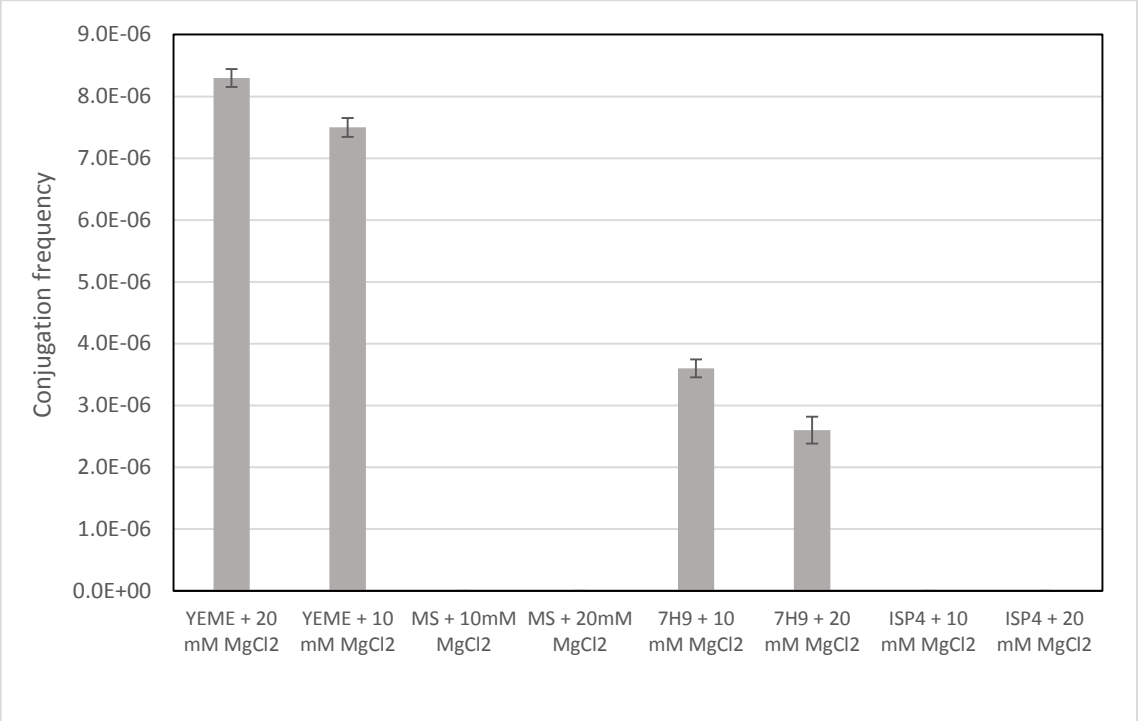

Figure 4.9 Effect of mycelial solid growth medium on intergeneric conjugation between *E. coli* ET12567/pUZ8002/pJN100 and *S. polyantibioticus* SPR^T. The data are presented as the mean ± SEM of three independent experiments of the number of *S. polyantibioticus* SPR^T exconjugants observed per 1 x 10^7 donor *E. coli* cells.

4.4.3 GENE DISRUPTION USING OPTIMIZED INTERGENERIC CONJUGATION METHOD

In order to determine which NRPS A domain is involved in DPO biosynthesis, in addition to testing the involvement of the genes encoding the phenylacetate-CoA ligase, thioesterase, cinnamate-CoA ligase, D-lactate dehydrogenase and acyl-CoA synthetase (as well as the putative NRPS Cy domain) in DPO synthesis, it was decided that all of these genes could be knocked-out in *S. polyantibioticus* SPR^T using the method of homologous recombination.

The shuttle vector, pOJ260, was genetically manipulated for use in the gene disruption experiments. Briefly, fragments of the genes encoding the A domain, Cy domain, phenylacetate-CoA ligase, thioesterase, cinnamate-CoA ligase, D-lactate dehydrogenase and

acyl-CoA synthetase were individually cloned into pOJ260 and transformed into *E. coli* ET12567/pUZ8002 for use in subsequent intergeneric conjugation with *S. polyantibioticus* SPR^T using the optimized, strain-specific conjugation method described in section 4.4.2.2.

The gene disruption experiments were based on the premise that the plasmid, pOJ260, carrying a homologous region to a gene in *S. polyantibioticus* SPR^T, would integrate into the genome by a single crossover event and thereby disrupt the target gene (Figure 4.10). It would then be possible to infer the gene's involvement in the production of DPO, should the mutant strain lose the ability to produce DPO. In order to detect the production of DPO after each respective gene knockout, DPO was extracted from the resulting mutants and subjected to bioautography against *M. aurum* A+. A loss in the detection of activity of the extract against *M. aurum* A+ indicated a loss of production of DPO and thereby confirmed the gene's involvement in DPO biosynthesis.

Fragments of the A domains designated AD2, A18, A16, A28, A7 and A99, in addition to the genes encoding the phenylacetate-CoA ligase (PAAK), the Cy domain (CYC), acyl-CoA synthetase (ACY), D-lactate dehydrogenase (LAC), cinnamate-CoA ligase (CIN) and thioesterase (THI), were PCR-amplified from the *S. polyantibioticus* SPR^T genome and individually cloned into pOJ260. The resulting plasmids were designated pOJAD2, pOJA18, pOJA16, pOJA28, pOJA7, pOJPAAK, pOJCYC, pOJA99, pOJACY, pOJLAC, pOJCIN and pOJTHI, and were individually transformed into *E. coli* ET12567/pUZ8002 and introduced into *S. polyantibioticus* SPR^T via intergeneric conjugation to yield the mutant strains *S. polyantibioticus* ΔAD2, *S. polyantibioticus* ΔA18, *S. polyantibioticus* ΔA16, *S. polyantibioticus* ΔA28, *S. polyantibioticus* ΔA7, *S. polyantibioticus* ΔPAAK, *S. polyantibioticus* ΔCYC, *S. polyantibioticus* ΔA99, *S. polyantibioticus* ΔACY, *S. polyantibioticus* ΔLAC, *S. polyantibioticus* ΔCIN and *S. polyantibioticus* ΔTHI, respectively. The location of DNA fragments used to make knock-out constructs within the genes they target, according to Figure 4.10, is provided in Appendix C.

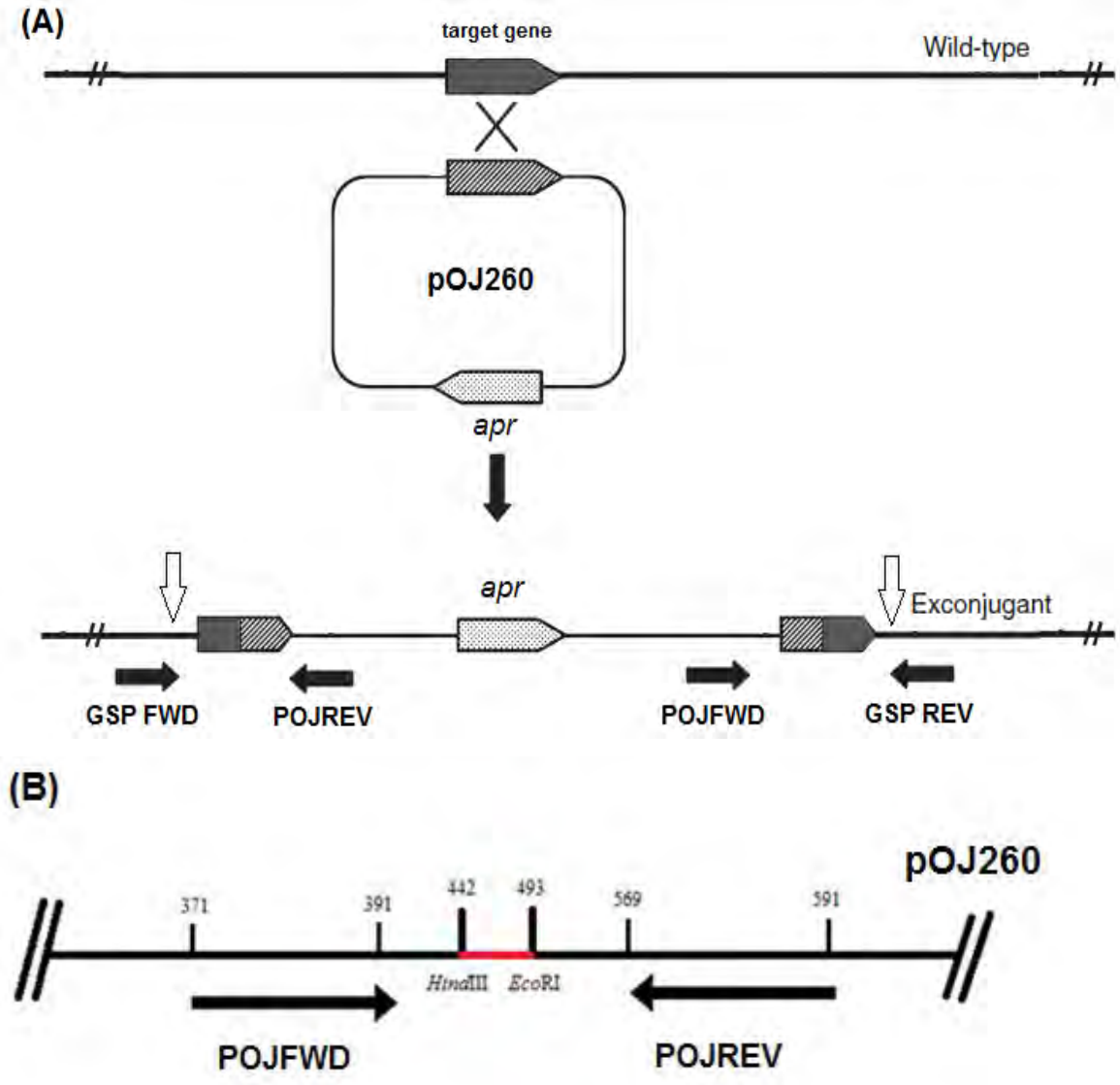

Figure 4.10 Schematic diagram representing the strategy employed in the construction of a theoretical gene A disruption mutant **(A)**. Gene disruption was achieved by inactivation of the target gene via single crossover homologous recombination. The horizontal black arrows represent the position of the binding sites of the PCR primers that were designed to confirm the integration of the plasmid DNA, carrying the target gene of interest, into the wild-type chromosome. GSP refers to the gene specific primer, depending on which target gene had been integrated into the *S. polyantibioticus* SPR^T genome. The GSP primer was used in conjunction with either the POJFWD or POJREV primer. The binding positions of the POJFWD and POJREV primers on the recombinant pOJ260 plasmid before integration of the whole plasmid into the genome is also depicted. The apramycin resistance gene is represented by the abbreviation, *apr*. The nucleotide sequences of the regions represented by the vertical arrows is provided in Appendix C for each knock-out construct. **(B)**. The red line indicates the target gene that has been cloned into the multiple cloning region of pOJ260 and the numbers indicate the exact binding positions on the plasmid according to the pOJ260 plasmid map (Figure 4.3).

Confirmation of the integration of the vector carrying the target gene of interest into the genome of each mutant strain was determined by PCR amplification across the junctions of each copy of the mutant gene derived from homologous recombination using the POJFWD primer and the respective reverse gene specific primer (Figure 4.11-4.12). Amplification using the forward gene specific primer of interest and POJREV in combination also confirmed the integration of the vector carrying the target gene of interest (data not shown). Due to the binding positions of the POJFWD and POJREV primers, the amplified products were 72 bp longer than the amplified product obtained using the gene specific primer set, when using the POJFWD primer in combination with the respective reverse target gene specific primer and 97 bp longer using the POJREV primer in combination with the forward target gene specific primer (Table 4.5).

Additionally, the full open reading frame of the acyl-CoA synthetase was PCR-amplified and cloned into the plasmid vector pJN100 to generate pJNACY, which was transformed into *E. coli* ET12567/pUZ8002 and introduced into *S. polyantibioticus* ΔACY via intergeneric conjugation in order to complement the acyl-CoA synthetase mutant strain, *S. polyantibioticus* ΔACY.

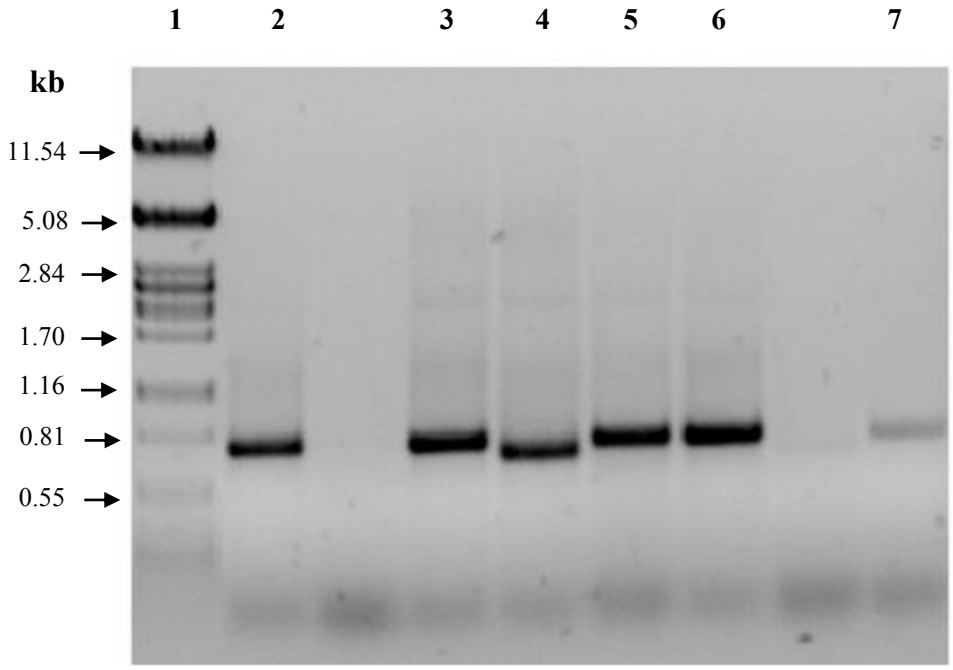

Figure 4.11 Gel electrophoresis of PCR amplified A domains from *S. polyantibioticus* SPR^T using the POJFWD and A7R primer set. Lanes: **1**- λ-*Pst*I molecular marker, **2-** A18 amplification product from *S. polyantibioticus* Δ18 gDNA, **3**- A28 amplification product from *S. polyantibioticus* Δ28 gDNA, **4**- A16 amplification product from *S. polyantibioticus* Δ16 gDNA, **5**- A99 amplification product from *S. polyantibioticus* Δ99 gDNA, **6**- A7 amplification product from *S. polyantibioticus* Δ7 gDNA, **7**- AD2 amplification product from *S. polyantibioticus* ΔAD2 gDNA.

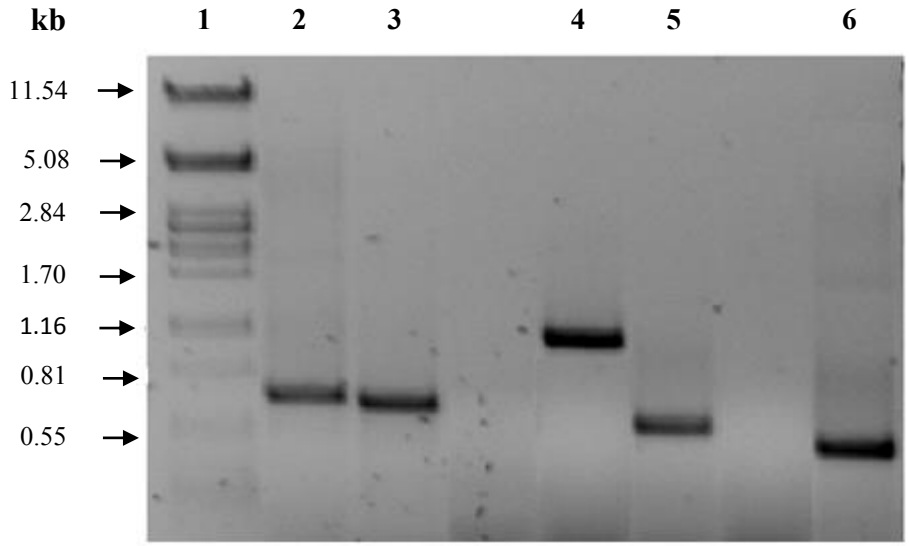

Figure 4.12 Gel electrophoresis of PCR amplified target genes from *S. polyantibioticus* SPR^T using the POJFWD and reverse gene specific primers in combination. Lanes: **1**- λ-*Pst*I molecular marker, **2**- CIN amplification product from *S. polyantibioticus* ΔCIN gDNA, **3**- LAC amplifcation product from *S. polyantibioticus* ΔLAC gDNA, **4**- CYC amplification product from *S. polyantibioticus* ΔCYC gDNA, **5**- PAAK amplification product from *S. polyantibioticus* ΔPAAK gDNA, **6**- THI amplification product, from *S. polyantibioticus* ΔTHI gDNA.

Table 4.5 PCR amplification product sizes of target genes using POJFWD and POJREV primers

Target gene	Product size using POJFWD and gene specific reverse primer (bp)	Product size using gene specific forward primer and POJREV (bp)
A18	742	767
A16	771	796
A28	780	805
AD2	794	817
A7	802	827
A99	792	817
CIN	721	746
LAC	680	705
THI	545	570
CYC	1052	1149
PAAK	588	613
ACY	993	1017

4.4.4 ISOLATION OF DPO FROM *S. POLYANTIBIOTICUS* SPR^T AND CONFIRMATION OF ITS ACTIVITY AGAINST *M. AURUM* A+

The antibacterial compound, DPO, was isolated from the fermentation broth and mycelial mass of *S. polyantibioticus* SPR^T. The activity of the compound was detected through bioautography, displaying an R*f* of 0.88 with *M. aurum* A+ as the test organism.

Chemically-synthesized 2,5-diphenyloxazole was used as a positive control for the detection of the bioactivity and its R*f* value correlated with the biologically-produced compound (Figure 4.13). Furthermore, the solvent used in the purification of DPO, toluene, was employed as a negative control in the detection of bioactivity in order to ensure that no contaminating compounds with similar bioactivity to DPO were present.

Reverse-phase HPLC was employed to corroborate the results obtained from the bioautography and the retention times measured at 280 nm of the biologically-produced DPO and chemically-synthesized DPO were identical. The results of the TLC-bioautography analysis and reverse-phase HPLC confirmed the production of DPO by *S. polyantibioticus* SPR^T and its activity against *M. aurum* A+.

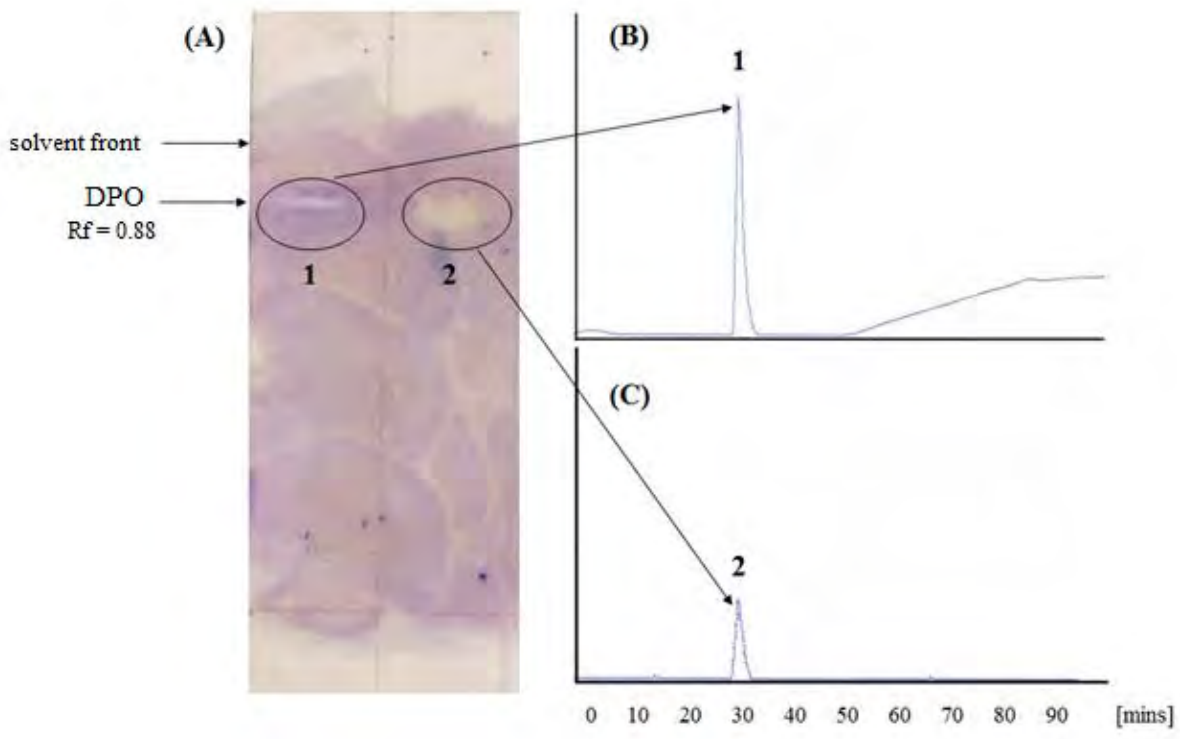

Figure 4.13 TLC bioautography and RP-HPLC analysis of the biologically-synthesized DPO sample isolated from *S. polyantibioticus* SPR^T and the chemically synthesized DPO sample. (A) TLC plates stained with MTT showing that both 1) the chemically synthesized DPO sample and 2) the biologically-synthesized DPO sample had antibacterial activity and an R*f* of 0.88 in the solvent system used (100% ethyl acetate). (B) RP-HPLC analysis depicting the retention time, 29.65 min, of the chemically synthesized DPO sample measured at 280 nm. (C) RP-HPLC analysis depicting the retention time, 29.65 min, of the biologically-synthesized DPO sample measured at 280 nm.

4.4.5 TLC BIOAUTOGRAPHY ANALYSIS TO DETERMINE THE PUTATIVE INVOLVEMENT OF THE TARGET GENES IN DPO BIOSYNTHESIS

In order to directly determine whether the A domains designated AD2, A18, A16, A28, A7 and A99, as well as the genes encoding the phenylacetate-CoA ligase (PAAK), the Cy domain (CYC), acyl-CoA synthetase (ACY), D-lactate dehydrogenase (LAC), cinnamate-CoA ligase (CIN) and thioesterase (THI), are involved in the biosynthesis of DPO, each gene was disrupted and the resulting mutant strain, lacking a functional copy of the target gene, was assayed for activity against *M. aurum* A+. A loss in activity would suggest the gene's involvement in DPO biosynthesis.

The DPO purification method was carried out on cultures of strains *S. polyantibioticus* SPR^T, *S. polyantibioticus* ΔAD2, *S. polyantibioticus* ΔA18, *S. polyantibioticus* ΔA16, *S. polyantibioticus* ΔA28, *S. polyantibioticus* ΔA7, *S. polyantibioticus* ΔPAAK, *S. polyantibioticus* ΔCYC, *S. polyantibioticus* ΔA99, *S. polyantibioticus* ΔACY, *S. polyantibioticus* ΔLAC, *S. polyantibioticus* ΔCIN and *S. polyantibioticus* ΔTHI. Each sample was subjected to UV spectrophotometry which confirmed the presence of a single peak at a wavelength of 280 nm in the samples from *S. polyantibioticus* SPR^T, *S. polyantibioticus* ΔAD2, *S. polyantibioticus* ΔA18, *S. polyantibioticus* ΔA16, *S. polyantibioticus* ΔA28, *S. polyantibioticus* ΔA7, *S. polyantibioticus* ΔPAAK, *S. polyantibioticus* ΔLAC, *S. polyantibioticus* ΔCIN and *S. polyantibioticus* ΔTHI. In contrast, the samples from strains *S. polyantibioticus* ΔCYC, *S. polyantibioticus* ΔA99 and *S. polyantibioticus* ΔACY displayed a reading of zero at a wavelength of 280 nm, thereby indicating that these strains had lost the ability to produce DPO.

TLC-bioautography analysis was performed on the extracts from each strain, which showed that *S. polyantibioticus* ΔAD2, *S. polyantibioticus* ΔA18, *S. polyantibioticus* ΔA16, *S. polyantibioticus* ΔA28, *S. polyantibioticus* ΔA7, *S. polyantibioticus* ΔPAAK, *S. polyantibioticus* ΔLAC, *S. polyantibioticus* ΔCIN and *S. polyantibioticus* ΔTHI all had activity against *M. aurum* A+ (Figures 4.14 and 4.15). This showed that disruption of these genes did *not* abolish DPO production, which indicated that none of these genes is involved in DPO biosynthesis. However, the extracts from strains *S. polyantibioticus* ΔCYC (Figure 4.16), *S. polyantibioticus* ΔA99 (Figure 4.17) and *S. polyantibioticus* ΔACY (Figure 4.18) showed no

activity against *M. aurum* A+, implying that the disruption of these genes resulted in the strains losing the ability to produce DPO.

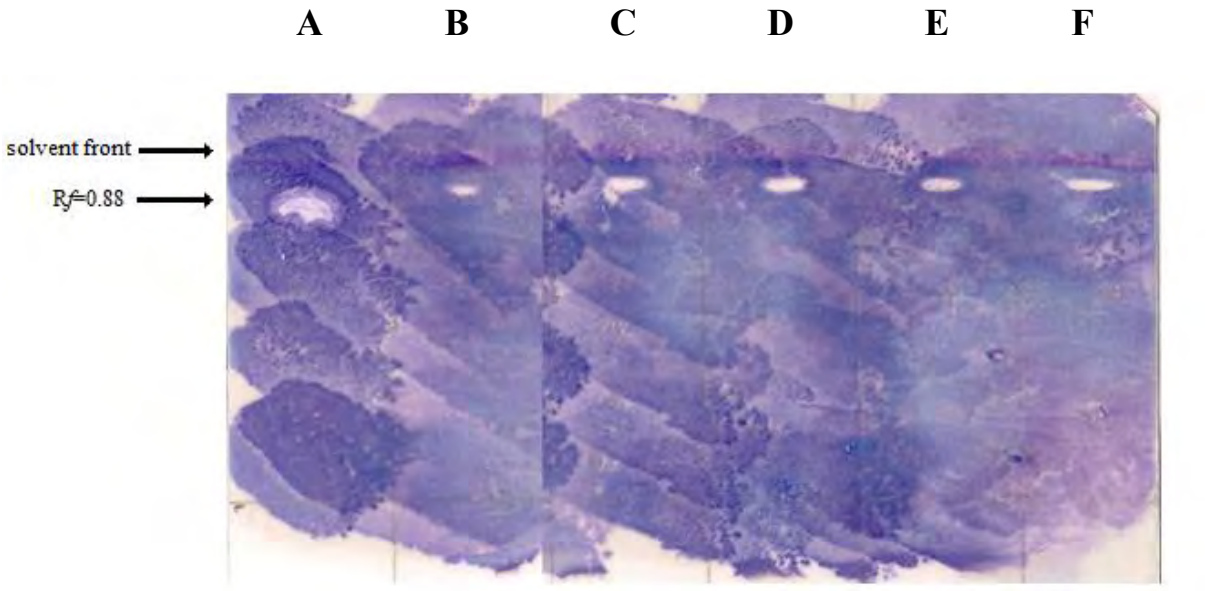

Figure 4.14 TLC plates stained with MTT showing the bioautographic analysis of DPO isolated from lane B) *S. polyantibioticus* ΔAD2, lane C) *S. polyantibioticus* ΔA18, lane D) *S. polyantibioticus* ΔA16, lane E) *S. polyantibioticus* ΔA28 and lane F) *S. polyantibioticus* ΔA7, where lane A) is chemically-synthesized DPO acting as a positive control.

Additionally, the introduction of pJNACY into *S. polyantibioticus* ΔACY via intergeneric conjugation in a complementation experiment, resulted in the restoration of DPO biosynthesis and the ability to inhibit *M. aurum* A+ (Figure 4.18).

These results demonstrated the involvement of the Cy domain encoded by gene SPR_53040, the A domain encoded by gene SPR_53060 and the acyl-CoA synthetase encoded by gene SPR_52860 in the production of DPO. Furthermore, these results excluded the involvement of the cinnamate-CoA ligase encoded by gene SPR_60150, the putative D-lactate dehydrogenase encoded by gene SPR_60250 and the TE domain encoded by gene SPR_53090

in the biosynthesis of benzoic acid, which is proposed to be required for DPO biosynthesis. It also showed that none of the other A domains investigated is involved in DPO biosynthesis.

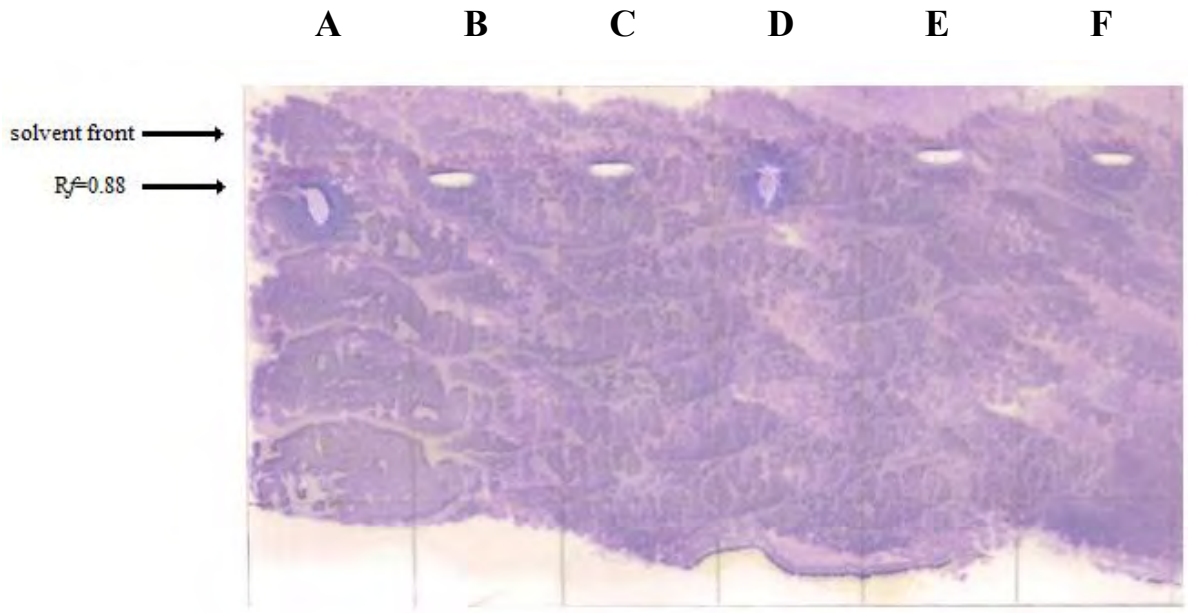

Figure 4.15 TLC plates stained with MTT showing the bioautographic analysis of DPO isolated from lane B) *S. polyantibioticus* ΔPAAK, lane C) *S. polyantibioticus* ΔLAC, lane E) *S. polyantibioticus* ΔCIN and lane F) *S. polyantibioticus* ΔTHI, where lanes A) and D) are chemically-synthesized DPO acting as a positive control.

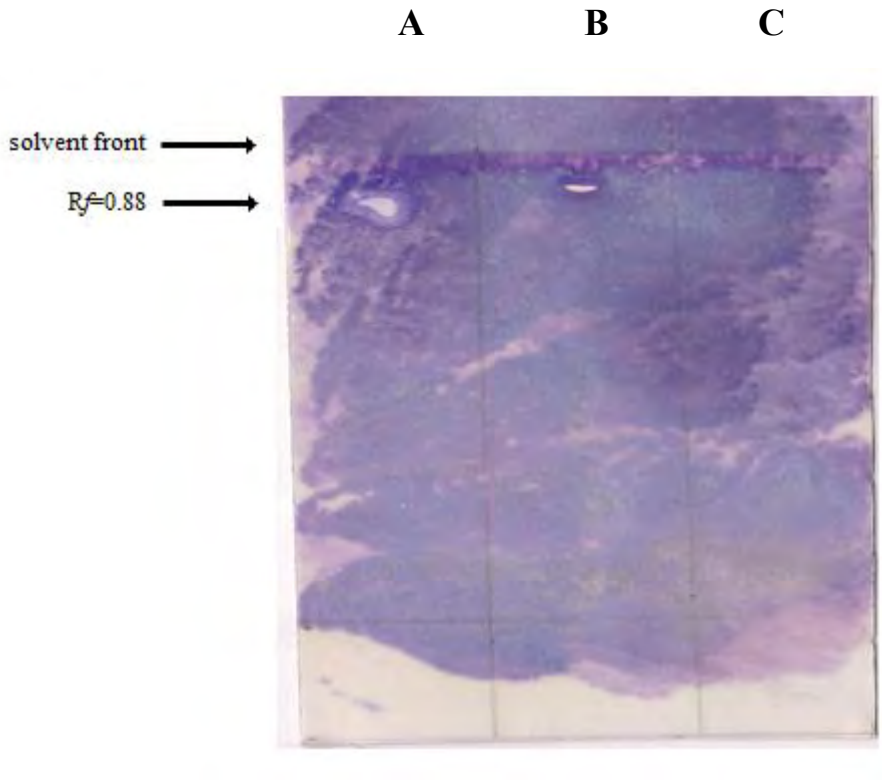

Figure 4.16 TLC plates stained with MTT showing the bioautographic analysis of DPO isolated from lane B) *S. polyantibioticus* SPR^T, lane C) *S. polyantibioticus* ΔCYC, where lane A) is chemically-synthesized DPO acting as a positive control.

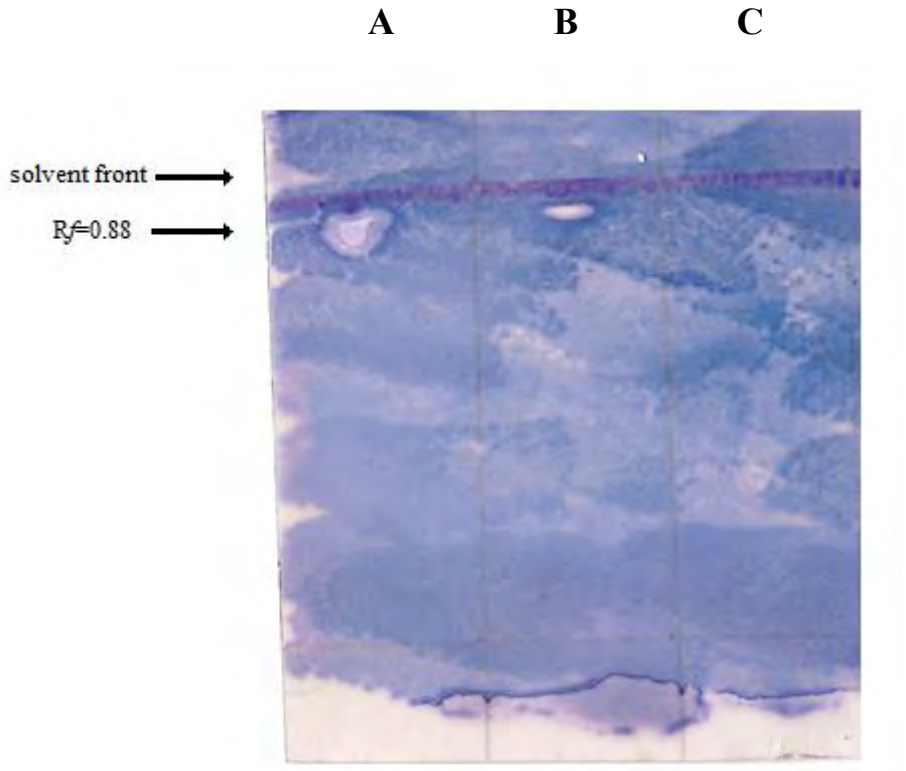

Figure 4.17 TLC plates stained with MTT showing the bioautographic analysis of DPO isolated from lane B) *S. polyantibioticus* SPR^T, lane C) *S. polyantibioticus* ΔA99, where lane A) is chemically-synthesized DPO acting as a positive control.

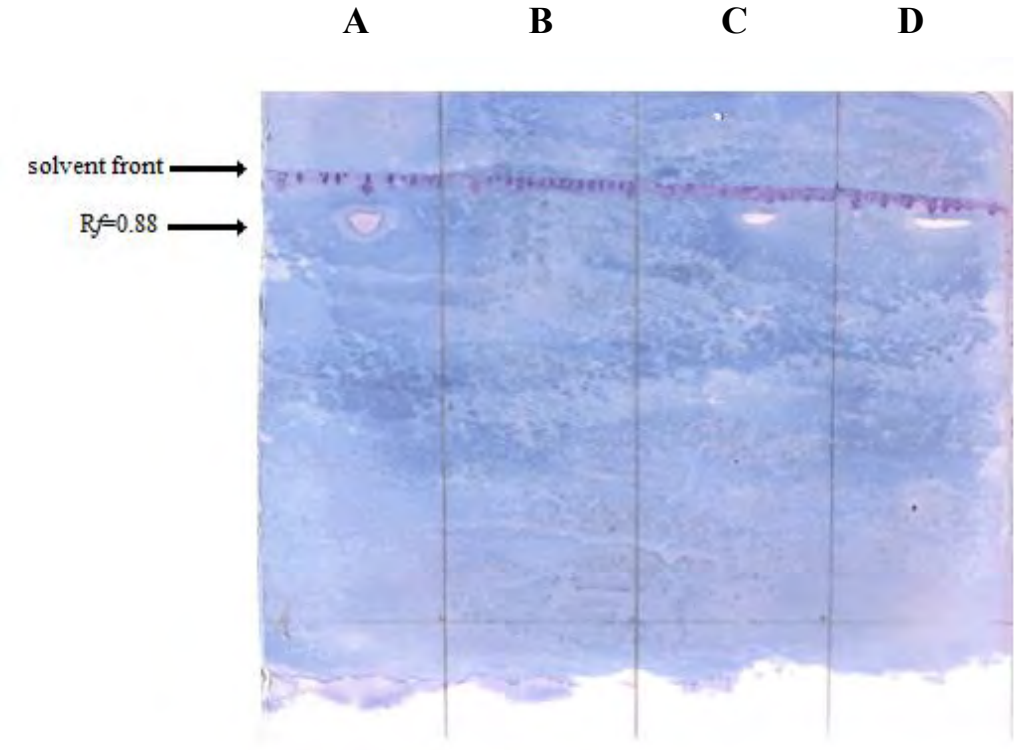

Figure 4.18 TLC plates stained with MTT showing the bioautographic analysis of DPO isolated from lane B) *S. polyantibioticus* ΔACY, lane C) *S. polyantibioticus* SPR^T and lane D) *S. polyantibioticus* PJNACY, where lane A) is chemically-synthesized DPO acting as a positive control.

4.5 CONCLUSION

An effective intergeneric conjugation method for the introduction of plasmid DNA from *E. coli* ET12567 into *S. polyantibioticus* SPR^T was developed and optimized. This method of transformation allowed for the identification and targeted manipulation of the biosynthetic pathway involved in DPO production. *S. polyantibioticus* mutant strains lacking a functional copy of the putative Cy domain encoded by gene SPR_53040, the A domain encoded by gene SPR_53060 and the acyl-CoA synthetase encoded by gene SPR_52860 exhibited a loss in the production of DPO and were thereby identified as putative members of the DPO biosynthetic pathway.

Due to the fact a gene encoding an acyl-CoA synthetase was not part of the initial hypothesis on how DPO is produced, it was deemed necessary to amend the original theory. The introduction of the revised hypothesis on how DPO is produced by *S. polyantibioticus* SPR^T, including an hypothesis on how benzoic acid is synthesized, is described in the following chapter.

4.6 REFERENCE LIST

Bailey, C.R. & Winstanley, D.J. (1986). Inhibition of restriction in *Streptomyces clavuligerus* by heat treatment. *Journal of General Microbiology*, **132:** 2945-2947.

Begg, K.J., Dewar, S.J. & Donachie, W.D. (1995). A new *Escherichia coli* cell division gene, *ftsK*. *Journal of Bacteriology,* **177**: 6211–6222.

Bibb, M.J., Ward, J.M. & Hopwood, D.A. (1978). Transformation of plasmid DNA into *Streptomyces* at high frequency. *Nature*, **274:** 398-400.

Bierman, M., Logan, R., O' Brien, K., Seno, E.T., Rao, R.N. & Schoner, B.E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene*, **116:** 43-49.

Bourn, W.R., Jansen, Y., Stutz, H., Warren, R.M., Williamson, A. & van Helden, P.D. (2007). Creation and characterisation of a high-copy-number version of the pAL5000 mycobacterial replicon. *Tuberculosis*, **87:** 481-488.

Choi, S.U., Lee, C.K., Hwang, Y.I., Kinoshita, H. & Nihira, T. (2004). Intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to *Kitasatospora setae*, a bafilomycin B1 producer. *Archives of Microbiology*, **181:** 294–298.

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

Du, L., Liu, R. & Zhao, G. (2012). An efficient intergeneric conjugation of DNA from *Escherichia coli* to mycelia of the lincomycin-producer *Streptomyces lincolnensis*. *International Journal of Molecular Sciences*, **13:** 4797-4806.

Flett, F., Mersinias, V. & Smith, C.P. (1997). High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS Microbiology Letters*, **155:** 223–229.

Frost, L.S., Ippen-Ihler, K. & Skurray, R.A. (1994). Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiological reviews*, **58(2):** 162-216.

Giebelhaus, L., Frost, L., Lanka, E., Gormley, E.P., Davies, J.E. & Leskiw, B. (1996). The Tra2 core of the InPα plasmid RP4 is required for intergeneric mating between Escherichia coli and Streptomyces lividans. *Journal of Bacteriology*, **178(21):** 6378-6381.

Gormley, E.P. & Davies, J. (1991). Transfer of plasmid RSF1010 by conjugation from Escherichia coli to Streptomyces lividans and Mycobacterium smegmatis. *Journal of Bacteriology*, **173:** 6705-6708.

Grohmann, E., Muth, G. & Espinosa, M. (2003). Conjugative plasmid transfer in gram-positive bacteria. *Microbiology and Molecular Biology Reviews*, **67(2):** 277-301.

Guan, D. & Pettis, G.S. (2009). Intergeneric conjugal gene transfer from *Escherichia coli* to the sweet potato pathogen *Streptomyces ipomoeae*. *Letters in Applied Microbiology*, 49: 67–72.

Hacène, H. & Lefebvre, G. (1995). AH17, a new non-polyenic antifungal antibiotic produced by a strain of *Spirillospora*. *Microbios,* **83:** 199 - 205.

Heinzelmann E., Berger, S., Puk, O., Reichenstein, B., Wohlleben, W. & Schwartz, D. (2003). A glutamate mutase is involved in the biosynthesis of the lipopeptide antibiotic friulimicin in Actinoplanes friuliensis. *Antimicrobial Agents and Chemotherapy*, **47(2):** 447-457.

Holo, H. & Nes, I.F. (1989). High-Frequency Transformation, by Electroporation, of *Lactococcus lactis*subsp. *cremoris* Grown with Glycine in Osmotically Stabilized Media. *Applied Environmental Microbiology*, **55(12):** 3119–3123.

Hopwood, D.A. (1989). Antibiotics: opportunities for genetic manipulation. *Philosophical Transactions of the Royal Society B*, **324:** 549-562.

Hopwood, D.A., Kieser, T. (1993). Conjugative plasmids of *Streptomyces*. In *Bacterial Conjugation*. Clewell, D.B. (ed.). New York: Plenum Press, 293–311.

Hussain, H.A., Ritchie, D.A. (1991). High frequency transformation of *Streptomyces niveus* protoplasts by plasmid DNA. *Journal of Applied Bacteriology*, **71:** 422-427.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K. & Hopwood, D.A. (2000). Practical *Streptomyces* Genetics. The John Innes Foundation: Norwich, UK.

Kieser, T. & Hopwood, D.A. (1991). Genetic manipulation of *Streptomyces*: integrating vectors and gene replacement. *Methods in Enzymology*, **204:** 430–458.

Kim, M.K., Ha, H.S. & Choi, S.U. (2008). Conjugal transfer using the bacteriophage phi C31 *att*/*int* system and properties of the *attB* site in *Streptomyces ambofaciens*. *Biotechnology Letters*, **30:** 695–699.

Kitani, S., Bibb, M.J. & Nihira, T. (2000). Conjugal transfer of plasmid DNA from *Escherichia coli* to *Streptomyces lavendulae* FRI-5. *Journal of Microbiology and Biotechnology*, **10:** 535–538.

Lacks, S., & Greenberg, B. (1977). Complementary specificity of restriction endonucleases of *Diplococcus pneumoniae* with respect to DNA methylation. *Journal of Molecular Biology*, **114:** 153-168.

Lanka, E. & Wilkins, B.M. (1995). DNA processing reactions in bacterial conjugation. *Annual Review of Biochemistry*, **64**: 141-169.

Lederberg, J. & Tatum, E.L. (1946). Gene recombination in *Escherichia coli*. *Nature,* **158:** 558.

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

Luo, J.M., Li, J.S., Wang, Y.T., Luo, S.H. & Wang, M. (2009). Protoplast Formation and Regeneration Conditions of *Streptomyces gilvosporeus*. In *Proceedings of 2009 3rd International Conference on Bioinformatics and Biomedical Engineering*, Beijing, China, 11–13 June 2009; IEEE: Piscataway, NJ, USA: 1647–1650.

Ma, Z., Liu, J., Shentu, X., Bian, Y. & Yu, X. (2013). Optimization of electroporation conditions for toyocamycin producer *Streptomyces diastatochromogenes* 1628. *Journal of Basic Microbiology*, **00:** 1–7.

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

MacNeil, D.J. (1988). Characterization of a Unique Methyl-Specific Restriction System in *Streptomyces avermitilis*. *Journal of Bacteriology*, **170(12):** 5607-5612.

Marcone, G.L., Carrano, L., Marinelli, F. & Beltrametti, F. (2010). Protoplast preparation and reversion to the normal filamentous growth in antibiotic-producing uncommon actinomycetes. *The Journal of Antibiotics*, **63:** 83-88.

Matshushima, P. & Baltz, R.H. (1985). Efficient plasmid transformation of *Streptomyces ambofaciens* and *Streptomyces fradiae* protoplasts. *Journal of Bacteriology*, **163:** 180-185.

Matshushima, P., Broughton, M.C., Turner, J.R. & Baltz, R.H. (1994). Conjugal transfer of cosmid DNA from *Escherichia coli* to *Saccharopolyspora spinosa*: effects of chromosomal insertions on macrolide A83543 production. *Gene*, **146:** 39-45.

Mazodier, P. & Davies, J. (1991). Gene transfer between distantly related bacteria. *Annual Review of Genetics*, **25:** 147-171.

Mazodier, P., Petter, R. & Thompson, C. (1989). Intergeneric conjugation between *Escherichia coli* and *Streptomyces* species. *Journal of Bacteriology*, **171**: 3583-3585.

Mazy-Servais, C., Backowski, D. & Dusart, J. (1997). Electroporation of intact cells of *Streptomyces parvulus* and *Streptomyces vinaceus*. *FEMS Microbiology letters*, **151:** 135-138.

McDonald, I.R., Riley, P.W., Sharp, R.J. & McCarthy, A.J. (1995). Factors affecting the electroporation of Bacillus subtilis. *Journal of Applied Bacteriology*, 79(2): 213-218.

Nikodinovic, J. & Priestley, N.D. (2006). A second generation *snp*-derived Escherichia coli-Streptomyces shuttle expression vector that is generally transferable by conjugation. *Plasmid*, **56:** 223-227.

Okanishi, M., Suzuki, K. & Umezawa, H. (1974). Formation and reversion of streptomycete protoplasts: cultural conditions and morphological study. *Journal of General Microbiology*, **80:**389-400.

Paget, M. S. B., Chamberlin, L., Atrih, A., Foster, S. J. & Buttner, M. J. (1999). Evidence that the extracytoplasmic function sigma factor sigma E is required 95 for normal cell wall structure in *Streptomyces coelicolor* A3(2). *Journal of Bacteriology*, **96(181):** 204-211.

Papagianni, M., **Avramidis, N. & Filioussis, G.. (2007).** High efficiency electrotransformation of *Lactococcus lactis*spp. *lactis* cells pretreated with lithium acetate and dithiothreitol. *BMC Biotechnology*, **7:** 15-23.

Paranthaman, S. & Dharmalingam, K. (2003). Intergeneric conjugation in *Streptomyces peucetius* and *Streptomyces* sp. strain C5: Chromosomal integration and expression of recombinant plasmids carrying the *chiC* gene. *Applied Environmental Microbiology*, **69:** 84–91.

Pigac, J. & Schrempf, H. (1995). A simple and rapid method of transformation of *Streptomyces rimosus* R6 and other streptomycetes by electroporation. *Applied Environmental Microbiology*, **61:**352-356.

Possoz, C., Ribard, C., Gagnat, J., Pernodet, J.L. & Guerineau, M. (2001). The integrative element pSAM2 from *Streptomyces*: kinetics and mode of conjugal transfer. *Molecular Microbiology,* **42:** 159–166.

Raleigh, E. A., & Wilson, G. (1986). *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. *Proceedings of the National Academy of Sciences USA*, **83:** 9070-9074.

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

Sanger, F., Nicklen, S. & Coulson, A., R. (1977). DNA Sequencing with chain-terminating inhibitors. *Biotechnology*. **24:** 104-108.

Shigekawa, K. & Dower, W.J. (1988). Electroporation of eukaryotes and prokaryotes: a general approach to the introduction of macromolecules into cells. *Biotechniques*, **6(8):**742-751.

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

Simon, R., Priefer, U. & Puhler, A. (1983). A broad host range mobilisation system for in vivo genetic engineering: transposon mutagenesis in Gram-negative bacteria. Biology/Technology, **1:** 784-791.

Spath, K., Heinl, S. & Grabherr, R. (2012). Direct cloning in *Lactobacillus plantarum*: Electroporation with non-methylated plasmid DNA enhances transformation efficiency and makes shuttle vectors obsolete. *Microbial Cell Factories*, **11**:141-147.

Stegmann, E., Pelzer, S., Wilken, K. & Wohlleben, W. (2001). Development of three different gene cloning systems for genetic investigation of the new species *Amycolatopsis japonicum* MG417-CF17, the ethylenediaminedisuccinic acid producer. *Journal of Biotechnology*, **92:** 195-204.

Sunairi, M., Iwabuchi, N., Murakami, K., Watanabe, F., Ogawa, Y., Murooka, H. & Nakajima, M. (1996). Effect of penicillin G on the electroporation of *Rhodococcus rhodochrous* CF222. *Letters in Applied Microbiology*, **22(1):** 66-69.

Trieu-Cuot, P., Carlier, C., Martin, P. & Courvalin, P. (1987). Plasmid transfer by conjugation from *Escherichia coli* to Gram-positive bacteria. *FEMS Microbiology Letters*, **48:** 289-294.

Tyurin, M., Starodubtseva, L., Kudryavtseva, H., Voeykova, T. & Livshits, V. (1995). Electrotransformation of germinating spores of *Streptomyces* spp. *Biotechnology techniques*, **9:** 737-740.

Vasu, K. & Nagaraja, V. (2013). Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. *Microbiology and Molecular Biology Reviews*, **77(1):** 53-72.

Voeikova, T.A. (1999). The conjugal transfer of plasmids from *Escherichia coli* to various strains of the order Actinomycetales. *Russian Journal of Genetics,* **35(12):** 1398-1404.

Vollmer, W. (2008). Structural variation in the glycan strands of bacterial peptidoglycan. *FEMS microbiology reviews*, **32(2):** 287-306.

Wilkinson, C.J., Hughes-Thomas, Z.A., Martin, C.J., Bohm, I., Mironenko, T., Deacon, M., Wheatcroft, M., Wirtz, G., Staunton, J. & Leadlay, P.F. (2002). Increasing the Efficiency of Heterologous Promoters in Actinomycetes. *Journal of Molecular Microbiology and Biotechnology*, **4(4):** 417–426.

Willetts, N. & Wilkins, B.M (1984). Processing of plasmid DNA during bacterial conjugation. *Microbiology Review*, **48(1):** 24-41.

Wu, L.J., Lewis, P.J., Allmansberger, R., Hauser, P.M. & Errington, J. (1995). A conjugation-like mechanism for prespore chromosome partitioning during sporulation in *Bacillus subtilis*. *Genes Development,* **9**: 1316–1326.

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.

Zhang, H.Z., Schmidt, H. & Piepersberg, W. (1992). Molecular cloning and characterization of two lincomycin-resistance genes, lmrA and lmrB, from *Streptomyces lincolnensls* 78-11. *Molecular Microbiology*, **6:** 2147–2157.

Zhang, P., Xu, P., Wang, J., Xiong, J. & Li, Y. (2013). Combined treatment with the antibiotics kanamycin and streptomycin promotes the conjugation of *Escherichia coli. FEMS Microbiology Letters*, **348(2):** 149-156.

CHAPTER 5

GENERAL DISCUSSION

CHAPTER 5

GENERAL DISCUSSION

CHAPTER 5

GENERAL DISCUSSION

5.1 Introduction

Natural products, such as bacterial secondary metabolites, have proven to be valuable sources of novel compounds exhibiting a diverse array of biological activities. Indeed, due to their diverse and complex chemical structures, bacterial secondary metabolites have demonstrated their use as pharmacologically active compounds that continue to serve medicine well by offering novel leads for novel therapeutic agents (Harvey, 1999; Bernan *et al.,* 1997). Additionally, the great structural and chemical diversity of natural products makes them well-suited for manipulation by chemical and/or genetic means to generate semi-synthetic drugs with improved antimicrobial and pharmacological properties (Ashforth *et al*., 2010). The first-line anti-TB drug, rifampicin, for example, was developed from the rifamycins, which are produced by members of the actinobacterial genus, *Amycolatopsis* (Wright, 2012).

The need for novel antibiotics with novel mechanisms of action has become urgent with the increased occurrence of bacterial resistance to known antibiotics and the emergence of new and old diseases (especially in the case of tuberculosis) (Bérdy, 2005).

It has been estimated that only 1–3 % of all antibiotics have been discovered, meaning that there is a high likelihood of identifying new antibiotic molecules from bacteria (and particularly actinobacteria) through a taxonomy-guided exploration of bacterial biodiversity. Certainly, the bacteria belonging to the suborders *Streptomycineae* (e.g. *Streptomyces*), *Micromonosporineae* (e.g. *Micromonospora*), *Pseudonocardineae* (e.g. *Amycolatopsis*) and *Streptosporangineae* (e.g. *Planobispora*) (all in the order *Actinomycetales*; class *Actinobacteria*) are the most prolific producers of novel antimicrobial antibiotics (Ashforth *et al*., 2010). To gain access to the biodiversity in these (and other) actinobacterial suborders, it is important to isolate

actinobacteria from many different environments, e.g. terrestrial soil and marine sediments, as well as from plants and animals.

Actinobacterial genomes (particularly those of the actinomycetes, such as *Streptomyces*) are larger than those of most other bacteria, possessing a high capacity for the synthesis of secondary metabolites and employing a substantial fraction of coding capacity (5–10 %) for the production of mostly cryptic secondary metabolites (Baltz, 2008). The sequenced actinobacterial genomes are a valuable resource, which have revealed that actinobacteria have the biosynthetic potential to make far more natural products than was realised (Ashforth *et al.,* 2010). The taxonomic diversity covered by these sequenced genomes has revealed that the more bacterial genomes that are sequenced, the more new gene families are discovered (Wu *et al.,* 2009). Thus, identifying taxonomic diversity reveals new genetic diversity and therefore new biosynthetic capabilities.

In light of the taxonomy-guided bacterial bioprospecting approach described above, *S. polyantibioticus* SPR^T was isolated as part of an antibiotic-screening programme and identified as the producer of several antibiotics (Le Roes-Hill & Meyers, 2009; Le Roes, 2005). One of these antibiotics possessed antitubercular activity and was identified as DPO (Le Roes, 2005). In this study, the *S. polyantibioticus* SPR^T genome was sequenced and explored in order to identify the biosynthetic pathway involved in DPO biosynthesis and thereby take the first step towards the generation of derivatives of DPO by combinatorial biosynthesis. Combinatorial biosynthesis is a powerful way to generate antibiotic derivatives with greater antibacterial activity and/or improved pharmacokinetic properties. It has advantages over the chemical derivatisation of antibiotic molecules, since the bacterial biosynthetic enzymes carry out site-specific and enantiomer-specific reactions, by-passing the need to protect reactive functional groups in each reaction step in the chemical approach. This ensures that only the desired product is produced, which enhances the product yield.

In order to determine whether the proposed biosynthetic pathway is correct, the genes coding for benzoic acid synthesis and the DPO NRPS had to be identified in the *S. polyantibioticus* SPR^T genome. The research aims of this study, as described in Chapter 1, are discussed in the following sections within the context of the original hypothesis for DPO biosynthesis in *S. polyantibioticus* SPR^T.

5.2 Identification and characterisation of the genes involved in joining benzoic acid and phenylalanine/3-hydroxyphenylalanine in *S. polyantibioticus* SPR^T to create a 2,5-disubstituted oxazole

Based on the structure of DPO, a biosynthetic scheme for the synthesis of this molecule, in which an NRPS condenses a molecule of benzoic acid with 3-hydroxyphenylalanine (also known as 3-phenylserine) to form an amide bond between them, was proposed (Chapter 1). This intermediate, known as N-benzoyl-β-hydroxyphenylalanine (Figure 5.1), is subsequently converted to a diphenyloxazole derivative by heterocyclization across the amide bond, after which a final decarboxylation step leads to DPO. An NRPS is believed to be the core enzyme in the synthesis of DPO due to the presence of an oxazole in the DPO structure.

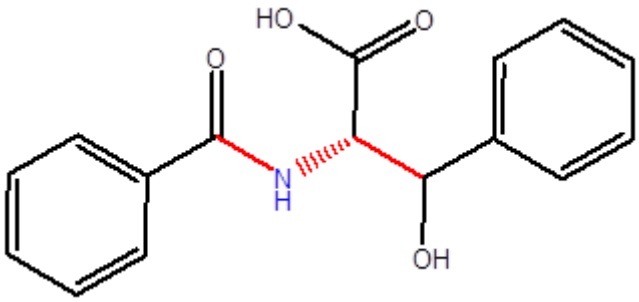

Figure 5.1 The proposed amide-bond-containing precursor to DPO, N-benzoyl-β-hydroxyphenylalanine. The atoms joined by the three red bonds (i.e. running from the carbonyl C atom of benzoic acid to the hydroxyl C atom of 3-hydroxyphenylalanine) are involved in forming a 5-membered heterocyclic ring, i.e. an oxazole.

Initial efforts to isolate and identify the DPO biosynthetic gene cluster in *S. polyantibioticus* SPR^T identified twelve unique NRPS A domains. The specificities of these A domains were identified using NRPSpredictor2 (Rausch et al., 2005). This approach has been used in numerous studies and is considered accurate. However, there have been examples of substrate-

specificity codes for which the predicted substrate specificity does not correspond to the activated amino acid identified *in vivo* (Lombó *et al.,* 2006). A study by McQuade *et al.* (2009) suggested that, in addition to the 10 key residues of the binding pocket, there may be other residues involved in controlling the substrate amino acid activated by an A domain.

In this study, the A domain contained within the inserts of the identical clones pGEMA-18 and pGEMA-66 was predicted to be specific for the activation of phenylalanine, while two A domains contained within the inserts of the clones pGEMA-5/pGEMA-7/pGEMA-12/pGEMA-22 and pGEMA-36/pGEMA-51/pGEMA-63, where predicted to be specific for serine. Serine is a common starter unit for the biosynthesis of oxazoles. However, if serine were used in DPO biosynthesis, it would require the addition of a phenyl group to the oxazole structure (Roy *et al.,* 1999). Importantly, β-phenylation of an activated serine has not yet been reported to occur in bacteria and it therefore seems more likely that β-hydroxylation of an activated phenylalanine residue by a P450 monooxygenase would be the mechanism involved in DPO biosynthesis. Nevertheless, it is possible that a β-phenylation reaction *is* used by *S. polyantibioticus* SPR^T (Le Roes, 2005).

Annotation of the *S. polyantibioticus* SPR^T draft genome sequence revealed that the A domain insert in clone pGEMA-18/pGEMA-66 is encoded by gene SPR_48330. This gene was predicted by antiSMASH 3.0 to be involved in the biosynthesis of an antimycin-type antibiotic and was therefore dismissed as being a member of the DPO biosynthetic cluster. Similarly, the A domains predicted to activate serine and contained within the clones pGEMA-5/pGEMA-7/pGEMA-12/pGEMA-22 and pGEMA-36/pGEMA-51/pGEMA-63, are encoded by the genes SPR_56140 and SPR_56150, respectively. Due to their predicted involvement in the biosynthesis of a complex PKS/NRPS hybrid secondary metabolite, genes SPR_56140 and SPR_56150 were deemed unlikely to be involved in DPO biosynthesis. Nevertheless, the A domains designated A-7 and A-18 (Chapter 2), in addition to the AD-2, A-16 and A-28 domains encoded by the genes SPR_6340, SPR_28920 and SPR_6040, respectively, were disrupted by homologous recombination in *S. polyantibioticus* SPR^T. Subsequent bioautography experiments confirmed that genes SPR_6340, SPR_28920 and SPR_6040 are not involved in DPO biosynthesis, as inactivation of these genes did not abolish DPO production by the mutant strains.

Conversely, the A domain, A-99, encoded by SPR_53060 was confirmed as a putative member of the *S. polyantibioticus* SPR^T DPO biosynthetic pathway due to the fact that an *S. polyantibioticus* mutant strain lacking a functional copy of this A domain exhibited a loss of production of DPO. The specificity prediction of this A domain was however inconclusive, as the software tools predicted specificity for both phenylalanine and tryptophan (Chapter 3). Nevertheless, the relaxed substrate specificity of the A domains in xenematide A biosynthesis has demonstrated the ability of A domains for aromatic amino acids to accept either tryptophan or phenylalanine as substrates. Furthermore, Schaffer & Otten (2009) demonstrated in the tyrocidine A biosynthetic pathway that although A domains are selective for specific substrates, many of these domains possess the ability to adenylate a number of different substrates. Thus, the A domain encoded by SPR_53060 might accept phenylalanine or 3-hydroxyphenylalanine as a substrate in the biosynthesis of DPO.

However, if 3-hydroxyphenylalanine is not the starter molecule in DPO biosynthesis, it is possible that an alternative starter molecule such as 1-phenyl-2-aminoethanol could be used in DPO biosynthesis (Figure 5.2). Indeed, 2-aminoethanol (ethanolamine) is a compound derived from cell membranes that certain bacteria can utilize as a source of carbon and/or nitrogen (Garsin, 2010). It is also possible that *S. polyantibioticus* SPR^T could utilize glycine to synthesize 2-aminoethanol by reduction of the carboxyl group. However, a phenylation reaction would need to occur to generate 1-phenyl-2-aminoethanol.

It is evident that, due to the weak predictions of the software tools and the contradictory results, the substrate specificity of the A domain encoded by SPR_53060 can only be elucidated by experimental analysis. Due to the fact that A domains consume ATP and release PPi during the activation of their substrates, and because the activation reaction is reversible, A domain selectivity can be assayed using the conventional ATP-^{32}PPi isotope exchange assay *in vitro*. During this assay, the A domain of interest is incubated with the potential substrate, ATP, Mg^{2+} and ^{32}P-labeled PPi. During the reverse activation reaction, ^{32}P from the labeled PPi is incorporated into ATP if the amino acid(s) present in the assay is/are activated by the A domain. The ^{32}P-ATP can then be separated via adsorption to activated charcoal. The amount of generated ^{32}P-ATP is proportional to the substrate activation reaction (Phelan *et al.,* 2009; Linne & Marahiel, 2004). However, Phelan *et al.* (2009) developed a non-radioactive mass spectrometry (MS)-based PPi exchange assay that is able to measure the consumption of

γ-$^{18}O_4$-labelled ATP and the formation of $^{16}O_4$-ATP by the isotopic back exchange of excess unlabelled PPi. The resulting mass shifts are detected by MALDI-TOF-MS and the activity of the A domain can therefore be measured. Incubation with different substrates and a comparison of each substrate's ATP-PPi exchange rate allows for the determination of the substrate specificity of the A domain under scrutiny. In contrast to the issues observed in the conventional systems, this exchange assay provides a rapid, sensitive and reproducible means to measure A domain specificity (Phelan et al., 2009). This technique has been used successfully to determine the A domain specificity in studies of *Streptomyces refuinius* (Phelan *et al.,* 2009) and *Streptomyces coeruleorubidus* NRRL 18370 (Zhang *et al.,* 2010), thereby providing an alternative option for the substrate prediction of A domains in future studies of *S. polyantibioticus* SPR^T.

Additionally, the A domain of the acyl-CoA synthetase encoded by SPR_52860 was predicted to be specific for the activation of an aromatic substrate such as 2,3-dihydroxybenzoic acid (or a derivative thereof), phenylalanine or tryptophan (Chapter 3). This gene was identified as a member of the DPO biosynthetic pathway via the gene disruption and complementation experiments described in Chapter 4.

The family of coenzyme A ligases includes the A domains of NRPSs, firefly luciferase and acyl- and aryl-CoA synthetases, such as the acyl-CoA synthetase encoded by SPR_52860. Reactions catalysed by this family of enzymes proceed in two steps in which the substrate reacts with ATP to form an acyl-adenylate intermediate with the simultaneous release of pyrophosphate. Thereafter, the adenylate group is replaced by CoA, accompanied by the release of AMP (Gulick, 2009). The CoA-ligases are known to initiate β-oxidation via the activation of fatty acids and are important in the biosynthesis of natural products such as the penicillins (Koetsier *et al.,* 2009) and lignin (Anterola & Lewis, 2002) (Figure 5.3).

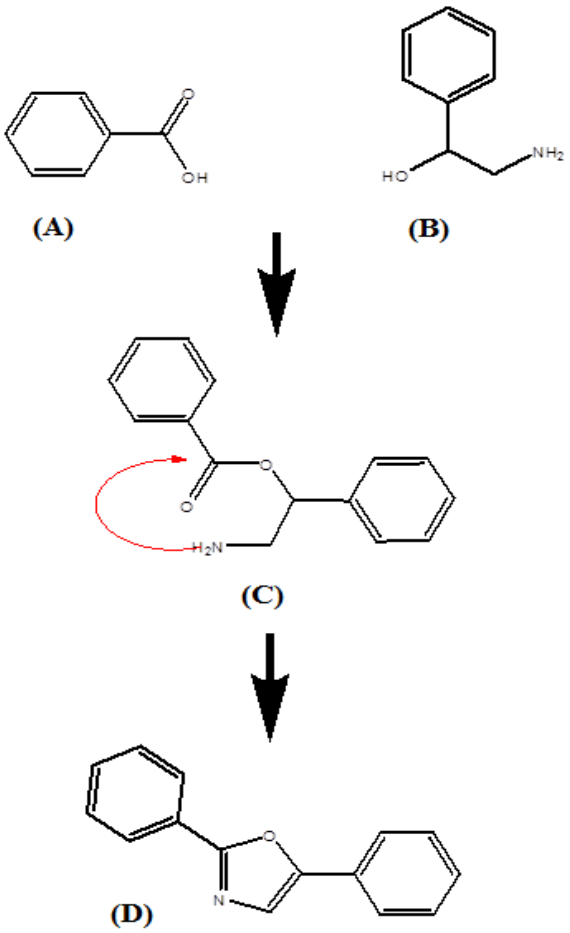

Figure 5.2 Proposed alternative route to DPO biosynthesis in *S. polyantibioticus* SPR^{T}. A molecule of benzoic acid (A) is condensed with 1-phenyl-2-aminoethanol (B) to form an ester bond between them. Subsequent heterocyclization and decarboxylation of the intermediate, *O*-benzoyl-1-phenyl-2-aminoethanol (C), would lead to DPO (D).

As the acyl-CoA synthetase encoded by SPR_52860 was identified as a member of the DPO biosynthetic pathway, it would activate its substrate, benzoic acid, as a CoA-thioester. Thereafter, the first reaction in DPO biosynthesis would proceed via amide bond formation between an aminoacyl-bound PCP-thioester and the CoA-thioester. Although the activation of a Cy domain substrate as a CoA-thioester instead of a PCP-Ppant-thioester is unusual, it has been reported in the biosynthesis of congocidine (Juguet *et al.,* 2009), the calcium-dependent antibiotics (Hojati *et al.,* 2002) and fengycin (Steller *et al.,* 1999). It is interesting to note that the aspartate residue (D_{235} according to the GrsA-PheA numbering convention) that is

conserved in the majority of A domains and which is involved in the binding of the α-amino acid substrate is replaced by a glycine in the substrate specificity binding pocket of the A domain encoded by SPR_52860. It is therefore suggested that the substrate of the SPR_52860 A domain does not possess an α-amino group, which would support the hypothesis that an aromatic acid, such as benzoic acid, is indeed activated.

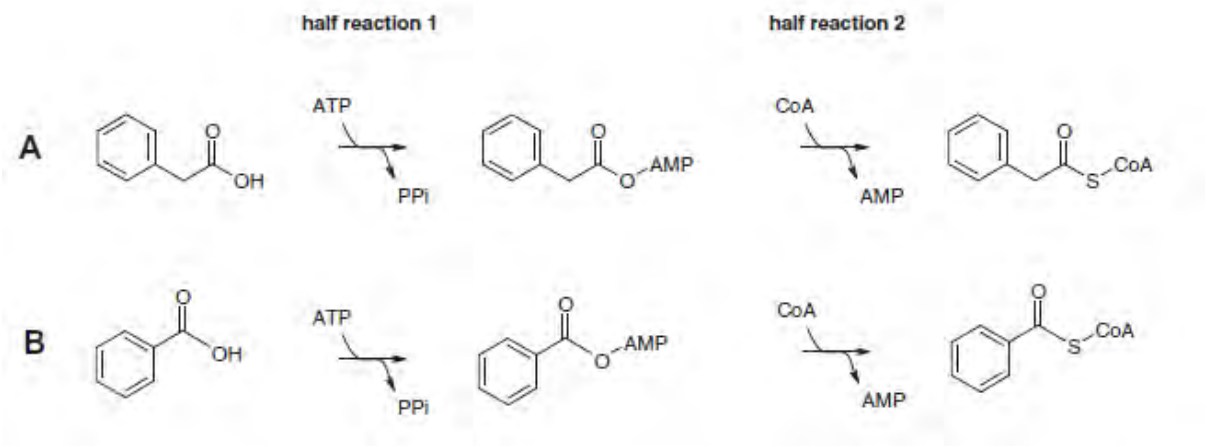

Figure 5.3 Examples of reactions catalysed by members of the family of adenylate-forming enzymes. (A) The activation of the pencillin G side chain phenylacetic acid is catalysed by a phenylacetate-CoA ligase in *Penicillium chrosygenum* (adapted from Koetsier *et al.,* 2011). (B) The activation of benzoic acid for DPO biosynthesis is believed to be catalysed by an acyl CoA-synthetase in *S. polyantibioticus* SPR^{T}.

Although, in the initial experiments described in Chapter 2, a Cy domain could not be amplified from the *S. polyantibioticus* SPR^{T} genome, a phylogenetic analysis of all of the identified C domains within the *S. polyantibioticus* SPR^{T} draft genome provided insight into the degree of evolutionary relatedness of these domains to well-characterised reference Cy domains. The domains encoded by SPR_53060, SPR_53040, SPR_6230, SPR_6240 and SPR_6360 shared the highest degree of evolutionary relatedness to the reference Cy domains and an analysis of the conserved signature motifs within these domains identified SPR_53040 and SPR_53060 as displaying an unusual putative tandem condensation/heterocyclization function in DPO biosynthesis. This was deduced because SPR_53040 lacks the catalytic second aspartate

residue found in the Cy domain consensus motif, whereas SPR_53060 lacks the first Cy domain aspartate residue.

It has been demonstrated that the second aspartate residue in the Cy domain motif D-x-x-x-x-D-x-x-S motif is critical for catalytic Cy activity, however the role of the first aspartate is still unclear. Mutation of the first aspartate residue to alanine resulted in the elimination of the activity in the first Cy domain of the yersiniabactin synthetase subunit HMWP2, while a comparable mutation in the Cy domain of the vibriobactin NRPS module VibF resulted in only a 10-fold reduction in activity (Marshall *et al.,* 2002; Keating *et al.,* 2000). The significance of the first aspartate residue of the consensus motif was also examined via an alanine mutation in the NRPS subunit EntF of enterobactin synthetase, where activity was only diminished 2-fold in comparison to the wild type strain. The authors surmised that this result implicates the residue as critical but not essential for heterocycle formation, with its importance varying according to the synthetase and substrate in question (Kelly *et al.,* 2005).

The serine residue found in the D-x-x-x-x-D-x-x-S motif is not strictly conserved among Cy domains and a mutation of the serine residue to alanine in the yersiniabactin synthetase subunit HMWP2 Cy domain resulted in no effect on heterocycle formation. This suggests that the lack of a serine residue in the signature motif found in the putative Cy domain encoded by SPR_53040 is of limited significance (Keating *et al*., 2000).

Peptide bond formation and heterocyclization usually depend on the action of a single Cy domain. However, a situation where these two reactions are split between two Cy domains was observed in the biosynthesis of the oxazoline-containing siderophore vibriobactin in *Vibrio cholerae.* Such a split-function Cy-domain pair is believed to be involved in the DPO biosynthetic pathway. The NRPS responsible for the formation of vibriobactin consists of an unusual Cy-Cy tandem arrangement, where the first domain is responsible for heterocyclization and the second domain is responsible for condensation (Schwarzer *et al.,* 2003; Marshall *et al.,* 2002).

Additionally, in the biosynthesis of holomycin by *Streptomyces clavuligerus*, a tridomain module with the organization Cy-A-PCP, together with a lone C domain, are proposed to be responsible for the selection, activation, heterocyclization and oxidation of two cysteine

residues in the biosynthesis of the final product. This further demonstrates the flexibility of NRPS systems, where instead of a Cy domain simply replacing the C domain as observed in conventional NRPS elongation modules for oxazoles and thiazoles, a pair of domains functions in tandem (Li & Walsh, 2010).

5.3 Identification of the genes responsible for the biosynthesis of benzoic acid in *S. polyantibioticus* SPR^T

In bacteria, only a handful of benzoic acid derived metabolites have been described, seemingly due to the rarity of PAL in these organisms. Indeed, despite the plant-like biosynthesis of benzoyl-CoA described in "*S. maritmus*" strain DSM 41777^T, where this rare starter unit is involved in the production of the structurally novel polyketide enterocin (Hertweck & Moore, 2000), unsubstituted benzoyl units have only been found in benzoyl α-L-rhamnopyranoside produced by *Streptomyces griseoviridis* Tü 3634 (Hoffman & Grond, 2004), as well as in the structures of aestivophoenin A (Kunigami *et al.,* 1998) and wailupemycin (Hertweck & Moore, 2000), all isolated from *Streptomyces* spp., and in the myxobacterial molecules soraphen (Hill *et al*., 2003), thiangazole and crocacin (Jansen *et al.,* 1992). In stark contrast to this, the occurrence of benzoyl units is common in secondary metabolites produced by plants and fungi, such as in the cancer drug taxol (Rohr, 1997) and the fungicide strobilurin (Nerud *et al.,* 1982).

In addition to the aerobic benzoyl-CoA biosynthetic process catalysed by PAL in "*S. maritmus*" strain DSM 41777^T, there is evidence for two PAL-independent pathways resulting in the biosynthesis of benzoic acid, namely, via: 1) the anaerobic degradation of L-phenylalanine described in the denitrifying bacterium *Thauera aromatica* (Breese *et al.,* 1998; Schneider *et al.,* 1997) and 2) directly from shikimate (Hoffman & Grond, 2004). Additionally, soraphen A biosynthesis in *S. cellulosum* is an example of an aerobic non-PAL pathway, which begins with a phenyl side group that is derived from phenylalanine, but the exact pathway for the generation of benzoic acid is unknown. It has been proposed that carboxybenzoyl-CoA is an intermediate in the biosynthesis of benzoic acid in *S. cellulosum* (Schupp *et al*., 1995, Ligon *et al*., 2002).

In this study, a number of approaches to obtaining the genes responsible for the biosynthesis of benzoic acid in *S. polyantibioticus* SPR^T were explored. The first approach focused on the search for an orthologue of the *encP* gene, which encodes PAL. Despite utilizing '*S. maritimus*' DSM 41777^T as a positive control, no *encP* orthologue was detected in the *S. polyantibioticus* SPR^T genome by PCR amplification and Southern blotting. However, a HAL, which is highly homologous to PAL and catalyzes the analogous deamination of histidine to *trans*-urocanate in prokaryotes, thereby initiating histidine degradation (Michal, 1999), was detected in the *S. polyantibioticus* SPR^T genome using PCR amplification and sequencing. Both HAL and PAL enzymes are also homologous to tyrosine ammonia lyase (TAL), an enzyme commonly found in plants (Kyndt *et al.,* 2002). All three enzymes contain the unique prosthetic group 4-methylidene imidazol-5-one, which may indicate a similar catalytic mechanism (Bode & Müller, 2003; Rother *et al.,* 2001; Schwede *et al.,* 1999). Due to the fact that PAL is the first enzyme in the plant phenylpropanoid biosynthetic pathway, it has received the greatest attention and is therefore the best studied member of this enzyme family (Schuster & Retey, 1995). Since a PAL was not detected in the *S. polyantibioticus* SPR^T genome, an alternative pathway for the production of cinnamic acid was proposed whereby the need for a PAL was bypassed (Chapter 3). This pathway was proposed on the basis of the discovery of a cinnamate CoA ligase, encoded by SPR_60150, in the *S. polyantibioticus* SPR^T draft genome. SPR_60150 displayed significant homology to the enzyme, 4-coumarate-CoA ligase (4CL, EC: 6.2.1.12), which has been characterized to occur in the biosynthetic route to benzoyl-CoA in "*S. maritmus*" strain DSM 41777^T. In addition, 4CL, which catalyzes the final reaction of the phenylpropanoid pathway in plants, has also been characterized in *S. coelicolor* A3(2) (Kaneko *et al.,* 2003). Similarly to *S. polyantibioticus* SPR^T, a PAL was not detected in the *S. coelicolor* A3(2) genome. Therefore, if cinnamate or 4-coumarate are physiological substrates, they may be supplied from dead plants in the environment, although no uptake system for either of these compounds has been described in *Streptomyces* spp. (Kaneko *et al*., 2003).

Three genes encoding a putative acetyl/propionyl-CoA carboxylase, a putative acyl-CoA dehydrogenase and a putative enoyl-CoA hydratase exist in the immediate vicinity of the gene encoding the 4CL enzyme in the genome of *S. coelicolor* A3(2) and, due to the fact that functionally related genes are normally clustered in bacterial genomes, the authors speculated that the 4CL, together with the enzymes encoded by the three other genes, are involved in the

production of a secondary metabolite (Kaneko *et al.,* 2003). In light of this and the fact that a putative enoyl-CoA hydratase and a putative acyl-CoA dehydrogenase were found next to the gene encoding a cinnamate-CoA ligase in the *S. polyantibioticus* SPR^T genome, this cluster was identified as being putatively responsible for the biosynthesis of benzoic acid.

The proposed pathway involved the catabolism of phenylalanine to cinnamic acid via phenylpyruvate and phenyllactate (Chapter 3). The D-lactate dehydrogenase encoded by SPR_60260 and the cinnamate CoA ligase encoded by SPR_60150 were disrupted and *S. polyantibioticus* single mutant strains lacking a functional copy of each of these genes were able to produce DPO, thereby proving that they are not involved in the DPO biosynthetic pathway.

Thus, benzoic acid does not appear to be synthesized in *S. polyantibioticus* SPR^T via a PAL-dependant pathway or via another cinnamate-containing pathway in which the PAL reaction is bypassed. As studies have demonstrated that antibiotic production in actinomycetes and other microorganisms makes use of amino acids as precursors, such as in the biosynthesis of oleandomycin in *S. antibioticus*, the second approach to determining the genes responsible for benzoic acid biosynthesis focussed on a novel variation of the phenylacetate degradation pathway (Tang *et al.,* 1994). Phenylacetic acid (PA) is a common intermediate in the microbial metabolism of a variety of aromatic substrates including phenylalanine but, although there are examples of aerobic, non-PAL pathways in nature (often utilising PA as the starter unit), none of them is fully understood.

In the denitrifying bacterium, *Thauera aromatica*, PA is catabolized under anoxic conditions to benzoyl-CoA via the intermediates phenylacetyl-CoA and phenylglyoxylate (Rhee & Fuchs, 1999). This mechanism is similar to the pathway described in *A. evansii*, in which it is speculated that an aerobic pathway may coexist in order to fully metabolise PA. Moreover, an aerobic pathway for the degradation of PA has been established in Gram-negative bacteria such as *Pseudomonas putida* and *E. coli,* whereby PA is first converted to PA-CoA, which subsequently undergoes ring hydroxylation, hydrolytic-ring opening and further degradation. In contrast, this pathway and its role have not been comprehensively characterised in Gram-positive bacteria, although a study of *Rhodococcus* sp. strain RHA1 provided conclusive evidence that it does have a functional PA degradation pathway that is partially responsible for

the degradation of styrene, ethylbenzene and 3-hydroxyphenylacetate (Navarro-Llorens *et al*., 2005). However, the existence of benzoic acid or benzoyl-CoA as a final product in these aerobic pathways has yet to be established.

Furthermore, the PA catabolon has been described to be involved in the biosynthesis of the seconday metabolite, antimycin A, produced by *Streptomyces albus* S4 (Seipke & Hutchings, 2013) and in the biosynthetic pathway for the production of neoantimycin, part of the antimycin family of depsipeptides, by *Streptoverticillium orinoci* (Li *et al.,* 2013).

In light of the fact that the PA catabolon has been associated with the production of secondary metabolites and because a PA-CoA ligase is responsible for the catalysis of the first reaction in this pathway, catalysing the activation of PA to PA-CoA, a homologue of the gene encoding this enzyme, *paaK*, was amplified from the *S. polyantibioticus* SPR^{T} genome. However, an *S. polyantibioticus* mutant lacking a functional copy of this gene did not lose the ability to synthesize DPO, thereby suggesting that a PA-CoA ligase is not involved in DPO biosynthesis. However, due to the fact that six paralogous PA-CoA ligases (with an amino acid similarity of 24 %) were annotated in the *S. polyantibioticus* SPR^{T} genome, it is plausible that one or more of these enzymes may have suppressed the effect of the knockout of the *paaK* gene in *S. polyantibioticus* strain ΔPAAK, which thereby resulted in a level of DPO production that was comparable to the wild-type strain. Future work may necessitate the need to measure the exact level of DPO production in both the *S. polyantibioticus* ΔPAAK and wild-type strains, as a lower level of DPO production in *S. polyantibioticus* ΔPAAK could serve as circumstantial evidence that a PA-CoA paralogue functions less efficiently than the dedicated enzyme in the biosynthesis of DPO, thereby implicating the involvement of the enzyme encoded by SPR_46390 in the production of DPO. Indeed, Geukens *et al.* (2006) demonstrated that Type I signal peptidase (SPase) paralogous enzymes in *S. lividans* can only partly complement each other, as clear differences in SPase substrate specificity resulted in a dramatic depletion of preprotein processing and secretion.

The attempts made in this study to identify the genes responsible for the biosynthesis of benzoic acid in *S. polyantibioticus* SPR^{T} were based on characterised bacterial benzoic acid biosynthetic pathways, as well as characterised bacterial amino acid degradation pathways. Due to the fact that the *in silico* gene identification using a genome mining approach and

subsequent *in vivo* gene disruption experiments disproved both hypotheses on how benzoic acid is synthesized in *S. polyantibioticus* SPR^T, the genes encoding this pathway are still unidentified.

However, since the shikimate and chorismate pathways (Figure 5.4) are commonly used to generate molecules with benzene rings in bacteria (and other organisms), *S. polyantibioticus* SPR^T could perhaps use a novel variation on one of these aromatic biosynthetic pathways to generate benzoic acid. Indeed, intermediates from the shikimate pathway serve as starting points for the biosynthesis of secondary metabolites in bacteria and plants (Herrman & Weaver, 1999; Herrman, 1995). Eukaryotic benzoate biosynthesis also proceeds via phenylalanine derived from the shikimate pathway and, most importantly, a study by Hoffmann & Grond (2004) postulated a direct conversion of shikimic acid to benzoate in *Streptomyces griseoviridis* strain Tü 3634. In *S. griseoviridis* strain Tü 3634, the hypothetical mechanism of benzoate formation consists of the dephosphorylation of shikimic acid-3-phosphate and subsequent dehydration steps involving the enzymes 3-dehydroquinate dehydratase and shikimate dehydrogenase. Shikimate is converted to benzoate in an alternative microbial route that has not been described before and was confirmed by ^{13}C feeding experiments. This proved that the unsubstituted benzoyl ring in the final product originates directly from shikimate, contrary to the plant-like conversion of shikimate to benzoate via prephenate, phenylalanine and cinnamate, in which the carboxylic carbon of shikimate is lost in the latter stages (Figure 5.5) (Hoffmann & Grond, 2004). However, the enzymes that convert shikimate to benzoic acid in *S. griseoviridis* strain Tü 3634 have yet to be characterized.

It is possible that *S. polyantibioticus* SPR^T employs a similar pathway for the biosynthesis of benzoic acid directly from shikimate. As would be expected for a central metabolic pathway, the enzymes involved in the shikimate pathway are all present in the *S. polyantibioticus* SPR^T draft genome, namely 3-deoxy-D-*arabino*-heptulosonate 7-phosphate (DAHP) synthase encoded by SPR_10690 and SPR_58840, 3-dehydroquinate synthase encoded by SPR_41260, 3-dehydroquinate dehydratase encoded by SPR_41270 and shikimate dehydrogenase encoded by SPR_41230. A suggestion for future work would be to disrupt the genes encoding enzymes in the shikimate pathway (Figure 5.4) in *S. polyantibioticus* SPR^T, followed by assays to determine whether DPO is still produced. If an *S. polyantibioticus* SPR^T shikimate mutant was unable to produce DPO, it would imply that the shikimate pathway is required to provide a

precursor for DPO biosynthesis (i.e. benzoic acid). Once this had been established, further work would be required to identify how benzoic acid is derived from shikimate (or one of its precursors).

Another possible source of benzoic acid could be the mandelate metabolic pathway (Figure 5.6), which has been well characterized in *Pseudomonas putida* ATCC 12633 and allows various pseudomonads to utilize mandelate as a sole carbon source via the oxidative degradation of R-mandelate to benzoate (Tsou *et al.,* 1990). Indeed the existence of homologous enzymes from the *P. putida* ATCC 12633 mandelate catabolic pathway in the draft *S. polyantibioticus* SPR^T genome, including a mandelate racemase (encoded by SPR_50530), a putative mandelate dehydrogenase (encoded by SPR_47860), a putative benzoylformate decarboxylase (encoded by SPR_19760) and a putative benzaldehyde dehydrogenase (encoded by SPR_9230) could confer on *S. polyantibioticus* SPR^T the ability to synthesize benzoic acid for DPO biosynthesis. However, due to the low degree of homology shared between the *S. polyantibioticus* SPR^T enzymes and the *P. putida* ATCC 12633 enzymes involved in mandelate catabolism, and due to the fact that the genes for the enzymes identified as putative mandelate pathway members in *S. polyantibioticus* SPR^T are not clustered, it seems unlikely that this method of benzoic acid synthesis is employed by *S. polyantibioticus* SPR^T. Another problem with any proposed mandelate-to-benzoate pathway is that an external source of mandelate would be required and *S. polyantibioticus* SPR^T is able to produce DPO in a conventional bacterial growth medium from which mandelate is absent.

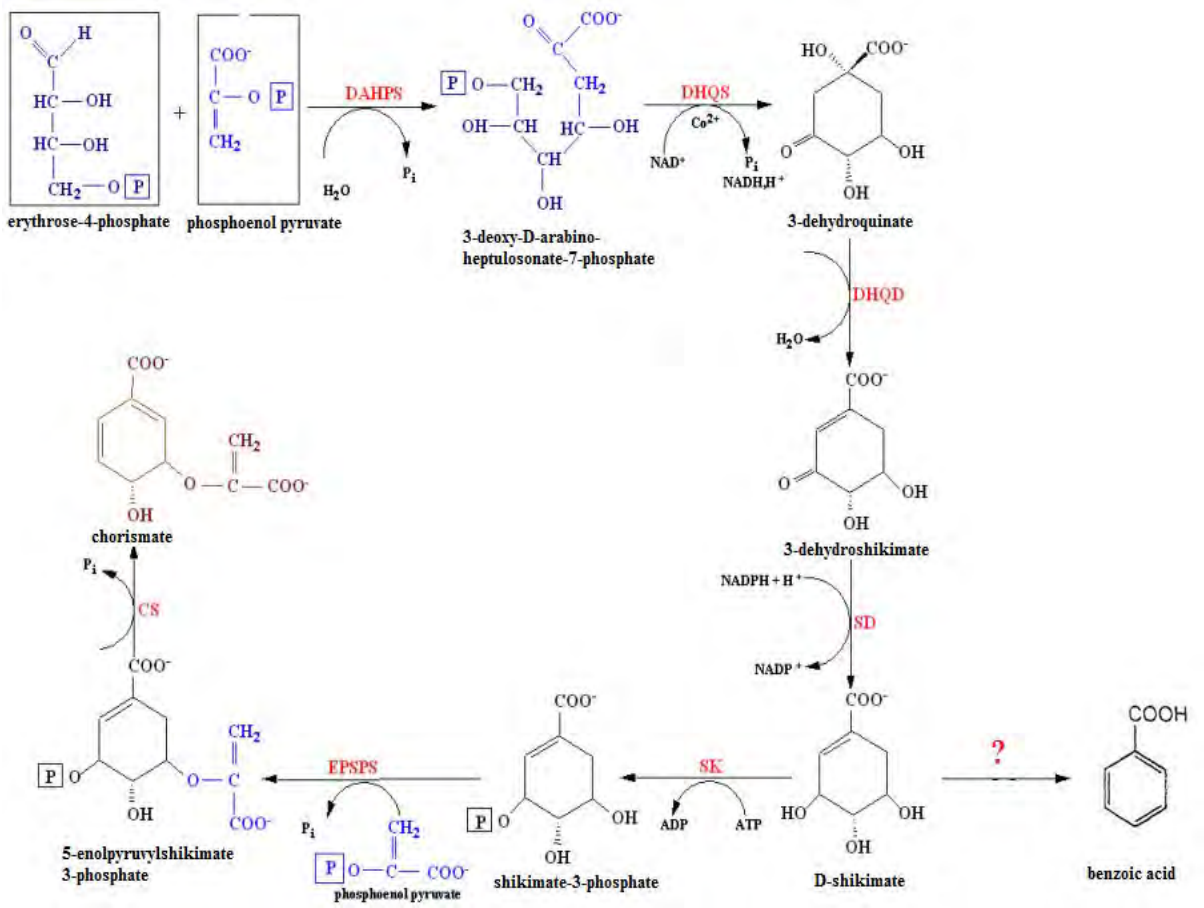

Figure 5.4 An illustration of the shikimate pathway found in microorganisms and plants. In a sequence of seven metabolic steps, phosphoenolpyruvate and erythrose 4-phosphate are converted to chorismate, the precursor of the aromatic amino acids and many aromatic secondary metabolites. All pathway intermediates can also be considered branch point compounds that may serve as substrates for other metabolic pathways. The key enzymes involved in the pathway are abbreviated in red: 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase (DAHPS), 3-dehydroquinate synthase (DHQS), 3-dehydroquinate dehydratase (DHQD), shikimate dehydrogenase (SD), shikimate kinase (SK), 5-enolpyruvylshikimate 3-phosphate synthase (EPSPS) and chorismate synthase (CS). The direct conversion of shikimate to benzoic acid, a branch point of the shikimate pathway, was postulated by Hoffman & Grond (2004) and is catalysed by unknown enzymes.

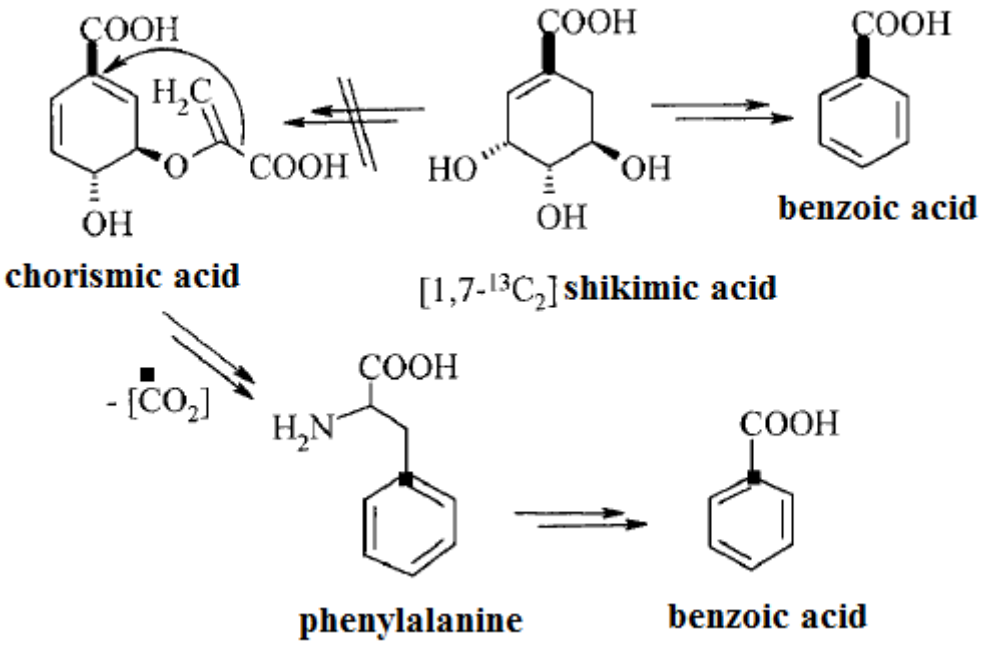

Figure 5.5 A depiction of the biosynthetic pathway leading to the synthesis of benzoic acid directly from shikimic acid in *S. griseoviridis* Tü 3634 and ruling out the plant-like pathway via phenylalanine described by Hertweck & Moore (2000), as established by ^{13}C labelling experiments (Hoffman & Grond, 2004).

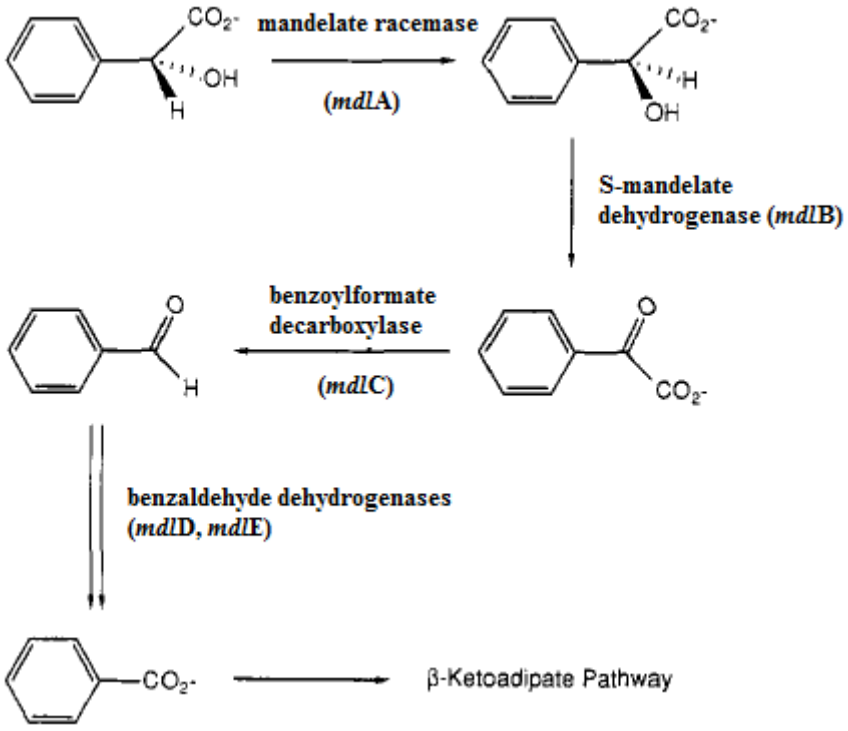

Figure 5.6 The *P. putida* ATCC 12633 mandelate metabolic pathway. Benzoate is subsequently converted to acetyl-coA and succinyl-CoA by the enzymes of the β-ketoadipate pathway (Tsou *et al.,* 1990).

Finally, benzoic acid biosynthesis in *S. polyantibioticus* SPR^T may occur via the key intermediate, chorismate. The chorismate pathway provides the aromatic building blocks for secondary metabolites such as enterobactin, pyochelin and yersiniabactin (Van Lanen *et al.,* 2008). Chorismate can be converted to 4-hydroxybenzoic acid (4HBA) by the action of chorismate lyase (Agarwal *et al.,* 2014), of which three homologues (SPR_8160, SPR_40730 and SPR_74890) have been identified in the *S. polyantibioticus* SPR^T draft genome. Dehydroxylation of 4-hydroxybenzoic acid would lead to the formation of benzoic acid. Indeed, the reductive dehydroxylation of 4-hydroxybenzoyl-CoA to benzoyl-CoA has been

reported under anaerobic conditions in *Pseudomonas* species (Glöckler *et al.,* 1989). However, there are currently no reports in the literature suggesting that 4HBA can be aerobically metabolized via dehydroxylation to benzoic acid (Fairley *et al.,* 2002).

5.4 Proposed pathway for the biosynthesis of DPO in *S. polyantibioticus* SPR^T

Most NPRS gene clusters abide by the so-called colinearity rule whereby each module is responsible for one discrete chain elongation step and the specific order of the modules defines the sequence of the incorporated amino acids (Wenzel & Müller, 2005). However, contrary to the hypothesis that a linear NRPS system with the standard C-A-PCP domain order would be identified for DPO biosynthesis in *S. polyantibioticus* SPR^T, the cluster exhibited a nonlinear arrangement with the core domains arranged as A-PCP-C, in addition to stand-alone A, C and PCP domains. Initially, these nonlinear NRPS systems were assumed to be exceptions to the colinearity rule, however, numerous examples of nonlinear NRPS systems have been reported in the past few years, and it has become increasingly apparent that they constitute a considerable portion of the range of NRPSs found in nature (Mootz *et al.,* 2002).

Based on the genome annotation analysis and gene disruption studies, a model for DPO biosynthesis can now be proposed. The source of benzoic acid is yet unknown, as the pathway for benzoic acid biosynthesis in *S. polyantibioticus* SPR^T has not been established, despite the efforts of this study to identify it. The acyl-CoA synthetase encoded by gene SPR_52860 is proposed to select and activate benzoic acid as a CoA-thioester in the first step of the DPO biosynthetic pathway (Figure 5.7).

As the substrate specificity of the A domain encoded by gene SPR_53060 is unclear (as it could bind either phenylalanine or 3-hydroxyphenylalanine), an *in trans* phenylalanine hydroxylation step is proposed. Indeed, FAD-dependent monooxygenases encoded by any of SPR_48520, SPR_49640, SPR_52630 and SPR_53180 could provide the hydroxylase activity to catalyse the β-hydroxylation of a phenylalanine residue in DPO biosynthesis. The A domain, encoded by gene SPR_53060, is proposed to be responsible for the selection and activation of 3-hydroxyphenylalanine through ATP hydrolysis. The β-hydroxyphenylalanyl-adenylate would then be transferred to the PCP domain, also encoded by SPR_53060, and formation of

the amide bond with benzoyl-CoA would be catalysed by the lone C/Cy domain encoded by SPR_53040. Heterocyclization of the intermediate, benozyl-β-hydroxyphenylalanine to 4-carboxy 2,5-diphenyloxazole would then be catalysed by the putative Cy domain encoded by SPR_53060 operating in tandem with the C/Cy domain encoded by SPR_53040. Indeed, the antiSMASH analysis, in addition to a multiple sequence alignment with the corresponding A-PCP-C NRPS module from *S. ambofaciens* ATCC 23877^{T}, revealed the existence of a large putative "recognition sequence" at the C terminus of the C domain (SPR_53040) and at the N terminus of the A domain encoded by SPR_53060. In contrast to standard COM domains which normally facilitate the interaction between PCP and C domains located on separate multienzymes (Hahn & Stachelhaus, 2006; Hahn & Stachelhaus, 2004), intersubunit protein-protein interactions in the DPO biosynthetic pathway may establish interfaces between the A (SPR_53060) and C/Cy (SPR_53040) domains to provide a functional pathway. Thus, the recognition sequence would mean that the C/Cy domain encoded by SPR_53040 would potentially interact with the A domain encoded by SPR_53060, thereby forming a conventional C-A-PCP arrangement.

The oxidation of the hydrolytically labile oxazoline to form the oxazole moiety in the heterocyclization reaction is proposed to be catalysed in a manner homologous to the oxidation process observed in pyrimidine biosynthesis.

Decarboxylation would be carried out *in trans* by a putative decarboxylase, such as the enzyme encoded by gene SPR_68606, which shares 30% amino acid homology to the decarboxylase found in the tautomycetin biosynthetic cluster from *Streptomyces griseochromogenes*. This reaction is required to convert 4-carboxy 2,5-diphenyloxazole to DPO.

The type II thioesterase, encoded by gene SPR_53090, would catalyse the release of DPO by breaking the covalent linkage between DPO and the 4'-phosphopantetheine (4'-PP) thiol arm. Even though the inactivation of the gene encoding this enzyme within the *S. polyantibioticus* SPR^{T} genome did not abolish DPO biosynthesis, the stand-alone type II TEs, often encoded within NRPS gene clusters, do not play an essential role in NRP synthesis because their removal from NRPS systems has been shown to decrease product titres, but not completely eliminate peptide synthesis (Schneider & Marahiel, 1998).

Finally, two NRPS domains found within the putative DPO biosynthetic cluster are believed to be inactive. The PCP domain encoded by SPR_53070 is postulated to be inactive due to the absence of the critical catalytic serine residue found in the conserved signature motifs of known PCP domains (Chapter 3; Figure 3.7) and the C domain encoded by SPR_52900 is believed to be inactive due to the absence of the H-H-x-x-x-D-G motif found in all active C domains.

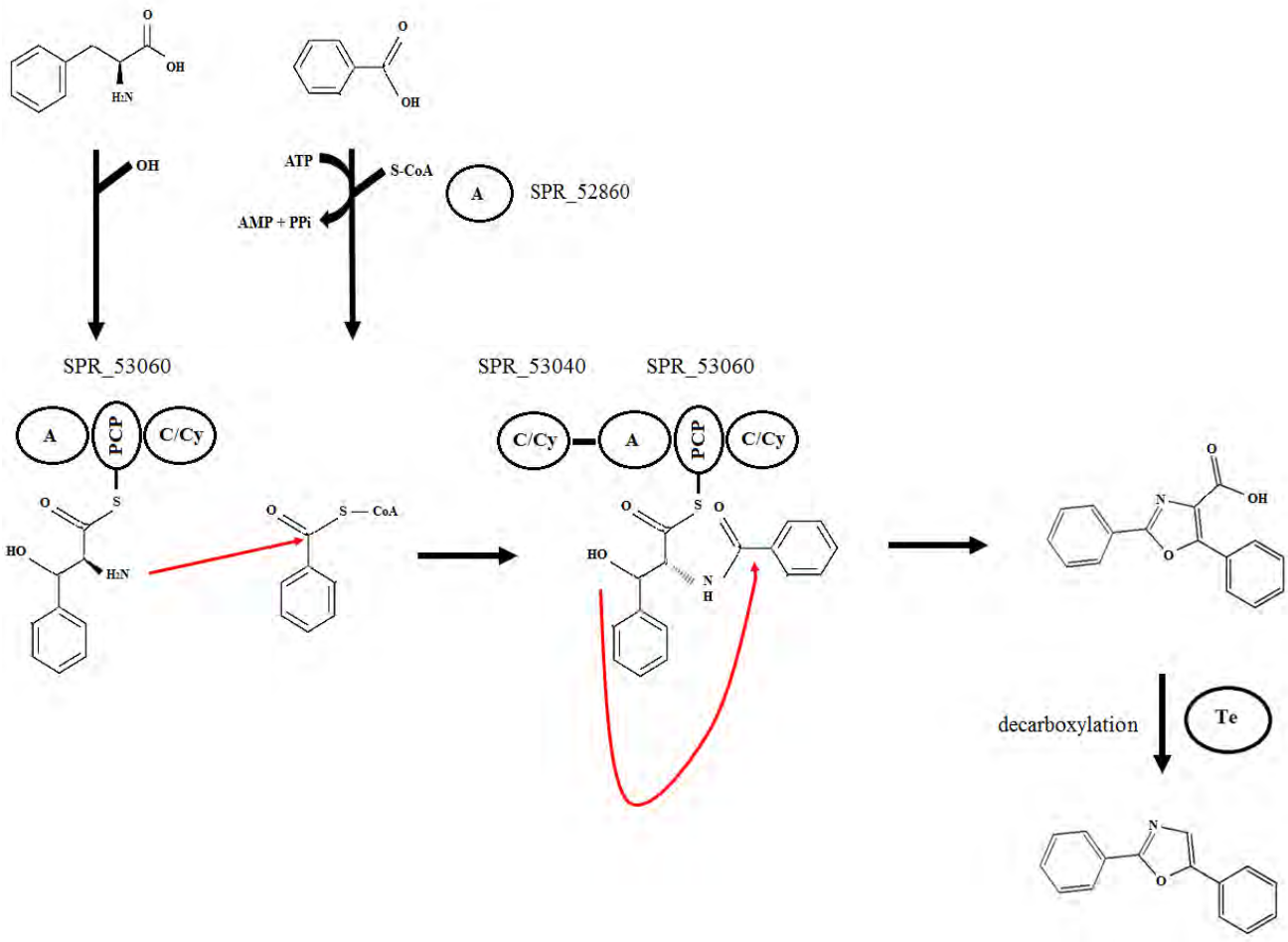

Figure 5.7 Proposed DPO biosynthetic pathway in *S. polyantibioticus* SPR^{T}.

5.5 Conclusion and future work

A gene cluster responsible for the biosynthesis of DPO was identified in this study and, based on the genome annotation analysis and gene disruption experiments, a model for DPO biosynthesis is proposed. At this stage, the model cannot account for the source of benzoic acid, as *in vivo* gene disruption experiments disproved both of the hypotheses on how benzoic acid is synthesized in *S. polyantibioticus* SPR^T. However, alternative hypotheses regarding benzoic acid biosynthesis in *S. polyantibioticus* SPR^T have been put forward and are suggested as the place to start in future studies to elucidate the production of this unusual starter unit in DPO biosynthesis. The heterologous expression of the putative DPO gene cluster, identified in this study, in host organisms such as '*Streptomyces maritimus*'strain DSM 41777^T, where benzoic acid biosynthesis occurs, could be used as a method to confirm the involvement of benzoic acid in DPO biosynthesis.

Future work should also involve the characterization of the substrate selectivity of the A domains encoded by SPR_52860 and SPR_53060, in order to confirm their specificity towards benzoic acid and 3-hydroxyphenylalanine, respectively, using an ATP-PPi exchange assay. Additional liquid chromatography-mass spectrometry (LC-MS) analyses of the antibacterial compounds produced by both the wild-type *S. polyantibioticus* strain SPR^T and constructed disruption mutants would confirm the identity of the bioactive compounds and further corroborate the results obtained from HPLC and TLC-bioautography analyses, thereby strengthening the preliminary hypothesis pertaining to DPO biosynthesis.

Additionally, DPO biosynthesis is most likely controlled by SPR_52890, the only gene in the putative DPO biosynthetic cluster predicted to be involved in regulation. This gene encodes a putative transcriptional regulator that contains a C-terminal DNA binding HTH domain and belongs to the two-component response regulator family. In order to identify the gene cluster responsible for the biosynthesis of benzoic acid in *S. polyantibioticus* SPR^T, the presence of cluster situated regulatory genes, such as SPR_52890, and/or promoter sequences can be used as a tool to search for any interconnected pathways in the genome.

Moreover, this study was successful in the optimization of a transformation method for the introduction of DNA into *S. polyantibioticus* SPR^T. In light of this, future work should entail

gene disruption experiments in order to determine the involvement of the genes postulated to be inactive, such as the PCP domain encoded by SPR_53070 and the C domain encoded by SPR_52900. In addition, gene disruption of the genes identified in Chapter 3 as having an unknown function in DPO biosynthesis will help to define the boundaries of the DPO gene cluster.

Furthermore, the identification of the gene cluster responsible for DPO biosynthesis has laid the foundation for combinatorial biosynthetic studies to create derivatives of DPO that might be used in the treatment of drug-resistant tuberculosis. Lastly, the *S. polyantibioticus* SPR^{T} genome sequence could be explored further to identify the antibiotic gene clusters for other potential antitubercular antibiotics that this organism produces.

5.6 REFERENCE LIST

Agarwal, V., El Gamal, A.A., Yamanaka, K., Poth, D., Kersten, R.D., Schorn, M., Allen, E.E. & Moore, B.S. (2014). Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nature Chemical Biology,* **10**: 640–647.

Anterola, A. M., & Lewis, N. G. (2002). Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. *Phytochemistry*, **61(3):** 221-294.

Ashforth, E. J., Fu, C., Liu, X., Dai, H., Song, F., Guo, H. & Zhang, L. (2010). Bioprospecting for antituberculosis leads from microbial metabolites. *Natural product reports*, **27(11):** 1709-1719.

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

Bérdy, J. (2005). Bioactive microbial metabolites. *The Journal of antibiotics*, **58(1):** 1-26.

Bernan, V. S., Greenstein, M. & Maiese, W. M. (1997). Marine microorganisms as a source of new natural products. *Advances in applied microbiology*, **43:** 57-90.

Bode, H. B., & Müller, R. (2003). Possibility of bacterial recruitment of plant genes associated with the biosynthesis of secondary metabolites. *Plant physiology*, **132(3):** 1153-1161.

Breese, K., Boll, M., Alt-Morbe, J., Schagger, H. & Fuchs, G. (1998). Genes coding for the benzoyl-CoA pathway of anaerobic aromatic metabolism in the bacterium *Thauera aromatica*. *European Journal of Biochemistry*, **256**: 148–154.

Fairley, D.J., Boyd, D.R., Sharma, N.D., Allen, C.C.R., Morgan, P. & Larkin, M.J. (2002). Aerobic Metabolism of 4-Hydroxybenzoic Acid in Archaea via an Unusual Pathway Involving an Intramolecular Migration (NIH Shift). *Applied Environmental Microbiology*, **68(12):** 6246-6255.

Garsin, D.A. (2010). Ethanolamine Utilization in Bacterial Pathogens: Roles and Regulation. *Nature Reviews Microbiology*, **8(4):** 290–295.

Geukens, N., Smitha Rao, V., Mellado, R.P., Frederix, F., Reekmans, G., De Keersmaeker, S., Vrancken, K., Bonroy, K., Van Mellaert, L., Lammertyn, E. & Anne, J. (2006). Surface plasmon resonance-based interaction studies reveal competition of Streptomyces lividans type I signal peptidases for binding preproteins. *Microbiology*, **152:** 1441-1450.

Glöckler, R., Tschech, A. & Fuchs, G. (1989). Reductive dehydroxylation of 4-hydroxybenzoyl-CoA to benzoyl-CoA in a denitrifying *Pseudomonas*species. *FEBS Letters*, **251:** 237-240.

Gulick, A.M. (2009). Conformational Dynamics in the Acyl-CoA Synthetases, Adenylation Domains of Non-Ribosomal Peptide Synthetases, and Firefly Luciferase. *ACS Chemical Biology*, **4(10):** 811-827.

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 USA*, **103:** 275-280.

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

Harvey, A.L. (1999). Medicines from nature: are natural products still relevant to drug discovery? *TiPS*, **20:** 196-198.

Herrman, K.M. (1995). The Shikimate Pathway as an Entry to Aromatic Secondary Metabolism. *Plant Physiology*, **107:** 7-1 2

Herrman, K.M. & Weaver, L.M. (1999). The Shikimate Pathway. *Annual Review of Plant Physiology and Plant Molecular Biology,* **50:** 473–503.

Hertweck, C. & Moore, B.S. (2000). A plant-like biosynthesis of benzoyl-CoA in the marine bacterium '*Streptomyces maritimus*'. *Tetrahedron*, **56:** 9115-9120.

Hill, A.M., Thompson, L.B., Harris, J.P. & Segret, R. (2003). Investigation of the early stages in soraphen A biosynthesis. *Chemical Communications*, **12:** 1358-1359.

Hoffmann, L., & Grond, S. (2004). Mixed Acetate-Glycerol Biosynthesis and Formation of Benzoate Directly from Shikimate in *Streptomyces* sp. *European Journal of Organic Chemistry*, **2004(23):** 4771-4777.

Hojati, Z., Milne, C., Harvey, B., Gordon, L., Borg, M., Flett, F. & Micklefield, J. (2002). Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chemistry & biology*, **9(11):** 1175-1187.

Jansen, R., Kunze, B., Reichenbach, H., Jurkiewicz, E., Hunsmann, G. & Höfle, G. (1992). Antibiotics from Gliding Bacteria, XLVII. Thiangazole: A Novel Inhibitor of HIV-1 from *Polyangium spec. European Journal of Organic Chemistry*, **4:** 357-359.

Juguet, M., Lautru, S., Francou, F. X., Nezbedová, Š., Leblond, P., Gondry, M. & Pernodet, J. L. (2009). An iterative nonribosomal peptide synthetase assembles the pyrrole-amide antibiotic congocidine in *Streptomyces ambofaciens*. *Chemistry & biology*, **16(4):** 421-431.

Kaneko, M., Ohnishi, Y. & Horinouchi, S. (2003). Cinnamate:Coenzyme A Ligase from the Filamentous Bacterium Streptomyces coelicolor A3(2). *Journal of Bacteriology*, **185(1):** 20–27.

Keating, T. A., Miller, D. A., & Walsh, C. T. (2000). Expression, purification, and characterization of HMWP2, a 229 kDa, six domain protein subunit of yersiniabactin synthetase. *Biochemistry,* **39:** 4729-4739.

Kelly, R. A., Scott, N. M., Díez-González, S., Stevens, E. D. & Nolan, S. P. (2005). Simple synthesis of CpNi (NHC) Cl complexes (Cp= cyclopentadienyl; NHC= N-heterocyclic carbene). *Organometallics*, **24(14):** 3442-3447.

Koetsier, M., Jekel, P., van den Berg, M., Bovenberg, R. & Janssen, D. (2009). Characterization of a phenylacetate-CoA ligase from *Penicillium chrysogenum*. *Biochemical Journal*, **417:** 467-476.

Koetsier, M.J., Jekel, P.A., Wijma, H.J., Bovenberg, R.A.L. & Janssen, D.B. (2011). Aminoacyl-coenzyme A synthesis catalysed by a CoA ligase from Penicillium chrysogenum. *FEBS letters*, ***585*(6):** 893-898.

Kunigami, T., Shin-Ya, K., Furihata, K., Hayakawa, Y. & Seto, H. (1998). A novel neuronal cell protecting substance, aestivophoenin C, produced by Streptomyces purpeofuscus. Journal of Antibiotics (Tokyo), **51(9):** 880-882.

Kyndt, J.A., Meyer, T.E., Cusanovich, M.A. & Van Beeumen, J.J. (2002). Characterization of a bacterial tyrosine ammonia lyase, a biosynthetic enzyme for the photoactive yellow protein. *FEBS Letters*, **512**: 240–244.

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.

Li, X., Zvanych, R., Vanner, S. A., Wang, W., & Magarvey, N. A. (2013). Chemical variation from the neoantimycin depsipeptide assembly line. *Bioorganic & medicinal chemistry letters*, **23(18):** 5123-5127.

Li, B. & Walsh, C.T. (2010). Identification of the gene cluster for the dithiolopyrrolone antibiotic holomycin in *Streptomyces clavuligerus.* Proceedings of the National Academy of Sciences, **107(46):** 19731-19735.

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.

Linne, U. & Marahiel, M. A. (2004). Reactions catalyzed by mature and recombinant nonribosomal peptide synthetases. *Methods in enzymology*, **388:** 293-315.

Lombó, F., Velasco, A., Castro, A., De la Calle, F., Braña, A. F., Sánchez-Puelles, J. M. & 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(2):** 366-376.

Michal, G. (1999). Biochemical Pathways. Spektrum Akademie Volume 1. Verlag, Heidelberg.

Marshall, C. G., Hillson, N. J., & Walsh, C. T. (2002). Catalytic mapping of the vibriobactin biosynthetic enzyme VibF, *Biochemistry*, **41:** 244-250.

McQuade, T. J., Shallop, A. D., Sheoran, A., DelProposto, J. E., Tsodikov, O. V. & Garneau-Tsodikova, S. (2009). A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. *Analytical biochemistry*, **386(2):** 244-250.

Mootz, H.D., Schwarzer, D. & Marahiel, M.A. (2002). Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem,* **3:** 490-504.

Navarro-Llorens, J.M., Patrauchan, M.A., Stewart, G.R., Davies, J.E., Eltis, L.D. & Mohn, W.W. (2005). Phenylacetate Catabolism in *Rhodococcus* sp. Strain RHA1: a Central Pathway for Degradation of Aromatic Compounds. *Journal of Bacteriology*, **187(13):** 4497-4504.

Nerud, F., Sedmera, P., Zouchov´a, Z., Mus´ılek, V. & Vondr´aˇcek, M. (1982). Biosynthesis of mucidin, an antifungal antibiotic from basidiomycete Oudemansiella mucida 2 H-, 13 C-, and 14 C-labelling study. *Collection of Czechoslovak Chemical Communications*, **47:** 1020–1025.

Phelan, V. V., Du, Y., McLean, J. A., & Bachmann, B. O. (2009). Adenylation enzyme characterization using γ-18 O 4-ATP pyrophosphate exchange. *Chemistry & biology*, **16(5):** 473-478.

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(18):** 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*, **262(2):** 507-515.

Rohr, J. (1997). Biosynthesis of Taxol. *Angewandte Chemie International Edition*, **36:** 2190-2195.

Rother, R., Poppe, L., Viergutz, S., Langer, B. & Retey, J. (2001). Characterization of the active site of histidine ammonia-lyase from *Pseudomonas putida*. European Journal of *Biochemistry*, **268**: 6011–6019.

Roy, R., Gehring, A., Milne, J., Belshaw, P. & Walsh, C. (1999). Thiazole and oxazole peptides: biosynthesis and molecular machinery. *Natural product reports*, **16(2):** 249-263.

Schaffer, M.L. & Otten, L.G. (2009). Substrate flexibility of the adenylation reaction in the Tyrocidine non-ribosomal peptide synthetase. *Journal of Molecular Catalysis B: Enzymatic*, **59:** 140-144.

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.

Schneider, A. & Marahiel, M. A. (1998). Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. *Archives of microbiology*, **169(5):** 404-410.

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.

Schuster, B. & Retey, J. (1995) The mechanism of action of phenylalanine ammonia-lyase: the role of prosthetic dehydroalanine. *Proceedings of the National Academy of Sciences USA*, **92**: 8433–8437.

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

Schwede, T.F., Retey, J. & Schulz, G.E. **(1999).** Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile. *Biochemistry,* **38**: 5355–5361.

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

Steller, S., Vollenbroich, D., Leenders, F., Stein, T., Conrad, B., Hofemeister, J. & Vater, J. (1999). Structural and functional organization of the fengycin synthetase multienzyme system from *Bacillus subtilis* b213 and A1/3.*Chemistry & biology*, **6(1):** 31-41.

Tang, L., Zhang, Y. & Hutchinson, C.R. (1994). Amino Acid Catabolism and Antibiotic Synthesis: Valine Is a Source of Precursors for Macrolide Biosynthesis in *Streptomyces ambofaciens* and *Streptomyces fradiae*. *Journal of Bacteriology*. **176(19):** 6107-6119.

Tsou, A.Y., Ransom, A.C. Gerlt, J.A., Buechter, D.D., Babbitt, P.C. & Kenyon, G.L. (1990). Mandelate Pathway of *Pseudomonas putida*: Sequence Relationships Involving Mandelate Racemase, (S)-Mandelate Dehydrogenase, and Benzoylformate Decarboxylase and Expression of Benzoylformate Decarboxylase in *Escherichia coli*. Biochemistry, **29:** 9856-9862.

Van Lanen, S. G., Lin, S., & Shen, B. (2008). Biosynthesis of the enediyne antitumor antibiotic C-1027 involves a new branching point in chorismate metabolism. *Proceedings of the National Academy of Sciences*, **105(2):** 494-499.

Wenzel, S.C. & Müller, R. (2005). Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from text book biosynthetic logic. *Current Opinions in Chemical Biology,* 9:447-458.

Wright, G. D. (2012). Back to the future: a new 'old'lead for tuberculosis. *EMBO molecular medicine*, **4(10):** 1029-1031.

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N. & Eisen, J. A. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature*, **462(7276):** 1056-1060.

Zhang, W., Ostash, B. & Walsh, C. T. (2010). Identification of the biosynthetic gene cluster for the pacidamycin group of peptidyl nucleoside antibiotics. *Proceedings of the National Academy of Sciences*, **107(39):** 16828-16833.

APPENDIX A

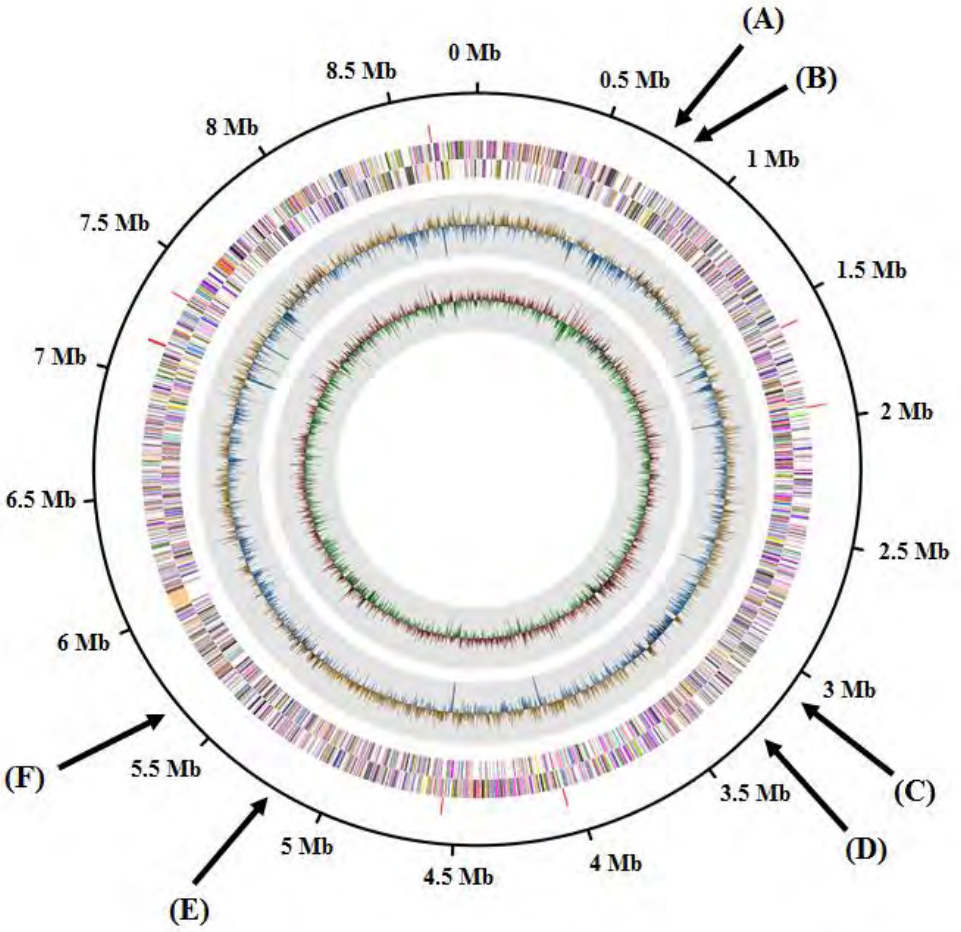

Figure A1. Genome map of the draft *S. polyantibioticus* SPR^T genome generated using the CLC Genomics Workbench 7.0.4 (CLCbio, Denmark). Arrows represent the relative positions of the secondary metabolite clusters discussed in Chapter 3, where (A) is the NRPS gene cluster spanning the region located from nucleotide 637827-688770, (B) is the NRPS gene cluster spanning the region located from nucleotide 669778-735243, (C) is the NRPS gene cluster spanning the region located from nucleotide 3093585-3154158, (D) is the NRPS gene cluster spanning the region located from nucleotide 3320540-3386004, (E) is the NRPS gene cluster spanning the region located from nucleotide 5343994-5398013 and (F) is the NRPS gene cluster spanning the region located from nucleotide 5634984-566500.

APPENDIX B

Nucleotide sequences of the genes preliminarily identified as components of the DPO biosynthetic gene cluster are provided in this appendix.

> SPR_52860

ATGCCCTCGACCACTGAATCCGCCGCCACGCTGTGGGAGCTCATCGACCGGGCCGTACGCCTCTCCCCCGCGGCC
ACCGCCGTACGCCAGGGAACGCGGGCCCTCACCTTCCGCGAACTCGCCGACCGCGTCGAGACCACAGCACGCCGC
CTGCGCACGCGCCTGCCCGCCGACGGCGGAGGAACCGTGGGCCTGCTCTTCGAGAACACCGTGGAGAGCACGGTG
GCCTTCCTGGGCGCGCTGTACGCGGGCGTACCCCTCACGCCCCTCGAACCCGACAGCACCGAGCCCCATCTCCTG
GGCGTGCACCGGGACTTGGGACCCCTGCACCTCGTCGGTCGGCAGGCGAGGCTCAGCACACTGGCCCCTCCCGCT
GCCACCTCACTCGCCGCCCGGTGGCAGGGCGGTGTCCTCATCGACGTGGACGAGCTCACCGCCCGCCCCGGCGCG
GCGGCCCCGTCGGCGCCGCTGCCCGCGCCGCCGCCCGACGCCCCCGCGCTCTACCAGTACACCTCCGGCTCGACC
GGCGAGCCGAGGGCCGCCGTGCACTCACAACACGACCTGGTCCGCGGTGGCGAGATCTACGCCCGCACCTACGGC
ATCACCCCGGCCGACCGGATTCTCGCGGCCGTACCCCTGCTGCACTCCTTCGGCATGGTCGCCGCGCTCGCCACC
GCGCTGCACGCCCGCGCGGAGCTCGTCCTGCTCGGCCGGTTCGCACCGGCCGAAATGCTCCGGGCACTGCACCAG
CACGCCTGCACCATCGTGGTGGGCACCCCGCTCGCCTATGACCTGGCGGCCCGTTCGGCGGCCTCGCGCAGCGAG
ACGTCCCGCCCGGGCGACACCGTACGGCTCTGTCTGTCCTCCGGGGCGGCCCTGCCCCCGGCCGTGGCGGACCGC
TTCGCCCAGCACTGCGGCCCGGCCGTCCAGCAGGTCTACGGAAGCACCGAGGCCGGTGTCGTCGCCGCGCAGCTC
CCGCAGCCGGACGGCACGGCCGATGCCGGGGTGGGCAGGCCGGTGCCCGGGGTACGGATCCGCCTGGTGGACGAG
GAGGACCGACCGGTTCCGCCGGGCGGCACGGGCGCCCTGCTCGTGCGCACCCCCGCCATGTTCACCCACTACCTC
GGCCACCCGGGCCCGTCCGAACGAGCCTTCCGCGACGGCTGGTACGTCACCGGCGACCTGGCGCGCCTCGACGGT
GACGGCCGTCTGCACCTGGTGGGCCGCAAGGAGTCCTTCATCAACGTCGGTGGCAAGAAGGTCAATCCCCTGGAG
GTCGAAGCGGCCCTGCTGGCCCACCCGGCCGTCGCCGAGGCCGTGGTCTGGGGCGAGGTCGTGGAGGAACAGACC
AGCGAGCGCGTACGGGCAGCCGTGGTGCCGTGCTCCCCGCTCAGCGGGGCCGCGATCACCGCGCACTGCCGTGAC
CGGCTCCTGGCCCATCAGGTGCCCGCGACGATCGAGTTCGTGGCGTCGCTGCCCAAGACCTCCCTGGGCAAGATC
CGCCGCGCGGCCGTGGCCGGAGCCGCCCCCGCACGGCACCGGGACGACACGGCCTGA

> SPR_52870

ATGGCACTTCTGGACGGCATCGAGGCGGGCGCCGAGGACGCCTACGACCGCCATTACGGCGACCGGCAGCTGCTG
CGCCGCATCACCCCCTACTTCCAGGGCCGCGGCGGCGCGGTGGCCTTCGTCACCGCGGCGGTGGTCGTGGCGGCG
GCCGCGGGCAGCGCCGTACCGATCCTGCTGGCCCGCACCGTCGACTCGGCCCTGGACGACGACACCCGGTCGGCG
ATGACACTGCTAGCCGTTGCCGTCCTCCTCACCGGCGCCGCGGCCTGGCTGTTCACCGTCCTCCAGCAGCGCCTC
ACCGGCACCCTTGTCGGCGACGCGGTGCTGCAACTGCGCGAACACGCCTTCGCCGCGGCCGTACGCCAGGACCTC
GCGTTCTACGACACCCACCCCACGGGCCTGGTGGTCAGCCGCGTCACCAACGACACCCAGACGTTCGGTGCCCTG
CTCTCCCTGGTGCTCTCCCTCATCGGTCAGGTCCTCGTCGTCCTGCTGATCCTGACGGCACTCTTCGTCATCAAC
GTGCCGCTGGCGCTGCTCACCACCGTGGTGGTCGCCCTGATCGTCGCGGTCAGCCTCGCCTTCAGGAAACGCGCG
CGCACCGCCTCCACCCGCCAGCAGCAGGCCCTCGCCGAGGTCAGCGGCTATGTGCAGGAGACCCTGCGCGGGATC
ACGGTGGCCCGCAACCACCGCGGCGAATCGGCGGCCGAAGCGGGGCTGCGGGCGGTCAACGGCCGCTGGTACCAG
GCCAGTGTGCGCCTCAACCGCCTCTTCAGCGGCATCTTCCCGCTGCTCATCTCCCTGACCGGACTGGGCACGACG
GCGGTGGTGCTGGCCGGCGGCCACCAGGTCGCATCCGGCGACATCTCGGCCGGCGAATGGGTCCTCTTCCTGGAG
GCCCTGGCCCTCTTCTGGTCGCCGGTCGCCACCATCGCCTCCTTCTGGCAGCAGCTCCAGCAGGGCCTCTCGGCC
GGGGAGCGGATCTTCGCGCTGATCGACCGGGAACCCGCCGTCCACCAGAGCGCCCAGCTGCCCCCGCCCCGTCTG
CGCGGCGCACTCGAACTGCGCGGCGTCAGCTTCCGCTACACCCCCGAGCGGCCCGTCCTCCACGACATCGACCTC
ACCGTCGCGGCCGGGGAGACCGTCGCACTGGTCGGCCACACAGGGGCGGGCAAGTCCAGCCTGGTGCGCCTGCTG
ACCCGCTCGTACGAGTTCCAGGAAGGCAGCCTGCTCGCCGACGGCCGGGACGTGCGGACCTTCGACCTGGCGCAG
TACCGGCGCTGCCTCGGCGTGGTCACCCAGACGCCCTACCTCTTCTCCGGCACCGTCGCCGAGAACATCGCGCTG
GGCCGGCCGGGCGCCGACCGCGCGGCCATCGAGGCCGCGGCCCACGCGGTCGCCTCCGGCGCCTGGCTGCGGCAG

CTGCCCGACGGCCTCGACACACCGACCGACGAAGGCGGCCGCAACCTGTCGACCGGGCAGCGGCAGATCGTGGCG
CTGGCCCGGGTGTTCCTGCAGGACGCCCCGATCTTCATCCTCGACGAGGCCACCGCCAGCGTCGACCCGCTCACC
GAGGCCCAGATCCAGGAAGGACTCGACGCGCTCGCGGCGGGGCGCACCACCATCGTCGTGGCGCACCGGCTCCCC
ACCGTCCGCAAGGCCGACCGCATCGTCGTCGTCGACGACGGGCGCCTCGTCGAACAGGGCGACCACGCCACCCTG
ATGCGGCAAGGCGGTGCCTACAGCCGCCTCTACCGGCAGTACTTCCACCATCAGCACCCCGACCACGACCCCGCC
GACGACCCGTCCGGCACCGGGCGGGCGCCGGTCGGCGCGGGGCCGCACCTAGCGGAGGCCAGCGATGCCCTCGAC
CACTGA

> SPR_52880

ATGACGACCGAGACCGCCGCGCAGACCCGCTCCGCGTCCGGACCGGCGCCGGCCACCGGGCCCGAACTGACCATC
GGCCACTGGCTCCGCCTGCACATGCGTCACTTCCGCCGGCACATCGCGGTCCATCTGACCGGAAGCCTGCTCTGG
CAGTCCCTCACCGCGGCCCTCCCCCTGGTCATCGGGCTGGCCTTCGACGCCGTACTGAAGACCAACGGGCGGGAA
CCTGATCTGGGGGCCCTGCACGCCGTGTGCGCGGCACTGCTCGCCCTGGTCCTCGTCCGGGGCGTCTGCGGCATC
GCCTCCACCTACGCCCTGGAGGCCTTCGCCAGCGGCCTTGAACGCGACGCGCGGGCCGGCGTGTTCGCCAGCATG
CTCCGCAAGAAGCAGGGCTTCTTCAACCGGCACCGCACCGGCGACCTGTCCACCCGCGCCACCGGCGACGCCGAG
GCACTCGGACTCATGGTGTCGCCGGGCTTCGACATGACCCTCGACATGGCGCTCAACACCCTCATTCCGCTGGTG
TTCATCGCGGCCGTCGACTGGCGGCTCCTGCTCTTCCCCGGGATCTTCGTCGTCCTGTTCGTCGCCGCCCTCATG
GATCACGGGCGGCGCCTCGAACCGGTCTCCGACCTCACCCGCGAGGAGTTCGGCGCGATGTCCGCCCAGGCCGCC
GAGAGCATCACCGGAATCGAGGCGGTCGAGGCGACCGACGGAGCGGAGCGCGAGCAGGACCGCTTCCGCCGCACC
GCCACCGCCTACCGGGACGCCGCCGTGGCGCAGGCACGCGTCCAAGCCCTGTCCTTCCCCCCGCTGCTGCTCGCC
GTGGCCACCGCCGGAACGCTGCTCCACGCCGTCTTCCTGCTGCGCGACGGCTCACTGAGCGTCGGTGAACTCGTC
GCCGTCCTGGGCCTGATGAGCACGCTGCGCGCCCCCACCCAGCTCGCCTCCTTCAGCATCGGGCTCATCTTCTTC
GGACTCTCCGGCGCACAGCGCATCCTGGAGATCATCAACGACCGGGCGGGAGAGGACGAGGAGGGAGGCCGGCAC
GAGGCGGTCATCACCGGCGAAGTGGTCTTCGAGAACGTCACCTTCGCCTACGGAAACGGCAAGCCCGTGCTGCGC
GACATCTCCTTCCGCGCCGCCCCCGGCACCACCGTCGCCGTCGTCGGCGCCACCGGCAGCGGCAAGTCCACCCTG
CTGCACCTGCTGAACCGCACCTACGTGCCCACCGAGGGCCGCGTCCTCGTCGACGACGTGGCCACCACCGCCTGG
GACCCCTCGGCACTGCGCGGCCAAGTAGCCGTCGTCGAGCAGGACGTGGTCCTGTTCTCCCGCACCATCGCCGAG
AACCTCGCCTTCGGCGCCGCACCGGACACCGGCCGGGCCCAACTGGAAGAGGCGGCCCGTACCGCCCAGGCGCAC
GACTTCGTCATGGCGAGCGAGCACGGCTACGACACCGTGCTGGGCGAGCGCGGCGTCACCCTCTCCGGCGGCCAG
CGCCAACGCCTCGCCATCGCCCGGGCGTTGGTGACCGACCCGCGCATCCTCGCCGTGGACGACGCGACCAGCGCC
GTCGACAGCGTGACCGAACACGAACTGCAGCTCGCCATGCGGCGCGCCGCCACGGGCCGTACCACCTTCCTCGTC
ACCCCCCGGCTGTCCCGCATCCGCGCCGCCGACCACATCGTCGTCCTCGACGCGGGCCGGGTCGTCGGCCAGGGC
ACCCACGAACACCTCCTGCGCACCTGCGCGTTCTACCGCACGATCTTCGCGCCGTACGCACTGCCGGCCGCGGAG
GGAACCACCACCGCCGCCGTCGCGGCCGGGGCAGAGGAAGGAACCCGCTGA

> SPR_52890

GTGAGCGATTTAGCATGCAAACCCCCCGCCCCGTCGCGACCGCTGGTCGTCGCGACTCTTGTCCACGACGGCCTC
GCGCGCCAGGAGCTGCAGACCGCTCTGGCACAGGCCGCCCTCGTGCGCGAAGCCCGGCACTGCGCCGACCGACGC
GAGGCGCGGGAGGCTCTGGACTCCGGCCGGGTCGACGTGTTCGTCGTGGCCAGGAGCGACTACGACGCCCATGCC
GACTTGCTTGACGACTTCCGCTCGCCCCTGCCGAGGATGCTTGTGCTGTTGAGTCATGCCGATCCGGACGAAGCG
ATGCTGTCCGGGCCACCAGCGCCCGATGGATTCCTGCTGCAGGGCGGTCTGAGCGCCGCCGATGTGGACAGCGCG
CTGCGGCGCATGGCCGCGGGCGAGGTACCGATGCCGGCCCTGCTCGCCCGCGCCCTGATGGACCGTGCCAGCACG
AGCCGCCTGCCGCACGGCCGACATGACATCGCTGTCACCGCCCGGGAGACCGAGGCGCTGCTCCTGCTCGCCGAG
GGGCTGAGCAACAAGCAGATGGCCCGGCGGCTGGAGATCACCAGCCATGGCGTGAAACGGCTCGTGGCCAGTCTG
CTGCTCAAGCTCGGGGCGCCGAACCGGACGGCTGCCGTGGTCCTGGCCATCCACGCGGGTCTGATCCGCACGACC
CGGCCGGGCCAGGGCACGGCGGACCGCCGGGTGGTACGGGTCCAGCTCGCGGGGGGTGGAGCGGACGTCCGCCTC
CCGGTCCCCGCCGTGCGCTTCCTCGGTGAGGAGGTGCCCGAGGAGCGCGCCTCCTGA

> SPR_52900

GTGGAAAAAATCGAGACGCGTGATGTCCCCTTCTACGGTCGCCGTTCGGTCATGGCGTCGGCCACCTGTGCACAG
CGGCAGATGTGGGACCTGATCCGACGGGAAATTCCCGACGGCTCCTTCTACGATTTCTGCCACGAGGTTTCCCTT
ACCGGCGAAGGAACCGTGGATGACGTTCTCGCCGTATGCGCGGAACTCCTCTCCCGGCATGAGTCGCTGCGTACG
GCATTTTTCTGCGGTGCGGACGGGAAATTGGTGCAGAGCGTGGCGCGGTCCGGATCCTTCGAGGCGGAGATATGC
GCCACCGGGCCCGGCGAGAGCGCGCAATTCGTGTACGACGCCTGGCGGCGGCGGATGCGGGACAGGACGTTCGAC
CTGCGGAACGGCCCGCCGCTGCGGGCGCTGGTCGTCGTCACCGGGCACGCGCCCGTCATGGCGGCCTTCTGTGTC
TCGCACCTGGCCGCGGACCTCATGTCGATGCGCACGCTGGCCGGTGAAGTCCTGCCCCTGCTGGCCGCGCGGATC
GCCCGTASCCCCGCCCCGCCGCCCGCCGTGACGCGCCAGCCGGTCGAACAGGCCGACTTCGAGCACTCCCCACAG
GGGCGCGCCCTGCTGGCCCGGGCCCACACACACTGGCGGCAGCAGCTGGCGAGCGCCCCCACCACGATGTTCCCC

GGCCGTCCCGCCCCGCAGGGGCCCGGCCGACCGGACCGCTACAGCGCGGCCATGGACTCCGCGGCCGCCTTCCTT
GCCATCCGTGCCCTGGCCGTCCGGTGGCGGGTGAGCACCTCCGCGGTGCTCCTCACAGCGGTCGCACTGCTGCTC
GGCGCCCGCTCCGGGCAACAGACCTGCGCCCTGAGGCTGTTGGCCGCCAACCGCACGCACCCCGCACTCCAGCGA
AGCGTGGCCAACCTGCACCAGGAGGTCCTCACCAGCTTCGACCTGCGCGGGGAGAACGTGCGGGCCGTCGCCCGC
CGCGCCTTCGCCGCGGGGACCCTCGCGTACGCCAACGGCCTGTTCGACCCCGACGGGGCGAACGAGCTGATCCGG
GCCGAAGGGCGCAGGAGGGGCGAGCCGCTCCAACTGTCCTGCTGCTTCAACGACATCCGCACCGACCACGATCCG
CGGAGTCCGGGCGGGACGGCGTCCGCCGAGCAGATCCGCGCCGCCCTGGCCCGTACCGTCGTCGCCTCCAGCGAC
TTCGAGGAAGCGGAGACCTTCTTCCTGGTCGTCGTCGACACCGACCCCGGGTGGCTCAGGTTCGTGCTGTGCGCC
GAGACGGCCGCCCTGTCGCCCGAGGAGGTGCACATCTTCCTCCGCGACCTCGAACGGCTCCTCGTGGACTGCGCC
GAGCAGCCCGAACGTTCCTGGCCACGACTCGATCAGGGCCGGGGCCCGCAGGCCGCCCAGCCACCGCGGACGGCT
GCCCGGGGATGA

> SPR_52910

ATGCCGCAGCACCGGGTGACGGGGGAGTCGACTGTGTTCACGGAAGCATTGGCACGTGAGGCTGCCATGGACGTG
GAGAACCCCGCGCACGGCACGGTGTTCGCGCAGGCTCCCCGCTGCACCGCCGCCGAACTCGACGCGGTCCTCGCC
GCCTCGGCCCGGGCGTTCCCCGGGTGGGCGGCACTGCCGCTGGAGAGCCGGCGTCCGTACCTGCTCGCCTGCCGG
GACGCGCTGCGCGACGCCGAGGACGACGTCGCGGACCTGCTGACCCGGGAACAGGGCAAGCCGCTGCGGCACGCG
CACCACGAAGTGCGGCTGGCCGCCGACTGGTTCGCGCACACCGCCGAACTCGCCCTGCACGCCGACCGGATCGTG
GACGAGCCGACGGCCCGCGTCACCCTGGAGCGGGTGCCGCACGGTGTCGTCGCGGCGATCGCCCCCTCCAACTAC
CCCGTCCTGCTGGCCGTGTGCAAGATCGCCCCGGCGCTCCTGGCGGGCAACACGGTCGTCCTCAAGCCCTCGCCC
GCCACCCCGCTGTCCAGCCTGCTCATGGGCGAGGTGCTGGGACGGGTCCTGCCGCCCGACGTCCTGCGGGTGATC
AGCGGCGACGCCGCGCTGGGGGCCCGGCTCACCGCACACCCGGCAGTGCGGCTGATCTCCTTCACCGGCTCGATC
GCCGCCGGCCGCGCCATCGCCCGGTCCGCCGCCGCGGACTTCAAACGGACCGTCCTGGAGCTGGGCGGCAACGAC
CCGGCCATCGTGCTGCCCGGCGCCGACGCACAGGCCCTGGCGCGTCCGCTGTTCGACCGGGCGATGGTCAACAGC
GGCCAGTTCTGCGCCGCGGTCAAGCGCGTCTACGTGCCGCGAGCCCAGCACGCCAAGCTCACCGCGGCCCTGGCC
GCGCCGGCCGAGGCCACCGTGGTCGGTGACGGCCTCGACGCCGCGACCGAACTCGGCCCGCTGGTCAGCCGTGAG
CAGCTCGCCCATGTGACCGCGCTGGTGGGCGACGCGGTGCGGCGGGGCGCCCGGCTCGTCACCGGGGGCCGCGCC
CTGGAGCGGCCCGGCCACTTCTATCCCCCCACCGTCGTCACGGACTTGCCGCAGGGGACGCGGCTTGAGGAGGAG
GAACAGTTCGGCCCCGTGATCCCCGTGATCGCGTACGACGATCTCGACGCCGTGACGGCCCGCGTCAACGCGTCC
CCCTACGGGCTGGGTTGCTCCCTGTGGGGCGATCCGGAACAGGCCGCCGCGGCGGCCCGGCGGCTGGACTGCGGA
ACGGTGTGGATCAACACCCACGGCGACCTCAGACACGACGTGCCGTTCGGCGGCCACCGGCACTCGGGCACGGGC
GTCGAGTACGGAGCGTGGGGCCTCCTGGAATACACCCAGATCACCATCCACCACCTCGCGCTGCGCGGGGGAGAT
GAGGGGAAGGCATGA

> SPR_52920

ATGAGCGAAACCACCTCCGGCGGCGCCCCCGGCACGAGCCTGGTCGATCTGTCGGGGCTGCGCGTGTTCGTCGGC
GGCCCCATCCAGTACGCCCTCGGGCACGACAGCTTCCACCCACCGCTGCGCGACACGATCCAGGCCATCGTCACG
GCGGTGACCGAGGCGGGGGCCACCGTCTTCTCCGCCCATGTCGTCGAGAGGTTCGGCCGGGACACACCGCACTTC
TCGCCGGAGGACGTCAGCGCACGGGACCTGGACTGGATGCGCCGGTGCGACGTGTTCGTGCCCGTCCTGCCGGCG
GACGACGGTGGCGGAGTGATGCGCACCGACGGCACGCACATCGAGATGGGCTGGGCGTCGGCGCTGGGCCGTCCC
ATCGTCATGGTGACCCCCCTGCCGGTCCCGGCCGGGGCCAGCCATCTGCTGCGCGGGCTGCCGTCCGTGGCGGAC
GTCGGCGGCGTCGACCTGGCCGAGCTGCGCCGCGACGGCCCCGGTGAACTGCTGCTGCGGCTGGGCAAGGTCGGC
CGCGACGCGGTGGTGACACCATGA

> SPR_52930

MTTVSHAASATEVLGLRLVSPLVVGSGLLTDQERNIRRLLTGGAGAVVTKTIHPGPLPAGDERLLHLPTGMINST
TYSRRGVDSWCATLSRFADDDLPVIASVHADSPAALADLAARVADTGCRALELGISCLNEDGGLADSARRVADYT
GAVRRATPLPFSVKLAVGEQVGERVSAAVEAGADAITLSDTIAGLAVSPDTGEVLLGRPFGYSGAGIKPLVLAAI
YELRRGGLSVPVMGSGGVRSGTDVAEYLTVGADAVQVYTALHTHMHQTLAEIRGGFDTWLGARGGSVADVVGRAL
DGGRGEVRAAADHERHGLR

> SPR_52940

GTGAGGTGCGTGCGGCAGCGGACCACGAACGTCACGGCTTACGGTGATCTGACCAGTCCCGAGACACCCGCGGCG
CTGAGCGGCGCCACTCTGGTCTGGCCGGTGGGCGGCCTTGAGCAGCACGGCCCGCACCTGCCCCTGTCGGTGGAC
TACGACATTCCCGACGCGCTCGCCCGGCAGCTGGTGGCGGACGTCGACGGCGTGCTCTTCCCGGGCCAGCCGCTG
TCCGCCAGGTCACTGCCGCACAGCGGCGGCGGACTGCGCTTTCCCGGCACCGTCCACATCGACGGGGGAACCTTC
ATCGACTACGTCGCCCAGTGCCTGCGGGCACTGGGCAGGCTCGCTCCCGCCCGGCTGGTGGTCGTCAACGGGCAC

TACGAGAACGAGGCCCTGCTGTTCGAGGCGATCGACGGACTCGACCCGGCGCGCACCTTCCCCGCCACCGAGATC
GTCGCCTTCAGCTGGTGGAGCCTGGTGGAGGAGGACTGGCTGCGCAAGCGGCTGCCCGAATTCCCGGGCTGGCAC
GCCGAGCACGCCGGGGTGACCGAGACCAGTCTCATGATGTACCTGCGTCCCGAGGTGGTCCGCCCGGTGCGGGTG
GACCACCCCACTCCGCCCGCCGCCGGTGTGTACCGCCACCCCGTCGACGCGGCCCGCATGTCCACCCAGGGGGTC
CTCACCACGACCTCCGGTGCCAGCGGCGAGCTGGGGGAGGAGCTGTTCTGGCACGTGATGGACGGCGTGGCACGG
ACCCTCGCGGACAGCCCGTCCCCGCGCACCGAGAAGCGCGCAAGCGAGTGA

> SPR_52950

ATGCGCCCGCCGACGACCGAACGCGACGACTTCGAAGCCCGGCTGCAAAGCCTCTTCGGCGACAAACTCCGCCCC
GTGGAGGAGAGTTTCGAGGCCATCCAGCACTTCAAGGACGGTTCCTTCGCGGTGGGGCAGCTCGGCCTCATGCTC
TACACCAACGGATACAACCTGCGGGGCACCAGCACCCGTCATCCGAACGGAATGGTCTTCCACGACGGGCAGAAC
CTCTTCGGCGTAGGCTACTTCAGCAAGGAACAGGACGAGCGCAAGCACCTGCACATCGTCGCTCCCAAGGGCAAG
GACCGGGTCGCCGCGGTCAGGTCGTTCATCTCCGCGACCCGCGAGGCGGGCCTGGCACACACCTCCGTCTACGTA
CGCCACCTCTCGCCGGACGACCACGCGCTGTTCCTGGCGGCCGGCTTCGAGCCGGTGGCGGCGGACCCCTGGCAC
CCCGAGGCGCCGGAAGAGGACGAGACCTACCCCAACCGCGTCTACCGCCTGGACGATCTGCTGGCGGTGGACGCC
GACGGCCGCCTGGTGGTGAAGAACCTGGCAGGGGACGGCAACCGGCGTCACAAGAACAAGAACCGGCTCGCCTAC
CGGCGTTTTGAGAACTTCCTCGCCCGCAACGACCACCTGGAGCTCCGCATCCGTCCCTACGGATACGGCCCCGAC
GAGGCGAAGATGGCCAGGGGCGTGGTCGAGGGCTACTTCGAAGCCCGCCGGGCACAGGGCGAGGTCGTCGGCTCC
ACCCCCGAGGACTACACGGCGATCGTCACCCAGCGGCCCGGAGGCCGCAACGAACACGACTACTTCGCCTACCTC
GGCGTCCTGGCCCAGCAGGGCGGCGAGGAGGTGCCCGTGATGTTCTTCGCGGGCGAGCGCACCGCGCCGCACAGG
GCATCGCTGTACTGCACGATGTCGATGCGCTTCGCGGACCGTATGAGCGGCCTGTTCAAGGACGCCACGGGATTC
ACCGCGATTCCGCAGTACATCTGGCTGACCGTGTTCAAGAAACTGTGGGACCGCGGCATCCGTGAGGTCGACGCG
GGCGGATCCGAGGTCAAAGGCCTCGACGACCAGAAACGACAGTTGGGCGGACGGCCCGAAAAGACCCACTGGGTC
GTGGGCTGA

> SPR_52960

ATGCAATCGCGCACCGCGCACCAGGACTTCACACCGTTCGACGTCTGCGTCGTGGGGATGGGATACGTCGGCGTC
ACCCTCGCCGCCGCCCTCCTGTCCACCGGCAAGCGGGTGCTGGGCTACGAGAGCGACCCGGCCGTCGCGGGCGAC
CTGGCACAAGGCCGCCTGAGGCTGGCGGAACCCGGCGTCGCCGAGCTCATCGAGCGGGGAGCCGCCGACGGCACG
CTCGCGGTCACGGCCGACATCTGCGGGCACCGTCTGCCACCCGTCGTCGTCATCTGCGTCGGCACACCCATCGCG
CCGGGCGGCACGACGCCCGAACTCGGCCATCTGTCGGCCGCCGCCGAGGCGGTCGCCGCGGGAGCGGACGAGAAC
ACCCTGGTGATCGTGCGCAGCACCGTCCCCGTGGGCACCACCCGCGAACTGGTCCTGCCGGCCCTGGCCCGCCGC
GTTCCCCAGCCGCTGCTCGCCTTCTGCCCCGAGCGCACCATCCAGGGCAAGGCACTGGCGGAACTGCTGTCCCTG
CCGCAGATCGTCGGCGGTCTGACCGAAGAGGCAACGAAGCTGGCCGCGGGGTTCTTCACCACCGTCAGCGGGCGC
GTCGTCCCCGTATCCACACTGGAGGCCGCCGAGCTGGTCAAACTGGTCAACAACTGTCACACCGACCTCATCTAC
GGCTTCGGCAACGAAGTGGCCCTCATCGCCGAGAAGCTCGGCCTGGACGCCATGGAGGTCATCACCTCCGCCAAC
ACGGACTATCCGCGCCCCGATCTGAGCAGGCCCGGCTTCGTCGGCGGCAGCTGCCTCACCAAGGACCCCCACCTC
CTCGCCTACTCCCTCGCCCGCCACGACCACACCCCGCAGATGGTCATGGCGGCCCGCACCCTCAACGAGTCCATG
CCCCGCAGGGTCGGCGAGCGCGTGCTCGACGCGCTGCGCCGGGACGGCCAGGACCCGCAGCGGTCCACCGTCCTC
GTCTCCGGATTCGCCTACAAGGGCCGGCCCGAGACGGACGACCTGCGCGGCGCCCCCTGTGTGCCGCTGCTGGAG
TTCCTGCGCGGCAAGGTCGCCCGGGTGGTCGGGCACGACTTCGTGATCCCGCCCGAACGCATCGCGTCACTGGGC
GTACACCCCGTCACCCTCACCGAGGGGTTCACCGGCGCCCACGCCGCCATCCTGCTCAACGACCACGCCCGGTAC
GGCGAGCTGCCGGCCGACGAGCTCATCGCCCGGATGAGCCCGCCGGCGCTGGTGTACGACGCCTGGCGGGTGCTT
CCGCAGACGACGAAGACAATGAGGCTCGGCAGTGCCTGA

> SPR_52970

GTGCCTGAGAAGATCCTGGTCACCGGGGGAGCGGGGTTCATCGGGCTGCATCTGGCCGCGGAACTCGCCTCCCAC
CACGACGTCACCCTCCTCGACGACTTCAGCAGGGGCCGGCGCGACGCGTTGCTCGACTCCCTGCTCGACCGTGTC
ACGCTCGTCGAGCACGACCTGACCACCCCGATTCCCGACACCCTTCTGCCCGACGACTTCGACACCGTTCACCAT
CTGGCGGCCGTCGTCGGCGTAGTCCACTCCAACGAGGAACCGCAGCGGGTGCTGCGGACCAACCTTCTGTCCACC
GTGCACGTCCTGGACTGGTTCACCGGGCGGCCGGGCATGTCCGGCGCCACCTTCTGCTTCGCCTCCTCCAGCGAG
GCGTACGCGGGAAGCGTCGCGGCCGGTCTCGCCGCGCTGCCCACCGGCGAGGACGTCCCGCTGCTGGTCCCCGAC
CCCGAGGTACCCCGCTCCTCGTACGGCTTCAGCAAGATCGCCGGAGAGCTGCTCTGCCGCACCTACGCCCAAGTC
CACGGCTTCGCCCTGCGCATGGTGCGCTTCCACAACGTCTACGGACCGCGCATGGGCTACGAACACGTCATCCCG
CAGTTCATCGAGCGCGTCCTCGGCGGCGCCGACCCCTTCGCCGTCTACGGCGGCGACCAGACCCGGGCCTTCTGC
CACGTCGACGACGCCGTGGCCGCCCTGATGGCGCTGGCCGCGCTGCCCACCAAGGAGACCCTGCTCGTCAACATC
GGCAACGACCAGGAGGAAGTCCGCATGGACGACCTCGCCCGGATGGTCTTCGAGACCGCCGGGCGCCGGCCGCGC
ATCGCCGCGCACCCGGCACCGCCCCTGTCCCCGGTGCGCCGCCTGCCCGACCTGACCCTGCTGCGCGAACTCACC

GGATACCGCCCCTCGGTCGATCTGCGCGAGGGGCTGCGCCGCACCTACGCGTGGTACGCGCACGACCTCGCCTCC
CGTGGAGCGGGGCGGTGA

> SPR_52980

GTGAGCGGCATGCGGGTGCGGTACGTGCACCAGGGGTACTTCCCCGCGCGAGCGGGCGCCGAGCTGATGACCCGG
TCCCTCGCCGTGGCCATGAGCCGCCGCGGGCTGCGCGTGGGCCTGTACGGCGGCGAGGGGGACCCGGAGGACGAG
CGGCTGATGAAGGCCGCGGGGATCGGCGTCGAACCGCTGCCGGTCCGGGACGGCGAGGAGCGTGCGGCGGACCTG
GTGCACGCCGTCGACGCCTTCCAGCCCGAGAACATCAGGACCGGCCTGCGTCTGGCCCGGGCCTGGGGCGTGCCG
TTCGCCGTGACACCGGCCTCCGCGCCGGACGTGTGGCCGCACCGCGCCGCCGTACTGGAGGGCTGCCGCCGCGCG
GACGCCGTGTTCGTCCTGACCGACGCGGAGCGCGACATGCTGCGCGCCGAAGGCGTGGCGGACTCCGTCCTGCAC
CGGATCGGCCAGGGAGCGCACCTGCCGGGCACCGCCGACCCGGAGGGGTTCCGTGCCGCGCACGGCATCAGCGGG
CCCGTGGTGCTCTTCCTCGGCCGCAAGATGCGCTCCAAGGGGTACCGGGTGCTGCTGGAGGCGACGCGGCACGTC
TGGGCCCGCCACCCCGAGGCGCACTTCGTCTTCCTCGGCCCGCGCTGGGACGAGGACTGGGCGCAGTGGTTCGCC
GCCCACGCCGACCCCCGGATCACCGAACTGGACCGGGTGGACGAGGACACCAAACTCAGTGCCCTGGCGGCCTGC
GACCTGCTCTGCCTGCCCTCGACGGTCGACGTCTTCCCCCTGGTGTTCGTCGAGGCGTGGATGTGCGGCAAGCCC
GTGATCGGCTCCGCGTTCATGGGCAGCGCGGAAGTGATCGCGGACGGGCGGGACGGGCTGATCGTCCCGCCCGCC
GCCCGGCCGGTGGCGGACGCGGTCAGCCGCCTCCTCGCCGACCCGGCCGAACGGGCGCGGATGGGGCGCGAGGGC
CATGACAGGGCGCGCCGCGAGCTCACCTGGGACGCGGTCGCCGCACAGGTGCACCGCGTCTACACCGAACTGGTC
CCGGCCCGGAGTTGA

> SPR_52990

GTGAAAGCAGTCGTACTGGCCGGCGGCGAGGGGCGGCGCCTGAGGCCCGCCACCTTCACCGTCCCCAAGCCCCTC
ATCGAGGTCGACGGCACCCCGATCCTGCACATCATCCTGCGACAGCTCAGAAGCGCCGGATTCACCCAGGTCACG
CTCTCGCTGGGCTATCGCGCCCAGCTGATCGAGGCCAGCTTCGACGGCCCCCGCTGGGCGGGGCTCGACCTGCGC
TTCTCCCTGGAGCACGAGCCGCTGGGGACCGCCGGGCCGCTGGGTCTGCTCGCGCCGCTGGAGGACTCCACCCTG
GTGATGAACGCGGACCTGCTCACCGACATCGACTTCGCCGATCTGTTCAGGCGGCACAAGAAGTCCGAGGCGGCG
GCGACGATCGCCCTCGTTCCGCGCCATGTCGACCTCGCCCACGGTGTGGTGGAGCTCGACGGCGAGAACCGGGTG
GCCGGCTTCCGGGAGAAACCGCGTATGAGCTTCCTGGCCAGCAGCGGCATCTACGTCCTGGAACCCTCGGTCCTG
CGTCTGCTGCCCCGGCGGGCGCGTTACGACATGCCGGCGCTGCTCAAGGACGCCTCCGCCCGGGGCGAGCGCGTC
GAGGGCCATGTCCTCGACGCCGCCTGGCACGACATCGGCACCCCCGAACAGCTCGCGGCGGCCGACGCCGCGCTC
CGTGCCGACCGGGCGCGCTATCTGGGCGCCCTGGAGCGGCACGGCCCCGGTGCCGAAACGGACTCCACGGAGGAG
GCGGTGCGCGGATGA

> SPR_53000

ATGAGCGACTGGAAGATCCCCCTGTACGGGCCGAGCACCGGTGCGGCGGAGGCCCGCGCCGTCGCCGAGGTCCTG
CGCGCCAACTGGCTGTCCGTGGGCCGGGTGACCCAGGACTTCGAGGAGCGCTACGCGGCCGCGCTGGACGTCGAG
GACGCGATCGCGGTCAGCAGCGGGACCGCGGCGCTGCACCTGGCCGTCCTCGCTCTGGGGATCGGCCCCGGCGAC
GAGGTCGTCCTGCCGTCACTGAGCTTCGTCTCGGCGGCCGCCGTGGTCGCCCTGTGCGGGGCCACACCGGTGTTC
GCCGAGGTGAGCGGCGCCCATGACCTGTGTGTGGACCCGGCGGACGTGGCCGCCCGGATCACGTCGCGCACCCGT
GCCGTCGTGGCGGTGCACTACGGCGGCCATACGGCCGACCTGCCCGCCCTGACCGAACTGGCCCGGCGGCACGGT
CTGGCCCTCATCGAGGACGCCGCGCACGCTCCGGTCACCAAGACCGCGCACGGCGTCCTGGGCACGGTCGGCGAC
ATCGGCTGCTACAGCTTCTTCGCCACCAAGAACCTGGCCATGGGGGAGGGCGGCGCCGTCGTCGCCCGCGATCCC
GCGGTGCGCGCCCGGATCAGGCGGCTGCGCTCGCACGCGCTGACCGTGGGCGCCGAGCAACGCCACCGCGGCGGG
CCCTCGGCGTACGACGTGGACGGGTTCGGCCTCAACTACCGCCCCACCGAGATCGCCTGCGCCCTCGGTCGCGTC
CAGTTGGAAGCGCTGGCGGAGCGCCGCATCCTGCGCCAGGAGGCGGTGCGGGCCTATCGCACGCTGCTCTCGGGG
CTGCCCGGCCTGGAAGTGCCCTTCGCGGAGCGCCCGGTGGAGGAGGGCGCGCACCATCTGTTCCCGCTCGTGCTG
CCCGACGGCCTCGACCGCGAAGACCTCCAGGCGCAGCTGCGGGCGGCCGGTGTGCAGAGCGGTGTGCACTACCCG
CCCACCCATCTGTTCACCGCCTACCGCGAACGTTTCGGCACCCGCCCCGGCAGCCTCCCGGTCACCGAGAGCGTC
GCCGCCCGGCAACTGTCGCTGCCGCTGCACGCCGGGACCGGGCTCCAGGACGTCGCCCATGTGGCCGAGGCGGTG
AGCCGGACATGGCCGCCGCAGCGGTGA

> SPR_53010

ATGGCCGCCGCAGCGGTGAGCCCCGCCTGTGGCACCCGTCTCGTCGAGGGGGTGCTGCACCGCGACGGACGTCCC
CTGTTCTGCGTCGGCGTCAACTACTTCCCCTCGCGGGCGGGCTGCGACTACTGGCGGGACTGGGACCCGGCCGTC
CTCGACGCGGACTTTGCCCGTATGGCCGCGCTCGGCTTCAACACCGTGCGGATCTTCGTGTTCTGGGCCGACTTC
GAGCCCACCGAGGGCAGCTACGACCCGCGTATGACCGCCAGGCTGCGTGAACTGGCGGTCCTGGCGGAACGACAC
CACCTGCTGGTTCTGCCCTCGCTGCTGACCATATGGATGAACGGGCAGCTCTTCGACCCGCCGTGGCGGGCGGGC

CGGGACCTGTGGCGCGATCCGGTCATGGCCGAGCGGCAGCGCGCCTTCGTCGGGCACATCGCGGGCACGCTGCGC
CACGCGCCCAACATCCTCGCCTACGACATCGGCGACGAGATCCCGCACGTCGACCCGGCCGCGTCCCACTCGCTC
GGCGCGCACGAGGTACGGGCATGGTGGGCCGACCTCGCCGAAGCGATCCGGACGGCGGACCCGGGGGCTTTGGTC
CTCCAGGCCAACGAAGGCTCGGCGGTCTTCGGCGACCACGCCTTCCGGCCCGAGCACGCCCGGCCGCTGGACCTG
GTCGCGCTGCACGGCTTCCCGCTGTGGACCCCCTTCCACATCGAATCGGCCGCCGCCGAGAAGGCCACCGCCTAT
CTGCCCTACCTGGTGCGGCGCGGGCGGGCCCACGCCCCGGTGCTCGTCGACGAGATGGGCAGCTACGGCTGCGAC
GAGGCCACCGCGGCCCGCTATCTGCGCGCCGCGGCCCACAGTGCCTTCGCCGCGGGCGCCGTCGGTATCTGCGTG
TGGTGCTGGCAGGACTTCACCTCCGAGAGCAAGCCGTACGCACTGCGGCCCGGCGAACGCTTCGTGGGGCTGCTG
GACATGGACGGCCGCGAGAAGCCCGCCATGGACGCCTTCCGCGGCTTCGCGCGCCGGGTGACCGGGGAACTCGCC
GGTTTCCGCCCGCTGCCGGCCCGGGTGGGCGTGTTCCTGCCGGAGCGGGCCCGCGACCACGACGGCGGGTACCTG
GCGTCCKGGGCCGACACGGACGCCGCCGCCTTCTACGCGCACTGTCTGCTCCAACAGGCCCATCTGCCCTACGAG
TTCACGGGGACCGAGGACCTGGAGCGGTACGCGATGGTGATCTGCCCCTCCGTACGGCCCCTGCCCCTGGCCGCG
CAGCGACGGCTCGGCGAGTACACGGTCGCGGGCGGGGTGCTGCTGTACTCCACCGGCGATCCGCTGGGCTCGGCG
GGCCTGGAGGAGCTCTTCGGTGTGCGGATCCGGGACTTCACCCTGAACACGGCCGAGCAGGACCGCTTCACCTGG
GCCGACACCCCTTTCCCGGTGCACTGGCCCGCCGGACGTATCCCCGTGGTCGACAGCACGCTCGCCGAGACGCTC
GCCCGCTACCCCAACGGCGCGCCCGCGCTGGCCCGCCGGCAGCGGGGCGGCGGCGTCGCCTACTTCCTCAACGCA
CCGCTGGAAGCCCTGCTGAACGCGCCCTACCGGCTCCAGGAGGCGCCCTGGCACCGGCTGTACGCCGCGATCGGT
GAGAAGGAAGGCATACGGCCGGAGCTGTTCGCCGACGAGCCGCTGGTGGAGACCACCGTCCTGGCGCGCGGCGAC
GAACGCCGCGGTGTGGTCGTCAACCACGCCACCGCGCCCGTGACGACGACGGTGTACCGCACGTCCGCCGCCGGA
GATCCACGGGCGGTGCGGCACCTGAGCCTGGAGCCCAAGGGCGTCGCGCTCGTGCGCTGGCACGACGGACCGCCC
CCACACCCCGCCGCCACGGGCTCCACAGCCGGGCAGGACGGCGAGCAATGA

> SPR_53020

ATGATCATCGACGCCCATGTCCACGCCGGTGAGTACTACCGCCACTTCACCGCCCGCTTCGCCGACCAGATGATG
GCCACCACCGGTCTGCCCCCCGAGGCGCTGTCCGCGCCGGAGGACAAGCTCCTGGCGGAGATGGACGCCGCCGGG
GTGGACCACGCCTTCCTGCTCGCCTTCGAGGTGCGGCGCGTCGAGGGGTTCTCGGTGCCGAACACCTTCGTCGCC
GAGCTGTGCGCCCGCCATCCCCAGCGGTTCACCGGCTTCGCCTCGGTGGACGCGGGCCGGCCCGGCGCCGCCGAG
GAACTACGCCACTCCGTGACCGAACTGGGCCTGCGCGGGCTGAAGACGGCCCCCTGCTACCTGCGGATGTCCCCG
GCCGACCCCCGCTGGTTCGAGGTGTACCGGACCGCCCAGGACCTGGGCATCCCGGTCCTCGTCCACACCGGCTAC
ACCCCCGCCAAGAACGCCGACGCCCGCTTCTTCTCCCCCTTGCTGCTGGAGCCGGTAGCGAAGCGGTTCCCGGAA
CTGCGGCTGATCCTGGCGCACTTGGGCACCCCGTGGACGGCCCAGTGCATCGATCTGCTCGCCCGCCATCCCCAT
CTGTACGCGGACCTGTCGATCTTCGGCTCCTACCAGTCGCCCCCCACGGTGGCCGCGGCGCTCGCCCACGCCCGC
GAACGGGGAGTGCTGGACCGCCTCCTGTGGGGCACCGACTTCCCCTTCGCCACCATGTCCGCCGCCGTGGCCCGG
ATGACCCGTCTCACCACCGACGCCGCGCCGTGGCCGTCGGACAGCGCCCCGCTGACACCCGACGAACACCGGGCG
GTCATGGGCGGCACCGCCGCACTACTGACCCAGAAATGA

> SPR_53030

GTGGCCTTCATCCAGGGATACGAATTCGACGCCATGTCGCGCCGCCCCCTGTACGGGGACGTGACCCAGCGGCTC
GTCGACCTGTGCGCGTGCGCCCCCGGCAGCGTCGTCGCGGACGTCGGCTGCGGATCGGGCCTGGCCACCGAGCTG
CTGCTGGAACGGTTCCCGCAGCTCGGCTCCGTGGTGGGCATCGACCCCTCCGACCACGAACTGGCCCTGGCCAAG
GAGCGGCTGCGGCACCACCCCCGGGCCCGCCTGGTCACCGGCCGGGCCCAGGACGTGGGCGCGATCGTCGGGCCG
GTGGACGCCGTGGTGCTCAGCAACGTCATGCACCAGATCCCCAGGGCCGAGCGCCCGGCAGTGCTGCGCGGCTGC
CACGATCTGCTCGGACCCGGCGGCCGGCTCGCTCTGAACACGCTGTTCTACGAGGGCGCCGTGCTGCCCGAGAGC
CGCCGCTTCTACGCCGCGTGGCTGCACGCCACCCGGACGTGGCTGCGCAGCCGGGGCGGCGACCTGGTGCTGGAG
CGGGAGGCCCCCATCGCCCTTGAGACCCTGACCCCGGCCGACCACGAAGCCCTCATGGCCGAGGCGGGCTTCGCC
TCCGTACGGTCCGAGGAGGACGTCTACCCGTGGACCCTGGAGGACTGGCAGGCCCTGTGCGGATACTCGGTCTTC
GTCGAGGGCGCCACCGGACTGACGGATGTGGACCTCGGCTCCCGCGCGCTGACGGCCGCCCTGCACACCGTCTTC
CGCGAGCTGGGGCTCAGCAGCGTGCCCCGGCGCTGGCTGTTCGTCCACGGCACGAGGCAGGGGTGA

> SPR_53040

GTGATCGCGGCGACGCGGCCCGCGGTGGGCACCGCCACCGGGCCGCTGTCCCACGCCCAGCTGGGCATGTGGTTC
CTGGAGCGGCGCGCGGGAACCACCTTGTACGCGGAACCGATGACCTTCCGCCTGGTCGGCGAGGTGAACGTGCGC
GCCCTGCACGCCGCGGTCGAGGAGGTCGTGCGCCGCCACGACGCGCTGCGCACCCGTTTCACCGTGGTCGACGGA
GCTCCCCGCCAGCACGTCGTCCCCGACGCGACGGTACCGATGGCCCGGATCGACCTGGCGTATCTGCCCCCCGCC
CGGCGCCACCAGCGCGCCGAGCGGCTCCTGCGCGCCGAGCTCGGCCGGCCCATCGACCTCGCCCAGGGACCGCTG
GCGCGGGCCACCCTGATCCGGCTCGCCCCGCACGAACACGTCTTCCGCCTCACCCTGCACCATCTCGTCTGCGAC
GCCTGGTCGTGGTGGATGGTCGTCCTGCGCGAGCTGGAGGAGGGCTACACCCGCCGGGTCCGCGGCGGGCACGCG
CCCCTGGAGCCCGTCCGCACCCACTACGCGGACTTCGTCGACTGGCAGGGACGGTGGCTTCAGAGCCCCGCGCAC

ACCCGGCAGGTCGCCCACTGGCGCCAGGCCCTCGCCGGTGTCGCCCCACTGCCCGAGCTCACCCTGGCCCGTACG
ACCGCGCCGGTGCCGTACACCTCCTCCGCGACCGAGTGGACGGCCTTCCCGCAGCCGCTGCACGACCGGCTGCGT
GCCCTCGGCCGCCGCGAACACGTCACCCTGTACATGCTGCTGCTCACCGCCTTCACGCGGGTGCTGCGCCGCCAT
GTGCCGGTGGACGACATCCTGGTCGGCACCCGGGGCGGGTTCCGCAGCCGGCCCGAGTTCGAGAAGACCGTCGGC
CTGTTCGTCAACATGCTGCCCATCCGCACACGTGTCACCGAGGGCACCGGGTTCCGCGAGCTGCTGCCGCGGGTG
CGCGACACCCTCCTGGGCACGTATCTGAACCGGGACCTGCCCTTCGAACGCCTCGTGGCCGAGCTGGGCCTGCGC
CGGACCGGACCCCGGCCGGTGATCAACGTGTGCGTCTCCTTCCAGACCACCCCCGAGGTCATGCCGCGACTCCCG
GGCCTGGACGTCACCTTGCTCAACCACGACCCCTACTCGCCGTTCGACCTCGATCTCGGCTTCTACACCGAGGAC
GGCGGGCTGCGCGCCCTGATGATCTACAACCCGGCCCGGTACACCCGCGACGCGGTGCGGCACCTGCTCGACGCG
CTGCACCGGGAGCTGTGGGACGCCACCCGCGCACCCGGGCCCGTCCGCGCCGACGACGAGAAGGAGACCACCGTA
TGA

> SPR_53050

ATGAGGGCTTTGCAGTGGCACGGCGCCCACCGGGTGGCCCTGACCGACGACGCCCCGCCGCCGCGCGTCGAGACA
CCGGACGACGCGGTCGTACGGGTGGTCAGAAGCGCGATCTGCGGCACCGACCTGCACCCCTACCGGGGCGAGATA
GCCTCCTTCACTCCGGGCACCGTCACCGGCCACGAGTTCACCGGGGTCGTCGAGTCCGTCGGCTCCGGGGTGCGC
GGCCTGCGGCCGGGACAGCGGGTCGTGGCCTCGGACATCATCGCCTGTGGCCGCTGCTGGTGGTGCCGGCACGGC
GGGCACTACCAGTGCGAGCAGGTGGGACTGTTCGGCTACGACACGGTCGTGGGCGCGCGCCCGTTCGCGGGCGGC
CATGCCGAGCTGGTCCGTGTCCCGTTCGCCGATGTGGTCCTCTTCCCGATCCCCGACGACGTCGGCGACGAGCGG
GCGCTGCTCATCGGGGACGCGCTGGCCACCGGGTACGCCTGCGCCCTGGGCGGCGAGGTGCGCCCCGGCGACACC
GTCGCGGTGATCGGGTGCGGCGCGGTGGGTCTGCTGGCGGCCGAGGCCGCCCGGCTGCTGGGAGCCGCCCGGATC
CTGGCCGTCGACCCCGTCGAGGCCCGCCGCAAAGCGGCCGCCGAGCGCGGTGCGACGGCCCTGGCCCCCGGGGAC
GACCTCTCGGAGCGGGTGCGCGAGCTGACACGGGGACGCGGCGCCGACGCCGTCCTTGAGGCGGTCGGCACGGAC
GCCGCCCTGCTGAGCGCGCTGGAGATCGTACGGCCGCGCGGCGCGGTGTGCGCGGTCGGCGCCCACGCCTCCACG
GCGATGCCGCTGCCGACCATCCAGGCCTTCGCGAAGGAGGTGACCCTGCGGTTCGCCGTGGGCGACCCCATCGCC
ACGCGTGAGCCGCTGATGGAGCTCGTCCGCCAGGACCGCGTCGACCCCTCCTTCGTCATCACCCACCGCCTGCCG
CTCGCCGAAGCCCCCGAAGCCTTCCGGCTGTTCGACGCCGGCGAGGCCATCAAGGTGGTGCTCGTCCCGTGA

> SPR_53060

GTGACCACCCGGCTCCCCGCGTCACCGGCGCGCCCGCCCGCCCCGGATCCGCTCGACGGCGCCGTTCACCGTCTC
GTCGAGGAGCGGGCCGCGCGCACACCGGACGCCGACGCCGTGCTGTGCGGGACCGAACGCCTCAGCTATGCGCAG
TTGGACCTGCGGGCCGAGCGTCTCGCACGCCGCCTGCGCCGCCACGGAGTGGGCCCCGAAGTCCCGGTGGGTGTA
CATCTGACGCGCTCGCCCGAACTGGCGGTGGCCTGTCTGGCGGTGCTCAAGGCGGGCGGGGCCTGCCTCCCGCTG
GACCCCGCCCATCCGCCCGGCCGTCTGGCCGACGCCCTGCGCGACGCCGGTGCCCCGCTGGTCCTGAGCCGGCGC
GCCCTGGCGGCGGGCCTGGGCCCGCACACGGCAGCCGTCCTCTACACGGACTCGGAGGCTGACGACTCGTCAGCC
CACGAGCGTCCGGCGCCCGGCCCGGCGGCAGGATCGCCGTCACCGGGTGCCCCGGCAGCCGAGGACCGGCGAGCG
GAGGAGCGGCCGGGCGACGCGGGGGCACCCCCCGCCCCGCCTCTTGCACGCCGCCTCGCCTGGGTCGCGTACACC
TCGGGATCCACCGGCCGACCCAAGCCCGTCGGCCTCGAACACGGGCCGCTCGCCAACCTCGCCGTGCAGATCGGC
CGCCGTCTGGACCTGGGCCCCGACGACCGCGTGCTGCAGTTCGCCTCCATCGGCTTCTCCGTCGCGGCCGAGGAG
ATGTTCTCGACCTGGGCCGCCGGAGCCTGTCTCGTCATCGATCCCGACGACACCCTCGCCGACAGCGCGGGGCTG
CTCACCGCCGTGGACAAGTACGCCGTCACCGTCCTCCAGCTCACCCCGTCCCTGTGGTACGAGTGGCTGCGCGAG
CTGAGCCGGGACGGCACTCTGCGCCCGCCCGCCTCGCTGCGGCTGCTCGTGGTGGGCAGCGAGCAGGCCGCACCC
GACCGGGCGGCCGACTGGCTGGCCACCGGGGTGCGCCTGGTGCACGAGTACGGTGCCACCGAGGGCACGGTCTCC
CAGCTGCTGTACGAACCCGACGTGTCACCGGCCGAGTTGAGGACCTGGCCGCGTCTGCCCGTCGGTACGCCGTTG
CCGGGTGTCCGCGTCCACATCCTCGACGCGCGGCGCCGACCGGTCCCCCCGGGCGAGCCCGGCGAGCTCCACCTC
GCCGGAGACACCCTGGCGCGCGGCTACCTGGGGCAGCCCGAGCTCACGGCCGAGCGCTTCCTGCCCGACCCGTTC
GCCGATCGGCCCGGCGCGCGCATGTACCGCACGGGCGACCTCGCACGACAGCGTGCGGACGGCACCATCGAGTTC
CTGGGTCGCGTCGACCACCAGATCACGCTGCGCGGCCTGCGTATCGAGCCGGGCGAGGTCGAGTCGGCGATCGGC
CGGTACCCCGGCGTCACCGAGAGCGCCGTGCTGGCCCGCACCACCGCACGGGGCGACGATCAGCTGTGCGCGTGC
GTGGTGTGGGAGAAGGAGCGCGACGAGGCCGGGCTGCGCGCCCACCTGCGCGCCGCGCTCCCGCCCGCGCTCGTC
CCCGCCCGCCTGCTGGCCCTGCCCGACCTGCCGCTGACCCCCCATGGCAAGGTGGACCGCCAGGCCCTGCGGGCC
CTGCGCTGGGAACCCGACCCCGTCGATTCACGGGGCGAGGCGCCGCGCACCGCGCTGGAGGCCGCCCTCGCCGCG
CTCTGGGCCTGGACCCTCGACGTGTCCCGGATCGGCCTCGACGACAGCCTCTTCGACCGGGGCGGCGACTCCCTG
ACCGCCACCCGGCTCGCGGCCCGACTGCGCGACGTCCTGATGGCGGACGTACGGCAGCGGACCGTGTTCGACGCG
CCGACCGTACGGGCCATGGCCGAGGTGATCTCGCGACAGCGCCGCCGCGCTCGCGCCGACGAGCCCGTGACGGCG
CCGGACGCGTACGAGGACGCCCCACTGACGTCGGCGCAGTGGCGCATGTGGCTGCACCACCGCAGATCCCCGACG
AGCGCCGCGTATCACGAACCCGTGGCCCTGCGGCTACGGGGCCCGCTCGACCCCGACCACCTCGTACGGGCACTG
CGGCGGACGGTGGGGCGGCATGCGGCTCTGCGCACCACCCTCGCCACGGCGGCGGGGCAGCCGCTCCAGCGGGTC
GCGCCGCCCGCCACCGCCGAGGCGTTCCCCCTGGCCCGCTTGGACCTGAGCGGCCACGCGGAGGGTGAACTTGAC

CGAGTGGTCGAGGAGTTGGCCGTGGCGCCGTTCGACCTGCGGCGTGGCCCCCTGATCCGCGCGCACCTGCTGCGG
GTCGCGGCCGACGAGCACGTCCTCGTGCTGGTCCTGCACCACATCGTCTTCGACGGATGGTCGATGGACGTGCTC
TTCCGGGATCTGGCGGCCTTCCACGACGAGGCCGCGACCGGAGTCGCCGCCGCGCCGGCGCCCCTGCCCTCCTCG
CCGCCCGCGCGCCCTCCGCTGCCCGCGGACGACCCGGGCGCCGAGCGCCGCGCCCAGGAACAGCGCGCGTACTGG
AAAAAGCAGCTGGCGGGGGCACCGCCCGTCACACGGCTGCCGACCGGCCGACCCACGTCCCGGCAGGGCACGGTG
CACTGCCCGTTCACCGTCCCGGCGGCCACCGCGGCGCGGCTGCGCGAGCTGGCGGCGGAGGAAGGCGCCACCCCG
TTCATGGTGATCCTCGCCGTCTTCACGGCAGTCGCCCACCAGAGCTCCGGCGCCCGCGACATCGTGGTGGGCACG
CTGCTGGCAGGCCGCGACGACCCGGACACCGCCGACCTGATCGGTCTGTTCACCCACACCGTTCCGCTGCGCACA
CAGGTGCCCGAGGGGACGACGGTCCGTCAGCTGGTGCGGCGGGTACGTGAGACGGTTCTGGGCGCGGTGGCCCAT
CAGGAGGTGCCGTTCGAGGACATCGTGGCGCACCTGGGACCGCCACGGGATCCCCGGCACAACCCGGTGTTCCAG
ATGGTGTTCTGCCACGGCGCCGCGTCCGCCCGCGCCCCGCAGCTGCGCGGCCTGAGCGTGACCCCCGTGGAGGTC
GCCGTGCCCTTCGCCAAGTTCGACGTCACCGTCATGGTCGACGAGGACCGGCACGGCGGATACGCCATCGACCTC
GCCTACGACCTCGCGCTCTACGACGAGACGACCGCTCACCGCCACACCGCGGCGTTCCGGTCCCTGCTGGAAGCC
GCCGCGGACCGACCGGAAGCAGGGGTGTGA

> SPR_53070

GTGTCCACAGAATCGACCGCCCGCACCACCGAATCCCCGGCCGCCGCCGCCCCGACGAACGCCGACGCACCCCCG
ACGCCCGAGGCCGCGTCGGTCCGGGTGGCCGAATCGGTTCGCCGGATCTGGGCCGAACTCCTGCAGATCGACGTC
GAAGCGATCGACGTCCGCCACAGCGACTTCTTCGAGCTCGGTGGCTATTCACTGCTCGCGCTCCAGGCCATCGGC
AGGCTGCTGGAGGAGCGGGGCTTCGACGAGTTCGAGGCGGCGGAGCTGGAGGGCGCGTTGCTCAACCGCCTCTTC
GAGGAGCCCACACCGTTGGCCCAGGCGGAGTGCCTGCAGTCCGCGCTGGCCGCCGGCGGGGCACCGCGCGCGTGA

> SPR_53080

ATGCACCACCGCTCCAAGGCCCTGTACGGGGTTTTCGTCACCGTCCTCGGACTCACGCTGACCGCCTGCGGCGGT
GACGGAGGAGGCGGGAAGTCCGGTGTTTCCGAGGGCGGGGCGGCGAAGGGCGAGACCGGCCAGGAAACCCCGTCC
GCGAGCCCGTCCGAGAAGCCGGAGTGGGGCCCCGCGCTGGCGATGGGCCAGCCGGCCCCGAAGCTGTACGAACCC
CTGAACACCAGCGGCGGGAAGTTCCAGGTAACCGCCAAGAAGATCGTCAAGGGAACTCCCCAGCAGGCAAAGGAC
CTGAACTGGGACACGCTCGCCCCGGGCGAACTCAAGGGCAGTACGCCGTACTTCGTGTACTTCACGTACACGCTC
AAGGAAGGCAAGCCGCAGGTCGCCAACCCTGACCTCGGGGCGCATGCCGCCGCCCTCGACATGAAGGGCGTGAAA
GTCGCGGAGCGGCCGGTCGTCCACTCGGGATTCGTCGACGGCGGTTGCGAAGTCCCACCCGTATACATGGGCTGG
GACATCGGCGAGACCCACACCCTCTGCACGGTCTTCGCCGGGGACGAGGCGCATCCCCCTGCCCGACTGGCCTGG
GGTACGGACATAGAAACCGACACGGACTACATGAAGGCCGCATCGTGGAAGTGGACCACCCAGTGA

> SPR_53090

ATGCCCGGTATGCGTGCGTTCTTCACCGTTGATCAGCGGCGGTTGTCGTTCGTGGACTTCGGCGGCCCGGGGCTC
CCGTTGCTGGCGTTTCACGGGCATTACAACGAGGCATCGGTGTTCGCCCCGCTGGCTGAAGCGCTGGCGCCGCAG
TGGCGGGTGATTGCCCTGGACCAGCGTGGGCATGGGGAATCCGACCGCGCCCCGAGCTACCAGCGGTCCGAGTAC
GGCGCTGACATCGCCGCCTTCCACCGCCATCTGGGACTCGGTCCGGTGGCGGTACTCGGGCATTCCCTGGGCGGC
GTGAACGCCTATCAGTACGCCGTCCGGCACGCGGACCGGGTGACAGCCCTGATCGTCGAGGACGTCGGCGCCGTG
GTGGATGCCGACTGGTCGCTCACCACGCAACTGCCTCGCAGAGCACCCTCCCGTCGGGCACTGGCCTCGGCTCTC
GGCCCGCTGGCGCCCTATCTCGAATGCACCTTCCGTCAGTTCGGTGACGGCTGGGGATTCTCCTTCGACATCAAG
GACACGGTGCTGTCCCACGAAGCCCTCACCGGGGACTACTGGGGCGACTGGGAGGCGGTGCCCTGCCCGACCCTG
CTGATCCGGGGAAGGGACAGCGTCGTCCTGGCGTCGGACCACGCCCGGGAGATGATAGCCCGCCGGGCAGGTCAG
GCCGAGCTGGTCGAACTTCCGGCCGGACACGTCGTGCACTGCGACGCTCCCGACGAGTTCGCCGCCGCTGTGCGC
ACCTTCCTGTCCCAACTGCCCGAGAGGTGA

APPENDIX C

The nucleotide sequences of the DNA fragments used to make each of the knock-out constructs within the genes they target are provided in this appendix. The black text represents the regions of the target gene within the wild-type strain, while the blue and green text represents the sequences of the target gene that were integrated into the genome of each mutant strain via homologous recombination. The position of the pOJ260 sequence within the target gene of each mutant strain is also provided.

> *S. polyantibioticus* ΔCIN

CCGCATCACCGGCCTGTACGTGGCCCCGCCGATCGTGCTCGCCCTCGCCAAGCACCCGGCGGTCG//CCCAGTAC
GACCTCTCCTCCGTGGAGTACGTCGTCAGCGCCGCCGCCCCGCTGGACGCCGGGCTCGCCGCCGCCTGCTCGGCC
CGCCTCAAGGTCCCGCCGGTGCGCCAGGCGTACGGCATGACCGAGCTCTCGCCCGGCACCCACGCCGTCCCCCTG
TCGGCACAGAACCCGCCGCCCGGCACCGTCGGCACACTCCTGCCCGGCACCGAGATGCGGCTGCTCTGCCTCGAC
GACCCCGGCCGCGACGCCGCCCCCGGCGAGCGGGGCGAGATCGCCATCCGGGGCCCGCAGGTCATGAAGGGCTAC
CTCGGCCGCCCCGAGGCCACCGCCGAGATGATCGACGCCGAGGGCTGGGTGCACACCGGCGACGTCGGCCACGTA
GACGCCGACGGCTGGCTGTTCGTCGTCGACCGCGTCAAGGAGCTCATCAAGTACAAGGGCTTCCAGGTCGCCCCC
GCCGAACTGGAGGCGCTGCTGCTCACCCATGACGCCGTCGCGGACGCGGCGGTGGTCGGGGTGTACGACGAGGAC
GGCACCGAGGTGCCGTGCGCCTATGTCGTACGGGCACCCGGCGCCCCGGACCTCACCGCCGAGGACGTCATGGCG
TACGTGGCCGAACGCGTCGCCCCGTACAAGAAGATCCGCCGCCGC//pOJ260//GCCCAGTACGACCTCTCCTC
CGTGGAGTACGTCGTCAGCGCCGCCGCCCCGCTGGACGCCGGGCTCGCCGCCGCCTGCTCGGCCCGCCTCAAGGT
CCCGCCGGTGCGCCAGGCGTACGGCATGACCGAGCTCTCGCCCGGCACCCACGCCGTCCCCCTGTCGGCACAGAA
CCCGCCGCCCGGCACCGTCGGCACACTCCTGCCCGGCACCGAGATGCGGCTGCTCTGCCTCGACGACCCCGGCCG
CGACGCCGCCCCCGGCGAGCGGGGCGAGATCGCCATCCGGGGCCCGCAGGTCATGAAGGGCTACCTCGGCCGCCC
CGAGGCCACCGCCGAGATGATCGACGCCGAGGGCTGGGTGCACACCGGCGACGTCGGCCACGTAGACGCCGACGG
CTGGCTGTTCGTCGTCGACCGCGTCAAGGAGCTCATCAAGTACAAGGGCTTCCAGGTCGCCCCCGCCGAACTGGA
GGCGCTGCTGCTCACCCATGACGCCGTCGCGGACGCGGCGGTGGTCGGGGTGTACGACGAGGACGGCACCGAGGT
GCCGTGCGCCTATGTCGTACGGGCACCCGGCGCCCCGGACCTCACCGCCGAGGACGTCATGGCGTACGTGGCCGA
ACGCGTCGCCCCGTACAAGAAGATCCGCCG//GGTGGAGTTCGTCGCCGGGGTGCCGCGCGCGGCCACCGGGAAG

> *S. polyantibioticus* ΔLAC

GGAGCACGGTCTGCGCGGCACCCCGTACGGGCACTTCGGCGACGGCTGTGTCCACGTCCGGATCG//ACTTCGAC
CTGCTGAGTACGGCGGGTGTGGCGCGCTTCCGCCGGTTCTCCGAGGAGCTGGCGGACCTGGTGGCCGCGCACGGC
GGCTCGCTCTCCGGCGAGCACGGTGACGGGCAGGCGCGGGCCGAGCTGCTGCCGAGGATGTACGGGGACGAACTG
GTCGGGCTGTTCGGGCGGTTCAAGGACGTGTGGGACCCGTCGGGCCTGCTCAACCCCGGGATGCTGGCCCGTCCG
GCCCGGCTGGACGAGAACCTGCGCTTCGCGGTGCTGCCGAAGGGGCCGGTGGAGGTGGAGTTCGGCTATCCGCAG
GACGGCGGGGACTTCTCGGCGGCGGTACGGCGGTGTGTCGGGGTGGCCAAGTGCCGTACTGTGTCCGGGACTTCG
GGTTCGTCGGTGATGTGCCCGTCCTTCCGGGCCACCGGCGAGGAGGCGCACTCGACGCGGGGCCGGGCGCGGCTG
CTGCACGAGATGCTGGCGGGCGAGGTCGTCACCGGCGGCTGGCGCTCGCCGGAGGTGCGTGACGCGCTCGATCTG
TGCCTGTCCTGCAAGGGCTGCCGCAGCGACTGCCCGGTCGGGGTGGACATGGCCACGTACAAGGCGGAGTTCCTG
CCGC//pOJ260//GACTTCGACCTGCTGAGTACGGCGGGTGTGGCGCGCTTCCGCCGGTTCTCCGAGGAGCTGG
CGGACCTGGTGGCCGCGCACGGCGGCTCGCTCTCCGGCGAGCACGGTGACGGGCAGGCGCGGGCCGAGCTGCTGC
CGAGGATGTACGGGGACGAACTGGTCGGGCTGTTCGGGCGGTTCAAGGACGTGTGGGACCCGTCGGGCCTGCTCA
ACCCCGGGATGCTGGCCCGTCCGGCCCGGCTGGACGAGAACCTGCGCTTCGCGGTGCTGCCGAAGGGGCCGGTGG
AGGTGGAGTTCGGCTATCCGCAGGACGGCGGGGACTTCTCGGCGGCGGTACGGCGGTGTGTCGGGGTGGCCAAGT

GCCGTACTGTGTCCGGGACTTCGGGTTCGTCGGTGATGTGCCCGTCCTTCCGGGCCACCGGCGAGGAGGCGCACT
CGACGCGGGGCCGGGCGCGGCTGCTGCACGAGATGCTGGCGGGCGAGGTCGTCACCGGCGGCTGGCGCTCGCCGG
AGGTGCGTGACGCGCTCGATCTGTGCCTGTCCTGCAAGGGCTGCCGCAGCGACTGCCCGGTCGGGGTGGACATGG
CCACGTACAAGGCGGAGTTCCTG//CACCACCACTACCGGGGCCGGCTGCGCCCCGCCTCCCACTACGCGATGGG

> *S. polyantibioticus* ΔTHI

ATGCCCGG//TATGCGTGCGTTCTTCACCGTTGATCAGCGGCGGTTGTCGTTCGTGGACTTCGGCGGCCCGGGGC
TCCCGTTGCTGGCGTTTCACGGGCATTACAACGAGGCATCGGTGTTCGCCCCGCTGGCTGAAGCGCTGGCGCCGC
AGTGGCGGGTGATTGCCCTGGACCAGCGTGGGCATGGGGAATCCGACCGCGCCCCGAGCTACCAGCGGTCCGAGT
ACGGCGCTGACATCGCCGCCTTCCACCGCCATCTGGGACTCGGTCCGGTGGCGGTACTCGGGCATTCCCTGGGCG
GCGTGAACGCCTATCAGTACGCCGTCCGGCACGCGGACCGGGTGACAGCCCTGATCGTCGAGGACGTCGGCGCCG
TGGTGGATGCCGACTGGTCGCTCACCACGCAACTGCCTCGCAGAGCACCCTCCCGTCGGGCACTGGCCTCGGCTC
TCGGCCCGCTGGCGCCCTATCTCGAATGCACCTTCCGC//pOJ260//GTATGCGTGCGTTCTTCACCGTTGATC
AGCGGCGGTTGTCGTTCGTGGACTTCGGCGGCCCGGGGCTCCCGTTGCTGGCGTTTCACGGGCATTACAACGAGG
CATCGGTGTTCGCCCCGCTGGCTGAAGCGCTGGCGCCGCAGTGGCGGGTGATTGCCCTGGACCAGCGTGGGCATG
GGGAATCCGACCGCGCCCCGAGCTACCAGCGGTCCGAGTACGGCGCTGACATCGCCGCCTTCCACCGCCATCTGG
GACTCGGTCCGGTGGCGGTACTCGGGCATTCCCTGGGCGGCGTGAACGCCTATCAGTACGCCGTCCGGCACGCGG
ACCGGGGTGACAGCCCTGATCGTCGAGGACGTCGGCGCCGTGGTGGATGCCGACTGGTCGCTCACCACGCAACTGC
CTCGCAGAGCACCCTCCCGTCGGGCACTGGCCTCGGCTCTCGGCCCGCTGGCGCCCTATCTCGAATGCACCTT//
CCGTCAGTTCGGTGACGGCTGGGGATTCTCCTTCGACATCAAGGACACGGTGCTGTCCCACGAAGCCCTCACC

> *S. polyantibioticus* ΔA99

GGAGCGGCCGGGCGACGCGGGGGCACCCCCCGCCCCGCCTCTTGCACGCCGCCTCGCCTGGGTCGC//GTACACC
TCGGGATCCACCGGCCGACCCAAGCCCGTCGGCCTCGAACACGGGCCGCTCGCCAACCTCGCCGTGCAGATCGGC
CGCCGTCTGGACCTGGGCCCCGACGACCGCGTGCTGCAGTTCGCCTCCATCGGCTTCTCCGTCGCGGCCGAGGAG
ATGTTCTCGACCTGGGCCGCCGGAGCCTGTCTCGTCATCGATCCCGACGACACCCTCGCCGACAGCGCGGGGCTG
CTCACCGCCGTGGACAAGTACGCCGTCACCGTCCTCCAGCTCACCCCGTCCCTGTGGTACGAGTGGCTGCGCGAG
CTGAGCCGGGACGGCACTCTGCGCCCGCCCGCCTCGCTGCGGCTGCTCGTGGTGGGCAGCGAGCAGGCCGCACCC
GACCGGGCGGCCGACTGGCTGGCCACCGGGGTGCGCCTGGTGCACGAGTACGGTGCCACCGAGGGCACGGTCTCC
CAGCTGCTGTACGAACCCGACGTGTCACCGGCCGAGTTGAGGACCTGGCCGCGTCTGCCCGTCGGTACGCCGTTG
CCGGGTGTCCGCGTCCACATCCTCGACGCGCGGCGCCGACCGGTCCCCCCGGGCGAGCCCGGCGAGCTCCACCTC
GCCGGAGACACCCTGGCGCGCGGCTACCTGGGGCAGCCCGAGCTCACGGCCGAGCGCTTCCTGCCCGACCCGTTC
GCCGATCGGCCCGGCGCGCGCATGTACCGCACGGGCGAG//pOJ260//TTAAGTACACCTCGGGATCCACCGGC
CGACCCAAGCCCGTCGGCCTCGAACACGGGCCGCTCGCCAACCTCGCCGTGCAGATCGGCCGCCGTCTGGACCTG
GGCCCCGACGACCGCGTGCTGCAGTTCGCCTCCATCGGCTTCTCCGTCGCGGCCGAGGAGATGTTCTCGACCTGG
GCCGCCGGAGCCTGTCTCGTCATCGATCCCGACGACACCCTCGCCGACAGCGCGGGGCTGCTCACCGCCGTGGAC
AAGTACGCCGTCACCGTCCTCCAGCTCACCCCGTCCCTGTGGTACGAGTGGCTGCGCGAGCTGAGCCGGGACGGC
ACTCTGCGCCCGCCCGCCTCGCTGCGGCTGCTCGTGGTGGGCAGCGAGCAGGCCGCACCCGACCGGGCGGCCGAC
TGGCTGGCCACCGGGGTGCGCCTGGTGCACGAGTACGGTGCCACCGAGGGCACGGTCTCCCAGCTGCTGTACGAA
CCCGACGTGTCACCGGCCGAGTTGAGGACCTGGCCGCGTCTGCCCGTCGGTACGCCGTTGCCGGGTGTCCGCGTC
CACATCCTCGACGCGCGGCGCCGACCGGTCCCCCCGGGCGAGCCCGGCGAGCTCCACCTCGCCGGAGACACCCTG
GCGCGCGGCTACCTGGGGCAGCCCGAGCTCACGGCCGAGCGCTTCCTGCCCGACCCGTTCGCCGATCGGCCCGGC
GCGCGCATGTACCGCACGGGCGA//CCTCGCACGACAGCGTGCGGACGGCACCATCGAGTTCCTGGGTCGCGTCG

> *S. polyantibioticus* ΔCYC

GCGCGCCCTGCACGCCGCGGTCGAGGAGGTCGTGCGCCGCCACGACGCGCTGCGCACCCGTTTCAC//CGTGGTC
GACGGAGCTCCCCGCCAGCACGTCGTCCCCGACGCGACGGTACCGATGGCCCGGATCGACCTGGCGTATCTGCCC
CCCGCCCGGCGCCACCAGCGCGCCGAGCGGCTCCTGCGCGCCGAGCTCGGCCGGCCCATCGACCTCGCCCAGGGA
CCGCTGGCGCGGGCCACCCTGATCCGGCTCGCCCCGCACGAACACGTCTTCCGCCTCACCCTGCACCATCTCGTC
TGCGACGCCTGGTCGTGGTGGATGGTCGTCCTGCGCGAGCTGGAGGAGGGCTACACCCGCCGGGTCCGCGGCGGG
CACGCGCCCCTGGAGCCCGTCCGCACCCACTACGCGGACTTCGTCGACTGGCAGGGACGGTGGCTTCAGAGCCCC
GCGCACACCCGGCAGGTCGCCCACTGGCGCCAGGCCCTCGCCGGTGTCGCCCCACTGCCCGAGCTCACCCTGGCC
CGTACGACCGCGCCGGTGCCGTACACCTCCTCCGCGACCGAGTGGACGGCCTTCCCGCAGCCGCTGCACGACCGG
CTGCGTGCCCTCGGCCGCCGCGAACACGTCACCCTGTACATGCTGCTGCTCACCGCCTTCACGCGGGTGCTGCGC
CGCCATGTGCCGGTGGACGACATCCTGGTCGGCACCCGGGGCGGGTTCCGCAGCCGGCCCGAGTTCGAGAAGACC
GTCGGCCTGTTCGTCAACATGCTGCCCATCCGCACACGTGTCACCGAGGGCACCGGGTTCCGCGAGCTGCTGCCG
CGGGTGCGCGACACCCTCCTGGGCACGTATCTGAACCGGGACCTGCCCTTCGAACGCCTCGTGGCCGAGCTGGGC
CTGCGCCGGACCGGACCCCGGCCGGTGATCAACGTGTGCGTCTCCTTCCAGACCACCCCCGAGGTCATGCCGCGA

CTCCCGGGCCTGGACGTCACCTTGCTCAACCACGACCCCTACTCGCCGTTCGACCTCGATCTCGGCTTCTACACC
G//pOJ260//TTAACGTGGTCGACGGAGCTCCCCGCCAGCACGTCGTCCCCGACGCGACGGTACCGATGGCCCG
GATCGACCTGGCGTATCTGCCCCCCGCCCGGCGCCACCAGCGCGCCGAGCGGCTCCTGCGCGCCGAGCTCGGCCG
GCCCATCGACCTCGCCCAGGGACCGCTGGCGCGGGCCACCCTGATCCGGCTCGCCCCGCACGAACACGTCTTCCG
CCTCACCCTGCACCATCTCGTCTGCGACGCCTGGTCGTGGTGGATGGTCGTCCTGCGCGAGCTGGAGGAGGGCTA
CACCCGCCGGGTCCGCGGCGGGCACGCGCCCCTGGAGCCCGTCCGCACCCACTACGCGGACTTCGTCGACTGGCA
GGGACGGTGGCTTCAGAGCCCCGCGCACACCCGGCAGGTCGCCCACTGGCGCCAGGCCCTCGCCGGTGTCGCCCC
ACTGCCCGAGCTCACCCTGGCCCGTACGACCGCGCCGGTGCCGTACACCTCCTCCGCGACCGAGTGGACGGCCTT
CCCGCAGCCGCTGCACGACCGGCTGCGTGCCCTCGGCCGCCGCGAACACGTCACCCTGTACATGCTGCTGCTCAC
CGCCTTCACGCGGGTGCTGCGCCGCCATGTGCCGGTGGACGACATCCTGGTCGGCACCCGGGGCGGGTTCCGCAG
CCGGCCCGAGTTCGAGAAGACCGTCGGCCTGTTCGTCAACATGCTGCCCATCCGCACACGTGTCACCGAGGGCAC
CGGGTTCCGCGAGCTGCTGCCGCGGGTGCGCGACACCCTCCTGGGCACGTATCTGAACCGGGACCTGCCCTTCGA
ACGCCTCGTGGCCGAGCTGGGCCTGCGCCGGACCGGACCCCGGCCGGTGATCAACGTGTGCGTCTCCTTCCAGAC
CACCCCCGAGGTCATGCCGCGACTCCCGGGCCTGGACGTCACCTTGCTCAACCACGACCCCTACTCGCCGTTCGA
CCTCGATCTCGGCTTCTACACCG//GAGGACGGCGGGCTGCGCGCCCTGATGATCTACAACCCGGCCCGGTACAC

> *S. polyantibioticus* ΔACY

ATGCCCTCGACCACTGAATCCGCCGCCACG//CTGTGGGAGCTCATCGACCGGGCCGTACGCCTCTCCCCCGCGG
CCACCGCCGTACGCCAGGGAACGCGGGCCCTCACCTTCCGCGAACTCGCCGACCGCGTCGAGACCACAGCACGCC
GCCTGCGCACGCGCCTGCCCGCCGACGGCGGAGGAACCGTGGGCCTGCTCTTCGAGAACACCGTGGAGAGCACGG
TGGCCTTCCTGGGCGCGCTGTACGCGGGCGTACCCCTCACGCCCCTCGAACCCGACAGCACCGAGCCCCATCTCC
TGGGCGTGCACCGGGACTTGGGACCCCTGCACCTCGTCGGTCGGCAGGCGAGGCTCAGCACACTGGCCCCTCCCG
CTGCCACCTCACTCGCCGCCCGGTGGCAGGGCGGTGTCCTCATCGACGTGGACGAGCTCACCGCCCGCCCCGGCG
CGGCGGCCCCGTCGGCGCCGCTGCCCGCGCCGCCGCCCGACGCCCCCGCGCTCTACCAGTACACCTCCGGCTCGA
CCGGCGAGCCGAGGGCCGCCGTGCACTCACAACACGACCTGGTCCGCGGTGGCGAGATCTACGCCCGCACCTACG
GCATCACCCCGGCCGACCGGATTCTCGCGGCCGTACCCCTGCTGCACTCCTTCGGCATGGTCGCCGCGCTCGCCA
CCGCGCTGCACGCCCGCGCGGAGCTCGTCCTGCTCGGCCGGTTCGCACCGGCCGAAATGCTCCGGGCACTGCACC
AGCACGCCTGCACCATCGTGGTGGGCACCCCGCTCGCCTATGACCTGGCGGCCCGTTCGGCGGCCTCGCGCAGCG
AGACGTCCCGCCCGGGCGACACCGTACGGCTCTGTCTGTCCTCCGGGGCGGCCCTGCCCCCGGCCGTGGCGGACC
GCTTCGCCCAGCACTGCGGCCCGGCCGTCCAGCAGGTCTACGGAAGCACCGAGG//pOJ206//TTAACTGTGGG
AGCTCATCGACCGGGCCGTACGCCTCTCCCCCGCGGCCACCGCCGTACGCCAGGGAACGCGGGCCCTCACCTTCC
GCGAACTCGCCGACCGCGTCGAGACCACAGCACGCCGCCTGCGCACGCGCCTGCCCGCCGACGGCGGAGGAACCG
TGGGCCTGCTCTTCGAGAACACCGTGGAGAGCACGGTGGCCTTCCTGGGCGCGCTGTACGCGGGCGTACCCCTCA
CGCCCCTCGAACCCGACAGCACCGAGCCCCATCTCCTGGGCGTGCACCGGGACTTGGGACCCCTGCACCTCGTCG
GTCGGCAGGCGAGGCTCAGCACACTGGCCCCTCCCGCTGCCACCTCACTCGCCGCCCGGTGGCAGGGCGGTGTCC
TCATCGACGTGGACGAGCTCACCGCCCGCCCCGGCGCGGCGGCCCCGTCGGCGCCGCTGCCCGCGCCGCCGCCCG
ACGCCCCCGCGCTCTACCAGTACACCTCCGGCTCGACCGGCGAGCCGAGGGCCGCCGTGCACTCACAACACGACC
TGGTCCGCGGTGGCGAGATCTACGCCCGCACCTACGGCATCACCCCGGCCGACCGGATTCTCGCGGCCGTACCCC
TGCTGCACTCCTTCGGCATGGTCGCCGCGCTCGCCACCGCGCTGCACGCCCGCGCGGAGCTCGTCCTGCTCGGCC
GGTTCGCACCGGCCGAAATGCTCCGGGCACTGCACCAGCACGCCTGCACCATCGTGGTGGGCACCCCGCTCGCCT
ATGACCTGGCGGCCCGTTCGGCGGCCTCGCGCAGCGAGACGTCCCGCCCGGGCGACACCGTACGGCTCTGTCTGT
CCTCCGGGGCGGCCCTGCCCCCGGCCGTGGCGGACCGCTTCGCCCAGCACTGCGGCCCGGCCGTCCAGCAGGTCT
ACGGAAGCACCGAG//GCCGGTGTCGTCGCCGCGCAGCTCCCGCAGCCGGACGGCACGGCCGATGCCGGGGTGGG

> *S. polyantibioticus* ΔPAAK

GGCGGTATGACGGCCCGCCAGGTCCAGCTGATCCAGGACTTCCGGCCCGAGGTCATCATGGTGACT//CCGTCGT
ACATGCTGACCCTCCTCGACGAGTTCGAGCGGCAGGGCGTCGACCCGCGCGCGACCTCGCTGAAAGTCGGGATCT
TCGGAGCCGAGCCGTGGACGGAGGAGATGCGCCGCGAGATCGAGGAGCGCTTCGCCATCGACGCGGTCGACATAT
ACGGGCTGTCGGAGGTGATCGGGCCCGGGGTGGCGCAGGAGTGCGTGGAGACCAAGGACGGGCTGCACATCTGGG
AGGACCACTTCTACCCGGAGATCGTCGACCCGCTCACCGGCGAGGTGCTGCCCGAGGGCGAGCGCGGCGAGCTGG
TCTTCACCTCGCTCACCAAGGAGGCCATGCCGGTGGTCCGCTACCGGACGCGGGACCTGACCCGGCTGCTGCCGG
GCTCGGCACGGGTGTTCCGGCGGATGGAGAAGGTGACCGGGCGCAGTGACGACCTGGTGATCCTGCGCGGGGTGA
ACCTGTTCCCCACCCAGATCGAGGAGATCGTGCTGCG//pOJ260//TTAACCGTCGTACATGCTGACCCTCCTC
GACGAGTTCGAGCGGCAGGGCGTCGACCCGCGCGCGACCTCGCTGAAAGTCGGGATCTTCGGAGCCGAGCCGTGG
ACGGAGGAGATGCGCCGCGAGATCGAGGAGCGCTTCGCCATCGACGCGGTCGACATATACGGGCTGTCGGAGGTG
ATCGGGCCCGGGGTGGCGCAGGAGTGCGTGGAGACCAAGGACGGGCTGCACATCTGGGAGGACCACTTCTACCCG
GAGATCGTCGACCCGCTCACCGGCGAGGTGCTGCCCGAGGGCGAGCGCGGCGAGCTGGTCTTCACCTCGCTCACC
AAGGAGGCCATGCCGGTGGTCCGCTACCGGACGCGGGACCTGACCCGGCTGCTGCCGGGCTCGGCACGGGTGTTC
CGGCGGATGGAGAAGGTGACCGGGCGCAGTGACGACCTGGTGATCCTGCGCGGGGTGAACCTGTTCCCCACCCAG

ATCGAGGAGATCGTGCTGC//GCACGCCCGGCCTCGCCCCCCACTTCCAGCTGCGGCTCACCAAGGAGGGCCGCC
TCGAC

> *S. polyantibioticus* ΔAD2

CCAGCCGACCGGGTGCTGGCTCAATGCCGGACCGAGCTGCCCGTCTACATCATTTATACCTCCGG//CTCCACCG
GCCTGCCGAAGGGTGTGGCAGTTCCGCACAGTTCGTGCGACAACATGGTGGAGTGGCAACGGACGCATTCGGTTC
GGCCCGACCTGAGGACTGCCCAGTACGCGCCGCTGAACTTCGATGTGTGCTTCCAGGAGATTCTCGGCACCCTGT
GCGGCGGCGGCACGCTCGTCGTAGTACCCGAGCGGCTCAGACGCGATCCGATCTCCCTGCTCGACTGGCTCGTGG
TGAACCGTATCGAGCGACTGTTCCTCCCCTGCGTGGCCCTTCATATGCTCACGGTCGCTGCCACCGCCGTGCATT
CACTCGCCGGTCTGGTCCTGGCGGAGATCAATGCCGCCGGTGAGCAGCTGGTCTGCACGCCGGCCATCAGAGATT
TCTTCGCGTTGCTGCCCGGTTGCCGACTGAACAACCACTACGGGCAGAGCGAATCGGCGATGGTCACGGTCCACA
CGCTGACCGGCCCCAGCCGGGAGTGGCCCGCGCTGGCTCCCATCGGCCGGCCACTGCCGGGATGCGAGGTGCTGA
TCGACCCCCCAGACCTCGAAGAGCCGGATGTCGGGGAACTGTTGGTGGCCGGAGCGCCTCTGTCGGCGGGATACC
TCAATCAGCCCCAACTCAGCGCCGAGCGGTATGTCACCGTCGATTCAACTCCGCAGGGACACACTCGTGCCTTCC
GGACGGGTGACCTCG//pOJ260//TTAACTCCACCGGCCTGCCGAAGGGTGTGGCAGTTCCGCACAGTTCGTGC
GACAACATGGTGGAGTGGCAACGGACGCATTCGGTTCGGCCCGACCTGAGGACTGCCCAGTACGCGCCGCTGAAC
TTCGATGTGTGCTTCCAGGAGATTCTCGGCACCCTGTGCGGCGGCGGCACGCTCGTCGTAGTACCCGAGCGGCTC
AGACGCGATCCGATCTCCCTGCTCGACTGGCTCGTGGTGAACCGTATCGAGCGACTGTTCCTCCCCTGCGTGGCC
CTTCATATGCTCACGGTCGCTGCCACCGCCGTGCATTCACTCGCCGGTCTGGTCCTGGCGGAGATCAATGCCGCC
GGTGAGCAGCTGGTCTGCACGCCGGCCATCAGAGATTTCTTCGCGTTGCTGCCCGGTTGCCGACTGAACAACCAC
TACGGGCAGAGCGAATCGGCGATGGTCACGGTCCACACGCTGACCGGCCCCAGCCGGGAGTGGCCCGCGCTGGCT
CCCATCGGCCGGCCACTGCCGGGATGCGAGGTGCTGATCGACCCCCCAGACCTCGAAGAGCCGGATGTCGGGGAA
CTGTTGGTGGCCGGAGCGCCTCTGTCGGCGGGATACCTCAATCAGCCCCAACTCAGCGCCGAGCGGTATGTCACC
GTCGATTCAACTCCGCAGGGACACACTCGTGCCTTCCGGACGGGTGACCTC//GTCCGGGTCGACGGGGACGTGC

> *S. polyantibioticus* ΔA16

CTGCCCGTGGTGCGGGCGGCCGACGCGCCCGAGGCGGCGGTGGACCGCGTCGCGGACGCGCACACC//GCGTACG
TGATGTACACCTCGGGATCGACCGGCCGCCCCAAGGGCGTGGCCGTGACGCACCGCAACATCCTGGCGCTCGCCG
CCGACCCCCTGTGGGCCGACGGCAGCCACACCCGGGTCCTCGCGCACGCGCCGCACTCCTTCGACGCCTCCACCT
TCGAGGTGTGGGTGCCGCTCCTTTGCGGCGGCACGGTCGTGGTGGCCCCGCCGTCCGACTCGGCGGCGCACGCCC
TGGAGCGGACGGTGGCACAGCACCGCCCGACCAGCGCCTTCGTCACCGCGTCCTTGTTCAACTCCCTGGTCGCCG
AGGGCAGTCCCGCCCTCGCCGGGCTGCAACACGTCCTGGTCGGCGGCGAAGCCCCTTCGGCGGCGGCCGTGCGGC
AGTTCCTGGCCGCCTCTCCCGGCACGGCGCTGACCAACGCCTACGGGCCGACCGAGAACACCACCTTCACGACCT
GCCACCGCTACGAGCCGGGCGCCGACGGCAGCCCCACGATCGGCCGGCCGATGGCCAACACCCGGGCCTACGTCC
TGGACGAGCGGCTGCACCCCGTACCCACCGGCGTCGTCGGCGAGCTGTACATCGCGGGCGCGGGCCTGGCCCGCG
GCTACCTCCACAACCCTGGCCTCACCGCCGGGCGGTTCGTCGCCGACCCCTTCGGCGCGCCGGGCGAGCGGATG/
/pOJ260//TTAAGCGTACGTGATGTACACCTCGGGATCGACCGGCCGCCCCAAGGGCGTGGCCGTGACGCACCG
CAACATCCTGGCGCTCGCCGCCGACCCCCTGTGGGCCGACGGCAGCCACACCCGGGTCCTCGCGCACGCGCCGCA
CTCCTTCGACGCCTCCACCTTCGAGGTGTGGGTGCCGCTCCTTTGCGGCGGCACGGTCGTGGTGGCCCCGCCGTC
CGACTCGGCGGCGCACGCCCTGGAGCGGACGGTGGCACAGCACCGCCCGACCAGCGCCTTCGTCACCGCGTCCTT
GTTCAACTCCCTGGTCGCCGAGGGCAGTCCCGCCCTCGCCGGGCTGCAACACGTCCTGGTCGGCGGCGAAGCCCC
TTCGGCGGCGGCCGTGCGGCAGTTCCTGGCCGCCTCTCCCGGCACGGCGCTGACCAACGCCTACGGGCCGACCGA
GAACACCACCTTCACGACCTGCCACCGCTACGAGCCGGGCGCCGACGGCAGCCCCACGATCGGCCGGCCGATGGC
CAACACCCGGGCCTACGTCCTGGACGAGCGGCTGCACCCCGTACCCACCGGCGTCGTCGGCGAGCTGTACATCGC
GGGCGCGGGCCTGGCCCGCGGCTACCTCCACAACCCTGGCCTCACCGCCGGGCGGTTCGTCGCCGACCCCTTCGG
CGCGCCGGGCGAGCGGAT//TACCGCACCGGGGACCTCGCACGGTGGAACGCGGACGGCGACATCGTCTTCACAG

> *S. polyantibioticus* ΔA18

GCCTTACAAGCGCCCCACGACGTGACCGATGGTGAGCGGGCGAGCGGGCTCACCGCCGACCACCCG//GCGTACG
TCATCTACACCTCCGGCTCGACGGGCAAGCCCAAAGGCGTGGTGATGACCCACCGCGGGCTCGCCAGTCTGGCCG
CCGACCACATCGAGCGGTTCGGGATCGTCGAGGGCGACGGGGTGCTCCAGTTCGCCTCGTTCAACTTCGACTGCT
CGGTGGGCGACCTGGTGATGGCGCTGGCCTCGGGCTCGGCGCTCATCGTACGGCCGCAGGACTGTCTGTCCGGGC
ACCAGCTGGGCGAGTTGATCGAGCGGACGTCCGCCACCCATGTGACGATCCCGCCGCAGGTCCTCGCCGCGCTGC
CGCCGGCCGCGCACCCCACCCTGAAGTCGGTCGCCACCGCCGGTGACGTGCTCACCGCCGAGCTCGTGGCCCAAT
GGGCGCCGGGGCGGCGGATGTTCAACGCCTACGGCCCGACCGAGACCACCGTGGACTCGCTGGCCACCGAGGTCG
AGGCCGGTTCGGGTGCCCCGCCGATCGGGCGGCCCCTGGTGAACACCCGGGTGTATGTGCTCGACGACGACATGC
GGCCGCTGCCCGTCGGCGCCGAGGGCGAGCTGTTCATCGCGGGTGCGGGGCTCGCCCGCGGCTATCTGCGTCAAC
CCGGGCTCACCGCCGAGCGGTTCGTGCCCTGCCCGTTCGGGGAGCCGGGCGAGCGCATGG//pOJ260//TTAAG

CGTACGTCATCTACACCTCCGGCTCGACGGGCAAGCCCAAAGGCGTGGTGATGACCCACCGCGGGCTCGCCAGTC
TGGCCGCCGACCACATCGAGCGGTTCGGGATCGTCGAGGGCGACGGGGTGCTCCAGTTCGCCTCGTTCAACTTCG
ACTGCTCGGTGGGCGACCTGGTGATGGCGCTGGCCTCGGGCTCGGCGCTCATCGTACGGCCGCAGGACTGTCTGT
CCGGGCACCAGCTGGGCGAGTTGATCGAGCGGACGTCCGCCACCCATGTGACGATCCCGCCGCAGGTCCTCGCCG
CGCTGCCGCCGGCCGCGCACCCCACCCTGAAGTCGGTCGCCACCGCCGGTGACGTGCTCACCGCCGAGCTCGTGG
CCCAATGGGCGCCGGGGCGGCGGATGTTCAACGCCTACGGCCCGACCGAGACCACCGTGGACTCGCTGGCCACCG
AGGTCGAGGCCGGTTCGGGTGCCCCGCCGATCGGGCGGCCCCTGGTGAACACCCGGGTGTATGTGCTCGACGACG
ACATGCGGCCGCTGCCCGTCGGCGCCGAGGGCGAGCTGTTCATCGCGGGTGCGGGGCTCGCCCGCGGCTATCTGC
GTCAACCCGGGCTCACCGCCGAGCGGTTCGTGCCCTGCCCGTTCGGGGAGCCGGGCGAGCGCATG//TACCGCAC

> *S. polyantibioticus* ΔA28

CGGTGCGTACGACAGCGCCGAACCGGCCGCCGTCGCCGTCGACATCGACGACTGGGCGTACGGCGG//ACCTCGG
GTTCCACGGGCAAGCCCAAGGGCGTCGTGACCGAGTACGCCGGACTCACCAACATGCTGATCAACCACCAGCGCC
GGATCTTCGAGCCGGTGCTGGCGGAGCACGGCGACCGGGTGTTCCGGATCGCCCACACCGTGTCGTTCGCGTTCG
ACATGTCGTGGGAGGAGCTGCTGTGGCTCGCCGACGGCCACGAGGTGCACATCTGCGACGAGGAACTGCGCCGCG
ACGCGCCCGCCCTGGTCGAGTACTGCGGCGAGCACGGGATCGACGTCATCAACGTGGCCCCCACGTACACGCAGC
AGCTGGTGGCCGAGGGCCTGCTCGACAACCCGGCCCGGCGCCCCGCGCTGGTGCTGCTGGGCGGCGAGGCGGTCA
CCCCGACCCTGTGGCAGCGGCTCGCCGGCACCGAGGGAACGGTCGGCTACAACCTGTACGGACCCACCGAGTACA
CCATCAACACCCTGGGCGTCGGCACCTTCGAGTGCGAGGACCCGGTGGTGGGCGTGGCGATCGACAACACCGACG
TGTTCGTGCTGGACCCGTGGCTGCGGCCGCTCCCGGACGGAGTTCCGGGTGAGCTCCACGTCACGGGCGTCGGCA
TCGCCCGCGGCTATCTGGGCTAGCACGCCCAGACCGCGCACCGGTTCGTGGCGTCCCCGTTCGGCGAGCCCGGCG
AGCGCATGTACCGCACCGGTGACCTCG//POJ260//TTAAACCTCGGGTTCCACGGGCAAGCCCAAGGGCGTCG
TGACCGAGTACGCCGGACTCACCAACATGCTGATCAACCACCAGCGCCGGATCTTCGAGCCGGTGCTGGCGGAGC
ACGGCGACCGGGTGTTCCGGATCGCCCACACCGTGTCGTTCGCGTTCGACATGTCGTGGGAGGAGCTGCTGTGGC
TCGCCGACGGCCACGAGGTGCACATCTGCGACGAGGAACTGCGCCGCGACGCGCCCGCCCTGGTCGAGTACTGCG
GCGAGCACGGGATCGACGTCATCAACGTGGCCCCCACGTACACGCAGCAGCTGGTGGCCGAGGGCCTGCTCGACA
ACCCGGCCCGGCGCCCCGCGCTGGTGCTGCTGGGCGGCGAGGCGGTCACCCCGACCCTGTGGCAGCGGCTCGCCG
GCACCGAGGGAACGGTCGGCTACAACCTGTACGGACCCACCGAGTACACCATCAACACCCTGGGCGTCGGCACCT
TCGAGTGCGAGGACCCGGTGGTGGGCGTGGCGATCGACAACACCGACGTGTTCGTGCTGGACCCGTGGCTGCGGC
CGCTCCCGGACGGAGTTCCGGGTGAGCTCCACGTCACGGGCGTCGGCATCGCCCGCGGCTATCTGGGCTAGCACG
CCCAGACCGCGCACCGGTTCGTGGCGTCCCCGTTCGGCGAGCCCGGCGAGCGCATGTACCGCACCGGTGACCTC
//ACGAGCTGCG

> *S. polyantibioticus* ΔA7

GCGCCGCCACCGCGCGCAGCCTGGCGTACGTCATCTACACCTCCGGCTCCACCGGCCGCCCCAAGG//GCCTACG
TGATCTACACCTCGGGGTCCACCGGCACGCCCAAGGGCGTGATGGTCGAACACCGCAATGCCACCCGGCTGTTCA
CCGCCACCGAGCCGTGGTTCGGCTTCGGCCGCGACGACGTGTGGACGCTGTTCCACTCCTTCGCCTTCGACTTCT
CCGTGTGGGAGATCTGGGGCGCGCTGCTGCACGGCGGCCGTCTGGTGATCGTGCCGCAGGCCACCACCCGCAACC
CGAACGACTTCTACGCCCTGCTGTGCGCGGAGGGGGTGACCGTCCTCAACCAGACGCCCAGCGCCTTCCGGCAGC
TGATCGCGGCCCAGGGCGACAGCCCGGCCGCACACCGGCTGCGCACGGTCGTCTTCGGCGGCGAGGCCCTGGACG
TCGCCGCACTCAAGCCGTGGCTGCGCCGCGCGGCCAACAAGGGCACCCGGCTCGTCAACATGTACGGGATCACCG
AGACCACCGTGCATGTCACCTACCGGCCGCTGACCGAGGCCGACGCGGAGCTCGCGGTGAGCCCGATCGGCGAAC
GGATCCCCGACCTGCGCACCTACGTCCTGGACCGGCACGGGCGGCCCGCCCCCGTCGGCGCGGTCGGCGAGCTGT
ATGTCGGCGGTGACGGCGTGGCCCGCGGCTACCTCAACCGGCCCGAGCTCACCGCCGAGCGCTTCCTGGACGACC
CGTTCTGCCCCGAGCCCGATGAGCGGATGTACCGCACCGGCGACGTCG//POJ260//TTAAGCCTACGTGATCT
ACACCTCGGGGTCCACCGGCACGCCCAAGGGCGTGATGGTCGAACACCGCAATGCCACCCGGCTGTTCACCGCCA
CCGAGCCGTGGTTCGGCTTCGGCCGCGACGACGTGTGGACGCTGTTCCACTCCTTCGCCTTCGACTTCTCCGTGT
GGGAGATCTGGGGCGCGCTGCTGCACGGCGGCCGTCTGGTGATCGTGCCGCAGGCCACCACCCGCAACCCGAACG
ACTTCTACGCCCTGCTGTGCGCGGAGGGGGTGACCGTCCTCAACCAGACGCCCAGCGCCTTCCGGCAGCTGATCG
CGGCCCAGGGCGACAGCCCGGCCGCACACCGGCTGCGCACGGTCGTCTTCGGCGGCGAGGCCCTGGACGTCGCCG
CACTCAAGCCGTGGCTGCGCCGCGCGGCCAACAAGGGCACCCGGCTCGTCAACATGTACGGGATCACCGAGACCA
CCGTGCATGTCACCTACCGGCCGCTGACCGAGGCCGACGCGGAGCTCGCGGTGAGCCCGATCGGCGAACGGATCC
CCGACCTGCGCACCTACGTCCTGGACCGGCACGGGCGGCCCGCCCCCGTCGGCGCGGTCGGCGAGCTGTATGTCG
GCGGTGACGGCGTGGCCCGCGGCTACCTCAACCGGCCCGAGCTCACCGCCGAGCGCTTCCTGGACGACCCGTTCT
GCCCCGAGCCCGATGAGCGGATGTACCGCACCGGCGACGTC//GTACGGCGCCTCGCGGACGGCACCCTGGAATT