A Genomics-Led Approach to Deciphering Heterocyclic Natural Product Biosynthesis

Karen Hoi-Lam Chan
Churchill College
Department of Biochemistry
University of Cambridge
October 2018

This dissertation is submitted for the degree of Doctor of Philosophy.

Abstract

A Genomics-Led Approach to Deciphering Heterocyclic Natural Product Biosynthesis

Karen Hoi-Lam Chan

Heterocycles play an important role in many biological processes and are widespread among natural products. Oxazole-containing natural products possess a broad range of bioactivities and are of great interest in the pharmaceutical and agrochemical industries. Herein, the biosynthetic routes to the oxazole-containing phthoxazolins and the bis(benzoxaozle) AJI9561, were investigated.

Phthoxazolins A-D are a group of oxazole trienes produced by a polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS) pathway in *Streptomyces* sp. KO-7888 and *Streptomyces* sp. OM-5714. The phthoxazolin pathway was used as a model to study 5-oxazole and primary amide formation in PKS-NRPS pathways. An unusually large gene cluster for phthoxazolin biosynthesis was identified from the complete genome sequence of the producer strains and various gene deletions were performed to define the minimal gene cluster. *PhoxP* was proposed to encode an ATP-dependent cyclodehydratase for 5-oxazole formation on an enzyme-bound *N*-formylglycylacyl-intermediate, and its deletion abolished phthoxazolin production. *In vitro* reconstitution of the early steps of phthoxazolin biosynthesis was attempted to validate the role of PhoxP, but was unsuccessful. Furthermore, Orf3515, a putative flavin-dependent monooxygenase coded by a remote gene, was proposed to hydroxylate glycine-extended polyketide-peptide chain(s) at the α-position to yield phthoxazolins with the primary amide moiety.

On the other hand, an *in vitro* approach was employed to establish the enzymatic logic of the biosynthesis of AJI9561, a bis(benzoxazole) antibiotic isolated from *Streptomyces* sp. AJ9561. The AJI9561 pathway was reconstituted using the precursors 3-hydroxyanthranilic acid and 6-methylsalicylic acid and five purified enzymes previously identified from the pathway as key enzymes for benzoxazole formation, including two adenylation enzymes for precursor activation, an acyl carrier protein (ACP), a 3-oxoacyl-ACP synthase and an amidohydrolase-like cyclase. Intermediates and shunt products isolated from enzymatic reactions containing different enzyme and precursor combinations were assessed for their competence for various steps of AJI9561 biosynthesis. Further bioinformatic analysis and *in silico* modelling of the amidohydrolase-like cyclase shed light on the oxazole cyclisation that represents a novel catalytic function of the amidohydrolase superfamily.

Preface

The dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualitification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit of 60,000 words.

Karen Hoi-Lam Chan
Churchill College
Department of Biochemistry
University of Cambridge
October 2018

Acknowledgements

First and foremost, my sincere gratitude goes to my PhD supervisor, Professor Peter Leadlay. I thank Peter for offering the opportunity for me to embark on this challenging yet fun, fruitful and memorable journey, as well as for his immense support, guidance and encouragement throughout my PhD.

I could not have asked for a better research group to work with. I thank all past and present group members I have met for the warm welcome and for creating an enjoyable atmosphere at work. Special thanks to Dr. Fanglu Huang who is always so kind and generous with advices, for supervising much of my lab work, to Dr. Hui Hong for her advices and help with analytics, and to Dr. Annabel Murphy for her help with chemical synthesis. I express my gratitude to Dr. Youngjun Zhou, Dr. Qiulin Wu and Joachim Hug who started the oxazole and benzoxaozle projects. Their insights and results served as a foundation of my work. Special thanks to my good friend Vinzent Schulz for the much enjoyable time working on thiotetronates and for his constant support and encouragement. I thank Dr. Freddie Dudbridge and Dr. Marie Yurkovich for their help with microbiology and molecular biology, and Dr. Anya Luhavaya and Dr. Katharina Dornblut for proofreading parts of this thesis and their detailed feedback, and most importantly, for being great friends. I appreciate the sequencing work conducted by Shilo Dickens, Nataliya Scott, Anna Efimova and Reda Deglau of the DNA Sequencing Facility. I truly cherish the time spent and the friendships I have with all who are already mentioned, Felix Trottmann, Roel van Harten, Hannah Büttner, Rory Little, Oana Sadiq, Karen Lebe, Constance Wu, Carsten Schotte, Katsiaryna Usachova, Anna Reva, Jake Pollock, David Möller, Dr. Oksana Bilyk and Dr. Kwaku Kyeremeh. Additionally, I thank Professor Satoshi Ômura and Dr. Yuki Inahashi of the Kitasato Institute for providing the phthoxazolin producer strains and sharing their findings, Dr. Markiyan Samborskyy for his work on genome assembly and annotation, and the Mass Spectrometry Service, Department of Chemistry for performing accurate mass analyses. My thanks also go to all my other friends in Cambridge and around the world for their support.

Financial support from the Henry Lester Trust and the Herchel Smith Foundation aided me to complete this PhD.

Finally, I would like to thank and dedicate this work to my parents and my sister, who always believed in me and offered their support unconditionally. I hope I have made you proud.

Contents

Abbreviations

(k)Da	(kilo)dalton
(M/k)bp	(mega/kilo)basepair
2,3-AMBA	2,3-aminomethylbenzoic acid
3-HAA	3-hydroxyanthranilic acid
3-HBA	3-hydroxybenozic acid
3-HPA	3-hydroxypicolinic acid
4'-PP	4'-phosphopantetheine
6-dEB	6-deoxyerythronolide B
6-MSA	6-methylsalicylic acid
6-MSAS	6-methylsalicylic acid synthase
A	adenylation
ACP	acyl carrier protein
ADIC	2-amino-2-desoxyisochorismate
Ala	alanine
AMP	adenosine monophosphate
AMR	Review on Antimicrobial Resistance
antiSMASH	antibiotics & Secondary Metabolite Analysis Shell
APS	adenyl-5'-phosphosulfate
Arg	arginine
Asn	asparagine
Asp	aspartic acid
AT	acyltransferase
ATBH	aminohydroxybacteriohopane
ATP	adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
C	condensation domain
CDA	calcium dependent antibiotic
cMT/*C*-MT	*C*-methyltransferase
CoA	coenzyme A
CPC	2-carboxy-2-hydroxycyclopentanone
cryo-EM	cryo-electron microscopy
Cy	heterocyclisation

CYC	aromatase/cyclase
Cyp	cytochrome P450
Cys	cysteine
DAHP	3-deoxy-D-arabino-heptulosonic acid 7-phosphate
DEBS	6-deoxyerythronolide B synthase
DH	dehydratase
DHB	dihydroxybenzoate
DHHA	*trans*-2,3-dihydro-3-hydroxyanthranilic acid
DHPA	dihydroxyphenylacetate
DHPG	dihydroxyphenylglycine
DHt	truncated dehydratase
DMAPP	dimethylallyl diphosphate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide
DUF	domain of unknown function
E	epimerisation
E4P	D-erythrose-4-phosphate
EDTA	ethylenediaminetetraacetic acid
ER	enoylreductase
ESKAPE	*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumanii*, *Pseudomonas aeruginosa* and *Enterobacter* spp.
FA	formic acid
FAS	fatty acid synthase
FDP	farnesyl diphosphate
GDP	geranyl diphosphate
GGDP	geranylgeranyl diphosphate
Gln	glutamine
Glu	glutamic acid
Gly	glycine
H_4F	tetrahydrofolate
HCHO	formaldehyde
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid, *N*-(2-Hydroxyethyl)piperazine-*N*′-(2-ethanesulfonic acid)
His	histidine
HMM	Hidden Markov model

Ile	isoleucine
IMG	Integrated Microbial Genomes
IMS	ion-mobility spectrometry
IPP	isopentenyl diphosphate
IPTG	isopropyl β-D-1-thiogalactopyranoside
KR	ketoreductase
KS	ketosynthase
LC-MS	liquid chromatography-mass spectrometry
Leu	leucine
Lys	lysine
m/z	mass to charge ratio
MCoA	malonyl-CoA
MDR	medium-chain dehydrogenase/reductase
MEGA7	Molecular Evolutionary Genetics Analysis Version 7.0
Met	methionine
MFS	major facilitator superfamily
MLSA	multilocus sequence analysis
MOx	monooxygenase
MQ	Milli-Q®
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MT	methyltransferase
MUSCLE	MUltiple Sequence Comparison by Log-Expectation
N^{10}-fH_4F	N^{10}-formyltetrahydrofolate
$NAD(P)^+$	β-nicotinamide adenine dinucleotide (2'-phosphate)
NAD(P)H	β-nicotinamide adenine dinucleotide (2'-phosphate), reduced
NCBI	National Center for Biotechnology Information
NMR	nuclear magnetic resonance
nMT/*N*-MT	*N*-methyltransferase
NRPS	non-ribosomal peptide synthetase
OD	optical density
ORF	open reading frame
OSMAC	one strain, many compounds
Ox	oxidase/oxidoreductase
PAPS	3'-phosphoadenyl-5'-phosphosulfate
PCP	peptidyl carrier protein

PCR	polymerase chain reaction
PDB	Protein Data Bank
PEP	phosphoenolpyruvate
Phe	phenylalanine
PKS	polyketide synthase
PP_i	inorganic pyrophosphate
PPTase	4'-phosphopantetheinyltransferase
Pro	proline
PT	product template domain
QTOF	quadrupole time of flight
RiPP	Ribosomally synthesised and post-translationally modified peptide
RNA	ribonucleic acid
RPM	rounds per minute
rRNA	ribosomal ribonucleic acid
SA	salicylic acid
SAM	*S*-adenosyl methionine
SARP	Streptomyces antibiotic regulatory protein
SAXS	small-angle X-ray scattering
SDR	short-chain dehydrogenase/reductase
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Ser	serine
SFLD	Structure-Function Linkage Database
SH	cysteine lyase
TD	terminal reductase
TDP	thymidine-5'-diphosphate
TE	thioesterase
TFA	trifluoroacetic acid
THID	thioester hydrolase
Thr	threonine
TIM	triosephosphate isomerase
TLC	thin-layer chromatography
TOMM	thiazole/oxazole-modified microcin
Trp	tryptophan
Tyr	tyrosine
UV	ultraviolet

Val	valine
WHO	World Health Organization
XRE	xenobiotic response element
αANH	adenine nucleotide α hydrolase

Chapter 1

Introduction

1.1 A brief history of antibiotic discovery

1.1.1 Antibiotics of microbial origin

The first documentation of microbial antagonism was arguably Pasteur and Joubert's (1877) report on the inhibitory effect of Enterobacteriaceae on the growth of pathogenic *Bacillus anthracis*, and they coined the term "antibiose" for this antagonistic phenomenon. However, it was only in 1929 that Fleming famously observed, on an agar plate of *Staphylococcus*, the clear halo surrounding a mould later identified as *Penicillium notatum*. The antibacterial substance secreted by *P. notatum* was named penicillin (Fleming, 1929). It was later isolated and tested in mice infected with *Streptococci* (Chain *et al*., 1940), and introduced for therapeutic use in the same year (Abraham *et al*., 1941). As the antibiotic was in high demand during World War II, intensive efforts were made to establish its industrial production and it became widely available by 1943 (Richards, 1964). Soon afterwards, the unique ß-lactam pharmacophore of penicillin was revealed by the work of Hodgkin (1949).

Meanwhile, in 1939, tyrothricin had been isolated from *Bacillus brevis* (Dubos, 1939). Tyrothricin consists of two bactericidal agents, gramicidin and tyrocidine (Dubos and Hotchkiss, 1941). They were found to be highly effective against a range of Gram-positive pathogens, and are the first antibiotics discovered from the deliberate screening of soil bacteria. Waksman, a Rutgers soil microbiologist, also joined the search for antibiotics. From various strains of *Streptomyces*, his group discovered more than a dozen antibiotics, including the peptide antibiotic actinomycin, the broad-spectrum streptothricin, and most notably, the first anti-tuberculosis drug streptomycin (Waksman and Woodruff, 1940; Waksman and Woodruff, 1942; Schatz *et al*., 1944). Interest in natural product research grew rapidly and prompted systematic screening and isolation of bioactive molecules from

microorganisms. This period marks the start of the golden age of antibiotics. Chloramphenicol, cephalosporin, chlortetracycline, erythromycin, vancomycin, kanamycin and many more antibiotics of microbial origin were discovered and brought to clinical use, saving millions of lives over many decades (Ehrlich *et al*., 1947; Brotzu, 1948; Duggar, 1948; McGuire *et al*., 1952; Brigham and Pittenger, 1956; Umezawa *et al*., 1957). Figure 1.1 shows the structure of a selection of antibiotics mentioned above.

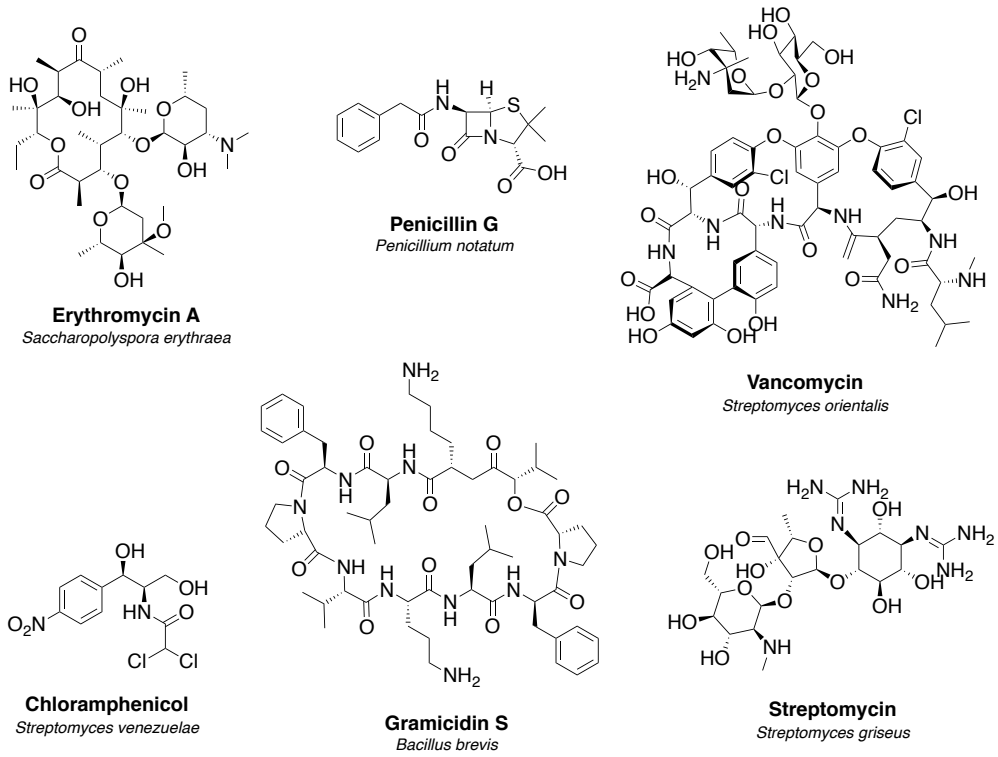


Figure 1.1. Structure and microbial origin of the antibiotics erythromycin A, penicillin G, vancomycin, chloramphenicol, gramicidin S and streptomycin.

1.1.2 Antibiotic classification and mechanisms of action

Antibiotics may either induce cell death (bactericidal) or inhibit cell growth (bacteriostatic). They may be classified according to their mechanisms of action, owing to their structural features and binding affinity to specific cellular targets. Known mechanisms of action include the inhibition of the processes of cell wall synthesis, DNA replication, RNA transcription, protein synthesis and essential metabolite synthesis; and the disruption of cell membrane function and of nucleic acid structures (Madigan and Martinko, 2006).

1.1.3 Mechanisms of antibiotic self-resistance

Microorganisms producing antibiotics develop self-resistance mechanisms to protect themselves from the lethal effect of their own products. To produce viable progeny, it is assumed that mechanisms of self-resistance have co-evolved with the biosynthesis of antibiotics. In fact, genes conferring self-resistance often cluster with the biosynthetic genes, and they are co-regulated by the same signals in response to environmental stress (Laskaris *et al*., 2010). In some pathways, multiple resistance-conferring genes are observed and a combination of different mechanisms contributes to provide greater protection (Mao *et al*., 1999; Kim *et al*., 2008; Luo *et al*., 2011).

Self-resistance may be achieved by the inactivation of antibiotics in the cytoplasm by chemical modification such as phosphorylation, adenylation or acetylation (Sugiyama *et al*., 1981; Sugiyama *et al*., 1985; Sugiyama *et al*., 1986); by the timely export of antibiotics using pathway specific-transporter proteins (Levy and McMurry, 1974; Tahlan *et al*., 2007); by the modification of endogenous cellular targets to render them insensitive to inhibition (Graham and Weisblum, 1979; Zalacain and Cundliffe, 1989; Kumagai *et al*., 1999); by the supply of an additional copy of a target enzyme whose active site is altered to abrogate antibiotic binding (Olano *et al*., 2004; Tao *et al*., 2016); or by the repair of damaged cellular targets (Lomovskaya *et al*., 1996).

1.1.4 Emergence of antibiotic resistance in pathogens

Soon after the introduction of antibiotics for clinical use, antibiotic resistance was noted to develop in the targeted pathogenic species. Antibiotic resistance is still an immense and growing threat to humanity, and poses a difficult challenge in modern medicine. Horizontal gene transfer promotes the spread of (self-)resistance mechanisms across different genera of bacteria, which is further accelerated by the overuse and misuse of antibiotics (Forsman *et al*., 1990; WHO, 2015). In 1940, a penicillin-resistant clinical strain of *Staphylococcus* was reported for the first time (Gardner, 1940). Abraham and Chain (1940) showed that bacteria utilise an enzyme named penicillinase to hydrolyse the β-lactam pharmacophore of the antibiotic. It soon became clear that antibiotic resistance was problematic not only with the use of penicillin, but also with that of other antibiotics. Tetracycline-resistant *Shigella*, methicillin-resistant *Staphylococcus*, penicillin-resistant *Pneumococcus* and erythromycin-resistant *Streptococcus* were identified, with many more examples following and an ever-diminishing interval between introduction and development of

resistance (Olarte and de la Torre, 1959; Jevons *et al.*, 1963; Kislak *et al*., 1965; Dixon, 1968).

While clinical cases of resistance have continued to climb (Knight and Collins, 1955), the rate of natural product discovery from bacteria has slowed down since the 1960s. The reasons for this include the frequent rediscovery of known molecules from bioactivity-guided screening, and the relative difficulty of conducting medicinal chemistry on complex natural product templates. Moreover, as antibiotics are typically taken for short, one-off courses and since resistance is known to develop quickly, research and development in antibiotics offer little incentive for pharmaceutical companies. With the rise of the ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumanii*, *Pseudomonas aeruginosa* and *Enterobacter* spp.) and other multidrug-resistant "superbugs" such as *Escherichia coli* O157:H7, *Clostridium difficile*, *Neisseria gonorrhoea* and *Mycobacterium tuberculosis*, it is estimated that by 2050, 10 million people each year could lose their lives to infections no longer treatable with existing medication (AMR Review, 2016). There is an urgent need for new antimicrobials and it is essential to revive the antibiotic discovery process via multidisciplinary approaches.

1.2 *Streptomyces* as a prolific source of bioactive natural products

1.2.1 The genus *Streptomyces*

Actinomycetes belong to the order Actinomycetales and represent a diverse group of Gram-positive, filamentous bacteria characterised by the high G+C content in their genomes and a complex life cycle involving morphological differentiation. Actinomycetes consist of the genera *Mycobacterium*, *Actinomyces*, *Nocardia*, *Streptomyces* and *Micromonospora* (Waksman and Henrici, 1943). In particular, *Streptomyces* spp. produce a wide range of bioactive natural products, including two-thirds of all known antibiotics (Bérdy, 2005), some of which have already been mentioned in Section 1.1.

Among bacteria, *Streptomyces* spp. have some of the largest genomes, typically around 8 to 12 Mbp. They have a linear chromosome with DNA termini consisting of long terminal inverted repeats that are prone to large deletions and rearrangements (Birch *et al*., 1990; Lin *et al*., 1993; Yu and Chen, 1993). The central part of the chromosome, which contains all essential genes for normal growth and replication, spans 4 to 5 Mbp and is flanked by regions that are enriched in genes governing natural product biosynthesis (Bentley *et al*., 2002; Ikeda *et al*., 2003).

Streptomyces spp. have a complex life cycle and undergo a programme of morphological differentiation (Figure 1.2) (Angert, 2005; Flärdh and Buttner, 2009). During germination, a spore swells and polarised growth takes place towards the tip of the emerging germ tube, which develops into a hypha. Hyphae extend by tip growth and branch into a vegetative mycelium that continues to grow into the solid substrate. In response to the increase in mycelial density, nutrient depletion and other stress signals, the growth of aerial hyphae is initiated by breaking the surface tension. Aerial hyphae differentiate into long chains of prespore compartments which develop thick walls, produce a grey spore pigment and acquire other characteristics of spores. The mature, dormant spores, which are desiccation-resistant but heat-labile, are dispersed in the environment and germinate when the environmental conditions are favourable.

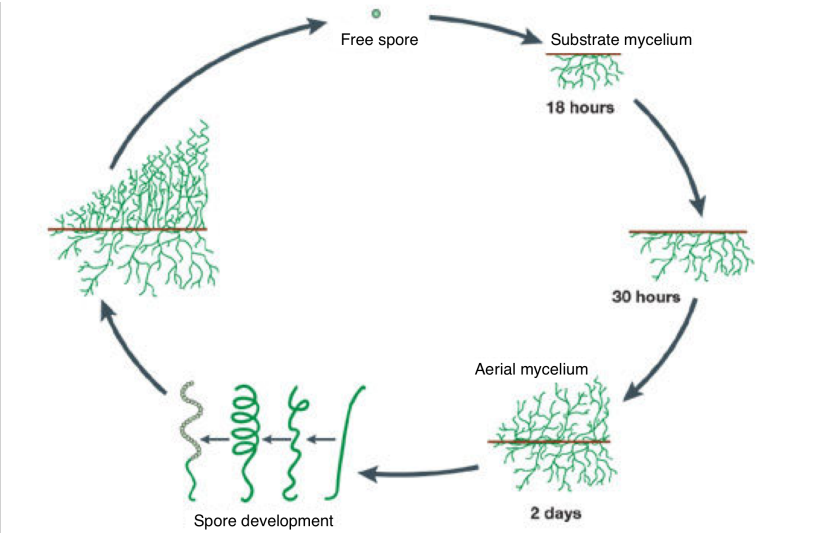


Figure 1.2. The life cycle of *Streptomyces*. Adapted from Angert (2005).

The onset of natural product biosynthesis in *Streptomyces* typically occurs during the stationary phase of growth and correlates, in growth on solid media, with the formation of aerial hyphae. Cluster-situated regulatory genes, like self-resistance genes, cluster with biosynthetic genes and are subject in turn to control by pleiotropic genes involved in quorum sensing, as exemplified by the A-factor signalling cascade in *S. griseus* (Horinouchi and Beppu, 1994; Horinouchi *et al*., 2007), and in other responses to environmental and physiological stress (Hood *et al*., 1992; Floriano and Bibb, 1996; Chater and Bibb, 1997; Challis and Ravel, 2000).

1.2.2 The ecology and evolution of natural product biosynthesis

Although a remarkable number of natural products of microbial origin have been shown to be useful in the clinic or as biocontrol agents, they are usually detected in their natural environment (for example, soil samples) at levels that seem too low to have any biological significance. However, Wellington and colleagues (Laskaris *et al*., 2010) have argued that production is easily underestimated, and adduce evidence for the role of antibiotic production in biocontrol of plant pathogens. Traxler and Kolter (2015) also argued that the limited diffusion of antibiotics away from the producing organism means that high local concentrations are achieved. They have reviewed a range of examples of soil bacteria using specialised metabolites for “competition sensing” under nutrient-limiting conditions. Some natural products play a role in morphological differentiation, quorum sensing and/or metal transport (Redshaw *et al*., 1979; Petersen *et al*., 1993; Schneider and Hantke, 1993;

Winkelman and Drechsel, 1997), whereas a few others were postulated to be involved in the symbiotic or mutualistic interactions with higher organisms, like insects and plants (Currie *et al*., 1999; Coombs and Franco, 2003; Haeder *et al*., 2009; Kaltenpoth *et al*., 2012). Still, there are many examples of natural products with no demonstrated bioactivity, or exhibiting bioactivities that could not be associated with any obvious selective advantage for the producer, and the ecological significance of natural product biosynthesis remains largely unknown.

Firn and Jones (2000) have proposed that the possession of genes encoding biosynthetic pathways is crucial, even though many individual products of the pathways may not necessarily confer an immediate selective advantage, as the ability to produce such compounds is itself the advantageous and selectable trait. Interestingly, it is frequently observed that different species may produce common natural products, while strains belonging to the same species may not necessarily share an identical metabolic profile. Comparison between the genomes of different bacterial strains offered much insight into how biosynthetic pathways may have distributed and evolved to expand chemical diversity through time and their significance for the survival of the individuals and/or population as a whole. At the species level, closely-related strains appear to possess highly conserved genomic islands enriched with genetic mobile elements for vertical inheritance, as a means of cluster exchange in addition to horizontal gene transfer, potentially to maximise their capacity to produce chemicals as a population (Ziemert *et al*., 2014; Medema *et al*., 2014; Bruns *et al*., 2018).

1.3 Harnessing the biosynthetic potential of microorganisms

1.3.1 Conventional methods for natural product discovery

As discussed earlier in this chapter, the first bioactive natural products were discovered and studied mostly via bioactivity-guided screening of microbial isolates. Positive isolates were then subjected to fermentation, extraction, purification and structure elucidation. Reverse genetics, study of blocked mutants obtained by random mutagenesis and heterologous expression further contributed to the identification and characterisation of the different genes and enzymes responsible for natural product biosynthesis (Feitelson *et al*., 1985; Malpartida and Hopwood, 1986; Baltz and Seno, 1981; Roberts *et al*., 1993; Cortés *et al*., 1995).

1.3.2 Mining for biosynthetic gene clusters in the age of genome sequencing

In bacteria, genes that belong to a common biosynthetic pathway often cluster in proximity to each other on the chromosome, forming a “biosynthetic gene cluster” (also referred to as “cluster” in text). Recent advances in DNA sequencing leading to the increased availability of whole-genome sequences and biosynthetic gene cluster databases, such as AntiSMASH (Blin *et al*., 2017), ClusterFinder (Cimermancic *et al*., 2014), Integrated Microbial Genomes (IMG) (Markowitz *et al*., 2012) and multigene BLAST (Medema *et al*., 2013), have greatly speeded up the identification of biosynthetic gene clusters. Tools are now widely available for the automated *in silico* annotation of putative protein function and the prediction of substrate specificity of enzymatic domains. This facilitates structural prediction of products of the clusters and provides insights into the biosynthetic logic of natural products. Examples of such tools include protein-protein BLAST (Altschul *et al*., 1990; Gish and States, 1993), the Natural Product Domain Seeker (NaPDoS) (Ziemert *et al*., 2012) and SANDPUMA (Chevrette *et al*., 2017).

Streptomyces coelicolor A3(2) was the first member of the genus *Streptomyces* to have its genome fully sequenced (Bentley *et al*., 2002). It was already well known that varying the growth conditions of a typical streptomycete strain could change the profile of specialised metabolites produced, and this was the basis for the "One Strain, Many Compounds" (OSMAC) approach to prospecting for novel metabolites (Bode *et al*., 2002). The results of the sequencing revealed the untapped biosynthetic potential of even well-studied strains, as 23 different clusters were identified at a time when it was then only known to produce four bioactive molecules, actinorhodin, undecylprodigiosin, methylenomycin and calcium-

dependent antibiotic (CDA). This finding initiated intensive research into harnessing the biosynthetic potential of microorganisms via whole-genome sequencing and genome mining.

Genome mining has opened up immense opportunities not only for the discovery and activation of new and/or cryptic clusters, but also for the identification and study of enzymes with novel functionality and/or responsible for the biotransformation of valuable chemotypes (Zhou *et al*., 2015; Saha *et al*., 2017). The combination of genome mining and the use of high-throughput LC-MS analysis, MS/MS fragmentation analysis (Kleigrewe *et al*., 2015), molecular network analysis (Nguyen *et al*., 2013; Cimermancic *et al*., 2014) and/or metabolomic maps (Goodwin *et al*., 2015) has proven powerful in efficiently establishing connections between clusters and their products, and in prioritising clusters for study. Additionally, metagenomic sequencing of complex environmental samples and single cell sorting and sequencing allow access to genetic data for the mining of uncultured bacteria. The direct capture and mobilisation of clusters via heterologous expression bypass the need to culture the "unculturable" (Owen *et al*., 2013; Iqbal *et al*., 2016; Hover *et al*., 2018).

Complete genome sequences are also particularly useful for the search of essential genes located away from a cluster and the identification and manipulation of duplicate or redundant copies of biosynthetic genes elsewhere on the chromosome (Lazos *et al*., 2010; Penwell *et al*., 2012; Cano-Prieto *et al*., 2015; Li *et al*., 2018).

1.3.3 Heterologous expression of biosynthetic pathways

As techniques in molecular genetics and synthetic biology continue to improve, the direct cloning, refactoring, mobilisation and heterologous expression of many whole biosynthetic gene clusters became possible (Pfeifer *et al*., 2001; Yamanaka *et al*., 2014; Yin *et al*., 2015). Engineered heterologous hosts, with endogenous clusters removed and/or pleiotropic regulatory elements manipulated, have proved useful for analysis of biosynthetic gene clusters from poorly-characterised or genetically-intractable strains (Gomez-Escribano and Bibb, 2011; Komatsu et al., 2010; Komatsu *et al*., 2013).

1.3.4 *In vitro* reconstitution of biosynthetic pathways

The isolation and characterisation of individual enzymatic components of biosynthetic pathways have facilitated complete or partial reconstitution *in vitro*, complemented by mutational, structural and kinetic studies (Miller *et al*., 2002; Lowry *et al*., 2013). Bacterial proteins are commonly overexpressed and purified from *E. coli*. Other heterologous hosts may be used depending on the origin of genes, codon usage and the requirement for post-translational modifications (Gomez-Escribano and Bibb, 2014; Ikeda *et al*., 2014). The reaction conditions of *in vitro* reconstitution are precisely controlled, and individual components may be substituted or removed. This allows the study of the (bio)chemistry of individual reactions and of the timing of biosynthetic steps, providing mechanistic insight to complement *in vivo* studies.

1.4 The enzymes of natural product biosynthesis

The study of the genes and enzymes responsible for natural product biosynthesis has been greatly facilitated by recent advances in molecular biology and sequencing technologies. In this section, the biosyntheses of various major classes of natural products from *Streptomyces* and related actinomycete genera, including terpenes, ribosomally-synthesised and post-translationally modified peptides, polyketides and non-ribosomal peptides, are discussed. The main focus will be on polyketides and non-ribosomal peptides as background for the work described in Chapters 3 to 7.

1.4.1 Terpenes and terpene synthases

In bacteria, terpenes arise from the acyclic precursor isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), generated from either the mevalonate pathway or the mevalonate-independent methylerythritol phosphate pathway (Rohmer *et al*., 1996). Terpene synthases catalyse the (consecutive) condensation of IPP- or DMAPP-derived intermediates to produce an isoprenyl backbone which is often subject to extensive oxidative modifications, cyclisation and/or rearrangement reactions, giving rise to an impressive array of cyclic structures that vary greatly in their functional groups and carbon skeleton. Terpenes are classified by the number of isoprene units in the carbon skeleton: monoterpenes (2 isoprene units; C_{10}), sesquiterpenes (3 isoprene units; C_{15}), diterpenes (4 isoprene units; C_{20}), triterpenes (6 isoprene units; C_{30}), tetraterpenes (8 isoprene units; C_{40}) and polyterpenes.

Production of the terpenes geosmin, 2-methylisobornel and albaflavenone is commonly observed in *Streptomyces* spp. In *S. coelicolor* A3(2), *sco6073* encodes a sesquiterpene cyclase, geosmin synthase, for the conversion of farnesyl diphosphate (FDP) to germacradienol, and from germacradienol to geosmin (Cane and Watt, 2003; Cane *et al*., 2006). The biosynthesis of both 2-methylisoborneol and albaflavenone involves tailoring enzymes: *sco7701* encodes a SAM-dependent *C*-methyltransferase to methylate geranyl diphosphate (GDP) at C-2, generating 2-methylgeranyl diphosphate, the substrate of 2-methylisoborneol synthase (Sco7700), to yield 2-methylisoborneol (Wang and Cane, 2008). A two-gene operon (*sco5222* and *sco5223*) is responsible for the biosynthesis of albaflavenols and albaflavenone. FDP is converted by the *epi*-isozizaene synthase (Sco5222) to *epi*-isozizaene, which is subjected to further oxidative reaction(s) catalysed by Sco5223, a cytochrome P450, to yield albaflavenols and albaflavenone (Lin *et al*., 2006) (Figure 1.3).

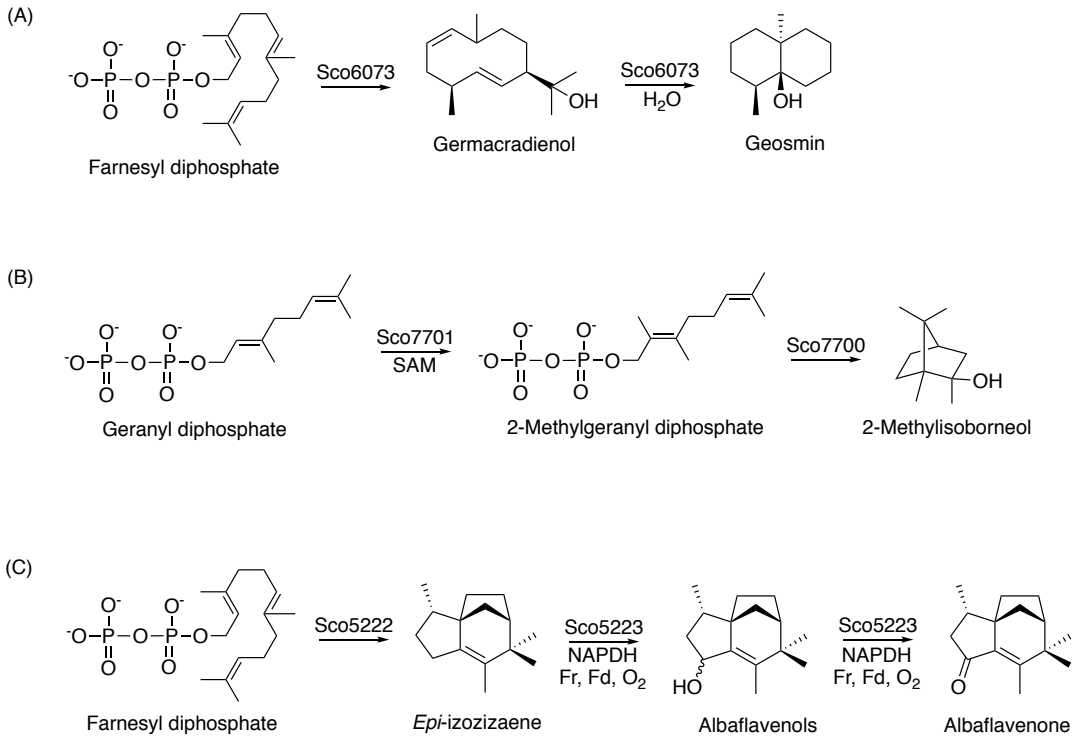


Figure 1.3. Biosynthetic pathways to terpenes in *S. coelicolor* A3(2). (A) Geosmin. (B) 2-Methylisoborneol. (C) Albaflavenols and albaflavenone.

S. coelicolor A3(2) also biosynthesises polyterpenoids, such as hopanoids and carotenoids. Squalene-hopene cyclase (Sco6764) catalyses the conversion of the acyclic squalene to the pentacyclic hopene (Poralla *et al*., 2000; Ghimire *et al*., 2009). It is postulated that four other enzymes, a SAM-dependent enzyme (Sco6766), a nucleotide phosphorylase (Sco6765), a phosphatase and a pyridoxal-dependent aminotransferase (Sco6769), are responsible for the installation of the aminohydroxylacyl side chain to yield the end product aminohydroxybacteriohopane (ATBH) (Figure 1.4). The isorenieratene cluster houses two operons, *crtEIBV* and *crtYTU*. CrtE, a geranylgeranyl diphosphate (GGDP) synthase produces GGDP using GDP and IPP as substrates and molecules of GGDP are condensed by CrtB, a phytoene synthase, to generate phytoene. CrtI, a phytoene desaturase, catalyses the conversion of phyotene to lycopene which is the substrate for cyclisation catalysed by CrtY, a lycopene cyclase, to yield β-carotene (Krügel *et al*., 1999). The last step of the isorenieratine pathway involves CrtU-catalysed intramolecular methyl transfer and desaturation of the aromatic end groups, to convert β-carotene to isorenieratine (Takano *et al*., 2005) (Figure 1.5). However, the role of the putative methyltransferases CrtV and CrtT remains unclear.

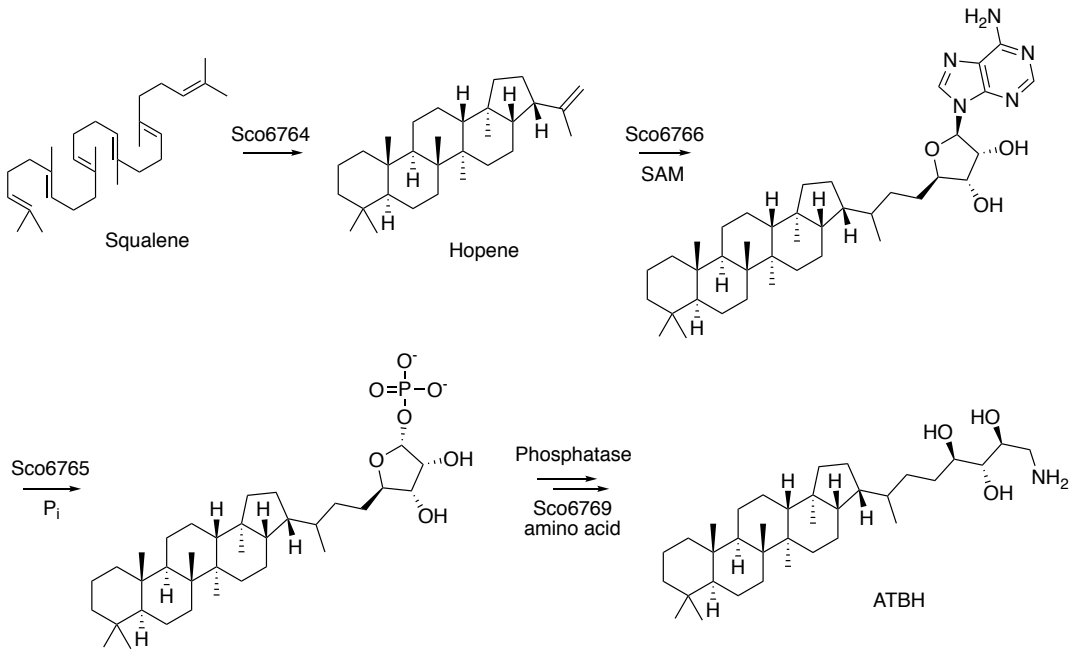


Figure 1.4. Biosynthetic pathway to hopene and ATBH in *S. coelicolor* A3(2).

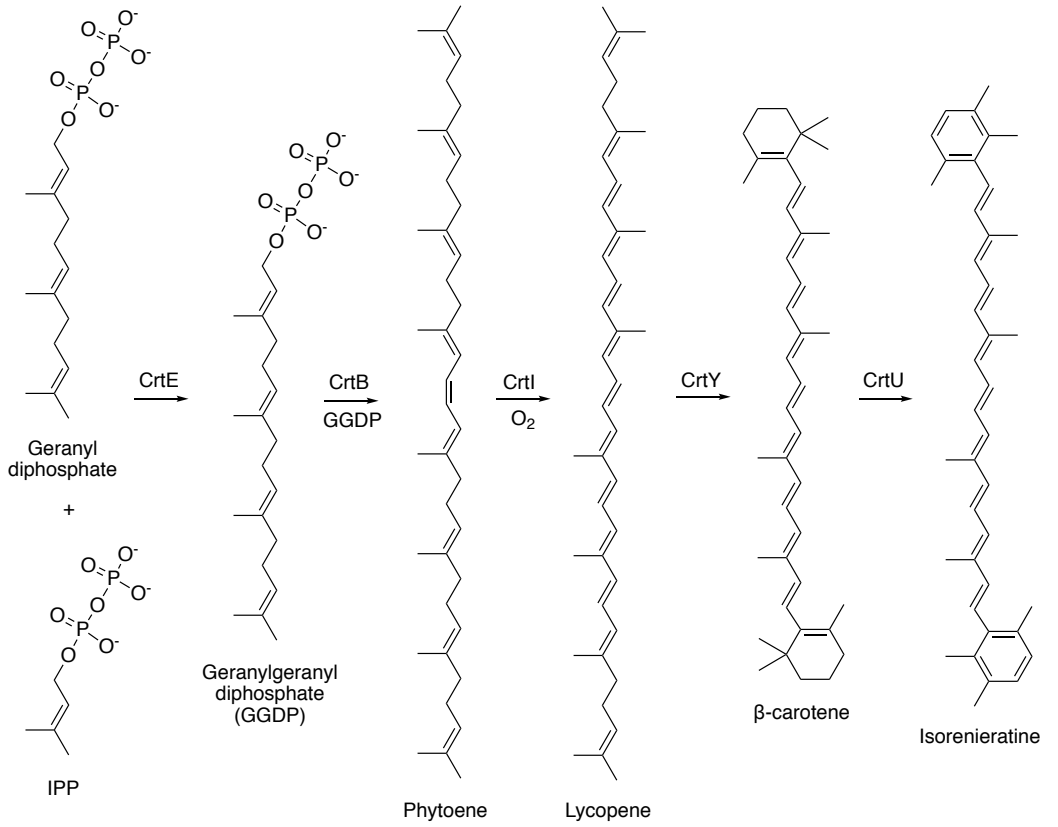


Figure 1.5. Biosynthetic pathway to carotenoids in *S. coelicolor* A3(2).

1.4.2 Ribosomally-synthesised and post-translationally modified peptides

Ribosomally synthesised and post-translationally modified peptides (RiPPs) are derived from ribosomally-translated precursor peptides. A precursor peptide contains an N-terminal leader peptide and a core peptide. The leader peptide is important for recognition by post-translational modification enzymes and transporter proteins, whereas the core peptide is subject to extensive post-translational modifications and proteolytic cleavage from the leader peptide, to yield the mature final product. RiPPs are divided into subfamilies according to their specific post-translational modifications and resulting structural features. Prominent subfamilies include lanthipeptides, lasso peptides, and thiazole/oxazole-modified microcins (TOMMs) (Figure 1.6). Post-translational modification contributes to the great structural diversity and rigid structures of mature RiPPs, compared to linear peptides.

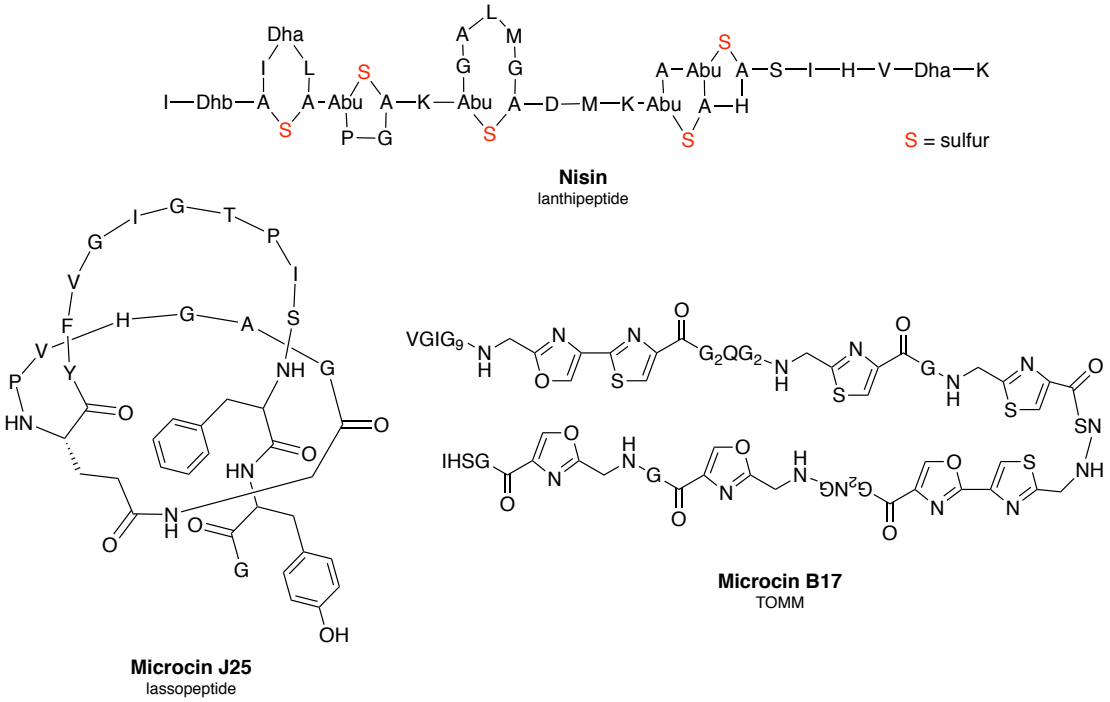


Figure 1.6. Structures of RiPPs belonging to different subfamilies. TOMM, thiazole/oxazole-modified microcin.

Among post-translational modifications in RiPPs, formation of covalent crosslinks, macrocyclisation and heterocyclisation are the most frequently observed. For instance, lanthipeptides are polycyclic RiPPs with intramolecular thioester cross-links arising from the nucleophilic addition of Cys thiols to dehydrated Ser or Thr residues (Schnell *et al*., 1988; Xie *et al*., 2004; Li *et al*., 2006). Lasso peptides, such

as microcin J25 and lassomycin, are macrolactam structures through which the C-terminal tail of the peptide is threaded. The macrolactams are formed from the intramolecular condensation of the *N*-terminal amino group with an ATP-activated side chain carboxylate of Glu or Asp in the middle of the peptide (Duquesne *et al*., 2007; Gavrish *et al*., 2014) to form an isopeptide bond. Furthermore, multiple subfamilies of RiPPs contain small azol(in)e heterocycles derived from Cys, Ser and Thr, including linear azol(in)e-containing peptides, thiopeptides and cyanobactins. Biochemical studies on the microcin B17 biosynthetic pathway reveal that the heterocycles are incorporated post-translationally using an ATP-dependent cyclodehydratase complex, McbDCB, comprised of a YcaO-like cyclodehydratase, a dehydrogenase and a partner protein (Dunbar *et al*., 2012). Upon nucleophillic attack by the side chains of Ser, Cys or Thr in the precursor peptide McbA, the YcaO-like cyclodehydratase uses ATP to activate the peptide backbone amide, followed by *N*-protonation and phosphate elimination to yield an oxazoline, a thiazoline or a methyl-oxazoline moiety, respectively. MbcC, a flavin-dependent dehydrogenase, further oxidises the azolines to produce azoles (Figure 1.7) (Dunbar *et al*., 2014).

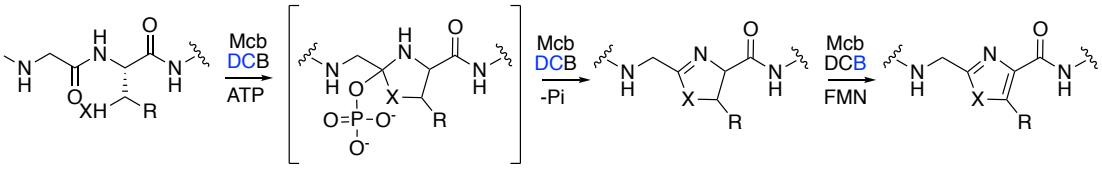


Figure 1.7. Biosynthetic pathway to the thiazole/oxazole-containing microcin B17 involving the ATP-dependent cyclodehydratase complex, McbDCB. Enzyme(s) involved in each step is/are highlighted in blue.

1.4.3 Polyketides and polyketide synthases

Polyketide natural products continue to offer an abundant source of biologically-relevant chemical diversity, with attractive potential for sustainable drug and agrochemical development. Clinically important polyketides include the antibiotics erythromycin A and rifamycin, the antiparasitic avermectin, the antifungal amphotericin B and the insecticide spinosyns A and D (Figure 1.8). Polyketides are biosynthesised by fatty acid synthase (FAS)-related enzymes called polyketide synthases (PKSs), from simple carboxylic acid-derived precursors (Donadio *et al*., 1991; Jenke-Kodama *et al*., 2005; Zheng *et al*., 2012). Three types of PKS are known to occur in microorganisms: type III, type II and type I (iterative and modular).

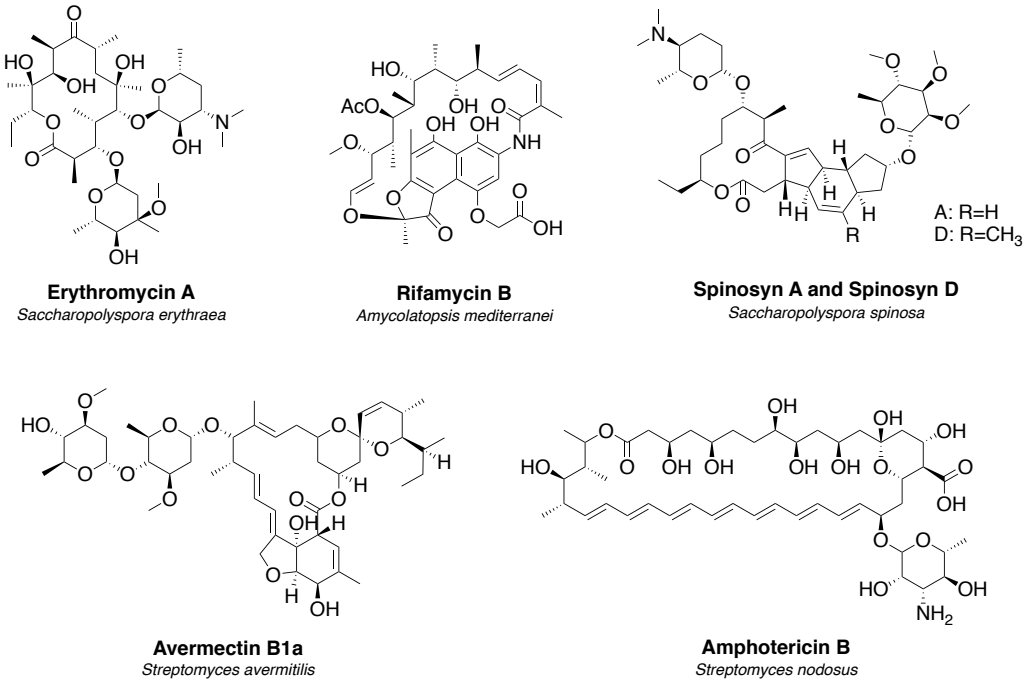


Figure 1.8. Structures and microbial origin of selected polyketide antibiotics.

Type III PKS

Type III PKSs, also known as chalcone synthase-like enzymes, are traditionally associated with plants but are also found in fungi and bacteria. Type III PKSs are single multifunctional enzymes that possess one active site to perform all required reactions, including the priming of CoA derivatives as extender units, chain elongation by intramolecular decarboxylative condensation and cyclisation, and they function in an iterative fashion (Figure 1.9) (Cortés *et al*., 2002; Achkar *et al*., 2005). Not only do type III PKSs produce chalcones, but also a variety of other molecules such as pyrones, acridones, and phloroglucinols.

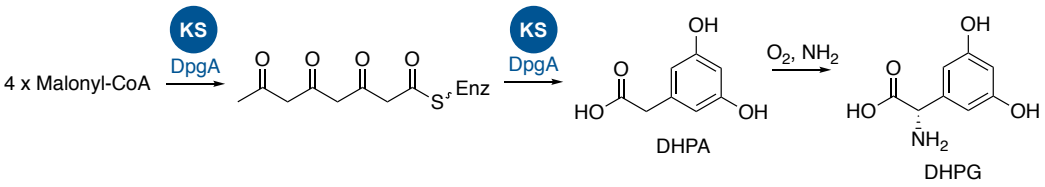


Figure 1.9. Type III PKS pathway to the unusual amino acid dihydroxyphenylglycine in *Amycolatopsis mediterranei*. DHPA, dihydroxyphenylacetate. DHPG, dihydroxyphenylglycine.

Type II PKS

Most aromatic polyketides are the product of type II PKSs, and are biosynthesised almost exclusively by bacteria and fungi. Type II PKSs consist of a single set of discrete, monofunctional enzymes that function iteratively for each round of chain elongation. A minimal set of enzymes includes two ketosynthase-like proteins (KSs), KS_α and KS_β, which together form an active heterodimeric complex, and an acyl carrier protein (ACP) (Shen and Hutchinson, 1996). During biosynthesis, the polyketide chain is tethered to and presented by an activated ACP via a flexible 4'-phosphopantetheine (4'-PP) arm, for assembly and elongation. KS_α catalyses the intramolecular condensation on each extender unit, while KS_β has been proposed to function as a decarboxylase during initiation (Bisang *et al*., 1999) and to contribute to determination of polyketide chain length (Tang *et al*., 2003). KS_β shows remarkable sequence homology to KS_α, with the catalytic Cys being replaced with Gln (Bisang *et al*., 1999). The full-length linear polyketide is modified by a ketoreductase (KR) and an aromatase/cyclase (CYC), yielding the aromatic core of the product. The regiospecificity of the aromatisation reaction is controlled through the precise orientation and correct folding of the linear polyketide in the active site pocket of the aromatase (Figure 1.10) (Zawada and Khosla, 1997; Ames *et al*., 2008; Lee *et al*., 2012). Enzymes in a type II PKS may form a complex in the bacterial cell but are readily dissociable *in vitro* (Zawada and Khosla, 1999).

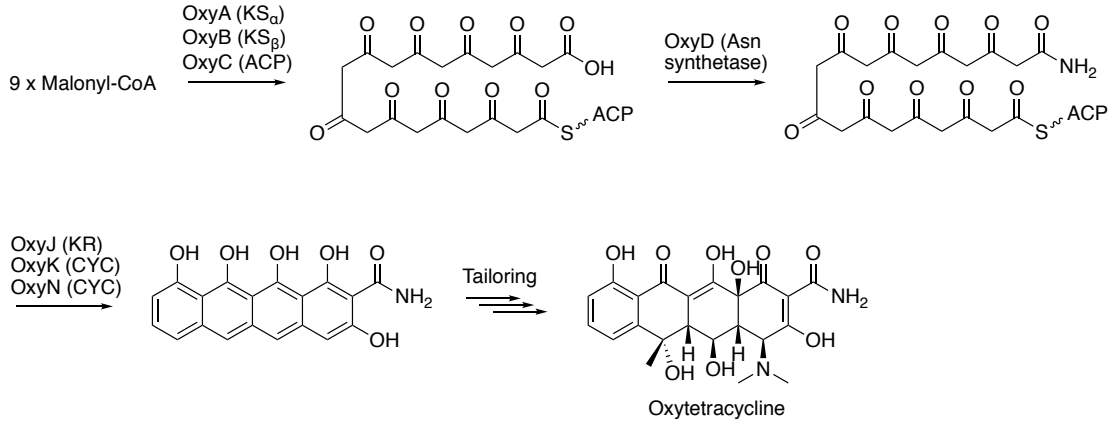


Figure 1.10. Type II PKS pathway to the aromatic polyketide antibiotic oxytetracycline.

Iterative type I PKS

Iterative type I PKSs are large, multifunctional enzymes organised into a module of FAS-related domains. They are commonly found in fungi but also play a role in

bacteria. Here, a minimal functional module consists of the KS, acyltransferase (AT) and ACP domains. The AT transfers (usually malonyl) extender units from an acyl-CoA precursor to an activated ACP, and KS catalyses a decarboxylative condensation reaction between the ACP-bound extender unit and the growing polyketide chain, elongating the polyketide backbone by two carbon units. A "reductive loop" containing the ketoreductase (KR), dehydratase (DH) and/or enoylreductase (ER) domains is often found between the AT and ACP domains. KR stereoselectively reduces the β-keto group, DH dehydrates the resulting β-hydroxy group to an αβ-double bond and ER reduces the double bond to a single bond. The multifunctional, unimodular enzyme is used iteratively with varying degree of reduction for each extender unit, as reductive domains may be optionally used, or additional reductive enzymes may, during specific extension cycles, be recruited *in-trans*, as in the lovastatin PKS (Kennedy *et al*., 1999; Auclair *et al*., 2001) (Figure 1.11). Release of the full-length polyketide chain is typically catalysed by a thioesterase/Claisen cyclase (TE/CLC) domain, situated at the C-terminal end of the module, via hydrolysis or cyclisation. Notably, no TE is present in LovB of the lovastatin pathway, instead a condensation (C) domain is responsible for chain release. Moreover, non-reducing iterative type I PKSs in fungi consist of the highly regioselective product template (PT) domain, to control the aromatisation of reactive poly-β-ketone backbones to unique ring patterns (Li *et al*., 2010; Barajas *et al*., 2017).

Modular type I PKS

Modular type I PKSs are large multifunctional enzymes organised as multiple modules of FAS-related domains, mostly found in bacteria. Each module catalyses one round of chain elongation of the polyketide backbone in an “assembly line” paradigm (Cortés *et al*., 1990; Donadio *et al*., 1991). Like iterative type I PKSs, the KS, AT and ACP domains comprise a minimal functional PKS module and optional reductive domains are often present. Domain composition of the reductive loop often differs in modules of the same pathway, conferring a variable degree of reduction for extender units incorporated by different modules. The substrate specificity and stereochemistry of PKS domains are outlined in Section 1.4.4. After the final round of elongation, a thioesterase domain (TE) releases the full-length polyketide chain through hydrolysis or macrocyclisation. A growing number of alternative release mechanisms, including the discrete, type II TE-catalysed chain release in polyether

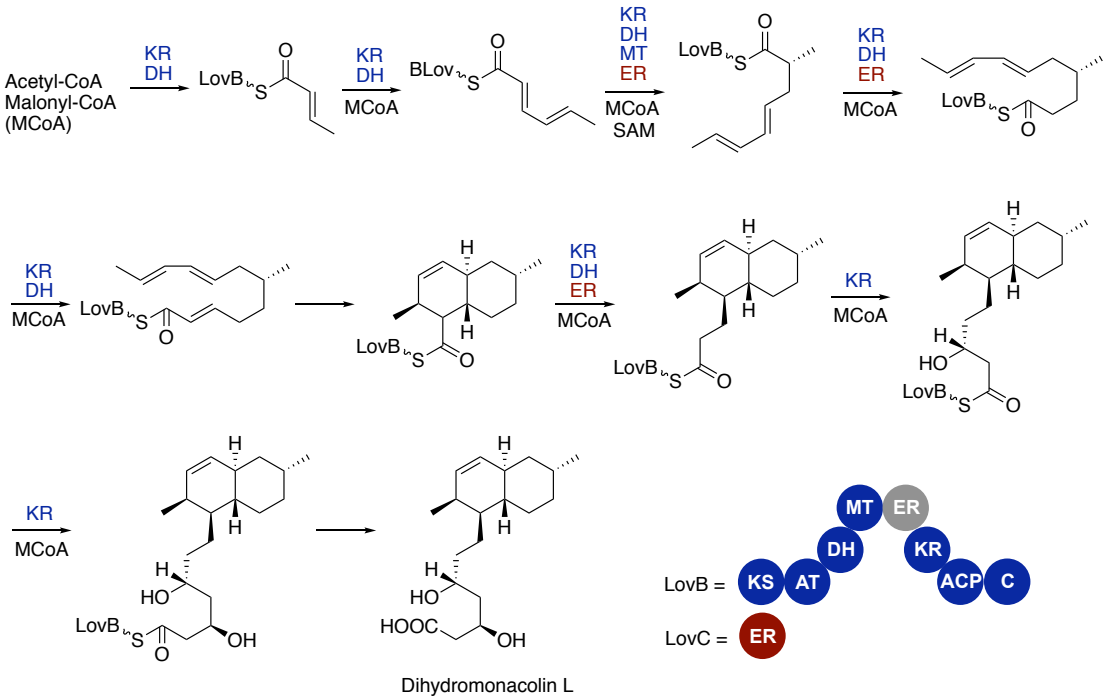


Figure 1.11. Iterative type I PKS pathway to dihydromonacolin L, the precursor to the cholesterol-lowering agent lovastatin, in *Aspergillus terreus*. Reductive and tailoring domain(s) involved in each elongation step is/are indicated in blue (LovB) and red (LovC).

biosynthesis (Harvey *et al*., 2006), thioester reduction in alkaloid biosynthesis (Awodi *et al*., 2017), Baeyer-Villiger oxidative release of the furanone moiety of aurafurons (Frank *et al*., 2007), formation of a tetronate ring (Vieweg *et al*., 2014), and the spontaneous chain release in gliotoxin biosynthesis (Balibar and Walsh, 2006), have also been proposed or reported.

The arrangement of modular type I PKS modules and domains dictates the order of catalytic reactions in the assembly line, a feature referred to as the colinearity rule (Cortés *et al*., 1995; Kao *et al*., 1995; Yu *et al*., 1999). This has been exploited to rationalise the chemical structure of polyketide products, as exemplified by many PKSs such as the 6-deoxyerythronolide B (6-dEB) synthase (DEBS) system that produces the aglycone of the macrolide antibiotic erythromycin A (Cortés *et al*., 1990; Donadio *et al*., 1991) (Figure 1.12). Such modularity and predictability of PKS has prompted intensive research into the generation and optimisation of natural products with desirable properties, via genetic engineering of biosynthetic pathways (Rowe *et al*., 2001; Kendrew *et al*., 2013).

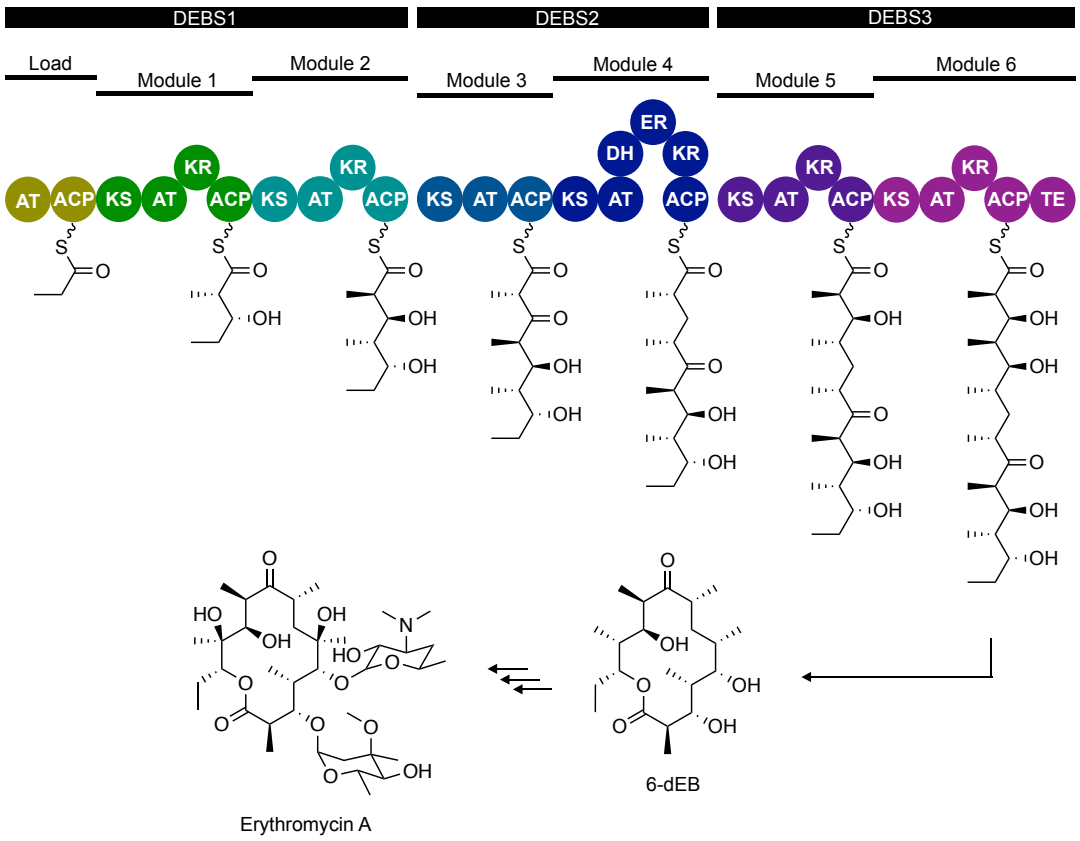


Figure 1.12. Modular type I PKS pathway to the 6-dEB aglycone of erythromycin A. 6-dEB, 6-deoxyerythronolide B.

Tailoring reactions

Additional enzymatic processing of either the PKS-bound or released polyketide intermediate may take place, to further expand the chemical diversity and complexity of polyketides. Such reactions are usually catalysed by tailoring enzymes encoded within the same gene cluster (Betlach *et al.*, 1998; Weitnauer *et al.*, 2001; Chen *et al.*, 2013). A broad range of tailoring reactions, including glycosylation, acylation, amidation, reduction, prenylation, carbamoylation and halogenation, is employed in PKS pathways. Some examples of the more prevalent oxygenation and methylation reactions are briefly described here.

Oxygenation. PKS pathways utilise a range of enzymes to catalyse oxygenation reactions. EryK, a cytochrome P450 monooxygenase in the erythromycin pathway, catalyses the C-12 hydroxylation of erythromycin D to generate erythromycin C, an immediate precursor of the final product, erythromycin A (Stassi *et al.*, 1993; Lambalot *et al.*, 1995). In oxytetracycline biosynthesis, a dioxygenase, OxyL, incorporates oxygen atoms at both C-4 and C12a in an NADPH-

dependent manner (Zhang *et al*., 2008). For the formation of cyclic polyethers, polyolefinic polyketides are oxidised to epoxide intermediates by flavin-dependent epoxidases, in preparation for the subsequent intramolecular ring-opening/cyclisation reactions to form tetrahydrofuran and tetrohydropyran rings (Gallimore *et al*., 2006).

Methylation. *C*- *O*- and *N*-methyltransferases (MTs) may be found in PKS pathways as discrete enzymes or enzymatic domains embedded in modules. In the erythromycin pathway, a discrete, SAM-dependent *O*-MT, EryG, is responsible for methylation at the C3 hydroxyl of the mycarose moiety of erythromycin C to yield erythromycin A (Paulus *et al*., 1990). Some PKS modules housing a SAM-dependent *C*-MT domain are capable of giving rise to the unusual geminal di-methyl group on the α-carbon of the PKS-bound intermediate. HMWP1 of the yersiniabactin pathway prefers malonyl-CoA over methylmalonyl-CoA and the two events of *C*-methylation may take place either before or after condensation is catalysed by KS (Miller *et al*., 2002), whereas EpoM8 of the epothilone pathway specifies methylmalonyl-CoA and KS catalyses the condensation reaction after the formation of the geminal di-methylmalonyl-ACP (Poust *et al*., 2015).

1.4.4 Substrate specificity and stereochemistry of PKS

PKSs are able to generate multiple chiral centres with various hydroxy and alkyl substituents along the carbon backbone of polyketides (Figure 1.13). At least some of the key features have been uncovered that contribute to the catalytic mechanism, substrate specificity and the highly-controlled stereochemistry of individual domains of modular type I PKSs. Many of these insights have been exploited for combinatorial biosynthesis and for the increasingly confident prediction of products of uncharacterised PKS biosynthetic gene clusters (genome mining).

ACP domain

The ACP domain contains a modified four-helix bundle and the catalytic Ser of the highly conserved GxDSL motif is situated at the *N*-terminal end of the second helix. ACP in its activated form (*holo*-ACP) has a flexible 4'-PP arm attached at the Ser residue, and is able to tether the substrate via a thioester linkage and present it to PKS active sites (Vance *et al*., 2016).

AT domain

Each round of polyketide chain elongation is initiated by the AT domain loading an extender unit onto the 4'-PP arm of ACP. The active site of AT is located between the

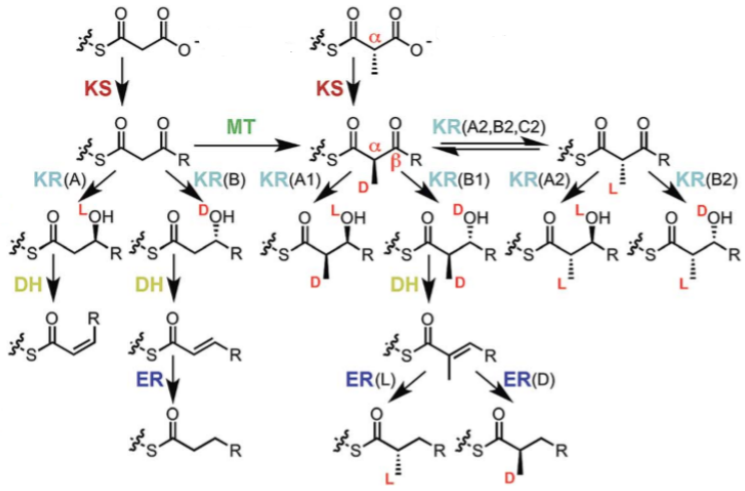


Figure 1.13. Stereocentres controlled by PKS domains. Adapted from Keatinge-Clay (2016).

Malonyl-CoA	Motif I	Motif II	Motif III
avr (5)	QTPYAQ	GHSLGE	HAFH
srm (5)	RTEFAQ	GHSVGE	HGFH
nid (5)	RTEYTQ	GHSVGE	HAFH
epo (4)	QTAFTQ	GHSIGE	HAFH
sor (3)	QTAFTQ	GHSIGE	HAFH
520 (2)	DTLYAQ	GHSIGE	HAFH
rap (7)	ETGYAQ	GHSVGE	HAFH
pic (1)	ETRYTQ	GHSVGE	HAFH
epoAT4 (relaxed)	QTAFTQ	GHSAGE	HASH
Methylmalonyl-CoA			
avr (4)	RADVVQ	GHSQGE	YASH
ery (6)	RVDVVQ	GHSQGE	YASH
nid (1)	RVDVVQ	GHSQGE	YASH
epo (4)	RIDVVQ	GHSMGE	VASH
sor (2)	RVDVVQ	GHSQGE	YASH
520 (5)	RVEVVQ	GHSQGE	YASH
rap (7)	RVDVVQ	GHSQGE	YASH
pic (5)	RVDVVQ	GHSQGE	YASH
Ethylmalonyl-CoA			
srm (1)	RVDVVQ	GHSQGE	TAGH
nid (1)	RVDVVQ	GHSQGE	TAGH
tyl (1)	RVDVVQ	GHSQGE	TAGH
520 (1)	RVDVVH	GHSQGE	CPTH

Figure 1.14. Structural motifs I-III of AT domain that contribute to catalytic and/or extender unit specificity. Adapted from Reeves *et al*. (2001).

αβ-hydrolase subdomain and the ferredoxin-like subdomain. ATs consist of three structural motifs, I, II and III, that contribute to their catalytic activity and/or substrate specificity (Reeves *et al*., 2001) (Figure 1.14). In motif II (GHSxG), if "x", the residue immediately downstream of the catalytic Ser, is a branched hydrophobic amino acid,

the AT is malonyl-specific; and a Gln or Met residue are found in other ATs. Non-native starter or extender units may be incorporated to alter the α-carbon on the polyketide chains through swapping or site-directed mutagenesis of motifs II and/or III to alter the substrate binding motif of AT (Lau *et al*., 1999; Reeves *et al*., 2001; Del Vecchio *et al*., 2003; Lopanik *et al*., 2008).

KS domain

The KS domain belongs to the thiolase superfamily and has an αβαβα topology. KS elongates the polyketide chain via Claisen condensation between the growing polyketide chain and an extender unit. The catalytic triad Cys-His-His located within the conserved TACSSS, EAHGTG and KSNIGHT sequence motifs, is present in all functionally active KSs. Cys acts as a nucleophile to attack the thioester of the acyl-ACP, while the second His contribute to the decarboxylation of the extender unit and the other His functions as a base to activate a water molecule for the nucleophilic addition to the activated carboxylate (Robbins *et al*., 2016).

KSs that utilise potentially chiral extender units such as methylmalonate, are stereospecific. Methylmalonyl extender units are known to arise solely from (2*S*)-methylmalonyl-CoA (Marsden *et al*., 1994). Following the condensation catalysed by KS, the α-methyl group is expected to appear in the elongated polyketide chain in the opposite stereochemical orientation, i.e. with inversion of configuration (Weissman *et al*., 1997; Staunton and Weissman, 2001). In certain extension modules, there appears to be retention of configuration, and this requires an additional epimerisation step catalysed by a side activity of the KR domain of the same module (Xie *et al*., 2016).

As discussed in Section 1.4.5 below, the KS domains of so-called *trans*-AT modular PKSs differ in sequence in a way that reflects the specific type of substrate handed on from the previous module. In canonical modular PKS (*cis*-AT PKSs) KS domains tend to resemble each other much more closely, but it has recently been shown that a short sequence motif in the KS domain correlates with substrate structure (Zhang *et al*., 2017). Interestingly, the identified motif is in the same position as a point mutation that was recently shown to affect the selectivity of an engineered DEBS KS domain (Murphy *et al*., 2016).

KR, DH and ER domains

The KR domain belongs to the short-chain dehydrogenase/reductase (SDR) family. Both the catalytic and structural subdomains possess a highly conserved Rossmann fold. The catalytic subdomain contains the Gly-rich TGxxGxG motif and the NNAG motif for NADPH binding, as well as the YAAAN motif containing the catalytic Tyr. Hydride ion is delivered from NADPH to the β-keto group and the carbonyl oxygen accepts a proton from the catalytic Tyr. The transition state is further stabilised through the formation of a hydrogen bond with a conserved Lys residue in the active site.

The stereochemical orientation of the resulting β-hydroxyl group is set through both the selection of substrate and the stereoselective hydride addition. There are three types of KR: type A and type B KRs give rise to L-β-hydroxy and D-β-hydroxy, respectively, whereas type C KRs are unable to catalyse reduction due to the absence of either the catalytic Tyr or the NADPH binding motif (Caffrey, 2003). The LDD motif and several other fingerprint residues were identified through sequence alignment to distinguish different types of KRs for structural prediction (Reid *et al*., 2003; Baerga-Ortiz *et al*., 2006). Keatinge-Clay (2007) later extended this analysis on the basis of the crystal structure determined for an individual KR domain, to provide a predictive "code", based on additional fingerprint residues. In this code, a number denotes the stereospecificity of individual type A and type B KRs towards α-substituted extender units: A1/B1 and A2/B2 type KRs are specific for α-substituents in the D- and L- orientations, respectively (Figure 1.15).

The DH domain possesses a characteristic double hotdog fold and its catalytic dyad Asp-His is located in the hydrophobic tunnel between the two hotdog folds (Keatinge-Clay, 2008). The Asp in the HPALLD motif protonates the β-hydroxy of the substrate and the His in the HxxxGxxxxP motif abstracts the L-α-proton. The highly conserved Tyr in the YGP motif may facilitate substrate binding. Typically, DHs in canonical PKSs are paired with a type B KR and act on the resulting D-β-hydroxy to yield an α,β-*trans* double bond. There are also rare cases of the dehydration of L-β-hydroxy produced by a type A KR, generating a *cis* double bond through the removal of the L-α-hydrogen (Chang *et al*., 2004; Alhamadsheh *et al*., 2007). Dehydrating bimodules split between two polypepetides are often observed in *trans*-AT PKSs and may generate α,β-*cis* olefin, α,β-*trans*-γ,δ-*cis* or α-β-*cis*-γ,δ-*trans* diene moieties.

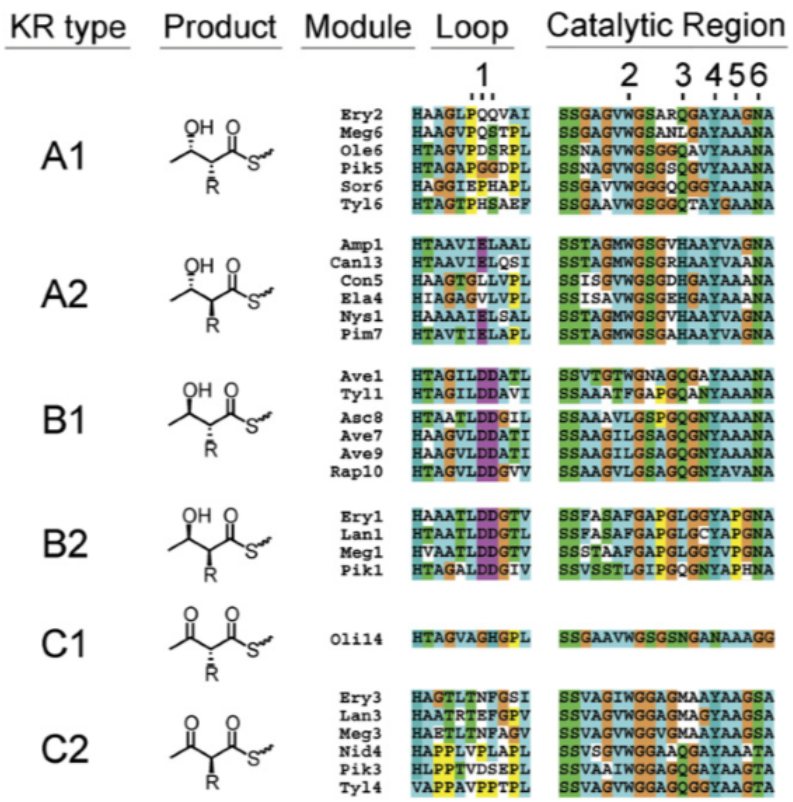


Figure 1.15. Partial sequence alignment of KR domains. The LDD motif and other fingerprint residues are indicated. Adapted from Keatinge-Clay (2007).

The ER domain is a member of the medium-chain dehydrogenase/reductase (MDR) family. It comprises an NADPH binding subdomain and a substrate binding catalytic subdomain. Similar to DH domains, its active site is located between the two subdomains. ERs typically act on trans olefin moieties that are products of canonical DHs paired with a type B KR. The hydride from NADPH is transferred to the β-carbon of the substrate and the resulting transient enolate intermediate is reprotonated at the α-carbon. The Lys and Asp residues in the active site are thought to be important for the binding and stabilisation of intermediates. A specific side-chain in the active site (either Tyr or Val) is postulated to play a role in setting the stereocentre of α-substituents (Kwan *et al*., 2010; Zheng *et al*., 2012).

TE domain

The TE domain has an α,β-hydrolase fold that forms an active site presenting its catalytic triad His-Asp-Ser for substrate binding and catalysis (Tsai *et al*., 2001). It is usually located at the C-terminus of the final module and catalyses either (i) the intramolecular cyclisation of a polyketide to form a macrolactone or (ii) hydrolysis to

release the linear polyketide as a carboxylic acid. The Ser residue acts as a nucleophile and attacks the thioester of the acyl-ACP of the final module, while the His and Asp residues accept the proton from Ser, resulting in an acyl-*O*-TE intermediate. The intermediate is subsequently attacked by either a hydroxy group on the polyketide chain activated by the catalytic His or by water for a hydrolytic release. At least some TEs exhibit broad substrate specificity and are able to act on various substrate mimics (Aggarwal *et al*., 1995).

1.4.5 *Trans*-AT PKS: a non-canonical class of modular type I PKS

Trans-AT PKSs constitute a non-canonical class of modular type I PKS. They are characterised by the lack of an integrated AT domain within modules. Instead, a discrete AT encoded within the same gene cluster acts iteratively *in-trans* in each round of chain elongation (Cheng *et al*., 2003). *Trans*-AT PKS was first described in *Bacillus subtilis* in 1993 (Scotti *et al*., 1993), and since then many more have been found across different genera of bacteria.

Unlike canonical modular type I PKS pathways, *trans*-AT PKSs often deviate from the colinearity rule. The arrangement of modules and domains does not necessarily represent the order of enzymatic reactions that take place, which makes structural prediction of the product challenging and poses a barrier to rational engineering. For instance, the presence of inactive domains (Molnár *et al*., 2000), split modules (Cheng *et al*., 2003), discrete reductive enzymes acting *in-trans* (Arakawa *et al*., 2005), dehydrating bimodules for the generation of diene moieties (Wagner *et al*., 2017), modules to introduce branched extension units (Heine *et al*., 2014), iterative use of modules and skipping of modules (Ng *et al*., 2018; Carvolho *et al*., 2005) are among the unusual features that have been observed in *trans*-AT PKSs Pathways containing a gene encoding more than one *trans*-AT are also known to occur. The additional AT may have a gatekeeping function through the hydrolytic cleavage of stalled intermediates from ACPs (Jensen *et al*., 2012).

Stringent substrate specificity of *trans*-AT KSs

Evolutionary studies suggest that *trans*-AT PKSs diversified through horizontal gene transfer both within *trans*-AT PKSs as well as with canonical type I PKSs, whereas canonical type I PKSs evolved by gene duplication (Nguyen *et al*., 2008; Lohman *et al*., 2015). Some *trans*-AT KSs were found to have remnants of the AT domain of various lengths, which potentially represent the “evolutionary intermediates” between canonical and *trans*-AT PKSs as the AT domain was lost gradually through time.

KSs from *trans*-AT PKSs display stringent substrate specificity and form distinct phylogenetic clades correlating to their substrates (Nguyen *et al*., 2008). Structural studies of various *trans*-AT KSs of different clades demonstrated that the combination of the subtle differences in some active site residues and the conformations adopted by three highly variable loops forming the active site tunnel contribute to their stringent requirement for substrate recognition and processing (Gay *et al*., 2014; Lohman *et al*., 2015) (Figure 1.16). In combination with the use of the colinearity rule, *trans*-AT KS phylogeny-based structural prediction represents a powerful tool to predict uncharacterised natural products governed by *trans*-AT PKSs, as exemplified by the previously uncharacterised thailandamide pathway (Nguyen *et al*., 2008) (Figure 1.17). It also provides useful insights for predicting incompatible domain combinations to optimise rational engineering.

Clade	151	175	215	220	237	280	281	317	321	362	363	365	433	434
I	A	A	M	M	F	T	A	L	I	A	A/T	A	F	S
II	H	A	M	M	L	T	A	L	V/I	F	A	S	M	S
III	A	A/Q/M	L	Q	H	A	A	V	A	E	S	S	A/F	G
IV	S	A	A/L/F	M	Y	T	A	L	I	E	P	A	F	G
V	V	A	F	M	Y	T	A	L	I	E	L	A	F	G
VI	A	A	F	M	Y	T	Y	K	I	E	A	A	F	G
VII	A	A	L	M	F	T	V	L	I	E	S	A	A	G
VIII	S	M	L	M	F	T	V	L	I	E	A	A	A	G
IX	S	M	L	F	Y	T	V	L	I	E	S	A	A	G
X	S	A	L	L	Y/W	G	A/V	L	I	E	A	S	A	G
XI	N	V	Y	M	Y	T	V	L	I	E	S	A	A	S/T
XII	D	A	L/M	A/M	Y	T	V	L	I	E	G/S	A	A	G
XIII	L	A	L	F	Y	T	V	L	V/I	E	A	A	A	G
XIV	G	A	F	L	V	A/S	S/T	V	L	L	C	E	D	G
XV	G	A	L	A	F	T	T	I	I	L	S	A	F	G
XVI	Q	N	H	N	M	Y	A	L	I	D	T	A	L	G

Figure 1.16. Fingerprint residues of different clades of trans-AT KSs. The most highly represented residues that vary between the 16 KS clades are shown for each clade. Adapted from Gay *et al*. (2014).

1.4.6 Overall architecture of PKS

Structural elucidation and mechanistic study of PKS multienzymes have proven challenging, due to their enormous size and the presence of multiple active sites. Early structural models of PKS modules were deduced from information available for the analogous FAS. Structures of high resolution have been resolved for individual domains and didomains by X-ray crystallography or NMR, which provided much insight into substrate specificity and stereochemical control in polyketide biosynthesis (Tsai *et al*., 2001; Alekseyev *et al*., 2007; Zheng *et al*., 2010; Tang *et al*., 2006; Zheng *et al*., 2012).

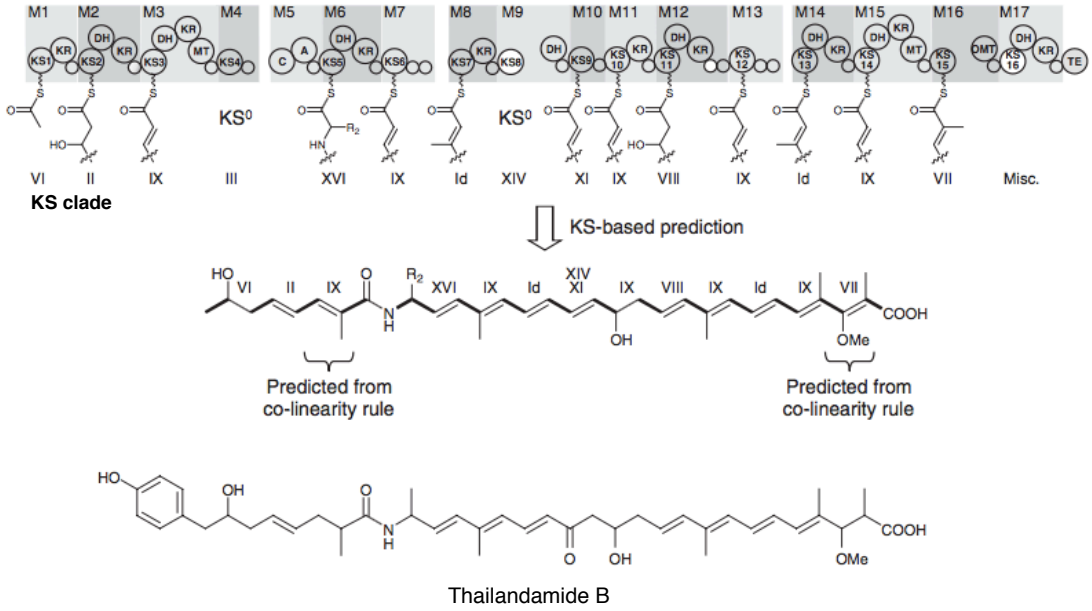


Figure 1.17. Trans-AT KS-phylogeny based structure prediction of thailandamide B. Adapted from Nguyen *et al*. (2008).

The first structures of intact PKS modules were made available in 2014 with the use of small-angle X-ray scattering (SAXS) and cryo-EM, respectively (Edwards *et al*., 2014; Davison *et al*., 2014, Dutta *et al*., 2014; Whicher *et al*., 2014). Similar to the previously established architecture of the porcine FAS module (Maier *et al*., 2008), the modified version of module 3 of DEBS (KS-AT-KR-ACP-TE, additional TE from the same pathway artificially fused to the C-terminus of the module) dimerises at the KSs and shows a flat, disc-shaped overall structure. The ATs are located away from the axis of symmetry and the ACPs reside towards the C-terminal end of the reaction chamber (Edwards *et al*., 2014). In an independent study, module 5 of the pikromycin PKS (KS-AT-KR-ACP) was shown to adopt an overall arch shape instead, with the reaction chamber in the centre of the homodimer and the ATs lying in parallel to the neighbouring KS and KR domains on the side of the reaction chamber (Dutta *et al*., 2014). The transfer of substrate between domains was also investigated in the same study: the ACP occupies different positions within the reaction chamber to access the active site of other domains for catalysis (Dutta *et al*., 2014; Whicher *et al*., 2014). Lowry and colleagues (2016), further proposed a turnstile mechanism for substrate turnover in a PKS module. The module adopts an open conformation for substrate entry and catalysis, the ACP is then ejected from the

central reaction chamber. As the polyketide intermediate exits, the module undergoes conformational change to a closed form, and the cycle repeats.

1.4.7 Non-ribosomal peptides and non-ribosomal peptide synthetases

Non-ribosomal peptides produced by non-ribosomal peptide synthetases (NRPSs) represent another large class of highly diverse natural products. Prominent non-ribosomal peptides in clinical use include the penicillins and other β-lactams, glycopeptide antibiotics like vancomycin, and the immunosuppressant cyclosporine used in transplantation surgeries. NRPSs employ an assembly-line paradigm for non-ribosomal peptide biosynthesis analogous to that of modular PKS. Each extension module selects and activates an amino acid and catalyses its condensation with the elongating peptide chain, before passing on the extended chain to the next module. In addition to proteinogenic amino acids, D-amino acids, methylated, formylated, halogenated and phosphorylated amino acids, and N-terminally attached fatty acid chains are often found in non-ribosomal peptides. As for PKS, the modular nature of NRPS offers opportunities for rational engineering (Mootz *et al*., 2002; Butz *et al*., 2008).

A minimal NRPS module consists of three core domains: a peptidyl carrier protein (PCP) that tethers and presents the elongating peptide chain via a flexible 4'-PP arm, an adenylation (A) domain that activates specific amino acids as an aminoacyl-adenylate for loading onto the PCP to form a thioester derivative, and a condensation (C) domain that catalyses the peptide bond formation between two PCP-bound thioesters. Finally, a TE or C domain situated at the C-terminal end of the final module is usually responsible for chain release via hydrolysis or macrocyclisation. In NRPS pathways lacking a TE or C domain, alternative release mechanims are employed. For example, thioester reductases in the form of a discrete enzyme or an enzymatic domain within an NRPS module have been reported to release an aldehyde product instead (Gaitatzis *et al*., 2001; Kessler *et al*., 2004). Unusual oxidative chain termination resulting in a terminal amide moiety, may also occur, catalysed by a final module whose A domain specifies glycine incorporation and which contains a flavin-dependent monooxygenase (MOx) domain, as described for example in the myxothiazol and melithiazol pathways (Weinig *et al*., 2003; Muller *et al*., 2006).

Tailoring reactions

As in PKS pathways, discrete tailoring enzymes or tailoring enzymatic domains integrated within modules are often present in NRPS pathways, to introduce a variety of functional groups or to modify amino acid side chains. In addition to some tailoring reactions common to PKS pathways, such as oxygenation and methylation, epimerisation and heterocyclisation reactions are often observed in NRPS pathways.

Epimerisation. Integrated epimerisation (E) domains are found immediately downstream of the PCP domain in some NRPS modules. D-amino acids in non-ribosomal peptides are often generated by the action of E domains on the PCP-bound L-aminoacyl-intermediates, via the extraction of the α-carbon proton. The C domain of the following module is usually stereospecific for the epimerised substrate (Kessler *et al*., 2004; Koumoutsi *et al*., 2004). However, there are examples of modules lacking an E domain where a D-amino acid is nevertheless inserted (Patel *et al*., 2003).

Heterocyclisation. In contrast to the heterocyclisation of RiPPs, which requires an ATP-dependent cyclodehydratase complex, heterocycle-forming NRPS modules utilise the heterocyclisation (Cy) domain to introduce azoline moieties. The Cy domain replaces the canonical C domain, and catalyses both peptide bond formation and the cyclodehydration, likely independent of each other (Bloudoff *et al*., 2017). A minimal heterocycle-forming module or bimodule, split between two successive polypeptides of the assembly line, appears to consists of an A domain specifying activation of either Ser, Cys or Thr; a Cy domain; and a PCP (Pulsawat *et al*., 2007). Additionally, an oxidase (Ox) domain is usually present within the module for the formation of azoles from azolines, whereas standalone reductases have been observed in pathways generating azolidines. Tandem copies of Cy and/or Ox occasionally exist, even if only one round of cyclodehydration and dehydrogenation is required to be carried out (Wakimoto *et al*., 2014).

1.4.8 Substrate specificity and overall architecture of NRPS domains

Each NRPS module typically utilises one particular amino acid as the extender unit and such substrate specificity is (at least initially) conferred by the A domain. Conti and colleagues (1997) first reported the structure of the A domain catalytic subunit of GrsA from the gramicidin S pathway, co-crystallised with phenylalanine and AMP. This work revealed that the amino acid binds in a hydrophobic pocket and the carboxyl and α-amino groups interact with the highly conserved Lys and Asp, respectively. Through sequence alignment, Stachelhaus *et al*. (1999) and Challis *et*

al. (2000) further identified 8 to 10 residues in the active site responsible for substrate binding. As the substrate binding pocket is highly conserved, the residues identified from an uncharacterised A domain could be compared to/with domains of known specificity, and used to infer its substrate specificity. From a combinatorial biosynthetic point of view, residues in the substrate binding pocket of an A domain may in principle be manipulated to alter, restrict or expand its specificity (Han *et al*., 2012).

The chirally selective C domains demonstrate a gatekeeping function through reinforcing specificity towards both the donor and acceptor substrates (Stindl and Keller, 1994; Luo *et al*., 2002; Meyer *et al*., 2016). *In vitro* studies by Belshaw *et al*. (1999) and Ehmann *et al*. (2000) have shown the substrate preference for C domains with the use of aminoacyl-CoAs or aminoacyl-*N*-acetylcysteamine thioesters (without the influence of substrate selection by A domains). C-domain structure and function has been recently reviewed (Bloudoff and Schmeing, 2017). It has also been possible to engineer NRPS assembly lines by alteration of PCP domains (Owen *et al*., 2016).

Tanovic and colleagues (2008) reported the structure of SrfA-C from the surfactin pathway, an intact termination module comprising all four essential NRPS domains (C-A-*apo*-PCP-TE) in its *apo*-form. They showed that a rigid catalytic platform forms between the C and A domains, on which the flexibly-tethered PCP (via the A-PCP linker) may adopt different conformations to enable the loading and transfer of substrates. In 2016, the crystal structures of two *holo*-termination modules (C-A-*holo*-PCP-TE), captured at different stages of catalysis, were solved by Drake and colleagues (2016). It was demonstrated that conformational change in the A domain directs PCP for the loading of an activated substrate and the subsequent delivery of the loaded PCP to the C domain to receive the elongating peptide intermediate. It was also revealed that the A and C domains simultaneously catalyse peptide bond formation and amino acid adenylation at their respective active sites. The relative positions of C domain to A domain and TE to other domains in the terminal module varies among different pathways, the authors proposed this could be due to the catalytic platform being highly dynamic and that the structural snapshots may have been captured when different conformations were adopted.

1.4.9 Hybrid PKS-NRPS

Hybrid natural products, containing both ketide and amino acid units, are biosynthesised by relatively uncommon hybrid PKS-NRPS “assembly lines”

containing both PKS and NRPS modules, as exemplified by the epothilone pathway. The epothilone gene cluster encodes nine PKS modules and one NRPS module (Tang *et al*., 2000) (Figure 1.18). Two hybrid interfaces are observed among the first three modules, EpoA (PKS), EpoB (NRPS) and EpoC (PKS). This requires the C domain in EpoB to catalyse a condensation reaction between an acyl building block from EpoA and a peptidyl-PCP, whereas the KS of EpoC has to accept a peptidyl intermediate rather than a polyketide chain from the previous module. Defined N- and C-terminal motifs rich in basic residues are identified at the hybrid interfaces, which are likely facilitating such unusual intermodular interactions and the successful elongation of hybrid polyketide-peptide chains (Gokhale *et al*., 1999; Liu *et al*., 2004).

Beyond hybrid polyketide-peptides, the structural diversity of polyketides is further expanded through the hybridisation of PKS with FAS (Metz *et al*., 2001; Kaulmann and Hertweck, 2002; Shulse and Allen, 2011), and PKS with both FAS and NRPS (Masschelein *et al*., 2013; Masschelein *et al*., 2015).

1.4.10 Heterocycles in polyketide and peptide natural products

Heterocycles are prominent features of bioactive natural products and serve as important scaffold components often required for bioactivity (Joule and Mills, 2010). The incorporation of heterocycles enhances structural complexity and confers conformational rigidity and stability for target binding (Storm and Strominger, 1973). The electronic distribution and molecular properties of heterocycles further contribute to their chemical versatility. Heterocyclic moieties may interact with electrophiles or nucleophiles; may behave as an acid or a base; and often act as ligands to coordinate metallic complexes (Liu and Sadler, 2011; Pozharskii *et al*., 2011).

In addition to azoline formation catalysed by Cy domains in NRPSs and cyclodehydratase complexes in RiPPs systems as discussed in Sections 1.4.7 and 1.4.2, Nature has evolved to afford a wide variety of heterocyclic moieties via diverse biosynthetic routes, and novel chemistry and enzymology continues to be uncovered (Walsh, 2015; Hemmerling and Hahn, 2016). The rest of this thesis will focus on oxazole-containing natural products possibly arising from novel enzymology. Oxazoles and the related azol(in)es are five-membered heterocyclic moieties found in many natural products that are endowed with remarkable biological activities. However, the biosynthetic routes to the relatively uncommon 5-substituted azol(in)e moieties in hybrid polyketide-non-ribosomal peptides and bicyclic oxazoles remain to be elucidated and are the focus of the work described in Chapters 3 to 7.

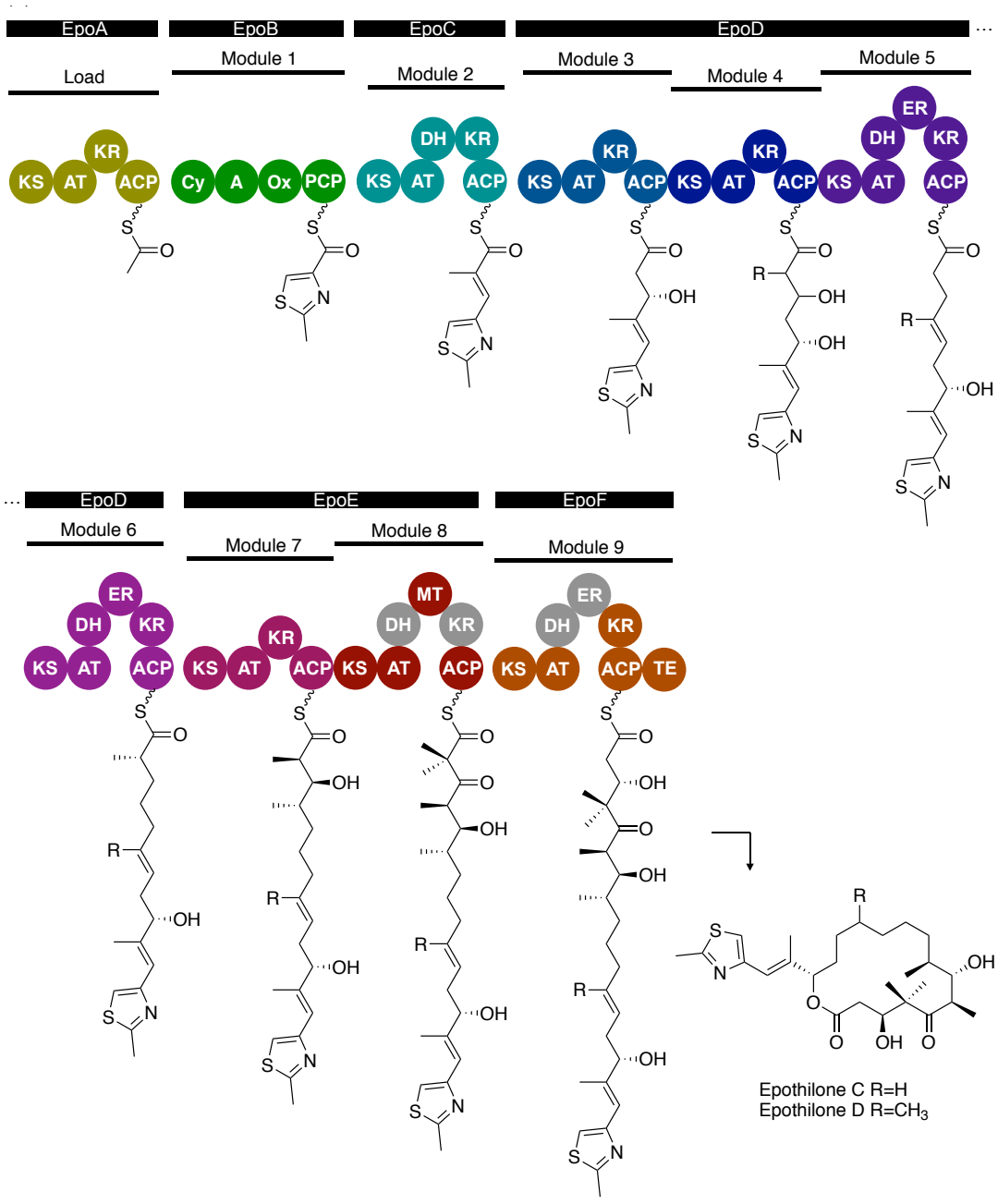
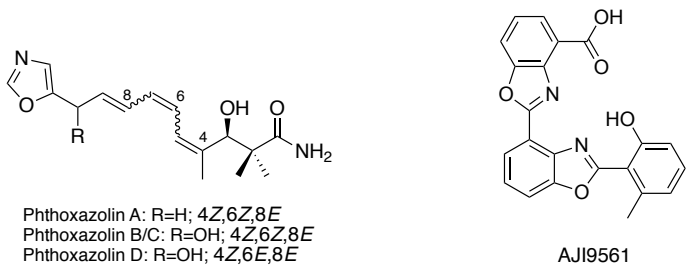


Figure 1.18. Hybrid PKS-NRPS biosynthetic pathway to epothilone.

1.5 Aims of this work

The overarching aim of this work was to use a genomics-led approach to elucidate the key features of the biosynthetic pathways of two (groups of) heterocyclic natural products: phthoxazolins A-D, a group of oxazole trienes with antifungal and herbicidal properties, and the bis(benzoxazole) antibiotic AJI9561 (Figure 1.19). Analysis of the whole-genome sequences of two known phthoxazolin producers, in conjunction with gene deletion experiments, was expected to lead to the identification and characterisation of the phthoxazolin biosynthetic gene cluster, and to shed light on the mechanism of 5-substituted oxazole formation. Meanwhile, a core set of candidate enzymes involved in benzoxazole formation had been revealed by comparative sequence analysis of the biosynthetic gene clusters governing AJI9561 and related benzoxazole antibiotics. The goal here was to establish the enzymatic logic of AJI9561 assembly via an *in vitro* approach.



Phthoxazolin A: R=H; 4*Z*,6*Z*,8*E*
Phthoxazolin B/C: R=OH; 4*Z*,6*Z*,8*E*
Phthoxazolin D: R=OH; 4*Z*,6*E*,8*E*

AJI9561

Figure 1.19. Structures of phthoxazolins A-D and AJI9561.

Chapter 2

Genome analysis of the phthoxazolin-producing strains *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888

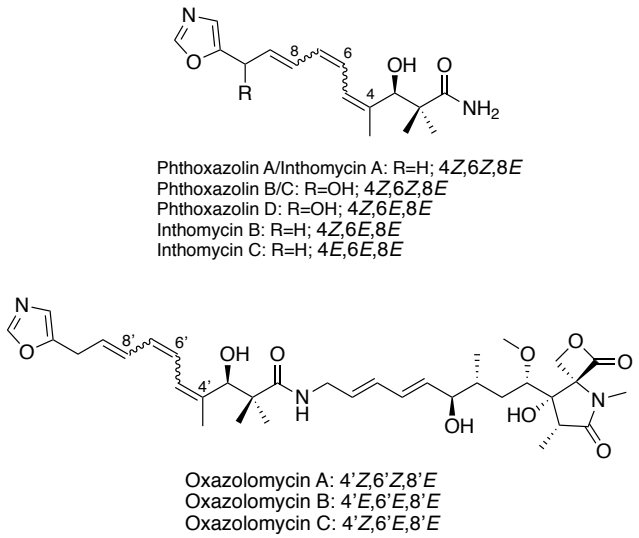
2.1 Introduction

As discussed in Section 1.3, advances in DNA sequencing have led to the increasing availability of complete microbial genome sequences. However, these are still greatly outnumbered in public databases by highly fragmented and poorly annotated sequences. For an accurate estimate of the biosynthetic potential of an antibiotic-producing actinomycete, and to recover the sequence of individual biosynthetic gene clusters intact, a high-quality genome sequence is essential. This is now the method of choice for identifying a target natural product gene cluster.

Phthoxazolin A (Figure 2.1) was discovered in 1989-1991 independently by laboratories in Japan (Ômura *et al.*, 1990), Germany (Henkel and Zeeck, 1991) and France (Legendre and Armau, 1989). Ômura and colleagues used bioactivity screening against cellulose-containing plant pathogens such as *Phytophthora* spp. to isolate phthoxazolin A, and showed it to be an oxazole triene produced by *Streptomyces* sp. OM-5714, a strain obtained from a soil sample collected in Hiroshima, Japan (Ômura *et al.*, 1990). They subsequently developed methods to increase its titre (Tanaka *et al*., 1993).

Phthoxazolin A was shown to inhibit cellulose biosynthesis in both a resting cell system and in a cell-free system from *Acetobacter aceti* subsp. *xylinum*, by targeting the incorporation of glucose and uridine diphosphate-glucose into cellulose (Ômura *et al*., 1990). It also exhibits considerable herbicidal activity against radish seedings and anti-proliferative activity against a human prostate cancer cell line (Ômura *et al*., 1990; Kawada *et al*., 2009). Ômura and colleagues also identified the strain *Streptomyces* sp. KO-7888 as a producer of phthoxazolins B-D (Figure 2.1) (Shiomi *et al*., 1995).

Meanwhile, chemical screening of strain *Streptomyces* Gö-2 by Zeeck and colleagues revealed three phthoxazolins which they named inthomycins A-C (inthomycin A is identical to phthoxazolin A) (Figure 2.1). They rigorously characterised the compounds as geometric E/Z isomers, demonstrated a pattern of ^{13}C incorporation from 1-^{13}C-acetate that confirmed their polyketide origin, and established the configuration at C-3 as (*R*)- (Henkel and Zeeck, 1991), an assignment later confirmed by total chemical synthesis (Hale *et al*., 2014). The herbicidal compound CL22T was originally isolated from the fermentation broth of *Streptomyces griseoaurantiacus* by screening for dicotyledon-specific herbicides and inhibitors of germination (Legendre and Armau, 1989) and it too proved to be identical with phthoxazolin A (Legendre *et al.*, 1995).



Phthoxazolin A/Inthomycin A: R=H; 4*Z*,6*Z*,8*E*
Phthoxazolin B/C: R=OH; 4*Z*,6*Z*,8*E*
Phthoxazolin D: R=OH; 4*Z*,6*E*,8*E*
Inthomycin B: R=H; 4*Z*,6*E*,8*E*
Inthomycin C: R=H; 4*E*,6*E*,8*E*

Oxazolomycin A: 4'*Z*,6'*Z*,8'*E*
Oxazolomycin B: 4'*E*,6'*E*,8'*E*
Oxazolomycin C: 4'*Z*,6'*E*,8'*E*

Figure 2.1. Structure of phthoxazolins A-D and inthomycins A-C, compared to that of oxazolomycins A-C from *Streptomyces albus* JA3453.

The compact structure of phthoxazolins shows several interesting features, including the 5-oxazole moiety, the 2,2-*gem*-dimethyl substituent, and the presence of a primary amide. Strikingly, too, the phthoxazolins are structurally identical to a portion of the oxazolomycin antibiotics (Mori *et al*., 1985; Kanzaki *et al*., 1998), (Figure 2.1) as discussed in more detail in Chapter 3 of this thesis. Elucidation of the biosynthetic pathway to the phthoxazolins was expected to shed light both on the mechanism of oxazole formation and on the relationship between phthoxazolin and oxazolomycin biosynthesis.

In this chapter, the whole-genome sequences of both *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888 are reported, clarifying their phylogenetic

relationship. Detailed analysis and annotation of the genomes led to the identification of the phthoxazolin cluster in both strains, as well as uncovered other previously uncharacterised clusters potentially giving rise to novel natural products.

2.2 Results and Discussion

2.2.1 Whole-genome sequencing of *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888, the phthoxazolin producers

Streptomyces sp. OM-5714 and *Streptomyces* sp. KO-7888 were kindly provided by Professor Satoshi Ômura and Dr. Yuki Inahashi from the Kitosato Institute, Japan. Whole-genome sequences were obtained in-house by Shilo Dickens and colleagues in the NextGen DNA Sequencing Facility, Department of Biochemistry, using both shotgun MiSeq and long-range mate-pair MiSeq data. Data were assembled, using an in-house pipeline, by Dr. Markiyan Samborskyy, and open reading frames were identified and visualised in the genome browser Artemis (Rutherford *et al*., 2000).

Streptomyces sp. OM-5714 has a genome size of 8.7 Mbp with 7789 putative protein-coding sequences (Figure 2.2). The final assembly had a single scaffold, A912_AS4_SC1_R2, covering over 91% of the genome and housing the majority of biosynthetic gene clusters. Six small unplaced scaffolds, which could not be mapped onto the core assembly due to the lack of overlapping regions, were also obtained. They are named A912_AS4_SC05, A912_AS4_SC06, A912_AS4_SC07, A912_AS4_SC08, A912_AS4_SC09 and A912_AS4_SC10, respectively, with biosynthetic gene clusters predicted in the former two unplaced scaffolds. The genome of *Streptomyces* sp. KO-7888 spans 8.5 Mbp, slightly smaller than that of *Streptomyces* sp. OM-5714, and has 7545 putative protein-coding sequences (Figure 2.3). In the final assembly, the single scaffold A913_AS4_SC1_R3 covers over 95% of the genome and also houses the majority of biosynthetic gene clusters. There are five small unplaced scaffolds, A913_AS4_SC08, A913_AS4_SC10, A913_AS4_SC11, A913_AS4_SC12 and A913_AS4_SC13, of which only A913_AS4_SC10 was predicted to contain biosynthetic gene clusters. The G+C content (~72%), coding density and average gene length of the two strains are similar. Neither strain appears to house a plasmid.

2.2.2 Phylogenetic studies and whole-genome comparisons of *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888

To uncover the phylogenetic context of the two phthoxazolin producers, an initial BLAST (Altschul *et al*., 1990) comparison was made of their near full-length 16S rRNA gene sequences to that of other *Streptomyces* spp. in the NCBI database. The dendrogram of the nearest neighbours (Figure 2.4) clearly showed the very close

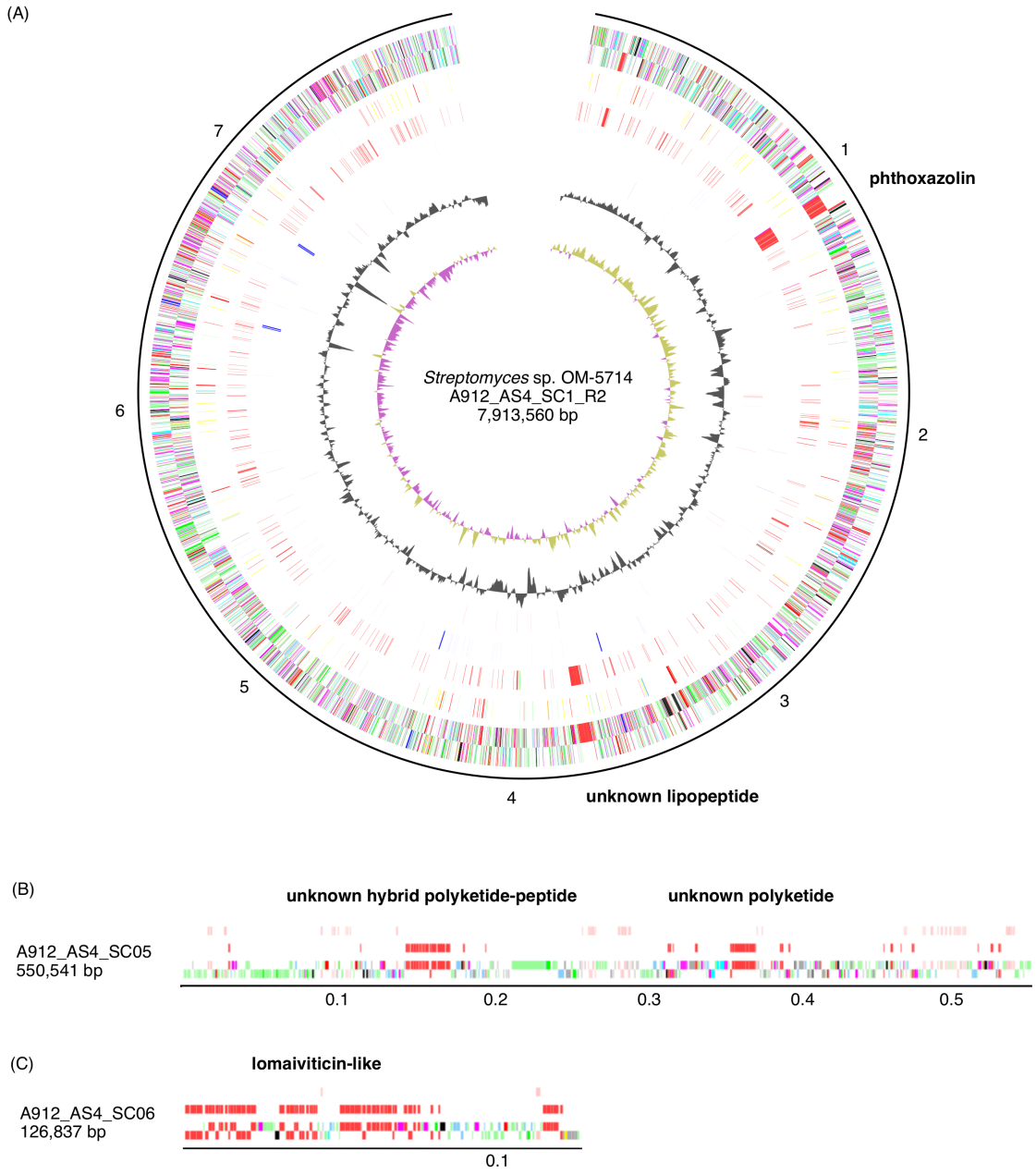


Figure 2.2. Map of the linear *Streptomyces* sp. OM-5714 chromosome. The scales are numbered in megabases. The outermost rings show genes on, respectively, the forward and reverse strands (red, secondary metabolism; black, energy metabolism; pale blue, regulators; yellow, central metabolism; cyan, degradation of large molecules; magenta, degradation of small molecules; dark green, surface associated; orange, conserved hypothetical; pale green, unknown; grey, miscellaneous); the third ring in shows the essential genes for cell division, DNA replication, transcription, translation and amino acid biosynthesis; the fourth ring in shows the genes of secondary metabolism, clearly identifying the position of the large gene clusters for phthoxazolin and an unknown lipopeptide; the fifth ring in

shows the mobile genetic elements (blue, transposases; red, prophages or integrated plasmids); the sixth ring in shows the G+C content; and the innermost ring shows the G+C content bias (khaki, >1; purple, <1). (A) Main assembly. (B-C) Unplaced scaffolds containing biosynthetic gene clusters.

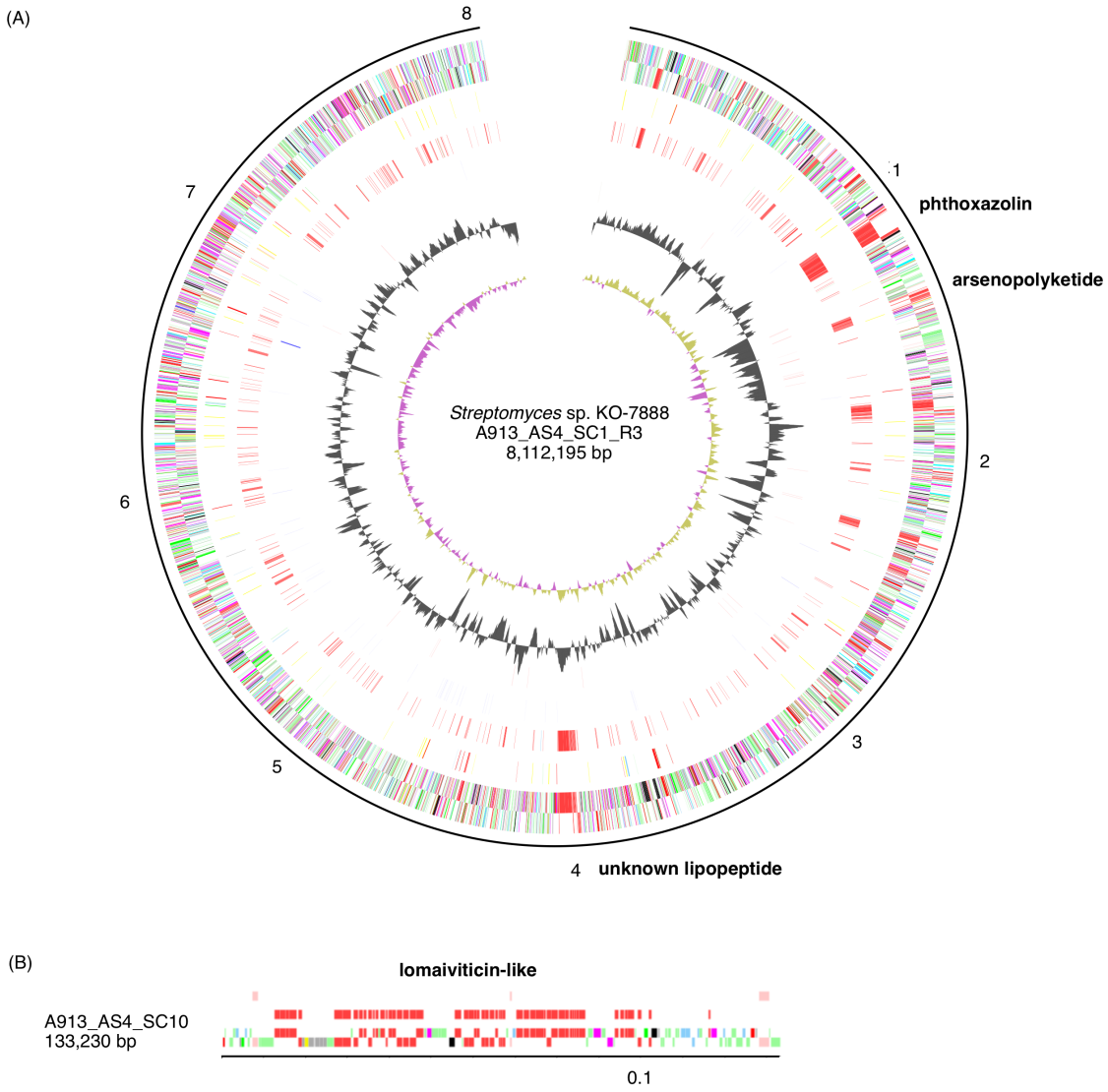


Figure 2.3. Map of the linear *Streptomyces* sp. KO-7888 chromosome. The outermost rings show genes on, respectively, the forward and reverse strands (red, secondary metabolism; black, energy metabolism; pale blue, regulators; yellow, central metabolism; cyan, degradation of large molecules; magenta, degradation of small molecules; dark green, surface associated; orange, conserved hypothetical; pale green, unknown; grey, miscellaneous); the third ring in shows the essential genes for cell division, DNA replication, transcription, translation and amino acid biosynthesis; the fourth ring in shows the genes of secondary metabolism, clearly

identifying the position of the large gene clusters for phthoxazolin, an arsenolipid and an unknown lipopeptide; the fifth ring in shows the mobile genetic elements (blue, transposases; red, prophages or integrated plasmids); the sixth ring in shows the G+C content; and the innermost ring shows the G+C bias (khaki, >1; purple, <1). (A) Main assembly. (B) Unplaced scaffold containing biosynthetic gene clusters.

relationship of *Streptomyces* sp. OM-5714 and *Streptomyces* KO-7888 to each other, and to the type strains of *Streptomyces coelicoflavus*, *Streptomyces thermocarboxydus* and *Streptomyces graminearus*, which all share 100% identity in their 16S rRNA. Although none of these type strains have been genome-sequenced, the genome sequence of *Streptomyces coelicoflavus* var. *nankaiensis* has been published (Geng *et al*., 2009), which differs in its 16S rRNA gene sequence from the type strain by only one base in 1480.

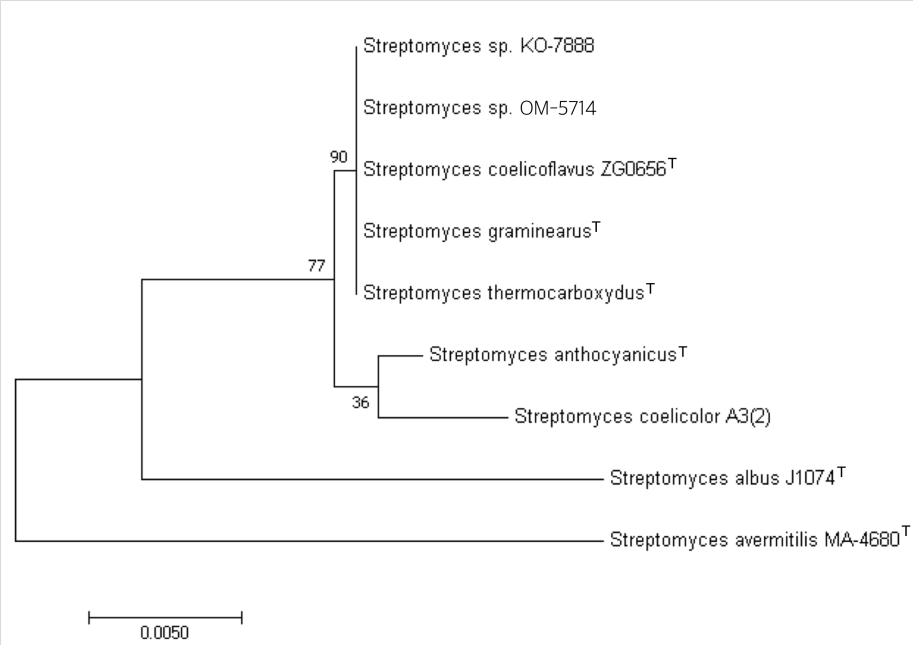


Figure 2.4. Phylogenetic tree based on 16S rRNA gene sequences of *Streptomyces* sp. OM-5714, *Streptomyces* sp. KO-7888 and other *Streptomyces* spp. Evolutionary history was inferred using the maximum likelihood method (Tamura and Nei, 1993). All positions containing gaps were eliminated and there were 1405 nucleotides in the final dataset. The scale bar represents 7 nucleotide substitutions per site. The bootstrap values were inferred from 500 replicates and are shown next to the branches (Felsenstein, 1985). Sequences were aligned with

MUSCLE (Edgar, 2004) and the evolutionary analysis was conducted in MEGA7 (Kumar *et al*., 2016). T, type strain.

Given the high 16S rRNA gene sequence similarity among different *Streptomyces* species that is illustrated in Figure 2.4, a complementary multilocus sequence analysis (MLSA) was performed using the sequences of the five housekeeping genes, *atpD*, *gyrB*, *recA*, *rpoB* and *trpB* (Guo *et al*., 2008; Labeda *et al*., 2017) (Figure 2.5). MLSA evolutionary distances were determined using MEGA7 (Kumar *et al*., 2016) by calculating the Kimura 2-parameter distance (Kimura, 1980). Strain pairs having ≤ 0.007 MLSA evolutionary distance (Kimura 2-parameter distance) are considered conspecific (Rong and Huang, 2012). This analysis suggests the two phthoxazolin producers are variants of the same species (0.003 MLSA evolutionary distance), and distinct from any described *Streptomyces* type strain (≥ 0.014 MLSA evolutionary distance).

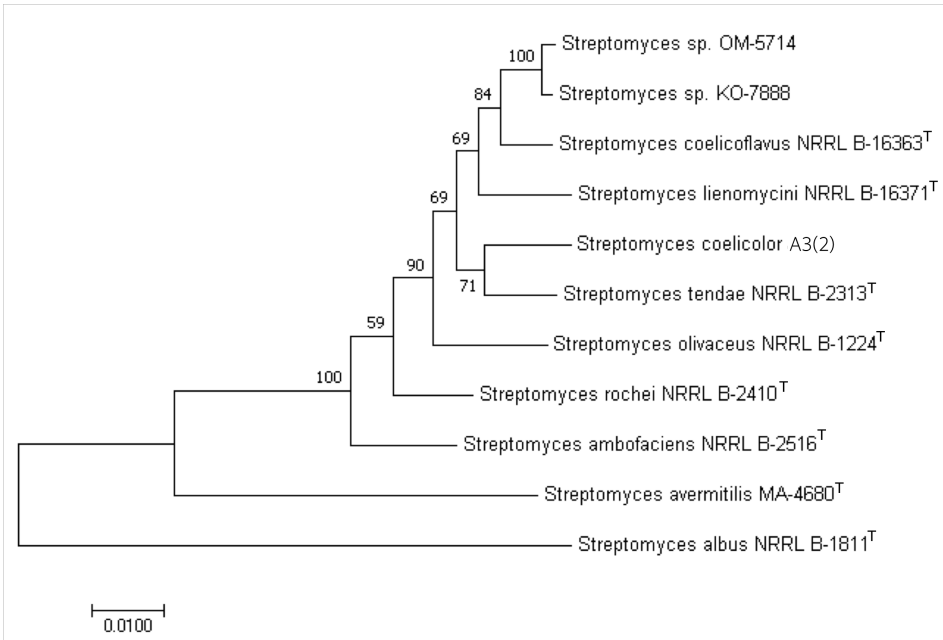


Figure 2.5. Phylogenetic tree based on MLSA of five housekeeping gene sequences of *Streptomyces* sp. OM-5714, *Streptomyces* sp. KO-7888 and other *Streptomyces* spp. Evolutionary history was inferred using the maximum likelihood method (Tamura and Nei, 1993). All positions containing gaps were eliminated and there were 2514 nucleotides in the final dataset. The scale bar represents 25 nucleotide substitutions per site. The bootstrap values were inferred from 500 replicates and are shown next to the branches (Felsenstein, 1985). Sequences were

aligned with MUSCLE (Edgar, 2004) and the evolutionary analysis was conducted in MEGA7 (Kumar *et al*., 2016). T, type strain.

Additionally, MUMmer 3, a bioinformatics software for the rapid alignment of whole (or draft) genome sequences (Kurtz *et al*., 2004), was used to compare the chromosome sequence of *Streptomyes* sp. KO-7888 against those of *Streptomyces* sp. OM-5714 and *Streptomyces coelicolor* A3(2). The alignments were generated and visualised in the form of dot plots by Dr. Markiyan Samborskyy (Figure 2.6). Figure 2.6A illustrates the differences on the chromosome of *Streptomyces* sp. OM-5714 compared to that of *Streptomyes* sp. KO-7888, which correspond to the location of unique biosynthetic gene clusters. No major DNA rearrangement between the two strains was observed. In contrast, various events of DNA inversion were indicated between *Streptomyces* sp. OM-5714 and *Streptomyes coelicolor* A3(2) (Figure 2.6B).

2.2.3 Annotation of the genomes of *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888

AntiSMASH 4.0 was used for the initial detection and analysis of biosynthetic gene clusters (Blin *et al*., 2017), and revealed a total of at least 26 and 24 biosynthetic gene clusters in *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888, respectively. In both phthoxazolin producers, biosynthetic gene clusters are largely located on either ends of the linear chromosome, as observed in many other *Streptomyces* spp. (Gomez-Escribano *et al*., 2016). Manual analysis was subsequently carried out to delimit (and frequently to revise) predicted cluster boundaries. However, the original antiSMASH numbering was retained as closely as possible. Some clusters were judged to be composite and were split into distinct clusters. The full lists of manually annotated clusters are shown in Tables 2.1 and 2.2. The properties of genes encoded in the individual clusters are detailed in Chapter 3 (cluster 4) and in Tables S1 to S46 (all other clusters; see Appendix).

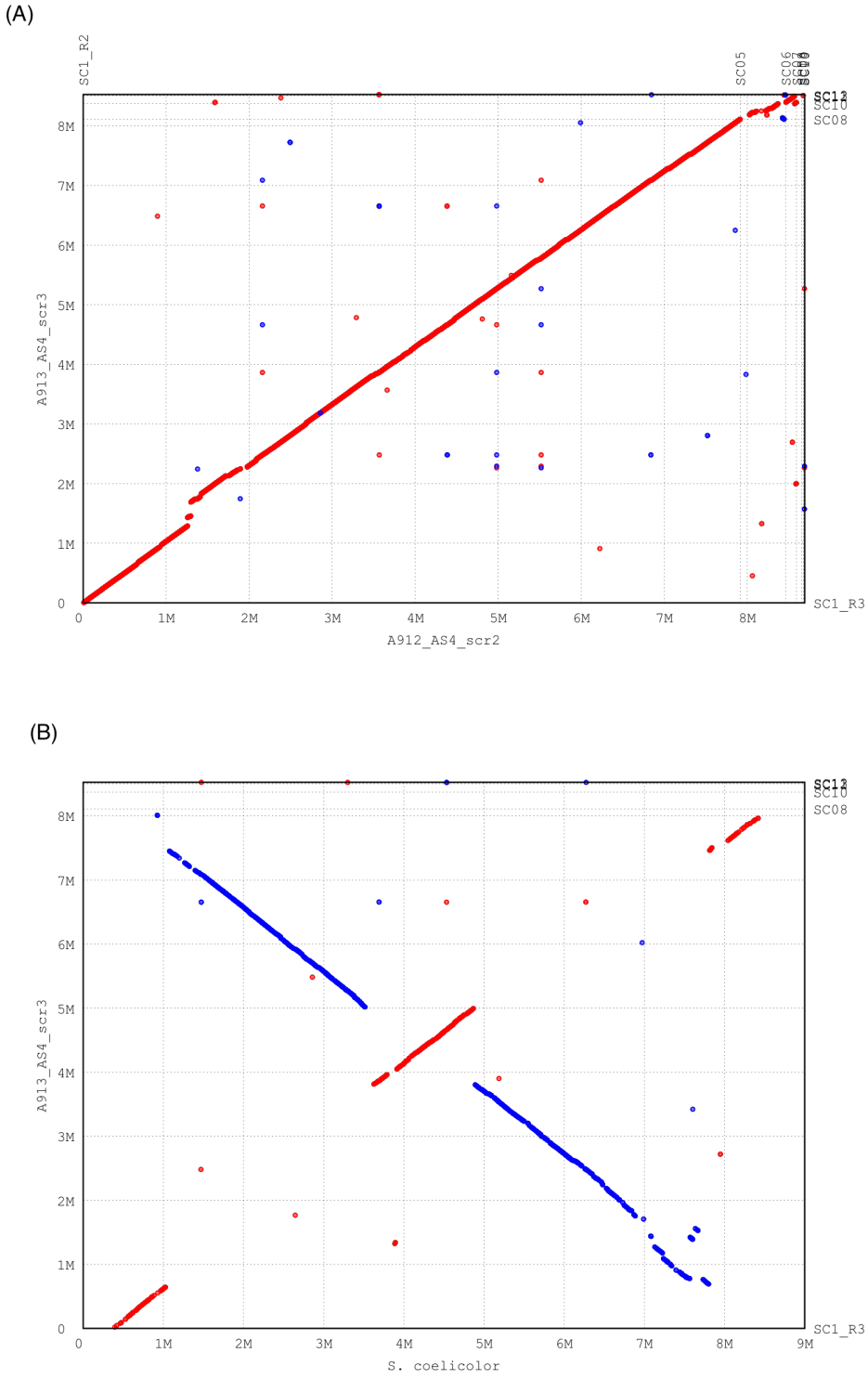


Figure 2.6 Whole-genome sequence alignments of *Streptomyces* sp. KO-7888. The analysis was conducted in MUMmer 3 with a minimum match value of 128 base pairs. Red, forward strand match; blue, reverse strand match. (A) Against *Streptomyces* sp. OM-5714. (B) Against *S. coelicolor* A3(2).

Table 2.1. Annotation of biosynthetic clusters encoded in the genome of *Streptomyces* sp. OM-5714 predicted using antiSMASH 4.0. Clusters were initially analysed using antiSMASH 4.0 but the limits shown here for the clusters were then determined by manual analysis. AntiSMASH numbering of clusters was kept but some clusters were judged to be composite and were split up manually into neighbouring but distinct clusters. ATBH, aminohydroxybacteriohopane. †Biosynthetic clusters unique to *Streptomyces* sp. OM-5714 (not present in *Streptomyces* sp. KO-7888).

Cluster	AntiSMASH 4.0 prediction	Nucleotide		Gene ID		Annotation
		From	To	From	To	
A912_AS4_SC1_R2						
1	Siderophore	126212	149144	0132	0122	Coelichelin
2	Bacteriocin	420880	431095	0394	0400	
3	Terpene	773291	800040	0721	0733	Hopene/ATBH
4	*Trans*-AT PKS-NRPS	1040620	1139534	0923	0955	**Phthoxazolin**
5a†	Butyrolactone	1345705	1346766	1130	-	
5b	-	-	-	-	-	
6a	-	-	-	-	-	
6b	Siderophore	1421033	1430152	1205	1213	Non-NRPS siderophore
7	Terpene	1622455	1631434	1390	1395	Geosmin
8	Bacteriocin	1646219	1657659	1412	1422	
9a†	Terpene	1959829	1968783	1660	1667	Isorenieratene
9b	PKS-NRPS	1978758	2011393	1677	1700	Undecylprodigiosin
10	Siderophore	2091793	2103046	1768	1774	Rhizobactin-like siderophore
11	Type II PKS	2621090	2630022	2227	2235	Spore pigment
12	Terpene	2715710	2718945	2315	2317	Albaflavenone
13	NRPS	3664031	3745384	3202	3236	Unknown lipopeptide
14	Siderophore	5241210	5253199	4615	4624	Desferrioxamine (NRPS-independent siderophore)
15	Melanin	5339462	5352844	4699	4711	
16	Ectoine	6306312	6317261	5562	5572	
17	Bacteriocin	6889825	6902500	6088	6098	
18	Type III PKS	7013081	7015895	6205	6207	Flaviolin
19	Terpene	7569239	7579638	6709	6716	Isorenieratene

20	Indole	7659409	7662093	6796	6797	
21	Terpene	7855680	7859245	6966	6986	2-methylisoborneol
A912_AS4_SC05						
22†	Type I PKS-NRPS	144935	175389	0155	0172	Unknown hybrid polyketide-peptide
23†	Type I PKS-other KS	355058	372595	0312	0317	Unknown polyketide
A912_AS4_SC06						
24	Type II PKS	1	74431	0001	0073	Lomaiviticin-like (partial)
24	Type II PKS	115054	126835	0118	0128	Lomaiviticin-like (partial)
25	Butyrolactone	74929	82201	0074	0080	

Table 2.2. Annotation of biosynthetic clusters encoded in the genome of *Streptomyces* sp. KO-7888 predicted using antiSMASH 4.0. Clusters were initially analysed using antiSMASH 4.0 but the limits shown here for the clusters were then determined by manual analysis. AntiSMASH numbering of clusters was kept but some clusters were judged to be composite and were split up manually into neighbouring but distinct clusters. ATBH, aminohydroxybacteriohopane. ‡Biosynthetic clusters unique to *Streptomyces* sp. KO-7888 (not present in *Streptomyces* sp. OM-5714).

Cluster	**AntiSMASH 4.0 prediction**	**Nucleotide**		**Gene ID**		**Annotation**
		From	To	From	To	
A913_AS4_SC1_R3						
1	NRPS	125731	148660	0124	0135	Coelichelin
2	Bacteriocin	421824	432608	0396	0402	
3	Terpene	792194	806816	0723	0735	Hopene/ATBH
4	*Trans*-AT PKS-NRPS	1072939	1171762	0938	0969	**Phthoxazolin**
5a	-	-	-	-	-	
5b‡	Type I PKS	1386674	1427061	1161	1193	Arsenopolyketide
6a‡	NRPS	1777986	1824374	1556	1577	
6b	Siderophore	1829833	1839074	1587	1595	
7	Terpene	2033328	2042285	1771	1776	Geosmin
8	Bacteriocin	2057332	2068914	1794	1804	
9a	-	-	-	-	-	
9b	PKS-NRPS	2282805	2315645	1972	1985	Undecylprodigiosin
10	Siderophore	2414374	2425627	2079	2085	Rhizobactin-like

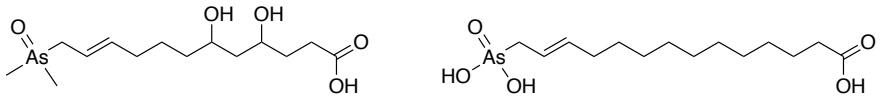
						siderophore
11	Type II PKS	2935852	2943655	2529	2536	Spore pigment
12	Terpene	3047162	3050532	2632	2634	Albaflavenone
13	NRPS	3963064	4044457	3489	3523	Unknown lipopeptide
14	Siderophore	5525881	5539012	4856	4866	Desferrioxamine (NRPS-independent siderophore)
15	Melanin	5624979	5638365	4940	4952	
16	Ectoine	6569243	6580182	5774	5784	
17	Bacteriocin	7132401	7145069	6294	6305	
18	Type III PKS	7255883	7258697	6412	6414	Flaviolin
19	Terpene	7772125	7782459	6888	6895	Carotenoid
20	Indole	7857553	7860237	6968	6969	
21	Terpene	8053745	8057309	7138	7140	2-methylisoborneol
22	-	-	-	-	-	
23	-	-	-	-	-	
A913_AS4_SC10						
24	Type II PKS	12688	94547	0015	0089	Lomaiviticin-like
25	Butyrolactone	95045	102432	0090	0096	

It was revealed that not only are the two strains closely-related to *S. coelicolor* A3(2) phylogenetically, but they also harbour many clusters common to *S. coelicolor* A3(2), including ones giving rise to the terpenes hopene and aminotrihydroxybacteriohopane (ATBH) (clusters 3), geosmin (clusters 7), albaflavanone (clusters 12), isorenieratene (clusters 19 and 9a) and 2-methylisoborneol (clusters 21), the non-ribosomal peptide-derived siderophore coelichelin (clusters 1), the polyketide-derived spore pigment (clusters 11), flaviolin (clusters 18) and undecylprodigiosin (clusters 9b), the siderophores desferrioxamines (clusters 14) and ectoine (clusters 16). The above clusters contain homologues of all essential genes required for the biosynthesis of the predicted products in *S. coelicolor* A3(2), many of which were described in Section 1.4, and will not be discussed any further here. Various clusters putatively encoding bacteriocin, siderophore, indole alkaloid, butyrolactone and melanin pathways were also identified.

Although the two highly-identical phthoxazolin-producing strains house 21 of the same clusters, some clusters were only found in one of them. The genome of *Streptomyces* sp. OM-5714 encodes a second isorenieratene cluster (cluster 9a), an

unknown hybrid type I PKS-NRPS (cluster 22), an unknown type I PKS (cluster 23) and an additional butyrolactone cluster (cluster 5a), none of which are present in *Streptomyces* sp. KO-7888; whereas clusters 5b and 6a, predicted to encode an arseno-PKS and an NRPS pathway, respectively, were only found in *Streptomyces* sp. KO-7888.

Cluster 5b is homologous to a gene cluster only recently discovered in *S. coelicolor* A3(2) (*sco6837-sco6812*) and *S. lividans* 66 (*sli_1077-sli_1103*) (Cruz-Morales *et al*., 2016). The clusters direct the biosynthesis of a novel arsenopolyketide. Homologues of the polyketide synthase (Orf1168), enzymes responsible for C-As bond formation (Orf1171, Orf1176-Orf1177), arsenic resistance and regulatory proteins (Orf1165, Orf1172) were identified in our cluster. The exact structure of the metabolite has not been reported, but two structures were deduced from liquid chromatography-mass spectrometry (LC-MS) data (Figure 2.7). By mining for the co-location of genes coding for C-As bond-forming enzymes, Cruz-Morales and colleagues (2016) revealed arseno-related clusters are unexpectedly prevalent in Actinobacteria.



Molecular formula: $C_{14}H_{27}AsO_5$

Figure 2.7. Structures of potential arsenopolyketides.

2.2.4 Bioinformatics analysis of uncharacterised biosynthetic gene clusters

Five previously uncharacterised biosynthetic gene clusters were identified from the whole-genome sequences of *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888. Present in both of the strains are a hybrid trans-AT PKS-NRPS (clusters 4), a lipopepide NRPS (clusters 13) and a lomaiviticin-like type II PKS (clusters 24). The remaining two clusters include another hybrid PKS-NRPS (cluster 22) and a cluster encoding polyunsaturated fatty acid synthase-related enzymes (cluster 23), both of which were only found in *Streptomyces* sp. OM-5714. In this section, these five clusters were analysed bioinformatically and evaluated for their capacity to generate new chemical diversity. The prediction of the incorporation of certain functional groups and the overall core structure of the plausible products were deduced from the substrate specificity of biosynthetic genes (whenever possible) and/or the

comparative sequence analysis on partial gene clusters with known homologues in the literature.

Clusters 13: a novel lipopeptide product of an assembly-line NRPS?

Analysis of clusters 13 suggests that it might encode biosynthesis of a novel lipopeptide related to known cyclic peptide and depsipeptide antibiotics, such as daptomycin and surfactin (Koumoutsi *et al*., 2004; Miao *et al*., 2005; Koumoutsi *et al*., 2004). Cluster 13 in *Streptomyces* sp. OM-5714 spans a region of 81.4 kbp and contains 34 contiguous opening reading frames predicted to encode three NRPSs, four transporters or proteins potentially conferring resistance, five regulators, four enzymes involved in fatty acid biosynthesis and various tailoring enzymes (Figure 2.8). Thirteen NRPS extension modules are encoded within the three multimodular NRPS-encoding genes (*orf3232-orf3230*), with one of the modules lacking an A domain and predicted to be inactive. The substrate specificity of A domains were predicted using SANDPUMA, a phylogenetics-inspired algorithm which combines the existing active-site motif-based algorithms for improved substrate prediction (Chevrette *et al*., 2017). A C-terminal TE domain is detected in Orf3230 and is potentially responsible for chain release.

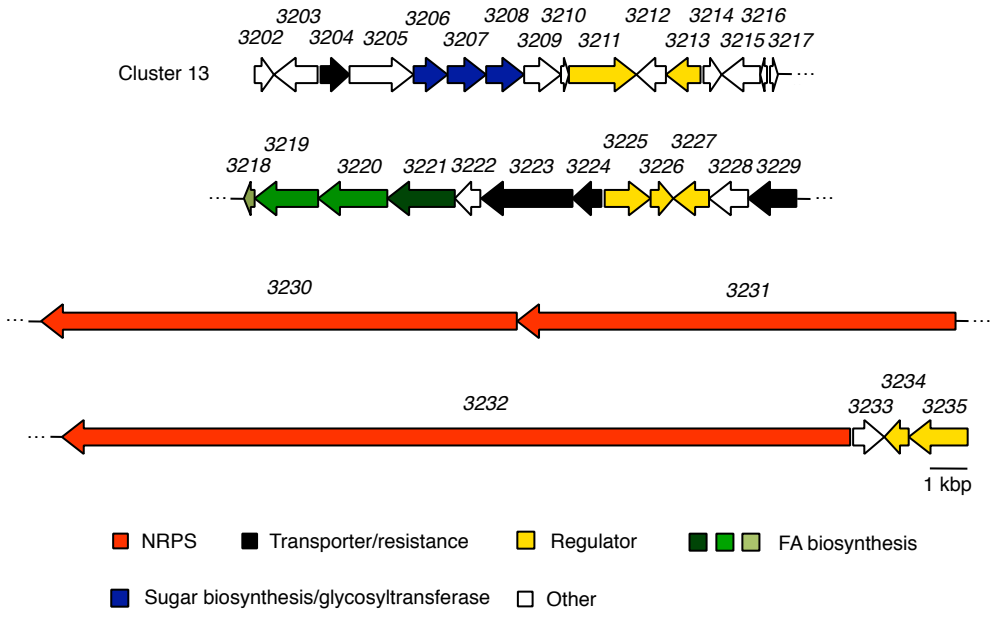


Figure 2.8. Gene organisation of cluster 13 in *Streptomyces* sp. OM-5714.

Genes encoding enzymes responsible for the activation and incorporation of a fatty acid moiety (*orf3221-orf3218*) were identified: acyl-CoA ligase (Orf3221) and

discrete PCP (Orf3218) are homologues of DptE and DptF from the daptomycin pathway. Daptomycin is a lipopeptide antibiotic comprised of a cyclised tridecapeptide coupled to a mid-chain fatty acyl side chain. DptE activates branched fatty acyl chains in an ATP-dependent manner and the fatty acid chains are transferred to DptF (Wittman *et al*., 2008). The C domain of the first NRPS module then catalyses the condensation between the fatty acid and the first amino acid to initiate daptomycin biosynthesis. Notably, two acyl-CoA dehydrogenase-encoding genes (*orf3220* and *orf3219*) are located in tandem immediately downstream of the acyl-CoA ligase-encoding gene. This gene arrangement has been observed in several biosynthetic gene clusters giving rise to unsaturated lipopeptides, such as ramoplanins, taromycins and enduracidins (McCafferty *et al*., 2002; Yamanaka *et al*., 2014; Yin *et al*., 2006) (Figures 2.9 and 2.10). In the enduracidin cluster, the acyl-CoA ligase is fused with the first acyl-CoA dehydrogenase.

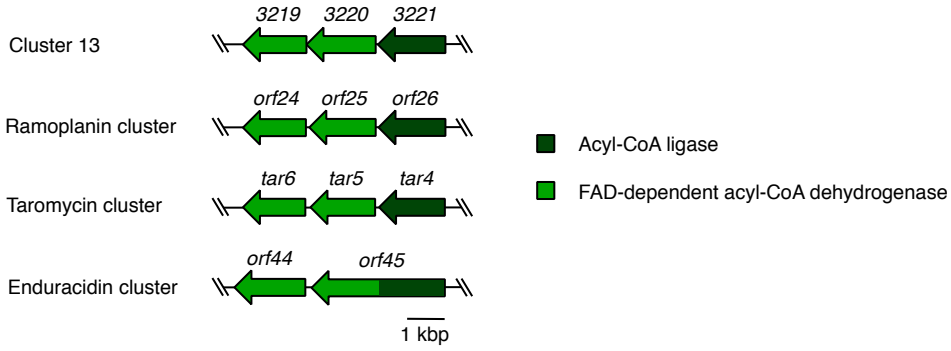


Figure 2.9. Set of candidate genes involved in the biosynthesis of unsaturated fatty acyl chains in ramoplanins, taromycins and enduracidins, and their homologues in cluster 13 of *Streptomyces* sp. OM-5714.

Cluster 13 also houses adjacent genes coding for a glucose-1-phosphate thymidylyltransferase (Orf3208) and a thymidine-5'-diphospho(TDP)-glucose 4,6-dehydratase (Orf3209), which would together convert glucose-1-phosphate into TDP-4-dehydro-6-deoxy-D-glucose. Other deoxysugar biosynthetic genes may be encoded elsewhere in the genome, so the exact identity of the eventual sugar donor is unknown. The glycosyltransferase encoded by *orf3207* may be responsible for the installation of such a sugar moiety onto the peptide backbone. In summary, cluster 13 may therefore goven the biosynthesis of a novel dodecapeptide incorporated with a rare unsaturated fatty acyl chain and a sugar moiety. A linear representation of the predicted lipopeptide backbone is shown in Figure 2.11.

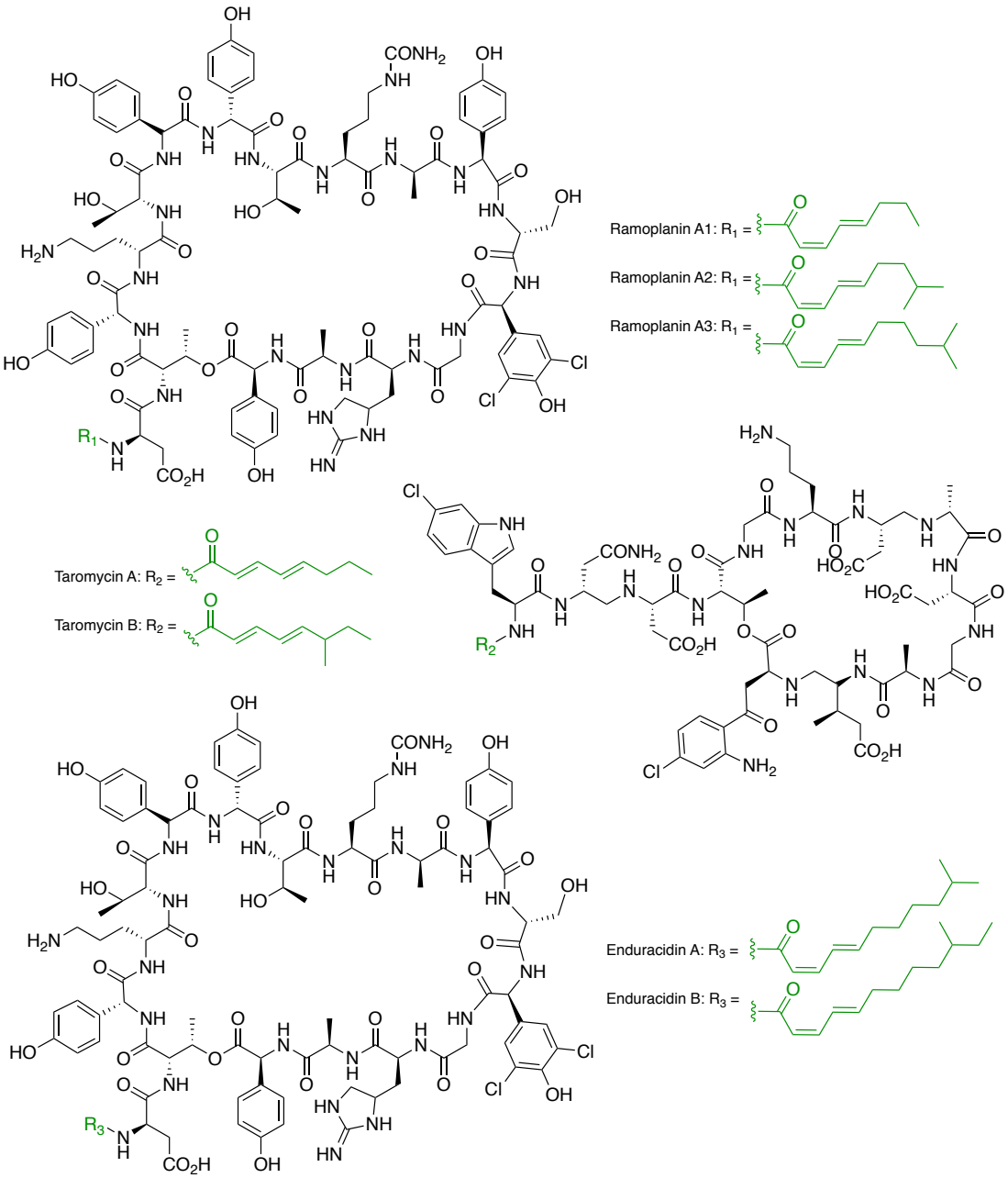


Figure 2.10. Structures of the unsaturated fatty acyl chain-containing lipopeptides, ramoplanins, taromycins and enduracidins.

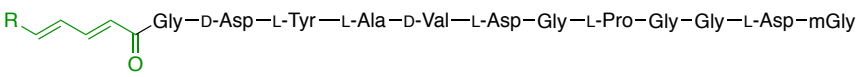


Figure 2.11. Linear representation of the acyl-peptide backbone of the lipopeptide predicted to be encoded by cluster 13. Substrate specificity for A domains was predicted using SANDPUMA (Chevrette *et al*., 2017). m, methylated.

Cluster 22: a hybrid polyketide-peptide featuring a novel starter unit derived from coronafacic acid biosynthesis?

Cluster 22 is unique to *Streptomyces* sp. OM-5714 and appears to encode an unusual PKS-NRPS biosynthetic pathway. The gene organisation of the cluster is shown in Figure 2.12. Notably, the cluster houses a set of genes (*orf0166*, *orf0163-orf0160*) encoding enzymes from the early part of the pathway to coronatine and the related coronafacoyl-amino acid conjugates. These are phytotoxins produced by the plant pathogens *Pseudomonas syringae* and *Streptomyces scabiei*, respectively (Rangaswamy *et al*., 1998; Bignell *et al*., 2010; Bignell *et al*., 2018) (Figures 2.13 and 2.14).

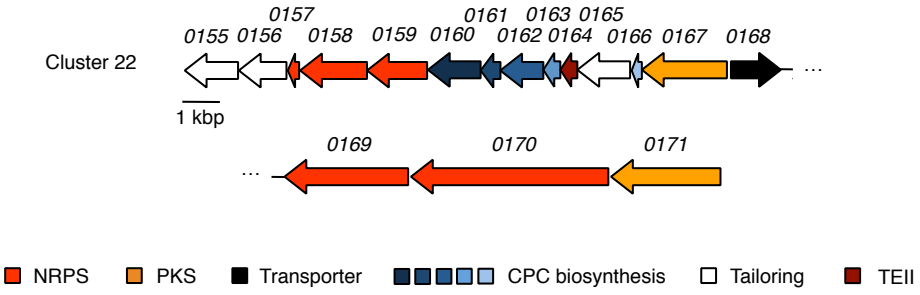


Figure 2.12. Gene organisation of cluster 22 in *Streptomyces* sp. OM-5714. CPC, 2-carboxy-2-hydroxycyclopentanone.

Cfa5, an acyl-CoA ligase, activates succinic semialdehyde, derived from 2-oxoglutarate catabolism, by adenylation and then Cfa3, a KS, catalyses a single round of decarboxylative condensation between the adenylate and a malonate extender unit. Cfa4, a putative cyclase, and Cfa2, a DH, subsequently catalyse the formation of the Cfa1-bound 2-carboxy-2-hydroxycyclopentanone (CPC) (Figure 2.13B). The homologues of Cfa1-Cfa5 in cluster 22 may assemble the enzyme-bound CPC as a starter unit for the hybrid PKS-NRPS assembly line. In any event, the association of cfa genes with a gene cluster for PKS-NRPS biosynthesis is unprecedented.

Cluster 23: polyunsaturated fatty acid synthase-related?

Cluster 23 is another previously-uncharacterised biosynthetic gene cluster unique to *Streptomyces* sp. OM-5714. The 17.5 kbp cluster harbours gene coding for polyunsaturated fatty acid synthase-related enzymes (Orf0316-Orf0313), a type II

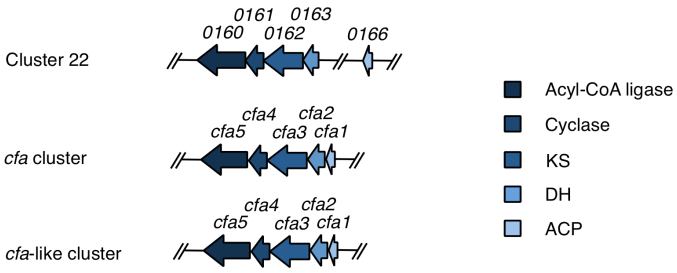


Figure 2.13. Comparative sequence analysis of *cfa1-5* in *cfa* cluster from *Pseudomonas syringae*, *cfa*-like cluster from *Streptomyces scabiei* and their homologues in cluster 22 of *Streptomyces* sp. OM-5714.

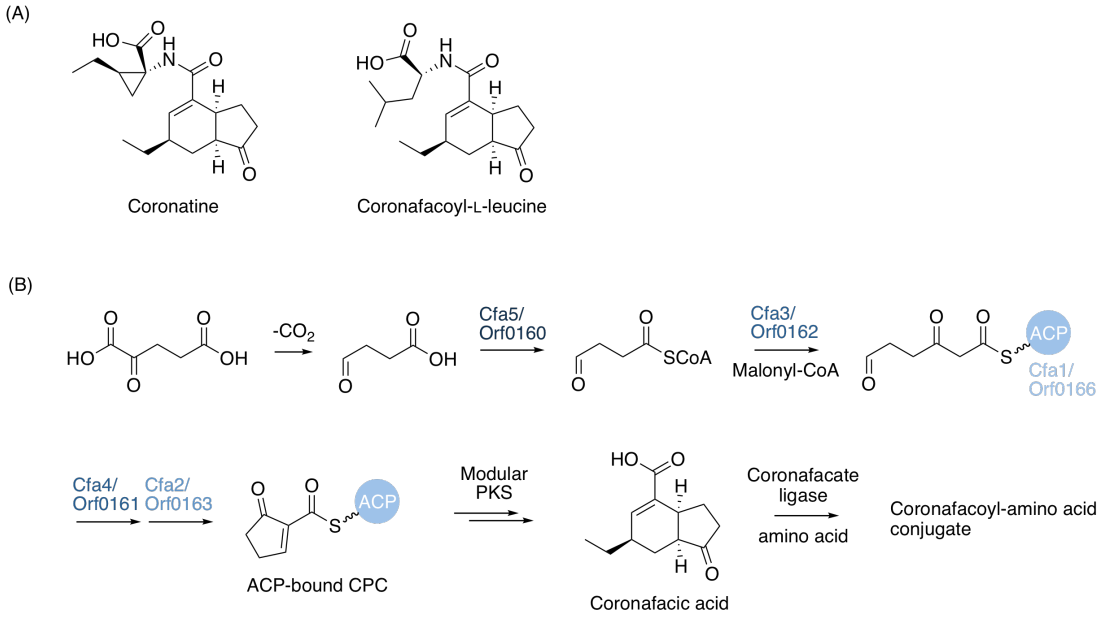


Figure 2.14. Biosynthesis of coronafacic acid-derived molecules. (A) Structures of coronatine and coronafacoyl-L-leucine. (B) Biosynthetic pathway to coronafacoyl-amino acid conjugates in *Streptomyces scabiei*, with the use of a CPC starter unit. CPC, 2-carboxy-2-hydroxycyclopentanone.

thioesterase (Orf0317), a discrete ACP (Orf0318) and a ROK-family transcriptional regulator (Orf0312). Homologues of *orf0318-orf0313* with an identical gene arrangment were identified using protein-protein BLAST in a variety of actinomyete bacteria, including *Saccharopolyspora*, *Kutzneria*, *Actinopolyspora* and *Kitasatospora*. In *Kitasatospora* sp. Root107, the homologues of *orf0316* and *orf0315* are fused into a single open reading frame (Figure 2.15). The high conservation of

the cluster across the Actinomycetales encourages the view that the cluster is expressed under at least certain conditions.

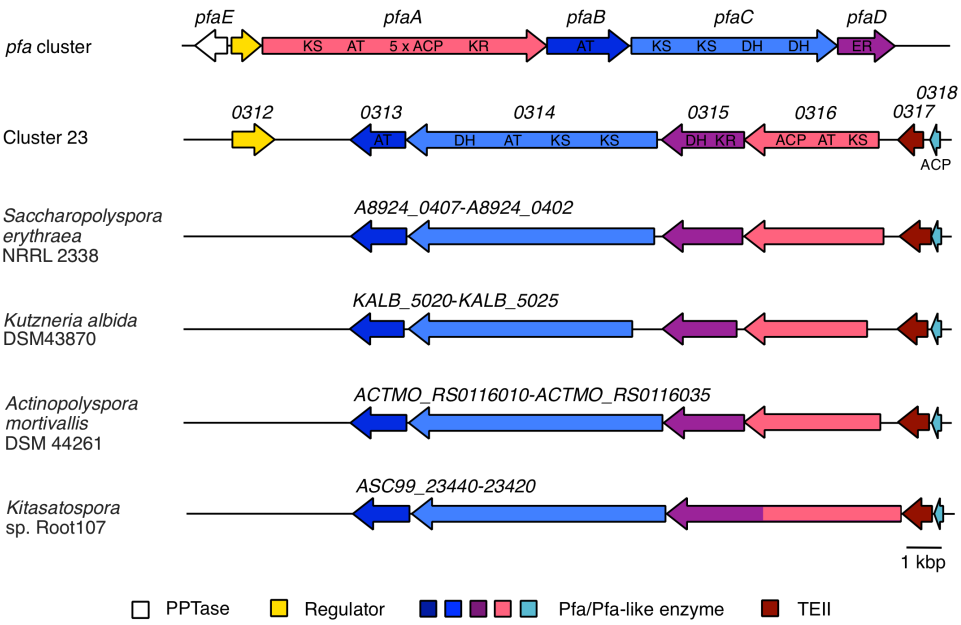


Figure 2.15. Gene organisation of cluster 23 in *Streptomyces* sp. OM-5714. Homologous clusters in *Saccharopolyspora erythraea* NRRL 2338, *Kutzneria albida* DSM 43870, *Actinopolyspora mortivallis* DSM 44261 and *Kitasatospora* sp. Root107. A canonical polyunsaturated fatty acid (*pfa*) cluster from *Shewanella pealeana* is shown. The fatty acid synthase(-related) domains are indicated.

Clusters 24: an atypical angucycline-like aromatic polyketide PKS?

Atypical argucyclines are a small but emerging group of aromatic polyketide natural products, typically characterised by a benzo[*b*]fluorene core. They have attracted considerable attention due to their potent anti-proliferative activity. Among the first to be described were the dimeric lomaiviticins from *Salinospora* spp. (He *et al.*, 2001; Colis *et al.*, 2014; Woo *et al.*, 2012; Kersten *et al.*, 2013; Janso *et al.*, 2014), kinamycins (Ito *et al.*, 1970; Gould *et al.*, 1997; Gould *et al.*, 1998; Bunet *et al.*, 2011), nenestatin A (Jiang *et al.*, 2017), and the related benzo[*a*]fluorene fluostatins from *Streptomyces* spp. and *Micromonospora* spp. (Akiyama *et al.*, 1998; Baur *et al.*, 2006) (Figure 2.16).

Lomaiviticins are potent antibacterial and cytotoxic agents comprising two *O*-glycosylated benzo[*b*]fluorene moieties symmetrically linked via a C2-C2' bond. Lomaiviticins A and B contain two diazofluroene moieties, whereas their analogues

lomaiviticins C-E are heterodimers of a diazofluorene and a hydroxyfulvene. Initial analysis of cluster 24 in both *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888 by antiSMASH 4.0 indicated that approximately half of the genes in this large cluster show considerable similarity to those in the lomaiviticin cluster. Although the gene order is very different, this prompted a closer analysis of the pathway to a putative kinamycin/lomaiviticin-like natural product.

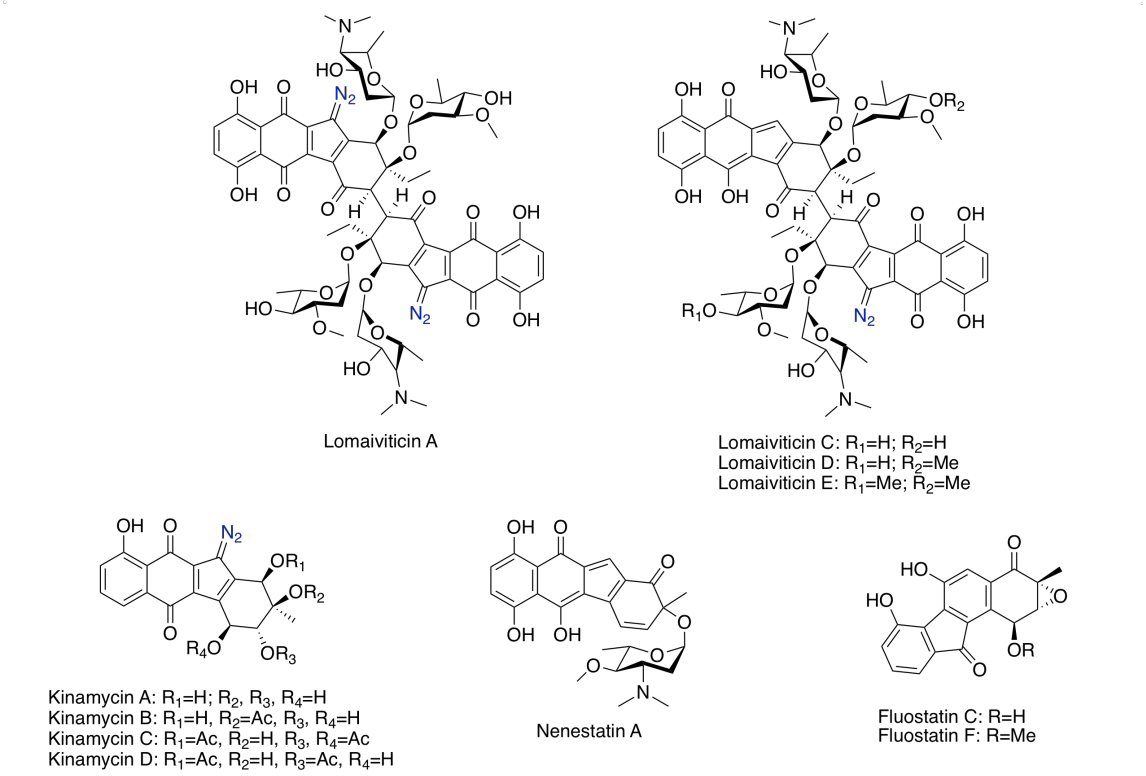


Figure 2.16. Structures of lomaiviticins and related benzofluorene-containing natural products.

The aromatic core of benzofluorene-containing natural products is commonly assembled by type II PKS systems and modified by a series of (redox) tailoring enzymes, as exemplified in many pathways giving rise to angucyclines and angucyclinones (Fernandez-Moreno *et al*., 1992; Decker and Haag, 2005; Yang *et al*., 1996). Cluster 24 contains the “minimal” type II PKS components, KS (Orf0016), KSβ (Orf0015), and ACP (Orf0079), as well as the homologues of tailoring enzymes used in the jadomycin pathway to yield the precursor to angucyclines (Orf0078, Orf0080, Orf0077, Orf0076, Orf0062) (Meurer *et al*., 1997; Kulowski *et al*., 1999). It has been demonstrated in the kinamycin pathway that an adjacent pair of oxidases, AlpJK, catalyse the B-ring opening, contraction and hydroxylation reactions (Wang *et*

al., 2015a); the homologues in cluster 24, Orf0082 and Orf0081, likely catalyse the same series of B-ring redox transformations to afford hydroquinone-kinobscurinone (Figure 2.17).

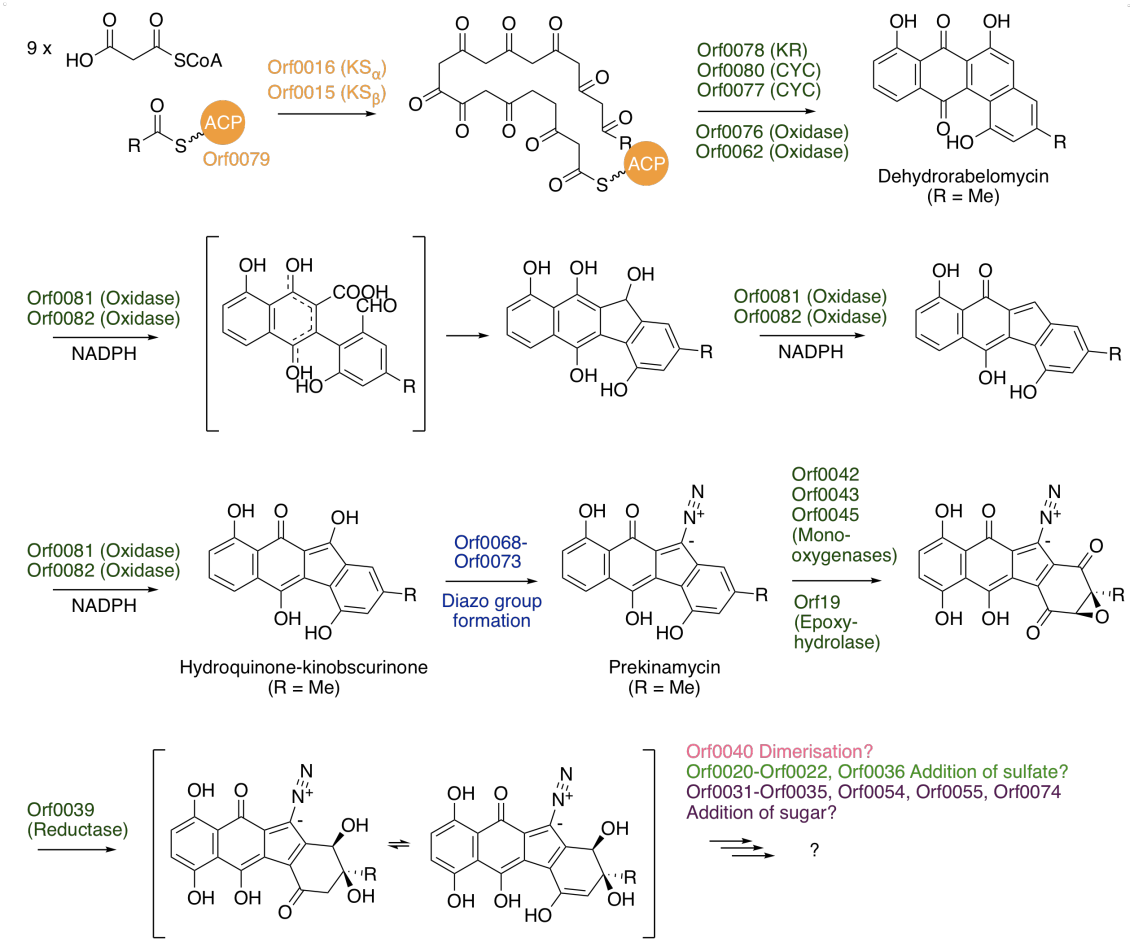


Figure 2.17. Proposed biosynthetic pathway to the product of cluster 24.

The biochemical role of various biosynthetic components, including those responsible for diazo formation and dimerisation, are yet to characterised. Through comparative sequence analysis, Janso and colleagues (2014) identified a set of six enzymes proposed to be responsible for diazo formation from the *lom* and *kin* clusters, including an *N*-acetyltransferase, an adenylosuccinate lyase), a glutamine synthase, an amidase, a 4Fe-4S ferredoxin and a hypothetical protein, to convert hydroquinone-kinobscurinone to prekinamycin. Homologues of all six candidate enzymes are, again, present in our cluster (Orf0068-Orf0073). Intriguingly, the same set of enzymes is also found in the *nes* and *fls* clusters whose products do not contain a diazo group (Figure 2.18). It is proposed that nenestatin and fluostatin

biosynthesis possibly involve diazo intermediates that later undergo diazo reduction (Janso *et al*., 2014; Jiang *et al*., 2017). It is unclear whether or not the product of cluster 24 would contain a diazofluorene moiety.

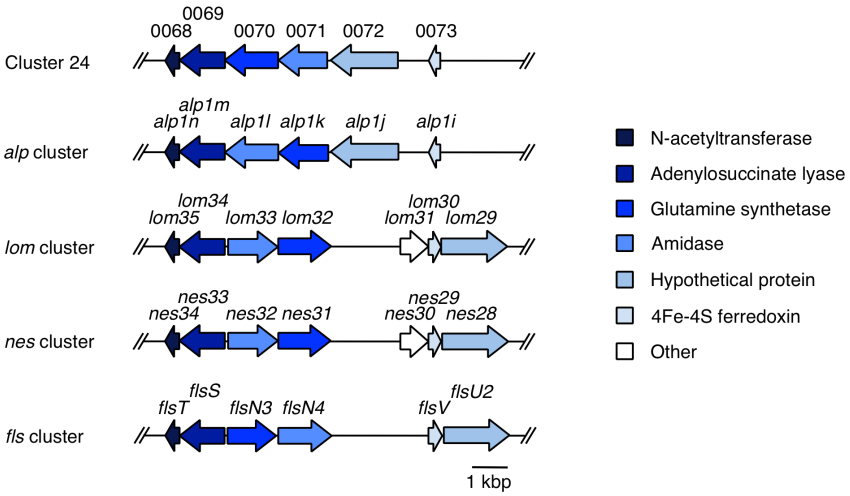


Figure 2.18. Organisation of candidate genes involved in diazo group installation in cluster 24 in Streptomyces sp. KO-7888. Homologous sets of genes from the kinamycin (*kin*) cluster in *Streptomyces ambofaciens* ATCC 23877, the lomaiviticin (*lom*) cluster in *Salinispora pacifica* DPJ-0019, the nenestatin (*nes*) cluster in *Micromonospora echinospora* SCSIO 04089 and the fluostatin (*fls*) cluster in *Micromonospora rosaria* SCSIO N160 are shown.

Following diazo formation, prekinamycin undergoes a series of oxidative A-ring modifications. Wang and colleagues (2015b) showed that the conversion of prekinamycin to kinamycin F involves an epoxy intermediate in kinamycin biosynthesis, and confirmed the epoxidase acitivty of an αβ-hydrolase, Alp1U. The hydrolases/monooxygenases (Orf0045, Orf0043-Orf0042) and the Alp1U-homologue (Orf0019) in cluster potentially play similar roles. Orf0039 may catalyse a subsequent reduction to generate substrates for dimerisation by an NADPH-binding protein homologous to Lom19 (0040). Lom19 was identified as the candidate dimerisation enzyme by both Kersten *et al*. (2013) and Janso *et al*. (2014), due to its significant sequence identity to ActVA-orf4 in actinorhodin biosynthesis (Taguchi *et al*., 2012). It was proposed that ActVA-orf4 and Lom19 may employ a similar strategy for catalysing the regiospecific C-C bond formation between the monomeric precursors to form the dimeric actinorhodins and lomaiviticins, respectively.

Homologues of some sugar biosynthetic genes involved in the TDP-L-oleandrose pathway are located throughout cluster 24 (*orf0033-orf0035*, *orf0054, orf0055*, *orf0074*). The TDP-L-oleandrose pathway and the closely related TDP-L-mycarose pathway in the biosynthesis of various polyketides linked with the respective sugar moieties have been extensively studied (Summer *et al*., 1998; Gaisser *et al*., 1998; Salah-Bey *et al*., 1998; Chen *et al*., 1999; Draegar *et al*., 1999). For instance, in the avermectin pathway, a set of seven enzymes, AveDCGIFHE, encoding a glucose-1-phosphate-thymidylyltransferase, a TDP-glucose 4,6-dehydratase, a TDP-4-keto-6-deoxyglucose 2,3-dehydratase, a TDP-4-keto-6-deoxy-L-hexose 3-ketoreductase, a TDP-4-keto-6-deoxyglucose 3-epimerase, a TDP-6-deoxy-L-hexose 3-*O*-methyltransferase and a TDP-4-ketohexulose reductase, respectively, catalyses the conversion of glucose-1-phosphate to TDP-L-oleandrose (Figure 2.19). Homologues of the first six enzymes in the TDP-L-oleandrose pathway are encoded in cluster 24, but no AveE homologue was found anywhere on the chromosome. A TDP-4-ketohexulose-reductase other than AveE may catalyse the reduction to yield TDP-L-oleandrose. One or more of the glycosyltransferases encoded in the cluster (Orf0031, Orf0032, Orf0051) may be responsible for the sugar transfer to the aglycone.

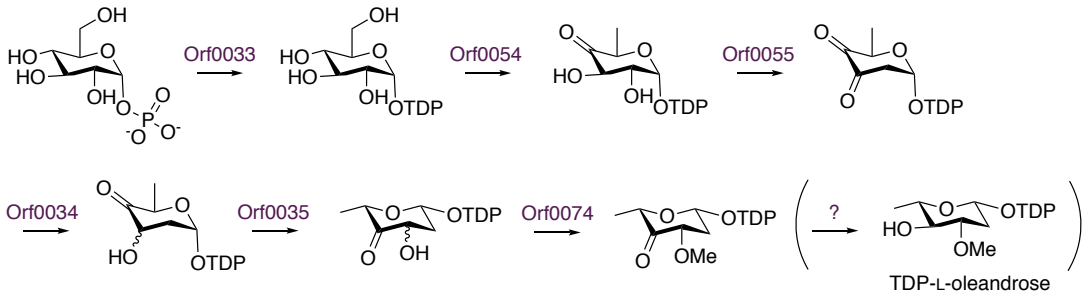


Figure 2.19. Biosynthetic pathway to TDP-L-oleandrose.

Intruguingly, cluster 24 also houses sulfate adenylyltransferase subunits (Orf0021 and Orf 0022) and an adenylylsulfate kinase (Orf0020) for generating the sulfo donor, 3'-phosphoadenyl-5'-phosphosulfate (PAPS) (Figure 2.20), and a putative sulfotransferase. The latter enzyme, encoded by *orf0036*, is a homologue of *lom10*. Since known lomaiviticins are not sulfated, the significance of the presence of these genes in cluster 24 is not clear.

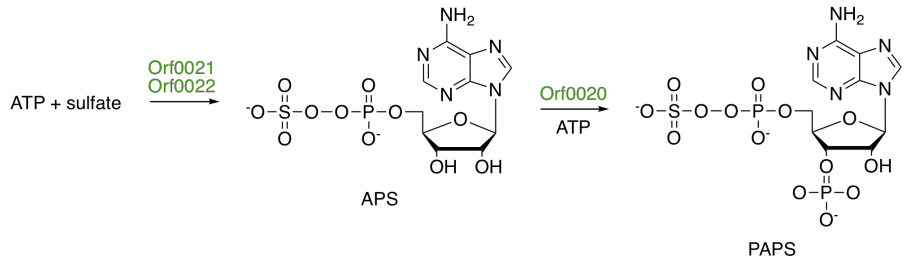


Figure 2.20. Biosynthetic pathway to PAPS. APS, adenyl-5'-phosphosulfate; PAPS, 3'-phosphoadenyl-5'-phosphosulfate.

2.2.5 Identification of the putative phthoxazolin biosynthetic gene cluster

The structure of phthoxazolins suggest a polyketide-peptide origin, as for the analogous portion of oxazolomycin (Zhao *et al*., 2010). Two clusters in *Streptomyces* sp. OM-5714 were predicted as hybrid PKS-NRPSs (clusters 4 and cluster 22). The genes and enzymes of cluster 22, which was only found in *Streptomyces* sp. OM-5714, were described earlier in Section 2.2.4. Cluster 4, a hybrid *trans*-AT PKS-NRPS, consists of the reductive (KR and DH) domains required for successive αβ-double bond formation to afford the triene moiety, as well as the methylation domains for *C*-methylation to generate the unusual germinal dimethyl group. Cluster 4 is present in the genome of both phthoxazolin producers and was determined as the only plausible candidate for phthoxazolin biosynthesis.

2.3 Concluding remarks

The whole-genome sequences of two phthoxazolin producers, *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888, have been obtained by shotgun nextgen DNA sequencing using the Illumina platform. Phylogenetic analysis has revealed the two phthoxazolin producers are variants of the same species and, although they are apparently distinct from previously-described type strains, they are nevertheless closely related to the model organism *Streptomyces coelicolor* A3(2). Detailed analysis and annotation of both genomes showed that most of the recognised biosynthetic gene clusters are shared with *S. coelicolor* A3(2) and with each other. However, certain shared clusters were apparently novel to this species, and a very few found only in either *Streptomyces* sp. OM-5714 or *Streptomyces* sp. KO-7888, housing novel combinations of biosynthetic genes. Meanwhile, our collaborators, Dr. Yuki Inahashi and Prof. Satoshi Ōmura (Kitasato Institute, Japan), used the whole-genome sequences we determined to identify cluster-situated regulatory genes. Overexpression of the SARP family transcriptional regulator located in the NRPS gene cluster (cluster 13) in *Streptomyces* sp. KO-7888 led to the production and isolation of two novel unsaturated lipopeptides. There are notable structural differences between the predicted product, based on bioinformatic analysis (see Section 2.2.4), and the elucidated structures (Figure 2.21). It was also a surprise that the lipopeptides contain a primary amide, as with phthoxazolins. Finally, the phthoxazolin gene clusters of each strain were identified with high confidence. The detailed analysis and characterisation of these clusters are discussed in Chapters 3, 4 and 5, the following chapters of this thesis.

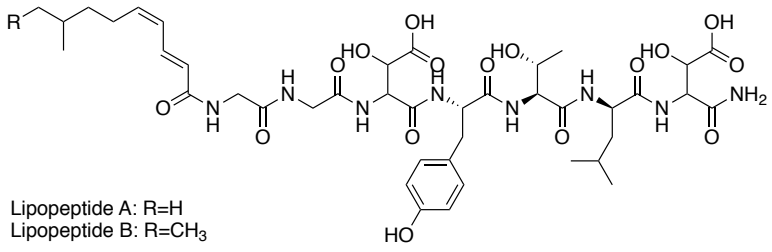


Figure 2.21. Structure of unsaturated lipopeptides isolated from *Streptomyces* sp. KO-7888.

Chapter 3

Characterisation of the unusual *trans*-AT PKS-NRPS gene cluster for phthoxazolin biosynthesis

3.1 Introduction

The putative phthoxazolin clusters from *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888 were identified as a new member of the growing list of hybrid *trans*-AT PKS NRPS clusters. Hybrid *trans*-AT PKS-NRPSs, often decorated with unusual biosynthetic and/or tailoring components, give rise to structurally-diverse, bioactive polyketide-peptides including mupirocin, lankacidin, and bryostatin. Notably, the putative phthoxazolin clusters were revealed to be unexpectedly large, consisting of many non-canonical features and PKS- and NRPS-encoding genes that do not appear necessary for phthoxazolin biosynthesis. This chapter describes detailed bioinformatic analysis of the cluster, and experiments using in-frame deletion and point mutagenesis of various genes, to establish the biosynthetic role of their gene products, leading to a proposal for the biosynthetic pathway to phthoxazolins.

3.2 Results and Discussion

3.2.1 Detection of phthoxazolins in *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888

Streptomyces sp. OM-5714 and *Streptomyces* sp. KO-7888 are known to produce phthoxazolin A, and phthoxazolins A-D, respectively. For both producer strains, peaks corresponding to the proton and sodium adduct ions of phthoxazolin A and phthoxazolins B-D were detected using liquid chromatography-mass spectrometry (LC-MS), with a retention time of 19.1 min and 16.3/16.9 min, respectively (Figures 3.1 and 3.2). In accord with the literature, they all showed the expected MS/MS fragmentation pattern and a UV absorption maximum at approximately 276 nm to 278 nm with shoulders on either side of the peak (characteristic of trienes). Phthoxazolins B, C and D could not be distinguished from each other using LC-MS, as the three stereoisomers share the same molecular weight and fragmentation pattern. Furthermore, the accurate masses of phthoxazolin A and phthoxazolins B-D were confirmed using IMS-MS-QTOF (MS Service, Department of Chemistry).

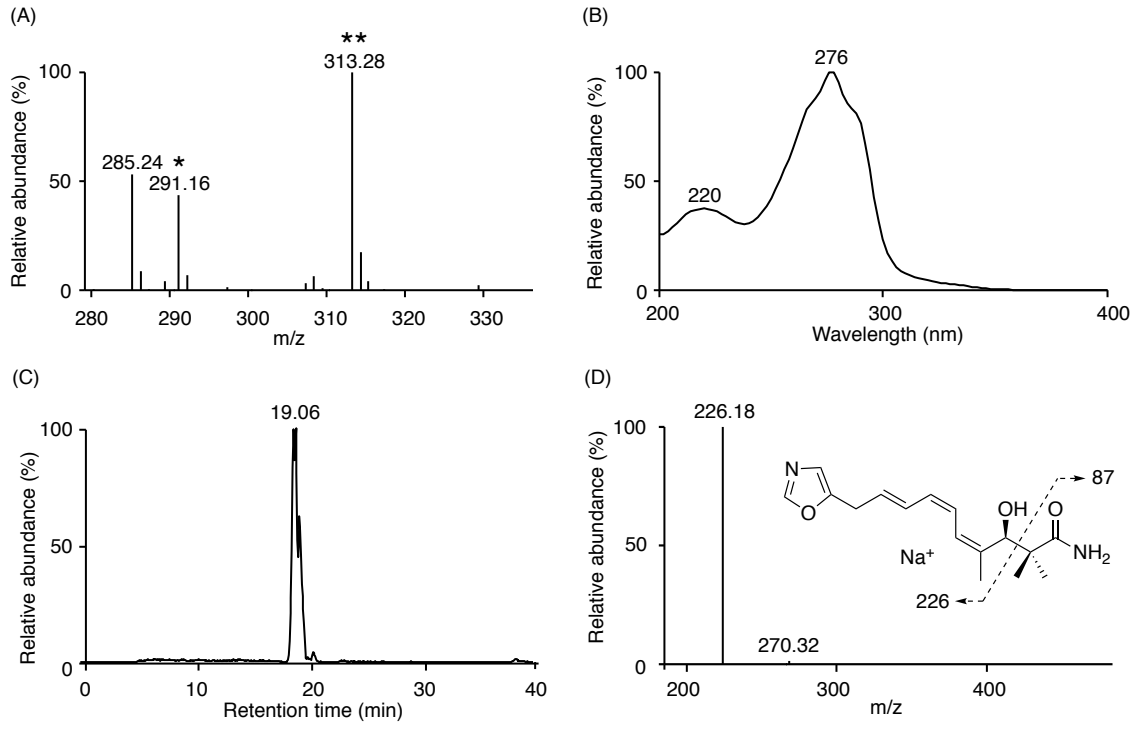


Figure 3.1. Detection of phthoxazolin A from *Streptomyces* sp. KO-7888 by LC-MS. (A) MS at 18.6-19.6 min. *, proton adduct ion; **, sodium adduct ion. (B) UV spectrum at 18.6-19.6 min. (C) Selected ion monitoring of *m/z* $[M+Na]^+$ 313.15. (D) MS/MS fragmentation of parent ion *m/z* $[M+Na]^+$ 313.15.

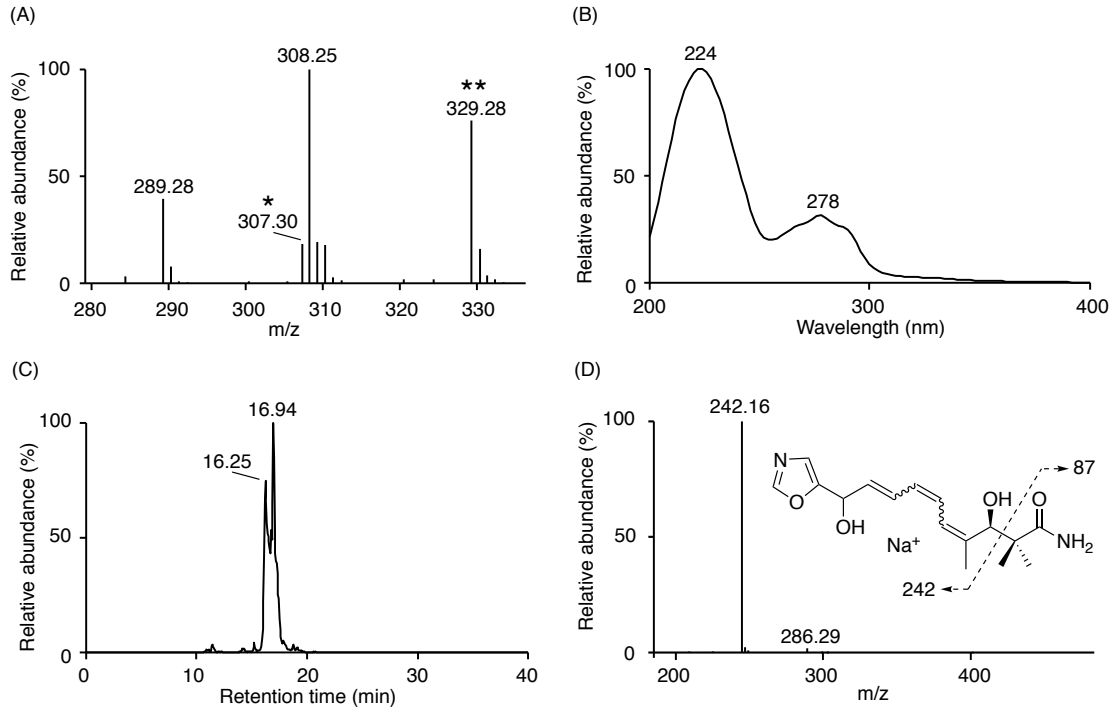


Figure 3.2. Detection of phthoxazolins B-D from *Streptomyces* sp. KO-7888 by LC-MS. (A) MS at 15.8-17.4 min. *, proton adduct ion; **, sodium adduct ion. (B) UV spectrum at 15.8-17.4 min. (C) Selected ion monitoring of *m/z* $[M+Na]^+$ 329.15. (D) MS/MS fragmentation of parent ion *m/z* $[M+Na]^+$ 329.15.

3.2.2 Gene organisation of the putative phthoxazolin clusters and preliminary definition of the cluster boundaries

The putative phthoxazolin clusters from the two *Streptomyces* spp. are nearly identical: they both span a region of 80.7 kbp, and house 21 open reading frames designated *phoxB-phoxV* (Figure 3.3). *PhoxB-phoxV* together encode four *trans*-AT PKSs, three NRPSs, three hybrid *trans*-AT PKS-NRPSs, a 4'-phosphopantetheinyltransferase (PPTase), three regulatory proteins, a membrane protein, a cytochrome P450 enzyme, and several other tailoring enzymes and proteins of presently unknown function. Although the exact boundaries of the gene clusters are yet to be determined experimentally, the gene products encoded by the regions immediately upstream and downstream of the initially chosen boundaries, nitrate respiratory enzyme subunits and cellulases, respectively, are unlikely to be involved. The properties and putative function of the gene products of genes within and flanking the clusters from *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888 are shown in Tables 3.1 and 3.2, respectively.

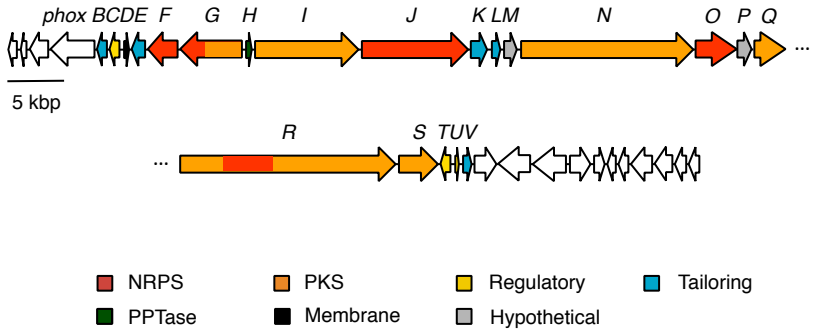


Figure 3.3. Gene organisation of the putative phthoxazolin cluster. Genes flanking the upstream and downstream regions of the cluster are shown in white.

Table 3.1. Properties of genes within and flanking the putative phthoxazolin cluster from *Streptomyces* sp. OM-5714.

Gene	Product size (aa)	% Identity/ similarity	Species	Putative function	GenBank accession no.
orf-4	240	99/100	*Streptomyces coelicoflavus*	Nitrate reductase gamma subunit	WP_007445619.1
orf-3	164	99/99	*Streptomyces* sp. CS159	Nitrate reductase assembly chaperone	WP_087788863.1
orf-2	528	99/100	*Streptomyces coelicoflavus*	Nitrate reductase beta subunit	WP_007445621.1
orf-1	1231	98/99	*Streptomyces diastaticus*	Nitrate reductase alpha subunit	WP_102932277.1
phoxB	222	99/99	*Streptomyces coelicoflavus*	SAM-dependent methyltransferase	WP_007445623.1
phoxC	287	99/99	*Streptomyces coelicoflavus*	XRE family transcriptional regulator	WP_007445625.1
phoxD	148	97/96	*Streptomyces spp.*	Membrane protein	WP_087788858.1
phoxE	386	97/98	*Streptomyces* spp.	Esterase	WP_007445627.1
phoxF	847	96/97	*Streptomyces coelicoflavus*	NRPS	WP_108988779.1
phoxG	1742	96/96	*Streptomyces* sp. CS159	PKS-NRPS	WP_087788856.1
phoxH	161	99/100	*Streptomyces coelicoflavus* ZG0656	4'-phosphopantetheinyl transferase	EHN75053.1
PhoxI	2910	92/93	*Streptomyces*	PKS	WP_108988777.1

			coelicoflavus		
phoxJ	3008	97/98	*Streptomyces* sp. CS159	NRPS	WP_087788854.1
phoxK	560	99/99	*Streptomyces coelicoflavus*	PKS	WP_108988775.1
phoxL	240	98/98	*Streptomyces coelicoflavus*	SDR family NAD(P)-dependent oxidoreductase	WP_108988774.1
phoxM	363	98/98	*Streptomyces* sp. CS159	Hypothetical protein	WP_087788851.1
phoxN	4867	97/97	*Streptomyces* sp. CS159	PKS	WP_087792773.1
phoxO	1158	97/97	*Streptomyces coelicoflavus*	NRPS	WP_108988772.1
phoxP	400	99/99	*Streptomyces* spp.	Hypothetical protein OzmP	WP_087788849.1
phoxQ	879	98/98	*Streptomyces* sp. CS159	PKS	WP_087788848.1
phoxR	6109	97/97	*Streptomyces* sp. CS159	PKS-NRPS	OWA21555.1
phoxS	1106	98/98	*Streptomyces* spp.	Malonyl-CoA-ACP transferase	WP_007388965.1
phoxT	272	98/98	*Streptomyces coelicoflavus*	XRE family transcriptional regulator	WP_007388964.1
phoxU	101	97/98	*Streptomyces* spp.	4a-Hydroxytetrahydrobiopterin dehydratase	WP_007388966.1
phoxV	246	98/98	*Streptomyces* sp. NRRL WC-3753	SAM-dependent methyltransferase	KPC88085.1
orf+1	611	97/99	*Streptomyces* sp. CS159	Rhamnogalacturonan lyase	WP_087788844.1
orf+2	891	92/94	*Streptomyces rubrogriseus*	Xyloglucanase	WP_109028527.1
orf+3	945	98/99	*Streptomyces coelicoflavus*	Cellulose 1,4-beta-cellobiosidase	WP_007388978.1
orf+4	579	98/98	*Streptomyces* spp.	Cellulose 1,4-beta-cellobiosidase	WP_011031002.1
orf+5	302	93/94	*Streptomyces* sp. CS113	SAM-dependent methyltransferase	WP_087805998.1
orf+6	233	96/98	*Streptomyces*	SDR family NAD(P)-	WP_102931503.1

			diastaticus	dependent oxidoreductase	
orf+7	277	99/99	*Streptomyces coelicoflavus*	Aldo/keto reductase	WP_007388979.1
orf+8	608	99/99	*Streptomyces* sp. CS159	Long-chain fatty acid-CoA ligase	WP_087788836.1
orf+9	504	97/98	*Streptomyces* sp. NRRL WC-3753	Multidrug MFS transporter	KPC88094.1
orf+10	327	99/99	*Streptomyces coelicoflavus* ZG0656	Lrp/AsnC family transcriptional regulator	EHN78694.1
orf+11	293	100/100	*Streptomyces* spp.	LysR family transcriptional regulator	WP_007388971.1
orf+12	336	98/98	*Streptomyces* sp. NRRL WC-3753	Bile acid-sodium symporter	KPC88096.1

Table 3.2. Properties of genes within and flanking the putative phthoxazolin cluster from *Streptomyces* sp. KO-7888.

Gene	Product size (aa)	% Identity/ similarity	Species	Putative function	GenBank accession no.
orf-4	240	99/100	*Streptomyces coelicoflavus*	Nitrate reductase gamma subunit	WP_007445619.1
orf-3	164	99/99	*Streptomyces* sp. CS159	Nitrate reductase assembly chaperone	WP_087788863.1
orf-2	528	99/100	*Streptomyces coelicoflavus*	Nitrate reductase beta subunit	WP_007445621.1
orf-1	1231	98/99	*Streptomyces diastaticus*	Nitrate reductase alpha subunit	WP_102932277.1
phoxB	273	98/98	*Streptomyces coelicoflavus*	SAM-dependent methyltransferase	WP_007445623.1
phoxC	287	99/99	*Streptomyces coelicoflavus*	XRE family transcriptional regulator	WP_007445625.1
phoxD	148	96/95	*Streptomyces* spp.	Membrane protein	WP_087788858.1
phoxE	386	97/98	*Streptomyces* spp.	Esterase	WP_007445627.1
phoxF	847	96/97	*Streptomyces* sp. CS159	NRPS	WP_087788857.1
phoxG	1744	95/96	*Streptomyces* sp. CS159	PKS-NRPS	WP_087788856.1
phoxH	161	99/100	*Streptomyces*	4'-phosphopantetheinyl	EHN75053.1

			coelicoflavus ZG0656	transferase	
PhoxI	2921	92/93	*Streptomyces coelicoflavus*	PKS	WP_108988777.1
phoxJ	3008	97/98	*Streptomyces* sp. CS159	NRPS	WP_087788854.1
phoxK	462	99/99	*Streptomyces* sp. NRRL WC-3753	Cytochrome P450	KPC72422.1
phoxL	240	98/98	*Streptomyces coelicoflavus*	SDR family NAD(P)-dependent oxidoreductase	WP_108988774.1
phoxM	363	98/98	*Streptomyces* sp. CS159	Hypothetical protein	WP_087788851.1
phoxN	4866	97/97	*Streptomyces* sp. CS159	PKS	WP_087792773.1
phoxO	1158	97/97	*Streptomyces* sp. NRRL WC-3753	NRPS	KPC71002.1
phoxP	400	99/99	*Streptomyces* spp.	Hypothetical protein OzmP	WP_087788849.1
phoxQ	879	98/98	*Streptomyces* sp. NRRL WC-3753	PKS	KPC71004.1
phoxR	6084	96/97	*Streptomyces* sp. CS159	PKS-NRPS	OWA21555.1
phoxS	1106	98/98	*Streptomyces* spp.	Malonyl-CoA-ACP transferase	WP_007388965.1
phoxT	272	99/98	*Streptomyces coelicoflavus*	XRE family transcriptional regulator	WP_007388964.1
phoxU	101	98/98	*Streptomyces* spp.	4a-Hydroxytetrahydrobiopterin dehydratase	WP_007388966.1
phoxV	246	99/99	*Streptomyces* sp. NRRL WC-3753	SAM-dependent methyltransferase	KPC88085.1
orf+1	611	98/99	*Streptomyces* sp. CS159	Rhamnogalacturonan lyase	WP_087788844.1
orf+2	891	92/94	*Streptomyces rubrogriseus*	Xyloglucanase	WP_109028527.1
orf+3	945	98/99	*Streptomyces coelicoflavus*	Cellulose 1,4-beta-cellobiosidase	WP_007388978.1
orf+4	575	97/98	*Streptomyces* spp.	Cellulose 1,4-beta-cellobiosidase	WP_011031002.1

orf+5	307	92/94	*Streptomyces* sp. CCM_MD2014	SAM-dependent methyltransferase	WP_061441401.1
orf+6	233	96/98	*Streptomyces* sp. M1013	SDR family NAD(P)-dependent oxidoreductase	WP_076977235.1
orf+7	277	98/99	*Streptomyces* sp. CS159	Aldo/keto reductase	WP_087788837.1
orf+8	608	99/99	*Streptomyces* sp. CS159	Long-chain fatty acid-CoA ligase	WP_087788836.1
orf+9	504	97/98	*Streptomyces* sp. NRRL WC-3753	Multidrug MFS transporter	KPC88094.1
orf+10	327	99/99	*Streptomyces coelicoflavus* ZG0656	Lrp/AsnC family transcriptional regulator	EHN78694.1
orf+11	293	100/100	*Streptomyces* spp.	LysR family transcriptional regulator	WP_007388971.1
orf+12	336	98/98	*Streptomyces* sp. NRRL WC-3753	Bile acid-sodium symporter	KPC88096.1

As the two clusters are highly identical, bioinformatics analysis of only the cluster from *Streptomyces* sp. KO-7888 is shown in the following sections of this chapter.

3.2.3 Organisation of PKS and NRPS modules and enzymatic domains

The unexpectedly large putative phthoxazolin cluster encodes a total of 16 PKS/NRPS modules. Fifty-nine PKS/NRPS domains were identified using antiSMASH 4.0 (Figure 3.4) and the *in silico* analysis of individual domains is detailed in Sections 3.2.5 and 3.2.6.

3.2.4 Comparative sequence analysis of the phthoxazolin and oxazolomycin clusters

As phthoxazolins are structurally identical to a portion of the oxazolomycin antibiotics, considerable similarity between the corresponding biosynthetic gene clusters was expected. Indeed, six conserved genes were identified, five of which encode PKS and NRPS enzymes (proposed to be involved in the early steps of oxazolomycin biosynthesis (Zhao *et al.*, 2010)) and one gene encodes a protein of unknown function (Figure 3.5). Not only do OzmN, OzmO, OzmQ and OzmH show a high

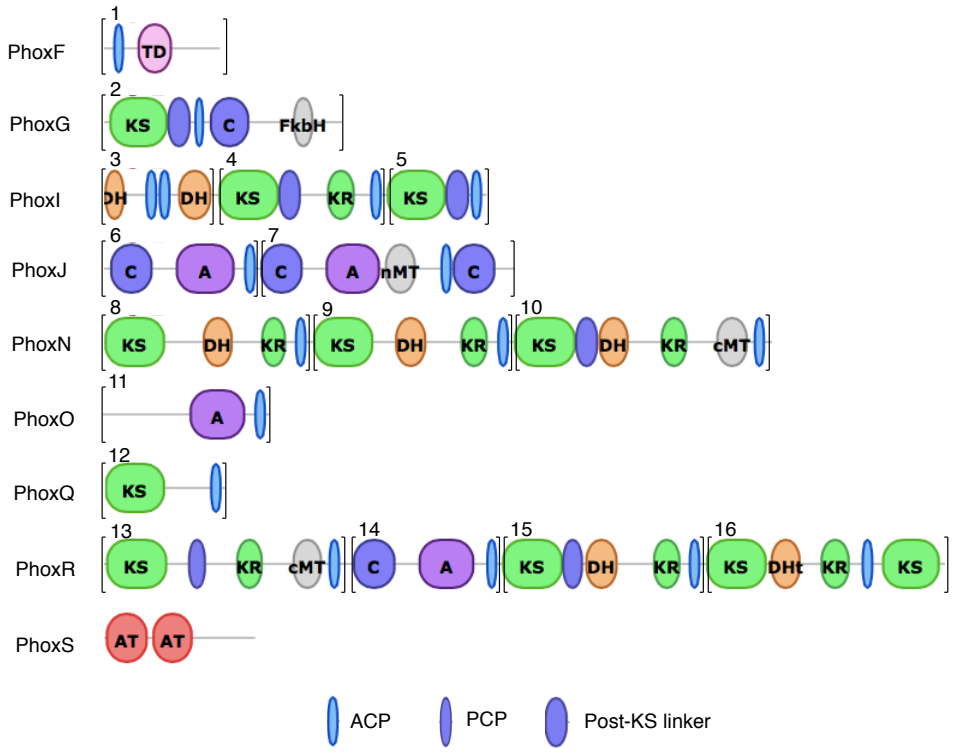


Figure 3.4. Organisation of PKS and NRPS modules and enzymatic domains. Detection and annotation of PKS and NRPS domains were performed using antiSMASH 4.0 (Blin *et al*., 2017). Modules are indicated with brackets and numbered 1 to 16 according to gene order. ACP, acyl carrier protein; PCP, peptidyl carrier protein; AT, acyltransferase; TD, terminal reductase; KS, ketosynthase; C, condensation; FkbH, FkbH-like; DH, dehydratase; KR, ketoreductase; A, adenylation; nMT, *N*-methyltransferase; cMT, *C*-methyltransferase; DHt, truncated dehydratase.

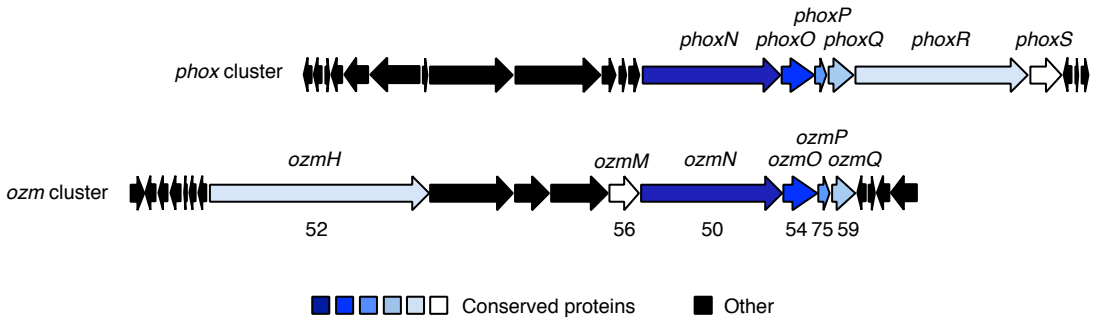


Figure 3.5. Comparison of the gene organisation of the phthoxazolin (*phox*) cluster to that of oxazolomycin (*ozm*). Sequence identities of conserved proteins are indicated in percentages.

amino acid sequence identity to PhoxN, PhoxO, PhoxQ and PhoxR, respectively, their PKS/NRPS domain organisation is also comparable. Noticeably, the gene organisation differs between the two clusters, and the putative phthoxazolin cluster consists of many genes that are not present in the oxazolomycin cluster. The role of these genes in the biosynthetic pathway is unclear.

3.2.5 *In silico* analysis of PKS domains

Trans-AT domains

PhoxS encodes a discrete AT-AT-Oxidoreductase (Ox) tri-domain (Figure 3.6). Both *trans*-AT domains contain the highly conserved GQGSQ loop, and the presence of the GHSxG and APFH/YAYH motifs signifies that the domains are specific for malonyl-CoA rather than methylmalonyl-CoA extender units (Figure 3.7). In the oxazolomycin pathway, Zhao *et al*. (2010) demonstrated that the second AT domain of OzmM (AT-AT-Oxidoreductase) is solely responsible for the loading to all PKS modules, via in-frame deletion experiments. Based on sequence homology, the second AT domain of PhoxS presumably plays a similar role in phthoxazolin

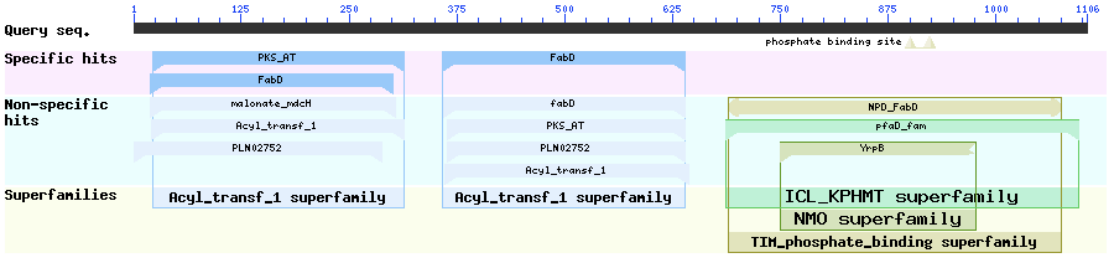


Figure 3.6. Protein-BLAST analysis of PhoxS. Putative conserved domains are shown. Note that the C-terminal oxidoreductase domain was detected using protein-BLAST but not antiSMASH 4.0. Blue, AT domain; green, oxidoreductase domain.

biosynthesis. *Trans*-AT domains with an oxidoreductase domain fused at the C-terminal end are occasionally observed, but their function is unknown (Piel *et al*., 2002; El-Sayed *et al*., 2003; Zhao *et al*., 2010).

```
PhoxS-AT2       -FPGQGSQQRGMGKELFTRYPRET-------AIADRVLGYSIEELCVHDPER----RLGR
OzmM-AT2        -FPGQGSQAKGMGKDLFEEFPEET-------ALADSVLGHSIRELCVEDPRR----ELKL
PhoxS-AT1       MFSGQGSQYYRMGQELHESDEVFRTALLRYDAVVAGLLGESVLARIL-DPDRRKNEPFVD
OzmM-AT1        ----------MGSALYETEPVFRRVMDRLDAAASGELGESVLAALY-APGRGRAEPFDD

PhoxS-AT2       TEYTQPALYVVSALTHLDRLAEDAEPADYLIGHSLGEYAALFAAGVFDFETGLRLVQRRG
OzmM-AT2        TRFTQPALYVVSALAWLRERRENPVPPDFLVGHSLGEYVALFAAGSFDFETGLRLVARRG
PhoxS-AT1       TRLTHPAIVMI-ELALAETLRAKGIEPDYLLGSSLGEYAAAVVSGSIDAETCLRLLVRQA
OzmM-AT1        IAFTHPAIVMV-ELAAAETLIASGIRPDLLLGASLGEFTASVLAGVLDADACLRILVRQA

PhoxS-AT2       ALMAAAGGGGMAAVVGCDEATVLRVLADSGIDELDLANYNAPDQFVLSGPAERVDAARDA
OzmM-AT2        ELMSRADGGRMAAVLKCDLETVEGALAGHGLTGLDIANHNAPGQFVIAGPTEQITAAKPV
```

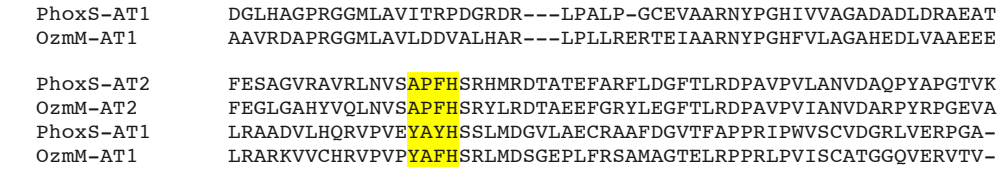


Figure 3.7. Partial sequence alignment of AT domains. Conserved motifs discussed in the text are highlighted in yellow.

ACP domains

The highly conserved signature motif GxDS(L/I) of ACP is present in PhoxN-ACP8, ACP10, PhoxQ-ACP12, PhoxR-ACP13, ACP15 and ACP16, and is modified by one residue in PhoxG-ACP2, PhoxI-ACP3a, ACP3b, ACP4, ACP5 and PhoxN-ACP9 (GxDSV, GxGSA, GxDSV and GxNSL, respectively). As the Ser residue essential for 4'-PP attachment was identified in all ACP domains, they are all potentially functional (Figure 3.8).

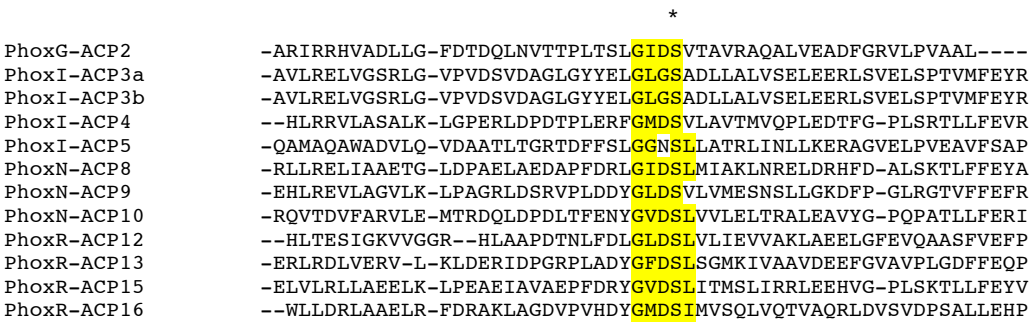


Figure 3.8. Partial sequence alignment of ACP domains. The highly conserved motif GxDS(L/I) is highlighted in yellow. The Ser residue essential for 4'-PP attachment is marked with an asterisk.

KS domains

The Cys-His-His triad essential for the catalytic activity of KS was observed in 8 out of 11 KS domains. The His residue in the highly conserved HGTGT motif essential for catalysing decarboxylative condensation is absent in PhoxI-KS5 (Cyc-Ala-His) and PhoxR-KS16b (Cys-Asn-His), suggesting they may be inactive and that modules containing these KS domains are not involved in chain elongation.

As a clade assignment system is available for substrate prediction of *trans*-AT KSs (Nguyen *et al*., 2008; Gay *et al*., 2014), all 11 KS domains in the gene cluster were aligned against 211 other KS domains from 16 previously characterised *trans*-

AT PKS pathways[1] (Figure 3.9). PhoxI-KS4, PhoxN-KS9, PhoxN-KS10 and PhoxR-KS16a are predicted to accept intermediates containing an αβ-double bond (clade IX). They are strong candidates for accepting intermediates containing the first two double bonds of the triene moiety. PhoxR-KS13 specifies for substrates with a double bond and an α-methyl group (clade VII), presumably accepting intermediates after the third double bond of the triene moiety is generated. In fact, PhoxN-KS9, PhoxN-KS10 and PhoxR-KS13 are highly identical to the KS domains belonging to modules 4, 5 and 6 of the oxazolomycin pathway, whereas PhoxQ-KS12 and PhoxN-KS8, corresponding to the KS domains in modules 2 and 3 of the oxazolomycin pathway, are predicted to accept substrates from an NRPS module (clade XVI).

KR domains

All eight KR domains show the N-terminal Rossmann fold required for the binding of the cofactor NADPH. The Lys-Ser-Tyr-Asn tetrad located in the active site was observed in the majority of KRs except for PhoxN-KR10 (Ser-Tyr-Cys) and PhoxR-KR13 (Ser-Tyr-Ser). It is unclear whether or not these two type-C2 domains are functional. PhoxN-KR8, KR9, KR10 show high sequence homology with OzmN-KR3, KR4 and KR5, respectively, which are found in the triene-producing modules 3-5 of the partly analogous oxazolomycin pathway. Further analysis of KR domains was carried out to predict the stereochemistry of the potential β-hydroxyl group produced (Keatinge-Clay, 2007; discussed in Section 1.4.4) (Figure 3.10). PhoxI-KR4, PhoxN-KR8 and PhoxR-KR15 contain a modified LDD motif (IED or LRD) at fingerprint position 1 and no Pro residue at fingerprint position 5, and thus, were predicted to be type B1 KRs giving rise to D-β-hydroxyl groups. This is consistent with the general rule that DH domains typically operate on products of type B instead of type A KRs. The stereochemistry of the putative products of PhoxN-KR9 and PhoxR-KR16 is ambiguous from inspection of the residues, as neither the LDD motif (characteristic of type B KRs) nor a Trp residue at fingerprint position 2 (characteristic of type A KRs) was observed.

DH domains

The highly conserved N-terminal HxxxGxxxxP motif was observed in all DH domains except in the truncated PhoxR-DHt16. The downstream YGP motif and a modified

[1] Bacillaene, bryostatin, chivosazol, difficidin, disorazol, lankacidin, leinamycin, macrolactin, mupirocin, myxovirescin, onnamide, pederin, rhizoxin, thailandamide and virginiamycin pathways

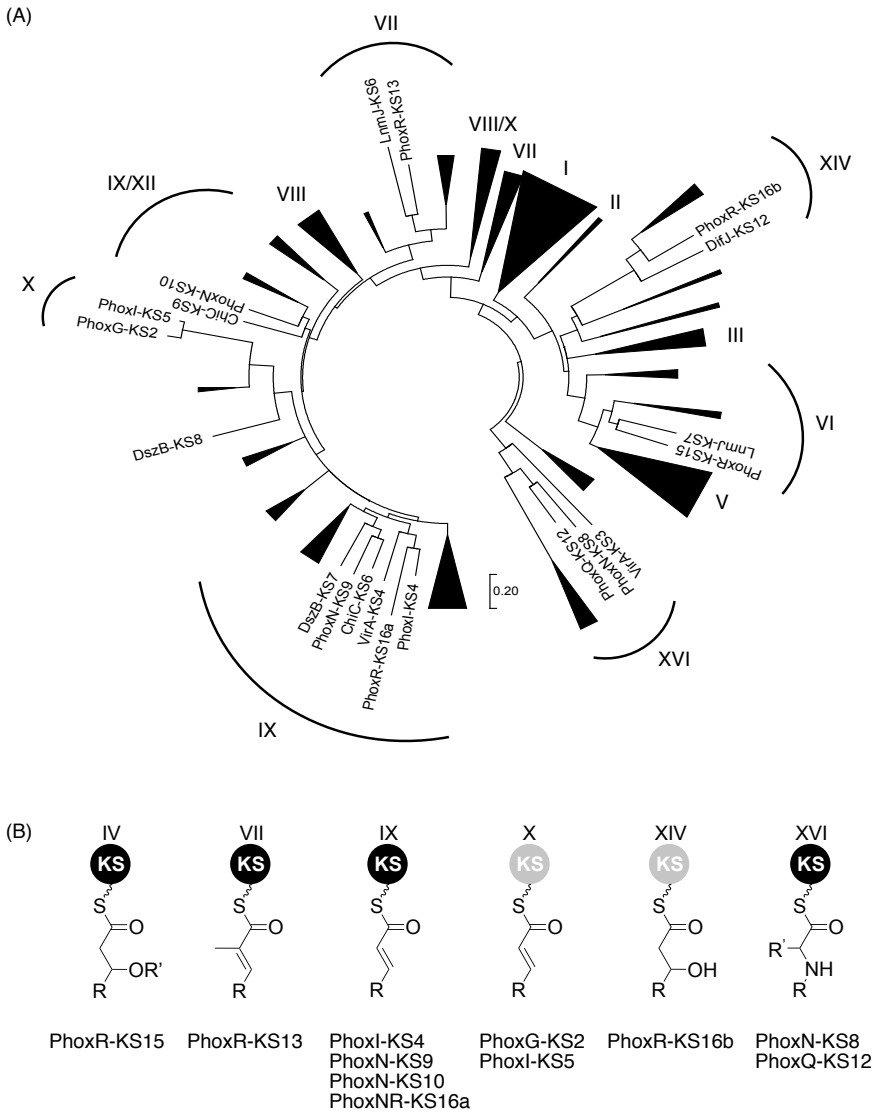


Figure 3.9. Clade assignment of KS domains for substrate specificity prediction. (A) Phylogenetic tree of *trans*-AT PKS KS domains. Evolutionary history was inferred using the maximum likelihood method (Jones *et al*., 1992). All positions containing gaps were eliminated and there were 97 amino acids in the final dataset. Sequences were aligned with MUSCLE (Edgar, 2004) and the evolutionary analysis was conducted in MEGA7 (Kumar *et al*., 2016). Clade assignments are taken from Nguyen *et al*. (2008). (B) Predicted substrate structures of the KS domains. Domains lacking the catalytic His residue essential for decarboxylative condensation are coloured in grey.

HPxxLDxxxQ motif were found in PhoxN-DH8, DH9 and DH10 (NPxxLDxxxH, PPxxLDxxxR or HPxxLDxxxH). These three DHs situated in consecutive modules of

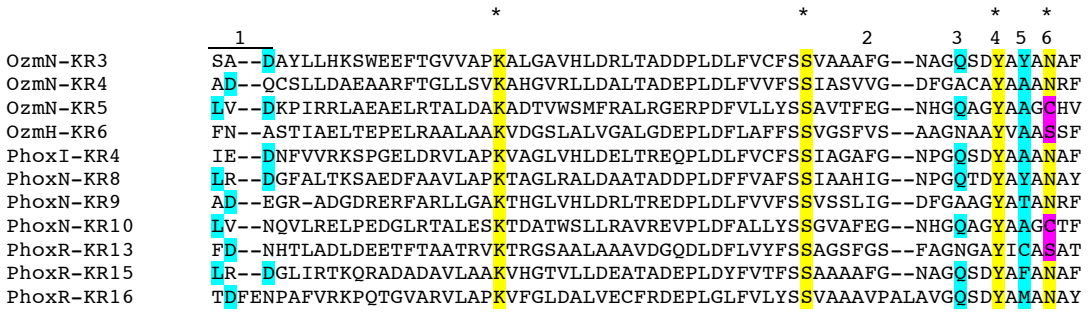


Figure 3.10. Partial sequence alignment of KR domains. Fingerprints 1 to 6 for the prediction of β-hydroxy stereochemistry are indicated (Keatinge-Clay, 2007). The Lys-Ser-Tyr-Asn tetrad in the active site is marked with asterisks and highlighted in yellow. Fingerprints characteristic of type B1 and type C2 KRs are highlighted in cyan and magenta, respectively.

PhoxN are presumably responsible for catalysing the formation of the three double bonds, giving rise to the triene moiety in phthoxazolins. The modified motifs may play a role in the formation of the third double bond in the non-canonical *cis*-conformation, as well as the second double bond in either conformation to afford a mixture of the geometrically isomeric phthoxazolins. Alternatively, a presently unknown domain exhibiting epimerase activity may be responsible. The DH3a-ACP3a-ACP3b-DH3b module in PhoxI shows unusual domain architecture and contains two additional DHs. The second and third motifs are poorly conserved in PhoxI-DH3a (LGG, HPxxFExxxL), suggesting inactivity; whereas PhoxI-DH3b may potentially be able to offer *trans*-DH activity as both motifs are present with all three catalytically important residues, Tyr, Asp and Gln intact (Figure 3.11).

MT domains

Two *C*-MT domains were found in the third module of the PKS PhoxN and the first module of the hybrid PKS-NRPS PhoxR, respectively. Both domains contain the His-Glu catalytic dyad, the highly conserved Tyr and Asn residues (Skiba *et al*., 2016), as well as the *S*-adenosyl-L-methionine (SAM) cofactor binding motifs (V/I/L)(L/V)(D/E)(V/I)G(G/C)G(T/P)G and LL(R/K)PGG(R/I/L)(L/I)(L/F/I/V)(I/L) (Kagan and Clarke, 1994; Miller *et al*., 2003) (Figure 3.12), and are potentially capable of methylating at positions C-2 and C-4 (geminal di-methyl). There is also an *N*-MT domain in the NRPS PhoxJ. The role of this domain is unclear as *N*-methylation is not thought to be required for phthoxazolin biosynthesis.

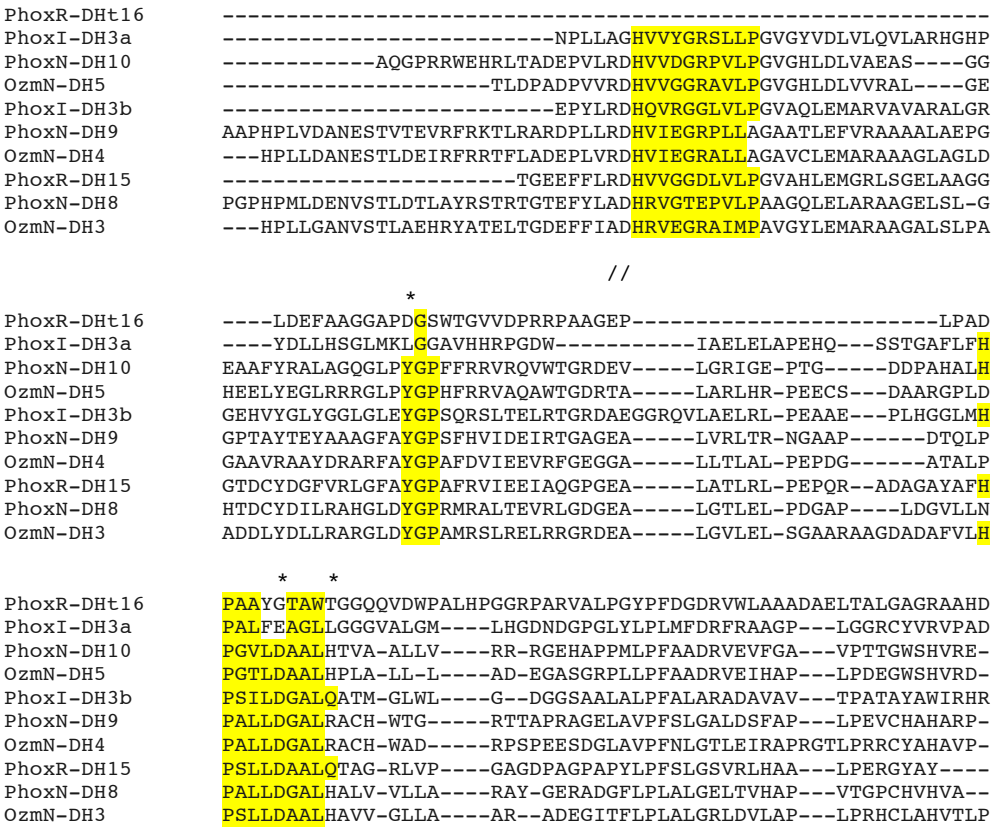


Figure 3.11. Partial sequence alignment of DH domains. Conserved motifs discussed in the text are highlighted in yellow. Catalytic residues are marked with an asterisk. DH domain sequences from the oxazolomycin (*ozm*) pathway were included for comparison.

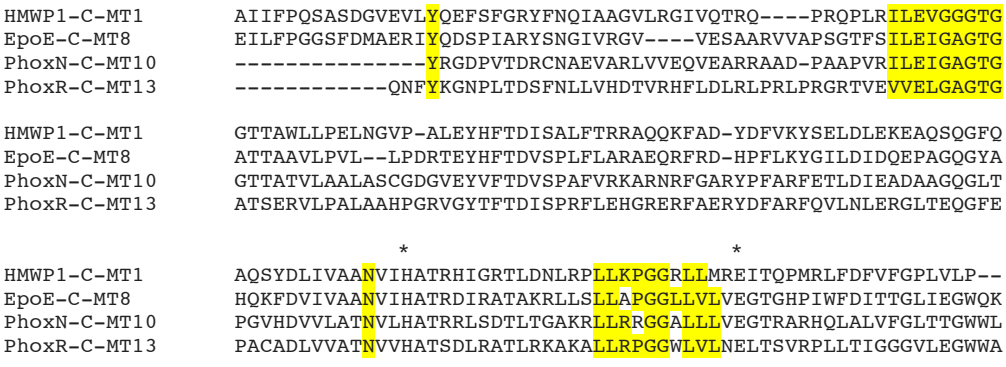


Figure 3.12. Partial sequence alignment of *C*-MT domains. The conserved residues and SAM cofactor binding motifs discussed in the text are highlighted in yellow. The His-Glu catalytic dyad is marked with asterisks. Sequences of *C*-MT domains involved in geminal di-methyl formation from the yersiniabactin (HMWP) and epothilone (*epo*) pathways were included for comparison.

3.2.6 *In silico* analysis of NRPS domains

A domains

All four A domains contain the two invariant Asp and Lys residues for the correct positioning of amino acids. The substrate specificity of individual domains was predicted based on the eight variable residues present in the substrate binding pocket (Figure 3.13 and Table 3.3). The key residues displayed by Ser-specific A domains are found in PhoxJ-A6, including the highly conserved His-Ser dyad for hydrogen bonding with the hydroxyl side chain of the substrate (Challis *et al*., 2000); whereas PhoxO-A11 and PhoxR-A14 were predicted to specify for Gly. The residues in the binding pocket of PhoxJ-A7 do not show significant similarity to those identified in characterised A domains, so its substrate specificity could not be predicted.

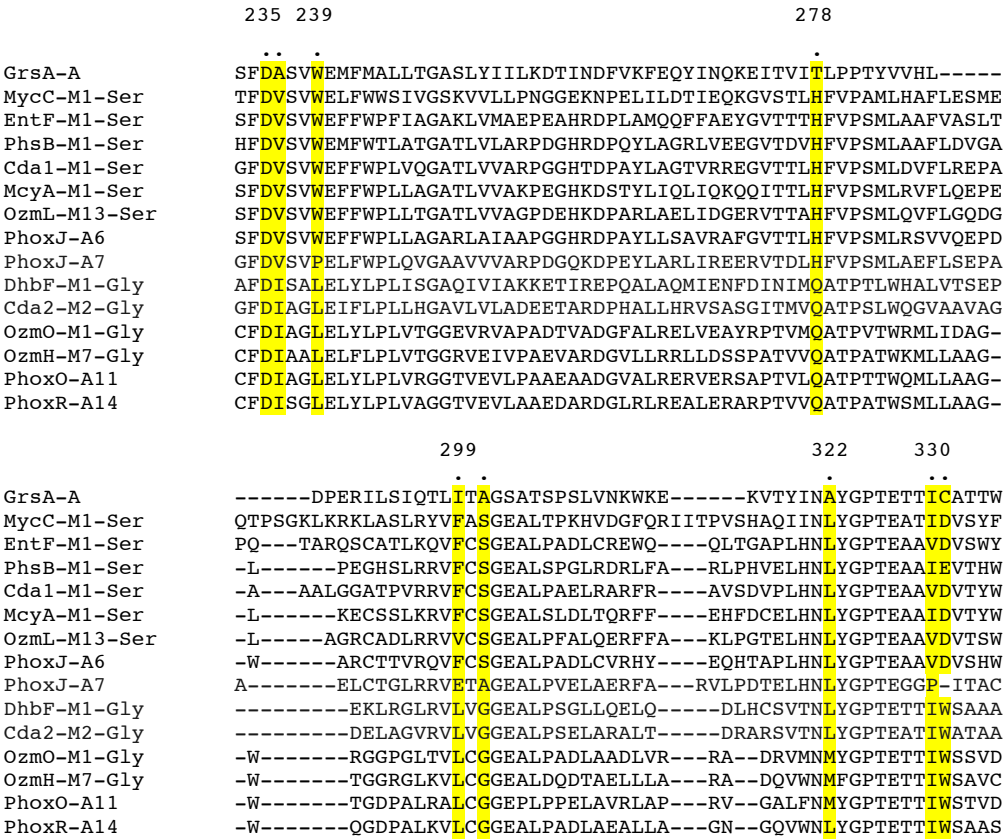


Figure 3.13. Partial sequence alignment of A domains. Sequences of characterised A domains from the mycosubtilin (*myc*), enterobactin (*ent*), phosphinothricin (*phs*), calcium-dependent antibiotic (*cda*), microcystin (*mcy*), (2,3-dihydroxybenzoate (*dhb*) and oxazolomycin (*ozm*) pathways were included for comparison. Residues in substrate binding pocket are highlighted in yellow. Residue numbering refers to the A domain of GrsA from the gramicidin pathway.

Table 3.3. Prediction of substrate specificity of A domains. Residue numbering refers to the A domain of GrsA from the gramicidin pathway.

A domain	Amino acid residues in substrate binding pocket										Predicted
	235	236	239	278	299	301	322	330	331	517	specificity
PhoxJ-A6	D	V	W	H	F	S	L	V	D	K	Ser
PhoxJ-A7	D	V	P	H	E	A	L	P	N/A	K	N/A
PhoxO-A11	D	I	L	Q	L	G	M	I	W	K	Gly
PhoxR-A14	D	I	L	Q	L	G	L	I	W	K	Gly

PCP domains

The Ser residue essential for 4'-PP attachment is present in all five PCP domains (Schlumbohm *et al.*, 1991) (Figure 3.14). The highly conserved signature motif LGGxS of PCP was identified in PhoxJ-PCP6 and PCP7, and is modified by one to three residues in PhoxF-PCP1, PhoxN-PCP11 and PhoxR-PCP14 (AGAxC, AGGxS, AGMxS).

```
                                          *
PhoxF-PCP1        ---IVSKAAALLDVK--EIDTETDLFDAGATSVDAVRLVAVLDRELNVRLSLDDVFADAR
PhoxJ-PCP6        ERHLAEVCAELLGFDADRVSAEDNFFALGGHSLMITVLVARLKD-SGLHITVQDVFTATT
PhoxJ-PCP7        -RTLCRIVADVLGLT--RVGLQDNFFDLGGHSLLATRLTLRIRKETGAELPLQLIFSGAT
PhoxN-PCP11       ERRIAQVWCEVLGLE--RVGAEDNFFDAGGDSLRLTSVVATLRERLGLQVTRLDMFGRPT
PhoxR-PCP14       ------MVAALVDARPEDIGAHTPLGEAGMNSVGFTALSAELRKAYGITEYPTLFYRRGT
```

Figure 3.14. Partial sequence alignment of PCP domains. The signature motif discussed in the text is highlighted in yellow. The Ser residue for 4'-PP attachment is marked with an asterisk.

C domains

PhoxJ-C6 and PhoxR-C14 both display the characteristic active site HHxxxDG motif, whereas a modified version is identified in PhoxG-C2 (HHxxxDL) and PhoxJ-C7a (HHxxxDA). PhoxJ-C7b is unlikely to be functional as the motif is poorly conserved (PWxxxDG), with both essential His residues missing (Figure 3.15). PhoxG-C2 is an additional C domain observed at the C-terminus of the hybrid PKS-NRPS module. Although the function of this domain is presently unclear, some NRPS systems are known or proposed to release the full-length peptide using a C domain via intermolecular nucleophilic attack of the thioester bond (Griffiths *et al.*, 1984; Weber *et al.*, 1994).

```
PhxoG-C2        RFELLEYPSG-GQVLLYGAHHAVSDLSSLALVAAEIGAELSG----EQLSSAPSLDDIAG
PhoxJ-C6        RVRLLRRSEQ-EHHLLVTTHHSVSDGWSVGVFFRDLVALYEAFGEGRPDPLEPLPVQYAD
PhoxJ-C7a       RLVVAEEPHSRRRYLLLTAHHLIEDATSLRLILAEL----AAHMAGRAD-LLAAPAPYRD
PhoxJ-C7b       RAHLLTVSDD-RHIGVLTRPWGVFDGWSVNVVLADLLELYRAFSRGRTAELPELPLSYAG
PhoxR-C14       RATLYSLGDG-RNALLLTFHHLVFDGVSIALLLRELESGYRALLEGRALSTERPARGYAD
```

Figure 3.15. Partial sequence alignment of C domains. The active site motif discussed in the text is highlighted in yellow.

Formylation (F) domain

In addition to the A domain and PCP domain detected using antiSMASH 4.0, protein-BLAST analysis revealed the presence of an N-terminal formylation (F) domain in the NRPS PhoxO (Figure 3.16). PhoxO-F11 contains the Asn-His-Asp catalytic triad, and the HxSLLPxxxG loop for N^{10}-formyltetrahydrofolate (N^{10}-fH_4F) cofactor binding, which is known to be required for the activity of formyltransferases involved in ribosomal peptide biosynthesis (Schmitt *et al.,* 1996; Gatzeva-Topalova *et al*., 2005; Williams *et al*., 2005), is modified by two residues (HxGPLPxxxG) (Figure 3.17).

Similar to the F domain-containing NRPS module (LgrA) involved in the initiation of linear gramicidin biosynthesis, PhoxO may possibly constitute the loading module of the phthoxazolin pathway, too. Notably, PhoxO and its homologues, OzmO and CongA, from the pathways governing oxazolomycin and another 5-oxazole-containing polyketide-peptide, conglobatin, show the same domain organization and predicted amino acid specificity. The significance of this finding is further discussed in Chapter 4.

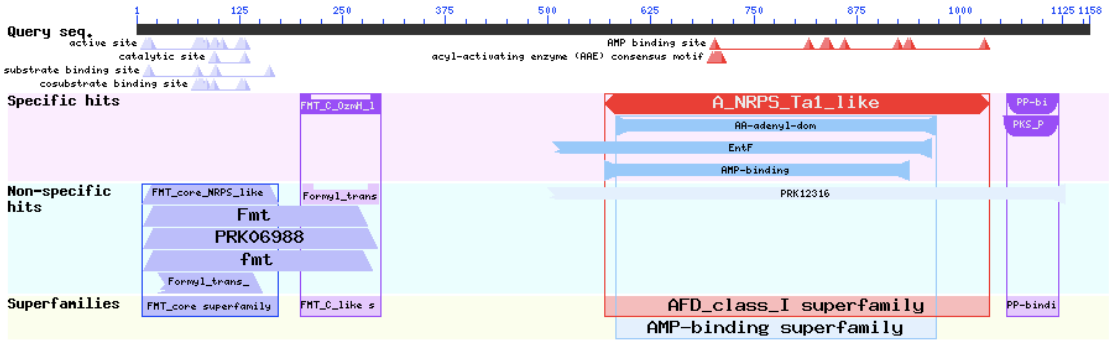


Figure 3.16. Protein-BLAST analysis of PhoxO. Putative conserved domains are shown. Blue, F domain; red, A domain; purple, PCP domain.

```
                              * *                                    *
ApdA-F          YVLPQEILELPRQFAINYHDAPLPRYAGVNATSWALMNQEKTHGVTWHIMAAMVDAGDIL
CongA-F         RMLPAALLDLPRELPVNFHDGPLPRHAGLNATTWAVLEREVSHGVTWHVMTEGADEGDVL
OzmO-F          RMLPDDVLALPERMPVNFHDALLPRHAGLHATSWAVLEGAAEHGVTWHVMEREADTGDVL
PhoxO-F11       RMLPEEVLALPTRLPVNFHDGPLPRHAGLFATSWAILDGDKQHGVTWHVMHLEADAGDVL
LgrA-F          YILDKEIVSRFRGRIINLHPSLLPWNKGRDPVFWSVWD-ETPKGVTIHLIDEHVDTGDIL
```

Figure 3.17. Partial sequence alignment of NRPS F domains. The conserved motif for N^{10}-fH_4F cofactor binding is highlighted in yellow. Catalytic residues are marked with asterisks.

Terminal reductase (TD) domain

PhoxF comprise a PCP domain and a rare terminal reductase (TD) domain. TD domains belong to the short-chain dehydrogenase/reductase (SDR) superfamily and in several NRPS systems, a TD domain integrated at the C-terminus of the last module catalyses the reductive release of the full-length peptide, yielding an aldehyde product/intermediate (Silakowski *et al*., 2000; Kessler *et al*., 2004; Read and Walsh, 2007). PhoxF-TD shows approximately 30% sequence identity to known TD domains from the myxochelin, linear gramicidin and lyngbyatoxin pathways. In PhoxF-TD, both the Asn-Ser-Tyr-Lys catalytic tetrad and the Gly-rich NAD(P)-binding motif (S/T)GxxGxxG required for activity are present (Figure 3.18). This domain could potentially be involved in chain release in the phthoxazolin pathway. There is no conventional C-terminal thioesterase (TE) domain in the phthoxazolin cluster that might mediate chain release. Investigation of the mode of chain release and primary amide formation is discussed further in Chapter 5.

```
PhoxF-TD        ALDDHAEDLSLIMADLAAADRLPWCGDPE----PVPPRRVLLTGATGFLGGHMLLDLLRH
MxcG-TD         MLADAELPEEIVPRLPTPGAEAPLAPSP---GPAAPLRQVLLTGATGFVGAHLLDQLLRQ
LgrD-TD         ------LKDEVVL-DPAIQAEHPYVGDP------SQFQAALLTGATGFLGAFLLRDLLQM
LtxA-TD         ------LAAEVVL-NPQIA---PYQSRPVELDRNTHPASVLLTGATGFLGAYLLYELLKQ

                                        //

                                        *                       *
PhoxF-TD        LLANDVDSVVSVAAAVDFLRGYPSLRRTNVLGVLSLAELAMTGRPKPLHHISSIAVFNEI
MxcG-TD         GLAAECDMILHNAAVVSVVREYGSLQATNVRGTRELLRLAASVRPKPLHYVSTLAVAPQA
LgrD-TD         ELAEKVDVLYHNGALVNFVYPYAALKKANVKGTEEILRLAVAKKTKPVHFVSTIFTFASE
LtxA-TD         KLTELIDAIYHNGAQVSAIEPYTYLKPTNVLGTSELLDFAARCRVKPLHFVSTAAVAVSS

                                      *   *
PhoxF-TD        GIAS---MGEDDPVAHVDRLTAGYDKSKWAAEAALRRARDRGLKATFLRPGGIVGHTRTG
MxcG-TD         NLSPEVP---EAFVPAHPGLRDGYQQSKWAAERLVEQASERGLPVTVYRLGRVSGALDSG
LgrD-TD         ETEESMAFREEDMPENSRVLTSGYTQSKWVAEHLVNLARERGIPAAIYRCGRMTGDSETG
LtxA-TD         KGNPDIIYENFR-LGADSVLPSGYVSSKWVAEELVWVASDRGLPVTVHRPGRISGDTTTG
```

Figure 3.18. Partial sequence alignment of NRPS TD domains. Mxc, myxochelin pathway from *Stigmatella aurantiaca* Sg a15; Lgr, linear gramicidin pathway from *Bacillus brevis*; Ltx, lyngbyatoxin pathway from *Lyngbya majuscula*. The conserved

motif discussed in the text is highlighted in yellow. Catalytic residues are marked with asterisks.

3.2.7 Investigation of the role of various candidate biosynthetic, tailoring and regulatory genes

Development and optimisation of genetic manipulation protocol for *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888

As a prerequisite for genetic manipulation, the antibiotic sensitivity profiles of the phthoxazolin producers were established in this work. Both strains were resistant to ampicillin, carbenicillin and thiostrepton at 100 μg/mL, the highest final concentration tested; and they were both sensitive to apramycin, kanamycin and streptomycin at 25 μg/mL[2]. Apramycin was used as a selection marker in all DNA constructs described in this chapter and in Chapters 4 and 5.

As the genetic manipulation of the strains had not been previously reported, initial attempts of conjugation were performed according to protocols established for well-characterised *Streptomyces* spp., such as *S. coelicolor* and *S. lividans*. However, they were unsuccessful as the strains grew and sporulated too rapidly, which led to a large number of false positive exconjugants being generated. The conjugation efficiency was improved by flooding the agar plate for selection at an earlier time point with a larger volume of antibiotic solution. As *Streptomyces* sp. KO-7888 appeared to be the more amenable target for conjugation, all mutants described in this work were produced from this strain.

In-frame deletion of candidate biosynthetic and tailoring genes

As the putative phthoxazolin cluster is unusually large for the production of such a relatively simple molecule, in-frame gene deletions were carried out to establish the role of individual genes in the pathway, and to define the boundaries of the gene cluster. DNA constructs for in-frame deletion of various biosynthetic and tailoring genes were prepared by cloning arms of homology flanking the targeted region into pYH7, a shuttle vector suitable for maintenance in both *Escherichia coli* and *Streptomyces* spp. The in-frame deletion constructs were introduced into the recipient *Streptomyces* sp. KO-7888 by conjugal DNA transfer from the methylation-deficient *E. coli* ET12567, which bears a helper plasmid. The resulting exconjugants were subjected to alternate rounds of plating on agar with and without the selection

[2] Agar plates showing no sign of growth were incubated for 21 days to confirm sensitivity.

antibiotic, to promote chromosomal integration and the subsequent loss of construct backbone by homologous recombination, and obtain the desired deletion. In-frame deletion mutants were confirmed by PCR (and DNA sequencing of the targeted regions), and were subjected to fermentation to check for the production of phthoxazolins.

From these mutagenesis attempts, in-frame mutants Δ*phoxG* (PKS-NRPS), Δ*phoxI* (PKS), Δ*phoxJ* (NRPS), Δ*phoxR* (PKS-NRPS), Δ*phoxU* (discrete dehydratase), Δ*phoxV* and Δ*orf+5* (SAM-dependent methyltransferases) were successfully generated. The production of phthoxazolins was not affected in Δ*phoxG*, Δ*phoxI*, Δ*phoxJ,* Δ*phoxU*, Δ*phoxV* and Δ*orf+5*, indicating that they are not involved in the pathway. On the other hand, the deletion of *phoxR* abolished production (Figure 3.19). This, for the first time, experimentally linked phthoxazolins to the cluster initially identified by genome mining.

Point mutagenesis of carrier protein domains in *phoxR*

PhoxR, which was shown to be essential for phthoxazolin biosynthesis, comprises three PKS modules and an NRPS module (Figure 3.20). The first PKS module ("module 13") in PhoxR was predicted by bioinformatic analysis to incorporate the fifth malonate extender unit required for phthoxazolin biosynthesis, whereas the role of the remaining modules ("modules 14, 15 and 16") are unclear. To investigate the biosynthetic role of individual modules in PhoxR, the attachment site of the 4'-PP "swinging arm" was precisely removed on each of the four ACP/PCP domains in turn, by carrying out point mutagenesis of the Ser residue to Ala.

As expected, inactivation of ACP13 abolished phthoxazolin production. Surprisingly, ΔPCP14-S2778A, ΔACP15-S4259A and ΔACP16-S5512A also produced little or no phthoxazolin (Figure 3.21). Traces of phthoxazolins could be seen in some but not all replicates of fermentation cultures. These results suggest that all four modules in PhoxR are involved in the biosynthetic pathway, potentially extending the polyketide-peptide chain by an additional Gly (incorporated by "module 14") and two additional malonate units (incorporated by "modules 15 and 16") before the phthoxazolin portion of the chain is released via a presently unknown mechanism.

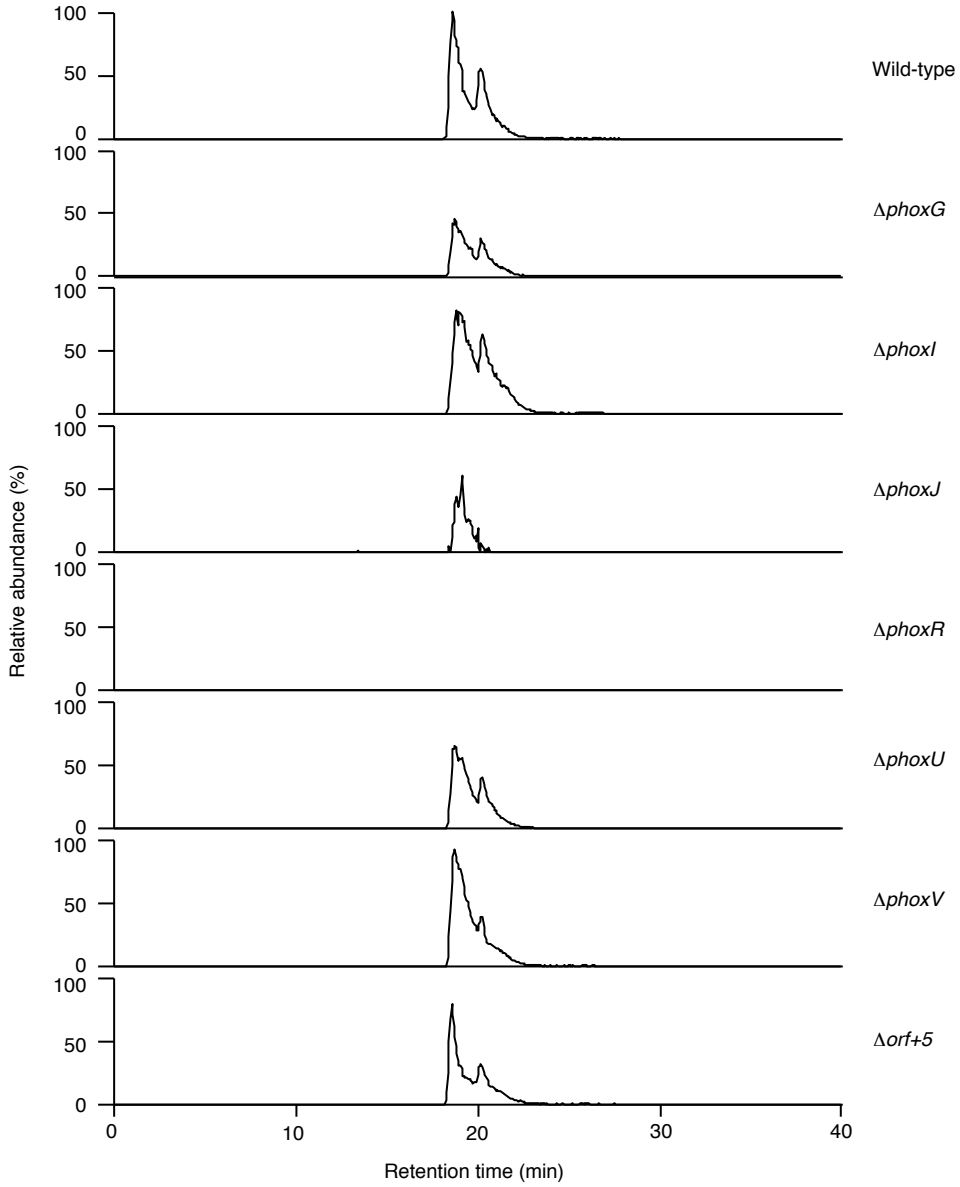


Figure 3.19. Phthoxazolin A production in *Streptomyces* sp. KO-7888 wild-type and in-frame deletion mutants. Selected ion monitoring of *m/z* $[M+Na]^+$ 313.15 is shown.

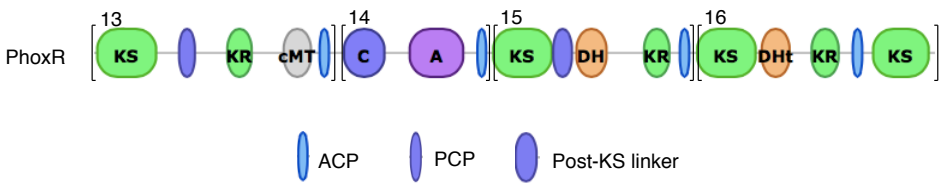


Figure 3.20. Module and domain organisation of the hybrid PKS-NRPS PhoxR. Modules are indicated with brackets and numbered according to order of genes in the phthoxazolin cluster (Section 3.2.3). ACP, acyl carrier protein; PCP, peptidyl

carrier protein; AT, acyltransferase; KS, ketosynthase; C, condensation; DH, dehydratase; KR, ketoreductase; A, adenylation; cMT, *C*-methyltransferase; DHt, truncated dehydratase.

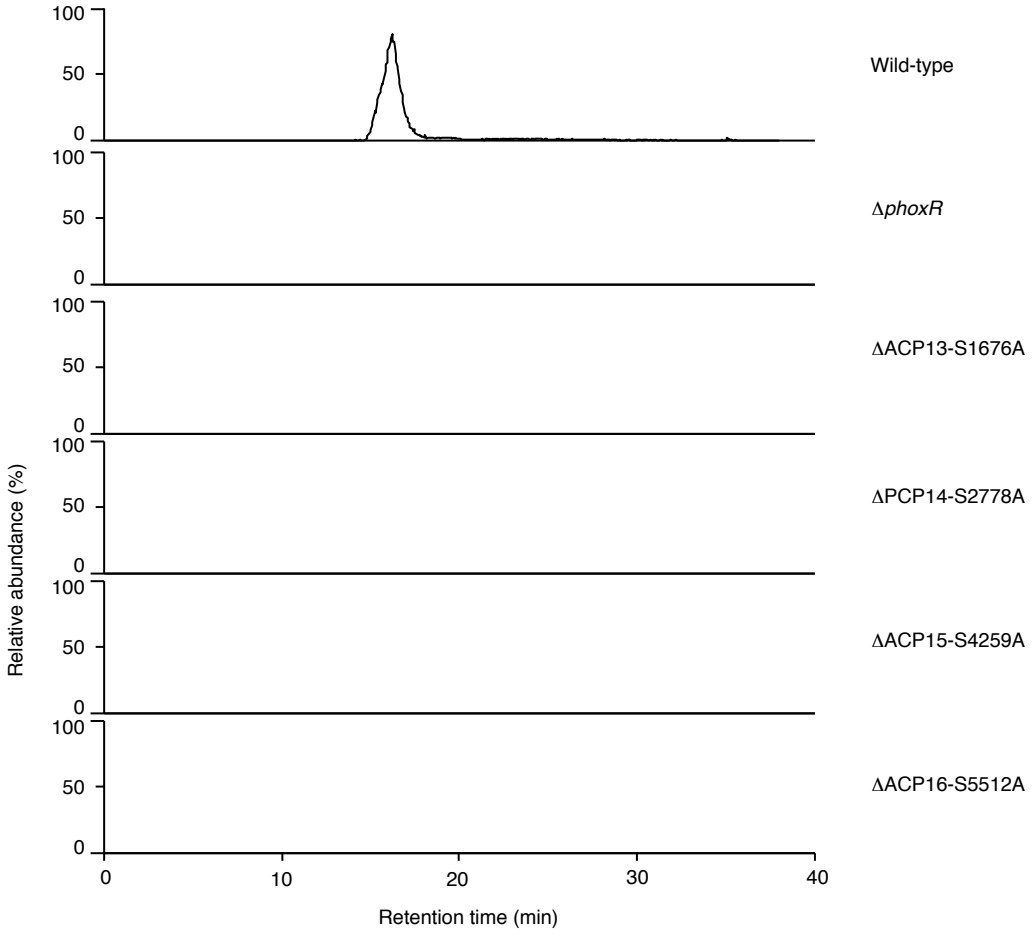


Figure 3.21. Phthoxazolin A production in *Streptomyces* sp. KO-7888 wild-type, *phoxR* in-frame deletion and *phoxR* carrier protein point mutants. Selected ion monitoring of *m/z* $[M+Na]^+$ 313.15 is shown.

In silico analysis and overexpression of putative regulatory genes

PhoxC and PhoxT are putative DNA binding proteins belonging to the xenobiotic response element (XRE) family of transcriptional regulators. Based on sequence homology with other XRE family-like regulators, both proteins were predicted to contain an N-terminal helix-turn-helix domain which presents residues for specific and non-specific DNA interactions as well as a conserved salt bridge to stabilise the helix-turn-helix structure (Beamer and Pabo, 1992; Padmanabhan *et al*., 1997; Luscombe *et al*., 2000) (Figure 3.22). Two additional genes, *orf+11* and *orf+12,* were identified downstream of the phthoxazolin cluster that potentially encode regulatory

proteins belonging to the Lrp/AsnC (leucine-responsive) and LysR families, respectively.

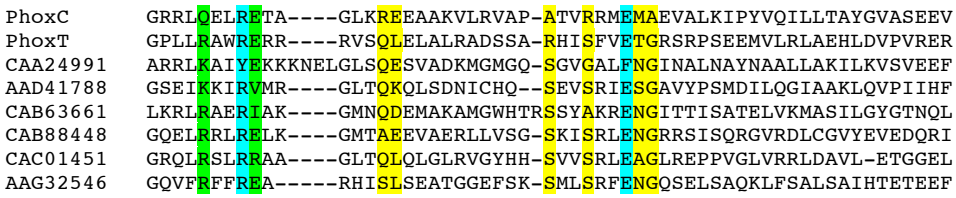


Figure 3.22. Partial sequence alignment of XRE-family like regulators. Residues involved in specific (yellow) and non-specific (green) DNA binding, and salt bridge formation (cyan) are highlighted.

PhoxC, *phoxT*, *orf+11* and *orf+12* were individually cloned into the shuttle vector pIB139, immediately downstream of a strong constitutive promoter P_{ermE}. Following conjugation, the overexpression constructs were introduced into the chromosome at the *attB* attachment site in an integrase-dependent manner and stably maintained in the presence of a selection antibiotic. Loss of phthoxazolin production was observed in the strain with *phoxC* overexpressed (Figure 3.23), implying that PhoxC is possibly acting as a negative regulator of the pathway. In the other three strains, no obvious alteration in the production level could be observed.

In-frame deletion of a putative global regulatory gene

It has recently been reported that cryptic phthoxazolin production in *S. avermitilis* KA-320 was activated via the disruption of an *avaR3* homologue (Suroto *et al*., 2017). AvaR3 is a pleiotropic, γ-butyrolactone autoregulator receptor that has been shown indirectly to control the biosynthesis of other polyketide natural products such as avermectin and filipin, either positively or negatively (Miyamoto *et al*., 2011). AvaR3 homologues (35% sequence identity) were identified in *Streptomyces* sp. KO-7888 and *Streptomyces* sp. OM-5714, both distant from the phthoxazolin clusters. To test the idea that the AvaR3 homologues may play a similar regulatory role in our producer strains, an in-frame deletion was carried out. The signals corresponding to phthoxazolins in the wild-type and Δ*avaR3* strains were comparable (no increase in production was seen, unlike in *S. avermitilis*) (Figure 3.24).

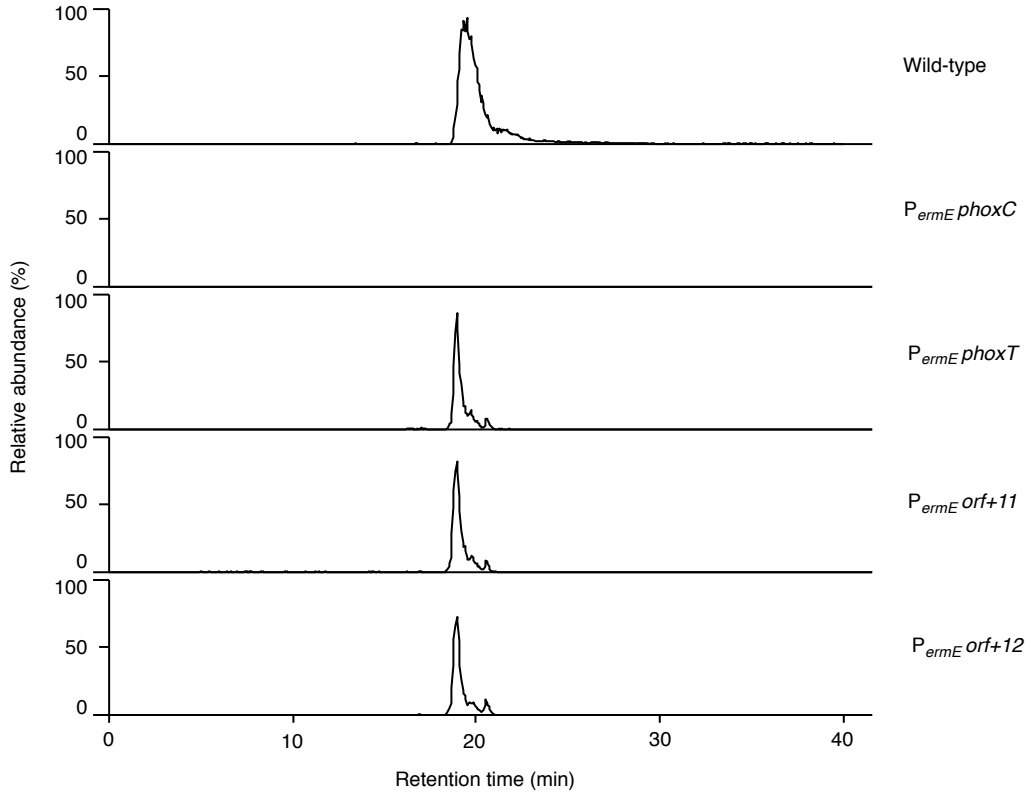


Figure 3.23. Phthoxazolin A production in *Streptomyces* sp. KO-7888 wild-type and deletion of a putative AvaR3-like regulator. Selected ion monitoring of *m/z* $[M+Na]^+$ 313.15 is shown. P_{ermE}, *ermE* promoter.

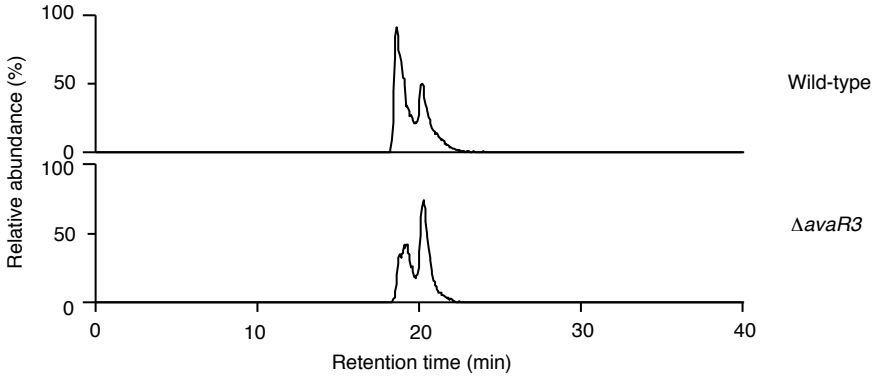


Figure 3.24. Phthoxazolin A production in *Streptomyces* sp. KO-7888 wild-type and *avaR3* homologue mutant. Selected ion monitoring of *m/z* $[M+Na]^+$ 313.15 is shown. P_{ermE}, *ermE* promoter.

3.2.8 Proposed biosynthetic pathway to phthoxazolins

Based on *in silico* analysis and results obtained from mutagenesis experiments detailed earlier in this chapter, a preliminary proposal could be made for the biosynthetic pathway to phthoxazolins (Figure 3.25).

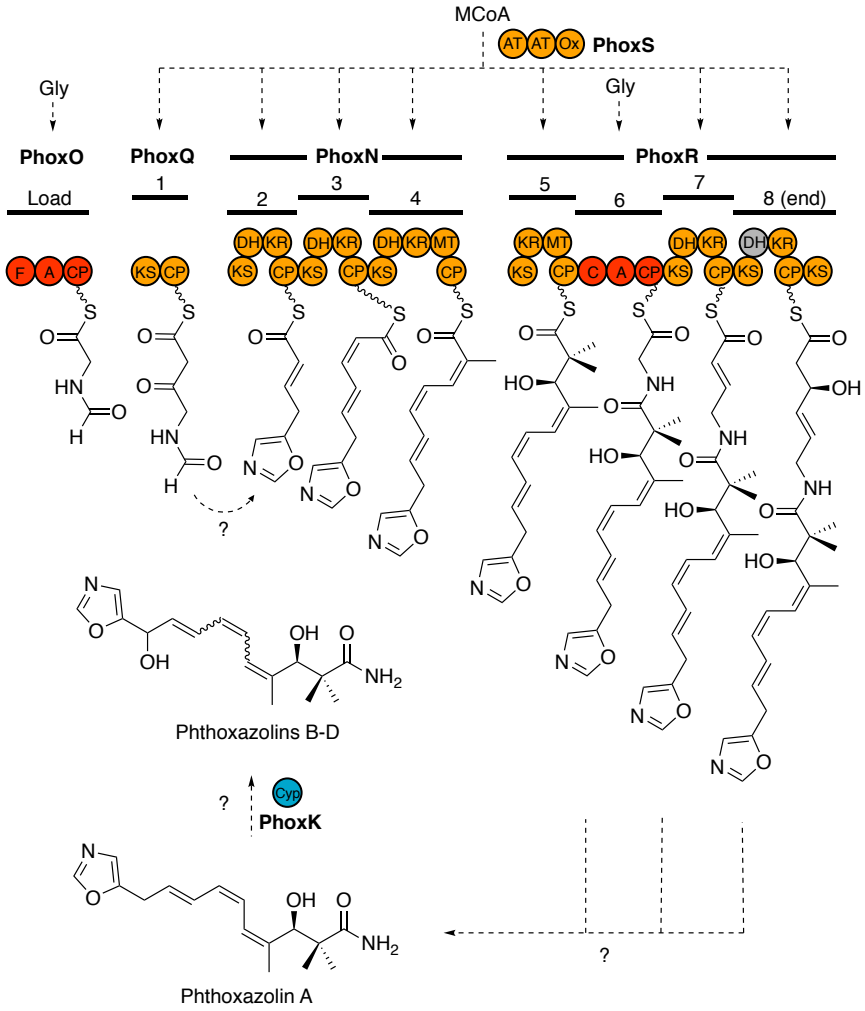


Figure 3.25. Proposed biosynthetic pathway to phthoxazolins. Loading and extension modules 1 to 8 are indicated. PKS (orange) and NRPS (red) domains proposed to be involved in biosynthesis are shown. The domain predicted to be inactive is coloured in grey. Module numbering from Section 3.2.3 (according to gene order) are indicated below the modules. CP, carrier protein; MCoA, malonyl-CoA; CYP, cytochrome P450.

The proposed pathway comprises two Gly-specific NRPS modules and seven malonyl-CoA-specific *trans*-AT PKS modules, which between them catalyse up to eight rounds of acyl/peptidyl chain extension. A presently unidentified enzyme

catalyses the release of the phthoxazolin A portion of the fully-extended chain. Phthoxazolin A may then be hydroxylated at C-10 by PhoxK, the only putative cytochrome P450 enzyme encoded within the cluster, to yield phthoxazolins B-D. Alternatively, PhoxK may catalyse the hydroxylation on PKS-NRPS-bound intermediates.

3.3 Concluding remarks and future work

The unusually large *trans*-AT PKS-NRPS clusters governing phthoxazolin biosynthesis in *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888 were subjected to detailed bioinformatic analysis and *in vivo* characterisation. All modules in the hybrid PKS-NRPS PhoxR were shown to be essential for phthoxazolin biosynthesis, whereas a number of other biosynthetic, tailoring and regulatory genes were found to be unnecessary for phthoxazolin biosynthesis. Through the overexpression of *phoxC*, the putative XRE-family like protein was identified as a likely candidate of the negative regulator of the cluster. The role of some genes in the cluster including *phoxB*, *phoxL* and *phoxM*, yet to be manipulated, still remain obscure and further work, ideally gene deletion/inactivation in combination with heterologous expression, would be required to delimit the cluster boundaries with confidence. Nevertheless, based on the preliminary work presented in this chapter, a biosynthetic pathway to phthoxazolins was proposed. The formation of the 5-oxazole and primary amide moieties, both key features of phthoxazolins, are investigated in Chapters 4 and 5, the following chapters of this thesis.

Chapter 4

Investigation of a key enzyme in phthoxazolin biosynthesis with relevance to 5-oxazole formation

4.1 Introduction

4.1.1 Oxazole-containing natural products

Oxazoles are five-membered heterocycles containing a nitrogen atom and an oxygen atom. They are found in many natural products that are endowed with remarkable biological activities. Prominent examples include the 50S ribosomal subunit inhibitor virginiamycin M (Di Giambattista *et al*., 1989; Pulsawat *et al*., 2007), the antimitotic agents rhizoxin (Scherlach *et al*., 2006; Partida-Martinez and Hertweck, 2007) and disorazoles (Carvalho *et al*., 2005; Tierno *et al*., 2009), the DNA gyrase inhibitor microcin B17 (Li *et al*., 1996; Parks *et al*., 2007) and plantazolicin which targets the plasma membrane of *Bacillus anthracis* (Lee *et al*., 2008; Molohon *et al*., 2016) (Figure 4.1).

4.1.2 Biosynthetic routes to oxazole and the related azol(in)e moieties

The biosynthetic routes to 2,4-substituted azol(in)es in RiPPs and (polyketide-) non-ribosomal peptides are well characterised. Early isotopic labelling studies revealed that the azol(in)e moieties originate from the amino acids Ser, Cys and Thr (Bayer *et al*., 1995; Lau and Rinehart, 1995). Although common biosynthetic precursors are utilised, the two classes of peptide natural products employ distinct enzymatic machineries and catalytic mechanisms to generate the heterocycles. In RiPPs, the heterocycles are created post-translationally using an ATP-dependent cyclodehydratase complex consisting of a YcaO-like cyclodehydratase, a dehydrogenase and a partner protein (Dunbar *et al*., 2012; Dunbar *et al*., 2014)

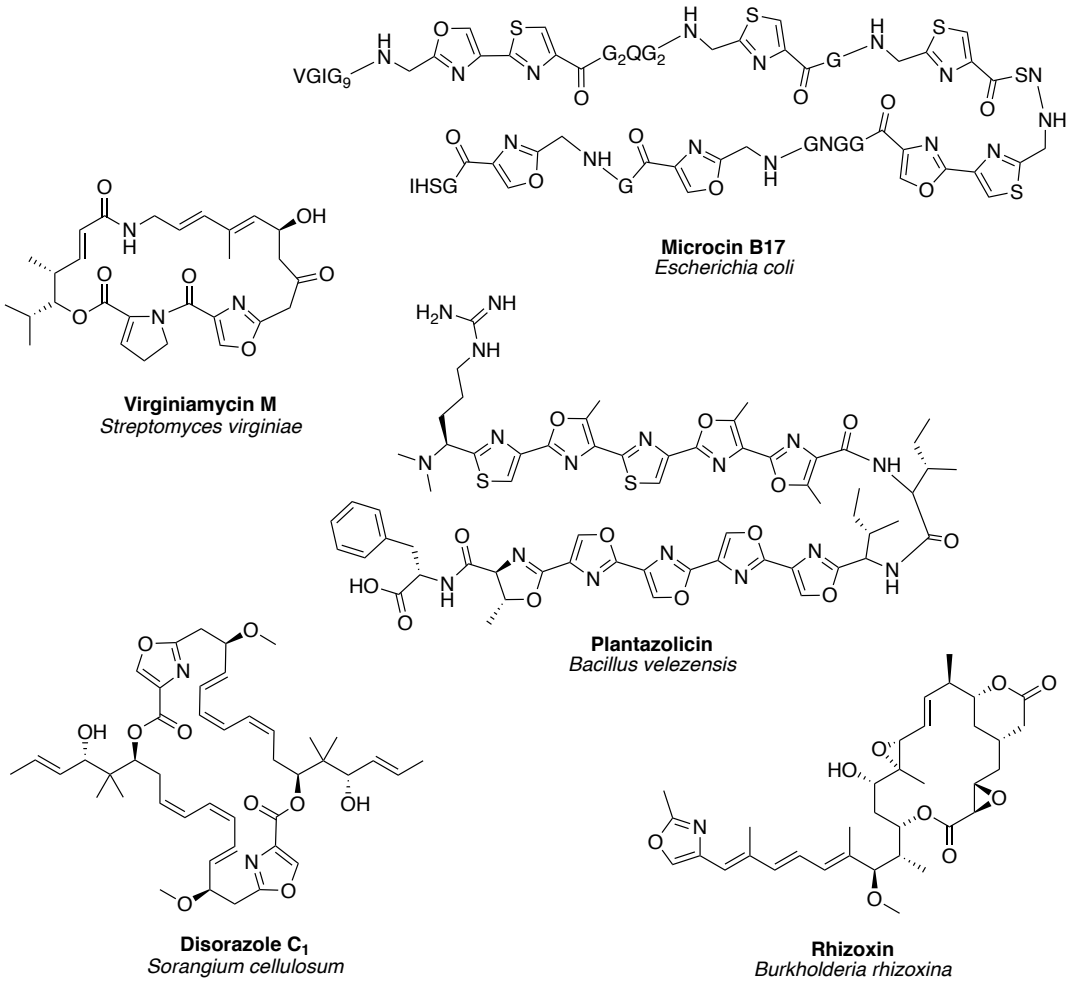


Figure 4.1. Structure and microbial origin of oxazole-containing natural products.

(discussed in Section 1.4.2). In contrast, NRPS modules utilise an embedded heterocyclisation (Cy) domain to introduce azoline moieties. An oxidase (Ox) domain is often present within the same module for the conversion of azolines to azoles (Bloudoff *et al*., 2017) (discussed in Section 1.4.7).

On the other hand, the biosynthetic route to the relatively uncommon 5-substituted oxazole moieties in polyketide-non-ribosomal peptides (including phthoxazolins, oxazolomycins and conglobatin), likely arising from a different mechanism, remains to be elucidated. This chapter presents the bioinformatic analysis, *in silico* structural modelling and in-frame deletion of *phoxP*, a proposed candidate for the cyclase catalysing 5-oxazole formation in phthoxazolin biosynthesis. An attempt at *in vitro* reconstitution of 5-oxazole formation is also described.

4.2 Results and discussion

4.2.1 Identification of candidate enzymes for 5-oxazole formation

Comparative sequence analysis of the phthoxazolin, oxazolomyin and conglobatin clusters revealed that the three contiguous genes *phoxO*, *phoxP* and *phoxQ* have counterparts in both the other clusters (Figure 4.2). As discussed in Chapter 3, PhoxO, comprising an F domain, an A domain specific for Gly and a PCP domain, potentially serves as the loading module in phthoxazolin biosynthesis; and PhoxQ was predicted as the first extension module. Sequences of PhoxOQ and homologues are approximately 50% identical, whereas PhoxP and homologues share more than 70% sequence identity, the highest amongst all three genes. This focused interest in the role and significance of PhoxP, a hypothetical protein of presently unknown function, in the initiation of phthoxazolin biosynthesis.

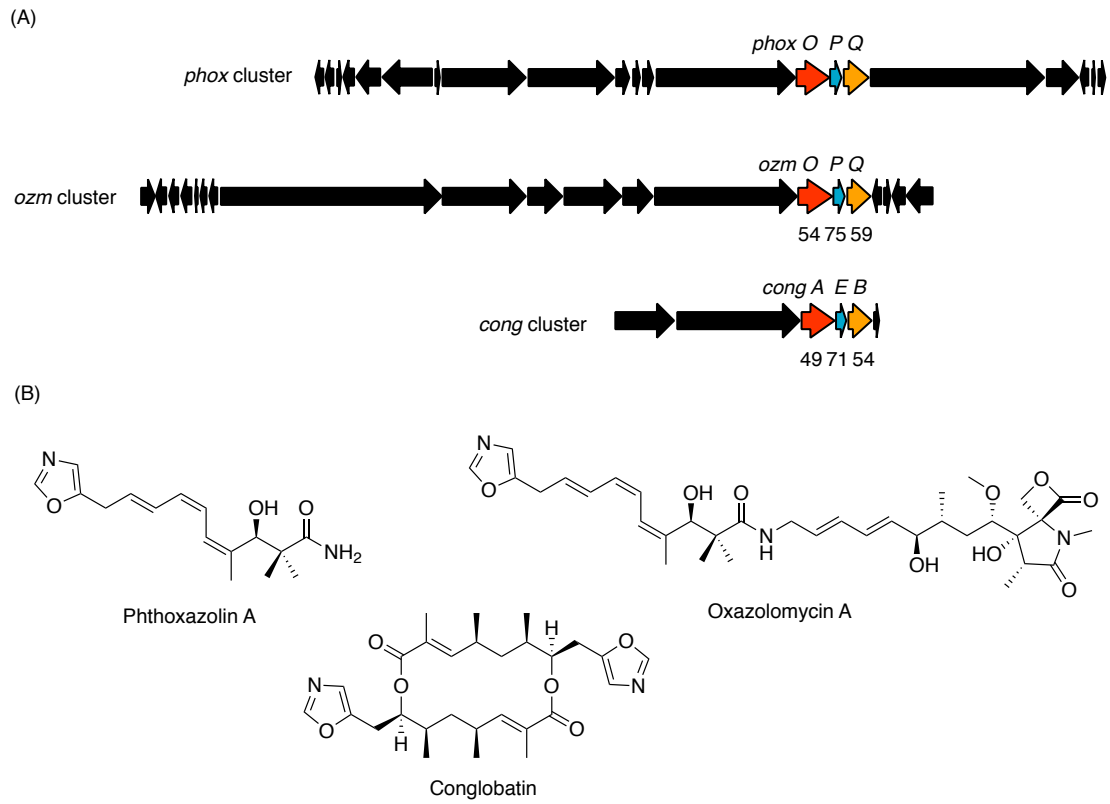


Figure 4.2. Comparison of gene organisation of the phthoxazolin (*phox*), oxazolomycin (*ozm*) and conglobatin (*cong*) clusters. (A) Genes conserved across all three clusters are coloured and their amino acid sequence identities are indicated in percentages. (B) Structures of phthoxazolin A, oxazolomycin A and conglobatin.

4.2.2 *In silico* analysis of PhoxP

Two conserved protein domains were detected in PhoxP using protein-BLAST: a large N-terminal domain belonging to a subfamily of the adenine nucleotide α hydrolase (αANH)-like superfamily (Aravind *et al*., 2002); and a small C-terminal domain resembling the N-terminal domain of some ATP-dependent DNA ligases (Sekiguchi and Shuman, 1997) (Figure 4.3). Members of the αANH-like superfamily include N-type ATP pyrophosphatases and ATP sulphurylases, and form an α-ß-α-fold which binds to the adenosine group. An N-terminal SGGKD motif is strongly conserved in multiple families of the superfamily, and this motif is also present in PhoxP and its homologues OzmP and CongE (Figure 4.4). Furthermore, phylogenetic anaylsis showed that PhoxP, OzmP and CongE form a distinct clade within the functionally-diverse αANH-like superfamily (Figure 4.5).

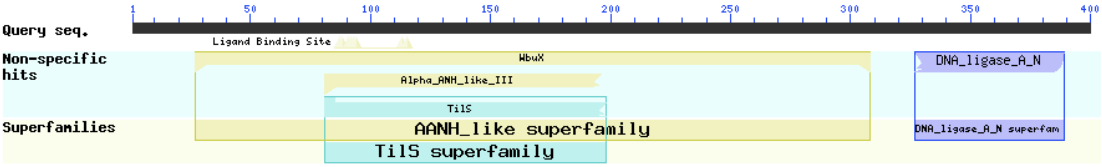


Figure 4.3. Protein-BLAST analysis of PhoxP. Yellow, αANH-like domain. Purple, DNA ligase N-terminal domain.

```
CongE       ------------------------MRVCQICALKEEHPGVSLDDDGVCSLCRLGVVGDLLE
PhoxP       MTGRPDATTPTAQPDPSVSPGGSVSSCQVCSLKAGHPGVVLDDQGVCNLCNLDFAEDLLV
OzmP        -------------------MEPTIVRCKICSLKEGHPGIVLDDQRVCSLCNLDFAGDLLV

CongE       NFSYTDQVHEEFTRSGPNPNGDYDCLFMYSGGKDSTYMLDKFVNEYGKRVLAYTFDVPFE
PhoxP       NYRYTSEVFTEFQQAPPG-TGEYDCLFMYSGGKDSTYMLDKFANEYGKRVLAFTFDVPFE
OzmP        NYKYTNEVFAEFQAAPPDPRGEFDCLFMYSGGKDSTYMLDKFVNEYGKRTLAYTFDVPFE

CongE       SSHAADNIALAREKIPATFVLDSDDDNIKTVMREVFNRPVT-KPGKYLDEKLPCVSCRTF
PhoxP       SEHAAQNIALAREKIPATFVLDADDDNIKLMMREVFNRPAPKKPGKYLDEKLPCVSCRTF
OzmP        STHAAQNIKLAQEKLPATFVVDADDDNIKKMMRDVFNRPRPKKPGQYLDEKLPCVSCRTF

CongE       FVIRAILYAFRNRIPYIALCADPQQILTMESDVREVVRGFYQTFGARLTDEVFRKEAEQI
PhoxP       FVLRAILMAYRQKIPYIALCADPQQILTMESNVREIVRDFYKTFGERLVDTLFGGQLEEL
OzmP        FVIRAILHAYREHIPYIALCADPQQILTMESNVREVVRGFYKSFGPELASELFDGEIEEI

CongE       LFADDADLPKIVFPFIAMRYSYDPDRIVADLKAKGLYNSSPLETHCTLFPLLNYYSFKNW
PhoxP       LFDDDENLPKIVFPFIAMRHEYDPDRIVEELRAKGLYQSSPMETHCTLFPLLNYYSYSNF
OzmP        LFAEDDELPKIVFPFIAMRHDYDPDRMVAELKEKGLYASSPMETHCTLFPLLNYYSFKNW

CongE       DCMFYKLNAASHRRAVRRNKNYDRSTFSITFPRAADLPSIEERMKRVVLDIAGTRGDPAA
PhoxP       DCMFYKLNASSHVRSVKRNASYDRNTFSIKFPRSLDLADVEDRLGKVVKEIAAGEGDRDE
OzmP        DCMFYKLNASSYTRAVSRNKDYDRSTFSIKFPRSLDLPEVEEKLKRVVLEIAAGEGDREQ

CongE       HRRELIDIFRELGATEDAARFVTEGFLGMRTVAADLGITL--------------------
PhoxP       HERSLIDLFQRMDASEDAARFVARSFLDMRSVAADLGIRLS-------------------
OzmP        HRAALVELFKQMDTTDDGAAFIAETFLDMRAVAADLGIKVG-------------------
```

Figure 4.4. Complete sequence alignment of PhoxP and its homologues. The highly conserved SGGKD motif is highlighted in yellow.

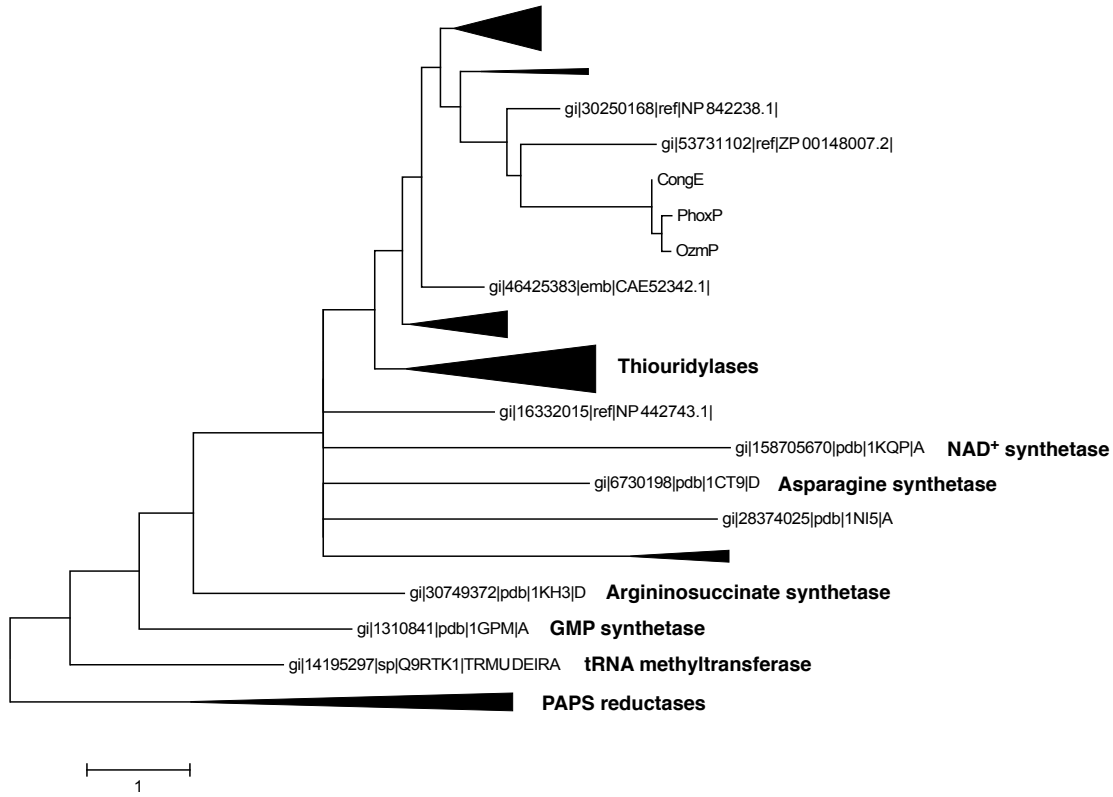


Figure 4.5. Phylogenetic tree of proteins belonging to the αANH-like superfamily. Evolutionary history was inferred using the maximum likelihood method (Jones *et al*., 1992). All positions containing gaps were eliminated and there were 210 amino acids in the final dataset. Sequences were aligned with MUSCLE (Edgar, 2004) and the evolutionary analysis was conducted in MEGA7 (Kumar *et al*., 2016). Functionally-characterised proteins are shown in bold. NAD^+, β-nicotinamide adenine dinucleotide; GMP, guanosine monophosphate; PAPS, phosphoadenosine phosphosulphate.

As PhoxOPQ, OzmOPQ and CongAEB are highly conserved, it is tempting to suggest that hybrid PKS-NRPS pathways governing 5-oxazole-containing natural products share a common mechanism for 5-oxazole formation. The F domain-containing loading module PhoxO possibly generate a formylated glycyl thioester which is subject to subsequent chain extension catalysed by the next module(s), PhoxQ (and PhoxN). PhoxP was proposed to catalyse 5-oxazole formation as a cyclodehydratase, via the nucleophilic attack of either the carbonyl keto group or the amide keto group, using ATP as activation energy. The co-location of *phoxOPQ* may

be a hint that PhoxP function alongside PhoxO and PhoxQ, early on in the phthoxazolin pathway, to incorporate the heterocycle (Figure 4.6).

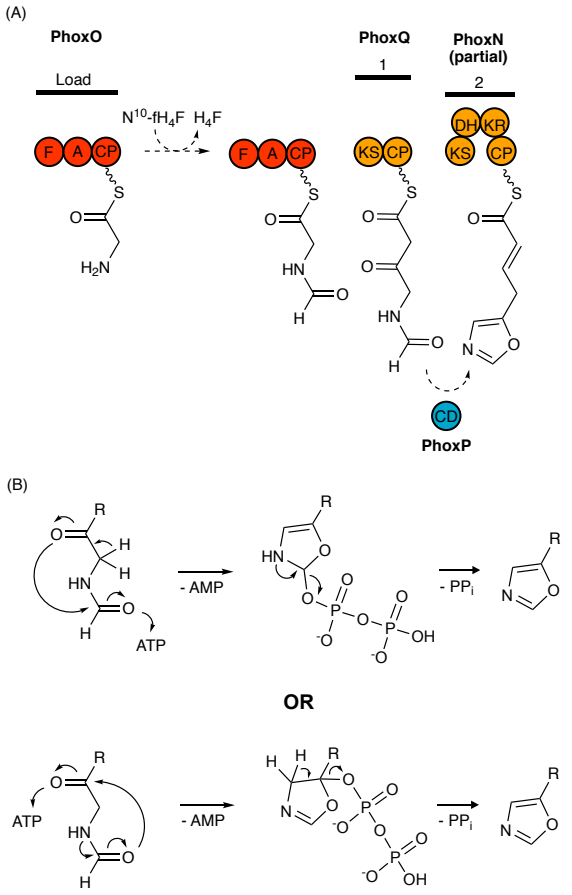


Figure 4.6. Proposed mechanism of 5-oxazole formation in phthoxazolin biosynthesis. (A) Involving PhoxP and early modules of the pathway. (B) Potential mechanisms of ATP-dependent oxazole ring closure catalysed by PhoxP. CD, cyclodehydratase.

4.2.3 *In silico* structural modelling of PhoxP and its homologues

The structure of PhoxP was modelled *in silico* using the protein fold recognition server, Phyre2 (Kelley *et al*., 2015). Multiple templates were selected to maximise confidence and alignment coverage , as no structural homologue of PhoxP was available (Table 4.1 and Figure 4.7).

An open α/ß structure with one parallel ß-sheet was identified as the conserved ATP binding domain (Asp83-Lys273) (Figure 4.8), which is a fold common across members of the N-type ATP pyrophosphatases. The PP-loop (YSGGKDS) crucial for interaction with the ß- and γ-phosphates of ATP was found in this domain, towards

the N-terminal end of α-helix-3. The sequences flanking the ATP binding domain appear to be unique to PhoxP and were modelled at a relatively low confidence, as none of the existing templates offer sufficient alignment coverage. They appear to form another domain that wraps around the side of ATP binding domain where the PP-loop is located. The structures of OzmP and CongE were also modelled using the same parameters, and they adopt a similar domain arrangement (Figures 4.9 and 4.10). This domain-domain interface in PhoxP, OzmP and CongE may contribute to forming the active site as in many N-type ATP pyrophosphatases, and may facilitate the interaction between the putative ATP-dependent cyclodehydratases and their respective PKS-NRPS partner(s).

Table 4.1. Alignment coverage, identify and confidence of templates used for the *in silico* structural modelling of PhoxP.

PDB ID	Function	% Alignment coverage	% Identity	% Confidence
5UDW	Sulfur transferase	40	14	99.4
3VRH	2-Thiouridine synthetase	44	18	99.3
5GHA	2-Thiouridine synthetase (TtuA)	44	15	99.2
3BL5	tRNA modifcation enzyme	35	18	99.1
1XNG	NAD^+ synthetase	36	12	99.1
2C5S	tRNA modifcation enzyme	30	24	99.1
1KH2	Argininosuccinate synthetase (AsS)	68	19	99.0

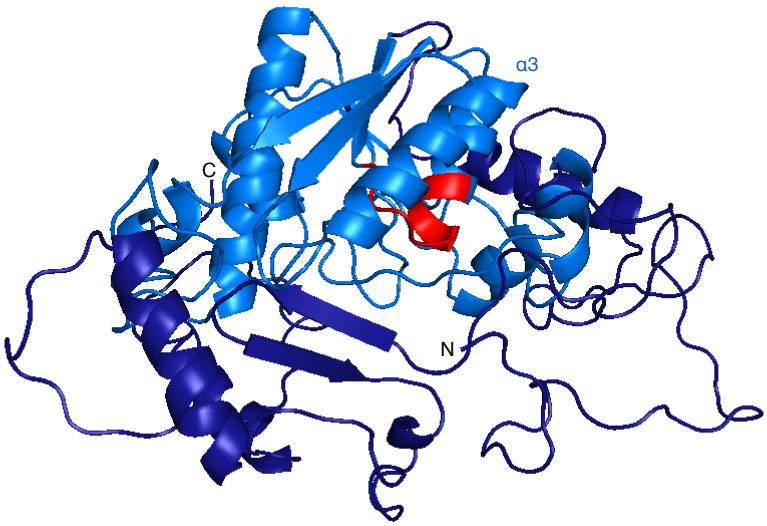


Figure 4.7. Structural model of PhoxP. The N- and C- terminal ends, and the PP-loop-containing α-helix-3 (α3) of the protein are indicated. Light blue, ATP-binding domain; dark blue, domain of unknown function; red, PP-loop.

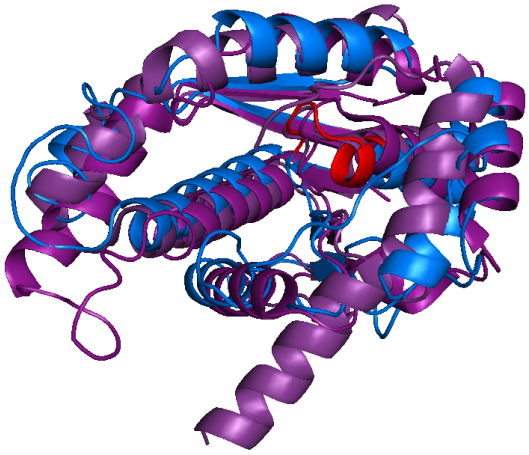


Figure 4.8. ATP-binding domain of PhoxP. Overlaid with the ATP-binding domain of two other N-type ATP pyrophosphatases, AsS (PDB: 1KH2) and TtuA (PDB: 5GHA). Blue, PhoxP; light purple, AsS; dark purple, TtuA.

Figure 4.9. Structural model of OzmP. The N- and C- terminal ends, and the PP-loop-containing α-helix-3 (α3) of the protein are indicated. Light green, ATP-binding domain; dark green, domain of unknown function; red, PP-loop.

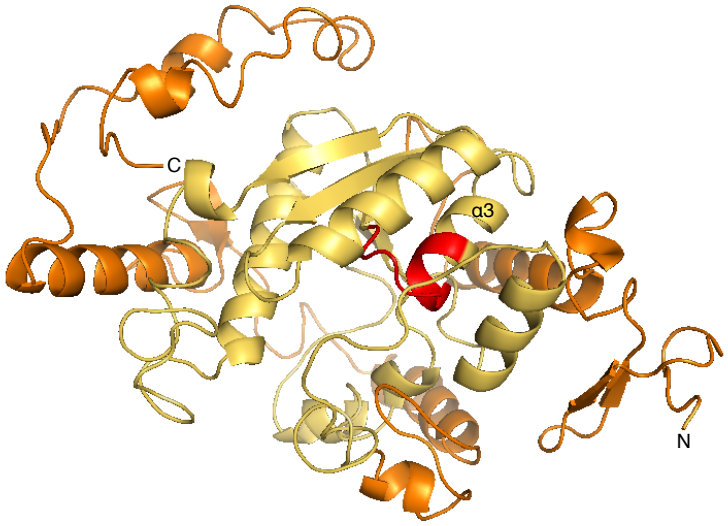


Figure 4.10. Structural model of CongE. The N- and C- terminal ends, and the PP-loop-containing α-helix-3 (α3) of the protein are indicated. Yellow, ATP-binding domain; orange, domain of unknown function; red, PP-loop.

4.2.4 In-frame deletion of *phoxP* and gene complementation

The in-frame deletion of *phoxP* yielded a mutant strain deficient in phthoxazolin production (Figure 4.11). In-*trans* gene complementations were performed with *phoxP* amplified from *Streptomyces* sp. KO-7888 and *Streptomyces* sp. OM-5714, expressed under the constitutive *ermE* promoter. Phthoxazolin production was restored in both complemented strains. This confirmed that PhoxP is essential for phthoxazolin biosynthesis. However, as no non-cyclised shunt products could be detected, the direct involvement of PhoxP in catalysing the oxazole ring closure could not be established at this stage.

OzmP and *congE*, were also tested for their ability to complement Δ*phoxP*. Δ*PhoxP*::*ozmP* was able to produce phthoxazolins at a lower level compared to that of the *phoxP*-complemented strains, whereas the introduction of *congE* failed to restore production (Figure 4.11). This is somewhat surprising as the homologues show a comparable level of sequence homology to *phoxP* and are predicted to act on the same substrate, an enzyme-bound *N*-formylglycyl intermediate. A possible explanation would be that CongE lacks the crucial residues, which are conserved in OzmP, for the interaction with the protein partner(s) of PhoxP.

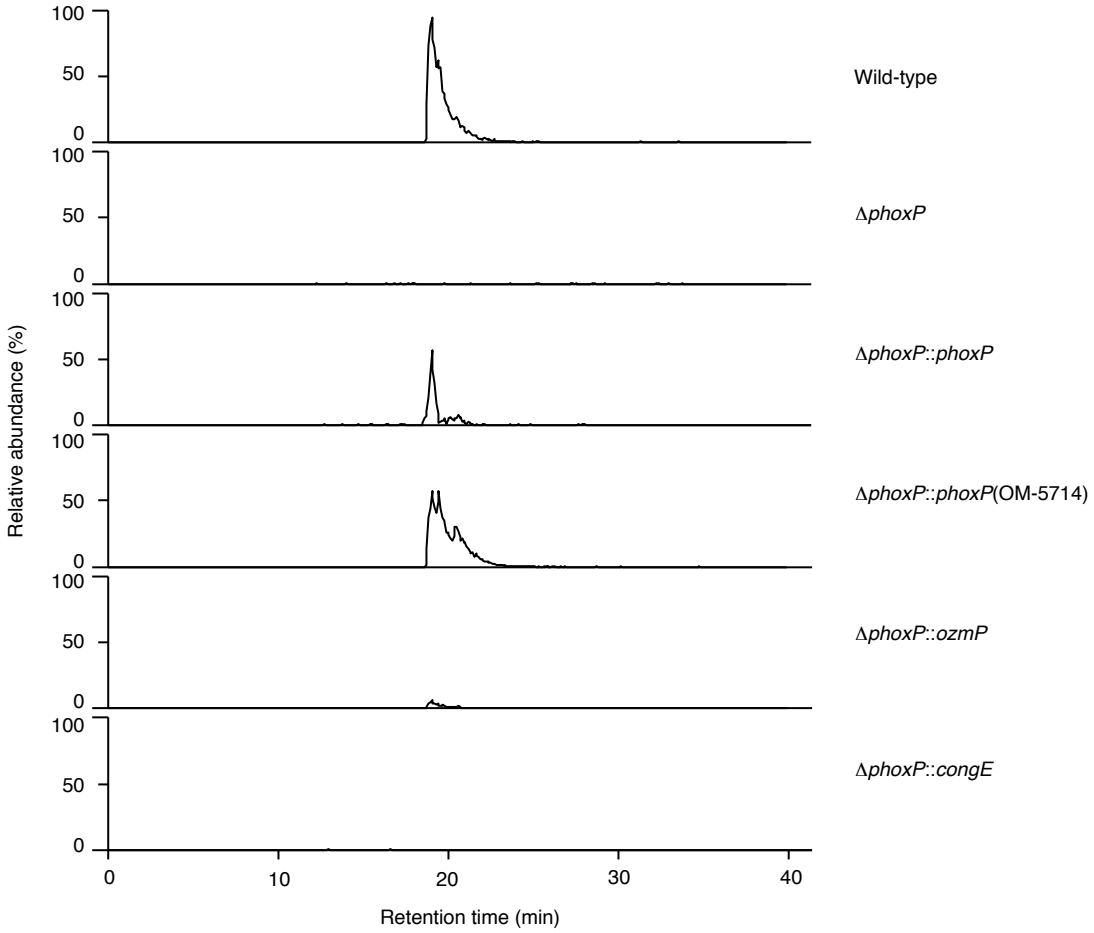


Figure 4.11. Phthoxazolin A production in *Streptomyces* sp. KO-7888 wild-type, *phoxP* mutant and complemented *phoxP* mutants.

4.2.5 Attempted reconstitution of 5-oxazole formation *in vitro*

Cloning, expression and purification of PhoxNOPQS

Genes encoding enzymes proposed to be involved in the early steps of phthoxazolin biosynthesis were amplified and cloned into the protein expression vectors, pET-28a(+) and pET-29b(+). *E. coli* BL21(DE3) was used as the heterologous host. The pET systems are driven by the IPTG-inducible bacteriophage T7 promoter and contain terminal His tags for protein purification by Ni^{2+} affinity chromatography. Proteins were eluted from the affinity column with buffer containing an increasing concentration of imidazole and clean protein fractions were subject to buffer exchange and concentration with centrifugal filter units of an appropriate molecular weight cut-off. The NRPS/PKS modules, PhoxO, PhoxQ and PhoxN (Val1-Gly1483), were co-expressed with MtaA, the 4'-phosphopantetheinyl transferase (PPTase) from

the myxothiazol pathway, to obtain the multienzymes in their *holo*-form (Silakowski *et al*., 1999).

The 300 mM imidazole elution fraction containing the 125.5 kDa PhoxO was concentrated using a 100 kDa cut-off filter (Figure 4.14), whereas the 100 mM to 300 mM imidazole elution fractions were used for the other four enzymes (Figure 4.12). One-hundred kilodalton cut-off filters were used for PhoxN (Val1-Gly1483) (157.4 kDa) and PhoxS (121.2 kDa), and 30 kDa cut-off filters were used for PhoxQ (93.1 kDa) and PhoxP (46.2 kDa). The yields of PhoxN (Val1-Gly1483), PhoxO, PhoxP, PhoxQ, and PhoxS were 9.89, 0.28, 1.20, 8.16 and 19.9 mg L^{-1} culture, respectively.

Cloning, expression and purification of FolD and the chemoenzymatic synthesis of the formyl donor N^{10}-formyltetrahydrofolate

FolD, encoding an NAD^+-depdent formyltrasferase, was used to synthesise the formyl donor N^{10}-formyltetrahydrofolate (N^{10}-fH_4F) required by the F domain from tetrahydrofolate (H_4F) and formaldehyde, chemoenzymatically (Schönafinger *et al*., 2006) (Figure 4.13). *FolD* was amplified from the genomic DNA of *E. coli* DH10B and cloned into pET-28(a). The 100 mM to 300 mM imidazole elution fractions containing the 33.2 kDa FolD were combined and concentrated using a 10 kDa cut-off filter (Figure 4.14). The yield of FolD was 39.5 mg L^{-1} culture.

The reaction mixture, consisting of H_4F, formaldehyde, NAD^+, β-mercaptoethanol and purified FolD in potassium phosphate buffer (pH 6.0) was incubated at 21 °C for 2 hours. Product formation was monitored by UV spectroscopy at 340 nm.

One-pot enzymatic assay

Due to the instability of some of the larger multienzymes, the *in vitro* assay was carried out in a one-pot manner, instead of adding individual enzymes stepwise, to minimise processing and incubation time. The one-pot reactions consist of purified PhoxO, PhoxQ, PhoxN (Val1-Gly1483), PhoxS and PhoxP, ATP, $MgCl_2$, Malonyl-CoA, Gly (Ser or no amino acid for the controls), NADPH and the FolD reaction in 25 mM Tris-HCl buffer (pH 7.5). PPTase and CoA were also added to convert modules still in their *apo*-form to *holo*-form. In a parallel set of experiments, pre-purification cell lysate of PhoxO, PhoxQ, PhoxN (Val1-Gly1483), PhoxS and PhoxP was used in place of purified enzymes. All reactions were incubated at 25 °C for an hour, then

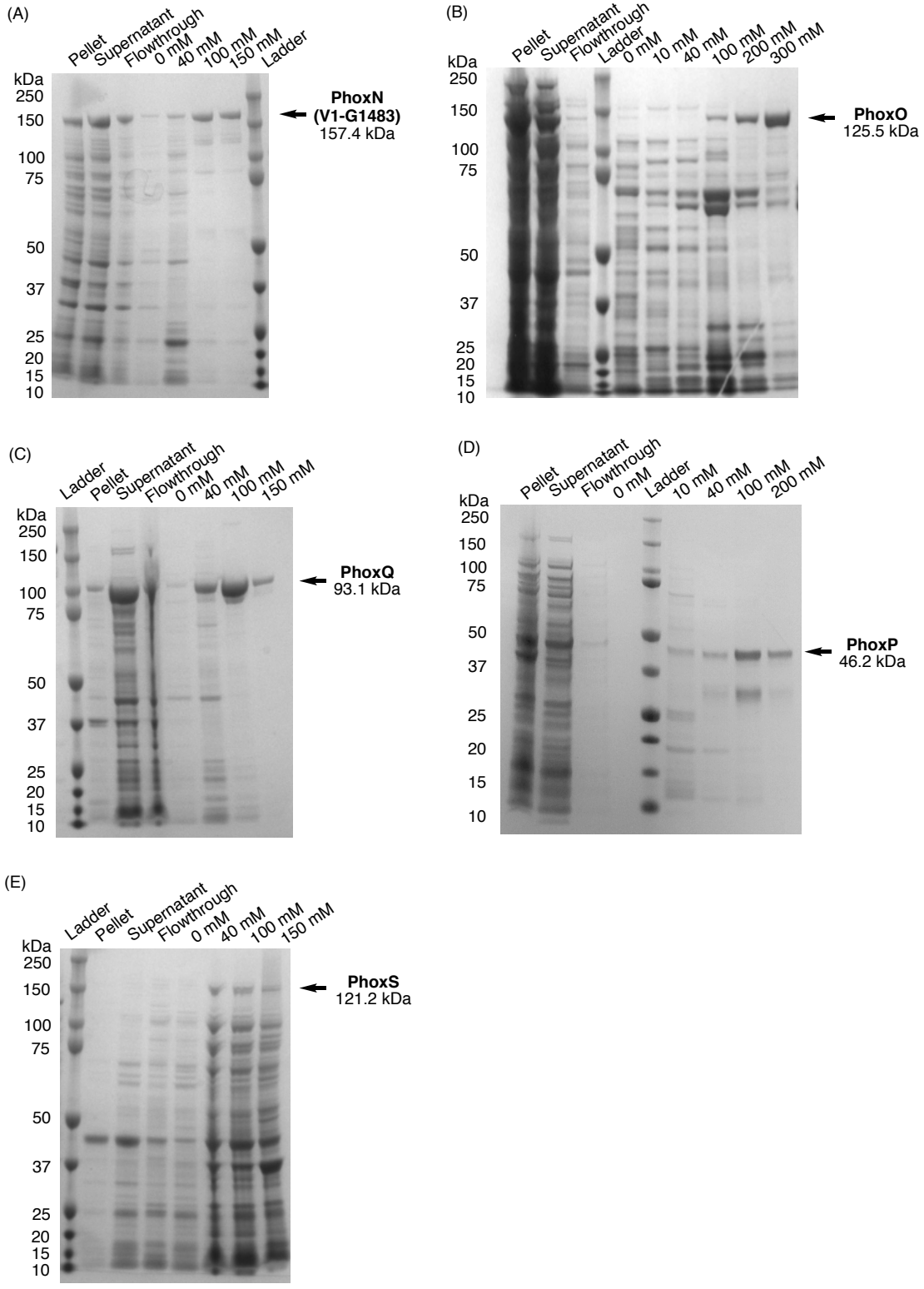


Figure 4.12. SDS-PAGE of purification fractions of PhoxN (Val1-Gly1483), PhoxO, PhoxQ, PhoxP and PhoxS. Imidazole concentration of elution buffers is indicated.

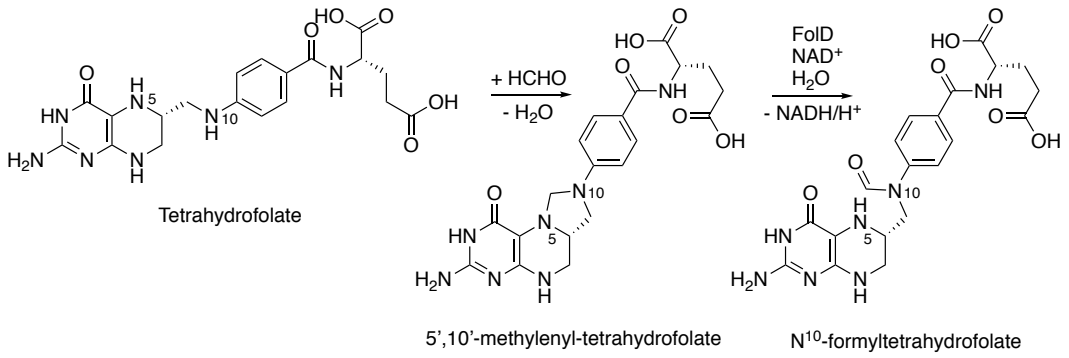


Figure 4.13. Chemoenzymatic synthesis of the formyl donor N^{10}-H_4F. HCHO, formaldehyde.

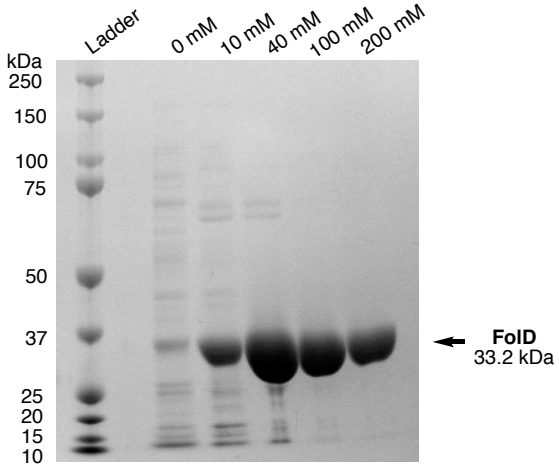


Figure 4.14. SDS-PAGE of purification fractions of FolD. Imidazole concentration of elution buffers is indicated.

precipitated for subsequent thioester cleavage with potassium hydroxide. The treated samples were subjected to LC-MS analysis, but unfortunately no m/z peak corresponding to any cyclised intermediate was detected.

4.3 Concluding remarks and future work

PhoxP was found to be essential for phthoxazolin production via an in-frame deletion experiment. PhoxP was proposed as an ATP-dependent cyclodehydratase to catalyse 5-oxazole formation on an enzyme-bound formylglycyl intermediate, unprecedented for oxazole formation, and possibly plays a key role in the initiation of phthoxazolin biosynthesis. PhoxP and its homologues from the oxazolomycin and conglobatin pathways show high sequence homology, indicative of a common and potentially novel biosynthetic route to incorporating the 5-oxazole(s) in hybrid PKS-NRPS systems. Moreover, they represent a new clade of enzymes within the functionally-diverse adenine nucleotide α hydrolase superfamily. *In silico* structural modelling further revealed a conserved ATP binding domain and a seemingly novel domain fold that could be involved in facilitating the interaction between the putative cyclodehydratases and their protein partner(s). *In vitro* reconstitution of the early steps of phthoxazolin biosynthesis was attempted in a one-pot manner to validate the proposed catalytic role of PhoxP and the timing of 5-oxazole formation, but was unsuccessful. For future work, validation and optimisation of the individual steps of the *in vitro* reaction should be carried out. Point mutagenesis and crystallographic studies of PhoxP could also provide important insights into the mechanism of and catalytic residues involved in 5-oxazole formation.

Chapter 5

Investigation of chain release and primary amide formation in phthoxazolin biosynthesis

5.1 Introduction

Another important question to address in phthoxazolin biosynthesis is the mode of chain release and primary amide formation. As discussed in Chapter 3, no thioesterase (TE) was identified in the phthoxazolin clusters; and the inactivation of the carrier proteins in PhoxR, the hybrid PKS-NRPS proposed to be involved in the last chain extension step(s), demonstrated that all four modules in the PhoxR are essential for phthoxazolin biosynthesis, implicating that a polyketide-peptide chain larger than phthoxaozlin may possibly be biosynthesised prior to release, via a presently unknown mechanism (see Section 3.2.7). Also, no obvious candidate for catalysis of the installation of the primary amide moiety could be identified within the clusters. These findings prompted a search for the enzyme(s) responsible.

5.1.1 Primary amides in natural product biosynthesis

Primary amide moieties are often observed in peptide natural products, and to a lesser extent, in polyketides (Figure 5.1). A number of distinct biosynthetic routes to primary amides, catalysed by different enzymes have been reported.

Asparagine synthetase B (AsnB)-like enzymes catalyse (primary) amide formation in various PKS, NRPS and RiPP pathways (Zhang *et al*., 2007; Wang *et al*., 2011, Tao *et al*., 2016). AsnB-like enzymes belong to a class of ATP-dependent glutamine amidotransferases and utilises Gln and ammonia as nitrogen donors (Figure 5.2A). For instance, OxyD in the oxytetracyline pathway (type II PKS) catalyses the amidation of a malonyl starter unit (Pickens and Tang, 2010), wheareas TsrC converts a carboxylate intermediate into the mature RiPP thiostrepton (Liao and Liu, 2011).

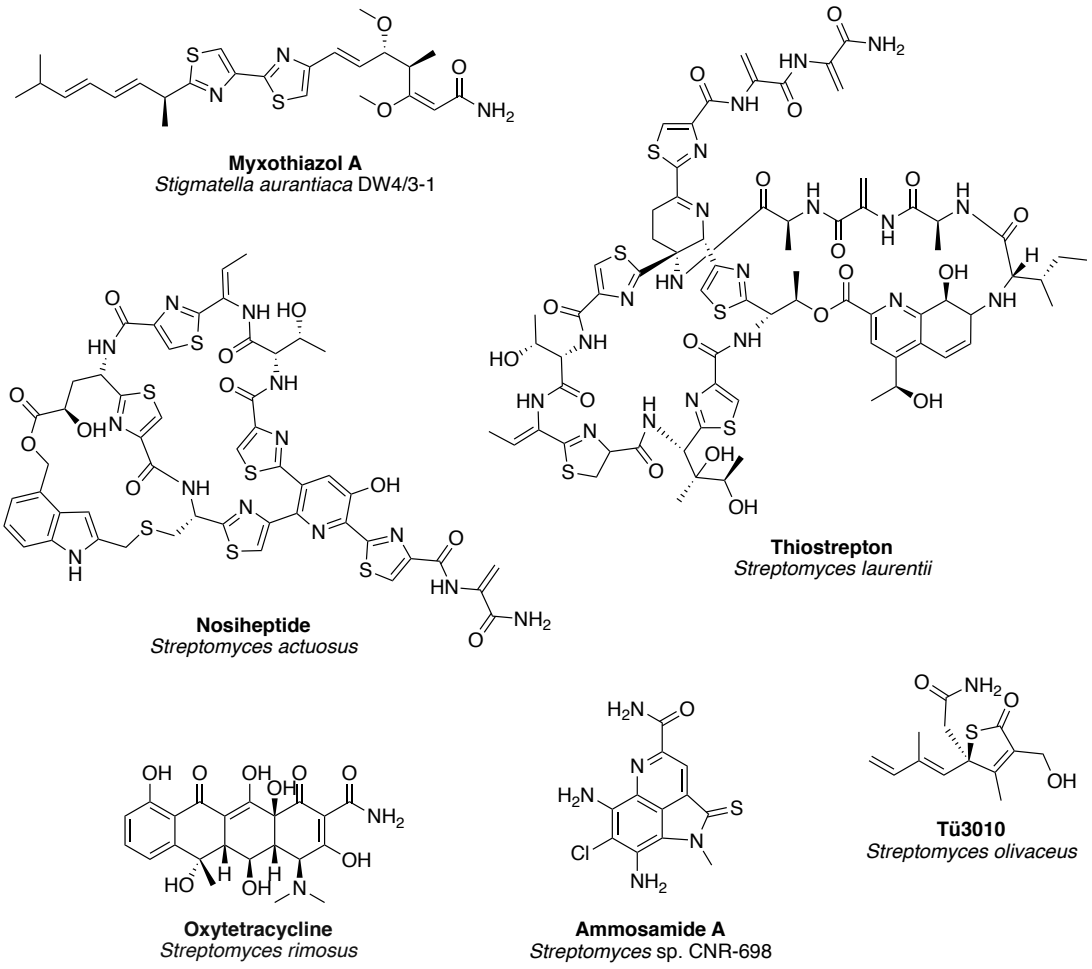


Figure 5.1. Structure and microbial origin of primary amide-containing natural products.

The biosynthesis of another RiPP nosiheptide employs a distinct strategy for amidation. The terminal dehydroalaninyl moiety derived from Ser of the extended precursor peptide was proposed to undergo tautomerisation to form a methyl imine intermediate, and NosA catalyses an enamide dealkylation to yield the mature nosiheptide with a primary amide moiety (Yu et al., 2011; Liu et al., 2015) (Figure 5.2B). The biosynthetic pathways to several other thiopeptides including GE2270A, nocathiacin and berninamycin A were also shown or postulated to share a common mechanism for primary amide formation, catalysed by NosA homolgoues in the respective clusters (Tocchetti et al., 2013; Ding et al., 2010, Malcolmson et al., 2013).

On the other hand, the hybrid PKS-NRPS pathways governing melithiazol and myxothiazol extend the polyketide-peptide chains with an additional Gly, and an integrated, luciferase-like monooxygenase (MOx) domain in the final NRPS module presumably catalyses the α-hydroxylation of Gly to generate an unstable

intermediate, leading to the release of melithiazol A and myxothiazol A, respectively (Weinig et al., 2003; Müller et al., 2006) (Figure 5.2C). Furthermore, Jordan and Moore (2016) also speculated the role of Amm4, an F420-dependent oxidase, in liberating from longer peptide chains the primary amide-containing precursors to ammosamides.

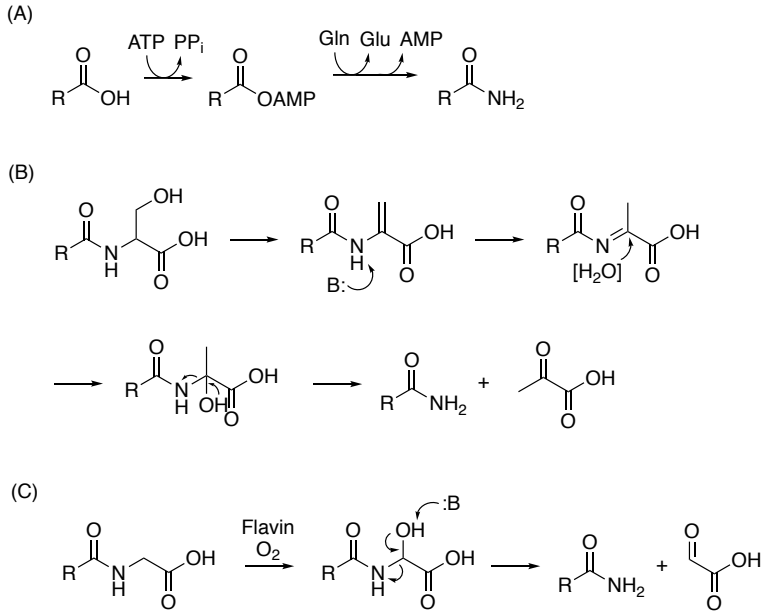


Figure 5.2. Alternative biosynthetic routes to primary amides. (A) ATP-dependent transamidation. (B) Enamide dealkylation. (C) Flavin-dependent oxidative peptide cleavage.

This chapter presents the bioinformatic analysis and in-frame deletion of various candidate genes, both within and outside of the phthoxazolin cluster, thought to be responsible for chain release and/or primary amide formation.

5.2 Results and discussion

5.2.1 Investigation of *phoxE* and *phoxF*

In silico analysis of PhoxE, a putative α/ß hydrolase

PhoxE belongs to the α/ß hydrolase fold superfamily (Figure 5.3). The α/ß hydrolase fold is common to a number of hydrolytic enzymes including bacterial lipases, serine peptidases, carbohydrate esterases and PKS/NRPS thioesterases (Lenfant *et al.*, 2012). As the only putative enzyme containing an α/ß hydrolase fold, PhoxE was regarded as a plausible candidate for the catalysis of chain release by hydrolysis.

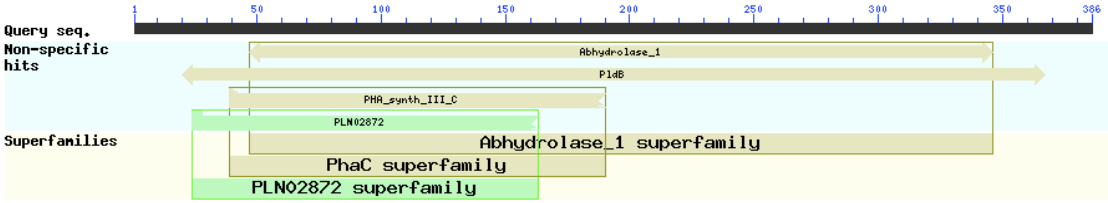


Figure 5.3. Protein-BLAST analysis of PhoxE. Yellow, αß-fold hydrolase domain.

In silico analysis of PhoxF, a putative NRPS containing a TD domain

As discussed in Section 3.2.6, PhoxF encodes an unusual NRPS module comprising a PCP domain and a TD domain. All conserved motifs characteristic of an active reductase were identified from PhoxF-TD. The motifs are either fully conserved or have up to three residues substituted with similar amino acids. The domain was suspected to be responsible for the release of an aldehyde product, as observed in several other NRPS systems (Gaitatzis *et al.*, 2001; Velasco *et al.*, 2005; Bailey *et al.*, 2007; Li *et al.*, 2008). However, how the primary amide moiety in phthoxazolins may arise from an aldehyde is unclear.

Another candidate for chain release previously identified within the cluster is the C-terminal C-domain in PhoxG. However, as described in Section 3.2.7, in-frame deletion of *phoxG* did not affect phthoxazolin production and so PhoxG-C2 was excluded from further investigation.

In-frame deletion of *phoxE* and *phoxF*

In-frame deletion mutants Δ*phoxE* and Δ*phoxF* were successfully generated. Neither of the mutants showed altered phthoxazolin production (Figure 5.4), and so PhoxE and PhoxF are not involved in the biosynthetic pathway and the enzyme responsible for the chain release remained to be identified.

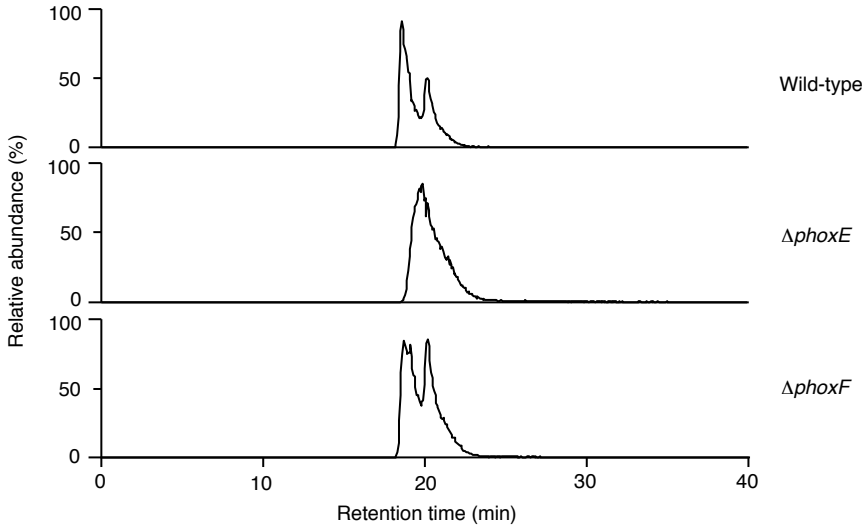


Figure 5.4. Phthoxazolin A production in *Streptomyces* sp. KO-7888 wild-type, *phoxE* mutant and *phoxF* mutant. Selected ion monitoring of *m/z* $[M+Na]^+$ 313.15 is shown.

5.2.2 Investigation of an asparagine synthetase-like gene

In silico analysis of Orf0034

As asparagine synthetase B (AsnB)-like enzymes have been shown or proposed to catalyse primary amide formation in the biosynthesis of various natural products, the amino acid sequence of OxyD, the AsnB-like enzyme from the oxytetracycline pathway, was used to detect homologues in the genomes of the phthoxazolin producers, using an in-house protein-BLAST tool created by Dr. Markiyan Samborskyy. *Orf0034*, a gene putatively encoding an AsnB-like enzyme 32% identical to OxyD, was identified more than 1 Mbp away from the phthoxazolin clusters. The N-terminal Cys-Arg-Asn-Asp required for glutamine-dependent activity and the SGGLD ATP-binding motif are present in Orf0034 (Boehlein *et al*., 1996; Larsen *et al*., 1999) (Figures 5.5 and 5.6). However, the Leu, Val and Gln residues essential for ß-aspartyl-AMP intermediate formation are replaced with Phe, Ile and Met, respectively. It is unclear whether or not this AsnB-like enzyme is functional.

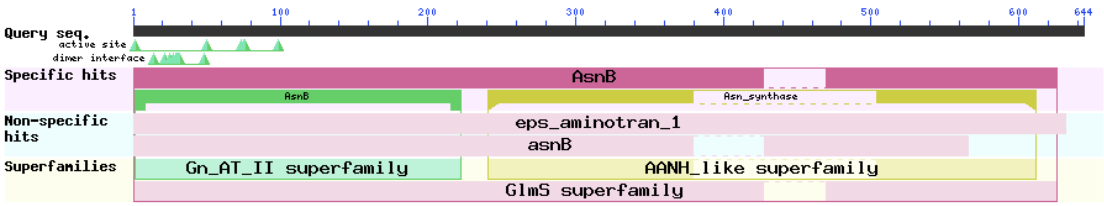


Figure 5.5. Protein-BLAST analysis of Orf0034. Pink, asparagine synthetase B/glutamine amidotransferase superfamily domain. Green, N-terminal domain for glutamine binding and hydrolysis. Yellow, C-terminal domain for ATP, Mg^{2+} and Asp binding.

```
AsnS        MCSIFGVFDIKTD-AVELRKKALELSRLMRHRGPDWSGIY----ASDNAILAHERLSIVD
Orf0034     MCGIAGTYHWPDG-----KAVTDRLTEVLAHRGPDGAGRYSHPAGAGEVHLGHRRLAIVD
FdmV        MCGIAGWVVRQDGRPREDPAVAEAMAGAMACRGPDEQGVR----GGRHVTLVHTRMAVID
OxyD        MCGIAGWIDFERN-LAQERATAWAMTDTMACRGPDDAGLW----TGGHAALGHRRLAVID
TueE        MCGIAGWVSYRRD-LTQQPQVLDGMNGTMACRGPDAGDTW----VGRHAALGHRRLAVID

AsnS        VNA-GAQPLYNQ---QKTHVLAVNGEIYNHQALRAEYGD-RYQFQTGSDCEVILALYQEK
Orf0034     LSGTGAQPMVSG-----GLALTYNGELYNAPELRRELESAGASFRGTSDTEVVLEAWRRW
FdmV        LLG-GRQPMAADESEDPAATLTYCGEIYNAAELRSDLAGRGHRFRTRSDTEVVLRAYLEW
OxyD        PAH-GRQPMHSTLPDGTSHVITFSGEIYNFRELRVELESQGHRFRTHCDTEVVLHGYTRW
TueE        LPG-GAQPMRVGTGDGD-VVMVYSGETYNYTELRDELRRRGHRFTTESDTEVVLRGYVEW

                                        //

AsnS        ---AVKDNVTDKNELRQALEDSVKSHLMSDVPYGVLLSGGLDSSIISAITKKYAARRVED
Orf0034     AEGQERARSGELPDLAAVVEESTRSHLLSDVPVATFLSGGLDSSYLTALAARH-------
FdmV        ---HEEDLAGTVATVRGLLESSVARELVSDVPLSVLLSGGLDSSTVAALAARALADG---
OxyD        ---HTDDLETTIATVRGLLTDRVRRQLVSDVPLCTLLSGGLDSSAVTALAARA---G---
TueE        ---HTDDRQTTVARVRELLEDSISRQLVADVPRCTLLSGGLDSSVITAVAAQRLGER---

AsnS        QERSEAWWPQLHSFAVG------------LPGSPDLKAAQEVANHLGTVHHEIHFTVQEG
Orf0034     -------RPGISAYTIGFRAEDARFEAM----PDDLRYARQVAARFGVDLHEIEIAPDVL
FdmV        -D-----GGPVRTTTVTYSGYGDNFQPDLVRSAPDSPYARAVAEHIGAEHLEIELTTADL
OxyD        -------DGPVRTFSVDFSGAGTRFQPDAVRGNTDAPYVQEMVRHVAADHTEVVLDSADL
TueE        -------GETVRSFNVDFPGQAENFKAIALAPTLDTPYAHAVAEHVRSDHRDILLHGSAL

AsnS        LDAIRDVIYHIETYDVTT----IRASTPMYLMSRKIKAMGIKMVLSGEGSDEVFGGYLYF
Orf0034     D-LLPRM-T----YHLDEPI-GDPAAINTFLICSAAREAGVKVMLSGMGADELFAGYRKH
FdmV        IDPVARR-TVLRAQDVPAPF-GDM-DTSTYQAFA-GVRRHSRVALTGESADEIFGGYSWV
OxyD        AAPEVRA-AVLGATDLPPAFWGDM-WPSLYLFFR-QVRQHCTVALSGEAADELFGGYRWF
TueE        ADPEVRR-AAVGARDFPNGI-GDR-DNSLHLLFK-AVREQSTVALSGESADELFGGYPWF
```

Figure 5.6. Partial sequence alignment of AsnB and AsnB-like enzymes. OxyD, FdmV and TueE, AsnB-like enzymes from the oxytetracycline pathway in *Streptomyces rimosus*, fredericamycin pathway in *S. griseus* and Tü3010 pathway from *S. olivaceus* Tü 3010, respectively. AsnB, asparagine synthetase B in *Escherichia coli* K12, was included for comparison. The ATP binding motif and residues required for ß-aspartyl-AMP intermediate formation are highlighted in yellow. Residues required for glutamine-dependent activity are highlighted in cyan.

In-frame deletion of *orf0034*

In-frame deletion of *orf0034* was carried out and phthoxazolin production was not affected in Δ*orf0034* (Figure 5.7). This result suggests that the phthoxazolin pathway employs a mechanism distinct from that of the oxytetracycline, fredericamycin and Tü3010 pathways to install its primary amide moiety.

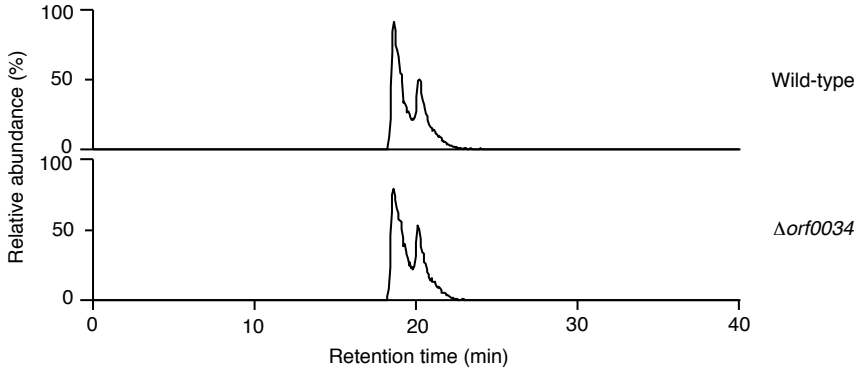


Figure 5.7. Phthoxazolin A production in *Streptomyces* sp. KO-7888 wild-type and *orf0034* mutant. Selected ion monitoring of *m/z* $[M+Na]^+$ 313.15 is shown.

5.2.3 Investigation of a luciferase-like monooxygenase gene in cluster 13

In the light of the discovery of the two primary amide-containing lipopeptides by our collaborators Dr. Yuki Inahashi and Prof. Satoshi Ōmura (Kitasato Institute, Japan) (as discussed in Section 2.3), a closer examination of cluster 13, the NRPS cluster likely governing the production of the lipopeptides (see Section 2.2.4), was carried out. A luciferase-like monooxygenase gene *orf3515* in *Streptomyces* sp. KO-7888 (homologue of *orf3228* in *Streptomyces* sp. OM-5714) was subjected to further investigation.

***In silico* analysis of Orf3515**

Orf3515 belongs to the flavin-utilising monooxygenase superfamily (Figure 5.8), members of which include alkanesulfonate monooxygenases, nitrilotriacetate monooxygenases, alkanal monooxygenases and tetrahydromethanopterin reductases. Some of the residues implicated to be involved in the binding of the flavin cofactor are present in Orf3515 (Baldwin *et al*., 1995) (Figure 5.9). With the integrated monooxygenase (MOx) domain shown to be catalysing the release of the amide precursor to melithiazol as a precedent, it is tempting to speculate that

Orf3515 may be responsible both the release of the lipopeptides and phthoxazolins, via a common mechanism from the extended (polyketide-) peptides.

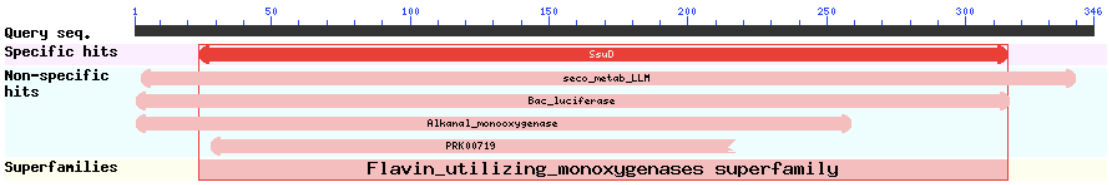


Figure 5.8. Protein-BLAST analysis of Orf3515. Red, luciferase-like monooxygenase domain.

```
LuxA          MKFGNFLLTYQPPELSQTEVMKRLVNLGKASEGCGFDTVWLLEHHFTEFG-LLGNPYVAA
Orf3515       MDLSLFYFAHDSTAPGEAGRYELLIEGAKLADRSGLAAVWTPERHFDPFGGAYPNPSVLG
MelG-MOx      --FSLFYFASDERE-RSGDKYRLLMEGARFADEHGFTAVWTPERHFHSFGGIYPNPSVVS

LuxA          AHLLGATETLNVGTAAIVLPTAHPVRQAEDVNLLDQMSKGRFRFGICRGLYDKDFRVFGT
Orf3515       AAVAMCTQRVGIRAGSVVAPLHHPARIAEEWSVVDNLSGGRAGVAFASGWNTVDFALRPE
MelG-MOx      AAIAATTRNLRIRAGSVVLPLHSPIRVAEEWSIVDNLSGGRVDLSFASGWHPNDFVLAPE

LuxA          DMDNSRALMDCWYDLMKEGFNEGYIAADNEHIKFPKIQLNPSAYTQGGAPVYVVAESAST
Orf3515       NHADRRALVRDTAEEVRRLWRGEALSTTDGLGRPVDIRAYPPA-VQPELPVWLTSAGGVD
MelG-MOx      RYAGARGQLMSQIEAFQKLWRGEAVVFPNGLGQDVEVRTLPRP-IQPDVAIWLTAAGNPE

LuxA          T-EWAAERGLPMILSWII-NTHEKKAQLDLYNEVATEHGYDVTKIDHCLSYITSVDHDSN
Orf3515       TFRAAARMKAGILTHLLGQSLEEVAGKIAEYRRVAA-EEHDGWSGHVVLMLHTLLGDDLD
MelG-MOx      TFRAAGERGLNVLTHLLGQNLAELAKKIQIYRDAWKAAGHGPGAGHVTLMLHTFLGEDRG

LuxA          RAKDICRNFLGHWYDSYVNATKI-FDDSDQTKGYDFNKGQWRDFVLKGHKDTNRRID-YS
Orf3515       QVREQVREPFTHYLKSSANLTAKSFT----GKDTDLSQLNADDMDYLA----ARAFDRYF
MelG-MOx      AVRQKIQGPLREYLKSSVGLLRSVIGP--LPHGAEFESLSEADIDVLL----SKAIERYF
```

Figure 5.9. Partial sequence alignment of flavin-dependent luciferase and luciferase-like monooxygenases. Residues involved in flavin-binding are highlighted in yellow. LuxA, luciferase α-subunit from *Vibrio harveyi*; MelG-MOx, luciferase-like monooxygenase domain from the melithiazol pathway.

In-frame deletion of *orf3515*

With the kind help of Dr. Katharina Dornblut, the in-frame deletion of *orf3515* was successfully carried out. Encouragingly, phthoxazolin production was abolished in Δ*orf3515* (Figure 5.10). Gene complementation of the mutant is currently underway to confirm its involvement in the phthoxazolin pathway. Based on this provisional result, the putative flavin-dependent monooxygenase, Orf3515, appears to be a highly attractive candidate for catalysing the α-hydroxylation of the Gly-extended polyketide-peptide chain(s) on PhoxR to directly yield phthoxazolin A via non-enzymatic cleavage cleavage of the resulting aminal.

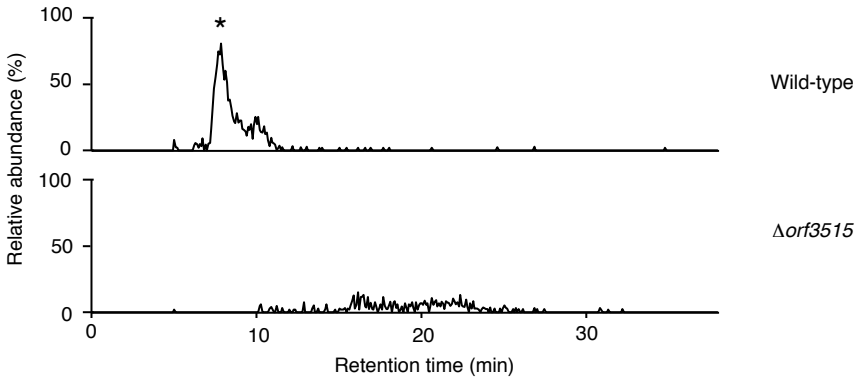


Figure 5.10. Phthoxazolin A production in *Streptomyces* sp. KO-7888 wild-type and *orf3515* mutant. Selected ion monitoring of *m/z* $[M+Na]^+$ 313.15 is shown. Peaks corresponding to phthoxazolin A is marked with an asterisk; the shift in retention time was due to the change of separation column in the LC-MS.

5.2.4 Proposed mechanism of primary amide formation in phthoxazolin biosynthesis

Similar to the melithiazol and myxothiazol pathways (Weinig *et al*., 2003; Müller *et al*., 2006), we propose that the Gly-extended polyketide-peptide(s) on the hybrid PKS-NRPS PhoxR are precursors to phthoxazolins. Orf3515, a flavin-dependent, luciferase-like monooxygenase coded by cluster 13, potentially catalyses α-hydroxylation of Gly, and subsequent oxidative cleavage directly yields the primary amide phthoxazolin A (Figure 5.11). The first AT domain in PhoxS, belonging to the acyl hydrolase clade of trans-ATs (Jensen *et al*., 2012) (as discussed in Section 3.2.5), may be responsible for the removal of the remnant chain(s) off the carrier proteins of PhoxR.

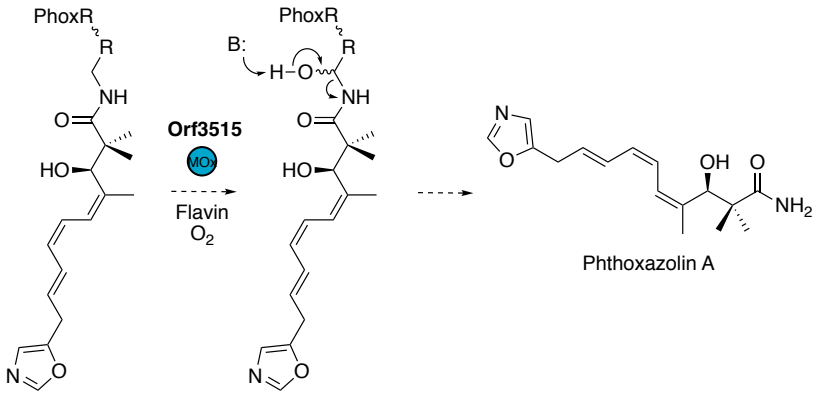


Figure 5.11. Proposed mechanism of chain release and primary amide formation from Gly-extended polyketide-peptide(s). MOx, monooxygenase.

5.3 Concluding remarks and future work

Orf3515, a putative luciferase-like monooxygenase encoded in cluster 13, was identified as a strong candidate for the catalysis of chain release and primary amide formation. Orf3515 was proposed as a flavin-dependent monooxygenase to α-hydroxylate the Gly-extended polyketide-peptide chain(s) on PhoxR, yielding phthoxazolin A via oxidative cleavage. Gene complementation of the *orf3515* in-frame deletion mutant is currently underway to confirm its involvement in the phthoxazolin pathway. If confirmed, this would be another intriguing example of a remotely located gene playing an essential role in another biosynthetic pathway (Lazos *et al*., 2010; Li *et al*., 2018). For future work, *in vitro* assays could be carried out using purified Orf3515 and the proposed enzyme-bound (or the mimic of) Gly-extended intermediates, to reconstitute the catalytic activity of the putative flavin-dependent monooxygenase in chain release and primary amide formation and to elucidate its substrate preference for oxidative cleavage.

Chapter 6

Final remarks on phthoxazolin biosynthesis

This short chapter summarises the findings on the characterisation of the *trans*-AT PKS-NRPS pathway to phthoxazolins, as described in Chapters 3, 4 and 5.

The phthoxazolin biosynthetic gene cluster

The unusually large *phox* clusters spanning 80.7 kbp were identified from the whole-genome sequence of the phthoxazolin producers, *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888. Bioinformatic analysis revealed the presence of more PKS/NRPS modules and domains than required for the biosynthesis of phthoxazolin (see Sections 3.2.2 and 3.2.3).

This prompted the examination of the functional roles of genes and domains not thought to be involved in phthoxazolin biosynthesis, including *phoxG*, *phoxI*, *phoxJ*, *phoxR* (partial), *phoxU* and *phoxV*, via in-frame deletion experiments (see Section 3.2.7). The deletion of these genes, with the exception of *phoxR*, did not affect phthoxazolin production and the reason behind their presence in the cluster remains unclear. The *phox* cluster was also compared to the recently published cryptic phthoxazolin (*ptx*) cluster from *S. avermetilis* KA-320 (Suroto *et al*., 2017) (Figure 6.1). A similar gene organisation was observed: *ptxHFEDCBA* were identified as the counterparts of *phoxINOPQRS* and *ptxG* potentially encodes an NRPS fused with a C-terminal cytochrome P450 enzymatic domain, which corresponds to *phoxJK*. Counterparts of *phoxLM* and genes flanking the upstream region of *phoxI* and downstream region of *phoxS* were not identified in *S. avermetilis* KA-320, and are unlikely to be involved in phthoxazolin biosynthesis.

Inactivation of individual acyl and peptidyl carrier protein domains in PhoxR was carried out and all four carrier proteins were shown to be essential for

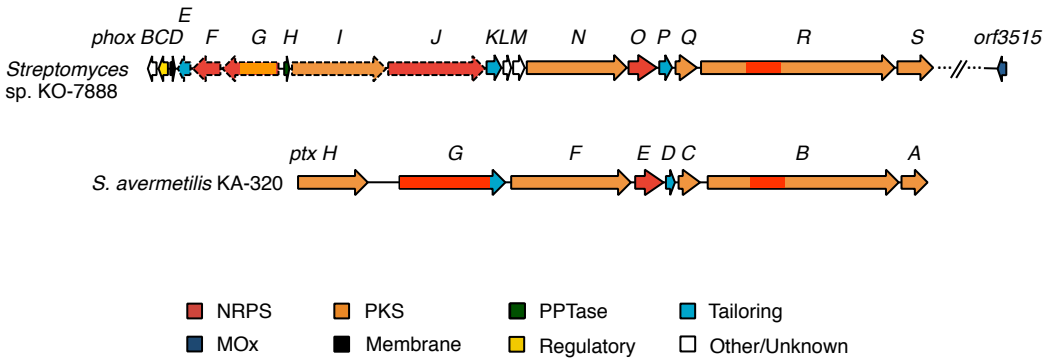


Figure 6.1. Comparison of the gene organisation of the phthoxazolin clusters in *Streptomyces* sp. KO-7888 (*phox*) to that in *S. avermetilis* KA-320 (*ptx*). *Phox* genes demonstrated to be not required for phthoxazolin biosynthesis are indicated in dashed lines.

biosynthesis. This provided first evidence that the polyketide-peptide chain may be extended beyond the phthoxazolin portion prior to release. This remains to be experimentally confirmed, perhaps by using substrate mimics *in vivo* to intercept, offload and detect intermediates from these modules (Riva et al., 2014; Ho et al., 2017). It has clear implications for the mechanism of chain release (see below).

Proposed mechanism for 5-oxazole formation

PhoxOPQ were proposed as the candidate enzymes for 5-oxazole formation, based on sequence homology with their counterparts in the oxazolomycin and conglobatin pathways, OzmOPQ and CongAEB. *PhoxP*, encoding a hypothetical protein, was deleted in-frame which led to the loss of phthoxazolin production. The restoration of production by the complementation of *phoxP in-trans* confirmed that PhoxP plays an essential role in phthoxazolin biosynthesis (see Section 4.2.4). Although the attempt to reconstitute the catalytic activity of PhoxP alongside other Phox proteins *in vitro* was unsuccessful (see Section 4.2.5), PhoxP is proposed to act as an ATP-dependent cyclodehydrase on a formylglycyl intermediate bound to one of the first modules in the pathway. PhoxP, OzmP and CongE together appear to represent a distinct clade of enzymes catalysing this novel function within the α-adenine nucleotide hydrolase superfamily.

Proposed mechanism for chain release and primary amide formation

The mechanism of chain release and primary amide formation were also investigated via the in-frame deletion of various candidate genes (see Sections 5.2.1 to 5.2.3).

The α-ß hydrolase PhoxE, the terminal reductase (TD) domain in PhoxF and the C-terminal condensation (C) domain in PhoxG were initially considered as candidates for chain release, however, each were shown in turn not to be required for phthoxazolin biosynthesis. Also, an asparagine synthetase-encoding gene, *orf0034*, was identified distant from the phox cluster and was considered as a candidate for catalysis of the formation of the primary amide moiety of phthoxazolins starting from the corresponding carboxylic acid, as seen with several polyketides and non-ribosomal peptides. However, the in-frame deletion of *orf0034* did not affect the production of phthoxazolins.

In light of the recent discovery of two novel lipopeptides, both containing a primary amide moiety, produced by the phthoxazolin producer *Streptomyces* sp. KO-7888 (Inahashi *et al*., manuscript in preparation), the corresponding NRPS cluster (cluster 13) that governs their biosynthesis (see Section 2.2.4) was re-examined. *Orf3515*, a luciferase-like monooxygenase-encoding gene was identified within the cluster, and it was subsequently deleted in-frame. The loss of phthoxazolin production was observed (see Section 5.2.3) and further work on the *in-trans* gene complementation of the mutant is currently underway to confirm its role in the phthoxazolin biosynthetic pathway.

As the inactivation of individual carrier proteins in PhoxR by point mutagenesis resulted in the loss of phthoxazolin production (see Section 3.2.7), it is tempting to suggest that polyketide-peptide chain(s) longer than phthoxazolin may be biosynthesised prior to cleavage by Orf3515. As no phthoxazolin derivative containing additional extender units was detected, the cleavage possibly takes place on the enzyme-bound intermediate(s). Chain release in PKS and NRPS pathways by truncation is rare but not unprecedented. The myxothiazol and melithiazol pathways utilise an integrated flavin-dependent monooxygenase domain in the last module to hydroxylate the Gly-extended polyketide-peptide at the α-position, leading to the oxidative release of myxothiazol and a melithiazol precursor, respectively, both of which bear a primary amide moiety (Weinig *et al*., 2003). Zwittermicin A is also proposed to be liberated in a similar fashion from an Ala-extended chain (Kevany *et al*., 2009), in this case by a standalone monooxygenase. A further documented example of α-hydroxylation of a Gly residue in an NRPS is provided by the biosynthesis of the cyclodepsipeptide skyllamicin A in *Streptomyces* sp. Acta 2897 (Pohle *et al*., 2011). The product of the gene *sky39* is a flavin-dependent monooxygenase with 51% sequence identity to the MOx domain of MtaG, and the adjacent *sky40* encodes a putative flavin mononucleotide reductase that would

regenerate FMN for the monooxygenase. In this case, the resulting aminal appears to be stable to hydrolysis, and it is suggested that cleavage is prevented by the cyclic nature of the peptide (Pohle *et al*., 2011).

Therefore, an analogous mechanism is proposed here for the phthoxazolin pathway: the remotely-encoded monooxygenase Orf3515 catalyses in-*trans* the α-hydroxylation of the Gly residue in the extended polyketide-peptide(s), and the subsequent spontaneous cleavage of the N-C bond yields phthoxazolin A directly. If this mechanism is confirmed, by showing restoration of phthoxazolin production by *in-trans* complementation of the *orf3515* mutant with a wild type copy, this would provide a further example of an remarkable emerging feature of certain biosynthetic pathways, in which a remotely-located gene is required for the function of a biosynthetic gene cluster. Examples of this include the requirement for a remote acetyltransferase gene for the production of the iron-chelating peptide erythrochelin in *S. erythraea* (Lazos *et al*., 2010) and the requirement for the remotely-encoded methyltransferase GenL for the terminal methyl transfer step in the formation of the gentamicin C complex in *Micromonospora echinospora* (Li *et al*., 2018).

Figures 6.1 and 6.2 show the gene organisation and the proposed biosynthetic pathway to phthoxazolins, respectively, revised according to this work. The hydroxylation by PhoxK, a putative cytochrome P450 enzyme, is tentatively proposed to be the terminal step in biosynthesis, but it cannot be excluded that hydroxylation takes place on the growing chain tethered to the assembly-line.

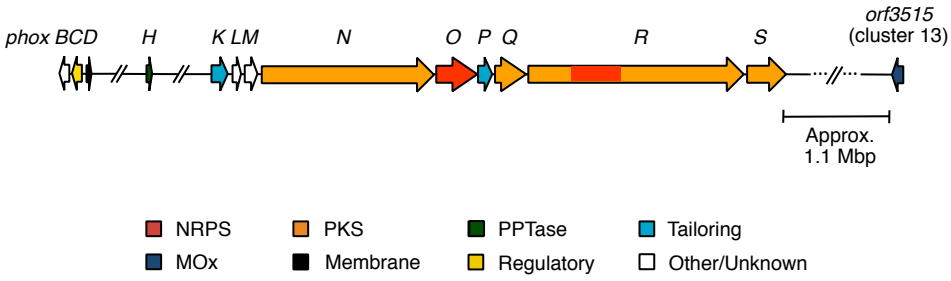


Figure 6.2. Gene organisation of the phthoxazolin cluster (revised). MOx, monooxygenase; PPTase, 4'-phosphopantetheinyl transferase.

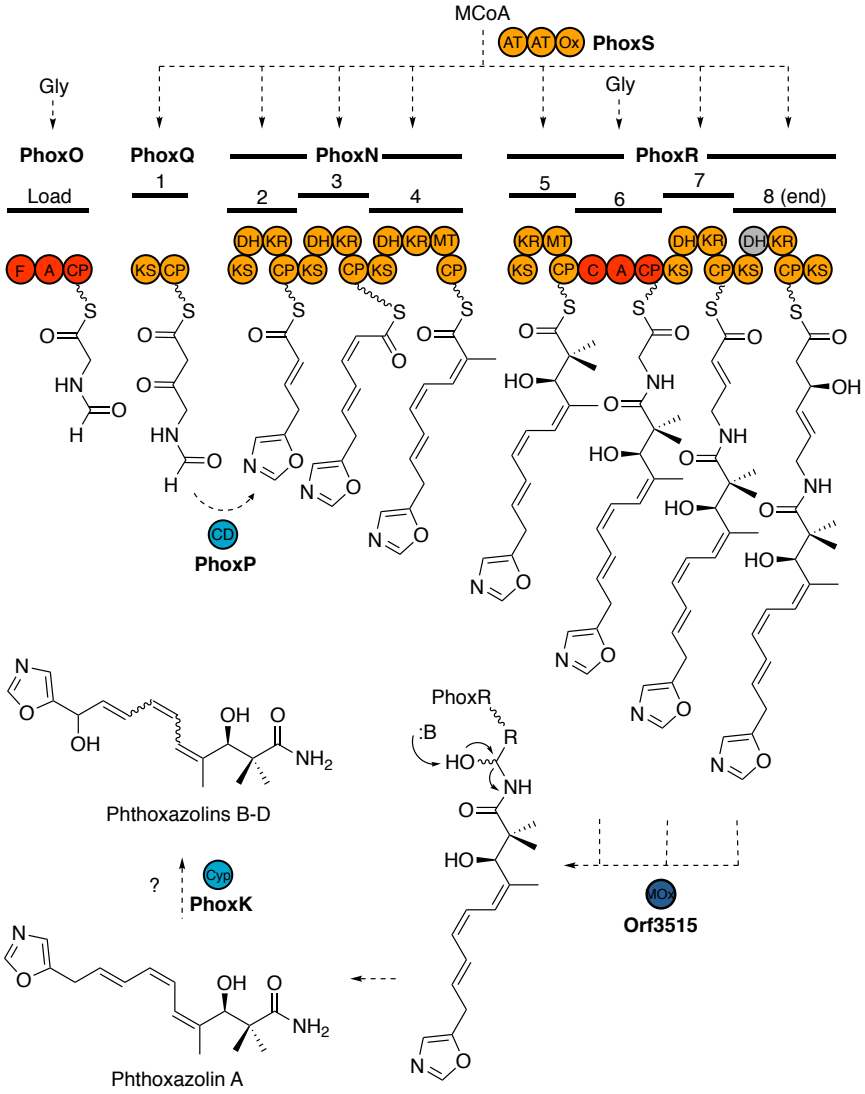


Figure 6.3. Proposed biosynthetic pathway to phthoxazolins (revised). CD, cyclodehydratase; MOx, monooxygenase; Cyp, cytochrome P450.

Chapter 7

Enzymatic logic of the biosynthesis of the bis(benzoxazole) antibiotic AJI9561

7.1 Introduction

7.1.1 Benzoxazole-containing natural products

Benzoxazoles are bicyclic heterocycles consisting of a benzene-fused oxazole moiety. Benzoxazole-containing natural products exhibit antibacterial, antiviral, antifungal, herbicidal, insecticidal and anti-proliferative activities (Figure 7.1). The aromatic and lipophilic planar structure of the benzoxazole moiety and the 1-oxygen and 3-nitrogen atoms potentially acting as hydrogen bond acceptors contribute to the molecular interactions with biological targets (Lin *et al*., 2004; Demmer and Bunch, 2015). Also, substituted benzoxazoles are the isosters of nucleotides and may interact with nucleic acids and/or other structures comprising nucleotides.

Streptomyces spp. produce many of the benzoxazole antibiotics including calcimycin, cezomycin, UK-1 and AJI9561. Due to their potent bioactivities, benzoxazole antibiotics attracted much interest in the fields of medicinal and synthetic chemistry (Kumar *et al*., 2002; McKee and Kerwin, 2008). However, the biosynthetic routes to this structurally unique class of natural products were rather underexplored until recently. Only the calcimycin cluster was reported prior to start of the work on AJI9561 in this laboratory (Wu *et al*., 2011).

7.1.2 AJI9561 biosynthetic gene cluster

AJI9561 is a bis(benzoxazole) antibiotic isolated from the soil bacterium *Streptomyces* sp. AJ9561 (Sato *et al*., 2001). It possesses antibacterial activity against *S. lividans* JT46, *S. albus* J1074 and *Micrococcus luteus*, and cytotoxic activity against human and murine leukaemia cell lines. Its mode of action has not

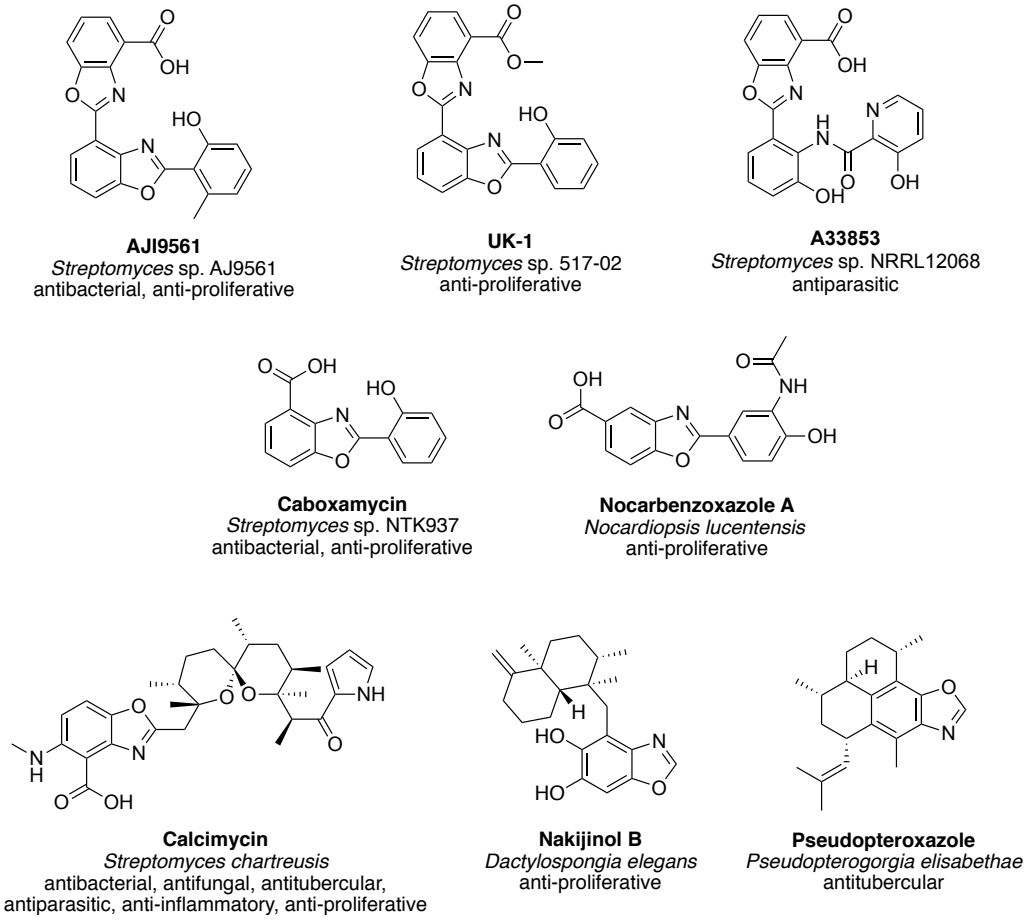


Figure 7.1. Structure, microbial origin and bioactivities of benzoxazole-containing natural products.

been elucidated but it is believed to share the same mechanism of metal ion-mediated DNA binding, as with the structurally similar UK-1 and nataxazole (Deluca and Kerwin, 1997).

The AJI9561 producer strain, *Streptomyces* sp. AJ9561, was sequenced and assembled in-house by the DNA Sequencing Facility (Department of Biochemistry). From the whole-genome sequence, the AJI9561 (*aji*) cluster was identified and annotated by Dr. Qiulin Wu (Figure 7.2 and Table 7.1). The *aji* cluster spans 20.0 kbp and comprises of 12 ORFs putatively encoding a major facilitator superfamily (MFS) transporter, a TetR family transcriptional regulator, a 6-methylsalicylic acid (6-MSA) synthase (6-MSAS), four enzymes responsible for 3-hydroxyanthranilic acid (3-HAA) biosynthesis, a 4'-phosphopantetheinyl transferase (PPTase), two ATP-dependent ligases, an acyl carrier protein (ACP), a 3-oxoacyl-ACP synthase and an amidohydrolase.

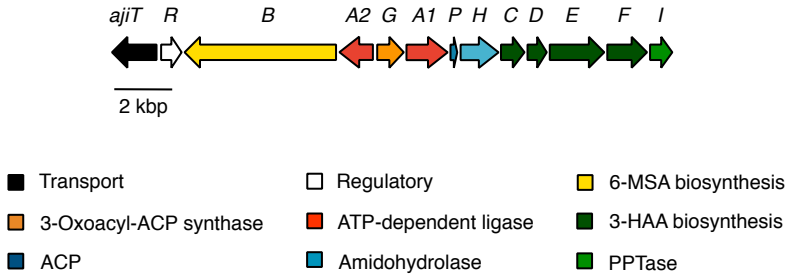


Figure 7.2. Gene organisation of the *aji* cluster in *Streptomyces* sp. AJ9561.

Table 7.1. Properties of genes within the *aji* cluster in *Streptomyces* sp. AJ9561. MFS, major facilitator superfamily; 6-MSAS, 6-methylsalicylic acid synthase; DHB, dihydroxybenzoate; ADIC, 2-amino-2-desoxyisochorismate; DAHP, 3-deoxy-D-arabino-heptulosonic acid 7-phosphate.

Gene	Size (aa)	Putative function	Species	Identity/ similarity (%)	Genbank accession
ajiT	516	MFS transporter	*Streptomyces* sp. SAT1	84/91	ANH93872.1
ajiR	235	TetR family transcriptional regulator	*Streptomyces* spp.	86/91	WP_018564636.1
ajiB	1751	6-MSAS	*Streptomyces* sp. Tu 6176	83/88	CEK42827.1
ajiA1	436	ATP-dependent ligase	*Streptomyces* spp.	90/94	WP_018564638.1
ajiG	347	3-oxoacyl ACP synthase	*Streptomyces* spp.	93/95	WP_018564639.1
ajiA2	*540*	ATP-dependent ligase	*Streptomyces* sp. PsTaAH-124	85/92	WP_018564640.1
ajiP	*88*	ACP	*Streptomyces* sp. PTY087I2	71/88	WP_079137835.1
ajiH	494	Amidohydrolase	*Streptomyces* sp. SAT1	87/91	WP_064535659.1
ajiC	271	2,3-dihydro-2,3-DHB dehydrogenase	*Streptomyces* sp. PTY087I2	82/88	WP_065485543.1
ajiD	*222*	Isochorismatase	*Streptomyces* sp. SAT1	80/85	WP_079159103.1
ajiE	628	ADIC synthase	*Streptomyces* sp. Tu 6176	86/91	CEK42818.1
ajiF	495	DAHP synthase	*Streptomyces* sp. SAT1	81/86	WP_107440699.1
ajiI	258	PPTase	*Streptomyces* sp. PTY087I2	55/63	WP_065485318.1

In both ATP-dependent ligases, AjiA1 and AjiA2, an adenylate forming domain containing an AMP binding site was identified. The family of adenylate forming enzymes, including the A domain in NRPSs, acyl-CoA ligases and aryl-CoA ligases, catalyses the ATP-dependent activation of a carboxylate substrate and the transfer of the resulting adenylate to the pantetheinyl group of either CoA or an/a acyl/peptidyl carrier protein. AjiA1 and AjiA2 showed high sequence similarities to coenzyme F390 synthetases and 2,3-dihydroxybenzoate-AMP ligases, respectively.

AjiG belongs to the initiating condensing enzyme subclass of the decarboxylating/condensing enzyme family. This family of enzymes are involved in the biosynthesis of fatty acids, phospholipids and polyketide natural products. AjiG shows a high sequence similarity to 3-oxoacyl-ACP and a dimer interface was detected using protein-BLAST. It possesses a Cys-His-His catalytic triad, which is characteristic of condensaiton enzymes that act on ACP thioesters (as opposed to CoA substrates) exclusively (Heath and Rock, 2002).

AjiP contains a putative phosphopantetheine-binding domain and shows high sequence similarity to other carrier proteins in PKS and NRPS pathways.

AjiH belongs to the metallo-amidohydrolase superfamily. Metallo-amidohydrolases are evolutionarily divergent and functionally versatile. They catalyse the hydrolysis of C-N, C-C, C-O, C-Cl, C-S and O-P bonds as well as non-hydrolytic reactions including decarboxylation, hydration and isomerisation (Holm and Sander, 1997; Seibert and Raushel, 2005). Moreover, dihydroorotases catalyse the cyclisation of *N*-carbamoyl-L-Asp to L-dihydroorotate in pyrimidine biosynthesis (Figure 7.3). Some members of the cyclic amidohydrolase family are also capable of catalysing cyclisation reactions as the reverse reaction.

The putative PPTase, AjiI is likely responsible for the conversion of *apo*-AjiB and *apo*-AjiP to *holo*-AjiB and *holo*-AjiP, respectively.

Biosynthesis of 6-methylsalicylic acid (6-MSA)

One of the precursors to AJI9561, 6-MSA, is predicted to be produced by AjiB, annotated as a 6-MSA synthase (6-MSAS). The type I iterative PKS AjiB comprises of five enzymatic domains including a ketosynthase (KS) domain, an acyltransferase (AT) domain, a thioester hydrolase (THID) domain, a ketoreductase (KR) domain and an ACP domain. With the use of an acetyl-CoA starter unit and three malonyl-CoA extender units, 6-MSAS catalyses the formation of a 5-hydroxytetraketide which undergoes intramolecular aldol condensation, dehydration, aromatisation and hydrolytic cleavage to yield 6-MSA (Parascandolo *et al*., 2016) (Figure 7.3).

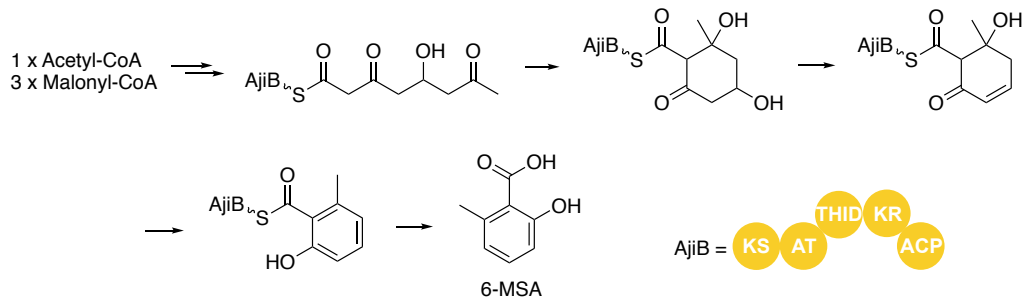


Figure 7.3. Biosynthesis of 6-MSA catalysed by 6-MSAS.

Biosynthesis of 3-hydroxyanthranilic acid (3-HAA)

AjiCDEF form a subcluster in the *aji* cluster and together encode enzymes potentially involved for the biosynthesis of the other precursor 3-HAA (Figure 7.4). AjiF is a 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthetase. DAHP synthetase catalyses the first step of the shikimate pathway for chorismate biosynthesis, the condensation of phosphoenolpyruvate (PEP) and D-erythrose-4-phosphate (E4P) to give rise to DAHP. Chorismate from the shikimate pathway is then converted to 2-amino-4-deoxychorismic acid (ADIC) by ADIC synthase AjiE via the nucleophilic attack of the side chain nitrogen of glutamine on C2 with the elimination of the C4 hydroxyl group. The isochorismatase AjiD cleaves the enoylpyruvate side chain of ADIC to yield *trans*-2,3-dihydro-3-hydroxyanthranilic acid (DHHA). AjiC, a short-chain oxidoreductase, catalyses an oxidation reaction, aromatising DHHA to 3-HAA (McDonald *et al.*, 2001).

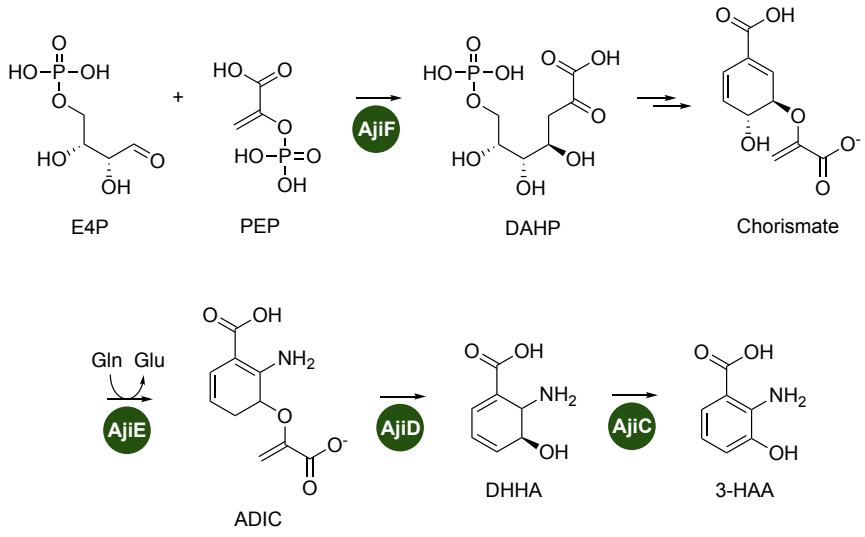


Figure 7.4. Biosynthesis of 3-HAA catalysed by AjiCDEF.

7.1.3 *In vitro* reconstitution of AJI9561 biosynthesis

Dr. Qiulin Wu and Joachim Hug purified Aji enzymes and reconstituted the AJI9561 biosynthetic pathway *in vitro*, allowing the functional role of individual enzymes to be established experimentally.

Activation of 3-HAA and 6-MSA

The precursors 3-HAA and 6-MSA were incubated individually with the ATP-dependent ligases, AjiA1 and AjiA2. In the presence of ATP and $MgCl_2$, the incubation of 6-MSA with AjiA2 enabled the activation and subsequent loading of 6-MSA to the ACP AjiP (Figure 7.5), whereas the addition of AjiA1 to 3-HAA lead to the complete conversion of 3-HAA to a 3-HAA dimer **(1)** (Figure. 7.6). The ATP-dependent ligases are substrate specific and did not act on the other precursor.

As the 3-HAA dimer **(1)** could either be amide-linked or ester-linked, a ^{15}N-labelled 3-HAA dimer was enzymatically synthesised from chorismate and ^{15}N-glutamine, using CalB1, CalB2, CalB3 (homologues of AjiE, AjiD and AjiC in the calcimycin pathway) and AjiA1 for structural elucidation. ^{15}N and ^{1}H NMR analysis of the ^{15}N-labelled 3-HAA dimer confirmed that the linkage is via amide bond.

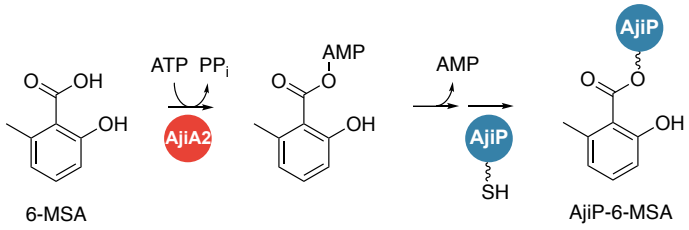


Figure 7.5. ATP-dependent activation and loading of 6-MSA.

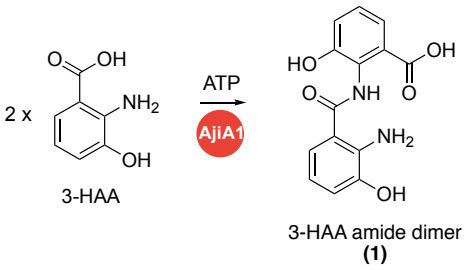


Figure 7.6. ATP-dependent activation and dimerisation of 3-HAA.

Oxazole ring closure catalysed by the amidohydrolase-like AjiH

The functional role of AjiH was first established via a one-pot reaction containing 3-HAA, AjiA1, ATP, $MgCl_2$ and the amidohydrolase-like AjiH (Figure 7.7). 3-HAA was converted efficiently to the 3-HAA oxazole dimer **(2)** containing the benzoxazole moiety and AjiH was named an "amidocyclase".

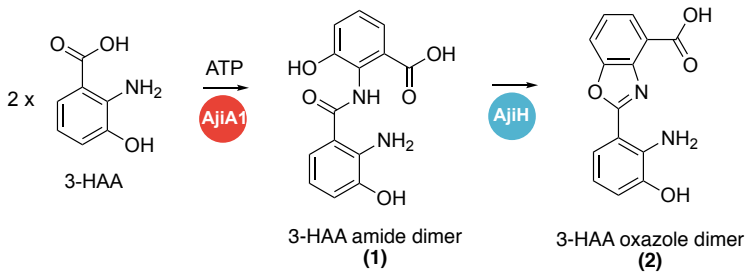


Figure 7.7. Proposed biosynthetic route to the 3-HAA oxazole dimer (2) catalysed by AjiA1 and AjiH.

Formation of 3-HAA oxazole dimer-6-MSA trimer and AJI9561

The 3-HAA oxazole dimer **(2)** was isolated and used as a substrate for downstream reactions. AjiG, the 3-oxoacyl-ACP synthase, catalysed the condensation between AjiP-6-MSA and the isolated 3-HAA oxazole dimer **(2)** to afford the 3-HAA oxazole dimer-6-MSA trimer **(3)**. Step-wise addition of amidocyclase AjiH to the same enzymatic reaction lead to the cyclisation of the second oxazole ring to form the final product, the bis(benzoxazole) AJI9561 (Figure 7.8). Both reactions have a conversion rate of less than 5%. AJI9561 was also detected in a one-pot reaction containing 6-MSA, 3-HAA and all five Aji enzymes, though the efficiency of this reaction is also low.

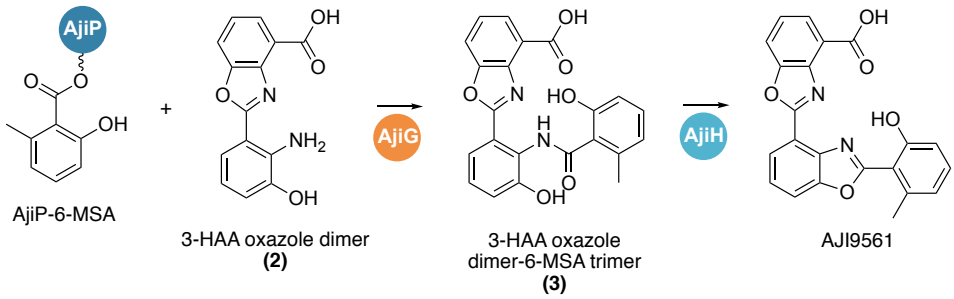


Figure 7.8. Proposed biosynthetic route to the 3-HAA oxazole dimer-6-MSA trimer (3) and AJI9561 catalysed by AjiG and AjiH.

7.1.4 Proposed biosynthetic pathway to AJI9561

The proposal for the biosynthetic pathway to AJI9561 was based on bioinformatic analysis, the *in vitro* reconstitution of AJI9561 biosynthesis using purified Aji enzymes and structural elucidation of isolated pathway intermediates (Figure 7.9). It is unclear whether or not AJI9561 could arise from alternative intermediates, such as the 3-HAA-6-MSA heterodimer and the 3-HAA dimer-6-MSA trimer.

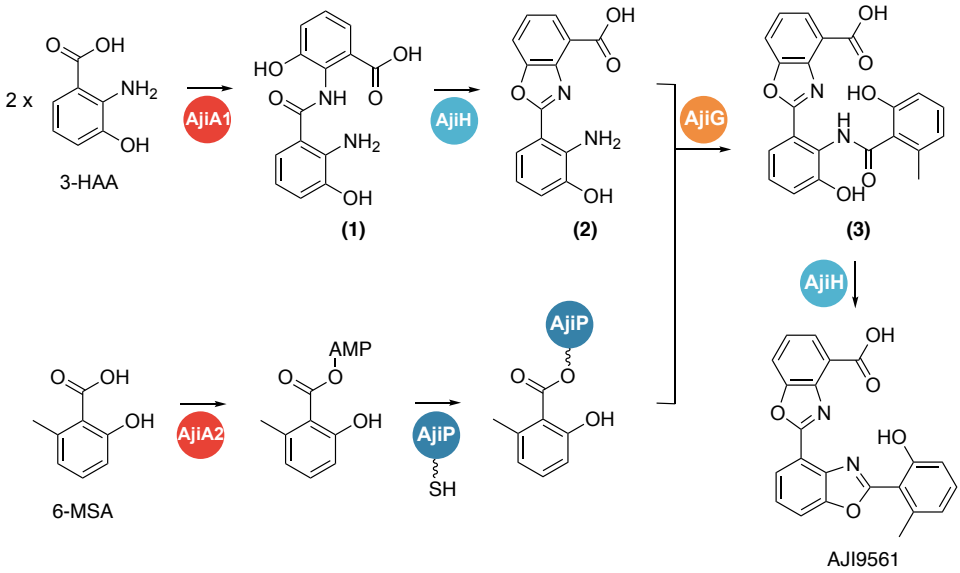


Figure 7.9. Proposed biosynthetic pathway to AJI9561.

7.1.5 Comparative sequence analysis of clusters governing AJI9561 and other (bis)benzoxazole natural products

During the course of this project, the genetic studies of biosynthetic pathways governing three other (bis)benzoxazole natural products, nataxazole, A33853 and caboxamycin, were published by two independent groups (Cano-Prieto *et al.*, 2015; Lv *et al.*, 2015; Losada *et al.*, 2017). Each gene in the AJI9561 biosynthetic gene cluster was aligned against these three clusters and several homologues were identified (Figure 7.10).

Nataxazole is a methyl ester derivative of AJI9561, produced by *Streptomyces* sp. Tü 6176. Like AJI9561, the bis(benzoxazole) nataxazole is also assembled from two units of 3-HAA and a unit of 6-MSA. NatL2/S/L1/AC1/AM were identified as the homologues of AjiA1/G/A2/P/H and show high levels of sequence

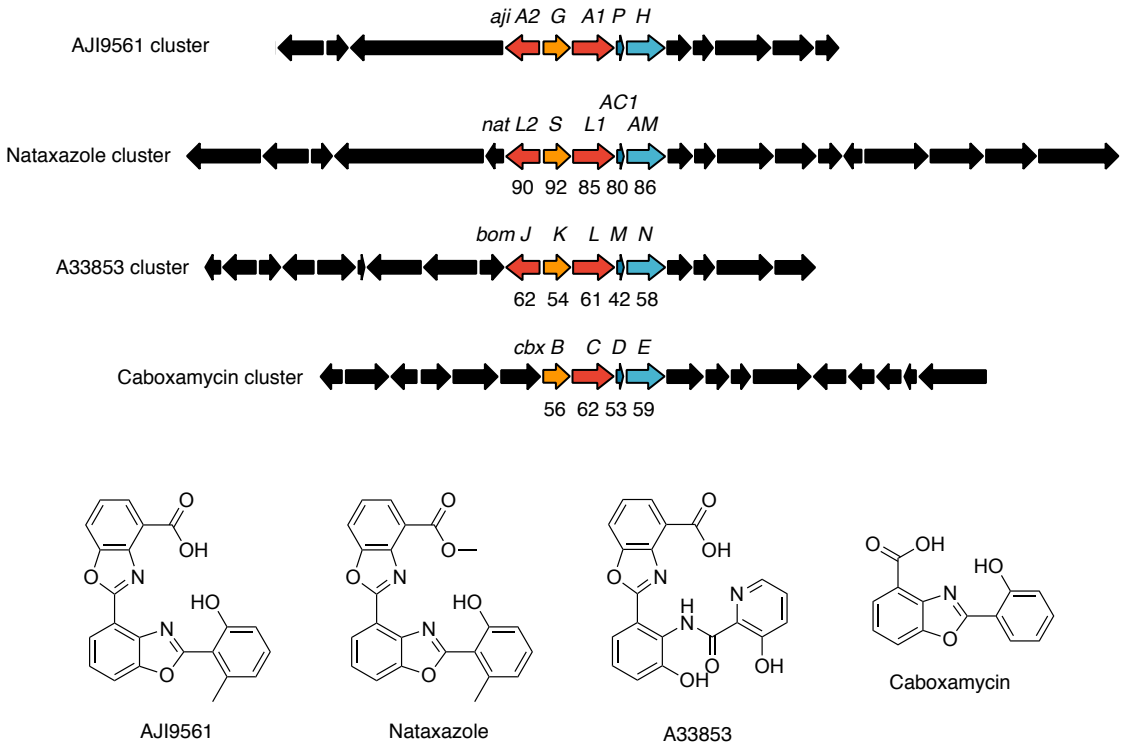


Figure 7.10. Comparative sequence analysis of clusters governing (bis)benzoxazole natural products.

identity (80-92%). Moreover, gene inactivation of *natR3*, encoding a TetR family transcriptional regulator, lead to an increase of nataxazole production (Cano-Prieto *et al*., 2015). As *ajiR* is highly similar to *natR3*, the biosynthesis of AJI9561 may be regulated in a similar fashion.

A33853 is composed of two units of 3-HAA and one unit of 3-hydroxypicolinic acid (3-HPA), and contains one benzoxazole moiety formed between the two 3-HAA units. The gene cluster responsible for A33853 biosynthesis was identified in *Streptomyces* sp. NRRL 12068 (Lv *et al*., 2015). In addition to 3-HAA biosynthetic genes, genes responsible for 3-HPA biosynthesis are present. The AjiA1/G/A2/P/H homologues in this pathway, BomJKLMN, display lower levels of sequence identity (42-62%) to the AJI9561 equivalents than do those in the nataxazole pathway. The amide between 3-HAA and 3-HPA is not cyclised, potentially due to the inability of BomN to accommodate 3-HPA. The bomN deletion mutant produced the non-cyclised 3-HAA dimer and 3HAA dimer-3HPA trimer, which is indicative of the role of BomN in oxazole ring formation and in accordance with our findings regarding AjiH *in vitro*.

Caboxamycin is produced by *Streptomyces* sp. NTK937 and its scaffold is formed from one unit of each of 3-HAA and salicylic acid (SA). Unlike 6-MSA, SA is not biosynthesised by a PKS but by a salicylate synthase which belongs to the superfamily of chorismate binding enzymes. The genes responsible for assembly and benzoxazole formation are also found in this gene cluster, with the homologue of AjiA1 missing as 3-HAA activation is not required in this pathway. It was shown that caboxamycin biosynthesis is positively regulated by CbxR, a SARP family transcriptional regulator (Losada *et al*., 2017).

Comparative genome analysis revealed that homologues of the set of genes involved in benzoxazole formation in the AJI9561 pathway were found in all three other biosynthetic gene clusters, suggesting a common biosynthetic route is shared among members of this unique class of (bis)benzoxazole natural products. A second copy of ATP-depedent ligase responsible for the activation and dimerisation of 3-HAA may or may not be present, depending on whether the final product is assembled from three or two precursor units.

This chapter begins with a detailed study of the amidocyclase AjiH via phylogenetic analysis and *in silico* structural modelling, to explore the relationship between AjiH and other members of the metallo-amidohydrolase superfamily and to gain insight into its mechanism of oxazole ring cyclisation. Attempts to reconstitute the single cyclisation reaction on the isolated 3-HAA amide dimer **(2)** catalysed by AjiH, and to reconstitute AJI9561 biosynthesis using the isolated 3-HAA amide dimer **(2)** (instead of 3-HAA) are also described. Moreover, the investigation of alternative routes to AJI9561 biosynthesis via the formation of 3-HAA-6-MSA heterodimers or 3-HAA dimer-6-MSA trimer is presented.

7.2 Results and discussion

7.2.1 Phylogenetic analysis of AjiH

As the metallo-amidohydrolase superfamily comprises a large set of evolutionarily-related but highly divergent enzymes catalysing many different reactions, the Sturcture-Function Linkage Database (SFLD) was used to investigate the amidocyclase AjiH by comparison to/against other members of the superfamily (Akiva *et al*., 2014). SFLD utilises manually-curated multiple sequence alignment of protein superfamilies and their corresponding Hidden Markov model (HMM) to generate sequence alignments. The metal-binding His and Asp residues essential for the superfamily's function were all identified in AjiH (Figure 7.11). Moreover, AjiH was shown to be the most closely related to the guanine deaminase, formimidoylglutamate deiminase and imidazolonepropionase protein families.

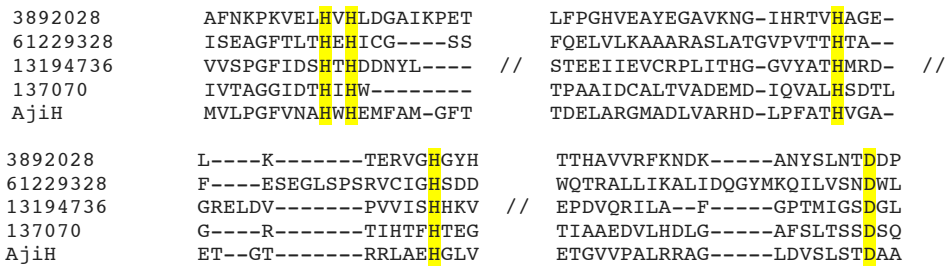


Figure 7.11. Partial alignment of AjiH and members of the metallo-amidohydrolase superfamily. The highly conserved metal-binding residues discussed in the text are highlighted in yellow. NCBI GI numbers of the enzymes are shown. The alignment was generated using the Structure-Function Linkage Database (SFLD) (Akiva *et al*., 2014).

A phylogenetic tree comprising selected members of the superfamily including AjiH and its counterparts in the nataxazole (*nat*), A33853 (*bom*) and carboxamycin (*cbx*) pathways, was generated using MEGA7 (Kumar *et al*., 2016) (Figure 7.12). The amidocyclases involved in benzoxazole formation form a distinct clade and certainly represent a novel catalytic function in the metallo-amidohydrolase superfamily.

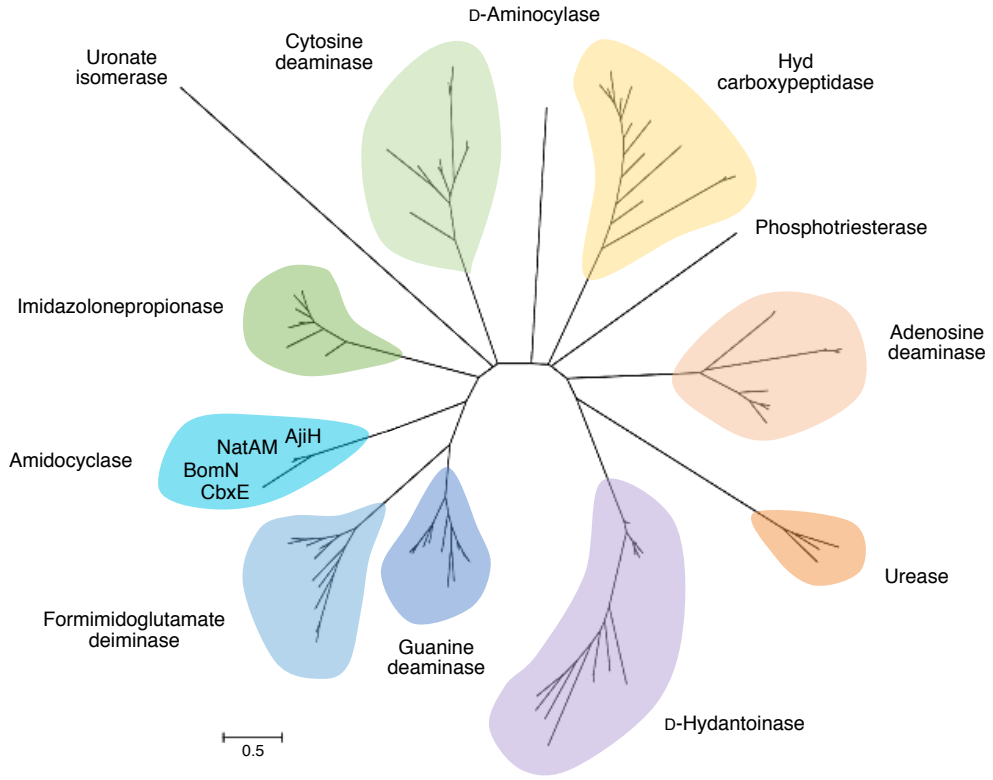


Figure 7.12. Phylogenetic tree of members belonging to the metallo-amidohydrolase superfamily. The evolutionary history was inferred using the maximum likelihood method (Jones *et al*., 1992). All positions containing gaps and missing data were eliminated. There were a total of 159 amino acids in the final dataset. The evolutionary analysis was conducted in MEGA7 (Kumar *et al*., 2016).

7.2.2 *In silico* structural modelling of AjiH

The structure of AjiH was modelled *in silico* using Phyre2 (Kelley *et al*., 2015). Multiple crystal structures of metallo-amidohydrolases were used as templates (Table 7.2).

Table 7.2. Alignment coverage, identify and confidence of templates used for the *in silico* structural modelling of AjiH.

PDB ID	Function	% Alignment coverage	% Identity	% Confidence
4V1X	Atrazine chlorohydrolase	97	22	100
4DYK	Adenosine deaminase	95	24	100
4DZH	Adenosine deaminase	95	22	100
4F0R	Adenosine deaminase	95	22	100
3LNP	Adenosine deaminase	95	21	100
3LSB	Atrazine chlorohydrolase	93	25	100

Although the amino acid sequence similarity is low within the superfamily, metallo-amidohydrolases shows a common overall three-dimensional structure, which was also observed for the structural model of AjiH. AjiH has a TIM barrel consisting eight alternating ß-sheets and α-helices, with the α-helices wrapped around the ß-barrel, and a smaller domain made of the N-terminus and C-terminus of the protein (Figure 7.13). Both domains are characteristic of the superfamily.

Using RaptorX binding site prediction (Källberg *et al*., 2012), a putative active site pocket comprising His75, His77 Phe80, Asp108, Ile112, Thr115, Trp178, Ser226, His253, Ala256, Lys315, Asp345 and Leu349 was identified from the middle four ß-sheets in the TIM barrel (Figure 7.14). AjiH likely constitutes a mononuclear divalent metal centre coordinated by the His and Asp residues in the active site, as the Lys or Glu residues required in the coordination of binuclear and multinuclear centres are absent. Above the metal binding site, a cavity potentially for substrate binding was observed. The aromatic residues, Phe80 and Trp178, on either side of the pocket may play a role in stabilising the substrates via hydrophobic interactions.

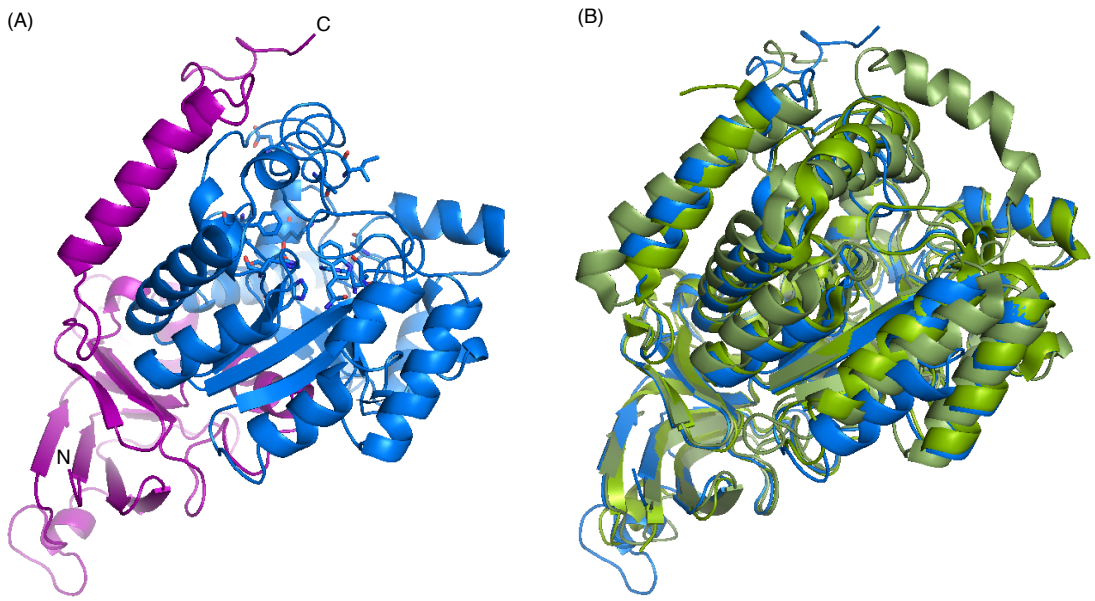


Figure 7.13. Structural model of AjiH. (A) The TIM barrel (blue) and the smaller domain (purple) of AjiH. The N-terminus and C-terminus are indicated. Resides in the predicted active site pocket are shown as sticks. (B) AjiH (blue) overlaid with AtzA (PDB ID: 4V1X; green), an atrazine chlorohydrolase from *Pseudomonas* sp. DSM 11735, and MtaD (PDB ID: 4F0R; light green), an adenosine deaminase from *Chromobacterium violaceum* ATCC12472.

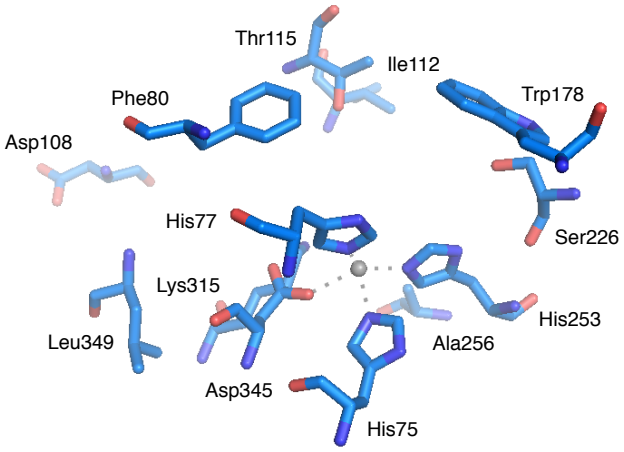


Figure 7.14. Predicted active site pocket of AjiH. The metal ion is shown as a grey sphere.

7.2.3 Attempted *in vitro* reconstitution of the single cyclisation reaction catalysed by AjiH

To confirm that the amidocyclase AjiH converts the 3-HAA amide dimer **(1)** to the 3-HAA oxazole dimer **(2)** as well as to determine its co-factor requirement, the reconstitution of the single reaction catalysed by AjiH was attempted under various conditions. Protein expression constructs and strains described in this chapter were designed and generated by Dr. Qiulin Wu and Joachim Hug (Hug, 2015).

Enzymatic synthesis of 3-HAA amide dimer

The ATP-dependent ligase AjiA1 was expressed in *Escherichia coli* BL21(DE3) bearing pET-28a(+)-*ajiA1* and purified using Ni^{2+} affinity chromatography (Figure 7.15). The 3-HAA dimerisation reaction contained 0.5 mM 3-HAA, 5 µM purified AjiA1, 5 mM ATP and 10 mM $MgCl_2$ in 25 mM HEPES buffer (pH 8.1). The reaction extract was subjected to analysis using LC-MS (Figure 7.16), and the eluate fraction containing the 3-HAA amide dimer **(1)** was collected.

Attempted reconstitution of the AjiH-catalysed cyclisation reaction

The amidocyclase AjiH was expressed in *Streptomyces coelicolor* CH999 bearing pCJW93-*ajiH* and purified using Ni^{2+} affinity chromatography (Figure 7.17). The AjiH reaction contained 0.25 mM 3-HAA amide dimer **(1)**, 1 µM purified AjiH, 5 mM ATP and 10 mM $MgCl_2$ in 25 mM HEPES buffer (pH 8.1). The reaction extract was

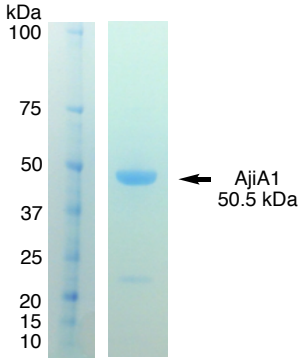


Figure 7.15. SDS-PAGE analysis of purified AjiA1.

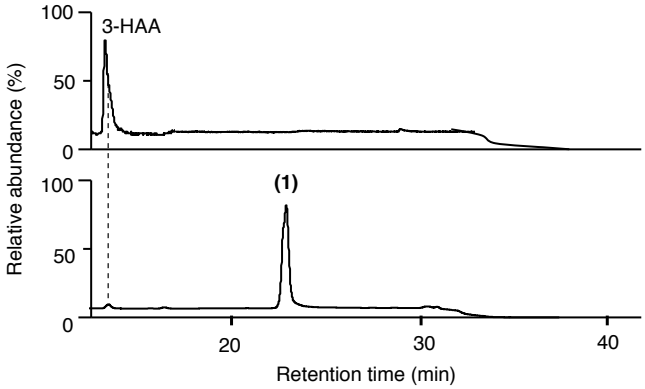


Figure 7.16. LC-MS analysis of the AjiA1 reaction. The reaction containing no AjiA1 (negative control; upper panel) is shown for comparison.

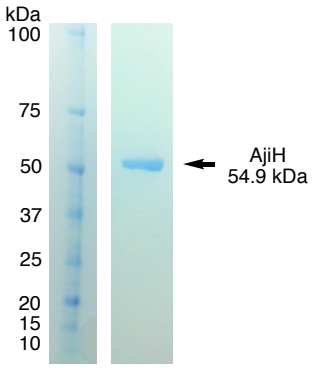


Figure 7.17. SDS-PAGE analysis of purified AjiH.

analysed using LC-MS. Unexpectedly, the 3-HAA oxazole dimer **(2)** could not be detected and no obvious decrease of the 3-HAA amide dimer **(1)** was observed, despite the successful and efficient conversion of 3-HAA to the 3-HAA oxazole dimer **(2)** in the one-pot reaction containing both AjiA1 and AjiH (positive control) (Figure 7.18).

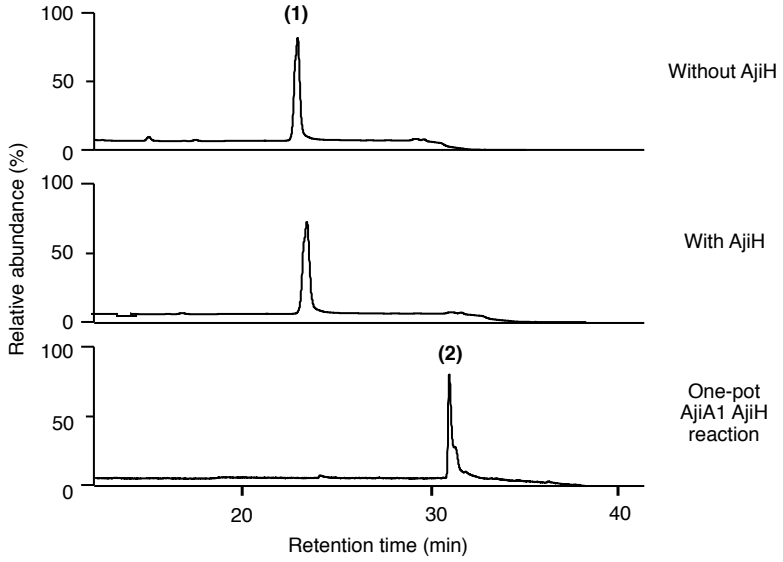


Figure 7.18. LC-MS analysis of the AjiH reactions. The presence or absence of AjiH is indicated and the one-pot reaction containing 3-HAA, AjiA1 and AjiH (positive control) is shown for comparison.

Various conditions were further tested in hope of enabling the cyclisation reaction to take place. At first, it was suspected that the interaction with protein partner(s) of AjiH (i.e. AjiA1 in this case) or the possible by-products of the AjiA1 reaction may be required for the catalytic function of AjiH. So, AjiA1 and AMP were added individually to the reaction but none of them lead to the formation of the 3-HAA oxazole dimer **(2)**. Furthermore, as it was reported that some members of the amidohydrolase superfamily favour the ring closing reactions at lower pH (Gojkovic *et al*., 2003; Porter *et al*., 2004; Hobbs *et al*., 2012; Grande-García *et al*., 2014), 25 mM MES buffer (pH 5.5), MOPS buffer (pH 6.5), HEPES buffer (pH 6.8), Tris-HCl buffer (pH 7.0 and 7.5) were used in place of 25 mM HEPES buffer (pH 8.1). However, the cyclisation still did not take place. The possibility of metal chelation by the aromatic 3-HAA amide dimer **(1)** in preventing it from being processed by AjiH was also considered. However, the addition of ethylenediaminetetraacetic acid (EDTA) had no effect.

Investigation of the true substrate of AjiH

These results raised questions regarding the identity of the true substrate of AjiH. In addition to the 3-HAA amide dimer **(1)**, two alternative substrates including the dimer adenylate and the 3-HAA ester dimer were considered as candidates.

First, to investigate whether or not AjiA1 is capable of producing an adenylate form of the dimer, the ATP/3-HAA stoichiometry was determined using the EnzChek™ Pyrophosphate Assay Kit (Figure 7.19). As only one molecule of ATP appeared to be required for two molecules of 3-HAA to be dimerised, and since the 3-HAA dimer adenylate was never observed from the AjiA1 reaction, it is unlikely that the adenylate was formed for AjiH to act on.

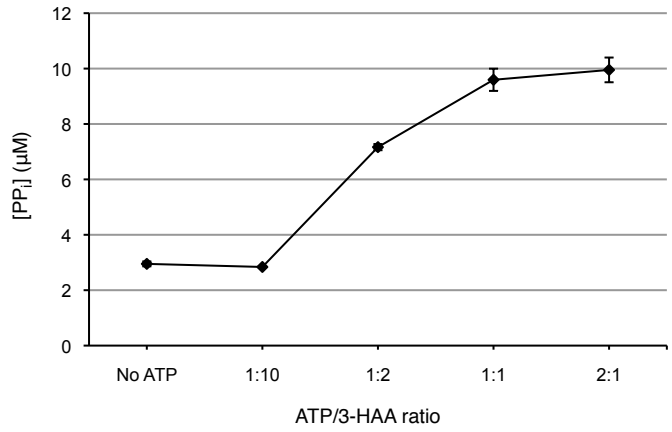


Figure 7.19. ATP/3-HAA stoichiometry of the AjiA1 reaction.

On the other hand, members of the metallo-amidohydrolase family are known to act on amide and/or ester substrates (Seibert and Raushel, 2005). As an ester carbonyl is more electrophilic and theoretically more prone to a nucleophilic attack than an amide carbonyl, it is tempting to propose that an ester-linked 3-HAA dimer would serve as a more energetically-favourable substrate. Although the isolated, ^{15}N-labelled product of the AjiA1 reaction was determined by NMR analysis (Hug, 2015), the possibility of an ester-linked dimer being formed first and undergoing an acyl shift to form the more stable amide (during isolation) could not be ruled out.

In silico docking of both the amide-linked and the ester-linked dimers to the AjiH model was attempted using SwissDock (Grosdidier *et al*., 2011a; 2011b), but was unsuccessful. The loops at the opening of the TIM barrel, were modelled at a relatively low level of confidence, and could possibly be restricting access of the substrate(s) to the channel for entry to the active site pocket.

To examine the potential of AjiA1 in ester formation, AjiA1 reactions were carried out with two alternative substrates, 3-hydroxybenozic acid (3-HBA) and 2,3-aminomethylbenzoic acid (2,3-AMBA), respectively. A peak corresponding to a dimer product was observed in the reaction containing 3-HBA but not 2,3-AMBA (Figure 7.20). The result demonstrates AjiA1's ability to form an ester bond in the absence of the amine group. Nevertheless, it does not necessarily imply an ester instead of an amide is formed when the native substrate 3-HAA is used. Anthranilic acid would have been a more suitable substrate for testing than 2,3-AMBA, as the methyl group may cause steric clash in AjiA1. Unfortunately, anthranilic acid is a category 2 substance (regulated drug precursor) and was not available.

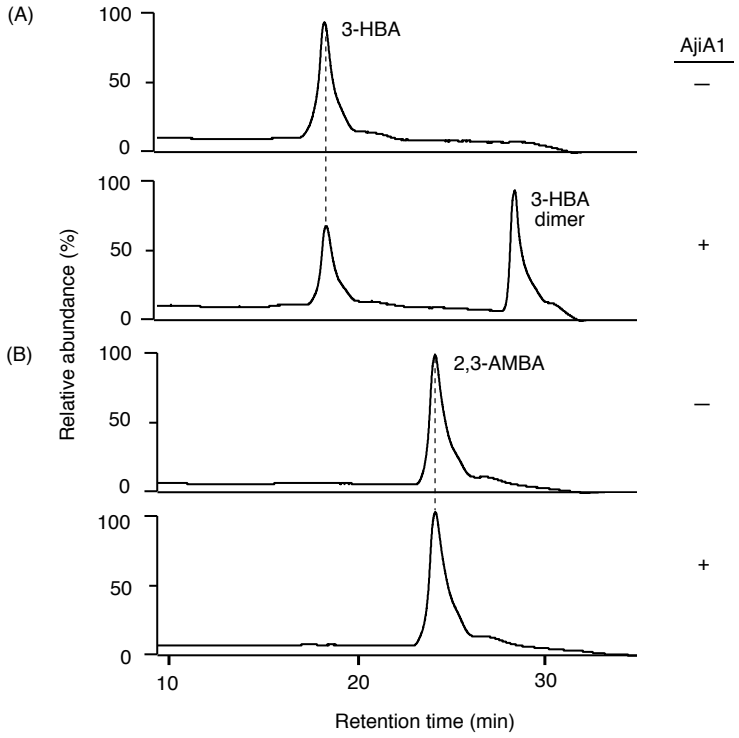


Figure 7.20. LC-MS analysis of AjiA1 reactions with alternative substrates. (A) 3-HBA. (B) 2,3-AMBA. The presence or absence of AjiA1 is indicated.

7.2.4 *In vitro* reconstitution of AJI9561 biosynthesis using 3-HAA amide dimer

Previously, the isolated 3-HAA oxazole dimer **(2)** was confirmed as a competent precursor to AJI9561 (Hug, 2015). A similar experiment was conducted with pre-isolated 3-HAA amide dimer **(1)** instead, to check if it could be processed by downstream enzymes in the pathway, in hope that the presence of 6-MSA and other

Aji enzymes in addition to AjiH would drive the overall reaction forward to yield AJI9561.

Chemical synthesis of 6-MSA

6-MSA was chemically synthesised from the commercially available ethyl-6-MSA, via alkaline hydrolysis of the ester, as described in Hug (2015) (Section 8.7.2).

Optimisation of one-pot *in vitro* reconstitution to AJI9561

As mentioned in Section 7.1.3, previous efforts on the *in vitro* reconstitution were frustrated by the incomplete conversion of intermediates to AJI9561 (<5% yield) (Hug, 2015). To enable the generation of a sufficient amount of products for further study of the pathway, various strategies were employed to optimise the yield of the reaction.

AjiA2 and AjiG were expressed in *E. coli* BL21(DE3) bearing pET-28a(+)-*ajiA2* and pET-28a(+)-*ajiG*, respectively. AjiP was co-expressed with MtaA, a PPTase from *Stigmatella aurantiaca* DW4/3-1 to enable the conversion of *apo*-AjiP to its *holo* form *in vivo*, in *E. coli* BL21(DE3) bearing pET-28a(+)-*ajiP* and pSU-*mtaA*. AjiA2, AjiG and AjiP were purified using Ni^{2+} affinity chromatography (Figure 7.21).

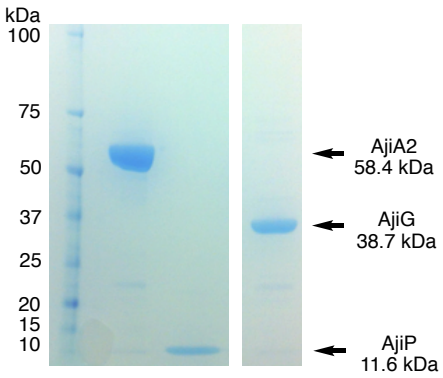


Figure 7.21. SDS-PAGE analysis of purified AjiA2, AjiG and AjiP.

LC-MS analysis of purified AjiP showed that although *holo*-AjiP was present, a significant proportion of the sample comprised of *apo*-AjiP which failed to be activated by the co-expressed MtaA *in vivo*. To activate the remaining *apo*-AjiP, Sfp was included in the *in vitro* reaction.

To facilitate the adenylation and loading of 6-MSA, the precursor was pre-incubated with AjiA2, AjiP, Sfp, CoA, ATP and $MgCl_2$ for 60 minutes prior to the mixing with other reagents and enzymes.

Moreover, although AJI9561 is assembled from two units of 3-HAA and only one unit of 6-MSA, an equal concentration of 6-MSA and 3-HAA increased the yield and maximised the utilisation of both precursors. This is likely due to the requirement of 6-MSA to be both activated and loaded onto the *holo*-AjiP to serve as a substrate for AjiG. A 6-MSA/3-HAA ratio of 1:1 resulted in the near-complete conversion to AJI9561 with no 3-HAA and minimal amounts of 6-MSA remaining (Figure 7.22).

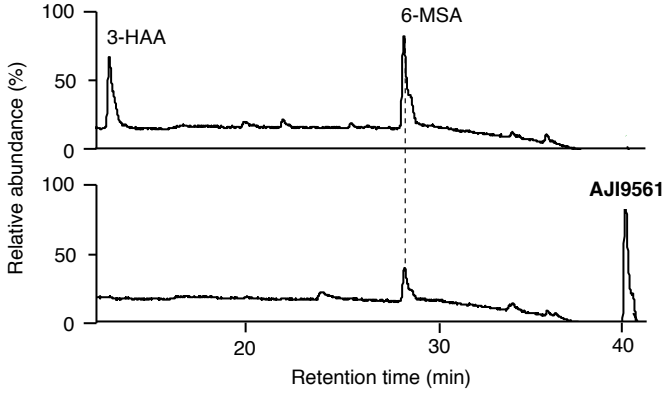


Figure 7.22. Optimised one-pot enzymatic conversion to AJI9561.

In vitro reconstitution of AJI9561 biosynthesis using the 3-HAA amide dimer and 6-MSA as substrates

The isolated 3-HAA amide dimer **(1)** (0.25 mM) was incubated with 1 mM 6-MSA, 10 µM AjiA2, 60 µM AjiP, 20 µM AjiG, 1 uM AjiH, 20 µM Sfp, 0.25 mM CoA, 5 mM ATP and 10 mM $MgCl_2$ in 25 mM HEPES buffer (pH 8.1), in a total reaction volume of 100 µL (as with the previous one-pot reactions) or 2 mL. A small amount of AJI9561 could be detected by LC-MS in the upscaled reaction of 2 mL (Figure 7.23), but not in 100 µL. Despite the near-complete conversion of 3-HAA and 6-MSA to AJI9561 in the one-pot reaction containing all five Aji enzymes (positive control), the efficiency was extremely low when the 3-HAA amide dimer **(1)** was used. Nevertheless, the 3-HAA amide dimer **(1)** was demonstrated for the first time to be a competent substrate for AJI9561 biosynthesis.

7.2.5 Investigation of alternative biosynthetic routes to AJI9561

Although the 3-HAA oxazole dimer (Hug, 2015) and the 3-HAA amide dimer were confirmed to be competent for the enzymatic conversion to AJI9561, alternative biosynthetic routes involving different orders of precursor assembly could not be

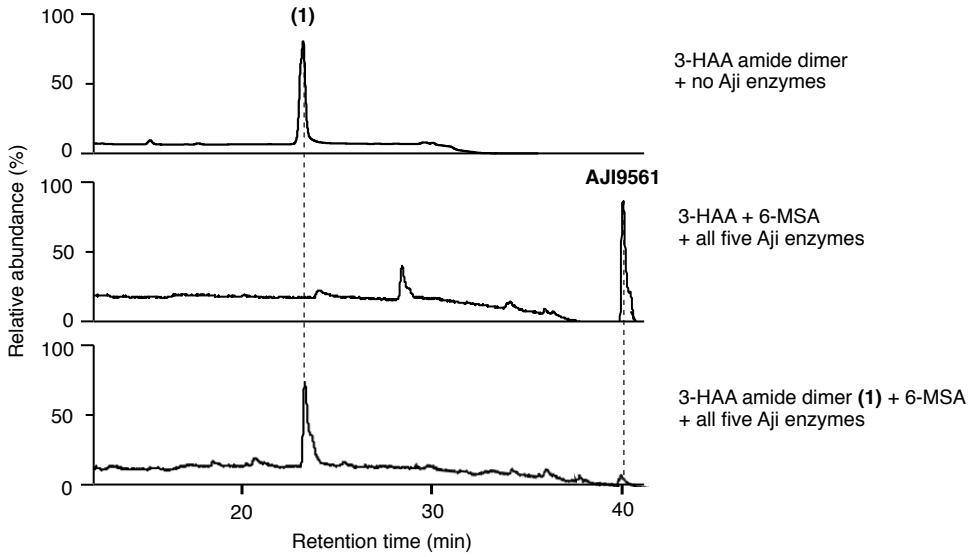


Figure 7.23. LC-MS analysis of the *in vitro* one-pot reconstitution of AJI9561 biosynthesis using the 3-HAA amide dimer and 6-MSA. A sample of the pre-isolated 3-HAA amide dimer **(1)** with no Aji enzymes was used as a standard and the one-pot reaction containing 3-HAA, 6-MSA and all five Aji enzymes was used as a positive control.

ruled out at this stage. For instance, the possibility of 3-HAA-6-MSA heterodimer formation and the heterodimer being the preferred dimer intermediate had not been assessed experimentally. Various combinations of Aji enzymes were incubated with 3-HAA and 6-MSA to check for the biosynthesis of previously unseen products. Subsequent testing of the isolated products, as alternative substrates for Aji enzymes, was carried out to determine their competence as intermediates in the AJI9561 pathway.

One-pot reaction with various combinations of Aji enzymes

Three different combinations of enzymes were incubated with 6-MSA and 3-HAA in a one-pot manner: (i) AjiA2, AjiP and AjiG, (ii) AjiA2, AjiP, AjiG and AjiH and (iii) AjiA2, AjiP, AjiG and AjiA1 (Figure 7.23).

Through the incubation of AjiA2, AjiP and AjiG with both precursors in reaction (i), m/z 310.07 which corresponds to the sodium adduct ion of the 3-HAA-6-MSA heterodimer **(4)** with a retention time of 29.24 min was detected (Figures 7.24 and 7.25).

With the addition of AjiH in reaction (ii), m/z 292.06 which corresponds to the sodium adduct ions of the 3-HAA-6-MSA oxazole heterodimer **(5)** at 35.32 min and a characteristic UV spectrum were observed (Figures 7.24 and 7.26).

In the absence of AjiH, the 3-HAA dimer **(1)** was the major product of reaction (iii) and m/z 445.10 which corresponds to the sodium adduct ion of the 3-HAA dimer-6-MSA trimer **(6)**, with neither of the oxazole rings closed, was detected at 30.73 min as a minor product (Figures 7.24 and 7.27). It is unclear whether the 3-HAA dimer **(1)** or the 3-HAA-6-MSA heterodimer **(4)** was the (preferred) precursor to the trimer **(6)**.

All three products **(4-6)** as well as AJI9561 were submitted for accurate mass analysis (Table 7.3). Unfortunately, the amounts were not sufficient for detailed structure elucidation.

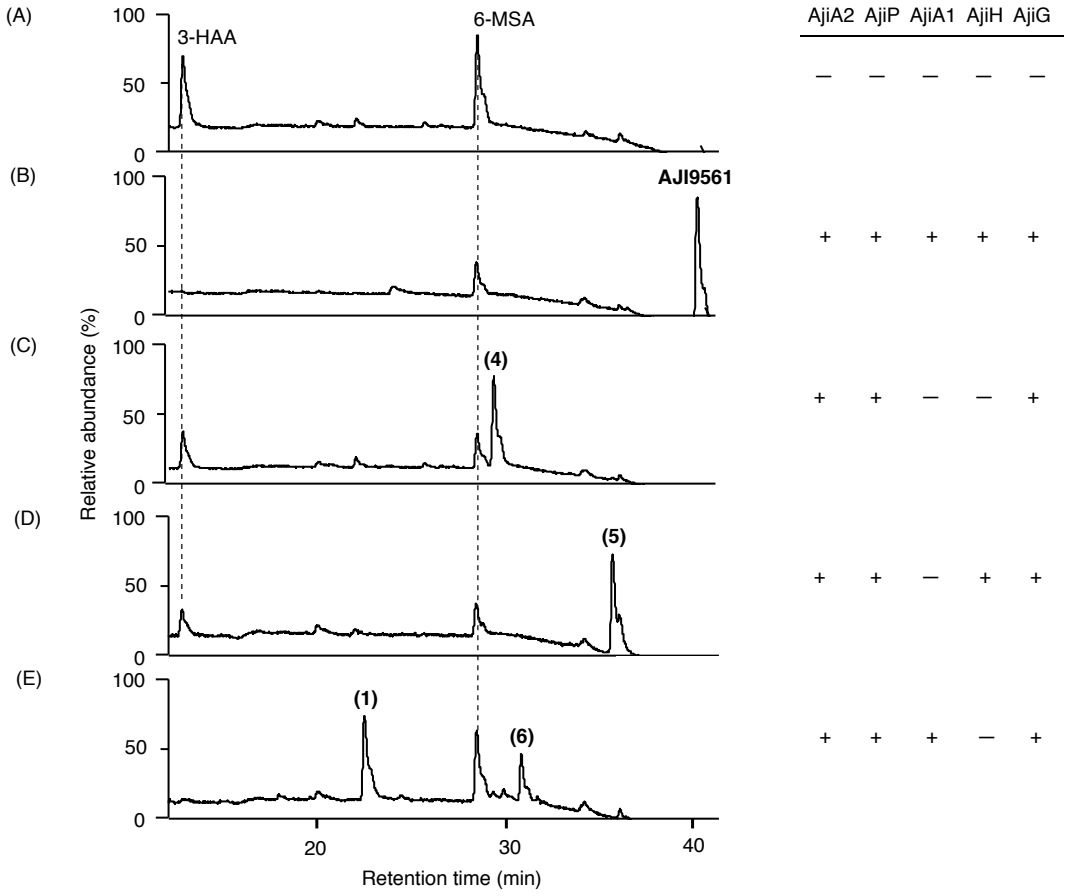


Figure 7.24. LC-MS analysis of one-pot reaction with various combinations of Aji enzymes. LC-MS-UV chromatogram of (A-B) the negative and positive controls, (C) reaction (i), (D) reaction (ii) and (E) reaction (iii). The presence and/or absence of Aji enzymes in the reactions are indicated.

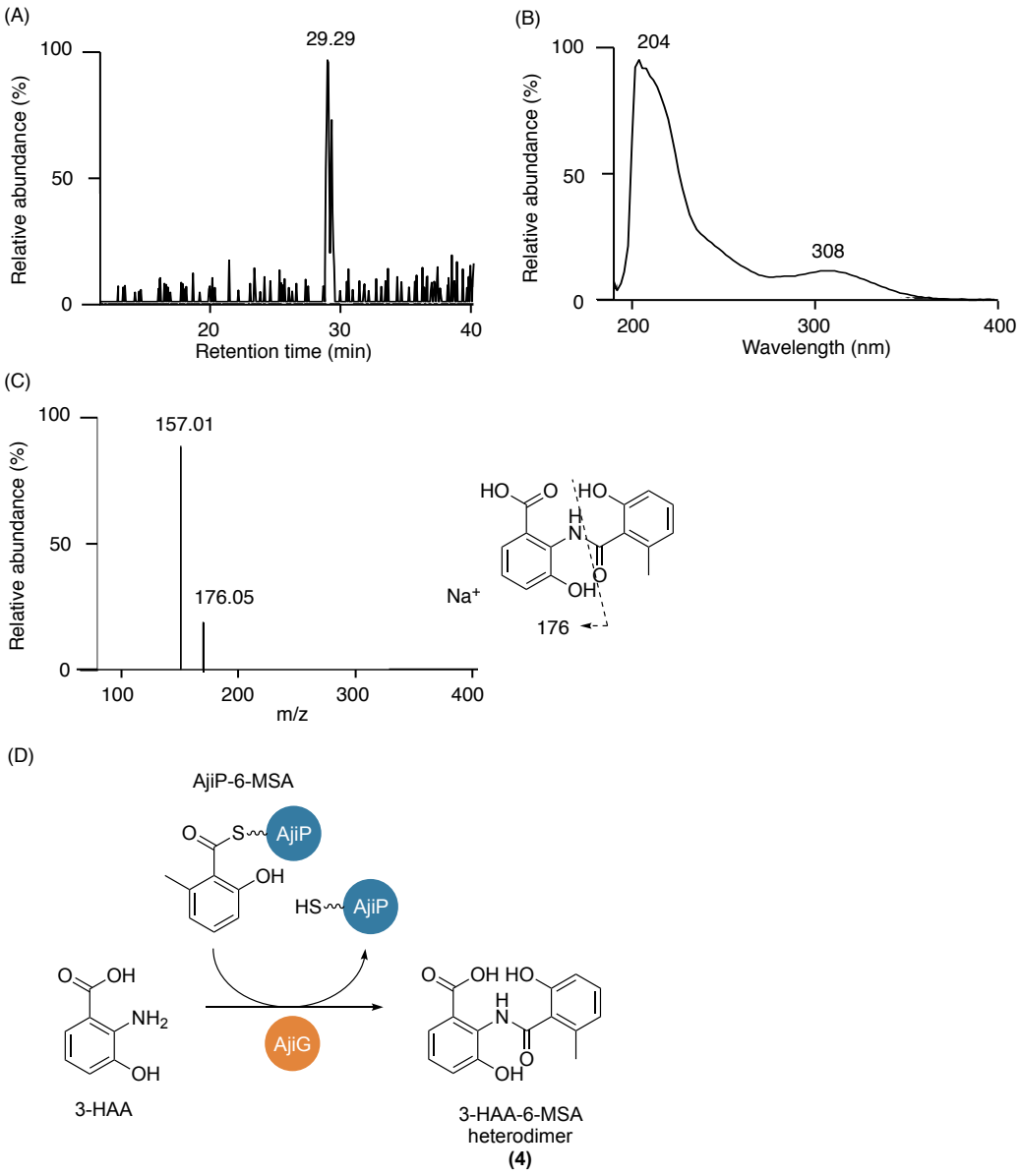


Figure 7.25. LC-MS analysis of 3-HAA-6-MSA heterodimer (4). (A) Selective ion monitoring, (B) UV spectrum and (C) MS/MS fragmentation of m/z $[M+Na]^+$ 310.07. (D) Proposed biosynthetic route to the 3-HAA-6-MSA heterodimer **(4)** catalysed by AjiG.

Notably, the detection of these three previously unseen products revealed the broad substrate specificity of AjiH and AjiG. Not only was AjiH capable of closing the oxazole ring in the 3-HAA dimer **(1)** and the benzoxazole trimer, but also in the 3-HAA-6-MSA heterodimer **(4)**, whereas AjiG catalyses the condensation between AjiP-6-MSA and any of: 3-HAA, the 3-HAA oxazole dimer **(2)** and the non-cyclised 3-HAA dimer **(1)**.

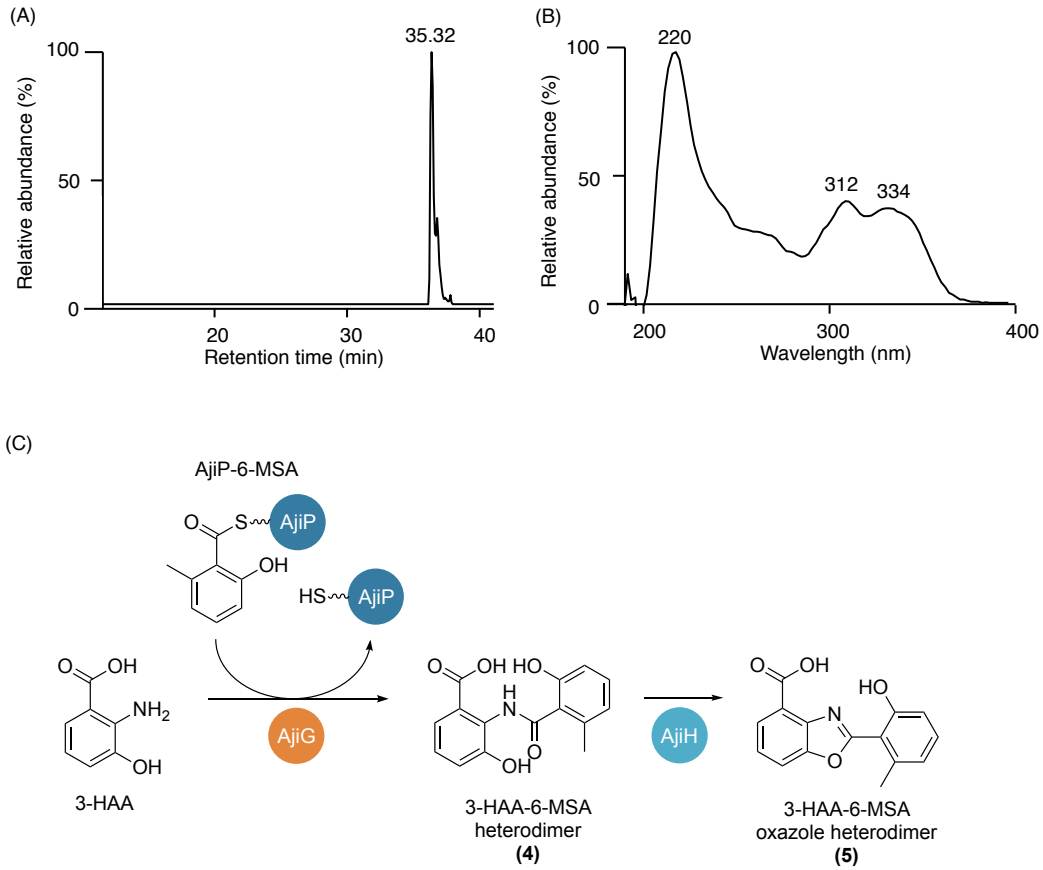


Figure 7.26. LC-MS analysis of 3-HAA-6-MSA oxazole heterodimer (5). (A) Selective ion monitoring and (B) UV spectrum of m/z $[M+Na]^+$ 292.06. (C) Proposed biosynthetic route to the 3-HAA-6-MSA oxazole heterodimer **(5)** catalysed by AjiG and AjiH.

Biosynthetic competency of 3-HAA-6-MSA dimer, 3-HAA-6-MSA oxazole dimer and 3-HAA dimer-6-MSA trimer

In light of the detection of previously unseen products, alternative biosynthetic routes to AJI9561 via a different order of precursor assembly were investigated.

The 3-HAA-6-MSA heterodimer **(4)**, the 3-HAA-6-MSA oxazole heterodimer **(5)** and the 3-HAA dimer-6-MSA trimer **(6)** were purified using LC-MS, and were incubated individually with downstream enzymes. None of them were competent for downstream processing to yield AJI9561, unlike the 3-HAA dimer **(1)** (discussed earlier in this section) and the 3-HAA oxazole dimer **(2)** (Hug, 2015), and were deemed to be shunt products of the pathway.

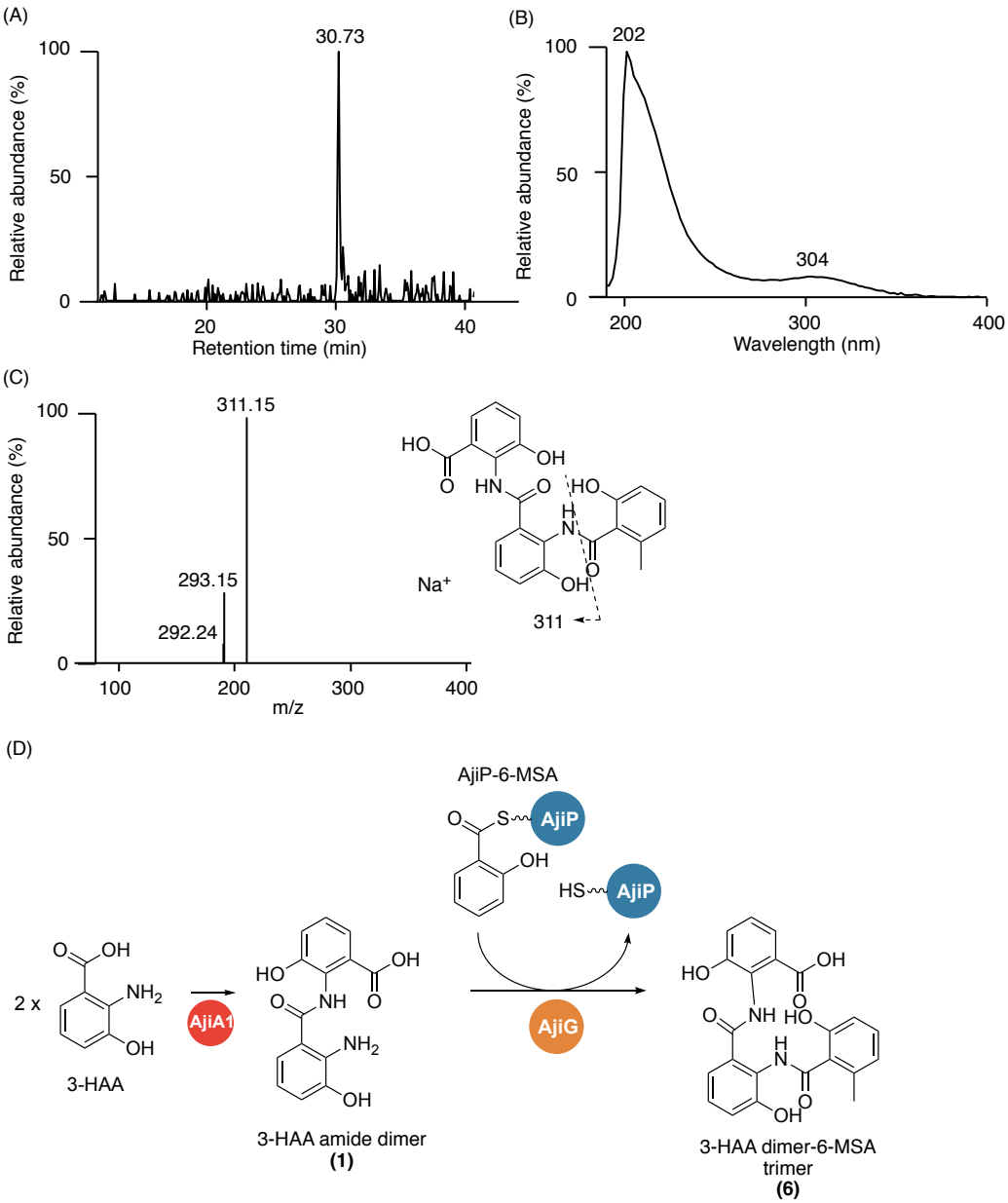


Figure 7.27. LC-MS analysis of 3-HAA dimer-6-MSA trimer (6). (A) Selective ion monitoring, (B) UV spectrum and (C) MS/MS fragmentation of m/z $[M+Na]^+$ 445.10. (C) Proposed biosynthetic route to the 3-HAA dimer-6-MSA trimer **(6)** catalysed by AjiA1 and AjiG.

Table 7.3. Accurate mass analyses of the 3-HAA-6-MSA heterodimer (4), the 3-HAA-6-MSA oxazole heterodimer (5), the 3-HAA dimer-6-MSA trimer (6) and AJI9561.

Product	Molecular formula	Theorectical neutral mass (Da)	Observed neutral mass (Da)
3-HAA-6-MSA heterodimer	$C_{15}H_{13}NO_5$	287.07937	287.0783
3-HAA-6-MSA oxazole heterodimer	$C_{15}H_{11}NO_4$	269.06881	269.0682
3-HAA dimer-6-MSA trimer	$C_{22}H_{18}N_2O_7$	422.11140	422.1101
AJI9561	$C_{22}H_{14}N_2O_5$	386.09027	386.0900

7.3 Concluding remarks and future work

In this chapter, the biosynthetic route to AJI9561, a potent, bis(benzoxazole) antibiotic and anti-proliferative agent produced by *Streptomyces* sp. AJ9561, was investigated via an *in vitro* approach.

A key enzyme in benzoxazole formation, the amidocyclase AjiH, was subjected to detailed study via phylogenetic analysis, structural modelling and *in vitro* reconstitution. AjiH and its counterparts in other (bis)benzoxazole antibiotic biosynthetic pathways form a distinct clade in the metallo-amidohydrolase superfamily. AjiH and its counterparts in the nataxazole, A33853 and caboxamycin pathways represent a novel catalytic function in the evolutionarily-divergent metallo-amidohydrolase superfamily. Unexpectedly, the cyclisation of the previously identified substrate 3-HAA amide dimer **(1)**, proposed to be catalysed by AjiH, could not be reconstituted as a single reaction despite multiple attempts under different assay conditions. Two alternative candidates to the amide-linked 3-HAA dimer **(1)**, the 3-HAA dimer adenylate and the ester-linked 3-HAA dimer, were considered. The adenylate was ruled out as it is unlikely to be produced by AjiA1, the ATP-dependent ligase involved in the activation and dimerisation of 3-HAA, and was never observed in the enzymatic assays. However, as the amide-linked and ester-linked 3-HAA dimers share the same molecular weight and MS/MS fragmentation pattern, they could not be distinguished from each other using LC-MS. Although AjiA1 was shown to catalyse the dimerisation of 3-HBA (in the absence of a 2-substituted amine group) via an ester linkage, the same may or may not be applicable in the case of 3-HAA, and hence, the true identity of AjiH's substrate for 3-HAA oxazole dimer **(2)** formation remains obscure.

On the other hand, the one-pot reconstitution of AJI9561 biosynthesis, using 3-HAA and 6-MSA as precursors, was successfully optimised with a yield improvement from <5% to more than 90%. This enabled the detection of the low amount of AJI9561 enzymatically-converted from the isolated 3-HAA amide dimer **(1)** and 6-MSA, demonstrating the competency of the amide as a substrate for AJI9561 biosynthesis for the first time. Alternative biosynthetic routes to AJI9561 were also investigated (Figure 7.30). From one-pot reactions containing different combinations of Aji enzymes, the 3-HAA-6-MSA heterodimer **(4)**, the 3-HAA-6-MSA oxazole heterodimer **(5)** and the 3-HAA dimer-6-MSA trimer **(6)** were isolated and assessed for their biosynthetic competency. They are likely shunt products (and not intermediates) of the AJI9561 pathway as none of them were suitable substrates for downstream processing by Aji enzymes *in vitro*.

For further work, metal substitution and crystallographic studies of the amidocyclase AjiH could shed light on its enzymology and mechanism for catalysis. If chemically-synthesised amide-linked and ester-linked 3-HAA dimers are available, they could be tested directly with AjiH and used as a standard for the other assays. Moreover, AjiH and AjiG exhibited a relaxed substrate specificity. The Aji enzymes could potentially be used for the generation of new benzoxazole antibiotics, via the use of benzoic acid derivatives as precursors.

Chapter 8

Materials and methods

8.1 Materials

8.1.1 Chemical and biological reagents

All chemical and biological reagents used in this work were purchased from Sigma-Aldrich, Thermo Fisher Scientific and New England Biolabs, unless stated otherwise.

8.1.2 Culture media

Culture media used in this work are shown in Table 8.1.

Table 8.1. Culture media used in this work.

Medium	Composition
2TY	1.6% (w/v) tryptone; 1% (w/v) yeast extract; 0.5% (w/v) NaCl; with or without 2% (w/v) agar; in distilled H_2O; pH 7.0
LB	1% (w/v) tryptone; 0.5% (w/v) yeast extract; 1% (w/v) NaCl; with or without 2% (w/v) agar; in distilled H_2O; pH 7.0
SFM	2% (w/v) soy flour; 2% (w/v) mannitol; with or without 2% (w/v) agar; in tap water
TSBY	3% (w/v) tryptic soy broth; 10.3% (w/v) sucrose; 0.5% (w/v) yeast extract; in distilled H_2O
Super YEME	0.3% (w/v) yeast extract; 0.5% (w/v) peptone; 1% (w/v) glucose; 0.3% (w/v) malt extract; 34% (w/v) sucrose; 0.5% (w/v) glycine; 0.235% (w/v) $MgCl_2$; 0.0075% (w/v) L-proline; 0.0075% (w/v) L-arginine; 0.0075% (w/v) L-cysteine; 0.01% (w/v) L-histidine; 0.0015% (w/v) uracil; distilled H_2O

8.1.3 Antibiotics

Antibiotics used in this work are listed in Table 8.2.

Table 8.2. Antibiotics used in this work.

Antibiotic	Solvent	Stock concentration (mg mL^{-1})	Working concentration (µg mL^{-1})
Apramycin	H_2O	50	50
Carbenicillin	H_2O	50	50
Chloramphenicol	Ethanol	25	25
Kanamycin	H_2O	50	50
Nalidixic acid	H_2O	25	25
Thiostrepton	Ethanol/DMSO 1:1 (v/v)	10	10

8.1.4 Bacterial strains

All bacterial strains were handled with aseptic techniques in laminar flow hoods. *Streptomyces* sp. OM-5714 and *Streptomyces* sp. KO-7888 were kind gifts from Professor Satoshi Omura (Kitasato Institute, Japan). Bacterial strains used in this work are listed in Table 8.3.

Table 8.3. Bacterial strains used in this work. Strains generated by previous group members and deposited in the PFL strain collection are indicated with an asterisk.

Strains	Genotype/Features	Reference
***Escherichia coli* strains**		
DH10B	F⁻ *mcrA* Δ(*mrr*⁻ *hsd* RMS⁻ *mcrBC*) Φ80*lac*Z ΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ (*ara leu*)7697 *gal*U *gal*K *rps*L *nup*G λ⁻. Host for cloning.	Invitrogen
ET12567/pUZ8002	F⁻ *dam*⁻ 13::Tn9 *dcm*⁻ 6 *hsdM hsdR* recF143 zjj-202::Tn10 *galK2 galT22 ara14 pacY1 xyl-5 leuB6 thi-1*. Donor strain for DNA conjugation into *Streptomyces*.	MacNeil (1992)
BL21(DE3)	F⁻ *ompT hsdS*$_B$(r_B^-, m_B^-) *gal dcm* (DE3). Host for protein expression.	Novagen
BL21(DE3)/pET28a-*ajiA1*	*E. coli* BL21(DE3) bearing the protein expression construct pET-28a(+) containing *ajiA1* with an N-terminal His_6-tag.	*
BL21(DE3)/pET28a-*ajiA2*	*E. coli* BL21(DE3) bearing the protein expression construct pET-28a(+) containing *ajiA2* with an N-terminal His_6-tag.	*
BL21(DE3)/pET28a-*ajiG*	*E. coli* BL21(DE3) bearing the protein expression construct pET-28a(+) containing *ajiG* with an N-terminal His_6-tag.	*
BL21(DE3)/pET28a-*ajiP* pSU-*mtaA*	*E. coli* BL21(DE3) bearing the protein expression constructs pET-28a(+) containing *ajiP* with an N-terminal His_6-tag and pSU containing *mtaA*.	Hug (2015)

***Streptomyces* strains**

Streptomyces sp. OM-5714	Phthoxazolins producer.	Omura *et al*. (1990)
Streptomyces sp. KO-7888	Phthoxazolins producer.	Shiomi *et al*. (1995)
S. conglobatus ATCC 31005	Conglobatin producer. For PCR amplification of *congE*.	Westley *et al*. (1979)
ΔACP13-S1676A	*Streptomyces* sp. KO-7888 with a point mutation of the catalytic Ser1676 to Ala in ACP13 of PhoxR.	This work
ΔACP15-S4259A	*Streptomyces* sp. KO-7888 with a point mutation of the catalytic Ser4259 to Ala in ACP15 of PhoxR.	This work
ΔACP16-S5512A	*Streptomyces* sp. KO-7888 with a point mutation of the catalytic Ser5512 to Ala in ACP16 of PhoxR.	This work
Δ*avaR3*	*Streptomyces* sp. KO-7888 with *avaR3* (outside of *phox* cluster) deleted in-frame.	This work
Δ*orf+5*	*Streptomyces* sp. KO-7888 with *orf+5* deleted in-frame.	This work
Δ*orf0034*	*Streptomyces* sp. KO-7888 with *orf0034* (outside of *phox* cluster) deleted in-frame.	This work
Δ*orf3515*	*Streptomyces* sp. KO-7888 with o*rf3515* (outside of *phox* cluster) deleted in-frame.	This work
ΔPCP14-S2778A	*Streptomyces* sp. KO-7888 with a point mutation of the catalytic Ser2778 to Ala in PCP14 of PhoxR.	This work
Δ*phoxE*	*Streptomyces* sp. KO-7888 with *phoxE* deleted in-frame.	This work
Δ*phoxF*	*Streptomyces* sp. KO-7888 with *phoxF* deleted in-frame.	This work
Δ*phoxG*	*Streptomyces* sp. KO-7888 with *phoxG* deleted in-frame.	This work
Δ*phoxI*	*Streptomyces* sp. KO-7888 with *phoxI* deleted in-frame.	This work
Δ*phoxJ*	*Streptomyces* sp. KO-7888 with *phoxJ* deleted in-frame.	This work
Δ*phoxK*	*Streptomyces* sp. KO-7888 with *phox*K deleted in-frame.	This work
Δ*phoxP*	*Streptomyces* sp. KO-7888 with *phoxP* deleted in-frame.	This work
Δ*phoxP*::*phoxP*	*Streptomyces* sp. KO-7888 with *phoxP* deleted in-frame and *phoxP* genomically integrated at phiC31 attB site.	This work
Δ*phoxP*::*phoxP*-OM	*Streptomyces* sp. KO-7888 with *phoxP* deleted in-frame and *phoxP* from *Streptomyces* sp. OM-5714 genomically integrated at phiC31 attB site.	This work
Δ*phoxP*::*ozmP*	*Streptomyces* sp. KO-7888 with *phoxP* deleted in-frame and *ozmP* genomically integrated at phiC31 attB site.	This work
Δ*phoxP*::*congE*	*Streptomyces* sp. KO-7888 with *phoxP* deleted in-frame and *congE* genomically integrated at phiC31 attB site.	This work

Δ*phoxR*	*Streptomyces* sp. KO-7888 with *phoxR* deleted in-frame.	This work
Δ*phoxU*	*Streptomyces* sp. KO-7888 with *phoxU* deleted in-frame.	This work
Δ*phoxV*	*Streptomyces* sp. KO-7888 with *phoxV* deleted in-frame.	This work
WT::*orf+10*	*Streptomyces* sp. KO-7888 with an additional copy of *orf+10* genomically integrated at phiC31 attB site.	This work
WT::*orf+11*	*Streptomyces* sp. KO-7888 with an additional copy of *orf+11* genomically integrated at phiC31 attB site.	This work
WT::*phoxC*	*Streptomyces* sp. KO-7888 with an additional copy of *phoxC* genomically integrated at phiC31 attB site.	This work
WT::*phoxT*	*Streptomyces* sp. KO-7888 with an additional copy of *phoxT* genomically integrated at phiC31 attB site.	This work
S. coelicolor CH999/pCJW93-*ajiH*	*S. coelicolor* CH999 bearing the protein expression construct pCJW93 containing *ajiH* between *Nde*I and *Eco*RI with an N-terminal His_6-tag. For thiostrepton-inducible expression of AjiH under the *tipA* promoter.	Hug (2015)

8.1.5 Plasmid constructs

All plasmid constructs were stored in MQ H_2O at -20 °C. Plasmid constructs used in this work are listed in Table 8.4.

Table 8.4. Plasmid constructs used in this work. Plasmid constructs generated by previous group members and deposited in the PFL strain collection are indicated with an asterisk.

Plasmid	Features	Reference
For in-frame deletions		
pYH7	*Amp, aac(3)IV, tsr, cos, oriT,* P_{T7}, P_{T3}, ori_{pIJ101}, ori_{ColE1}, *E. coli-Streptomyces* shuttle vector	Sun (2009)
pYH7-avaR3	Construct for in-frame deletion of *avaR3* with arms of homology flanking *avaR3* inserted at *Nde*I of pYH7.	This work
pYH7-orf+5	Construct for in-frame deletion of *orf+5* with arms of homology flanking *phoxAA* inserted at *Nde*I of pYH7.	This work
pYH7-orf0034	Construct for in-frame deletion of *orf0034* with arms of homology flanking *asnS* inserted at *Nde*I of pYH7.	This work
pYH7-orf3515	Construct for in-frame deletion of *orf3515* with arms of homology flanking *orf3515* inserted at *Nde*I of pYH7.	This work
pYH7-phoxE	Construct for in-frame deletion of *phoxE* with arms of homology flanking *phoxE* inserted at *Nde*I of pYH7.	This work

pYH7-phoxF	Construct for in-frame deletion of *phoxF* with arms of homology flanking *phoxF* inserted at *Nde*I of pYH7.	This work
pYH7-phoxG	Construct for in-frame deletion of *phoxG* with arms of homology flanking *phoxG* inserted at *Nde*I of pYH7.	This work
pYH7-phoxI	Construct for in-frame deletion of *phoxI* with arms of homology flanking *phoxI* inserted at *Nde*I of pYH7.	This work
pYH7-phoxJ	Construct for in-frame deletion of *phoxJ* with arms of homology flanking *phoxJ* inserted at *Nde*I of pYH7.	This work
pYH7-phoxP	Construct for in-frame deletion of *phoxP* with arms of homology flanking *phoxP* inserted at *Nde*I of pYH7.	This work
pYH7-phoxR	Construct for in-frame deletion of *phoxR* with arms of homology flanking *phoxR* inserted at *Nde*I of pYH7.	This work
pYH7-phoxU	Construct for in-frame deletion of *phoxU* with arms of homology flanking *phoxU* inserted at *Nde*I of pYH7.	This work
pYH7-phoxV	Construct for in-frame deletion of *phoxV* with arms of homology flanking *phoxV* inserted at *Nde*I of pYH7.	This work
For point mutagenesis		
pYH7-ACP13	Construct for inactivation of ACP13 in PhoxR with arms of homology containing the Ser1676 to Ala point mutation inserted at *Nde*I of pYH7.	This work
pYH7-ACP15	Construct for inactivation of ACP15 in PhoxR with arms of homology containing the Ser4259 to Ala point mutation inserted at *Nde*I of pYH7.	This work
pYH7-ACP16	Construct for inactivation of ACP16 in PhoxR with arms of homology containing the Ser5512 to Ala point mutation inserted at *Nde*I of pYH7.	This work
pYH7-PCP14	Construct for inactivation of PCP14 in PhoxR with arms of homology containing the Ser2278 to Ala point mutation inserted at *Nde*I of pYH7.	This work
For complementation of in-frame deletions		
pIB139	*aac(3)IV, oriT, attP*(ΦC31), *int*, P_{ermE*}, ori_{pUC}, ΦC31 site integrative vector with *erm*E* promoter.	Wilkinson (2002)
pIB139-phoxP	Complementation construct containing *phoxP* between *Nde*I and *Xba*I of pIB139.	This work
pIB139-phoxP-OM	Complementation construct containing *phoxP* from *Streptomyces* sp. OM-5714 between *Nde*I and *Xba*I of pIB139.	This work
pIB139-ozmP	Complementation construct containing *ozmP* from pYJ55 between *Nde*I and *Xba*I of pIB139.	This work
pIB139-congE	Complementation construct containing *congE* from	This work

	Streptomyces conglobatus between *Nde*I and *Xba*I of pIB139.	
For overexpression of (putative) regulatory genes		
pIB139-orf+10	Overexpression construct containing *orf+10* between *Nde*I and *Xba*I of pIB139.	This work
pIB139-orf+11	Overexpression construct containing *orf+11* between *Nde*I and *Xba*I of pIB139.	This work
pIB139-phoxC	Overexpression construct containing *phoxC* between *Nde*I and *Xba*I of pIB139.	This work
pIB139-phoxT	Overexpression construct containing *phoxT* between *Nde*I and *Xba*I of pIB139.	This work
For protein expression		
pET-28a(+)	*kan*, *lacI,* P_{T7}, ori_{pBR322}, ori_{F1}, N-terminal and C-terminal His_6-tag. For protein expression.	Novagen
pET28a-folD	Protein expression construct containing *folD* from *E. coli* DH10B between *Nde*I and *Xho*I of pET-28a(+) with an N-terminal His_6-tag.	This work
pET28a-phoxO	Protein expression construct containing *phoxO* between *Nde*I and *Xho*I of pET-28a(+) with an N-terminal His_6-tag.	This work
pET28a-phoxN (Val1-Gly1483)	Protein expression construct containing the first 4452 bp of p*hoxN* (encoding Val1-Gly1483) between *Nde*I and *Hind*III of pET-28a(+) with an N-terminal His_6-tag.	This work
pET28a-phoxS	A protein expression construct containing *phoxS* between *Nde*I and *Xho*I of pET-28a(+) with an N-terminal His_6-tag.	This work
pET28a-sfp	Protein expression construct containing *sfp* between *Nde*I and *Xho*I of pET-28a(+) with an N-terminal His_6-tag.	*
pET-29b(+)	*kan*, *lacI,* P_{T7}, ori_{pBR322}, ori_{F1}, N-terminal S-tag, C-terminal His_6-tag. For protein expression.	Novagen
pET29b-phoxP	Protein expression construct containing *phoxP* between *Nde*I and *Xho*I of pET-29b(+) with an N-terminal His_6-tag.	This work
pET29b-phoxQ	Protein expression construct containing *phoxQ* between *Nde*I and *Hind*III of pET-29b(+) with an N-terminal His_6-tag.	This work
For amplification of ozmP		
pYJ55	Protein expression construct containing *ozmP* from *S. albus* JA3453 between *Nde*I and *Xho*I of pET-29b(+) with an N-terminal His_6-tag.	*

8.1.6 Oligonucleotides

All DNA oligonucleotides were purchased from Eurofins and Sigma-Aldrich in lyophilised form. Diluted stocks (10 µM) were stored at -20 °C. DNA oligonucleotides used in this work are listed in Table 8.5.

Table 8.5. DNA oligonucleotides used in this work. Restriction sites are underlined. F, forward primer; R, reverse primer; U, for upstream flanking region; D, for downstream flanking region.

Primer	Sequence (5'-3')
For in-frame deletions	
pYH7_avaR3_UF	TGATCAAGGCGAATACTTCATATGCCCGCGTTGCGCATCGAAC
pYH7_avaR3_UR	CGGTCTGTTATGGTCGTGCGCGCAGCGCTCGCTCCTGTTTGAC
pYH7_avaR3_DF	GTCAAACAGGAGCGAGCGCTGCGCGCACGACCATAACAGACCG
pYH7_avaR3_DR	CCGCGCGGTCGATCCCCGCATATGGGTCTGCAACGTGGTGAGGG
pYH7_orf+5_UF	TGATCAAGGCGAATACTTCATATG CCGCTGCCGTGACCGTCAC
pYH7_orf+5_UR	GATCGCCAGGGTGCCGCCCCAGGTCATGGCCTGCCGGTA
pYH7_orf+5_DF	TACCGGCAGGCCATGACCTGGGGCGGCACCCTGGCGATC
pYH7_orf+5_DR	CCGCGCGGTCGATCCCCGCATATG GGTCGCCGCCTCCGGTC
pYH7_orf0034_UF	TGATCAAGGCGAATACTTCATATGGTCCTCTTGTCTCTGTGCCGGTTG
pYH7_orf0034_UR	CGGTAGCACATGACCGAGGTGGTCCTGGAGGCCTGG
pYH7_orf0034_DF	GACCACCTCGGTCATGTGCTACCGCGCCTCCTGGTC
pYH7_orf0034_DR	CCGCGCGGTCGATCCCCGCATATGGTTCCGGTGTCCCGGCCGTTGGTG
pYH7_orf3515_UF	TGATCAAGGCGAATACTTCATATGCGAGCACGAGATCCTCGCCATGAAG
pYH7_orf3515_UR	GTTCAGCTTGCCCGAGTCGTGCGCGAAATAGAACAGG
pYH7_orf3515_DF	GCGCACGACTCGGGCAAGCTGAACGAGACCATCTCCCGA
pYH7_orf3515_DR	CCGCGCGGTCGATCCCCGCATATGACCCTGTTGCACGTGGAGGTTCC
pYH7_phoxB_UF	TGATCAAGGCGAATACTTCATATG CCAGCGGTCCGGTCGCTC
pYH7_phoxB_UR	CGGTCCGTGCCCTCGCCGAGTCGCGCTGTCGCTGGTGG
pYH7_phoxB_DF	CCACCAGCGACAGCGCGACTCGGCGAGGGCACGGACCG
pYH7_phoxB_DR	CCGCGCGGTCGATCCCCGCATATG GCGATGACGCCGCGCGAG
pYH7_phoxF_UF	TGATCAAGGCGAATACTTCATATGCATGTTCCACCGCCGGTCTCACAG
pYH7_phoxF_UR	GCAAGCGCATGA GATGACACCACGACCGCTCTTTGC
pYH7_phoxF_DF	CGTGGTGTCATC TCATGCGCTTGCCTCCACGGTCTC
pYH7_phoxF_DR	CCGCGCGGTCGATCCCCGCATATGGTGTCTCCGGACCTGGGTCACATG

pYH7_phoxG_UF	TGATCAAGGCGAATACTTCATATG GGAGAGCGATGGCTGTCAGGATGA
pYH7_phoxG_UR	GTTTCATGCGCT CATCGGTGTCCG TCCCAGAAAGAG
pYH7_phoxG_DF	CGGACACCGATG AGCGCATGAAAC CCACCACCGACA
pYH7_phoxG_DR	CCGCGCGGTCGATCCCCGCATATG CGAGACGCTCGCTTCGACGATCTT
pYH7_phoxI_UF	TGATCAAGGCGAATACTTCATATG ATTCACAAATGGCAGCGCACGGTT
pYH7_phoxI_UR	CCCTCTGCCTCA GGTCATCGTTTT TCCGTCCCAAGA
pYH7_phoxI_DF	AAAACGATGACC TGAGGCAGAGGG ACGACGATGGGC
pYH7_phoxI_DR	CCGCGCGGTCGATCCCCGCATATG TCGGGGTGCAGGACGAGCAGGTGC
pYH7_phoxJ_UF	TGATCAAGGCGAATACTTCATATG CCGTGGACAGCATGTGCACCTCCT
pYH7_phoxJ_UR	CGAAGGGATTGG CGTCGTCCCTCT GCCTCAGCTCTC
pYH7_phoxJ_DF	AGAGGGACGACG CCAATCCCTTCG ACCGACCGGACA
pYH7_phoxJ_DR	CCGCGCGGTCGATCCCCGCATATG TCACCACCGCGAATTCCTTGTGCA
pYH7_phoxK_UF	TGATCAAGGCGAATACTTCATATG AGCTTCCGCTCCAGCTGATCTTCA
pYH7_phoxK_UR	CCCGTTCGGTCA GTTGGTGTCTCT CCTTTTGCTGCG
pYH7_phoxK_DF	AGAGACACCAAC TGACCGAACGGG CGGCACTCATTT
pYH7_phoxK_DR	CCGCGCGGTCGATCCCCGCATATG CGATCATGATGATGTTGGCGTCCT
pYH7_phoxP_UF	TGATCAAGGCGAATACTTCATATGACCGCCGACCTGCTCGTACCCGAG
pYH7_phoxP_UR	CTGCTCGGTGTC TCATGGGCGCCGCCCCCTGCGGTC
pYH7_phoxP_DF	CGGCGCCCATGA GACACCGAGCAGGAGAACCCACAC
pYH7_phoxP_DR	CCGCGCGGTCGATCCCCGCATATGGGCGAGGGCGTACTCGTGGCAGCA
pYH7_phoxR_UF	TGATCAAGGCGAATACTTCATATG CGCCGGAGCGGTCGTCCTC
pYH7_phoxR_UR	GTGCGTCGGGCGCATCGGAGGCCAGCAGGTCCAGCAGCAG
pYH7_phoxR_DF	CTGCTGCTGGACCTGCTGGCCTCCGATGCGCCCGACGCAC
pYH7_phoxR_DR	CCGCGCGGTCGATCCCCGCATATG GCTCGACGGGCCGCTCCTC
pYH7_phoxV_UF	TGATCAAGGCGAATACTTCATATG GGCTCGCTCGCCGAGCTG
pYH7_phoxV_UR	CGTCGTGCGCGGTGCTGGGCAGGCGGCGGGCCTGCC
pYH7_phoxV_DF	CGGCAGGCCCGCCGCCTGCCCAGCACCGCGCACGACG
pYH7_phoxV_DR	CCGCGCGGTCGATCCCCGCATATG AGGCGGGTGTCGCCGGAGG

For point mutagenesis

pYH7_ACP13_UF	TGATCAAGGCGAATACTTCATATGGCCGGCTCCTTCGGCAGCTTCG
pYH7_ACP13_UR	TCTTCATGCCGCTGAGGGCGTCGAAGCCGTAGTCG
pYH7_ACP13_DF	CGACTACGGCTTCGACGCCCTCAGCGGCATGAAGA
pYH7_ACP13_DR	CCGCGCGGTCGATCCCCGCATATGGACCGCGACCGCGTCGGGC
pYH7_ACP15_UF	TGATCAAGGCGAATACTTCATATGCCCTGGGCTCGGTCCGGCTGC

pYH7_ACP15_UR	ACTCATGGTGATGAGGGCGTCCACGCCGTACCG
pYH7_ACP15_DF	CGGTACGGCGTGGACGCCCTCATCACCATGAGT
pYH7_ACP15_DR	CCGCGCGGTCGATCCCCGCATATGGGCGACCTGCGCCGGGGTCG
pYH7_ACP16_UF	TGATCAAGGCGAATACTTCATATGTCGACGCGCCCGACGCGAGC
pYH7_ACP16_UR	CTGGGAGACCATGATGGCGTCCATGCCGTAGTC
pYH7_ACP16_DF	GACTACGGCATGGACGCCATCATGGTCTCCCAG
pYH7_ACP16_DR	CCGCGCGGTCGATCCCCGCATATGCCCGGTGGCCGCGGTCATCG
pYH7_PCP14_UF	TGATCAAGGCGAATACTTCATATGGGCGCGCCTCGGCACCCAC
pYH7_PCP14_UR	CGGTGAATCCGACGGCGTTCATGCCGGCCTC
pYH7_PCP14_DF	GAGGCCGGCATGAACGCCGTCGGATTCACCG
pYH7_PCP14_DR	CCGCGCGGTCGATCCCCGCATATGCGACGACGGCGAGCCGGTGC

For complementation of in-frame deletions

pIB139_phoxP_F	ATAATCATATGACCGGCCGGCCCGA
pIB139_phoxP_R	TTATATCTAGATCAGCTCAGCCGGATGCCGA
pIB139_congE_F	ATAATCATATGCGCGTGTGCCAGATAT
pIB139_congE_R	TTATATCTAGATTAGAGGGTGATGCCAAGA
pIB139_ozmP_F	ATAATCATATGGAGCCGACCATCGTGCGCT
pIB139_ozmP_R	TTATATCTAGATTAGCCCACCTTGATGCCGA

For overexpression of (putative) regulatory genes

pIB139_orf+10_F	AATCATATGGGAGAGCCTGCGTTGCG
pIB139_orf+10_R	TTATCTAGATCAGCGGGCGGGCCTC
pIB139_orf+11_F	AATCATATGTACGACCCCTCCCAC
pIB139_orf+11_R	TTATCTAGACTATCCGAGGCCCGCC
pIB139_phoxC_F	AATCATATGTGAGCGAGCGGCGGGCCGCA
pIB139_phoxC_R	TTATCTAGATCGTCGCAGGTCGTCGCCGG
pIB139_phoxT_F	AATCATATGACCGCACTCGCGC
pIB139_phoxT_R	TTATCTAGATCAGGGCAGCAGCGTGTG

For protein expression

pET28a_folD_F	ATAATCATATGGCAGCAAAGATTATTGAC
pET28a_folD_R	TATTACTCGAGTTACTCATCCTGTGGATCATG
pET28a_phoxN(Val1-Gly1483)_F	ATAATCATATGCATCACCATCACCATCACGTGAACAACCCGGTTCCGGT
pET28a_phoxN(Val1-Gly1483)_R	TTATAAAGCTTATGCGGACGTGACGGGTG
pET28a_phoxO_F	AATCATATGACTGCCGAGCCGTTCAGCTGT

pET28a_phoxO_R	TTACTCGAGCCCCCTGGTGATACGGACGGATCG
pET29b_phoxP_F	ATAATCATATGCATCACCATCACCATCACACCGGCCGGCCCGACGCCA
pET29b_phoxP_R	TTATACTCGAGTCAGCTCAGCCGGATGCCGAG
pET29b_phoxQ_F	ATAATCATATGCATCACCATCACCATCACACGCTTTCCGACTCCGACG
pET29b_phoxQ_R	TTATAAAGCTTTCAGGCCTGGCCGGTGTGGCG
pET28a_phoxS_F	ATAATCATATGCTGGGACACGTACCGATGAC
pET28a_phoxS_R	TTATACTCGAGAGTACCTTCACACGCTGCT

8.2 Microbiological methods

8.2.1 Growth and maintenance of *E. coli* strains

All *E. coli* strains were grown on solid or in liquid 2TY or LB medium at 37 ºC, with antibiotics where appropriate. Liquid cultures were aerated by shaking at 220 RPM. For long-term storage, liquid cultures suspended in 25 % (v/v) glycerol were kept at -80 ºC.

8.2.2 Growth and maintenance of *Streptomyces* strains

All *Streptomyces* strains were grown in liquid TSBY, on solid or in liquid SFM at 30 ºC, with antibiotics where appropriate. Liquid cultures were aerated in baffled flasks or flasks containing a coiled spring and by shaking at 220 RPM. For long-term storage, mycelia or spores suspended in 25 % (v/v) glycerol were kept at -80 ºC.

8.3 Molecular biology methods

8.3.1 Isolation of genomic DNA from *Streptomyces* strains

The *Streptomyces* strains were cultured for 2 days for the isolation of genomic DNA. The cells from 500 µL of culture were harvested by centrifugation at 10,000 x*g* for 10 minutes. The cell pellets were suspended in 500 µL of SET buffer (75 mM NaCl; 25 mM EDTA, pH 8.0; 20 mM Tris-HCl; in MQ H_2O; pH 7.5), added with 0.5 mg of lysozyme and incubated at 37 °C for 30 minutes. The mixtures were added with 0.2 mg of proteinase K and 60 µL of 10 % SDS and incubated at 55 °C for 2 hours. After cooling down the mixtures to room temperature (21 °C), 200 µL of 5 M NaCl was added and the mixtures were mixed by inversion. Subsequently, 500 µL of chloroform was added and the mixtures were again mixed by inversion, followed by centrifugation at 10,000 x*g* for 10 minutes. The upper phase was transferred to fresh Eppendorf tubes, added with 600 µL of isopropanol and mixed by inversion. The mixtures were centrifuged at 10,000 x*g* for 10 minutes. The resulting pellets of genomic DNA were washed with 200 µL of ethanol twice and dried at 37 °C. The dried pellets were left to dissolve in 200 µL of MQ H_2O at 4 °C overnight.

8.3.2 Isolation of plasmid DNA from *E. coli*

Plasmid DNA was isolated from overnight *E. coli* liquid cultures using the GeneJET Plasmid Midiprep Kit and eluted in MQ H_2O.

8.3.3 Polymerase chain reaction (PCR)

PCR amplifications were performed using the Geneamp 9700 PCR thermal cycler (Applied Biosystems). For cloning, the reactions contained 10 µL of 5X Q5 Reaction Buffer, 10 µL of 5X Q5 High GC Enhancer, 2.5 µL of 10 mM dNTPs, 2.5 µL of each 10 µM primer stock solution, 1 µL of DNA template, 0.5 µL of Q5® High-Fidelity DNA Polymerase and MQ H_2O to adjust to a final volume of 50 µL. For screening, the reactions contained 10 µL of MyTaq™ Red Mix, 1 µL of each 10 µM primer stock solution, 1 µL of DMSO, 1 µL of DNA template and MQ H_2O to adjust to a final volume of 50 µL. Thermocycling conditions are shown in Table 8.6.

Table 8.6. Thermocycling conditions for PCR amplifications.

Step	Temperature	Time	
Initial denaturation	98 °C	30 seconds	

Denaturation	98 °C	10 seconds	30 cycles
Annealing	66-72 °C	30 seconds	
Extension	72 °C	30 seconds/kbp	
Final extension	72 °C	2 minutes	
Hold	4 °C		

8.3.4 Agarose gel electrophoresis

Agarose gels were prepared with 0.8 to 1.2 % (w/v) agarose in TAE buffer (0.45 M Tris, 0.45 M acetic acid, 0.1 M EDTA, pH 8.0; in MQ H_2O) containing 2.5 % (v/v) of SYBR® Safe DNA Gel Stain (Invitrogen). DNA samples were mixed with 6X DNA Loading Dye and loaded to the agrose gels. The GeneRuler 1 kb Plus DNA Ladder was used as a molecular size marker. The gels were run at 120 to 150 V for 20 to 60 minutes. DNA was visualised under a UV illuminator at 254 nm and purified using the GeneJET Gel Extraction Kit and the DNA Clean & Concentrator-5 Kit (Zymo Research).

8.3.5 Restriction cloning

Restriction enzyme reactions contained 2 µL of 10X FastDigest (Green) Buffer, 1 µL of (each) restriction enzyme, up to 1 µg of DNA and MQ H_2O to adjust to a final volume of 20 µL. The reactions were incubated at 37 °C for 1 hour and subsequently heat-inactivated. To prevent the self-ligation of linearised plasmids, 1 µL of shrimp alkaline phosphatase was added directly to the completed digestion reactions and incubated further for 30 minutes prior to heat-inactivation at 65 °C or 80 °C. The reactions were then subjected to DNA separation by agarose gel electrophoresis or direct DNA purification using the DNA Clean & Concentrator-5 Kit (Zymo Research).

DNA ligation reactions contained 50 ng linearised and dephosphorylated plasmid with two- to five-fold excess of (each) DNA insert, 2 µL of 10X T4 DNA Ligase Buffer, 1 µL of T4 DNA Ligase and MQ H_2O to adjust to a final volume of 20 µL. Following incubation at 16 °C overnight, the reactions were heat-inactivated at 65 °C and desalted by micro-dialysis using mixed cellulose esters membranes (0.025 µm pore size) (MF-Millipore™) prior to transformation of chemically-competent *E. coli* DH10B cells with the entire reaction volume.

8.3.6 Isothermal DNA assembly

Linearisation of plasmids was performed as described in Section 8.3.5.

Isothermal assembly reactions were prepared on ice, and contained 50 ng of linearised and dephosphorylated plasmid with three-fold excess of each DNA insert, 10 µL of 2X NEBuilder® HiFi DNA Assembly Master Mix and MQ H_2O to adjust to a final volume of 20 µL. The reactions were incubated at 50 °C for 60 minutes and subsequently desalted by micro-dialysis using mixed cellulose esters membranes (0.025 µm pore size) (MF-Millipore™) prior to transformation of chemically-competent *E. coli* DH10B cells with the entire reaction volume.

8.3.7 Preparation and transformation of chemically-competent *E. coli* cells

For the preparation of chemically-competent *E. coli* cells, glycerol stocks were streaked on 2TY agar and incubated at 37 °C overnight. A single colony was inoculated into 5 mL of 2TY and incubated at 37 °C overnight. In 250 mL flasks, 1 mL of overnight culture was inoculated into 25 mL of 2TY and incubated at 37 °C until the cultures reached an OD_{600} of 0.3 to 0.5. The cells were harvested by centrifugation at 4 °C, 4,000 x*g* for 10 minutes. The cell pellets were resuspended with 50 mL of ice-cold buffer (50 mM CaCl2, 10 mM KAc, in MQ H_2O; pH 6.2) and incubated on ice with occasional agitation for 1 hour. The cells were centrifuged at 4 °C, 4,000 x*g* for 10 minutes. The cell pellets were resuspended in 2 mL of ice-cold buffer (50 mM $CaCl_2$, 10 mM KAc, 20 % (v/v) glycerol; in MQ H_2O; pH 6.2). Aliquots of the cell suspensions were snap-frozen with dry ice and stored at -80 °C.

For transformation of chemically-competent *E. coli* cells, aliquots of the cells were thawed in ice slurry and added with either 100 ng of purified plasmid or the entire volume of a micro-dialysed DNA assembly reaction. The mixtures were incubated on ice for 30 minutes, then heat-shocked at 42 °C for 1 minute and incubated on ice again for 2 minutes. The mixtures were added with 500 µL of 2TY and incubated at 37 °C for 1 hour. The cultures were plated on LB agar containing appropriate antibiotics and incubated at 37 °C overnight.

8.3.8 DNA conjugation into *Streptomyces* sp. KO-7888

Streptomyces sp. KO-7888 was inoculated in 25 mL of TSBY in 250 mL flasks containing a coiled spring and incubated at 30 °C for 2 days. The cells were harvested by centrifugation at 4 °C, 4,000 x*g* for 10 minutes. The cell pellets were resuspended in 20 mL of 2TY and centrifuged at 4 °C, 4,000 x*g* for 10 minutes. The final cell pellets were resuspended in 1 mL of 2TY and kept on ice. A single colony of *E. coli* ET12567/pUZ8002 bearing a pYH7 deletion or pIB139 complementation

plasmid was inoculated into 5 mL of 2TY with the appropriate antibiotics and incubated at 37 °C overnight. In 50 mL Falcon tubes, 500 µL of overnight cultures were inoculated into 10 mL of 2TY with antibiotics at half of the usual working concentrations and incubated at 37 °C until the cultures reached an OD_{600} of 0.3 – 0.5. The cells were harvested by centrifugation 4 °C, 4,000 x *g* for 10 minutes. The cell pellets were resuspended in 20 mL of 2TY and centrifuged at 4 °C, 4,000 x *g* for 10 minutes. This resuspension and centrifugation step was repeated for three times. The final cell pellets were resuspended in 500 µL of 2TY and kept on ice. For conjugation, 100 µL of *Streptomyces* sp. KO-7888 suspension and 100 µL of *E. coli* suspension were mixed, plated on 15 mL of SFM agar containing 10 mM $MgCl_2$ and incubated at 30 °C for 15 to 20 hours.

Following 15 to 20 hours of incubation, each plate was flooded with 2 mL of antibiotic solution containing 0.75 mg of Apr and 0.375 mg of Nal in autoclaved, MQ H_2O and incubated at 30 °C for 7 days for exconjugants to emerge. To confirm the presence of the plasmid, exconjugants were streaked on SFM agar with antibiotics. To obtain deletion mutants, antibiotic-resistant exconjugants were subject to two to four rounds of growth on SFM agar without antibiotics, to facilitate the double crossover via homologous recombination and the subsequent loss of the pYH7 deletion plasmid. Genomic DNA of antibiotic-sensitive colonies was extracted, screened by PCR and confirmed by DNA sequencing.

8.3.9 DNA sequencing

Sanger DNA sequencing and whole-genome sequencing were carried out using the 3730xl DNA Analyser (Applied Biosystems) and the MiSeq system (Illumina) respectively, at the DNA Sequencing Facility, Department of Biochemistry.

8.4 Protein expression and purification methods

8.4.1 Protein expression in *E. coli* BL21(DE3)

Protein expression constructs were introduced to *E. coli* BL21(DE3) by heat shock transformation. To prepare starter cultures, single colonies were inoculated to 10 mL LB medium with appropriate antibiotics in 50 mL Falcon tubes and incubated at 37 ºC 220 RPM overnight. The entire starter cultures were inoculated to 400 mL LB with appropriate antibiotics in 2 L conical flasks and incubated at 37 ºC 220 RPM until OD_{600} reached 0.8 to 1.0. Protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to the cultures at a final concentration of 0.5 mM. Following incubation at 16 ºC for another 16 hours, the cultures were centrifuged at 4,000 x*g* for 10 minutes. The cell pellets were suspended in 20 mL 1X binding buffer (0.5 M NaCl, 20 mM Tris-HCl; in MQ H_2O; pH 7.9) and sonicated on ice slurry, pulsed on for 2 seconds and off for 5.9 seconds, for a total sonication time of 4 minutes using the Vibra-CellTM system (Sonics & Materials Inc.). Cell lysate was centrifuged at 18000 x*g* for 30 minutes and the supernatant containing soluble proteins was subjected to purification by nickel affinity chromatography.

8.4.2 Protein expression in *S. coelicolor* CH999

To prepare starter cultures, 100 µL of mycelia suspension from glycerol stocks of *S. coelicolor* CH999 bearing the expression construct pCJW93-*ajiH* was inoculated to 30 mL of TSBY medium with appropriate antibiotics in 250 mL conical flasks and incubated at 28 ºC 220 RPM for 4 days. Two millilitres of starter culture were inoculated to 100 mL Super YEME medium with appropriate antibiotics in 500 mL conical flasks with a coil spring and incubated at 28 ºC 220 RPM for 2 days. Protein expression was induced by the addition of thiostrepton to the cultures at a final concentration of 5 µg mL^{-1}. Following incubation for another two days, the cultures were centrifuged at 18,000 x*g* for 30 minutes. The cell pellets were suspended in 50 mL of 1X binding buffer (0.5 M NaCl, 20 mM Tris-HCl; in MQ H_2O; pH 7.9) and sonicated on ice slurry, pulsed on for 5.9 seconds and off for 5.9 seconds, for a total sonication time of 4 minutes using the Vibra-Cell™ system (Sonics & Materials Inc.). Cell lysate was centrifuged at 18,000 x*g* for 30 minutes at 4 ºC and the supernatant containing soluble proteins was subjected to purification by nickel affinity chromatography.

8.4.3 Nickel affinity chromatography

The supernatant of cell lysate was loaded to columns containing 500 µL bed volume of His-bind resin (Bio-Rad) charged with 5 mL 50 mM $NiCl_2$ and equilibrated with one column volume of 1X binding buffer. After the supernatant passed through, the column was washed with 10 mL 1X binding buffer. Proteins were eluted with 1X elution buffer (1X binding buffer containing 10 to 300 mM imidazole), 2 to 4 mL of buffer at each concentration. Elution fractions containing the protein of interest were combined and concentrated using the Amicon® Ultra 4 mL Centrifugal Filter Units (Millipore) with a molecular weight cut-off of 3, 10, 30 or 100 kDa. Buffer exchange was performed in the filter units with 1X binding buffer until the estimated imidazole concentration fell below 0.1 mM. All protein purification steps were carried out at 4 ºC. Purified proteins were stored in 20 % (v/v) glycerol and stored at 4 ºC.

8.4.4 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples (7.5 µL) mixed with 2.5 µL of NuPAGE® 4X LDS Sample buffer were loaded to NuPAGE® Novex® 4 – 12 % Bis-Tris gels. The Precision Plus Protein™ Dual Color Standards (Bio-Rad) were used as a molecular weight marker. The gels were run in 1X NuPAGE® MES Buffer at 200 V for 40 minutes. Proteins were visualised by staining the gels with InstantBlue (Expedeon) for 30 to 60 minutes.

8.5 Enzymatic assay methods

8.5.1 *In vitro* reconstitution of the early steps of phthoxazolin biosynthesis

All *in vitro* enzymatic reactions were performed in 25 mM Tris-HCl buffer (pH 7.5) with a total volume of 500 µL. All reactions contained 5 mM ATP, 10 mM $MgCl_2$, 20 µM Sfp, 0.25 CoA, 0.25 mM malonyl-CoA, 0.5 mM NADPH, 1 mM Gly or Ser, purified or cell lysate containing PhoxO (2 µM), PhoxQ (5 µM), PhoxN (Val1-Gly1483) (1 µM), PhoxS (1 µM) and PhoxP (1 µM) and 25 % (v/v) of the N^{10}-fH_4F chemoenzymatic synthesis reaction (see Section 8.6.1). Following incubation at 25 °C for 1 hour, the proteins were precipitated by the addition of 4.5 mL of 10 % (v/v) TFA. The precipitated reactions were centrifuged at 4000 x*g* at 4 °C and the pellets were washed with 2 mL 3:1 (v/v) diethyl ether/ethanol. The solvent was dried at 30 °C using the Eppendorf™ Vacufuge™ Concentrator. Thioester cleavage was carried out by the addition of 500 µL of 0.1 M KOH. The reactions were incubated at 70 °C for 15 minutes, subsequently mixed with 2 mL of methanol and centrifuged at 4000 x*g*. The supernatant was dried at 45 °C using the Eppendorf™ Vacufuge™ Concentrator and subjected to liquid chromatography-mass spectrometry (LC-MS) analysis. Reactions that did not contain amino acids or PhoxP were used as negative controls. All reactions were performed in duplicate.

8.5.2 *In vitro* AjiA1 reactions

All *in vitro* AjiA1 reactions were performed in 25 mM HEPES buffer (pH 8.1). For the isolation of the 3-HAA amide dimer, the reactions contained 10 µM AjiA1, 5 mM ATP, 10 mM $MgCl_2$, 0.25 mM 3-hydroxyanthranilic acid (3-HAA) and MQ H_2O to adjust to a final volume of 1 to 4 mL. For the testing of alternative substrates, the reactions contained 10 µM AjiA1, 5 mM ATP, 10 mM $MgCl_2$, either 0.25 mM 3-hydroxybenzoic acid or 0.25 mM 2-amino-3-methylbenzoic acid and MQ H_2O to adjust to a final volume of 100 µL. Following incubation at 21 °C overnight, the reactions were inactivated by the addition of TFA to a final concentration of 0.5 % (v/v) and subjected to extraction and analysis. Reactions which did not contain AjiA1 were used as negative controls. All reactions were performed in duplicate.

8.5.3 *In vitro* reconstitution of (individual steps of) AJI9561 biosynthesis

All *in vitro* enzymatic reactions were performed in 25 mM HEPES buffer (pH 8.1) with a total reaction volume of 100 µL, unless stated otherwise. All reactions contained 5 mM ATP, 10 mM $MgCl_2$, both or either of 0.25 mM 3-HAA and 0.5 mM 6-

methylsalicylic acid (6-MSA), and one or more of 20 μM AjiA2, 60 μM AjiP, 10 μM AjiA1, 1 μM AjiH and 20 μM AjiG. Following incubation at 21 °C overnight, the reactions were inactivated by the addition of TFA to a final concentration of 0.5 % (v/v), then subjected to extraction and analysis. Reactions which did not contain the enzyme(s) of interest were used as negative controls. All reactions were performed in duplicate.

8.6 Extraction methods

8.6.1 Extraction of *Streptomyces* spp. liquid cultures or agar plates

Liquid cultures of *Streptomyces* spp. were extracted twice with equal volume of ethyl acetate. Agar plates (5 mL) of *Streptomyces* spp. were cut into pieces (1 cm x 1 cm) and incubated in 15 mL of ethyl acetate for 1 hour. The organic phases were dried and subjected to LC-MS analysis.

8.6.2 Extraction of *in vitro* enzymatic reactions

All *in vitro* enzymatic reactions were extracted twice with equal volume of ethyl acetate. The organic phases were dried and subjected to LC-MS analysis.

8.7 Chemoenzymatic and chemical synthesis methods

8.7.1 Chemoenzymatic synthesis of N^{10}-formyltetrahydrofolate (N^{10}-fH_4F)

The N^{10}-fH_4F chemoenzymatic synthesis reaction contained 20 µM FolD, 500 µM H_4F, 8 mM NAD^+ and 2 mM HCHO in potassium phosphate buffer (120 mM K3PO4, 50 mM β-mercaptoethanol; in MQ H_2O; pH 6.0), as described in Schönafinger *et al*. (2006). A reaction which did not contain FolD was used as the negative control. The reactions were monitored by UV spectroscopy at 340 nm.

8.7.2 Chemical synthesis of 6-MSA

Ethyl-6-MSA (300 mg) was refluxed at 120 °C in 15 mL of 1 M NaOH for 2 hours. The reaction was analysed by thin-layer chromatography (TLC) using 1:1 (v/v) THF/isopropanol as the liquid solvent phase and potassium permanganate solution for staining, and visualised under a UV illuminator at 254 nm. The completed reaction was incubated on ice and added with sulphuric acid to a final concentration of 10% (v/v). The precipitated product was filtered, dried by lyophilisation and dissolved in methanol.

8.8 Analytical methods

8.8.1 Liquid chromatography-mass spectrometry (LC-MS)

LC-MS analysis was performed using the Agilent 1200 Series LC system and the MAT LTQ mass spectrometer (Thermo/Finnigan) fitted with an electrospray ionisation source, at the positive ionisation mode. The system was fitted with a Prodigy® C18 Column (4.6 x 250 mm, 5 µm) (Phenomenex) and a Jupiter® C4 Column (4.6 x 250 mm, 5 µm) (Phenomenex) for the analysis of extracts and protein samples, respectively.

For the analysis of *Streptomyces* spp. extracts and *in vitro* enzymatic reactions regarding phthoxazolin biosynthesis, a solvent system containing MQ H_2O with 0.1 % (v/v) formic acid (FA) and acetonitrile with 0.1 % (v/v) formic acid, with a linear gradient of 95 to 5 % MQ H_2O with 0.1 % (v/v) FA at a flow rate of 0.7 mL min^{-1} over 30 minutes, was used. M/z values ranging from 100 to 1000 were scanned and m/z values of interest were subjected to MS/MS fragmentation at a normalised collision energy of 30 %. UV absorption was also monitored.

For the analysis of *in vitro* enzymatic reactions regarding AJI9561 biosynthesis, a solvent system containing MQ H_2O with 0.1 % (v/v) TFA and acetonitrile with 0.1 % (v/v) TFA, with a linear gradient of 95 to 5 % MQ H_2O with 0.1 % (v/v) TFA at a flow rate of 0.7 mL min^{-1} over 45 minutes, was used. M/z values ranging from 100 to 1000 were scanned and m/z values of interest were subjected to MS/MS fragmentation at a normalised collision energy of 30 %. UV absorption was also monitored.

LC-MS and LC-MS/MS data was processed (and deconvoluted) using Xcalibur (Thermo/Finnigan).

8.8.2 Accurate mass analysis

Accurate mass analysis was performed using Vion IMS-MS-QTOF in the positive ionisation mode at the Mass Spectrometry Facility, Department of Chemistry.

8.9 Bioinformatic methods

8.9.1 Genome annotation

Genome sequences were visualised and annotated using Artemis 16.0 (Rutherford *et al*., 2000). AntiSMASH 4.0 was used to detect biosynthetic gene clusters and PKS and NRPS enzymatic domains (Blin *et al*., 2017). Putative function of gene products was predicted using protein-BLAST (Altschul *et al*., 1990).

8.9.2 Multiple sequence alignments

Multiple protein sequence alignments were performed using Clustal Omega (Sievers *et al*., 2011), unless stated otherwise.

8.9.3 Phylogenetic analysis

DNA or protein sequences were aligned with MUSLE (Edgar, 2004) and the phylogenetic trees were created by the Maximum-likelihood Method using MEGA7 (Kumar *et al*., 2016).

8.9.4 *In silico* modelling of protein structures

In silico modelling of protein structures and prediction of active site pocket were performed using Phyre2 and RaptorX (Kelley *et al*., 2015; Källberg *et al*., 2012). Edu MacPyMOL (Schrödinger) was used to visualise structural models.

Appendix

Table S1. Properties of genes encoded within cluster 1 (coelichelin) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
0133R	315	98/99	*Streptomyces* sp. NRRL WC-3753	formyltrans-ferase CchA	EHN76207.1
0132R	451	98/99	*Streptomyces coelicoflavus*	L-ornithine mono-oxygenase CchB	EHN77998.1
0131R	328	99/99	*Streptomyces coelicoflavus*	iron transporter permease CchC	EHN77999.1
0130R	354	99/99	*Streptomyces coelicoflavus*	iron transporter permease CchD	EHN78000.1
0129R	287	95/97	*Streptomyces coelicoflavus*	ABC iron transporter CchE	EHN78001.1
0128R	350	98/99	*Streptomyces coelicoflavus*	iron-siderophore binding protein CchF	EHN78002.1
0127R	597	98/99	*Streptomyces coelicoflavus*	ABC transporter CchG	EHN78003.1
0126R	3672	96/97	*Streptomyces coelicoflavus*	NRPS (L-fh-Orn, hOrn, L-Thr) CchH	EHN78004.1
0125	566	99/99	*Streptomyces coelicoflavus*	ABC transporter CchI	EHN78005.1
0122R	70	94/97	*Streptomyces* sp. NRRL F4711	MbtH-like protein CchK	KOT96041.1

Table S2. Properties of genes encoded within cluster 2 (bacteriocin) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
0394	819	98/98	*Streptomyces* sp. NRRL WC-3753	PAS regulator	KPC76431.1
0395	408	98/99	*Streptomyces* sp. NRRL WC-3753	proteinase	KPC76432.1
0396	168	—	—	—	—
0397	71	—	—	—	—
0398	268	85/92	*Streptomyces azureus*	secreted protein	GAP45380.1

0399	720	99/99	*Streptomyces coelicoflavus*	ABC transporter	EHN78272.1
0400	937	99/99	*Streptomyces coelicoflavus*	ABC transporter	EHN78271.1

Table S3. Properties of genes encoded within cluster 3 (hopene/ATBH) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
0721	265	96/98	*Streptomyces* sp. NRRL WC-3753	protein phosphatase	KPC72758.1
0722R	39	89/97	*Streptomyces aureofaciens*	small hydrophobic protein	WP_052842011
0723R	204	99/99	*Streptomyces* sp. NRRL WC-3753	Xre regulator	KPC72756.1
0724R	461	99/100	*Streptomyces coelicoflavus*	aminotransfer-ase	EHN78561.1
0725R	650	98/99	*Streptomyces coelicoflavus*	DXP synthase	EHN78560.1
0726R	384	99/99	*Streptomyces coelicoflavus*	4-hydroxy-3-methyl-but--2-enyl diphosphate synthase	EHN78559.1
0727R	340	99/99	*Streptomyces* sp. FXJ7.023	radical SAM protein HpnH	WP_037772613
0728R	213	99/100	*Streptomyces coelicolor*	1-hydroxy-2-methyl-but-2-enyl 4-diphosphate reductase	CAB39698.1
0729R	684	98/99	*Streptomyces* sp. NRRL WC-3753	squalene-hopene cyclase	KPC72750.1
0730R	354	99/100	*Streptomyces coelicoflavus*	geranylgeranyl diphosphate synthase	EHN78555.1
0731R	478	99/99	*Streptomyces* sp. NRRL WC-3753	phytoene dehydrogenase	KPC72748.1
0732R	312	98/99	*Streptomyces coelicoflavus*	phytoene synthase	WP_042822674
0733R	300	99/99	*Streptomyces coelicoflavus*	squalene synthase	WP_042822672_

Table S4. Properties of genes encoded within cluster 5a (butyrolactone) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
1130	355	36/48	*Streptomyces avermitilis*	gamma butyrolactone synthase	BAC69980.1

Table S5. Properties of genes encoded within cluster 6b (NIS) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
1205	126	—	—	—	—
1206	506	97/98	*Streptomyces* sp. NRRL WC-3753	IucA-IucC siderophore synthetase	KPC87683.1
1207	516	94/96	*Streptomyces* sp. NRRL WC-3753	iron transporter	KPC87684.1
1208	421	94/96	*Streptomyces* sp. NRRL WC-3753	Von Willebrand Factor type A protein	KPC87685.1
1209	262	77/87	*Streptomyces sviceus*	secreted protein	EDY55527.1
1210R	218	92/94	*Streptomyces lividans*	HxlR regulator	EOY51317.1
1211	402	99/99	*Streptomyces coelicolflavus*	aminotransfer-ase	EHN77626.1
1212	199	99/100	*Streptomyces coelicoflavus*	phosphoglycer-ate mutase	EHN77625.1
1213	220	—	—	—	—

Table S6. Properties of genes encoded within cluster 7 (geosmin) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
1389	251	97/98	*Streptomyces* sp. NRRL WC-3753	phosphodiester-ase	KPC71997.1
1390R	193	95/96	*Streptomyces* sp. NRRL WC-3753	membrane protein	KPC71996.1
1391R	740	97/98	*Streptomyces* sp. NRRL WC-3753	geosmin synthase	KPC71995.1
1392R	405	75/83	*Streptomyces*	UPF0364 protein	GAP48695.1

			azureus		
1393R	215	99/99	*Streptomyces* sp. NRRL WC-3753	A-factor receptor	KPC71993.1
1394R	337	95/96	*Streptomyces lividans*	sugar ABC transporter	EFD65917.1
1395	340	95/96	*Streptomyces* sp. NRRL WC-3753	histidine kinase	KPC71991.1

Table S7. Properties of genes encoded within cluster 8 (bacteriocin) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
1412	507	98/99	*Streptomyces coelicoflavus*	iron-sulfur-binding protein	EHN75283.1
1413	292	90/93	*Streptomyces aureofaciens*	polysaccharide deacetylase	WP_052837412
1414	75	-	-	-	-
1415R	224	98/98	*Streptomyces* sp. NRRL WC-3753	ABC transporter	KPC71969.1
1416R	284	97/97	*Streptomyces coelicoflavus*	ABC transporter	EHN75332.1
1417	254	90/93	*Streptomyces* sp. FXJ7.023	peptidyl-tRNA hydrolase	WP_037762889
1418	461	93/93	*Streptomyces* sp. NRRL WC-3753	endonuclease	KPC71966.1
1419	265	93/93	*Streptomyces* sp. NRRL WC-3753	membrane protein	KPC71965.1
1420	535	98/98	*Streptomyces* sp. NRRL WC-3753	proteinase	KPC71964.1
1421R	243	93/96	*Streptomyces turgidiscabies*	chlorite dismutase	ELP64953.1
1422R	494	99/99	*Streptomyces coelicoflavus*	proto-porphyrinogen oxidase	EHN75320.1

Table S8. Properties of genes encoded within cluster 9a (isorenieratene) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
1660	389	83/87	*Streptomyces* sp.	geranylgeranyl	WP_031020605

			NRRL WC-3753	diphosphate synthase	
1661	520	89/92	*Streptomyces* sp. CNQ-509	phytoene dehydrogenase	AKH84866.1
1662	331	90/93	*Streptomyces* sp. CNQ-525	phytoene synthase	WP_037744984
1663	261	93/95	*Streptomyces violaceorubidus*	2Fe-2S ferredoxin	WP_030182286
1664R	522	89/92	*Streptomyces* sp. NRRL WC-3753	isorenieratene synthase	KPC71993.1
1665R	244	86/90	*Streptomyces lividans*	methyltransfer-ase	EFD65917.1
1666R	394	80/86	*Streptomyces* sp. NRRL WC-3753	lycopene cyclase	KPC71991.1
1667R	236	70/76	*Streptomyces thermolilicinus*	short chain dehydrogenase	WP_028964738

Table S9. Properties of genes encoded within cluster 9b (undecylprodigiosin) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
1677R	295	98/99	*Streptomyces* sp. NRRL WC-3753	acyltransferase	KPC71499.1
1678R	226	33/58	*Kibdelosporang-ium aridum*	membrane protein	WP_033381793
1679R	406	99/99	*Streptomyces* sp. NRRL WC-3753	oxidase (MarG-like)	KPC71501.1
1680R	939	97/97	*Streptomyces* sp. NRRL WC-3753	PEP-dependent enzyme (MarH-like)	KPC71502.1
1681R	362	95/99	*Streptomyces coelicoflavus*	methyltransfer-ase	EHN75476.1
1682R	280	95/96	*Streptomyces* sp. NRRL WC-3753	thioesterase II	KPC71504.1
1683R	347	95/96	*Streptomyces* sp. NRRL WC-3753	oxidoreductase	KPC71505.1
1684R	328	98/99	*Streptomyces lividans*	fatty acyl-AMP ligase	AIJ12834.1
1685R	1882	85/88	*Streptomyces coelicolor*	modular PKS	CAA16183.1
1686R	532	96/98	*Streptomyces* sp. NRRL WC-3753	NRPS (MarM-like)	KPC87186.1
1687R	636	97/98	*Streptomyces* sp.	8-amino-7-oxo-	KPC87185.1

			NRRL WC-3753	nonanoate synthase	
1688R	97	73/85	*Streptomyces peruviensis*	ACP/PCP	WP_030050065
1689R	329	99/99	*Streptomyces* sp. NRRL WC-3753	3-ketoacyl-ACP synthase	KPC87183.1
1690	81	97/98	*Streptomyces coelicoflavus*	ACP	EHN75488.1
1691	407	99/99	*Streptomyces* sp. NRRL WC-3753	3-ketoacyl-ACP synthase	KPC87181.1
1692	146	47/60	*Streptomyces katrae*	phosphoester-ase	WP_030293828
1693	296	64/75	*Streptomyces* sp. CNQ-617	?MarT-like protein	AHF22847.1
1694	265	58/67	*Streptomyces* sp. CNQ-617	4'-phospho-pantetheine transferase	AHF22846.1
1695	378	64/75	*Streptomyces coelicoflavus*	RedV nitroreductase	EHN75483.1
1696	197	58/67	*Streptomyces coelicoflavus*	LuxR regulator	EHN75482.1
1697R	106	98/98	*Streptomyces coelicoflavus*	RedY protein	EHN75490.1
1698R	391	99/99	*Streptomyces coelicoflavus*	RedN acyl-CoA dehydrogenase	EHN75489.1
1699R	1086	94/95	*Streptomyces* sp. NRRL WC-3753	RedX PKS	KPC87173.1
1700R	268	98/99	*Streptomyces coelicoflavus*	RedD regulator	EHN79898.1

Table S10. Properties of genes encoded within cluster 10 (NIS) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
1768R	230	64/74	*Streptomyces mirabilis*	siderophore biosynthetic protein	WP_037710783
1769R	610	83/87	*Streptomyces* sp. NRRL WC-3753	iron transporter	WP_030184400
1770R	422	86/87	*Streptomyces* sp. NRRL WC-3753	2,4-diamino-butyrate 4-amino-transferase	KPC87106.1
1771	408	80/88	*Streptomyces*	secreted protein	CCK27029.1

			davawensis		
1772R	497	98/98	*Streptomyces lividans*	ATP-binding protein	EFD66294.1
1773R	512	98/98	*Streptomyces* sp. NRRL WC-3753	zinc metalloproteinase	KPC87103.1
1774R	721	96/98	*Streptomyces* sp. NRRL WC-3753	ppGpp synthetase	KPC87189.1

Table S11. Properties of genes encoded within cluster 11 (spore pigment) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
2227R	591	90/92	*Streptomyces violaceoruber*	polyketide hydroxylase	WP_037710783
2228	405	98/98	*Streptomyces* sp. NRRL WC-3753	protein in whiE locus	WP_030184400
2229	158	99/99	*Streptomyces aureofaciens*	small protein in whiE locus	WP_052840703
2230	423	99/99	*Streptomyces* sp. NRRL WC-3753	PKS KSalpha	KPC88984.1
2231	424	98/99	*Streptomyces* sp. NRRL WC-3753	PKS KSbeta	KPC88985.1
2232	89	100/100	*Streptomyces coelicoflavus*	ACP	EHN75822.1
2233	159	98/98	*Streptomyces* sp. NRRL WC-3753	polyketide cyclase	KPC88987.1
2234	111	100/100	*Streptomyces coelicoflavus*	WhiE VII	EHN75820.1
2235	299	98/99	*Streptomyces coelicoflavus*	membrane protein	EHN75819.1

Table S12. Properties of genes encoded within cluster 12 (albaflavenone) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
2315	322	90/92	*Streptomyces coelicoflavus*	AraC regulator	EHN75894.1
2316R	461	98/98	*Streptomyces lividans*	*epi*-isoizaene-5-hydroxylase	EFD66820.1
2317R	289	99/99	*Streptomyces lividans*	*epi*-isozizaene	AIJ13444.1

Table S13. Properties of genes encoded within cluster 13 (lipopeptide) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
3202	168	94/95	*Streptomyces* sp. NRRL WC-3753	membrane protein	KPC77435.1
3203R	387	98/98	*Streptomyces* sp. NRRL WC-3753	aromatic acid decarboxylase	KPC77436.1
3204	253	98/99	*Streptomyces* sp. NRRL WC-3753	Clp proteinase	KPC77437.1
3205	563	97/98	*Streptomyces* sp. NRRL WC-3753	3'-5'-exo nuclease	KPC77438.1
3206	301	96/98	*Streptomyces* sp. NRRL WC-3753	formylTHF deformylase	KPC77439.1
3207	338	99/99	*Streptomyces coelicoflavus*	glycosyltrans-ferase	EHN72132.1
3208	355	97/99	*Streptomyces* sp. NRRL WC-3753	G1P thymidylyltrans-ferase	KPC77441.1
3209	324	99/99	*Streptomyces coelicoflavus*	dTDP-glucose 4,6-dehydratase	EHN72130.1
3210	72	99/99	*Streptomyces coelicoflavus*	MbtH-like protein	EHN72129.1
3211	597	99/99	*Streptomyces coelicoflavus*	SARP regulator	EHN72128.1
3212R	262	53/66	*Streptomyces avermitilis*	4'-phosphopan-tetheine transferase	BAB69338.1
3213R	305	96/98	*Streptomyces coelicoflavus*	SyrP-like protein	EHN78774.1
3214	159	96/98	*Streptomyces coelicoflavus*	NADPH-flavin oxidoreductase	EHN78775.1
3215R	339	82/86	*Streptomyces griseoruber*	ACP	KEG39421.1
3216R	59	—	—	—	—
3217R	72	47/60	*Streptomyces* sp. NRRL WC-3753	MbtH-like protein	KPC77449.1
3218R	93	97/97	*Streptomyces* sp. NRRL WC-3753	PCP	KPC77450.1
3219R	560	51/62	*Streptomyces ipomoeae*	acyl-CoA dehydrogenase	EKX68087.1
3220R	610	64/75	*Streptomyces coelicoflavus*	acyl-CoA dehydrogenase	AKZ58687.1
3221R	593	58/67	*Streptomyces*	fatty-acyl-AMP	AKZ58686.1

			coelicoflavus	synthase	
3222R	227	98/98	*Streptomyces coelicoflavus*	RedY protein	EHN75490.1
3223R	803	99/99	*Streptomyces coelicoflavus*	RedN acyl-CoA dehydrogenase	EHN75489.1
3224R	256	94/95	*Streptomyces* sp. NRRL WC-3753	RedX PKS	KPC87173.1
3225	405	99/99	*Streptomyces* sp. NRRL WC-3753	histidine kinase	KPC77455.1
3226	202	99/99	*Streptomyces* sp. NRRL WC-3753	LuxR regulator	KPC77472.1
3227R	314	99/99	*Streptomyces coelicoflavus*	SyrP-like protein	EHN72123.1
3228R	346	100/100	*Streptomyces coelicoflavus*	monooxygenase	EHN72122.1
3229R	427	98/98	*Streptomyces coelicoflavus*	MFS transporter	EHN72121.1
3230R	3865	50/62	*Streptomyces* sp. 769	NRPS (CATCATCAMTTe)	WP_039638095
3231R	4194	52/64	*Streptomyces viridochromo-genes*	NRPS (CATCATCATCAT)	AEG64697.1
3232R	6982	85/90	*Actinospica acidiphila*	NRPS CATCTCATTECA TCATCATE)	WP_051724693
3233	276	28/45	*Rhodococcus pyridinivorans*	ABC ATPase transporter	EHK85780.1
3234R	217	95/98	*S. aureofaciens*	LuxR regulator	WP_052836389
3235R	520	57/66	*Streptomyces scopuliridis*	2-component histidine kinase	WP_051746162
3236R	409	—	—	—	—

Table S14. Properties of genes encoded within cluster 14 (NIS) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
4615R	605	99/100	*Streptomyces coelicoflavus*	glutamine amidotransfer-ase	EHN73015.1
4616R	85	—	—	—	—
4617R	297	97/98	*Streptomyces* sp. NRRL WC-3753	TAT signal sequence protein	KPC85278.1
4618R	535	98/99	*Streptomyces* sp.	beta-N-	KPC85279.1

			NRRL WC-3753	acetylhexos-aminidase	
4619R	592	98/98	*Streptomyces* sp. NRRL WC-3753	lucA-lucC siderophore synthetase	KPC85280.1
4620R	182	98/99	*Streptomyces coelicoflavus*	acetyltransfer-ase	EHN73011.1
4621R	426	98/99	*Streptomyces* sp. NRRL WC-3753	alcaligin-like biosynthesis protein	KPC85282.1
4622R	480	99/99	*Streptomyces* sp. NRRL WC-3753	decarboxylase	KPC85283.1
4623R	286	97/98	*Streptomyces* sp. NRRL WC-3753	sialic acid transporter	KPC85284.1
4624R	350	99/99	*Streptomyces coelicoflavus*	ABC transporter	EHN73007.1

Table S15. Properties of genes encoded within cluster 15 (melanin) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
4699R	478	96/97	*Streptomyces* sp. NRRL WC-3753	undecaprenyl-phosphate glucose phosphotrans-ferase	KPC85353.1
4700R	422	74/79	*Streptomyces* sp. NRRL S-1831	glycosyl transferase	WP_031084094
4701	79	97/100	*Streptomyces ambofaciens*	chaplin	AKZ55784.1
4702	254	44/53	*Streptomyces sviceus*	secreted protein	EDY61093.1
4703	100	67/73	*Streptomyces scabiei*	secreted protein	CBG72938.1
4704	171	42/56	*Streptomyces venezuelae*	secreted protein	WP_015035837
4705	210	94/94	*Streptomyces* sp. NRRL WC-3753	tyrosinase	KPC85870.1
4706	288	100/100	*Streptomyces* sp. NRRL WC-3753	tyrosinase	KPC85358.1
4707R	90	93/97	*Streptomyces*	membrane protein	WP_030866270

			violaceoruber		
4708	211	83/88	*Streptomyces lividans*	secreted protein	EOY47759.1
4709R	323	96/97	*Streptomyces* sp. NRRL WC-3753	2-hydroxyacid dehydrogenase	KPC85361.1
4710	535	96/98	*Streptomyces aureofaciens*	ATP-binding protein	WP_052838090
4711	615	95/96	*Streptomyces* sp. NRRL WC-3753	ATP-binding protein	KPC85505.1

Table S16. Properties of genes encoded within cluster 16 (ectoine) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
5562	509	98/98	*Streptomyces coelicoflavus*	NADP-dependent fatty aldehyde dehydrogenase	EHN73793.1
5563	163	94/98	*Streptomyces* sp. NRRL WC-3753	regulation of enolase 1	KPC74171.1
5564	237	97/98	*Streptomyces* sp. NRRL WC-3753	protein-disulfide isomerase	KPC74170.1
5565	364	98/98	*Streptomyces* sp. NRRL WC-3753	beta-N-acetylhexos-aminidase	KPC74216.1
5566R	298	99/100	*Streptomyces* sp. NRRL WC-3753	lucA-lucC siderophore synthetase	KPC74169.1
5567R	132	99/100	*Streptomyces* sp. NRRL F-5650	L-ectoine synthase	WP_031047567
5568R	423	100/100	*Streptomyces* sp. NRRL WC-3753	diamino-butyrate-2-oxoglutarate transaminase	KPC74167.1
5569R	170	99/99	*Streptomyces* sp. NRRL WC-3753	L-2,4-diaminobutyric acid acetyl-transferase	KPC74166.1
5570	375	97/98	*Streptomyces* sp.	aspartate	KPC74165.1

			NRRL WC-3753	aminotransferase	
5571	176	50/62	*Streptomyces mirabilis*	membrane protein	WP_037710600
5572	363	99/99	*Streptomyces* sp. NRRL WC-3753	amidohydrolase	KPC74163.1

Table S17. Properties of genes encoded within cluster 17 (bacteriocin) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
6088	388	97/97	*Streptomyces coelicoflavus*	monooxygenase	EHN74214.1
6089	301	97/98	*Streptomyces* sp. NRRL WC-3753	oxidoreductase	EHN74213.1
6090R	524	97/97	*Streptomyces violaceorubidus*	acyltransferase	WP_030189328
6091	246	88/93	*Streptomyces mutabilis*	oxidoreductase	KFG76696.1
6092R	59	-	-	-	-
6093R	217	-	-	-	-
6094R	639	100/100	*Streptomyces ambofaciens*	B_{12}-binding domain-containing radical SAM protein	CAK50896.1
6095R	42	-	-	-	-
6096R	606	96/97	*Streptomyces ambofaciens*	putative peptidase	CAK50898.1
6097R	408	95/97	*Streptomyces ambofaciens*	MFS transporter	CAK50899.1
6098R	291	99/99	*Streptomyces coelicoflavus*	UDP pyrophosphate phosphatase	EHN74217.1

Table S18. Properties of genes encoded within cluster 18 (flaviolin) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
6205R	177	96/98	*Streptomyces violaceoruber*	cupin	WP_032866113
6206R	395	96/98	*Streptomyces* sp. NRRL F-5650	cytochrome P450	WP_031033656
6207R	357	98/99	*Streptomyces coelicoflavus*	RppA type III PKS	EHN79529.1

Table S19. Properties of genes encoded within cluster 19 (isorenieratene) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
6709	545	99/99	*Streptomyces coelicoflavus*	FAD-dependent oxidoreductase	EHN74595.1
6710	292	98/99	*Streptomyces coelicoflavus*	short-chain dehydrogenase	EHN74594.1
6711R	500	97/98	*Streptomyces* sp. NRRL WC-3753	isorenieratene synthase	KPC75211.1
6712R	372	80/83	*Streptomyces mutabilis*	methyltransferase	KFG77210.1
6713R	493	98/99	*Streptomyces* sp. NRRL WC-3753	lycopene cyclase	KPC75212.1
6714R	505	98/99	*Streptomyces* sp. NRRL WC-3753	phytoene dehydrogenase	KPC75087.1
6715R	331	97/98	*Streptomyces* sp. NRRL WC-3753	crtB phytoene/squal-ene synthetase	KPC75088.1
6716R	340	97/98	*Streptomyces* sp. NRRL WC-3753	2Fe-2S ferredoxin	KPC75089.1

Table S20. Properties of genes encoded within cluster 20 (indole) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
6796	449	99/99	*Streptomyces coelicoflavus*	cytochrome P450	EHN79639.1

6797	375	83/89	*Streptomyces* sp. NRRL F-5635	prenyltransferase	WP_031185738

Table S21. Properties of genes encoded within cluster 21 (2-methylisoborneol) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
6966R	291	98/100	*Streptomyces coelicoflavus*	methyltransferase	EHN74708.1
6967R	405	96/97	*Streptomyces coelicoflavus*	methylisoborneol synthase	EHN74707.1
6968R	458	97/98	*Streptomyces coelicoflavus*	Crp/Fnr regulator	EHN74706.1

Table S22. Properties of genes encoded within cluster 22 (PKS-NRPS) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
0155R	470	44/61	*Streptomyces* sp. AA4	CalE9 aminotransferase	EFL07741.1
0156R	420	86/91	*Streptomyces* sp. NRRL WC-3753	cytochrome P450	KPC76453.1
0157R	99	98/99	*Kibdelospor-angium* sp.	PCP/ACP	CEL23081.1
0158R	589	98/99	*Saccharothrix syringae*	NRPS C-T domains	WP_051766167
0159R	532	97/98	*Kibdelospor-angium* sp.	NRPS A (Pro) domain	CTQ90221.1
0160R	466	97/98	*Streptomyces* sp. NRRL WC-3753	CFA ligase (Cfa5)	WP_053745359
0161R	177	44/54	*Streptomyces scabiei*	CFA cyclase (Cfa4)	CBG74923.1
0162R	380	57/66	*Streptomyces scabiei*	ketosynthase (Cfa3)	CBG74922.1
0163R	151	86/92	*Streptomyces scabiei*	dehydratase CFA (Cfa2)	CBG74921.1
0164R	150	42/52	*Streptomyces celluloflavus*	thioesterase II	WP_052860305
0165R	463	85/91	*Streptomyces* sp. AS58	acetylornithine aminotransfer-ase	KOV54483.1
0166R	89	73/86	*Streptomyces* sp.	ACP CFA (Cfa1)	KOV54497.1

			AS58		
0167R	752	45/54	*Streptomyces* sp. AS58	EsmB4 pksD (KS-AT?)	AFB35635.1
0168	439	42/60	*Verrucosispora maris*	MFS transporter	AEB46034.1
0169R	1090	72/81	*Nonomuraea candida*	NRPS C-A(?)-T	WP_052422987
putative ORF	*70*			*possible MbtH protein?*	
0170R	1738	76/83	*Nonomuraea candida*	NRPS C-A(cys)-T-Te; *an additional C-terminal ~450 amino acids resembles an MFS transporter domain*	WP_043623389
0171R	963	46/60	*Stackebrandtia nassauensis*	PKS KS-AT(mal)-ACP (*the active site of the AT is mutated Ser to Gly*)	ADD43707.1
0172R	498	90/94	*Streptomyces toyocaensis*	phosphoester-ase	KES04890.1

Table S23. Properties of genes encoded within cluster 23 (PUFA-related) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
0312	404	69/81	*Glaciibacter* sp.	ROK regulator	WP_022884622
0313R	521	66/76	*Streptomyces viridochromo-genes*	unsaturated fatty acid synthase PfaD; malonyl-CoA:ACP acyltransferase	KMS68667.1
0314R	2364	59/71	*Saccharopoly-spora erythraea*	unsaturated fatty acid synthase PfaB/PfaC; KS-KS-AT(mal)-?DH	CAL99376.1
0315R	778	53/62	*Saccharopoly-spora erythraea*	KR-DH domains	CAL99375.1
0316R	1271	59/71	*Saccharopoly-spora erythraea*	KS-AT domains	CAL99374.1

0317R	239	50/63	*Streptomyces* sp. SPB74	thioesterase	EFG64483.1
0318R	96	67/85	*Saccharopoly-spora erythraea*	ACP	CAL99372.1

Table S24. Properties of genes encoded within cluster 24 (lomaiviticin-like) in *Streptomyces* sp. OM-5714.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
0118	415	97/98	*Streptomyces* sp. NRRL WC-3753	KSβ Lom59	KPC71304.1
0119	421	100/100	*Streptomyces coelicoflavus*	KSα Lom58	EHN77732.1
0120	268	98/98	*Streptomyces* sp. NRRL WC-3753	thioesterase	KPC71296.1
0121	573	69/78	*Streptomyces* sp. NRRL WC-3753	NRPS	EDY66720.2
0122	258	99/100	*Streptomyces* sp. NRRL WC-3753	αβ-hydrolase	KPC71297.1
0123	194	96/97	*Streptomyces coelicoflavus*	adenylylsulfate kinase	EHN80224.1
0124	312	98/99	*Streptomyces coelicoflavus*	sulfate adenylyltransferase subunit 2	EHN80225.1
0125	445	97/98	*Streptomyces coelicoflavus*	sulfate adenylyltransferase subunit 1	EHN80226.1
0126	353	96/97	*Streptomyces coelicoflavus*	sulfate ABC transporter substrate-binding	EHN80227.1
0127	282	97/99	*Streptomyces coelicoflavus*	sulfate ABC transporter ATP-binding	EHN80228.1
0128	156 (partial)	97/98	*Streptomyces coelicoflavus*	sulfate ABC transporter permease	EHN80229.1

end of contig start of contig			*nuclease*		
0001	471	51/59	*Streptomyces* sp. NRRL S-646	lycopene cyclase	WP_037884983
0002	67	-	-	-	-
0003	592	99/99	*Streptomyces coelicoflavus*	multidrug ABC transporter ATPase	EHN77703.1
0004	644	98/98	*Streptomyces* sp. NRRL WC-3753	phenoxazinone synthase	EHN77704.1
0005R	87	-	-	-	-
0006	391	47/61	*Salinispora pacifica*	glycosyltrans-ferase	AHF72785.1
0007R	358	52/66	*Salinispora pacifica*	glycosyltrans-ferase Lom55	AHF72791.1
0008R	265	99/100	*Streptomyces coelicoflavus*	glucose-1-phosphate thymidylyltransfera se	EHN77692.1
0009	197	99/99	*Streptomyces* sp. NRRL WC-3753	NDP-hexose-3-ketoreductase	KPC71380.1
0010R	208	99/100	*Streptomyces coelicoflavus*	dTDP-4-dehydro-rhamnose 3,5-epimerase	EHN77693.1
0011R	220	63/77	*Salinispora pacifica*	sulfotrans-ferase Lom10	AHZ61844.1
0012R	146	94/95	*Streptomyces* sp. NRRL WC-3753	alkylhydro-peroxidase	EHN77695.1
0013	340	98/99	*Streptomyces coelicoflavus*	O-methyltrans-ferase Lom23	EHN77696.1
0014	208	97/98	*Streptomyces* sp. NRRL WC-3753	DSBA oxidoreductase Lom20	KPC71395.1

0015R	301	99/99	*Streptomyces coelicoflavus*	NmrA regulator Lom19	EHN77697.1
0016	242	99/99	*Streptomyces* sp. NRRL WC-3753	dehydrogenase Lom18	KPC71375.1
0017	491	95/96	*Streptomyces coelicoflavus*	FAD-binding monooxygen-ase Lom17	EHN77708.1
0018	497	97/98	*Streptomyces coelicoflavus*	FAD-binding mono-oxygenase Lom16	EHN77709.1
0019	475	98/99	*Streptomyces* sp. NRRL WC-3753	protoporphyrin-ogen oxidase Lom14	KPC71373.1
0020R	573	98/98	*Streptomyces coelicoflavus*	aromatic ring hydroxylase Lom13	EHN77699.1
0021R	160	99/99	*Streptomyces coelicoflavus*	pyridoxamine 5'-phosphate oxidase	EHN77700.1
0022R	421	99/99	*Streptomyces coelicoflavus*	MFS transporter	EHN77701.1
0023R	464	39/53	*Streptomyces* sp. NRRL S-813	protein kinase	WP_030184461
0024R	259	27/47	*Frankia* sp. CN3	phytanoyl-CoA dioxygenase	WP_007511940
0025R	407	37/49	*Streptomyces olindensis*	protein kinase	KDN77276.1
0027	438	42/59	*Streptomyces bingcheng-gensis*	glycosyl transferase	ADI11531.1
0028	510	97/98	*Streptomyces* sp. NRRL WC-3753	transporter	KPC71367.1
0029	104	99/100	*Streptomyces coelicoflavus*	glyoxalase	EHN77725.1
0030	326	99/100	*Streptomyces coelicoflavus*	dTDP-glucose 4,6-dehydratase	EHN77726.1

0031	450	98/98	*Streptomyces coelicoflavus*	NDP-hexose 2,3-dehydratase Lom9	EHN77727.1
0032R	176	100/100	*Streptomyces coelicoflavus*	alkylhydro-peroxidase Lom11	EHN77723.1
0033R	245	97/99	*Streptomyces coelicoflavus*	peptidase C26 Lom12	EHN77724.1
0034	308	99/99	*Streptomyces* sp. NRRL WC-3753	ABC transporter ATP-binding Lom4	KPC71362.1
0035	279	99/99	*Streptomyces* sp. NRRL WC-3753	ABC transporter Lom3	KPC71391.1
0036	312	98/99	*Streptomyces* sp. NRRL WC-3753	nuclease Lom15	KPC71361.1
0037R	263	94/95	*Streptomyces* sp. NRRL WC-3753	AraC regulator Lom7	*0061R*
0038	140	98/98	*Streptomyces coelicoflavus*	glyoxalase Lom5	EHN77715.1
0039	156	96/98	*Streptomyces* sp. NRRL WC-3753	AraC regulator	KPC71358.1
0040	210	85/90	*Streptomyces* sp. NRRL S-813	glyoxalase	WP_030184415
0041R	138	99/99	*Streptomyces coelicoflavus*	apoprotein precursor	EHN77713.1
0042	113	-	-	-	-
0043	151	99/99	*Streptomyces azureus*	transposase	GAP50821.1
0044R	132	94/97	*Streptomyces* sp. NRRL F-5008	HxlR regulator	KPC71361.1
0045	222	94/95	*Streptomyces* sp. NRRL F-5650	esterase	WP_037836837
0046	145	100/100	*Streptomyces* sp. Amel2xE9	peptidase S51	WP_037725129
0047R	262	100/100	*Streptomyces* sp. NRRL WC-3753	AraC regulator	KPC71390.1
0048R	101	85/90	*Streptomyces*	Xre regulator	KOT38454.1

			caelestis		
0049R	138	90/98	*Streptomyces griseoflavus*	toxin RelE	EFL40766.1
0050R	244	-	-	-	-
0051	161	44/62	*Mycobacterium tuberculosis*	hypothetical protein	CNE79824.1
0052R	134	98/100	*Streptomyces coelicoflavus*	N-acetyltrans-ferase GCN5 Lom35	EHN77656.1
0053R	428	99/99	*Streptomyces coelicoflavus*	adenylo-succinate lyase Lom34	EHN77655.1
0054R	506	97/97	*Streptomyces* sp. NRRL WC-3753	glutamine synthetase Lom32	KPC71351.1
0055R	462	99/99	*Streptomyces coelicoflavus*	amidase Lom33	EHN77687.1
0056R	635	-	-	Lom29?	-
0057R	109	99/100	*Streptomyces coelicoflavus*	Fe-4S ferredoxin	EHN77682.1
0058	360	99/99	*Streptomyces coelicoflavus*	O-methyltrans-ferase	EHN77684.1
0059R	514	98/98	*Streptomyces* sp. NRRL WC-3753	FAD-mono-oxygenase	KPC71387.1
0060R	510	97/97	*Streptomyces* sp. NRRL WC-3753	FAD-mono-oxygenase	KPC71348.1
0061R	314	99/99	*Streptomyces coelicoflavus*	cyclase/ dehydratase	EHN77676.1
0062R	261	100/100	*Streptomyces coelicoflavus*	ketoacyl reductase	EHN77677.1
0063R	91	100/100	*Streptomyces coelicoflavus*	ACP act-like	EHN77678.1
0064R	109	100/100	*Streptomyces coelicoflavus*	polyketide cyclase	EHN77679.1
0065R	231	98/99	*Streptomyces* sp. NRRL WC-3753	anthrone oxidase	KPC71343.1
0066R	477	98/99	*Streptomyces* sp. NRRL WC-3753	FAD-mono-oxygenase	KPC71342.1

0067	185	59/70	*Streptomyces murayamaensis*	JadX-like protein	AAO65342.1
0068R	335	98/98	*Streptomyces coelicoflavus*	Ppan transferase	EHN77668.1
0069R	604	58/67	*Streptomyces* sp. NRRL S-87	acyl-CoA dehydrogenase	WP_051796199
0070	326	99/99	*Streptomyces coelicoflavus*	cyclase/act dimerase	EHN80232.1
0071	422	96/97	*Streptomyces* sp. NRRL WC-3753	MFS transporter	KPC71339.1
0072	74	-	-	-	-
0073R	273	87/93	*Streptomyces pristinaespiralis*	SARP regulator Lom7	CBW45731.1
0074R	276	80/88	*Streptomyces azureus*	SARP regulator	GAP47966.1
0075R	208	96/97	*Streptomyces coelicoflavus*	TetR regulator	EHN80235.1
0076R	530	99/99	*Streptomyces coelicoflavus*	methylmalonyl-CoA carboxyl-transferase	EHN80236.1
0077R	59	-	-	-	-
0078	321	63/78	*Streptomyces aureofaciens*	Afs-like γ-butyrolactone synthase	ADM72848.2
0079R	286	97/97	*Streptomyces* sp. NRRL WC-3753	nucleoside-diphosphate sugar epimerase	KPC71259.1
0080	222	98/99	*Streptomyces coelicoflavus*	γ-butyrolactone receptor protein	EHN80268.1
0081R	495	59/68	*Streptomyces* sp. SM8	aromatic ring hydroxylase	EKC95639.1

Table S25. Properties of genes encoded within cluster 1 (coelichelin) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
0135	315	98/99	*Streptomyces* sp. NRRL WC-3753	formyltrans-ferase CchA	KPC76207.1
0134R	451	98/99	*Streptomyces* sp. NRRL WC-3753	L-ornithine mono-oxygenase CchB	KPC76206.1
0133R	328	99/99	*Streptomyces coelicoflavus*	iron transporter permease CchC	EHN77999.1
0132R	354	99/99	*Streptomyces* sp. NRRL WC-3753	iron transporter permease CchD	KPC76204.1
0131R	287	95/97	*Streptomyces* sp. NRRL WC-3753	ABC iron transporter CchE	KPC76203.1
0130R	350	98/99	*Streptomyces coelicoflavus*	iron-siderophore binding protein CchF	EHN78002.1
0129R	597	98/99	*Streptomyces coelicoflavus*	ABC transporter CchG	EHN78003.1
0128R	3672	96/97	*Streptomyces coelicoflavus*	NRPS (L-fh-Orn, hOrn, L-Thr) CchH	EHN78004.1
0127R	566	99/99	*Streptomyces coelicoflavus*	ABC transporter CchI	EHN78005.1
0124R	70	92/97	*Streptomyces* sp. NRRL F4711	MbtH-like protein CchK	KOX46430.1

Table S26. Properties of genes encoded within cluster 2 (bacteriocin) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
0396R	828	98/99	*Streptomyces* sp. NRRL WC-3753	PAS regulator	KPC76431.1
0397R	408	98/99	*Streptomyces* sp. NRRL WC-3753	proteinase	KPC76432.1
0398	168	—	—	—	—
0399	71	—	—	—	—
0400	268	85/92	*Streptomyces azureus*	secreted protein	GAP45380.1
0401	720	99/99	*Streptomyces coelicoflavus*	ABC transporter	EHN78272.1
0402	937	99/99	*Streptomyces coelicoflavus*	ABC transporter	EHN78271.1

Table S27. Properties of genes encoded within cluster 3 (hopene/ATBH) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
0723	265	97/98	*Streptomyces* sp. NRRL WC-3753	protein phosphatase	KPC72758.1
0724R	39	89/97	*Streptomyces aureofaciens*	small hydrophobic protein	WP_052842011
0725R	204	99/100	*Streptomyces coelicoflavus*	Xre regulator	EHN78562.1
0726R	461	99/99	*Streptomyces coelicoflavus*	aminotransfer-ase	EHN78561.1
0727R	650	98/99	*Streptomyces coelicoflavus*	DXP synthase	EHN78560.1
0728R	384	99/100	*Streptomyces* sp. NRRL WC-3753	4-hydroxy-3-methyl-but--2-enyl diphosphate synthase	KPC72753.1
0729R	340	99/99	*Streptomyces* sp. FXJ7.023	radical SAM protein HpnH	WP_037772613
0730R	213	98/99	*Streptomyces coelicolor*	1-hydroxy-2-methyl-but-2-enyl 4-diphosphate reductase	CAB39698.1
0731R	684	98/99	*Streptomyces* sp. NRRL WC-3753	squalene-hopene cyclase	KPC72750.1
0732R	354	98/100	*Streptomyces coelicoflavus*	geranylgeranyl diphosphate synthase	EHN78555.1
0733R	478	99/100	*Streptomyces* sp. NRRL WC-3753	phytoene dehydrogenase	KPC72748.1
0734R	312	98/99	*Streptomyces coelicoflavus*	phytoene synthase	KPC72747.1
0735R	300	98/98	*Streptomyces coelicoflavus*	squalene synthase	KPC72746.1

Table S28. Properties of genes encoded within cluster 5b (arsenolipid) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
1161	354	94/96	*Streptomyces lividans*	transcriptional activator	EFD66080.1
1162R	108	98/99	*Streptomyces*	ion tolerance	ABP49085.1

			lividans	protein	
1163R	534	99/99	*Streptomyces coelicolor*	methylmalonyl-CoA mutase	CAB71920.1
1164R	91	44/55	*Streptomyces bingcheng-gensis*	histidine kinase	ADI11540.1
1165R	111	97/98	*Streptomyces coelicolor*	ArsR repressor	CAB71918.1
1166	196	97/98	*Streptomyces coelicolor*	oxidoreductase	CAB71917.1
1167R	493	99/99	*Streptomyces coelicolor*	flavoprotein	CAB71916.1
1168	2358	97/97	*Streptomyces coelicoflavus*	PKS	EHN74897.1
1169	346	45/61	*Streptomyces avermitilis*	FabH	BAC71233.1
1170	223	97/99	*Streptomyces coelicoflavus*	membrane protein	EHN74895.1
1171	367	98/99	*Streptomyces coelicoflavus*	phosphono-pyruvate decarboxylase	EHN74894.1
1172R	133	99/99	*Streptomyces lividans*	ArsR regulator	EOY45801.1
1173R	521	99/99	*Streptomyces coelicolor*	membrane efflux protein	CAB71269.1
1174	240	90/93	*Streptomyces griseorubens*	CTP synthase	KEG37908.1
1175	745	98/99	*Streptomyces coelicolor*	oxidoreductase	CAB71267.1
1176	440	99/100	*Streptomyces coelicoflavus*	3-phospho-shikimate 1-carboxyvinyl-transferase	EHN74891.1
1177	511	96/97	*Streptomyces coelicoflavus*	putative C-As bond-forming enzyme	EHN74890.1
1178	119	—	—	hypothetical protein	—
1179	290	98/99	*Streptomyces coelicolor*	ABC phosphate transporter	CAB71263.1
1180	603	98/99	*Streptomyces coelicoflavus*	ABC phosphate transporter permease	EHN74887.1
1181	229	98/99	*Streptomyces coelicoflavus*	ABC phosphate transporter ATP-	EHN74886.1

				binding	
1182R	92	100/100	*Streptomyces coelicolor*	ArsR regulator	CAB71260.1
1183	429	97/98	*Streptomyces coelicolor*	secreted protein	CAB71258.1
1184	152	94/96	*Streptomyces coelicolor*	glyoxalase	CAB71257.1
1185	409	97/97	*Streptomyces coelicolor*	MFS transporter	CAB71256.1
1186	120	99/100	*Streptomyces coelicoflavus*	ArsR regulator	EHN74882.1
1187R	352	97/98	*Streptomyces coelicoflavus*	mono-oxygenase	EHN74294.1
1188R	65	—	—	—	—
1189R	369	98/98	*Streptomyces coelicoflavus*	ArsB arsenic resistance transport protein	EHN74922.1
1190R	102	100/100	*Streptomyces coelicoflavus*	ArsR regulator	EHN74921.1
1191	137	97/100	*Streptomyces coelicoflavus*	ArsC arsenate reductase	EHN74881.1
1192	328	97/98	*Streptomyces coelicoflavus*	ArsT putative thioredoxin reductase	EHN74880.1
1193	107	89/92	*Streptomyces cattleya*	thioredoxin	YP_006054723

Table S29. Properties of genes encoded within cluster 6a (NRPS) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
1556R	304	79/84	*Streptomyces viridosporus*	aminotransfer-ase	WP_016823965
1557R	233	100/100	*Streptomyces coelicoflavus*	LuxR regulator	EHN79569.1
1558R	113	99/99	*Streptomyces coelicoflavus*	MbtH-like	EHN79570.1
1559R	61	100/100	*Streptomyces coelicoflavus*	MbtH-like	EHN79571.1
1560R	207	96/98	*Streptomyces sp. NRRL F-4711*	3-isopropyl-malate synthase	KOU01073.1
1561R	467	100/100	*Streptomyces coelicoflavus*	3-isopropyl-malate synthase	EHN79573.1

1562R	463	30/45	*Streptomyces mobaraensis*	3-isopropyl-malate synthase	EMF01221.1
1563R	435	98/99	*Streptomyces coelicoflavus*	aspartate aminotransfer-ase	EHN79575.1
1564R	375	99/99	*Streptomyces coelicoflavus*	3-isopropyl malate dehydratase	EHN79576.1
1565R	432	99/99	*Streptomyces coelicoflavus*	esterase	EHN79577.1
1566R	3259	98/98	*Streptomyces coelicoflavus*	NRPS (ATE-CAT-CAT)	EHN79578.1
1567R	3631	77/82	*Streptomyces coelicoflavus*	NRPS (CAMT-CATE-C)	EHN79578.1
1568	152	-	-	-	-
1569	398	98/99	*Streptomyces coelicoflavus*	peptidase M23	EHN75119.1
1570R	235	99/100	*Streptomyces coelicoflavus*	ABC transporter	EHN75124.1
1571R	325	100/100	*Streptomyces coelicoflavus*	ABC transporter	EHN75123.1
1572R	381	99/99	*Streptomyces coelicoflavus*	peptidase	EHN75122.1
1573R	241	96/97	*Streptomyces coelicoflavus*	2-component system regulator	EHN75121.1
1574R	199	100/100	*Streptomyces coelicoflavus*	LuxR regulator	EHN75120.1
1575R	936	33/45	*Streptomyces* sp. MBT28	LuxR regulator	WP_046251400
1576	788	34/48	*Streptomyces* sp. NRRL B-24891	LuxR regulator ATP-binding	WP_046495161
1577	841	30/40	*Streptomyces* sp. PRh5	LuxR regulator	EXU65116.1

Table S30. Properties of genes encoded within cluster 6b (butyrolactone) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
1587R	504	97/98	*Streptomyces* sp. NRRL WC-3753	IucA-IucC siderophore synthetase	KPC87683.1
1588	516	94/96	*Streptomyces* sp. NRRL WC-3753	iron transporter	KPC87684.1
1589	421	94/96	*Streptomyces* sp. NRRL WC-3753	Von Willebrand Factor type A	KPC87685.1

				protein	
1590	262	77/87	*Streptomyces sviceus*	secreted protein	EDY55527.1
1591R	218	92/94	*Streptomyces lividans*	HxlR regulator	EOY51317.1
1592	402	99/99	*Streptomyces coelicolflavus*	aminotransferase	EHN77626.1
1593	199	99/100	*Streptomyces coelicoflavus*	phosphoglycerate mutase	EHN77625.1
1594	216	—	—	—	—
1595	132	91/96	*Streptomyces* sp.	thioesterase	WP_030403422

Table S31. Properties of genes encoded within cluster 7 (geosmin) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
1771R	193	95/96	*Streptomyces* sp. NRRL WC-3753	membrane protein	KPC71996.1
1772R	740	98/99	*Streptomyces* sp. NRRL WC-3753	geosmin synthase	KPC71995.1
1773R	405	75/83	*Streptomyces azureus*	oxidoreductase	GAP48695.1
1774R	215	99/99	*Streptomyces* sp. NRRL WC-3753	A-factor receptor	KPC71993.1
1775	337	95/96	*Streptomyces lividans*	sugar ABC transporter	EFD65917.1
1776	860	94/96	*Streptomyces* sp. NRRL WC-3753	histidine kinase	KPC71991.1

Table S32. Properties of genes encoded within cluster 8 (bacteriocin) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
1794	507	98/98	*Streptomyces coelicoflavus*	iron-sulfur-binding protein	EHN75283.1
1795	296	87/91	*Streptomyces* sp. FXJ7.023	polysaccharide deacetylase	WP_037762876
1796	75	-	-	-	-
1797R	224	98/98	*Streptomyces* sp. NRRL WC-3753	ABC transporter	KPC71969.1
1798	284	95/96	*Streptomyces*	ABC transporter	EHN75332.1

			coelicoflavus		
1799	254	90/93	*Streptomyces* sp. FXJ7.023	peptidyl-tRNA hydrolase	WP_037762889
1800	469	93/93	*Streptomyces* sp. NRRL WC-3753	endonuclease	KPC71966.1
1801	265	93/93	*Streptomyces* sp. NRRL WC-3753	membrane protein	KPC71965.1
1802	535	96/97	*Streptomyces* sp. NRRL WC-3753	proteinase	KPC71964.1
1803R	243	93/96	*Streptomyces turgidiscabies*	chlorite dismutase	ELP64953.1
1804R	494	96/98	*Streptomyces coelicoflavus*	protoporphyrinogen oxidase	EHN75320.1

Table S33. Properties of genes encoded within cluster 9b (undecylprodigiosin) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
1972R	295	98/99	*Streptomyces* sp. NRRL WC-3753	acyltransferase	KPC71499.1
1973R	226	33/58	*Kibdelosporang-ium aridum*	membrane protein	WP_033381793
1974R	406	99/99	*Streptomyces* sp. NRRL WC-3753	oxidase (MarG-like)	KPC71501.1
1975R	936	97/97	*Streptomyces* sp. NRRL WC-3753	PEP-dependent enzyme (MarH-like)	KPC71502.1
1976R	362	99/99	*Streptomyces coelicoflavus*	methyltransfer-ase	EHN75476.1
1977R	280	96/97	*Streptomyces* sp. NRRL WC-3753	thioesterase II	EHN75475.1
1978R	347	98/99	*Streptomyces* sp. NRRL WC-3753	oxidoreductase	EHN75474.1
1979R	328+	98/98	*Streptomyces lividans*	fatty acyl-AMP ligase	AIJ12834.1
1980R	1885	85/89	*Streptomyces coelicolor*	modular PKS	CAA16183.1
1981R	532	96/97	*Streptomyces* sp. NRRL WC-3753	NRPS (MarM-like)	KPC87186.1
1982R	642	96/96	*Streptomyces* sp. NRRL WC-3753	8-amino-7-oxo-nonanoate synthase	KPC87185.1

1983R	97	73/85	*Streptomyces peruviensis*	ACP/PCP	WP_030050065
1984R	329	99/99	*Streptomyces* sp. NRRL WC-3753	3-ketoacyl-ACP synthase	KPC87183.1
1985	81	98/98	*Streptomyces coelicoflavus*	ACP	EHN75488.1
1986	407	99/99	*Streptomyces* sp. NRRL WC-3753	3-ketoacyl-ACP synthase	EHN75487.1
1987	146	47/60	*Streptomyces katrae*	phosphoester-ase	WP_030293828
1988	296	64/75	*Streptomyces* sp. CNQ-617	?MarT-like protein	AHF22847.1
1989	265	58/67	*Streptomyces* sp. CNQ-617	4'-phospho-pantetheine transferase	AHF22846.1
1990	389	64/75	*Streptomyces coelicoflavus*	RedV 4'-Ppan transferase	EHN75483.1
1991	197	98/99	*Streptomyces coelicoflavus*	LuxR regulator	EHN75482.1
1992R	106	98/98	*Streptomyces coelicoflavus*	RedY protein	EHN75490.1
1993R	391	99/99	*Streptomyces coelicoflavus*	RedW acyl-CoA dehydrogenase	EHN75489.1
1994R	1071	94/95	*Streptomyces* sp. NRRL WC-3753	RedX PKS	KPC87173.1
1995R	268	98/99	*Streptomyces coelicoflavus*	RedD regulator	EHN79898.1

Table S34. Properties of genes encoded within cluster 10 (NIS) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
2079R	230	72/84	*Streptomyces fulvoviolaceus*	siderophore biosynthetic protein	WP_030605409
1769R	673	79/82	*Streptomyces violaceorubidis*	iron transporter	WP_030184400
1770R	422	86/87	*Streptomyces* sp. NRRL WC-3753	2,4-diamino-butyrate 4-amino-transferase	KPC87106.1
1771	408	80/89	*Streptomyces davawensis*	secreted protein	CCK27029.1
1772R	497	98/98	*Streptomyces*	ATP-binding	EFD66294.1

			lividans	protein	
1773R	512	98/98	*Streptomyces* sp. NRRL WC-3753	zinc metalloprotein-ase	KPC87103.1
1774R	721	96/98	*Streptomyces* sp. NRRL WC-3753	ppGpp synthetase	KPC87189.1

Table S35. Properties of genes encoded within cluster 11 (spore pigment) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
2529R	591	90/92	*Streptomyces violaceoruber*	polyketide hydroxylase	WP_030872674
2530	405	98/98	*Streptomyces* sp. NRRL WC-3753	protein in whiE locus	KPC89504.1
2531	158	99/99	*Streptomyces aureofaciens*	small protein in whiE locus	WP_052840703
2532	423	99/99	*Streptomyces* sp. NRRL WC-3753	PKS KSalpha	KPC88984.1
2533	424	98/99	*Streptomyces* sp. NRRL WC-3753	PKS KSbeta	KPC88985.1
2534	89	100/100	*Streptomyces coelicoflavus*	ACP	EHN75822.1
2535	159	98/98	*Streptomyces* sp. NRRL WC-3753	polyketide cyclase	KPC88987.1
2536	111	100/100	*Streptomyces coelicoflavus*	WhiE VII	EHN75820.1
2537	299	98/99	*Streptomyces coelicoflavus*	membrane protein	EHN75819.1

Table S36. Properties of genes encoded within cluster 12 (albaflavenone) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
2632	322	99/100	*Streptomyces coelicoflavus*	AraC regulator	EHN75894.1
2633R	458	97/98	*Streptomyces violaceorubidis*	*epi*-isoizaene-5-hydroxylase	WP_030145938
2634R	307	99/99	*Streptomyces* sp. WC-3735	*epi*-isozizaene synthase	KPC89510.1

Table S37. Properties of genes encoded within cluster 13 (lipopeptide) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
3489	172	96/97	*Streptomyces* sp. NRRL WC-3753	membrane protein	KPC77435.1
3490R	387	98/99	*Streptomyces* sp. NRRL WC-3753	aromatic acid decarboxylase	KPC77436.1
3491	253	98/99	*Streptomyces* sp. NRRL WC-3753	Clp proteinase	KPC77437.1
3492	563	97/98	*Streptomyces* sp. NRRL WC-3753	3'-5'-exo nuclease	KPC77438.1
3493	301	96/98	*Streptomyces* sp. NRRL WC-3753	formylTHF deformylase	KPC77439.1
3494	338	98/99	*Streptomyces* sp. NRRL WC-3753	glycosyltrans-ferase	KPC77440.1
3495	352	98/99	*Streptomyces* sp. NRRL WC-3753	G1P thymidylyltrans-ferase	KPC77441.1
3496	324	91/95	*Streptomyces* sp. NRRL WC-3753	dTDP-glucose 4,6-dehydratase	KPC77442.1
3497	72	99/99	*Streptomyces* sp. NRRL WC-3753	MbtH-like protein	KPC77443.1
3498	597	97/98	*Streptomyces coelicoflavus*	SARP regulator	EHN72128.1
3499R	262	53/66	*Streptomyces avermitilis*	4'-PPan transferase	BAC71349.1
3500R	305	96/98	*Streptomyces coelicoflavus*	SyrP-like protein	EHN78774.1
3501	159	96/98	*Streptomyces* sp. NRRL WC-3753	NADPH-flavin oxidoreductase	KPC77447.1
3502R	339	82/86	*Streptomyces griseoruber*	ACP	KEG39421.1
3503R	59	—	—	—	—
3504	72	98/100	*Streptomyces coelicoflavus*	MbtH-like protein	EHN77811.1
3505R	93	97/97	*Streptomyces coelicoflavus*	PCP	EHN77810.1
3506R	560	52/63	*Streptomyces ipomoeae*	acyl-CoA dehydrogenase	EKX68087.1
3507R	610	46/57	*Streptomyces ambofaciens*	acyl-CoA dehydrogenase	AKZ58687.1
3508R	593	58/67	*Streptomyces ambofaciens*	fatty-acyl-AMP synthase	AKZ58686.1

3509R	227	98/98	*Streptomyces* sp. NRRL WC-3753	RedY protein	KPC77452.1
3510R	803	99/99	*Streptomyces* sp. NRRL WC-3753	RedN acyl-CoA dehydrogenase	KPC77453.1
3511R	256	94/95	*Streptomyces* sp. NRRL WC-3753	RedX PKS	KPC77454.1
3512	405	99/99	*Streptomyces* sp. NRRL WC-3753	histidine kinase	KPC77455.1
3513	202	99/99	*Streptomyces* sp. NRRL WC-3753	LuxR regulator	KPC77456.1
3514R	314	99/99	*Streptomyces* sp. NRRL WC-3753	SyrP-like protein	KPC77457.1
3515R	346	100/100	*Streptomyces* sp. NRRL WC-3753	monooxygenase	KPC77458.1
3516R	427	98/98	*Streptomyces* sp. NRRL WC-3753	MFS transporter	KPC77459.1
3517R	3867	50/62	*S.* sp. 769	NRPS (CAT-CAT-CAMT-Te)	AJC59844.1
3518R	4188	52/64	*Streptomyces viridochromo-genes*	NRPS (CAT-CAT-CAT-CAT)	AEG64697.1
3519R	6986	85/90	*Actinospica acidiphila*	NRPS (CAT-CT-CATTE-CAT-CAT-CATE)	WP_051724693
3520	276	28/46	*Rhodococcus pyridinivorans*	ABC ATPase transporter	EHK85780.1
3521R	217	95/97	*S. aureofaciens*	LuxR regulator	WP_052836389
3522R	537	58/66	*Streptomyces acidiscabies*	2-component histidine kinase	WP_0103521761
3523R	400	35/49	*Prauserella* sp.	Clp proteinase	KID29624.1

Table S38. Properties of genes encoded within cluster 14 (NIS) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
4856R	605	99/99	*Streptomyces* sp. NRRL WC-3753	glutamine amidotransfer-ase	KPC85276.1
4857R	85	—	—	—	—
4858R	297	98/98	*Streptomyces* sp. NRRL WC-3753	TAT signal sequence protein	KPC85278.1
4859R	535	98/99	*Streptomyces* sp. NRRL WC-3753	beta-N-acetylhexos-aminidase	KPC85279.1

4860R	592	97/98	*Streptomyces* sp. NRRL WC-3753	lucA-lucC siderophore synthetase	KPC85280.1
4861R	182	98/99	*Streptomyces coelicoflavus*	acetyltransferase	EHN73011.1
4862R	426	98/99	*Streptomyces* sp. NRRL WC-3753	alcaligin-like biosynthesis protein	KPC85282.1
4863R	480	99/99	*Streptomyces* sp. NRRL WC-3753	decarboxylase	KPC85283.1
4864R	286	98/99	*Streptomyces* sp. NRRL WC-3753	sialic acid transporter	KPC85284.1
4865R	350	99/99	*Streptomyces coelicoflavus*	ABC transporter	EHN73007.1
4866R	345	56/66	*Streptospor-angium roseum*	Ser/Thr protein kinase	ACZ87621.1

Table S39. Properties of genes encoded within cluster 15 (melanin) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
4940R	480	99/100	*Streptomyces* sp. NRRL WC-3753	undecaprenyl-phosphate glucose phosphotrans-ferase	KPC85353.1
4941R	447	77/80	*Streptomyces violaceoruber*	glycosyl transferase	WP_030866265
4942	72	97/98	*Streptomyces ambofaciens*	chaplin	AKZ55784.1
4943	259	44/53	*Streptomyces sviceus*	secreted protein	EDY61093.1
4944	100	67/73	*Streptomyces scabiei*	secreted protein	CBG72938.1
4945	171	42/56	*Streptomyces venezuelae*	secreted protein	WP_015035837
4946	210	93/95	*Streptomyces* sp. NRRL WC-3753	tyrosinase	KPC85870.1
4947	288	99/100	*Streptomyces* sp. NRRL WC-3753	tyrosinase	KPC85358.1
4948R	90	93/97	*Streptomyces*	membrane protein	WP_030866270

			violaceoruber		
4949	211	83/88	*Streptomyces lividans*	secreted protein	EOY47759.1
4950R	323	96/97	*Streptomyces* sp. NRRL WC-3753	2-hydroxyacid dehydrogenase	KPC85361.1
4951	535	96/98	*Streptomyces aureofaciens*	ATP-binding protein	WP_052838090
4952	615	95/96	*Streptomyces* sp. NRRL WC-3753	ATP-binding protein	KPC85505.1

Table S40. Properties of genes encoded within cluster 16 (ectoine) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
5774	509	98/98	*Streptomyces coelicoflavus*	NADP-dependent fatty aldehyde dehydrogenase	EHN73793.1
5775	163	94/98	*Streptomyces* sp. NRRL WC-3753	regulation of enolase 1	KPC74171.1
5776	237	97/97	*Streptomyces* sp. NRRL WC-3753	protein-disulfide isomerase	KPC74170.1
5777	364	98/99	*Streptomyces* sp. NRRL WC-3753	beta-*N*-acetylhexos-aminidase	KPC74216.1
5778R	298	99/100	*Streptomyces* sp. NRRL WC-3753	lucA-lucC siderophore synthetase	KPC74169.1
5779R	132	99/100	*Streptomyces* sp. NRRL F-5650	L-ectoine synthase	WP_031047567
5780R	423	100/100	*Streptomyces* sp. NRRL WC-3753	diamino-butyrate-2-oxoglutarate transaminase	KPC74167.1
5781R	170	99/99	*Streptomyces* sp. NRRL WC-3753	L-2,4-diaminobutyric acid acetyl-transferase	KPC74166.1
5782R	375	97/98	*Streptomyces* sp.	aspartate	KPC74165.1

			NRRL WC-3753	aminotransferase	
5783	176	50/62	*Streptomyces mirabilis*	membrane protein	WP_037710600
5784	363	99/99	*Streptomyces* sp. NRRL WC-3753	amidohydrolase	KPC74163.1

Table S41. Properties of genes encoded within cluster 17 (bacteriocin) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
6294	388	97/97	*Streptomyces coelicoflavus*	monooxygenase	EHN74214.1
6295	301	97/98	*Streptomyces* sp. NRRL WC-3753	oxidoreductase	EHN74213.1
6296R	524	97/97	*Streptomyces violaceorubidus*	acyltransferase	WP_030189328
6297	246	88/93	*Streptomyces mutabilis*	oxidoreductase	KFG76696.1
6298R	59	-	-	-	-
6299R	217	-	-	-	-
6300R	639	100/100	*Streptomyces ambofaciens*	B_{12}-binding domain-containing radical SAM protein	CAK50896.1
6301R	42	-	-	-	-
6302R	606	96/97	*Streptomyces ambofaciens*	putative peptidase	CAK50898.1
6303R	408	95/97	*Streptomyces ambofaciens*	MFS transporter	CAK50899.1
6304R	33	-	-	-	-
6305R	291	99/99	*Streptomyces coelicoflavus*	UDP pyrophosphate phosphatase	EHN74217.1

Table S42. Properties of genes encoded within cluster 18 (flaviolin) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
6412R	177	97/98	*Streptomyces violaceoruber*	cupin	WP_030866113
6413R	395	96/98	*Streptomyces* sp. NRRL F-5650	cytochrome P450	WP_031633656
6414R	357	98/99	*Streptomyces coelicoflavus*	type III PKS, RppA-like	EHN79529.1

Table S43. Properties of genes encoded within cluster 19 (isorenieratene) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
6888	545	99/99	*Streptomyces* sp. NRRL WC-3753	FAD-dependent oxidoreductase	KPC75085.1
6889	292	99/99	*Streptomyces coelicoflavus*	short-chain dehydrogenase	KPC75086.1
6890R	500	97/98	*Streptomyces* sp. NRRL WC-3753	isorenieratene synthase	KPC75087.1
6891R	372	81/85	*Streptomyces mutabilis*	methyltransferase	KFG77210.1
6892R	493	98/98	*Streptomyces* sp. NRRL WC-3753	lycopene cyclase	KPC75212.1
6893R	505	98/99	*Streptomyces* sp. NRRL WC-3753	phytoene dehydrogenase	KPC75087.1
6894R	331	97/98	*Streptomyces* sp. NRRL WC-3753	crtB phytoene/squal-ene synthetase	KPC75088.1
6895R	340	97/98	*Streptomyces* sp. NRRL WC-3753	2Fe-2S ferredoxin	KPC75089.1

Table S44. Properties of genes encoded within cluster 20 (indole) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
6968	449	98/99	*Streptomyces coelicoflavus*	cytochrome P450	EHN79639.1
6969	375	83/89	*Streptomyces* sp.	prenyl transferase	WP_031185738

			NRRL F-5635		

Table S45. Properties of genes encoded within cluster 21 (MIB) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
7138R	291	98/100	*Streptomyces* sp. WC-3753	methyltransferase	KPC86710.1
7139R	405	96/97	*Streptomyces coelicoflavus*	methylisoborneol synthase	EHN74707.1
7140R	458	97/98	*Streptomyces coelicoflavus*	Crp/Fnr regulator	EHN74706.1

Table S46. Properties of genes encoded within cluster 24 (lomaiviticin-like) in *Streptomyces* sp. KO-7888.

ORF	Product size (aa)	% identity/ similarity	Species	Putative Function	Database entry
0015R	415	97/98	*Streptomyces* sp. NRRL WC-3753	KSβ	KPC71304.1
0016R	421	100/100	*Streptomyces coelicoflavus*	KSα	EHN77732.1
0017R	268	98/98	*Streptomyces* sp. NRRL WC-3753	thioesterase	KPC71296.1
0018R	615	69/78	*Streptomyces* sp. NRRL WC-3753	NRPS	EDY66720.2
0019	258	99/100	*Streptomyces* sp. NRRL WC-3753	αβ-hydrolase	KPC71297.1
0020	184	96/97	*Streptomyces coelicoflavus*	adenylylsulfate kinase	EHN80224.1
0021	312	98/99	*Streptomyces coelicoflavus*	sulfate adenylyltransferase subunit 2	EHN80225.1
0022	445	97/98	*Streptomyces coelicoflavus*	sulfate adenylyltransferase subunit 1	EHN80226.1
0023	353	96/97	*Streptomyces coelicoflavus*	sulfate ABC transporter	EHN80227.1

				substrate-binding	
0024	282	97/99	*Streptomyces coelicoflavus*	sulfate ABC transporter ATP-binding	EHN80228.1
0025	296	97/98	*Streptomyces coelicoflavus*	sulfate ABC transporter permease	EHN80229.1
0026	471	73/85	*Streptomyces* sp. HGB0020	lycopene cyclase	WP_037780904
0027	67	-	-	-	-
0028	592	98/98	*Streptomyces coelicoflavus*	multidrug ABC transporter ATPase	EHN77703.1
0029	644	98/98	*Streptomyces* sp. NRRL WC-3753	phenoxazinone synthase	EHN77704.1
0030R	87	-	-	-	-
0031	391	47/61	*Salinispora pacifica*	glycosyltrans-ferase	AHF72785.1
0032R	358	52/66	*Salinispora pacifica*	glycosyltrans-ferase Lom55	AHF72791.1
0033R	265	99/100	*Streptomyces coelicoflavus*	glucose-1-phosphate thymidylyltransferase	EHN77692.1
0034	197	99/99	*Streptomyces* sp. NRRL WC-3753	NDP-hexose-3-ketoreductase	KPC71380.1
0035R	208	99/100	*Streptomyces coelicoflavus*	dTDP-4-dehydro-rhamnose 3,5-epimerase	EHN77693.1
0036R	220	63/77	*Salinispora pacifica*	sulfotrans-ferase Lom10	AHZ61844.1
0037R	146	94/95	*Streptomyces* sp. NRRL WC-3753	alkylhydro-peroxidase	EHN77695.1

0038	340	98/99	*Streptomyces coelicoflavus*	*O*-methyltrans-ferase Lom23	EHN77696.1
0039	208	97/98	*Streptomyces* sp. NRRL WC-3753	DSBA oxidoreductase Lom20	KPC71395.1
0040R	301	99/99	*Streptomyces coelicoflavus*	NmrA regulator Lom19	EHN77697.1
0041	242	99/99	*Streptomyces* sp. NRRL WC-3753	dehydrogenase Lom18	KPC71375.1
0042	492	95/96	*Streptomyces coelicoflavus*	FAD-binding monooxygen-ase Lom17	EHN77708.1
0043	497	97/98	*Streptomyces coelicoflavus*	FAD-binding mono-oxygenase Lom16	EHN77709.1
0044	475	98/99	*Streptomyces* sp. NRRL WC-3753	protoporphyrin-ogen oxidase Lom14	KPC71373.1
0045R	573	98/98	*Streptomyces coelicoflavus*	aromatic ring hydroxylase Lom13	EHN77699.1
0046R	160	99/99	*Streptomyces coelicoflavus*	pyridoxamine 5'-phosphate oxidase	EHN77700.1
0047R	421	99/99	*Streptomyces coelicoflavus*	MFS transporter	EHN77701.1
0048R	464	39/53	*Streptomyces* sp. NRRL S-813	protein kinase	WP_030184461
0049R	259	27/47	*Frankia* sp. CN3	phytanoyl-CoA dioxygenase	WP_007511940
0050R	407	37/49	*Streptomyces olindensis*	protein kinase	KDN77276.1
0051	438	42/59	*Streptomyces bingcheng-gensis*	glycosyl transferase	ADI11531.1

0052	510	97/98	*Streptomyces* sp. NRRL WC-3753	transporter	KPC71367.1
0053	104	99/100	*Streptomyces coelicoflavus*	glyoxalase	EHN77725.1
0054	326	99/100	*Streptomyces coelicoflavus*	dTDP-glucose 4,6-dehydratase	EHN77726.1
0055	450	98/98	*Streptomyces coelicoflavus*	NDP-hexose 2,3-dehydratase Lom9	EHN77727.1
0056R	176	100/100	*Streptomyces coelicoflavus*	alkylhydro-peroxidase Lom11	EHN77723.1
0057R	245	97/99	*Streptomyces coelicoflavus*	peptidase C26 Lom12	EHN77724.1
0058	308	99/99	*Streptomyces* sp. NRRL WC-3753	ABC transporter ATP-binding Lom4	KPC71362.1
0059	279	99/99	*Streptomyces* sp. NRRL WC-3753	ABC transporter Lom3	KPC71391.1
0060	312	98/99	*Streptomyces* sp. NRRL WC-3753	nuclease Lom15	KPC71361.1
0061R	263	94/95	*Streptomyces* sp. NRRL WC-3753	AraC regulator Lom7	KPC71360.1
0062	140	98/98	*Streptomyces coelicoflavus*	glyoxalase Lom5	EHN77715.1
0063	156	96/98	*Streptomyces* sp. NRRL WC-3753	AraC regulator	KPC71358.1
0064	210	85/90	*Streptomyces* sp. NRRL S-813	glyoxalase	WP_030184415
0065R	138	99/99	*Streptomyces coelicoflavus*	apoprotein precursor	EHN77713.1
0066	139	80/87	*Streptomyces azureus*	transposase	GAP50821.1|
0067R	262	100/100	*Streptomyces* sp. NRRL WC-3753	two-component system response regulator	KPC71390.1
0068R	134	98/100	*Streptomyces coelicoflavus*	*N*-acetyltrans-ferase GCN5 Lom35	EHN77656.1

0069R	428	99/99	*Streptomyces coelicoflavus*	adenylo-succinate lyase Lom34	EHN77655.1
0070R	506	97/97	*Streptomyces* sp. NRRL WC-3753	glutamine synthetase Lom32	KPC71351.1
0071R	462	99/99	*Streptomyces coelicoflavus*	amidase Lom33	EHN77687.1
0072R	635	-	-	Lom29?	-
0073R	109	99/100	*Streptomyces coelicoflavus*	Fe-4S ferredoxin	EHN77682.1
0074	360	99/99	*Streptomyces coelicoflavus*	*O*-methyltrans-ferase	EHN77684.1
0075R	514	98/98	*Streptomyces* sp. NRRL WC-3753	FAD-mono-oxygenase	KPC71387.1
0076R	510	97/97	*Streptomyces* sp. NRRL WC-3753	FAD-mono-oxygenase	KPC71348.1
0077R	314	99/99	*Streptomyces coelicoflavus*	cyclase/ dehydratase	EHN77676.1
0078R	261	100/100	*Streptomyces coelicoflavus*	ketoacyl reductase	EHN77677.1
0079R	91	100/100	*Streptomyces coelicoflavus*	ACP act-like	EHN77678.1
0080R	109	100/100	*Streptomyces coelicoflavus*	polyketide cyclase	EHN77679.1
0081R	231	98/99	*Streptomyces* sp. NRRL WC-3753	anthrone oxidase	KPC71343.1
0082R	477	98/99	*Streptomyces* sp. NRRL WC-3753	FAD-mono-oxygenase	KPC71342.1
0083	185	59/70	*Streptomyces murayamaensis*	JadX-like protein	AAO65342.1
0084R	335	98/98	*Streptomyces coelicoflavus*	thiamine biosynthesis protein ThiF	EHN77668.1
0085R	604	58/67	*Streptomyces* sp. NRRL S-87	acyl-CoA dehydrogenase	WP_051796199
0086	326	99/99	*Streptomyces coelicoflavus*	cyclase/act dimerase	EHN80232.1

0087	418	96/97	*Streptomyces* sp. NRRL WC-3753	MFS transporter	KPC71339.1
0088	74	-	-	-	-
0089R	273	87/93	*Streptomyces pristinaespiralis*	SARP regulator Lom7	CBW45731.1
0090R	276	80/88	*Streptomyces azureus*	SARP regulator	GAP47966.1
0091R	208	96/97	*Streptomyces coelicoflavus*	TetR regulator	EHN80235.1
0092R	521	99/99	*Streptomyces coelicoflavus*	methylmalonyl-CoA carboxyl-transferase	EHN80236.1
0093R	69	-	-	-	-
0094	321	63/78	*Streptomyces aureofaciens*	Afs-like γ-butyrolactone synthase	ADM72848.2
0095R	289	97/97	*Streptomyces* sp. NRRL WC-3753	nucleoside-diphosphate sugar epimerase	KPC71259.1
0096	222	98/99	*Streptomyces coelicoflavus*	γ-butyrolactone receptor protein	EHN80268.1

Bibliography

Abraham, E.P., and Chain, E. (1940). An enzyme from bacteria able to destroy penicillin. *Nature* **146**, 837.

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

Achkar, J., Xian, M., Zhao, H., and Frost, J. (2005). Biosynthesis of phloroglucinol. *J. Am. Chem. Soc.* **127**, 5332–5333.

Aggarwal, R., Caffrey, P., Leadlay, P.F, Smith, C.J., and Staunton, J. (1995). The thioesterase of the erythromycin-producing polyketide synthase: mechanistic studies *in vitro* to investigate its mode of action and substrate specificity *J. Chem. Soc., Chem. Commun.* **15**, 1519–1520.

Akiva, E., Brown, S., Almonacid, D.E., Barber, A.E. 2nd, Custer, A.F., Hicks, M.A., Huang, C.C., Lauck, F., Mashiyama, S.T., Meng, E.C., Mischel, D., Morris, J.H., Ojha, S., Schnoes, A.M., Stryke, D., Yunes, J.M., Ferrin, T.E., Holliday, G.L., and Babbitt, P.C. (2014). The Structure-Function Linkage Database. *Nucl. Acids Res.* **42**, D521–D530.

Akiyama, T., Harada, S., Kojima, F., Takahashi, Y., Imada, C., Okami, Y., Muraoka, Y., Aoyagi, T., and Takeuchi, T. (1998). Fluostatins A and B, new inhibitors of dipeptidyl peptidase III, produced by Streptomyces sp. TA-3391. I. Taxonomy of producing strain, production, isolation, physico-chemical properties and biological properties. *J. Antibiot.* **51**, 553–559.

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

Alhamadsheh, M.M., Palaniappan, N., Daschouduri, S., and Reynolds, K.A. (2007). Modular polyketide synthases and cis double bond formation: establishment of activated *cis*-3-cyclohexylpropenoic acid as the diketide intermediate in phoslactomycin biosynthesis. *J. Am. Chem. Soc.* **129**, 1910–1911.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic

local alignment search tool. *J. Mol. Biol*. **215**, 403–410.

Ames, B.D., Korman, T.P., Zhang, W., Smith, P., Vu, T., Tang, Y., and Tsai, S.C. (2008). Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5349–5354.

AMR Review (Review on Antimicrobial Resistance). (2016). Tackling drug-resistant infections globally: Final report and recommendations. London: Review on Antimicrobial Resistance.

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

Arakawa, K., Sugino, F., Kodama, K., Ishii, T., and Kinashi, H. (2005). Cyclization mechanism for the synthesis of macrocyclic antibiotic lankacidin in *Streptomyces rochei*. *Chem. Biol.* **12**, 249–256.

Aravind, L., Anantharaman, V., and Koonin, E.V. (2002). Monophyly of class I aminoacyl tRNA synthetase, USPA, ETFP, photolyase, and PP-ATPase nucleotide-binding domains: implications for protein evolution in the RNA. *Proteins* **48**, 1–14.

Auclair, K., Kennedy, J., Hutchinson, C.R., and Vederas, J.C. (2001). Conversion of cyclic nonaketides to lovastatin and compactin by a lovC deficient mutant of *Aspergillus terreus*. *Bioorg. Med. Chem. Lett*. **11**, 1527–1531.

Awodi, U.R., Ronan, J.L., Masschelein, J., de Los Santos, E.L.C., and Challis, G.L. (2017). Thioester reduction and aldehyde transamination are universal steps in actinobacterial polyketide alkaloid biosynthesis. *Chem. Sci*. **8**, 411–415.

Baerga-Ortiz, A., Popovic, B., Siskos, A.P., O'Hare, H.M., Spiteller, D., Williams, M.G., Campillo, N., Spencer, J.B., and Leadlay, P.F. (2006). Directed mutagenesis alters the stereochemistry of catalysis by isolated ketoreductase domains from the erythromycin polyketide synthase. *Chem. Biol.* **13**, 277–285.

Bailey, A.M., Cox, R.J., Harley, K., Lazarus, C.M., Simpson, T.J., and Skellam, E. (2007). Characterisation of 3-methylorcinaldehyde synthase (MOS) in *Acremonium strictum*: first observation of a reductive release mechanism during polyketide biosynthesis. *J. Chem. Soc., Chem. Commun.* **39**, 4053–4055.

Baldwin, T.O., Christopher, J.A., Raushel, F.M., Sinclair, J.F., Ziegler,

M.M., Fisher, A.J., and Rayment, I. (1995). Structure of bacterial luciferase. *Curr. Opin. Struct. Biol.* **5,** 798–809.

Balibar, C.J., and Walsh, C.T. (2006). GliP, a multimodular nonribosomal peptide synthetase in Aspergillus fumigatus, makes the diketopiperazine scaffold of gliotoxin. *Biochemistry* **45**, 15029–15038.

Baltz, R.H., and Seno, E.T. (1981). Properties of Streptomyces fradiae mutants blocked in biosynthesis of the macrolide antibiotic tylosin. *Antimicrob. Agents Chemother.* **20**, 214–225.

Barajas, J.F., Shakya, G., Moreno, G., Rivera, H. Jr., Jackson, D.R., Topper, C.L., Vagstad, A.L., La Clair, J.J., Townsend, C.A., Burkart, M.D., and Tsai, S.C. (2017). Polyketide mimetics yield structural and mechanistic insights into product template domain function in nonreducing polyketide synthases. *Proc. Natl. Acad. Sci. U.S.A*. **114**, E4142–E4148.

Baur, S., Niehaus, J., Karagouni, A.D., Katsifas, E.A., Chalkou, K., Meintanis, C., Jones, A.L., Goodfellow, M., Ward, A.C., Beil, W., Schneider, K., Süssmuth, R.D., and Fiedler, H.P. (2006). Fluostatins C-E, novel members of the fluostatin family produced by *Streptomyces* strain Acta 1383. *J. Antibiot.* **59**, 293–297.

Bayer, A., Freund, S., and Jung, G. (1995). Post-translational heterocyclic backbone modifications in the 43-peptide antibiotic microcin B17. Structure elucidation and NMR study of a ^{13}C,^{15}N-labelled gyrase inhibitor. *Eur. J. Biochem*. **234**, 414–426.

Beamer, L.J., and Pabo, C.O. (1992). Refined 1.8 Å crystal structure of the lambda repressor-operator complex. *J. Mol. Biol.* **227**, 177–196.

Belshaw, P.J., Walsh, C.T., and Stachelhaus, T. (1999). Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284**, 486–489.

Bentley, S.D., Chater, K.F., Cerdeño-Tárraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J., and Hopwood, D.A.

(2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141–147.

Bérdy, J. (2005). Bioactive microbial metabolites. *J. Antibiot.* **58**, 1–26.

Betlach, M.C., Kealey, J.T., Ashley, G.W., and McDaniel, R. (1998). Characterization of the macrolide P-450 hydroxylase from *Streptomyces venezuelae* which converts narbomycin to picromycin. *Biochemistry* **37**, 14937–14942.

Bignell, D.R., Seipke, R.F., Huguet-Tapia, J.C., Chambers, A.H., Parry, R.J., and Loria, R. (2010). *Streptomyces scabies* 87-22 contains a coronafacic acid-like biosynthetic cluster that contributes to plant-microbe interactions. *Mol. Plant Microbe Interact.* **23**, 161–175.

Bignell, D.R., Cheng, Z., and Bown, L. (2018). The coronafacoyl phytotoxins: structure, biosynthesis, regulation and biological activities. *Antonie Van Leeuwenhoek* **111**, 649–666.

Bindman, N.A., and Van Der Donk, W.A. ed., (2014). RiPPs: Ribosomally Synthesized and Posttranslationally Modified Peptides. In: Natural Products: Discourse, Diversity, and Design, first edition. Hoboken: John Wiley & Sons, Inc., 197–218.

Birch, A., Häusler, A., and Hütter, R. (1990). Genome rearrangement and genetic instability in *Streptomyces* spp. *J. Bacteriol.* **172**, 4138–4142.

Bisang, C., Long, P.F., Cortés, J., Westcott, J., Crosby, J., Matharu, A.L., Cox, R.J., Simpson, T.J., Staunton, J., and Leadlay, P.F. (1999). A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* **401**, 502–505.

Blin K, Medema MH, Kottmann R, Lee SY, Weber T. (2017). The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucl. Acids Res.* **45**, D555–D559.

Bloudoff, K., and Schmeing, T.M. (2017). Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim. Biophys. Acta.* **1865**, 1587–1604.

Bode, H.B., Bethe, B., Höfs, R., and Zeeck, A. (2002). Big effects from small changes: possible ways to explore nature's chemical diversity. *ChemBioChem* **3**,

619–627.

Boehlein, S.K., Schuster, S.M., and Richards, N.G. (1996). Glutamic acid gamma-monohydroxamate and hydroxylamine are alternate substrates for Escherichia coli asparagine synthetase B. *Biochemistry* **35**, 3031–3037.

Brigham, R.B., and Pittenger, R.C. (1956). *Streptomyces orientalis*, n. sp., the source of vancomycin. *Antibiot. Chemother*. **6**, 642–647.

Brotzu, G. (1948). Ricerche su di un nuovo antibiotico. Lav. Ist. Ig. Cagliari, 1–11.

Bruns, H., Crüsemann, M., Letzel, A.C., Alanjary, M., McInerney, J.O., Jensen, P.R., Schulz, S., Moore, B.S., and Ziemert, N. (2018). Function-related replacement of bacterial siderophore pathways. *ISME J.* **12**, 320–329.

Bunet, R., Song, L., Mendes, M.V., Corre, C., Hotel, L., Rouhier, N., Framboisier, X., Leblond, P., Challis, G.L., and Aigle, B. (2010). Characterization and manipulation of the pathway-specific late regulator AlpW reveals *Streptomyces ambofaciens* as a new producer of Kinamycins. *J. Bacteriol*. **193**, 1142–1153.

Butz D, Schmiederer T, Hadatsch B, Wohlleben W, Weber T, Süssmuth RD. (2008). Module extension of a non-ribosomal peptide synthetase of the glycopeptide antibiotic balhimycin produced by *Amycolatopsis balhimycina*. *ChemBioChem* **9**, 1195–1200.

Caffrey, P. (2003). Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *ChemBioChem* **4**, 654–657.

Cane, D.E., and Watt, R.M. (2003). Expression and mechanistic analysis of a germacradienol synthase from *Streptomyces coelicolor* implicated in geosmin biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1547–1551.

Cane, D.E., He, X., Kobayashi, S., Ōmura, S., and Ikeda, H. (2006). Geosmin Biosynthesis in *Streptomyces avermitilis*. Molecular Cloning, Expression, and Mechanistic Study of the Germacradienol/Geosmin Synthase *J. Antibiot*. **59**, 471–479.

Cano-Prieto, C., García-Salcedo, R., Sánchez-Hidalgo, M., Braña, A.F., Fiedler, H.P., Méndez, C., Salas, J.A., and Olano, C. (2015). Genome Mining of Streptomyces sp. Tü 6176: Characterization of the Nataxazole Biosynthesis Pathway. *ChemBioChem* **16**, 1461–1473.

Carvalho, R., Reid, R., Viswanathan, N., Gramajo, H., and Julien, B. (2005) The

biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation. *Gene* **359**, 91–98.

Chain, E., Florey, H.W., Gardner, A.D., Heatley, N.G., Jennings, M.A., Orr-Ewing, J., and Sanders, A.G. (1940). Penicillin as a chemotherapeutic agent. *Lancet* **236**, 226–228.

Challis, G.L., and 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 Microbiol. Lett.* **187**, 111–114.

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

Chang, Z., Sitachitta, N., Rossi, J.V., Roberts, M.A., Flatt, P.M., Jia, J., Sherman, D.H., and Gerwick, W.H. (2004). Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod*. **67**, 1356–1367.

Chater, K.F., and Bibb, M.J. (1997). Regulation of bacterial antibiotic production. In Biotechnology, Vol 7: Products of Secondary Metabolism. Kleinkauf, H., and Von Döhren, H., eds. Weinheim: Wiley-VCH, 57–105.

Chen, D., Zhang, L., Pang, B., Chen, J., Xu, Z., Abe, I., and Liu, W. (2013). FK506 maturation involves a cytochrome p450 protein-catalyzed four-electron C-9 oxidation in parallel with a C-31 *O*-methylation. *J. Bacteriol*. **195**, 1931–1919.

Cheng, K., Rong, X., and Huang, Y. (2016). Widespread interspecies homologous recombination reveals reticulate evolution within the genus *Streptomyces*. *Mol. Phylogenet. Evol.* **102**, 246–254.

Cheng, Y.Q., Tang, G.L., and Shen, B. (2003). Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3149–3154.

Chevrette, M.G., Aicheler, F., Kohlbacher, O., Currie, C.R., and Medema, M.H. (2017). SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across Actinobacteria. *Bioinformatics* **33**, 3202–3210.

Cimermancic, P., Medema, M.H., Claesen, J., Kurita, K., Wieland Brown, L.C.,

Mavrommatis, K., Pati, A., Godfrey, P.A., Koehrsen, M., Clardy, J., Birren, B.W., Takano, E., Sali, A., Linington, R.G., and Fischbach, M.A. (2014). Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412–421.

Colis, L.C., Woo, C.M., Hegan, D.C., Li, Z., Glazer, P.M., and Herzon, S.B. (2014). The cytotoxicity of (-)-lomaiviticin A arises from induction of double-strand breaks in DNA. *Nat. Chem*. **6**, 504–510.

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

Coombs, J.T., and Franco, C.M. (2003). Visualization of an endophytic Streptomyces species in wheat seed. *Appl. Environ. Microbiol.* **69**, 4260–4262.

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

Cortés, J., Velasco, J., Foster, G., Blackaby, A.P., Rudd, B.A.M., and Wilkinson, B. (2002). Identification and cloning of a type III polyketide synthase required for diffusible pigment biosynthesis in *Saccharopolyspora erythraea*. *Mol. Microbiol*. **44**, 1213–1224.

Cortes, J., Wiesmann, K.E., Roberts, G.A., Brown, M.J., Staunton, J., and Leadlay,P.F. (1995). Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science* **268**, 1487–1489.

Cruz-Morales, P., Kopp, J.F., Martínez-Guerrero, C., Yáñez-Guerra, L.A., Selem-Mojica, N., Ramos-Aboites, H., Feldmann, J., and Barona-Gómez, F. (2016). Phylogenomic Analysis of Natural Products Biosynthetic Gene Clusters Allows Discovery of Arseno-Organic Metabolites in Model *Streptomycetes*. *Genome Biol. Evol.* **8**, 1906–1916.

Currie, C.R., Scott, J.A., Summerbell, R.C., and Malloch, D. (1999). Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. *Nature* **398**, 701–704.

Davison, J., Dorival, J., Rabeharindranto, H., Mazon, H., Chagot, B., Gruez, A.,

and Weissman, K.J. (2014). Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module *Chem. Sci.* **5**, 3081–3095

Decker, H., and Haag, S. (1995). Cloning and characterization of a polyketide synthase gene from *Streptomyces fradiae* Tü2717, which carries the genes for biosynthesis of the angucycline antibiotic urdamycin A and a gene probably involved in its oxygenation. *J. Bacteriol*. **177**, 6126–6136.

Del Vecchio, F., Petkovic, H., Kendrew, S.G., Low, L., Wilkinson, B., Lill, R., Cortés, J., Rudd, B.A., Staunton, J., and Leadlay, P.F. (2003). Active-site residue, domain and module swaps in modular polyketide synthases. *J. Ind. Microbiol. Biotechnol.* **30**, 489–494.

Deluca, M.R., and Kerwin, S.M. (1997). Total synthesis of UK-1. *Tetrahedron Lett.* **38**, 199–202.

Demmer, C.S., and Bunch, L. (2015). Benzoxazoles and oxazolopyridines in medicinal chemistry studies. *Eur. J. Med. Chem.* **97**, 778–785.

Di Giambattista, M., Chinali, G., and Cocito, C. (1989). The molecular basis of the inhibitory activities of type A and type B synergimycins and related antibiotics on ribosomes. *J. Antimicrob. Chemother.* **24**, 485–507.

Ding, Y., Yu, Y., Pan, H., Guo, H., Li, Y., and Liu, W. (2010). Moving posttranslational modifications forward to biosynthesize the glycosylated thiopeptide nocathiacin I in *Nocardia* sp. ATCC202099. *Mol. Biosyst*. **6**, 1180–1185.

Dixon, J.M. (1968). Group A Streptococcus resistant to erythromycin and lincomycin. *Can. Med. Assoc. J*. **99**, 1093–1094.

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

Drake, E.J., Miller, B.R., Shi, C., Tarrasch, J.T., Sundlov, J.A., Allen, C.L., Skiniotis, G., Aldrich, C.C., and Gulick, A.M. (2016). Structures of two distinct conformations of *holo*-nonribosomal peptide synthetases. *Nature* **529**, 235–238.

Dubos, R.J. (1939). Studies on a bactericidal agent extracted from a soil *Bacillus*: I.

Preparation of the agent. Its activity *in vitro*. *J. Exp. Med*. **70**, 1–10.

Dubos, R.J., and Hotchkiss, R.D. (1941). The production of bactericidal substances by aerobic sporulating *Bacilli*. *J. Exp. Med.* **73**, 629–640.

Duggar, B.M. (1948). Aureomycin-a New Antibiotic. *Ann. N. Y. Acad. Sci.* **51**, 177–342.

Dunbar, K.L., Chekan, J.R., Cox, C.L., Burkhart, B.J., Nair, S.K., and Mitchell, D.A. (2014). Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. *Nat. Chem. Biol*. **10**, 823–829.

Dunbar, K.L., Melby, J.O., and Mitchell, D.A. (2012). YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nat. Chem. Biol.* **8**, 569–575.

Duquesne, S., Destoumieux-Garzón, D., Zirah, S., Goulard, C., Peduzzi, J., and Rebuffat, S. (2007). Two enzymes catalyze the maturation of a lasso peptide in *Escherichia coli*. *Chem. Biol.* **14**, 793–803.

Dutta, S., Whicher, J.R., Hansen, D.A., Hale, W.A., Chemler, J.A., Congdon, G.R., Narayan, A.R., Håkansson, K., Sherman, D.H., Smith, J.L., and Skiniotis, G. (2014). Structure of a modular polyketide synthase. *Nature* **510**, 512–517.

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

Edwards, A.L., Matsui, T., Weiss, T.M., and Khosla, C. (2014). Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase. *J. Mol. Biol.* **426**, 2229–2245.

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

Ehrlich, J., Bartz, Q.R., Smith, R.M., Joslyn, D.A., and Burkholder, P.R. (1947). Chloromycetin, a New Antibiotic From a Soil Actinomycete. *Science* **106**, 417.

Feitelson, J.S., Malpartida, F., and Hopwood, D.A. (1985). Genetic and biochemical characterization of the red gene cluster of *Streptomyces coelicolor* A3(2). *J. Gen. Microbiol.* **131**, 2431–2441.

Fernández-Moreno, M.A., Martínez, E., Boto, L., Hopwood, D.A., and Malpartida,

F. (1992). Nucleotide sequence and deduced functions of a set of cotranscribed genes of Streptomyces coelicolor A3(2) including the polyketide synthase for the antibiotic actinorhodin. *J. Biol. Chem*. **267**, 19278–19290.

Firn, R.D., and Jones, C.G. (2000). The evolution of secondary metabolism - a unifying model. *Mol. Microbiol*. **37**, 989–994.

Flärdh, K, and Buttner MJ. (2009). *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat. Rev. Microbiol.* **7**, 36–49.

Fleming, A. (1929). On the antibacterial action of cultures of a Penicillium, with special reference to their use in the isolation of B. influenzæ. *Br. J. Exp. Pathol*. **10**, 226–236.

Floriano, B., and Bibb, M. (1996). *AfsR* is a pleiotropic but conditionally required regulatory gene for antibiotic production in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **21**, 385–396.

Forsman, M., Häggström, B., Lindgren, L., and Jaurin, B. (1990). Molecular analysis of β-lactamases from four species of *Streptomyces*: comparison of amino acid sequences with those of other β-clactamases. *J. Gen. Microbiol*. **136**, 589–598.

Frank, B., Wenzel, S.C., Bode, H.B., Scharfe, M., Blöcker, H., and Müller, R. (2007). From genetic diversity to metabolic unity: studies on the biosynthesis of aurafurones and aurafuron-like structures in myxobacteria and streptomycetes. *J. Mol. Biol.* **374**, 24–38.

Gaisser, S., Bohm, G.A., Doumith, M., Raynal, M.C., Dhillon, N., Cortés, J., and Leadlay, P.F. (1998). Analysis of *eryBI*, *eryBIII* and *eryBVII* from the erythromycin biosynthetic gene cluster in *Saccharopolyspora erythraea*. *Mol. Gen. Genet*. **258**, 78–88.

Gaitatzis, N., Kunze, B., and Müller, R. (2001). In vitro reconstitution of the myxochelin biosynthetic machinery of *Stigmatella aurantiaca* Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11136–11141.

Gallimore, A.R., Stark, C.B., Bhatt, A., Harvey, B.M., Demydchuk, Y., Bolanos-Garcia, V., Fowler, D.J., Staunton, J., Leadlay, P.F., and Spencer, J.B. (2006). Evidence for the role of the monB genes in polyether ring formation during

monensin biosynthesis. *Chem. Biol.* **13**, 453–460.

Gardner, A.D. (1940). Morphological effects of penicillin on bacteria. *Nature* **146**, 837–838.

Gatzeva-Topalova, P.Z., May, A.P., and Sousa, M.C. (2005). Crystal structure and mechanism of the *Escherichia coli* ArnA (PmrI) transformylase domain. An enzyme for lipid A modification with 4-amino-4-deoxy-L-arabinose and polymyxin resistance. *Biochemistry* **44**, 5328–5338.

Gavrish, E., Sit, C.S., Cao, S., Kandror, O., Spoering, A., Peoples, A., Ling, L., Fetterman, A., Hughes, D., Bissell, A., Torrey, H., Akopian, T., Mueller, A., Epstein, S., Goldberg, A., Clardy, J., and Lewis, K. (2014). Lassomycin, a ribosomally synthesized cyclic peptide, kills *Mycobacterium tuberculosis* by targeting the ATP-dependent protease ClpC1P1P2. *Chem. Biol.* **21**, 509–518.

Gay, D.C., Gay, G., Axelrod, A.J., Jenner, M., Kohlhaas, C., Kampa, A., Oldham, N.J., Piel, J., and Keatinge-Clay, A.T. (2014). A close look at a ketosynthase from a *trans*-acyltransferase modular polyketide synthase. *Structure* **22**, 444–451.

Geng, P., Bai, G., Shi, Q., Zhang, L., Gao, Z., and Zhang, Q. (2009). Taxonomy of the *Streptomyces* strain ZG0656 that produces acarviostatin alpha-amylase inhibitors and analysis of their effects on blood glucose levels in mammalian systems. *J. Appl. Microbiol.* **106**, 525–533.

Ghimire, G.P., Oh, T.J., Lee, H.C., and Sohng, J.K. (2009). Squalene-hopene cyclase (Spterp25) from *Streptomyces peucetius*: sequence analysis, expression and functional characterization. *Biotechnol. Lett*. **31**, 565–569.

Gish, W., and States, D.J. (1993). Identification of protein coding regions by database similarity search. *Nat. Genet.* **3**, 266–272.

Gojkovic, Z., Rislund, L., Andersen, B., Sandrini, M.P., Cook, P.F., Schnackerz, K.D., and Piskur, J. (2003). Dihydropyrimidine amidohydrolases and dihydroorotases share the same origin and several enzymatic properties. *Nucl. Acids Res.* **31**, 1683–1692.

Gokhale, R.S., Tsuji, S.Y., Cane, D.E., and Khosla, C. (1999). Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **28**4, 482–485.

Gomez-Escribano, J.P., Alt, S., and Bibb, M.J. (2016). Next generation sequencing

of actinobacteria for the discovery of novel natural products. *Mar. Drugs*. **14**, 78.

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

Gomez-Escribano, J.P., and Bibb, M.J. (2014). Heterologous expression of natural product biosynthetic gene clusters in Streptomyces coelicolor: from genome mining to manipulation of biosynthetic pathways. *J. Ind. Microbiol. Biotechnol*. **41**, 425–431.

Goodwin, C.R., Covington, B.C., Derewacz, D.K., McNees, C.R., Wikswo, J.P., McLean, J.A., and Bachmann, B.O. (2015). Structuring microbial metabolic responses to multiplexed stimuli via self-organizing metabolomics maps. *Chem. Biol.* **22**, 661–670.

Gould, S.J., Hong, S.T., and Carney, J.R. (1998). Cloning and heterologous expression of genes from the kinamycin biosynthetic pathway of *Streptomyces murayamaensis*. *J. Antibiot*. **51**, 50–57.

Gould, S.J., Melville, C.R., Cone, M.C., Chen, J., and Carney, J.R. (1997). Kinamycin biosynthesis. Synthesis, isolation, and incorporation of stealthin C, an aminobenzo[*b*]fluorene. *J. Org. Chem*. **62**, 320–324.

Graham, M.Y., and Weisblum, B. (1979). Altered methylation of adenine in 23 S ribosomal RNA associated with erythromycin resistance in Streptomyces erythreus and *Streptococcus faecalis. Contrib. Microbiol. Immunol.* **6**, 159–164.

Grande-García, A., Lallous, N., Díaz-Tejada, C., and Ramón-Maiques, S. (2014). Structure, functional characterization, and evolution of the dihydroorotase domain of human CAD. *Structure* **22**, 185–198.

Griffiths, G.L., Sigel, S.P., Payne, S.M., and Neilands, J.B. (1984). Vibriobactin, a siderophore from *Vibrio cholerae*. *J. Biol. Chem.* **259**, 383–385.

Grosdidier, A., Zoete, V., and Michielin, O. (2011a). Fast docking using the CHARMM force field with EADock DSS. *J. Comput. Chem*. **32**, 2149–2159.

Grosdidier, A., Zoete, V., and Michielin, O. (2011b). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucl. Acids Res*. **39**, W270–W277.

Guo, Y., Zheng, W., Rong, X., and Huang, Y. (2008). A multilocus phylogeny of the

Streptomyces griseus 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics. *Int. J. Syst. Evol. Microbiol*. **58**, 149–159.

Haeder, S., Wirth, R., Herz, H., and Spiteller, D. (2009). Candicidin-producing *Streptomyces* support leaf-cutting ants to protect their fungus garden against the pathogenic fungus *Escovopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4742–4746.

Hale, K.J., Hatakeyama, S., Urabe, F., Ishihara, J., Manaviazar, S., Grabski, M., and Maczka, M. (2014). The absolute configuration for inthomycin C: revision of previously published work with a reinstatement of the (3*R*)-configuration for (−)-inthomycin C. *Org. Lett*. **16**, 3536–3539.

Han, J.W., Kim, E.Y., Lee, J.M., Kim, Y.S., Bang, E., and Kim, B.S. (2012). Site-directed modification of the adenylation domain of the fusaricidin nonribosomal peptide synthetase for enhanced production of fusaricidin analogs. *Biotechnol. Lett.* **34**, 1327–1334.

Harvey, B.M., Hong, H., Jones, M.A., Hughes-Thomas, Z.A., Goss, R.M., Heathcote, M.L., Bolanos-Garcia, V.M., Kroutil, W., Staunton, J., Leadlay, P.F., and Spencer, J.B. (2006) Evidence that a novel thioesterase is responsible for polyketide chain release during biosynthesis of the polyether ionophore monensin. *ChemBioChem* **7**, 1435–1442.

He, H., Ding, W.D., Bernan, V.S., Richardson, A.D., Ireland, C.M., Greenstein, M., Ellestad, G.A., and Carter, G.T. (2001). Lomaiviticins A and B, potent antitumor antibiotics from *Micromonospora lomaivitiensis*. *J. Am. Chem. Soc*. **123**, 5362–5363.

Heath, R.J., and Rock, C.O. (2002). The Claisen condensation in biology. *Nat. Prod. Rep.* **19**, 581–596.

Heine, D., Bretschneider, T., Sundaram, S., and Hertweck, C. (2014). Enzymatic polyketide chain branching to give substituted lactone, lactam, and glutarimide heterocycles. *Angew. Chem. Int. Ed. Engl.* **53**, 11645–11649.

Hemmerling, F., and Hahn, F. (2016). Biosynthesis of oxygen and nitrogen-containing heterocycles in polyketides. Beilstein. *J. Org. Chem*. **12,** 1512–1250.

Henkel, T., and Zeeck, A. (1991). Secondary metabolites by chemical screening, 15. Structure and absolute configuration of naphthomevalin, a new dihydro-naphthoquinone antibiotic from *Streptomyces* sp. *J. Antibiot*. **44**, 665–669.

Henkel, T., and Zeeck, A. (1991). Sekundärstoffe aus dem chemischen Screening, 16. Inthomycine, neue oxazol-triene aus *Streptomyces* sp. *Liebigs Ann. Chem*. 367–373.

Ho, Y.T.C., Leng, D.J., Ghiringhelli F, Wilkening I, Bushell DP, Köstner O, Riva E, Havemann J, Passarella D, Tosin M. (2017). Novel chemical probes for the investigation of nonribosomal peptide assembly. *J. Chem. Soc., Chem. Commun.* **53**, 7088–7091.

Hobbs, M.E., Malashkevich, V., Williams, H.J., Xu, C., Sauder, J.M., Burley, S.K., Almo, S.C., and Raushel, F.M. (2012). Structure and catalytic mechanism of LigI: insight into the amidohydrolase enzymes of cog3618 and lignin degradation. *Biochemistry* **51**, 3497–3507.

Hodgkin, D.C. (1949). The X-ray analysis of the structure of penicillin. *Adv. Sci.* **6**, 85–89.

Holm, L., and Sander, C. (1997). An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. *Proteins* **28**, 72–82.

Hood, D.W., Heidstra, R., Swoboda, U.K., and Hodgson, D.A. (1992). Molecular genetic analysis of proline and tryptophan biosynthesis in *Streptomyces coelicolor* A3(2): interaction between primary and secondary metabolism - a review. *Gene* **115**, 5–12.

Horinouchi, S., and Beppu, T. (1994). A-factor as a microbial hormone that controls cellular differentiation and secondary metabolism in *Streptomyces griseus*. *Mol. Microbiol*. **12**, 859–864.

Horinouchi, S., and Beppu, T. (2007). Hormonal control by A-factor of morphological development and secondary metabolism in *Streptomyces*. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci*. **83**, 277–295.

Hover, B.M., Kim, S.H., Katz, M., Charlop-Powers, Z., Owen, J.G., Ternei, M.A., Maniko, J., Estrela, A.B., Molina, H., Park, S., Perlin, D.S., and Brady, S.F. (2018). Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat. Microbiol*. **3**, 415–422.

Hug, J. (2015). Enzymatic logic in biosynthesis of benzoxazole antibiotics. Diplom-Pharmazeut. Albert-Ludwigs-University Freiburg, Freiburg.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., and Omura, S. (2003). Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **21**, 526–531.

Ikeda, H., Kazuo, S.Y., and Omura, S. (2014). Genome mining of the Streptomyces avermitilis genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. *J. Ind. Microbiol. Biotechnol.* **41**, 233–250.

Iqbal, H.A., Low-Beinart, L., Obiajulu, J.U., and Brady, S.F. (2016). Natural product discovery through improved functional metagenomics in *Streptomyces*. *J. Am. Chem. Soc.* **138**, 9341–9344.

Itŏ, S., Matsuya, T., Ômura, S., Otani, M., and Nakagawa, A. (1970). A new antibiotic, kinamycin. *J. Antibiot.* **23**, 315–317.

Janso, J.E., Haltli, B.A., Eustáquio, A.S., Kulowski, K., Waldman, A.J., Zha, L., Nakamura, H., Bernan, V.S., He, H., Carter, G.T., Koehn, F.E., and Balskus, E.P. (2014). Discovery of the lomaiviticin biosynthetic gene cluster in *Salinispora pacifica*. *Tetrahedron* **70**, 4156–4164.

Jenke-Kodama, H., Sandmann, A., Müller, R., and Dittmann, E. (2005). Evolutionary implications of bacterial polyketide synthases. *Mol. Biol. Evol*. **22**, 2027–2039.

Jensen K, Niederkrüger H, Zimmermann K, Vagstad AL, Moldenhauer J, Brendel N, Frank S, Pöplau P, Kohlhaas C, Townsend CA, Oldiges M, Hertweck C, Piel J. (2012). Polyketide proofreading by an acyltransferase-like enzyme. *Chem. Biol.* **19**, 329–339.

Jevons, M.P., Coe, A.W., and Parker, M.T. (1963). Methicillin resistance in staphylococci. *Lancet.* **1**, 904–907.

Jiang, X., Zhang, Q., Zhu, Y., Nie, F., Wu, Z., Yang, C., Zhang, L., Tian, X., and Zhang, C. (2017). Isolation, structure elucidation and biosynthesis of benzo[b]fluorene nenestatin A from deep-sea derived Micromonospora echinospora SCSIO 04089. *Tetrahedron* **73**, 3585–3590.

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

Joule, J.A., and Mills, K. (2010). Heterocyclic Chemistry, fifth edition. Wiley, 2010. Oxford: Wiley-Blackwell.

Kagan, R.M., and Clarke, S. (1994). Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. *Arch. Biochem. Biophys.* **310**, 417–427.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Y., Lu, H., and Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc*. **7**, 1511–1522.

Kaltenpoth, M., Yildirim, E., Gürbüz, M.F., Herzner, G., and Strohm, E. (2012). Refining the roots of the beewolf-Streptomyces symbiosis: antennal symbionts in the rare genus Philanthinus (Hymenoptera, Crabronidae). *Appl. Environ. Microbiol*. **78**, 822–827.

Kanzaki, H., Wada, K., Nitoda, T., and Kawazu, K. (1998). Novel bioactive oxazolomycin isomers produced by *Streptomyces albus* JA3453. *Biosci. Biotechnol. Biochem.* **62**, 438–442.

Kao, C.M., Luo, G., Katz, L., Cane, D.E., and Khosla, C. (1995). Manipulation of macrolide ring size by directed mutagenesis of a modular polyketide synthase. *J. Am. Chem. Soc.* **117**, 9105–9106.

Kaulmann, U., and Hertweck, C. (2002). Biosynthesis of polyunsaturated fatty acids by polyketide synthases. *Angew. Chem. Int. Ed. Engl.* **41**, 1866–1869.

Kawada, M., Inoue, H., Usami, I., and Ikeda, D. (2009). Phthoxazolin A inhibits prostate cancer growth by modulating tumor-stromal cell interactions. *Cancer Sci.* **100**, 150–157.

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

Keatinge-Clay, A.T. (2007). A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol*. **14**, 898–908.

Keatinge-Clay, A.T. (2016). Stereocontrol within polyketide assembly lines. *Nat. Prod. Rep.* **33**, 141–149.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis *Nat. Protoc.* **10**, 845–858.

Kendrew, S.G., Petkovic, H., Gaisser, S., Ready, S.J., Gregory, M.A., Coates, N.J., Nur-E-Alam, M., Warneck, T., Suthar, D., Foster, T.A., McDonald, L., Schlingman, G., Koehn, F.E., Skotnicki, J.S., Carter, G.T., Moss, S.J., Zhang, M.Q., Martin, C.J., Sheridan, R.M., and Wilkinson, B. (2013). Recombinant strains for the enhanced production of bioengineered rapalogs. *Metab. Eng.* **15**, 167–4173.

Kennedy, J., Auclair, K., Kendrew, S.G., Park, C., Vederas, J.C., and Hutchinson, C.R. (1999). Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science* **284**, 1368–1372.

Kersten, R.D., Lane, A.L., Nett, M., Richter, T.K., Duggan, B.M., Dorrestein, P.C., and Moore, B.S. (2013). Bioactivity-guided genome mining reveals the lomaiviticin biosynthetic gene cluster in *Salinispora tropica*. *ChemBioChem* **14**, 955–962.

Kessler, N., Schuhmann, H., Morneweg, S., Linne, U., and Marahiel, M.A. (2004). The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. *J. Biol. Chem.* **279**, 7413–7419.

Kevany, B.M., Rasko, D.A., and Thomas, M.G. (2009). Characterization of the complete zwittermicin A biosynthesis gene cluster from *Bacillus cereus*. *Appl. Microbiol. Biotechnol*. **75**, 1144–1155

Kim, K.R., Kim, T.J., and Suh, J.W. (2008). The gene cluster for spectinomycin biosynthesis and the aminoglycoside-resistance function of *spcM* in *Streptomyces spectabilis*. *Curr. Microbiol*. **57**, 371–374.

Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120.

Kislak, J.W., Razavi, L.M., Daly, A.K., Finland, M. (1965). Susceptibility of pneumococci to nine antibiotics. *Am. J. Med. Sci*. **250**, 261–268.

Kleigrewe, K., Almaliti, J., Tian, I.Y., Kinnel, R.B., Korobeynikov, A., Monroe, E.A., Duggan, B.M., Di Marzo, V., Sherman, D.H., Dorrestein, P.C., Gerwick, L., and Gerwick, W.H. (2015). Combining mass spectrometric metabolic profiling with genomic analysis: a powerful approach for discovering natural products from Cyanobacteria. *J. Nat. Prod*. **78**, 1671–1682.

Knight, V., and Collins, H.S. (1955). A current view on the problem of drug resistant staphylococci and staphylococcal infection. *Bull. N. Y. Acad. Med*. **31**, 549–568.

Komatsu, M., Komatsu, K., Koiwai, H., Yamada, Y., Kozone, I., Izumikawa, M., Hashimoto, J., Takagi, M., Omura, S., Shin-Ya, K., Cane, D.E., and Ikeda, H. (2013). Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. *ACS Synth. Biol.* **2**, 384–396.

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

Koumoutsi, A., Chen, X.H., Henne, A., Liesegang, H., Hitzeroth, G., Franke, P., Vater, J., and Borriss, R. (2004). Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus amyloliquefaciens* strain FZB42. *J. Bacteriol*. **186**, 1084–1096.

Krügel, H., Krubasik, P., Weber, K., Saluz, H.P., and Sandmann, G. (1999). Functional analysis of genes from *Streptomyces griseus* involved in the synthesis of isorenieratene, a carotenoid with aromatic end groups, revealed a novel type of carotenoid desaturase. *Biochim. Biophys. Acta*. **1439**, 57–64.

Kulowski, K., Pienkowski, E.W., Han, L., Yang, K.Q., Vining, L.C., and Hutchinson, C.R. (1999). Functional Characterization of the *jadI* Gene As a Cyclase Forming Angucyclinones. *J. Am. Chem. Soc*. **121**, 1786–1794.

Kumagai, T., Nakano, T., Maruyama, M., Mochizuki, H., and Sugiyama, M. (1999). Characterization of the bleomycin resistance determinant encoded on the transposon Tn5. *FEBS Lett*. **8**, 34–38.

Kumar, D., Jacob, M.R., Reynolds, M.B., and Kerwin, S.M. (2002). Synthesis and evaluation of anticancer benzoxazoles and benzimidazoles related to UK-1. *Bioorg. Med. Chem.* **10**, 3997–4004.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. *Mol. Biol. Evol*. **33**, 1870–1874.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12.

Kwan, D.H., and Leadlay, P.F. (2010). Mutagenesis of a

modular polyketide synthase enoylreductase domain reveals insights into catalysis and stereospecificity. *ACS Chem. Biol*. **5**, 829–838.

Labeda, D.P., Dunlap, C.A., Rong, X., Huang, Y., Doroghazi, J.R., Ju, K.S., and Metcalf, W. W. (2017). Phylogenetic relationships in the family Streptomycetaceae using multi-locus sequence analysis. *Antonie van Leeuwenhoek* **110**, 563–583.

Lambalot, R.H., Cane, D.E., Aparicio, J.J., and Katz, L. (1995). Overproduction and characterization of the erythromycin C-12 hydroxylase, EryK. *Biochemistry* **34**, 1858–1866.

Larsen, T.M., Boehlein, S.K., Schuster, S.M., Richards, N.G., Thoden, J.B., Holden, H.M., and Rayment, I. (1999). Three-dimensional structure of Escherichia coli asparagine synthetase B: a short journey from substrate to product. *Biochemistry* **38**, 16146–16157.

Laskaris, P., Tolba, S., Calvo-Bado, L., and Wellington, E.M. (2010). Coevolution of antibiotic production and counter‐resistance in soil bacteria. *Environ. Microbiol*. **12**, 783–796.

Lau, J., Fu, H., Cane, D.E., and Khosla, C. (1999). Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units. *Biochemistry* **38**, 1643–1651.

Lau, R.C., and Rinehart, K.L. (1994). Berninamycins B, C, and D, minor metabolites from *Streptomyces bernensis*. *J. Antibiot*. **47**, 1466–1472.

Lautru, S., Deeth, R.J., Bailey, L.M., and Challis, G.L. (2005). Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nat. Chem. Biol.* **1**, 265–269.

Lazos, O., Tosin, M., Slusarczyk, A.L., Boakes, S., Cortés, J., Sidebottom, P.J., and Leadlay, P.F. (2010) Biosynthesis of the putative siderophore erythrochelin requires unprecedented crosstalk between separate nonribomsal peptide gene clusters. *Chem. Biol*. **17**, 160–173.

Lee, M.Y., Ames, B.D., and Tsai, S.C. (2012). Insight into the molecular basis of aromatic polyketidecyclization: crystal structure and in vitro characterization of WhiE-ORFVI. *Biochemistry* **51**, 3079–3091.

Lee, S.W., Mitchell, D.A., Markley, A.L., Hensler, M.E., Gonzalez, D., Wohlrab, A., Dorrestein, P.C., Nizet, V., and Dixon, J.E. (2008). Discovery of a widely distributed

toxin biosynthetic gene cluster. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5879–5884.

Legendre, F., and Armau, E. (1989). Nouveaux composés herbicides, le CL22T et ses derivés, procédés de préparation de ces composés, composition les contenant et procédés de traitement du desherbage les utilisant. French Patent 89,08615, June 28, 1989.

Legendre, F., Maturano, M.D., Etienne, G., Klaebe, A., and Tiraby, G. (1994). Synthesis and biological activity of the herbicidal metabolite CL22T (phthoxazolin) *J. Antibiot.* **48**, 341–343.

Lenfant, N., Hotelier, T., Bourne, Y., Marchot, P., and Chatonnet, A. (2013). Proteins with an alpha/beta hydrolase fold: Relationships between subfamilies in an ever-growing superfamily. *Chem. Biol. Interact*. **203**, 266–268.

Levy, S.B., and McMurry, L. (1974). Detection of an inducible membrane protein associated with R-factor-mediated tetracycline resistance. *Biochem. Biophys. Res. Commun.* **27**, 1060–1068.

Li, B., Yu, J.P., Brunzelle, J.S., Moll, G.N., Van Der Donk, W.A., and Nair, S.K. (2006). Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis. *Science* **311**, 1464–1467.

Li, L., Deng, W., Song, J., Ding, W., Zhao, Q.F., Peng, C., Song, W.W., Tang, G.L., and Liu, W. (2007). Characterization of the saframycin A gene cluster from *Streptomyces lavendulae* NRRL 11002 revealing a nonribosomal peptide synthetase system for assembling the unusual tetrapeptidyl skeleton in an iterative manner. *J. Bacteriol.* **190**, 251–263.

Li, S., Guo, J., Reva, A., Huang, F., Xiong, B., Liu, Y., Deng, Z., Leadlay, P.F., and Sun, Y. (2018). Methyltransferases of gentamicin biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1340–1345.

Li, Y., Xu, W., and Tang, Y. (2010). Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains. *J. Biol. Chem*. **285**, 22764–22773.

Li, Y.M., Milne, J.C., Madison, L.L., Kolter, R., and Walsh, C.T. (1996). From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. *Science* **274**, 1188–1193.

Liao, R., and Liu, W. (2011). Thiostrepton maturation involving a deesterification-

amidation way to process the C-terminally methylated peptide backbone. *J. Am. Chem. Soc.* **133**, 2852–2855.

Lin, L.S., Lanza, T.J. Jr., Castonguay, L.A., Kamenecka, T., McCauley, E., Van Riper, G., Egger, L.A., Mumford, R.A., Tong, X., MacCoss, M., Schmidt, J.A., and Hagmann, W.K. (2004). Bioisosteric replacement of anilide with benzoxazole: potent and orally bioavailable antagonists of VLA-4. *Bioorg. Med. Chem. Lett*. **14**, 2331–2334.

Lin, X., Hopson, R., and Cane, D.E. (2006). Genome mining in *Streptomyces coelicolor*: molecular cloning and characterization of a new sesquiterpene synthase. *J. Am. Chem. Soc*. **128**, 6022–6023.

Lin, Y.S., Kieser, H.M., Hopwood, D.A., and Chen, C.W. (1993). The chromosomal DNA of Streptomyces lividans 66 is linear. *Mol. Microbiol*. **10**, 923–33.

Liu, F., Garneau, S., and Walsh, C.T. (2004). Hybrid nonribosomal peptide-polyketide interfaces in epothilone biosynthesis: minimal requirements at N and C termini of EpoB for elongation. *Chem. Biol.* **11**, 1533–1542.

Liu, H.K., and Sadler, P.J. (2011). Metal complexes as DNA intercalators. *Acc. Chem. Res*. **44**, 349–359.

Liu, S., Guo, H., Zhang, T., Han, L., Yao, P., Zhang, Y., Rong, N., Yu, Y., Lan, W., Wang, C., Ding, J., Wang, R., Liu, W., and Cao, C. (2015). Structure-based mechanistic insights into terminal amide synthase in nosiheptide-represented thiopeptides biosynthesis. *Sci. Rep*. **5**, 12744.

Lohman, J.R., Ma, M., Osipiuk, J., Nocek, B., Kim Y, Chang, C., Cuff, M., Mack, J., Bigelow, L., Li, H., Endres, M., Babnigg, G., Joachimiak, A., Phillips, G.N. Jr., and Shen, B. (2015). Structural and evolutionary relationships of "AT-less" type I polyketide synthase ketosynthases. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12693–12698.

Lomovskaya, N., Hong, S.K., Kim, S.U., Fonstein, L., Furuya, K., and Hutchinson, R.C. (1996). The Streptomyces peucetius drrC gene encodes a UvrA-like protein involved in daunorubicin resistance and production. *J. Bacteriol.* **178**, 3238–3245.

Lopanik, N.B., Shields, J.A., Buchholz, T.J., Rath, C.M., Hothersall, J., Haygood, M.G., Håkansson, K., Thomas, C.M., and Sherman, D.H. (2008). *In vivo* and *in vitro trans*-acylation by BryP, the putative bryostatin pathway acyltransferase

derived from an uncultured marine symbiont. *Chem. Biol.* **15**, 1175–1186.

Losada, A.A., Cano-Prieto, C., García-Salcedo, R., Braña, A.F., Méndez, C., Salas, J.A., and Olano, C. (2017). Caboxamycin biosynthesis pathway and identification of novel benzoxazoles produced by cross-talk in Streptomyces sp. NTK 937. *Microb. Biotechnol.* **10**, 873–885.

Lowry, B., Li, X., Robbins, T., Cane, D.E., and Khosla, C. (2016). A turnstile mechanism for the controlled growth of biosynthetic intermediates on assembly line polyketide synthases. *ACS Cent. Sci*. **2**, 14–20.

Lowry, B., Robbins, T., Weng, C.H., O'Brien, R.V., Cane, D.E., and Khosla, C. (2013). *In vitro* reconstitution and analysis of the 6-deoxyerythronolide B synthase. *J. Am. Chem. Soc.* **135**, 16809–16812.

Luo, L., Kohli, R.M., Onishi, M., Linne, U., Marahiel, M.A. and Walsh, C.T. (2002). Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry* **41**, 9184–9196.

Luo, Y., Ruan, L.F., Zhao, C.M., Wang, C.X., Peng, D.H., and Sun, M. (2011). Validation of the intact zwittermicin A biosynthetic gene cluster and discovery of a complementary resistance mechanism in *Bacillus thuringiensis*. *Antimicrob. Agents Chemother.* **55**, 4161–4169.

Luscombe, N.M., Austin, S.E., Berman, H.M., and Thornton, J.M. (2000). An overview of the structures of protein-DNA complexes. *Genome Biol*. **1**, REVIEWS001.

Lv, M., Zhao, J., Deng, Z., and Yu, Y. (2015). Characterization of the biosynthetic gene cluster for benzoxazole antibiotics A33853 reveals unusual assembly logic. *Chem. Biol.* **22**, 1313–1324.

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

Madigan, M.T., and Martinko, J.M. (2006). Brock Biology of Microorganisms, eleventh edition. New Jersey: Prentice Hall.

Maier, T., Leibundgut, M., and Ban, N. (2008). The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315–1322.

Malcolmson, S.J., Young, T.S., Ruby, J.G., Skewes-Cox, P., and Walsh, C.T. (2013). The posttranslational modification cascade to the thiopeptide berninamycin generates linear forms and altered macrocyclic scaffolds. *Proc. Natl. Acad. Sci. U.S.A*. **110**, 8483–8488.

Malpartida, F., and Hopwood, D.A. (1986). Physical and genetic characterisation of the gene cluster for the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2). *Mol. Gen. Genet*. **205**, 66–73.

Mao, Y., Varoglu, M., and Sherman, D.H. (1999). Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from *Streptomyces lawendulae* NRRL 2564. *Chem. Biol*. **6**, 251–263.

Markowitz, V.M., Chen, I.M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., Huntemann, M., Anderson, I., Mavromatis, K., Ivanova, N.N., and Kyrpides, N.C. (2012). IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucl. Acids Res.* **40**, D115–D122.

Marsden, A.F., Caffrey, P., Aparicio, J.F., Loughran, M.S., Staunton, J., and Leadlay, P.F. (1994). Stereospecific acyl transfers on the erythromycin-producing polyketide synthase. *Science* **263**, 378–380.

Masschelein, J., Clauwers, C., Awodi, U.R., Stalmans, K., Vermaelen, W., Lescrinier, E., Aertsen, A., Michiels, C., Challis, G.L., and Lavigne, R. (2015). A combination of polyunsaturated fatty acid, nonribosomal peptide and polyketide biosynthetic machinery is used to assemble the zeamine antibiotics. *Chem. Sci.* **6**, 923–929.

Masschelein, J., Mattheus, W., Gao, L.J., Moons, P., Van Houdt, R., Uytterhoeven, B., Lamberigts, C., Lescrinier, E., Rozenski, J., Herdewijn, P., Aertsen, A., Michiels, C., and Lavigne, R. (2013). A PKS/NRPS/FAS hybrid gene cluster from *Serratia plymuthica* RVH1 encoding the biosynthesis of three broad spectrum, zeamine-related antibiotics. *PLoS One* **8**, E54143.

McCafferty, D.G., Cudic, P., Frankel, B.A., Barkallah, S., Kruger, R.G., and Li, W. (2002). Chemistry and biology of the ramoplanin family of peptide antibiotics. *Biopolymers* **66**, 261–284.

McDonald, M., Mavrodi, D.V., Thomashow, L.S., and Floss, H.G. (2001). Phenazine biosynthesis in *Pseudomonas fluorescens*: branchpoint from the

primary shikimate biosynthetic pathway and role of phenazine-1,6-dicarboxylic acid. *J. Am. Chem. Soc.* **123**, 9459–9460.

McGuire, J.M., Bunch, P.L., Anderson, R.C., Boaz, H.E., Flynn, E.H., Powell, E.H., and Smith, J.W. (1952). Ilotycin, a new antibiotic. *Antibiot. Chemother.* **2**, 281–283.

McKee, M.L., and Kerwin, S.M. (2008). Synthesis, metal ion binding, and biological evaluation of new anticancer 2-(2'-hydroxyphenyl)benzoxazole analogs of UK-1. *Bioorg. Med. Chem*. **16**, 1775–1783.

Medema, M.H., Cimermancic, P., Sali, A., Takano, E., and Fischbach, M.A. (2014). A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis. *PLoS Comput. Biol*. **10**, e1004016.

Medema, M.H., Takano, E., and Breitling, R. (2013). Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol. Biol. Evol.* **30**, 1218–1223.

Metz, J.G., Roessler, P., Facciotti, D., Levering, C., Dittrich, F., Lassner, M., Valentine, R., Lardizabal, K., Domergue, F., Yamada, A., Yazawa, K., Knauf, V., and Browse, J. (2001). Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. *Science* **293**, 290–293.

Meurer, G., Gerlitz, M., Wendt-Pienkowski, E., Vining, L.C., Rohr, J., and Hutchinson, C.R. (1997). Iterative type II polyketide synthases, cyclases and ketoreductases exhibit context-dependent behavior in the biosynthesis of linear and angular decapolyketides. *Chem. Biol.* **4**, 433–443.

Meyer, S., Kehr, J.C., Mainz, A., Dehm, D., Petras, D., Süssmuth, R.D., and Dittmann, E. (2016). Biochemical dissection of the natural diversification of microcystin provides lessons for synthetic biology of NRPS. *Cell Chem Biol*. **23**, 462–471.

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

Miller, D.A., Luo, L., Hillson, N., Keating, T.A., and Walsh, C.T. (2002). Yersiniabactin synthetase: a four-protein assembly line producing the

nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*. *Chem. Biol.* **9**, 333–344.

Miller, D.J., Ouellette, N., Evdokimova, E., Savchenko, A., Edwards, A., and Anderson, W.F. (2003). Crystal complexes of a predicted S-adenosylmethionine-dependent methyltransferase reveal a typical AdoMet binding domain and a substrate recognition domain. *Protein Sci.* **12**, 1432–1442.

Molnár, I., Schupp, T., Ono, M., Zirkle, R., Milnamow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D.D., Gorlach, J., Mayo, J.M., Hu, A., Goff, S., Schmid, J., and Ligon, J.M. (2000). The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. *Chem. Biol.* **7**, 97–109.

Molohon, K.J., Blair, P.M., Park, S., Doroghazi, J.R., Maxson, T., Hershfield, J.R., Flatt, K.M., Schroeder, N.E., Ha, T., and Mitchell, D.A. (2016). Plantazolicin is an ultra-narrow spectrum antibiotic that targets the Bacillus anthracis membrane. *ACS Infect. Dis*. **2**, 207–220.

Mootz, H.D., Kessler, N., Linne, U., Eppelmann, K., Schwarzer, D., and Marahiel, M.A. (2002). Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes. *J. Am. Chem. Soc.* **124**, 10980–10981.

Mori, T., Takahashi, K., Kashiwabara, M., Uemura, D., Katayama, C., Iwadare, S., Shizuri, Y., Mitomo, R., Nakano, F., and Matsuzaki, A. (1985). Structure of oxazolomycin, a novel β-lactone antibiotic. *Tetrahedron Lett*. **26**, 1073-1076.

Müller I, Weinig S, Steinmetz H, Kunze B, Veluthoor S, Mahmud T, Müller R. (2006). A unique mechanism for methyl ester formation via an amide intermediate found in myxobacteria. *ChemBioChem* **7**, 1197–1205.

Murphy, A.C., Hong, H., Vance, S., Broadhurst, R.W., and Leadlay, P.F. (2016). Broadening substrate specificity of a chain-extending ketosynthase through a single active-site mutation. *J. Chem. Soc., Chem. Commun*. **52**, 8373–8376.

Ng, B.G., Han, J.W., Lee, D.W., Choi, G.J., and Kim, B.S. (2018) The chejuenolide biosynthetic gene cluster harboring an iterative *trans*-AT PKS system in *Hahella chejuensis* strain MB-1084. *J. Antibiot*. **71**, 495–505.

Nguyen, D.D., Wu, C.H., Moree, W.J., Lamsa, A., Medema, M.H., Zhao, X.,

Gavilan, R.G., Aparicio, M., Atencio, L., Jackson, C., Ballesteros, J., Sanchez, J., Watrous, J.D., Phelan, V.V., van de Wiel, C., Kersten, R.D., Mehnaz, S., De Mot, R., Shank, E.A., Charusanti, P., Nagarajan, H., Duggan, B.M., Moore, B.S., Bandeira, N., Palsson, B.Ø., Pogliano, K., Gutiérrez, M., and Dorrestein, P.C. (2013). MS/MS networking guided analysis of molecule and gene cluster families. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E2611–E2620.

Nguyen, T., Ishida, K., Jenke-Kodama, H., Dittmann, E., Gurgui, C., Hochmuth, T., Taudien, S., Platzer, M., Hertweck, C., and Piel, J. (2008). Exploiting the mosaic structure of *trans*-acyltransferase polyketidesynthases for natural product discovery and pathway dissection. *Nat. Biotechnol*. **26**, 225–233.

Olano, C., Wilkinson, B., Sánchez, C., Moss, S.J., Sheridan, R., Math, V., Weston, A.J, Braña, A.F., Martin, C.J., Oliynyk, M., Méndez, C., Leadlay, P.F., and Salas, J.A. (2004). Biosynthesis of the angiogenesis inhibitor borrelidin by *Streptomyces parvulus* Tü4055: cluster analysis and assignment of functions. *Chem. Biol*. **11**, 87–97.

Olarte, J., and de la Torre, J.A. (1959). Resistance of Shigella flexneri to tetracyclines, chloramphenicol and streptomycin; a study of 131 freshly isolated strains. *Am. J. Trop. Med. Hyg*. **8**, 324–326.

Ômura, S., Tanaka, Y., Kanaya, I., Shinose, M., and Takahashi, Y. (1990). Phthoxazolin, a specific inhibitor of cellulose biosynthesis, produced by a strain of *Streptomyces* sp. *J. Antibiot.* **43**, 1034-1036.

Owen, J.G., Calcott, M.J., Robins, K.J., Ackerley, D.F. (2016). Generating functional recombinant NRPS enzymes in the laboratory setting via peptidyl carrier protein engineering. *Cell Chem. Biol*. **23**, 1395–1406.

Owen, J.G., Reddy, B.V., Ternei, M.A., Charlop-Powers, Z., Calle, P.Y., Kim, J.H., and Brady, S.F. (2013). Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11797–11802.

Parascandolo, J.S., Havemann, J., Potter, H.K., Huang, F., Riva, E., Connolly, J., Wilkening, I., Song, L., Leadlay, P.F., and Tosin, M. (2016). Insights into 6-methylsalicylic acid bio-assembly by using chemical probes. *Angew. Chem. INt. Ed. Engl.* **55**, 3463-3467.

Parks, W.M., Bottrill, A.R., Pierrat, O.A., Durrant, M.C., and Maxwell, A. (2007).

The action of the bacterial toxin, microcin B17, on DNA gyrase. *Biochimie* **89**, 500–507.

Pasteur, L., and Joubert, J.F. (1877). Charbon et septicémie. C. R. Hebd. *Séances Acad. Sci.* **85**, 101–115.

Patel, H.M., Tao, J., and Walsh, C.T. (2003). Epimerization of an L-cysteinyl to a D-cysteinyl residue during thiazoline ring formation in siderophore chain elongation by pyochelin synthetase from *Pseudomonas aeruginosa*. *Biochemistry* **42**, 10514–10527.

Paulus, T.J., Tuan, J.S., Luebke, V.E., Maine, G.T., DeWitt, J.P., and Katz, L. (1990). Mutation and cloning of eryG, the structural gene for erythromycin O-methyltransferase from Saccharopolyspora erythraea, and expression of eryG in *Escherichia coli*. *J. Bacteriol*. **172**, 2541–2546.

Penwell, W.F., Arivett, B.A., and Actis, L.A. (2012). The *Acinetobacter baumannii entA* gene located outside the acinetobactin cluster is critical for siderophore production, iron acquisition and virulence. *PLoS One* **7**, e36493.

Petersen, F., Zähner, H., Metzger, J.W., Freund, S., and Hummel, R.P. (1993). Germicidin, an autoregulative germination inhibitor of *Streptomyces viridochromogenes* NRRL B-1551. *J. Antibiot*. **46**, 1126–1138.

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

Pickens, L.B., and Tang, Y. (2010). Oxytetracycline biosynthesis. *J. Biol. Chem.* **285**, 27509–27515.

Poralla, K., Muth, G., and Härtner, T. (2000). Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). *FEMS Microbiol. Lett.* **189**, 93–95.

Porter, T.N., Li, Y., and Raushel, F.M. (2004). Mechanism of the dihydroorotase reaction. *Biochemistry* **43**, 16285–16292.

Poust, S., Phelan, R.M., Deng, K., Katz, L., Petzold, C.J., and Keasling, J.D. (2015). Divergent mechanistic routes for the formation of *gem*-dimethyl groups in the biosynthesis of complex polyketides. *Angew. Chem. Int. Ed. Engl*. **54**, 2370–2373.

Pozharskii, A.F., Soldatenkov, A.T., and Katritzky, A.R. (2011). Heterocycles in Life and Society: An Introduction to Heterocyclic Chemistry, Biochemistry and Applications, second edition. Hoboken: John Wiley & Sons, Ltd.

Pulsawat, N., Kitani, S., and 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.

Rangaswamy, V., Mitchell, R., Ullrich, M., and Bender, C. (1998). Analysis of genes involved in biosynthesis of coronafacic acid, the polyketide component of the phytotoxin coronatine. *J. Bacteriol*. **180**, 3330–3338.

Read, J.A., and Walsh, C.T. (2007). The lyngbyatoxin biosynthetic assembly line: chain release by four-electron reduction of a dipeptidyl thioester to the corresponding alcohol. *J. Am. Chem. Soc*. **129**, 15762–15763.

Redshaw, P.A., McCann, P.A., Pentella, M.A., and Pogell, B.M. (1979). Simultaneous loss of multiple differentiated functions in aerial mycelium-negative isolates of streptomycetes. *J. Bacteriol*. **137**, 891–899.

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

Reid, R., Piagentini, M., Rodriguez, E., Ashley, G., Viswanathan, N., Carney, J., Santi, D.V., Hutchinson, C.R., and McDaniel, R. (2003). A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry* **42**, 72–79.

Richards, A.N. (1964). Production of penicillin in the United States (1941-1946). *Nature* **201**, 441–445.

Riva, E., Wilkening, I., Gazzola, S., Li, W.M., Smith, L., Leadlay, P.F., and Tosin, M. (2014). Chemical probes for the functionalization of polyketide intermediates. *Angew. Chem. Int. Ed. Engl.* **53**, 11944–11949.

Robbins, T., Kapilivsky, J., Cane, D.E., and Khosla, C. (2016). Roles of conserved active site residues in the ketosynthase domain of an assembly Line polyketide synthase. *Biochemistry* **55**, 4476–4484.

Roberts, G.A., Staunton, J., and Leadlay, P.F. (1993). Heterologous expression in

Escherichia coli of an intact multienzyme component of the erythromycin-producing polyketide synthase*. Eur. J. Biochem*. **214**, 305–311.

Rohmer, M., Seemann, M., Horbach, S., Bringer-Meyer, S., and Sahm, H. (1996). Glyceraldehyde 3-phosphate and pyruvate as precursors of isoprenic units in an alternative non-mevalonate pathway for terpenoid biosynthesis. *J. Am. Chem. Soc*. **118**, 2564–2566.

Rong, X., and Huang, Y (2012). Taxonomic evaluation of the *Streptomyces hygroscopicus* clade using multi-locus sequence analysis and DNA-DNA hybridization, validating the MLSA scheme for the systematics of the whole genus. *Syst. Appl. Microbiol*. **35**, 7–18.

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

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945.

Saha, S., Zhang, W., Zhang, G., Zhu, Y., Chen, Y., Liu, W., Yuan, C., Zhang, Q., Zhang, H., Zhang, L., Zhang, W., and Zhang, C. (2017). Activation and characterization of a cryptic gene cluster reveals a cyclization cascade for polycyclic tetramate macrolactams. *Chem. Sci.* **8**, 1607–1612.

Salah-Bey, K., Doumith, M., Michel, J.M., Haydock, S., Cortés, J., Leadlay, P.F., and Raynal, M.C. (1998). Targeted gene inactivation for the elucidation of deoxysugar biosynthesis in the erythromycin producer *Saccharopolyspora erythraea*. *Mol. Gen. Genet*. **25**7, 542–553.

Sato, S., Kajiura, T., Noguchi, M., Takehana, K., Kobayashi, T., and Tsuji, T. (2001). AJI9561, a new cytotoxic benzoxazole derivative produced by *Streptomyces* sp. *J. Antibiot.* **54**, 102–104.

Schatz, A., Bugle, E., and Waksman, S.A. (1944). Streptomycin, a substance exhibiting antibiotic activity against Gram-positive and Gram-negative bacteria. *Exp. Biol. Med*. **55**, 66–69.

Scherlach, K., Partida-Martinez, L.P., Dahse, H.M., and Hertweck, C. (2006). Antimitotic rhizoxin derivatives from a cultured bacterial endosymbiont of the rice pathogenic fungus *Rhizopus microsporus*. *J. Am. Chem. Soc*. **128**, 11529–11536.

Schlumbohm, W., Stein, T., Ullrich, C., Vater, J., Krause, M., Marahiel, M.A., Kruft, V., and Wittmann-Liebold, B. (1991). An active serine is involved in covalent substrate amino acid binding at each reaction center of gramicidin S synthetase. *J. Biol. Chem.* **266**, 23135–23141.

Schmitt, E., Blanquet, S., and Mechulam, Y. (1996). Structure of crystalline *Escherichia coli* methionyl-tRNA(f)Met formyltransferase: comparison with glycinamide ribonucleotide formyltransferase. *EMBO J*. **15**, 4749–4758.

Schneider, R., and Hantke, K. (1993). Iron-hydroxamate uptake systems in Bacillus subtilis: identification of a lipoprotein as part of a binding protein-dependent transport system. *Mol. Microbiol*. **8**, 111–121.

Schnell, N., Entian, K.D., Schneider, U., Götz, F., Zähner, H., Kellner, R., and Jung, G. (1988). Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings. *Nature* **333**, 276–278.

Schönafinger, G., Schracke, N., Linne, U., and Marahiel, M.A. (2006). Formylation domain: an essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin. *J. Am. Chem. Soc*. **128**, 7406–7407.

Scotti, C., Piatti, M., Cuzzoni, A., Perani, P., Tognoni, A., Grandi, G., Galizzi, A., and Albertini, A.M. (1993). A *Bacillus subtilis* large ORF coding for a polypeptide highly similar to polyketide synthases. *Gene* **130**, 65–71.

Seibert, C.M., and Raushel, F.M. (2005). Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry* **44**, 6383–6391.

Sekiguchi, J., and Shuman, S. (1997). Nick sensing by DNA ligase requires a 5′ phosphate at the nick and occupancy of the adenylate binding site on the enzyme. *J. Virol.* **71**, 9679–9684.

Shen, B., and Hutchinson, C.R. (1996). Deciphering the mechanism for the assembly of aromatic polyketides by a bacterial polyketide synthase. *Proc. Natl. Acad. Sci. U.S.A*. **93**, 6600–6604.

Shiomi, K., Arai, N., Shinose, M., Takahashi, Y., Yoshida, H., Iwabuchi, J., Tanaka, Y., and Ômura, S. (1995). New antibiotics phthoxazolins B, C and D produced by

Streptomyces sp. KO-7888. *J. Antibiot*. **48**, 714–719.

Shulse, C.N., and Allen, E.E. (2011). Diversity and distribution of microbial long-chain fatty acid biosynthetic genes in the marine environment. *Environ. Microbiol*. **13**, 684–695.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., and Higgins, D.G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol*. **7**, 539–544.

Silakowski, B., Kunze, B., and Müller, R. (2000). Stigmatella aurantiaca Sg a15 carries genes encoding type I and type II 3-deoxy-D-arabino-heptulosonate-7-phosphate synthases: involvement of a type II synthase in aurachin biosynthesis. *Arch. Microbiol*. **173**, 403–411.

Silakowski, B., Schairer, H.U., Ehret, H., Kunze, B., Weinig, S., Nordsiek, G., Brandt, P., Blöcker, H., Höfle, G., Beyer, S., and Müller, R. (1999). New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of Stigmatella aurantiaca DW4/3-1. *J. Biol. Chem*. **274**, 37391–37399.

Skiba, M.A., Sikkema, A.P., Fiers, W.D., Gerwick, W.H., Sherman, D.H., Aldrich, C.C., and Smith, J.L. (2016). Domain organization and active site architecture of a polyketide synthase *C*-methyltransferase. *ACS Chem. Biol*. **11**, 3319–3327.

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

Stassi, D., Donadio, S., Staver, M.J., and Katz, L. (1993). Identification of a Saccharopolyspora erythraea gene required for the final hydroxylation step in erythromycin biosynthesis. *J. Bacteriol*. **175**, 182–189.

Staunton, J., and Weissman, K.J. (2001). Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380–416.

Stindl, A., and Keller, U. (1994). Epimerization of the D-valine portion in the biosynthesis of actinomycin D. *Biochemistry* **33**, 9358–9364.

Storm, D.R., and Strominger, J.L. (1973). Complex formation between bacitracin peptides and isoprenyl pyrophosphates. The specificity of lipid-peptide interactions. *J. Biol. Chem*. **248**, 3940–3945.

Sugiyama, M., Mochizuki, H., Nimi, O., and Nomi, R. (1981). Mechanism of protection of protein synthesis against streptomycin inhibition in a producing strain. *J. Antibiot.* **34**, 1183–1188.

Sugiyama, M., Paik, S.Y., Nomi, R. (1985). Mechanism of self-protection in a puromycin-producing micro-organism. *J. Gen. Microbiol*. **131**, 1999–2005.

Sugiyama, M., Takeda, A., Paik, S.Y., Nimi, O., and Nomi, R. (1986). Acetylation of blasticidin S by its producing actinomycetes. *J. Antibiot.* **39**, 827–832.

Summers, R.G., Donadio, S., Staever, M.J., Wendt-Pienkowski, E., Hutchinson, C.R., and Katz, L. (1997). Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of *Saccharopolyspora erythraea* that are involved in L-mycarose and D-desosamine production. *Microbiology* **143**, 3251–3262.

Sun, Y., He, X., Liang, J., Zhou, X., and Deng, Z. (2009). Analysis of functions in plasmid pHZ1358 influencing its genetic and structural stability *in Streptomyces lividans* 1326. *Appl. Microbiol. Biotechnol*. **82**, 303–310

Suroto, D.A., Kitani, S., Miyamoto, K.T., Sakihama, Y., Arai, M., Ikeda, H., and Nihira, T. (2017). Activation of cryptic phthoxazolin A production in Streptomyces avermitilis by the disruption of autoregulator-receptor homologue AvaR3. *J. Biosci. Bioeng.* **124**, 611–617.

Taguchi, T., Ebihara. T., Furukawa, A., Hidaka, Y., Ariga, R., Okamoto, S., and Ichinose, K. (2012). Identification of the actinorhodin monomer and its related compound from a deletion mutant of the *actVA-ORF4* gene of *Streptomyces coelicolor* A3(2). *Bioorg. Med. Chem. Lett.* **22**, 5041–5045.

Tahlan, K., Ahn, S.K., Sing, A., Bodnaruk, T.D., Willems, A.R., Davidson, A.R., and Nodwell, J.R. (2007). Initiation of actinorhodin export in Streptomyces coelicolor. *Mol. Microbiol.* **63**, 951–961.

Takano, H., Obitsu, S., Beppu, T., and Ueda, K. (2005). Light-induced carotenogenesis in Streptomyces coelicolor A3(2): identification of an extracytoplasmic function sigma factor that directs photodependent transcription of the carotenoid biosynthesis gene cluster. *J. Bacteriol*. **187**, 1825–1832.

Tanaka, Y., Kanaya, I., Takahashi, Y., Shinose, M., Tanaka, H., and Ômura, S. (1993). Phthoxazolin A, a specific inhibitor of cellulose biosynthesis from microbial

origin. I. Discovery, taxonomy of producing microorganism, fermentation, and biological activity. *J. Antibiot.* **46**, 1208–1213.

Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C., and Julien, B. (2000). Cloning and heterologous expression of the epothilone gene cluster. *Science* **287**, 640–642.

Tang, Y., Lee, H.Y., Tang, Y., Kim, C.Y., Mathews, I., and Khosla, C. (2006). Structural and functional studies on SCO1815: a beta-ketoacyl-acyl carrier protein reductase *from Streptomyces coelicolor* A3(2). *Biochemistry* **45**, 14085–14093.

Tang, Y., Tsai, S.C., and Khosla, C. (2003). Polyketide chain length control by chain length factor. *J. Am. Chem. Soc*. **125**, 12708–12709.

Tao, W., Yurkovich, M.E., Wen, S., Lebe, K.E., Samborskyy, M., Liu, Y., Yang, A., Liu, Y., Ju, Y., Deng, Z., Tosin, M., Sun, Y., and Leadlay, P.F. (2016). A genomics-led approach to deciphering the mechanism of thiotetronate antibiotic biosynthesis. *Chem. Sci.* **7**, 376–385.

Tierno, M.B., Kitchens, C.A., Petrik, B., Graham, T.H., Wipf, P., Xu, F.L., Saunders, W.S., Raccor, B.S., Balachandran, R., Day, B.W., Stout, J.R., Walczak, C.E., Ducruet, A.P., Reese, C.E., and Lazo, J.S. (2009). Microtubule binding and disruption and induction of premature senescence by disorazole C. *J. Pharmacol. Exp. Ther*. **32**8, 715–722.

Tocchetti, A., Maffioli, S., Iorio, M., Alt, S., Mazzei, E., Brunati, C., Sosio, M., and Donadio, S. (2013). Capturing linear intermediates and C-terminal variants during maturation of the thiopeptide GE2270. *Chem. Biol*. **20**, 1067–1077.

Traxler, M. F. and Kolter, R. (2015) Natural products in soil microbe interactions and evolution. *Nat. Prod. Rep.* **32**, 956–970.

Tsai, S.C., Lu, H., Cane, D.E., Khosla, C., and Stroud, R.M. (2002). Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry* **41**, 12598–12606.

Umezawa, H., Ueda, M., Maeda, K., Yagishita, K., Kondo, S., Okami, Y., Utahara, R., Osato, Y., Nitta, K., and Takeuchi, T. (1957). Production and isolation of a new antibiotic: kanamycin. *J. Antibiot*. **10**, 181–188.

Vance, S., Tkachenko, O., Thomas, B., Bassuni, M., Hong, H., Nietlispach, D., and

Broadhurst, W. *Biochem. J.* **473**, 1097–1110.

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

Vieweg, L., Reichau, S., Schobert, R., Leadlay, P.F., and Süssmuth, R.D. (2014) Recent advances in the field of bioactive tetronates. *Nat. Prod. Rep.* **31**, 1554–1584.

Wagner, D.T., Zeng, J., Bailey, C.B., Gay, D.C., Yuan, F., Manion, H.R., and Keatinge-Clay, A.T. (2017). Structural and functional trends in dehydrating bimodules from *trans*-acyltransferase polyketide synthases. *Structure* **25**, 1045–1055.

Wakimoto, T., Egami, Y., Nakashima, Y., Wakimoto, Y., Mori, T., Awakawa, T., Ito, T., Kenmoku, H., Asakawa, Y., Piel, J., and Abe, I. (2014). Calyculin biogenesis from a pyrophosphate protoxin produced by a sponge symbiont. *Nat. Chem. Biol*. **10**, 648–655.

Waksman, S.A., and Henrici, A.T. (1943). The Nomenclature and Classification of the Actinomycetes. *J. Bacteriol.* **46**, 337–341.

Waksman, S.A., and Woodruff, H.B. (1940). The soil as a source of microorganisms antagonistic to disease-producing bacteria. *J. Bacteriol*. **40**, 581–600.

Waksman, S.A., and Woodruff, H.B. (1942). Selective antibiotic action of various substances of microbial origin. *J. Bacteriol*. **44**, 373–384.

Walsh, C.T. (2016). Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat. Prod. Rep*. **33**, 127–135.

Wang, B., Guo, F., Ren, J., Ai, G., Aigle, B., Fan, K., and Yang, K. (2015a). Identification of Alp1U and Lom6 as epoxy hydrolases and implications for kinamycin and lomaiviticin biosynthesis. *Nat. Commun*. **6**, 7674.

Wang, B., Ren, J., Li, L., Guo, F., Pan, G., Ai, G., Aigle, B., Fan, K., and Yang, K. (2015b). Kinamycin biosynthesis employs a conserved pair of oxidases for B-ring contraction. *J. Chem. Soc., Chem. Commun*. **51**, 8845–8848.

Wang, C.M., and Cane, D.E. (2008) Biochemistry and molecular genetics of the biosynthesis of the earthy odorant methylisoborneol in *Streptomyces coelicolor*. *J. Am. Chem. Soc.* **130**, 8908–8909.

Weber, G., Schörgendorfer, K., Schneider-Scherzer, E., and Leitner, E. (1994). The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. *Curr. Genet.* **26**, 120–125.

Weinig, S., Hecht, H.J., Mahmud, T., and Müller, R. (2003). Melithiazol biosynthesis: further insights into myxobacterial PKS/NRPS systems and evidence for a new subclass of methyl transferases. *Chem. Biol*. **10**, 939–952.

Weissman, K.J., Timoney, M., Bycroft, M., Grice, P., Hanefeld, U., Staunton, J., and Leadlay, P.F. (1997). The molecular basis of Celmer's rules: the stereochemistry of the condensation step in chain extension on the erythromycin polyketide synthase. *Biochemistry* **36**, 13849–13855.

Weitnauer, G., Mühlenweg, A., Trefzer, A., Hoffmeister, D., Süssmuth, R.D., Jung, G., Welzel, K., Vente, A., Girreser, U., and Bechthold, A. (2001). Biosynthesis of the orthosomycin antibiotic avilamycin A: deductions from the molecular analysis of the avi biosynthetic gene cluster of *Streptomyces viridochromogenes* Tü57 and production of new antibiotics. *Chem. Biol.* **8**, 569–581.

Westley, J.W., Liu, C.M., Evans, R.H., and Blount, J.F. (1979). Conglobatin, a novel macrolide dilactone from *Streptomyces conglobatus* ATCC 31005. *J. Antibiot.* **32**, 874–877.

Whicher, J.R., Dutta, S., Hansen, D.A., Hale, W.A., Chemler, J.A., Dosey, A.M., Narayan, A.R., Håkansson, K., Sherman, D.H., Smith, J.L., and Skiniotis, G. (2014). Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560–564.

Wilkinson C.J., Hughes-Thomas Z.A., Martin C.J., Bohm I., Mironenko T., Deacon M., Wheatcroft, M.P., Wirtz, G., Staunton, J., and Leadlay, P.F. (2002). Increasing the efficiency of heterologous promoters in actinomycetes. *J. Mol. Microbiol. Biotechnol.* **4**, 417–426.

Winkelman, G., and Drechsel, H. (1997). Microbial siderophores. In Biotechnology, Vol 7: Products of Secondary Metabolism. Kleinkauf, H., and von Döhren, H., eds. Weinheim: Wiley-VCH, 200–246.

Wittmann, M., Linne, U., Pohlmann, V., and Marahiel, M.A. (2008). Role of DptE and DptF in the lipidation reaction of daptomycin. *FEBS J.* **275**, 5343–5354.

Woo, C.M., Beizer, N.E., Janso, J.E., and Herzon, S.B. (2012). Isolation of lomaiviticins C-E, transformation of lomaiviticin C to lomaiviticin A, complete structure elucidation of lomaiviticin A, and structure-activity analyses. *J. Am. Chem. Soc.* **134**, 15285–15288.

World Health Organization (WHO). (2015). Worldwide country situation analysis: response to antimicrobial resistance. Geneva: World Health Organization.

Wu, Q., Liang, J., Lin, S., Zhou, X., Bai, L., Deng, Z., and Wang, Z. (2011). Characterization of the biosynthesis gene cluster for the pyrrole polyether antibiotic calcimycin (A23187) in *Streptomyces chartreusis* NRRL 3882. *Antimicrob. Agents Chemother*. **55**, 974–982.

Xie, L., Miller, L.M., Chatterjee, C., Averin, O., Kelleher, N.L., Van Der Donk, W.A. (2004). Lacticin 481: *in vitro* reconstitution of lantibiotic synthetase activity. *Science* **303**, 679–681.

Xie, X., Garg, A., Keatinge-Clay, A.T., Khosla, C., and Cane, D.E. (2016). Epimerase and reductase activities of polyketide synthase ketoreductase domains utilize the same conserved tyrosine and serine residues. *Biochemistry* **55**, 1179–1186.

Yamanaka, K., Reynolds, K.A., Kersten, R.D., Ryan, K.S., Gonzalez, D.J., Nizet, V., Dorrestein, P.C., and Moore, B.S. (2014). Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 1957–1962.

Yang, K., Han, L., Ayer, S.W., and Vining, L.C. (1996). Accumulation of the angucycline antibiotic rabelomycin after disruption of an oxygenase gene in the jadomycin B biosynthetic gene cluster of *Streptomyces venezuelae*. *Microbiology* **142**, 123–132.

Yin, J., Hoffmann, M., Bian, X., Tu, Q., Yan, F., Xia, L., Ding, X., Stewart, A.F., Müller, R., Fu, J., and Zhang, Y. (2015). Direct cloning and heterologous expression of the salinomycin biosynthetic gene cluster from *Streptomyces albus* DSM41398 in *Streptomyces coelicolor* A3(2). *Sci. Rep*. **5**, 15081.

Yin, X., and Zabriskie, T.M. (2006). The enduracidin biosynthetic gene cluster from

Streptomyces fungicidicus. *Microbiology* **152**, 2969–2983.

Yu, T.W., and Chen, C.W. (1993). The unstable melC operon of *Streptomyces antibioticus* is codeleted with a Tn4811-homologous locus. *J. Bacteriol*. **175**, 1847–1852.

Yu, T.W., Shen, Y., Doi-Katayama, Y., Tang, L., Park, C., Moore, B.S., Hutchinson, C.R., and Floss, H.G. (1999). Direct evidence that the rifamycin polyketide synthase assembles polyketide chains processively. *Proc. Natl. Acad. Sci. U.S.A*. **96**, 9051–9056.

Yu, Y., Guo, H., Zhang, Q., Duan, L., Ding, Y., Liao, R., Lei, C., Shen, B., and Liu, W. (2010). NosA catalyzing carboxyl-terminal amide formation in nosiheptide maturation via an enamine dealkylation on the serine-extended precursor peptide. *J. Am. Chem. Soc*. **132**, 16324–16326.

Zalacain, M., and Cundliffe, E. (1989). Methylation of 23S rRNA caused by tlrA (ermSF), a tylosin resistance determinant from *Streptomyces fradiae*. *J. Bacteriol*. **171**, 4254–4260.

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

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

Zhang ,W., Watanabe, K., Cai, X., Jung, M.E., Tang, Y., and Zhan, J. (2008). Identifying the minimal enzymes required for anhydrotetracycline biosynthesis. *J. Am. Chem. Soc.* **130**, 6068–6069.

Zhang, L., Hashimoto, T., Qin, B., Hashimoto, J., Kozone, I., Kawahara, T., Okada, M., Awakawa, T., Ito, T., Asakawa, Y., Ueki, M., Takahashi, S., Osada, H., Wakimoto, T., Ikeda, H., Shin-Ya, K., and Abe, I. (2017). Characterization of giant modular PKSs provides insight into genetic mechanism for structural diversification of aminopolyol polyketides. *Angew. Chem. Int. Ed. Engl*. **56**, 1740–1745.

Zhang, W., Watanabe, K., Wang, C.C., and Tang, Y. (2007). Investigation of early tailoring reactions in the oxytetracyclinebiosynthetic pathway. *J. Biol. Chem*. **282**, 25717–25725.

Zhao, C., Coughlin, J.M., Ju, J., Zhu, D., Wendt-Pienkowski, E., Zhou, X., Wang, Z., Shen, B., and Deng, Z. (2010). Oxazolomycin biosynthesis in *Streptomyces albus* JA3453 featuring an "acyltransferase-less" type I polyketide synthase that incorporates two distinct extender units. *J. Biol. Chem*. **285**, 20097–20108.

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

Zheng, J., Taylor, C.A., Piasecki, S.K., and Keatinge-Clay, A.T. (2010). Structural and functional analysis of A-type ketoreductases from the amphotericin modular polyketide synthase. *Structure* **18**, 913–922.

Zhou, Z., Xu, Q., Bu, Q., Guo, Y., Liu, S., Liu, Y., Du, Y., and Li, Y. (2015). Genome mining‐directed activation of a silent angucycline biosynthetic gene cluster in *Streptomyces chattanoogensis*. *ChemBioChem* **16**, 496–502.

Ziemert, N., Lechner, A., Wietz, M., Millán-Aguiñaga, N., Chavarria, K.L., and Jensen, P.R. (2014). Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. P*roc. Natl. Acad. Sci. U.S.A.* **111**, E1130–E1139.

Ziemert, N., Podell, S., Penn, K., Badger, J.H., Allen, E., and Jensen, P.R. (2012). The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* **7**, e34064.