

2016

Studies Towards the Discovery of Novel Natural Products Through Functional Analysis of Environmental DNA-Encoded Type II Polyketide Synthases

Debjani Chakraborty

Follow this and additional works at: http://digitalcommons.rockefeller.edu/student_theses_and_dissertations

 Part of the [Life Sciences Commons](#)

Recommended Citation

Chakraborty, Debjani, "Studies Towards the Discovery of Novel Natural Products Through Functional Analysis of Environmental DNA-Encoded Type II Polyketide Synthases" (2016). *Student Theses and Dissertations*. Paper 295.



STUDIES TOWARDS THE DISCOVERY OF NOVEL NATURAL PRODUCTS
THROUGH FUNCTIONAL ANALYSIS OF ENVIRONMENTAL DNA-
ENCODED TYPE II POLYKETIDE SYNTHASES

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the degree of Doctor of Philosophy

by

Debjani Chakraborty

June 2016

STUDIES TOWARDS THE DISCOVERY OF NOVEL NATURAL PRODUCTS
THROUGH FUNCTIONAL ANALYSIS OF ENVIRONMENTAL DNA-
ENCODED TYPE II POLYKETIDE SYNTHASES

Debjani Chakraborty, Ph.D.

The Rockefeller University 2016

Natural products have historically served as a major source for most of the therapeutically relevant compounds. A significant fraction of these natural product-based drugs and leads are derived from microorganisms. Although the traditional culture-based strategy has proven very successful over the course of more than a century of investigation, it excludes a significant portion of the molecules present in environmental samples because majority of the bacteria are not readily amenable to culture-based strategies. This uncultured majority is believed to represent 99% of bacterial species in the environment. Therefore, it has remained a significant challenge to access the chemical diversity present in these microorganisms using culture-based methods. Most modern natural product discovery efforts are faced with the hurdle of high rediscovery rates (>99%) of known metabolites produced by easily cultured bacteria.

Metagenomics has evolved as an alternative approach to conventional microbial screening. By directly cloning environmental DNA (eDNA or metagenome) in a surrogate host, one can exhaustively investigate the metagenome, independent of the culturability of the source organisms. Metagenomic approaches can thus provide access to previously untapped pools of

chemical diversity. These techniques have been employed in the current thesis to search for a family of therapeutically relevant, polycyclic aromatic natural products, synthesized by Type II polyketide synthases (PKSs) in soil-borne bacteria. Through the functional characterization of soil eDNA-derived Type II PKS containing clones, herein we attempt to identify novel PKS systems that encode structurally diverse metabolites and explore their biological activity (Chapter 1).

My doctoral work towards the discovery of novel type II PKS derived natural products has led to the elucidation of the mode of action of eDNA encoded antibiotics Fasamycins A and B. To identify the target of the fasamycins, fasamycin A resistant mutants were generated. Next-generation sequencing of these mutants, along with *in vitro* biochemical assays, showed that fasamycins inhibit the FabF gene of the type II fatty acid biosynthetic system. Subsequently, candidate gene overexpression studies consolidated the mode of action of fasamycins. (Chapter 2)

Insights into the genomes of uncultured microorganisms indicate that they are rich in Type II PKS gene clusters. In our attempt to identify novel aromatic polyketide natural products, we investigated the eDNA of soil samples collected from Texas region for unique Type II PKSs using homology-based metagenomic screening techniques. The eDNA-derived clones containing unique type II polyketide synthase genes identified in this screening were examined for the ability to produce clone-specific metabolites in *Streptomyces*. This led to the identification of hitherto unknown biosynthetic gene cluster for Seitomycin, a Type II PKS derived antibiotic. (Chapter 3)

In addition to the work using homology-based metagenomic techniques outlined above, three unique Type II PKS clones were identified based on bioinformatics analysis and comparison with previously characterized PKS gene clusters. These novel pathways were functionally reconstructed from multiple overlapping clones using Transformation Associated Recombination (TAR). Upon complete reassembly, all of the three PKS pathways were demonstrated to produce clone-specific metabolites during heterologous expression studies in *Streptomyces* host. In future studies, characterization of these metabolites is likely to lead to novel Type II PKS derived natural products. (Chapter 4)

This thesis is dedicated to

My Family.

ACKNOWLEDGMENTS

I have had the fortune of working with talented, driven and innovative scientists during my graduate studies. I am grateful to Drs. Minkui Luo, David Jin, Lorraine Gudas and Sean Brady for allowing me to work in their laboratories as a rotation student and providing diverse scientific background that enriched my training and experience significantly. I am very much thankful to Dr. Sean Brady for accepting me as graduate student in his laboratory and introducing me to the exciting field of natural product discovery through metagenomics.

I am greatly indebted to my committee members Drs. Vincent Fishetti, Derek Tan and Luciano Marraffini for their guidance and advice that fostered my scientific development over the years. I cannot express how grateful I am for having had the opportunity to learn from them. I would like to thank Dr. Yueming Li for kindly serving on my committee and reviewing the thesis. I cannot thank enough Drs. Sidney Strickland and Emily Harms for their constant support and advice both on scientific and personal front. I would like to thank all the Brady lab members for being incredibly supportive during my studies.

I am very thankful to the Rockefeller University, the Tri-Institutional Training Program in Chemical Biology, the Howard Hughes Medical Institute and the National Institute of Health for kindly supporting my doctoral education and research.

Finally, I would like to thank my family for their love, support and encouragement.

TABLE OF CONTENTS

TABLE OF CONTENTS	v
LIST OF FIGURES	viii
LIST OF TABLES	xii
LIST OF ABBREVIATIONS	xiii
CHAPTER 1	1
Introduction and Background	
1.1 Natural Products	1
1.1.1 Polyketides	3
1.1.2 Type II polyketides	5
1.2 Metagenomics and uncultured microorganisms	8
1.2.1 Functional metagenomic screening	12
1.2.2 Homology based screening	15
1.2.2.1 Novel metabolites	16
1.2.2.2 Congeners of known metabolites	18
1.3 Future of metagenomic mining	19
CHAPTER 2	20
Target Identification of eDNA-Encoded Antibiotics Fasamycins A and B	
2.1 Introduction	20
2.2 Results	22
2.3 Discussion	31
2.4 Materials and Methods	33
2.4.1 Isolation of Fasamycins A (1) and B (2)	33
2.4.2 Selection of <i>Enterococcus faecalis</i> resistant to Fasamycin A	33
2.4.3 Genome sequencing and analysis methods	34
2.4.4 Cloning FASII genes from <i>E. faecalis</i> OG1RF	35
2.4.5 Bioactivity: Growth on elevated concentrations of Fasamycin A	35
2.4.6 Bioactivity: <i>Ex Vivo</i> studies	36
2.4.6 Bioactivity: <i>In Vitro</i> studies	37

CHAPTER 3 41

Homology-Based Metagenomic Screening Efforts Towards the Identification of Novel Type II Polyketide Synthase Gene Clusters

3.1 Introduction	41
3.2 Results	44
3.2.1 Screening for type II PKS pathways	44
3.2.2 Recovery, retrofitting and heterologous expression of eDNA-derived Type II PKS pathways	45
3.2.3 Sequencing of PKS containing eDNA-derived clones	48
3.2.4 Overlapping-clone recovery and transformation associated recombination (TAR) of eDNA-derived Type II PKS pathways	50
3.2.5 Isolation and structural characterization of eDNA encoded natural products: Seitomycin, Rabelomycin, 8-Methylrabelomycin, and 1-Deoxy-1-hydroxy-8-O-methylrabelomycin	51
3.2.6 Proposed biosynthesis of Seitomycin	59
3.3 Discussion	62
3.4 Materials and Methods	64
3.4.1 General experimental procedures	64
3.4.2 PCR screening of eDNA libraries for KS α genes and clone recovery	64
3.4.3 Cosmid retrofitting and integration into <i>Sreptomycetes albus</i>	65
3.4.4 eDNA clone sequencing	65
3.4.5 Transformation-associated recombination (TAR) reassembly of overlapping clones	66
3.4.6 Heterologous expression and analysis	67
3.4.7 Compound extraction and purification Seitomycin, Rabelomycin, 8-O-methylrabelomycin, and 1-Deoxy-1-hydroxy-8-O-methylrabelomycin	68

CHAPTER 4 87

Phylogeny-Guided Discovery and Heterologous Expression Studies of Novel Type II PKS Gene Clusters

4.1 Introduction	87
4.2 Results	90
4.2.1 KS β sequence-guided screening of soil eDNA libraries	90
4.2.2 Overlapping clone recovery	91
4.2.3 Gene cluster analysis of complete AB1215, AB1449 and AZ1237 pathways	93

4.2.4 TAR cloning for reconstruction of type II PKS pathways AB1215, AB1449 and AZ1237	97
4.2.5 Heterologous expression of the eDNA-derived gene clusters	99
4.2.6 Complementation of sugar biosynthesis	101
4.3 Discussion	102
4.4 Materials and Methods	106
4.4.1 Bioinformatic analysis of KS β sequence tags	106
4.4.2 Clone recovery and bioinformatics analysis	106
4.4.3 Cosmid retrofitting and integration into <i>Streptomyces albus</i>	107
4.4.4 Overlapping clone recovery and bioinformatic analysis	107
4.4.5 eDNA clone sequencing	109
4.4.6 Transformation-associated recombination (TAR) reassembly of three overlapping eDNA clones	109
4.4.7 Heterologous expression and analysis	111
4.4.8 Cloning of the rhamnose biosynthetic genes	112
 CHAPTER 5	 113
 Future Directions	
5.1 Better bioinformatics	113
5.2 Large insert metagenomic libraries	114
5.3 Activation of silent biosynthetic gene clusters	115
 REFERENCES	 117

LIST OF FIGURES

Figure 1. Natural products as therapeutic agents.	3
Figure 2. Polyketide biosynthesis.	5
Figure 3. Classification of polyketides.	6
Figure 4: Cyclization pattern of bacterial type II PKS derived polyketides.	7
Figure 5: Microbial natural product isolation: Traditional culture-based method.	9
Figure 6: Metagenomic library construction and screening.	11
Figure 7: Natural products identified through functional metagenomic screening	13
Figure 8: PCR-based screening for novel natural products (type II PKS).	16
Figure 9: Metabolites isolated in homology-based metagenomic screening.	17
Figure 10: Derivatives of known natural products identified in homology-based metagenomic screening.	19
Figure 11. Antibiotics Fasamycins A (1) and B (2).	22
Figure 12. Alignment of fabT mutations with reference genome sequence.	25
Figure 13. Sequence of genomic DNA from nine unique resistant strains.	26
Figure 14. Activity of the Fasamycins A and B was tested in an <i>in vitro</i> fatty acid elongation assay.	27
Figure 15. Schematic of FASII biosynthesis.	28
Figure 16. <i>In vitro</i> fatty acid elongation assay.	29
Figure 17. Mechanisms of Fasamycin A activity.	30
Figure 18. The gem-dimethyl-anthracenone substructure in selected natural products.	31

Figure 19. Iterative use of the minimal PKS for the assembly of aromatic type II polyketide backbone.	43
Figure 20. eDNA derived PKS probes.	44
Figure 21. An overview of TX library screening results.	45
Figure 22. A ClustalW-based phylogenetic tree of the eDNA-derived KS _a genes is shown.	46
Figure 23. ClustalW-based phylogenetic tree showing cosTX451 derived KS _a amplicon sequence and known PKS gene clusters.	46
Figure 24. <i>S. albus</i> transformed with eDNA-derived PKS clone cosTX451 was screened for the production of clone specific metabolites.	48
Figure 25. ORF diagram and gene table for TX451/2512 pathway.	49
Figure 26. TAR cloning of two overlapping eDNA clones carrying the type II PKS pathway BAC:TX451/2512.	50
Figure 27. HPLC analysis of crude ethyl acetate extracts of <i>S. albus</i> harboring BAC:TC451/2512 and cosTX451 clones.	52
Figure 28. Structural characterization of Seitomyacin (1).	53
Figure 29. Structural characterization of Rabelomyacin (2).	55
Figure 30. Structural characterization of 8- <i>O</i> -methylrabelomyacin (3).	57
Figure 31. Structural characterization of 1-Deoxo-1-hydroxy-8- <i>O</i> -methylrabelomyacin (4).	58
Figure 32. Proposed biosynthetic scheme for Seitomyacin (1).	60
Figure 33. ¹ H NMR spectrum of Seitomyacin (1)	71
Figure 34. ¹³ C NMR spectrum of Seitomyacin (1)	72
Figure 35. ¹ H- ¹ H COSY spectrum of Seitomyacin (1)	73
Figure 36. ¹ H- ¹³ C HMQC spectrum of Seitomyacin (1)	74

Figure 37. ^1H - ^{13}C HMBC spectrum of Seitomycin (1)	75
Figure 37a. High-resolution mass spectrum of Seitomycin (1)	76
Figure 38. ^1H NMR spectrum of Rabelomycin (2)	77
Figure 39. ^1H - ^1H COSY spectrum of Rabelomycin (2)	78
Figure 40. ^1H NMR spectrum of 8- <i>O</i> -Methylrabelomycin (3)	79
Figure 41. ^{13}C NMR spectrum of 8- <i>O</i> -Methylrabelomycin (3)	80
Figure 42. ^1H - ^{13}C HMQC spectrum of 8- <i>O</i> -Methylrabelomycin (3)	81
Figure 43. ^1H - ^{13}C HMBC spectrum of 8- <i>O</i> -Methylrabelomycin (3)	82
Figure 44. ^1H NMR spectrum of 1-deoxo-1-hydroxy-8- <i>O</i> -methylrabelomycin (4)	83
Figure 45. ^{13}C NMR spectrum of 1-deoxo-1-hydroxy-8- <i>O</i> -methylrabelomycin (4)	84
Figure 46. HMQC spectrum of 1-deoxo-1-hydroxy-8- <i>O</i> -methylrabelomycin (4)	85
Figure 47. HMBC spectrum of 1-deoxo-1-hydroxy-8- <i>O</i> -methylrabelomycin (4)	86
Figure 48. Metagenomic approach to natural product discovery.	87
Figure 49. Overview of the phylogeny-guided strategy for the discovery of new natural products.	89
Figure 50. Biosynthetic marker phylogeny.	90
Figure 51. Maximum likelihood tree of KS β genes from type II PKS gene clusters.	92
Figure 52. A partial maximum likelihood tree of partial KS β gene sequences.	93
Figure 53. ORF diagram and gene table for Type II PKS pathway AB1215/183.	94
Figure 54. ORF diagram and gene table for Type II PKS pathway AB1449/647.	95
Figure 55. ORF diagram and gene table for <i>Type II PKS pathway</i> AZ2250/1237/2224.	97
Figure 56. Transformation Associated Reassembly (TAR) of overlapping eDNA	

clones in yeast.	98
Figure 57. Clone specific metabolites for BAC-AB1215/183.	99
Figure 58. Clone specific metabolites for BAC-AB1449/647.	100
Figure 59. Clone specific metabolites for BAC-AZ22250/1237/2250.	101

LIST OF TABLES

Table 1: Biological activity of Fasamycins A (1) and B (2)	43
Table 2: Primers used for cloning <i>E. faecalis</i> FASII genes	57
Table 3: List of genes not used in biosynthetic hypothesis of seitomycin	61

LIST OF ABBREVIATIONS

AD: Adenylation domain

AMP: Adenosine monophosphate

ATP: Adenosine triphosphate

ACP: Acyl carrier protein

AB: Anza-Borrogo (eDNA-derived cosmid mega-library)

BAC: Bacterial artificial chromosome

Bla: Beta-lactamase (resistance gene)

BLAST: Basic local alignment search tool

CEN/ARS: Centromere/autonomous replicating sequence

CoA: Coenzyme A

COSY: COrrelation SpectroscopY

CSM: Complete synthetic medium

DNA: Deoxyribonucleic acid

DMSO: Dimethylsulfoxide

eDNA: Environmental DNA

ESI: Electrospray ionization

ESI/MS: Electrospray ionization mass spectrometry

Erd: Erdacin

FAD: Flavin adenine dinucleotide

HPLC: High pressure liquid chromatography

HPLC-MS: High pressure liquid chromatography–mass spectrometry

HRMS: High resolution mass spectrometry

IPP: Isopentanyl pyrophosphate

Kan: Kanamycin (or associated resistance gene)

Kb: Kilobase

KS: Ketosynthase

KS α : α -Ketosynthase

KS β : β -Ketosynthase

LB: Luria Bertani medium

LC-MS: Liquid chromatography–mass spectrometry

MS: Mass spectrometry

HMBC: Heteronuclear multiple bond correlation

NMR: Nuclear magnetic resonance spectroscopy

HMQC: Heteronuclear multiple-quantum correlation

MB: Megabase

NCBI: National center for biotechnology information

NRPS: Non ribosomal peptide synthetase

ORF: Open reading frame

OUT: Operational taxonomic unit (i.e. unique sequence)

PCA: Polymerase cycling assembly

PCR: Polymerase chain reaction

PKS: Polyketide synthase

PPTase: 4'-phosphopantetheinyltransferase

rRNA: Ribosomal ribonucleic acid

s.d. Standard deviation

SMM: Supplemented minimal medium

TAR: Transformation associated recombination

TLC: Thin-layer chromatography

T_m: Melting temperature

TPS: Terpene synthase

UT: Utah (eDNA-derived cosmid mega-library)

YAC: Yeast artificial chromosome

YCP: Yeast centromeric plasmid

CHAPTER 1

Introduction and Background

1.1 Natural Products

Throughout most of history, small molecule natural products have been a source of novel therapeutics for a wide spectrum of diseases. Since the early 1800s, discovery of natural products has started a new era of medicine where drugs could be purified from plants or microorganisms and administered in precise dosage. The term natural product really means any naturally occurring substance but is generally taken to mean a secondary metabolite - a molecule that has no essential function in metabolism. It's thought that these chemicals exist to increase the likelihood of an organism's survival by attracting or repelling other organisms. Among the first reported examples of natural products were strychnine, morphine, atropine, quinine, and colchicine. In 1826, E. Merck developed the first commercial natural product, morphine, and this was followed by the first semisynthetic, pure drug based on a natural product, aspirin, by Bayer in 1899.

The next milestone was the discovery of penicillin from the filamentous fungus *Penicillium notatum* by Alexander Fleming in 1929. This truly ushered in the "Golden Age of Antibiotics", the time period from the 1940s to the 1970s. Pharmaceutical companies launched extensive investigations of microbes as sources of novel antibiotics and led to the discovery of a host of other antibacterial antibiotics, including the tetracyclines (e.g., doxycycline), cephalosporins, aminoglycosides (e.g., streptomycin), lipopeptides (e.g., daptomycin), glycopeptides (e.g., vancomycin), and macrocyclic compounds such as erythromycin. (Newman *et al.* 2000, Cragg

and Newman 2001). Microbial natural products have also been the source of other anti-infective drugs in the form of antibiotics having antifungal (e.g., amphotericin, nystatin) and antiparasitic (e.g., ivermectin, fumagillin) activities. (Newman *et al.* 2000, Cragg and Newman 2001, Newman and Cragg 2012). Microbes have provided the template for the development of anticholesterolemic drugs, known as statins. The discovery of the fungal metabolites mevastatin (compactin) and lovastatin as inhibitors of HMG-CoA reductase led to the development of synthetic statin analogues, such as lovastatin and simvastatin (Mevacor). Antitumor antibiotics such as anthracyclines (e.g., doxorubicin), enediynes (e.g., calicheamicin), ansamycins (e.g., geldanamycin), peptolides (e.g., dactinomycin), epothilones (e.g., ixabepilone), mitosanes (e.g., mitomycin C), etc. are all microbial natural products that continue to play significant role in the drug discovery and development process. (Newman and Cragg 2012, Cragg *et al.* 2009, Cragg *et al.* 2012). Rapamycins are another important class of immunosuppressive drugs that complement the role of another microbial metabolite cyclosporine. (Cragg and Newman 2001). It is estimated that almost 70% of the small molecule drugs have their origin in natural products. (Newman and Cragg 2007). Thus, natural products have been, and continue to be, an indispensable source of novel drugs and pharmaceutical leads for the treatment of many human diseases.

Natural products isolated from cultured organisms have provided a large proportion of the most important pharmacophores discovered to date (Figure 1) (Newman, Cragg *et al.* 2003; Newman and Cragg 2004; Newman and Cragg 2007). Their bioactivity is attributed, to a large extent, to selective pressures that culminated in the evolution of structural and chemical properties of natural products to function effectively in biological contexts. In the native microbial hosts, the metabolites have been shown to play critical roles in processes like self-defense, nutrient scavenging and virulence. (Omura, Ikeda *et al.* 2001; Wolfgang, Kulasekara *et al.* 2003;

Nougayrède, Homburg *et al.* 2006; Wyatt, Wang *et al.* 2010). Soil-dwelling bacteria, especially actinomycetes, have been an abundant source of bioactive natural products. In fact, an estimated two thirds of all clinically useful antibiotics come from one bacterial genus alone, *Streptomyces* (Kieser *et al.* 2000). As a single gram of soil is predicted to harbor over 10,000 individual bacterial species, soil-bacteria seem to represent a promising reservoir of novel natural products with tremendous therapeutic potential (Torsvik *et al.* 1990; Curtis *et al.* 2002; Rappe and Giovannoni 2003)

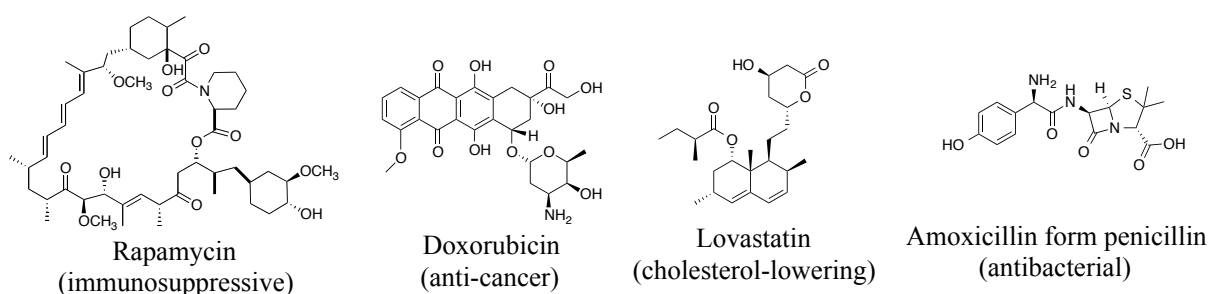


Figure 1. Natural products as therapeutic agents

There are hundreds of thousands of different natural products reported in literature. The *CRC Dictionary of Natural Products* (DNP) lists 190,939 records (DNP, Chapman and Hall, 2005). There is no rigidly defined scheme for classifying natural products. Their immense diversity in structure, function and biosynthesis is too vast to allow them to fit into a few simple categories. However, widely accepted are five major classes of natural products: polyketide, non-ribosomal peptides, alkaloids, terpenoids and enzyme co-factors.

1.1.1 Polyketides

Polyketides represent one of the largest families of secondary metabolites, especially abundant among actenimycete bacteria, eukaryotic fungi and higher plants. They are a highly diverse group of natural products having structurally intriguing carbon skeletons, which comprises

polyphenols, macrolides, polyenes, enediynes and polyethers. Although their exact roles in biological contexts are not known in all cases, it is believed that they function as pigments, virulence factors, infochemicals, or for defense. From a pharmacological viewpoint, polyketides are an important source of novel therapeutics. In particular, they are used in medicine mainly as antibiotics, immunosuppressants, antiparasitics, cholesterol-lowering, and antitumor agents. (O'Hagan 1991). Some notable examples include tetracyclines and erythromycin (antibiotics), doxorubicin (anti-cancer), avermectin and monencin (antiparasitics), FK506 (immunosuppressant), all from the bacterial group that produces a large proportion of such compounds, the actinomycetes.

These molecules are biologically synthesized by multifunctional polyketide synthases (PKSs), which catalyze repeated decarboxylative condensations between acyl thioesters (usually acetyl, propionyl, malonyl, or methylmalonyl) (Figure 2). Whereas the total chemical synthesis of polyketides is highly challenging, it is remarkable that their vast structural and functional diversity results from the controlled assembly of some of the simplest biosynthetic building blocks. Since the first biosynthetic investigations by Collie, who also coined the term 'polyketide', and the acetogenin hypothesis by Barton (Staunton and Weissman 2001), biosynthesis of polyketides has become a fertile area of research and discovery. With emerging techniques in molecular biology (Hopwood 1997), it has now become possible to understand the biosynthetic logic of polyketide diversity, and engineer biosynthetic pathways for the production of novel drug candidates. (Wilkinson and Micklefield 2007, Weissman and Leadlay 2005).

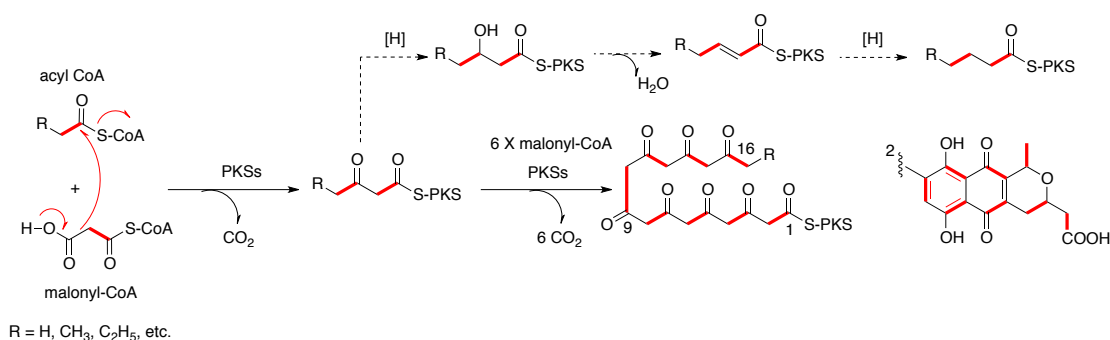


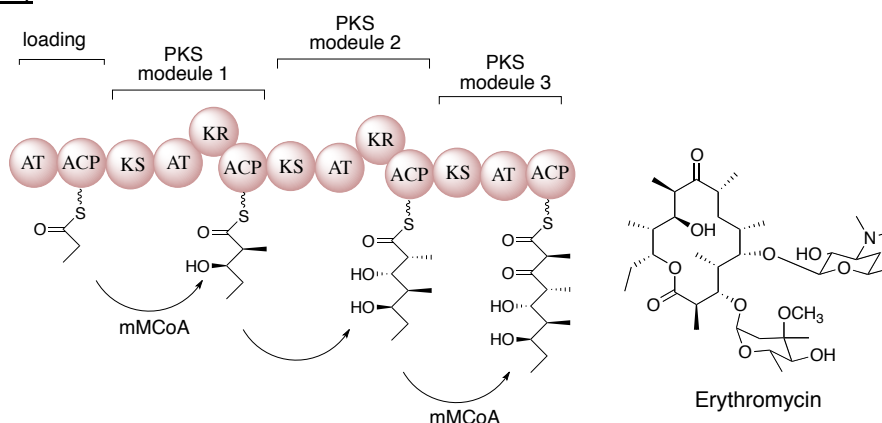
Figure 2. Polyketide biosynthesis: Concatenation of acetate units during polyketide biosynthesis, the decarboxylation of malonyl-CoA (or methylmalonyl-CoA), generates a nucleophilic enolate thioester which attacks an acyl thioester (or other starting unit) in a Claisen condensation

On the basis of architecture type and mode of action of the assembly lines, PKSs are classified in various types (Figure 3). Type I refers to linearly arranged and covalently fused catalytic domains within large multifunctional enzymes, whereas the term type II indicates dissociable complex of discrete and usually monofunctional enzymes. Furthermore, a third group of multifunctional enzymes of the chalcone synthase type is denoted as type III PKSs (Figure 3). Aside from the structures of the enzymes or enzyme complexes, the PKSs are also categorized as iterative or non-iterative, that is, whether or not each KS domain catalyze more than one round of elongation.

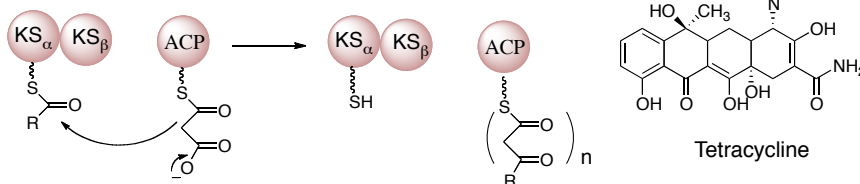
1.1.2. Type II polyketides

A number of polycyclic aromatic natural products are synthesized by polyketide synthases (PKSs) in soil-borne bacteria. They exhibit anticancer (e.g. doxorubicin), antibacterial (e.g. oxytetracycline), antifungal (e.g. pradimicin), antiviral (e.g. A-74528), antiparasitic (e.g. frenolicin) and other related activities. These PKSs, also called type II PKSs because of their relationship to type II fatty acid synthases from bacteria and plants, are a fascinating family of multifunctional catalysts that synthesize a wide range of interesting and medicinally important natural products.

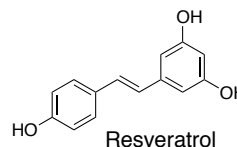
Type I (modular)



Type II (iterative heterodimers)



Type III (iterative homodimers) – chalcone/stillbene synthases



AT: acyl transferase ACP: acyl carrier protein KS: keto synthase KR: keto reductase DH: dehydratase
ER: enoyl reductase TE: thioesterase

Figure 3. Following the reaction similarity with fatty acid synthases (FASs), PKSs are classified into three types, type I, type II, and type III, based on their enzyme architecture. Type I PKSs consist of large multifunctional proteins with individual functional domains. These are further classified into single modular iterative type I PKS (iPKS) and multi-modular PKS (mPKS) in which one module consists of a set of catalytic enzyme domains (such as ketosynthase (KS), acyltransferase (AT) and acyl carrier protein (ACP), at the minimum), together with other modification domains. Type II PKSs consist of individual separate dissociable enzymes that form a multi-enzyme complex to carry out the reaction similar to bacterial FASs. In contrast to these multi-functional or multi-enzyme systems, type III PKSs are enzymes that maintain a simple architecture, i.e. a homodimer of condensation subunits, corresponding with KS domains of other types of PKS. This type of PKS uses acyl-CoA thioester directly, without using ACPs. Therefore, type III PKSs are the smallest and so-far well characterized in PKSs .

Iterative type II PKS systems are common in prokaryotic organisms. Methods for genetic manipulation of Gram-positive actinomycetes, pioneered by D. A. Hopwood and coworkers, (Malpartida and Hopwood 1984) provided the technical foundation for the cloning and

sequencing of a number of aromatic polyketide biosynthetic gene clusters. For example, the complete *act* gene cluster responsible for actinorhodin biosynthesis was cloned in 1984 (Malpartida and Hopwood 1984), and its PKS genes were fully sequenced by 1992 (Fernandez-Moreno, *et al.* 1992). A ketosynthase (KS), chain length factor (CLF) and acyl carrier protein (ACP) are encoded within the *act* gene cluster; these subunits were designated the ‘minimal PKS’ (Figure 3). Additional PKS subunits, including ketoreductases (KR), cyclases (CYC), and aromatases (ARO) define the folding pattern of the nascent poly-beta keto intermediate. In

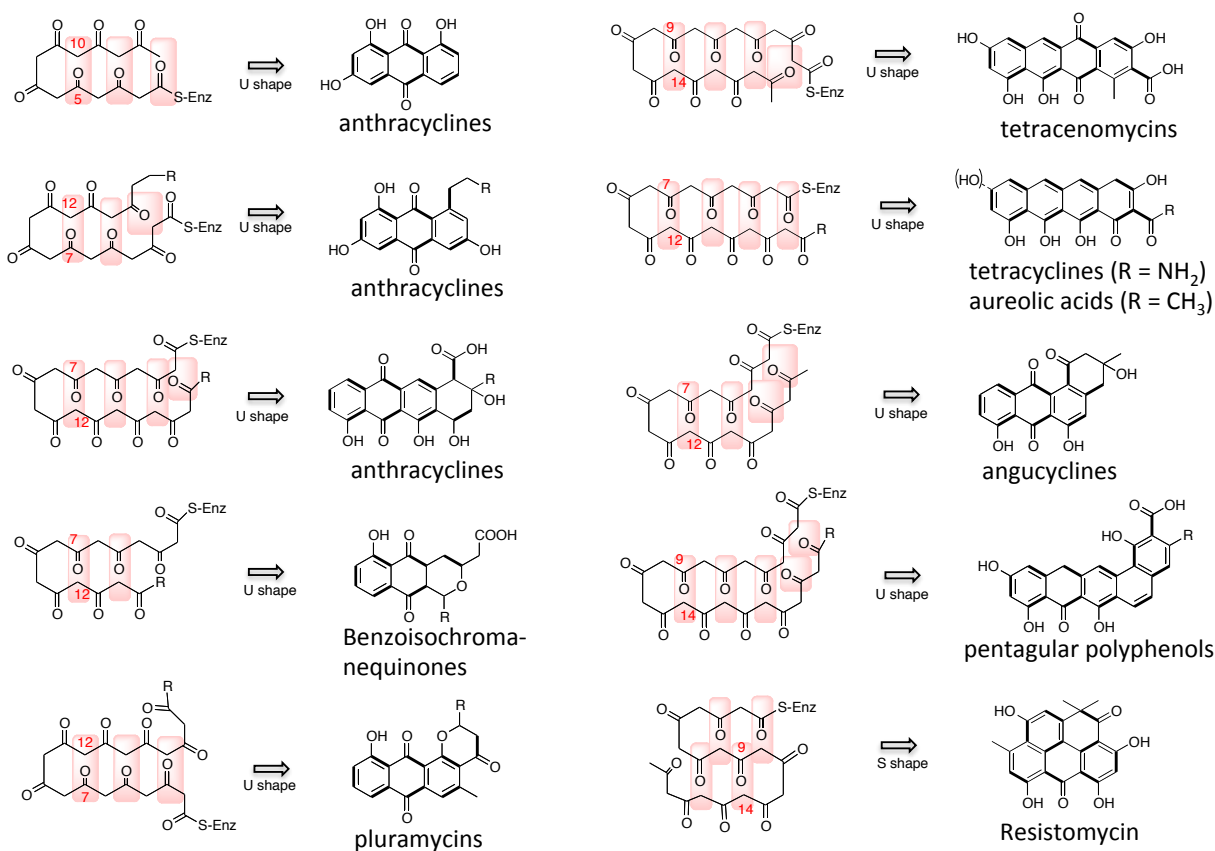


Figure 4. Cyclization pattern of bacterial type II PKS derived polyketides. On the basis of the carbocycles produced by polyketide folding and condensation, type II PKS products are further classified into various subtypes.

almost all type II polyketide pathways, cyclases mediate polyketide cyclization to generate the aromatic ring system (Figure 4). Studies indicate that the first cyclization (i.e. C9–C14 or C7–C12) is regulated by the PKS and the presence or absence of a KR has some impact on the preformation of a bend in the carbon chain. Cyclase functions are attributed to individual cyclization steps, such as second and third ring formation, and so on. While cyclases catalyze specific aldol condensations, aromatases support the cyclodehydration process (Hertweck 2009) (Figure 4). Given the clinical and commercial success of a diverse collection of aromatic compounds that arise from iterative type II polyketide synthases, considerable interest has developed to gain access to these microbial secondary metabolites. Accessing this particular class of polyketide natural products is the primary focus of the current doctoral work.

1.2. Metagenomics and uncultured microorganisms

Charting biologically relevant chemical space is of utmost importance for drug discovery and development. However, the discovery of new natural product antibiotics has declined over the past several decades (Fox 2006). There is more than one factor contributing to this decline. The isolation of small molecules from bacteria or fungi from environmental sources mostly relies on the decade-old methods and techniques (Figure 5). The traditional protocol used to isolate and identify natural products from bacteria is to first culture an individual bacterium from the environment, extract bacterial cultures using organic solvents, and isolate pure compounds through activity-guided fractionations (Figure 5). However, the traditional culture-based paradigm itself seems to be one of the major hurdles to new natural product discovery. Despite decades of historical productivity, pharmaceutical companies have recently deemphasized natural product discovery efforts due to the increasing rediscovery rates of common metabolites produced by easily cultured bacteria - a rate that now exceeds 99%. (Newman *et al.* 2004,

Newman *et al.* 2007, Newman *et al.* 2012). The difficulties associated with culturing environmental bacteria prevent traditional culture-dependent methods from being applied to a significant fraction of the bacterial species found in nature. Microbial 16S rRNA sequencing efforts indicate that only a small fraction of the world's environmental microbes have been cultured using standard microbiological methods. (Torsvik *et al.* 1990, Kaeberlein *et al.* 2002, Zengler *et al.* 2002, Gans *et al.* 2005, Tringe *et al.* 2005). From the bacterial kingdom, this minority is believed to represent less than 1% of all species in existence. These same analyses also indicate that more than 80 major bacterial divisions exist, but only less than half are represented by cultured isolates. (Schloss *et al.* 2004, Keller *et al.* 2004, DeSantis *et al.* 2006). Such insights suggest that uncultured bacteria are likely the largest remaining pool of genetic and chemical diversity on the planet. Thus, it is apparent that the vast majority of natural products produced by microorganisms found in the biosphere are beyond the scope of the traditional culture-dependent paradigm.

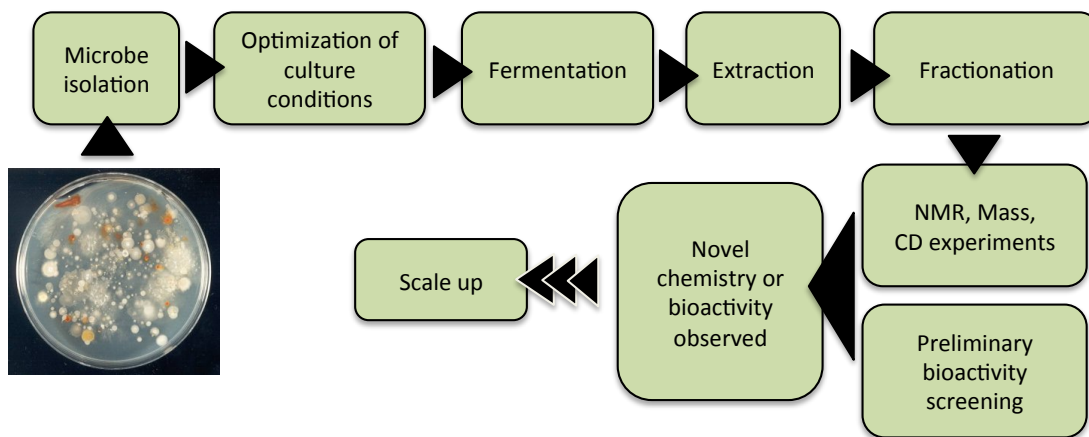


Figure 5. Microbial natural product isolation: Traditional culture-based method. Cultured environmental samples (in this example, bacteria) are isolated and crude extracts are generated. Iterative rounds of fractionation and activity assays are used until a pure compound is obtained for structural elucidation.

There are a number of different strategies, using both culture-dependent and culture-independent

methods, which are now being developed to access this untapped reservoir of chemical diversity. Typically, culture-independent strategies involve direct cloning of the genetic material from an environmental sample, which contains the genes encoding for small molecule (Figure 6). The genetic diversity of all the bacteria in any given environmental sample has been termed the “metagenome” (Handelsman, Rondon, Brady, *et al.* 1998). While genomics is the study of an individual organism's DNA, metagenomics transcends the individual genome. Taking advantage of the advances in DNA-sequencing technology, this new approach allows the study of entire communities of microbes simultaneously, bypassing the need for isolation and culture. It therefore seems likely that we can capture the biosynthetic potential of uncultured bacteria through metagenomics and gain access to unidentified bioactive metabolites for future drug discovery efforts. Metagenomics is fast evolving as an alternative to conventional microbial screening for natural product discovery.

Studying natural product gene clusters using metagenomic techniques requires eDNA (environmental DNA) cloning and sequence enrichment techniques. DNA extracted from environmental sample, such as soil and water, contains large quantities of genetic material, including secondary metabolite biosynthetic gene clusters (Figure 6). DNA extracted from environmental sample, such as soil and water, contains large quantities of genetic material, including secondary metabolite biosynthetic gene clusters. This eDNA is purified and cloned into a vector that allows the high efficiency transformation of a tractable library host, typically *E. coli*. To accomplish this, eDNA can be ligated into a cosmid vector, packaged into lambda phage and transfected into *E. coli* (Brady 2007). This cloning method can capture up to 50 kb pieces of eDNA in a single clone. Since the genes responsible for the biosynthesis of natural products are typically clustered together, it is possible to capture an entire biosynthetic gene cluster or a large

part of it on a single cosmid (Handelsman, Rondon, Brady *et al.* 1998).

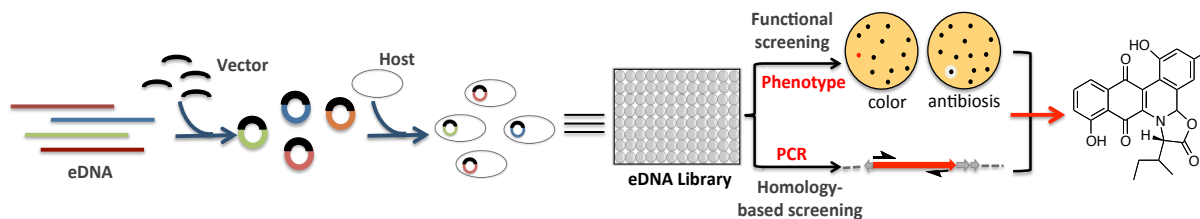


Figure 6. Metagenomic library construction and screening: eDNA is purified and cloned into a vector that allows high efficiency transformation into a culturable library host. One of two primary screening methodologies (functional/homology-based) are used to identify clone-specific molecules or biosynthetic gene clusters for downstream heterologous expression and characterization efforts.

Metagenomic methods are universally applicable and have already been applied to microbial communities from soils/sediments (Rondon *et al.* 2000; Brady *et al.* 2002; Voget *et al.* 2003), rumen gut (Brulc *et al.* 2009), planktonic marine microbial assemblages (Beja *et al.* 2000, Breitbart 2002), deep sea microbiota (Sogin *et al.* 2006), an acid mine site (Tyson *et al.* 2004), arctic sediments and the Sargasso Sea (Venter *et al.* 2004). The foundation of all metagenomic approaches is the isolation and subsequent examination of DNA extracted directly from naturally occurring microbial populations, which avoids the difficulties associated with culturing environmental bacteria (Handelsman *et al.* 1998). As previously mentioned, secondary metabolite biosynthetic gene clusters are often clustered in the bacterial genome. Metagenomics is particularly appealing to natural product researchers because of this characteristic clustering, since most secondary metabolite biosynthetic gene clusters are under 100 kb, making it possible to capture biosynthetic gene clusters on individual or, at most, a small number of overlapping eDNA clones (Handelsman *et al.* 1998). Large scale cosmid cloning of environmental DNA therefore provides a platform from which one can investigate secondary metabolite biosynthetic gene clusters from uncultured bacteria.

1.2.1. Functional metagenomic screening

One approach to access the chemical diversity of uncultured microbes is through expression-dependent or functional screening (Figure 6). In functional metagenomic studies, eDNA libraries are examined in simple high throughput assays designed to identify clones that have phenotypes traditionally associated with the production of small molecules, such as pigment production, antibiosis, or altered colony morphology. One of the simplest strategies used to detect eDNA clones that might produce small molecule antibiotics has been to screen libraries hosted in *E. coli* for clones that generate zones of growth inhibition against test microbes in top agar overlay assays. The isolation of clone specific metabolites produced by active eDNA clones identified from bacterial top agar overlay assays has led to the characterization of a variety of new longchain N-acylated amines (**3**), as well as a new isonitrile functionalized indole antibiotic (**5**) (Figure 7) (Brady and Clardy 2000; Brady and Clardy 2005). Antibiotics have also been found by examining pigmented eDNA clones, and also through the direct examination of fermentation extracts from randomly selected clones (Wang, Graziani *et al.* 2000; Brady, Chao *et al.* 2001; MacNeil, Tiong *et al.* 2001; Gillespie, Brady *et al.* 2002; Lim, Chung *et al.* 2005; Long, Dunlap *et al.* 2005). Compounds with bioactivity identified from these types of studies include the antibiotic pigments violacein (**2**) and indigo (**1**) all recovered from soil libraries (Figure 7).

Functional metagenomics has also been used to identify clones that produce proteins with potential anti-infective properties (Schipper, Hornung *et al.* 2009). Although nearly all small-molecule-focused functional metagenomic studies have been carried out in *E. coli*, it is likely that the majority of the biosynthetic diversity present in an environmental sample is not functionally accessible using the same heterologous host. A computational analysis of promoters

and ribosomal binding sites used by a taxonomically diverse group of sequenced bacteria found that at most, 40% of the enzymatic activities present within a typical metagenomic sample could be accessed using *E. coli* as a heterologous host (Gabor, Alkema *et al.* 2004). Therefore, libraries initially constructed in *E. coli* were later shuttled into different hosts including *S. lividans*, *Ralstonia metallodurans*, *Rhizobium leguminosraum*, *Agrobacterium tumefaciens*, *Burkholderia graminis*, *Caulobacter vibrioides*, and *Pseudomonas putida* (Wang, Graziani *et al.* 2000; Martinez, Kolvek *et al.* 2004; Li, Wexler *et al.* 2005; Craig, Chang *et al.* 2009; Craig, Chang *et al.* 2010).

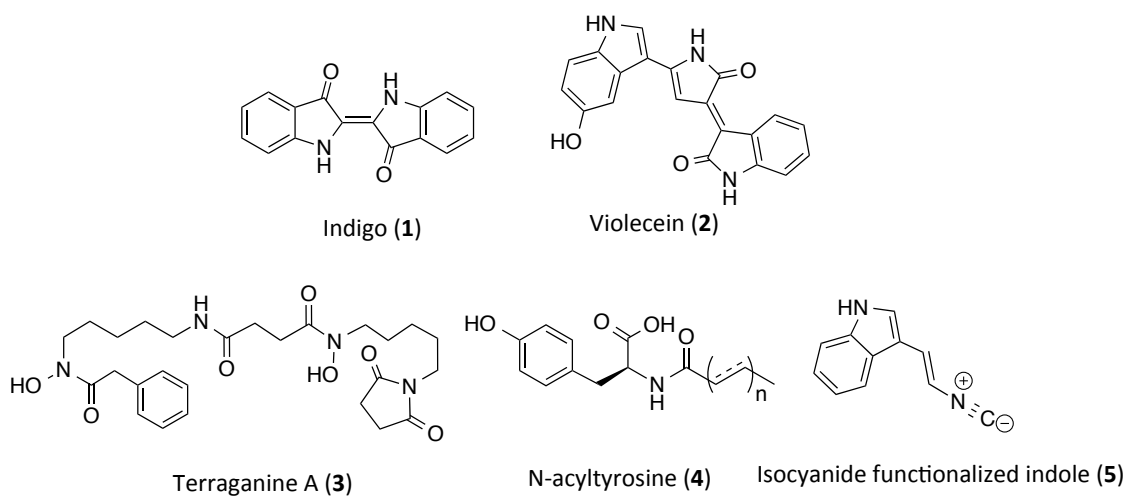


Figure 7. Natural products identified through functional metagenomic screening

Functional metagenomic screening strategies have been used to identify eDNA clones carrying natural product biosynthetic gene clusters. One primary advantage of a functional screen is that no *a priori* knowledge of the biosynthetic enzymes is required to discover novel metabolites. In functional screens, a secondary metabolite is directly linked to its biosynthetic genes allowing the unbiased discovery of novel biosynthetic sequences that have not been characterized before. Although several novel metabolites have been uncovered using functional screens, successful

expression of entire biosynthetic gene clusters requires the coordinated production of multiple proteins in a clone under laboratory culture conditions. Because the gene clusters isolated from a metagenomic sample are of diverse phylogenetic origin, the likelihood that a gene cluster meets all of these requirements is very low, and hit rates for functional screens of metagenomic libraries are generally around 0.01% (Courtois, Cappellano *et al.* 2003; Williamson, Borlee *et al.* 2005; Guan, Ju *et al.* 2007; Brady, Simmons *et al.* 2009). Due to these limitations, functional screens of metagenomic libraries have been designed so that they can be easily carried out on a large number of clones.

Although functional metagenomic screening is a powerful technique to gain access to the chemical diversity encoded in uncultured microorganisms, this strategy still requires some optimization. As described earlier, metagenomic cloning relies primarily on cosmid-based vectors. However, many canonical biosynthetic gene clusters are too large (>50 kb) to be captured on single cosmids and are truncated. This technical hurdle presents a major limitation of using functional assays to discover new natural products from cosmid-based eDNA libraries. Aside from this, the diversity present in environmental microbiomes makes the selection of an optimal heterologous screening host quite challenging (Torsvik, Øvreås *et al.* 2002). Heterologous expression hosts are inherently limited in their capacity to functionally process foreign DNA. The expanded use of phylogenetically diverse heterologous expression hosts will, therefore, most likely continue to benefit functional screening efforts. Exploring vector-host pairs that allow for the screening of metagenomic libraries in phylogenetically diverse microorganisms have the potential to expand the chemical diversity accessed through functional metagenomic studies.

1.2.2. Homology based screening

Expression-independent or homology-based screening of metagenomes is based on PCR amplification of conserved natural product biosynthetic gene sequences to identify and recover gene clusters from environmental DNA (eDNA) libraries. This allows targeted recovery of specific biosynthetic pathways from the genomes of all bacterial species present within an environmental sample.

Homology-based screens rely on the homology of the unknown, eDNA-derived secondary metabolite biosynthetic pathways to sequenced clusters of known metabolites. In these studies, eDNA libraries are probed to identify clones that contain conserved sequences associated with secondary metabolite biosynthesis. Bioinformatics analyses of sequences recovered through homology screening allow for the exclusion of sequences that are likely associated with gene clusters encoding for previously encountered metabolites, significantly reducing the likelihood for redundant isolation and increasing the likelihood of finding hitherto rare metabolites. Thus, homology-based screening coupled with phylogenetic analyses of sequences captured in large eDNA libraries should permit routine discovery of biosynthetic gene clusters encoding for metabolites belonging to even previously rare families of bioactive natural products.

Homology-based screening utilizes degenerate oligonucleotides to amplify sequence homologs via the polymerase chain reaction (PCR) (Seow, Meurer *et al.* 1997). Degenerate PCR primers are designed from previously sequenced conserved regions of a biosynthetic gene of interest. Screening metagenomic libraries with these degenerate primers enables the identification and recovery of individual clones with homologous genes from an eDNA library. These type of screening can be used to find new derivatives of known metabolites by targeting pathway-

specific biosynthetic genes. These primers can also be designed to identify novel structures by targeting conserved general biosynthetic genes found in a class of compounds. For example, degenerate primers based on conserved regions of minimal PKS KS_{α} gene are used to amplify KS_{α} sequences captured in the Texas eDNA library. The unique (<85% identity to previously known KS_{α} genes) eDNA-derived KS_{α} sequences then used as probes to identify and recover the cosmid clones containing type II PKS biosynthetic gene clusters from the metagenomic library (Figure 8). There have been multiple reports in the literature describing the identification of novel metabolites derived from type II PKS biosynthetic systems from eDNA samples (Hertweck 2009).

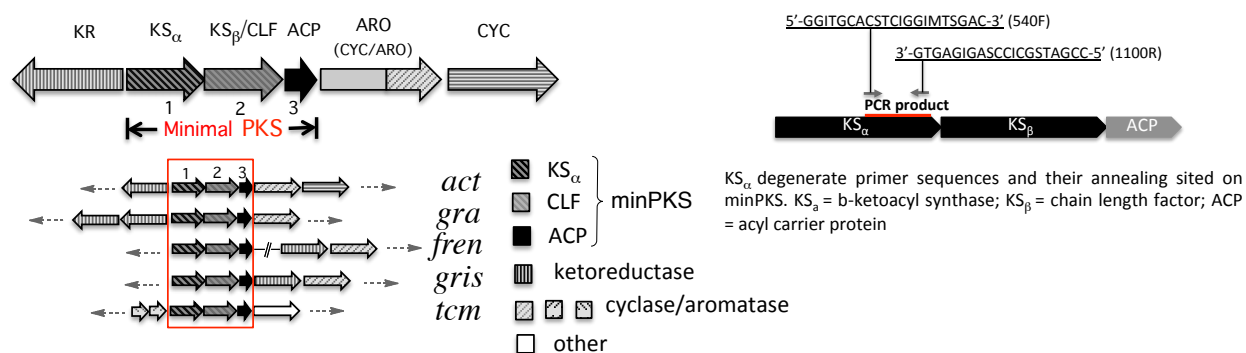


Figure 8. PCR-based screening for novel natural products (type II PKS): Degenerate PCR primers designed to recognize a conserved region in KS_{α} sequences from minPKS were used to amplify partial KS_{α} sequences from eDNA (Wawrik *et al.* 2005).

1.2.2.1. Novel metabolites

Homology-based metagenomic screening is proving to be a rewarding avenue to access the structural diversity encoded by uncultured bacteria. It has been successfully used to find novel structures by targeting conserved general biosynthetic genes found in a class of compounds. One of the first groups of molecules targeted using this strategy was the iterative (type II, aromatic)

polyketides. Type II polyketides utilize a small collection of biosynthetic enzymes to generate a diverse array of scaffolds, which are modified by various downstream enzymes. Although the gene clusters that encode the biosynthesis of these diverse metabolites can differ substantially in gene content, they all encode a highly conserved minimal PKS composed of three proteins:

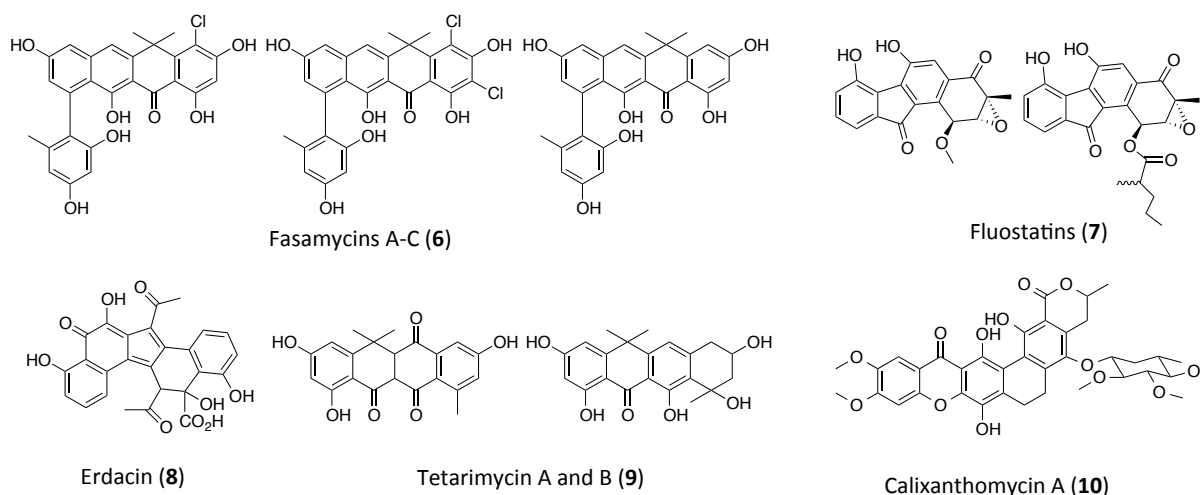


Figure 9. Metabolites isolated in homology-based metagenomic screening. A number of aromatic polyketides have been isolated from soil eDNA libraries.

ketosynthase alpha (KS_{α}); ketosynthase beta/chain length factor (KS_{β}); and acyl carrier protein (ACP). This “minimal PKS”, which is responsible for the iterative condensation of malonyl-CoA into a nascent polyketide chain, possesses a conserved organization, and is therefore an ideal target to identify using homology-based metagenomics (Macpherson *et al.* 1994). The examination of eDNA libraries and metagenomic sequencing data for relatives of known biosynthetic systems is likely to be a generally applicable strategy for identifying new structural variants of many bacterially derived antibiotics, potentially providing ready access to compounds with improved spectrum of activity (Figure 9). Kang and Brady recently identified a collection of novel bioactive pentangular polyphenols, calixanthomycin A, arenimycins C-D, Arixanthomycins A-C (Kang and Brady 2014, Kang and Brady 2013) through homology-based

screening. Using similar strategy, the antibiotic fasamycins (**6**) were isolated from soil and found to inhibit FabF of type II fatty acid biosynthesis (Feng, Kallifidas *et al.* 2011; Feng, Chakraborty *et al.* 2012). King *et al.* expressed a pathway encoding for the biosynthesis of a type II polyketide and characterized a new molecule erdacin (**8**) (King *et al.* 2009) with previously unknown pentacyclic skeleton, lending credence to the idea that uncultured bacteria may contain chemical diversity unlike that seen in cultured bacteria (King *et al.* 2009). A collection of fluostatins (**7**) and the methicillin-resistant *Staphylococcus aureus* (MRSA)-active antibiotic tetarimycin A (**9**) have also been isolated from soil libraries (Feng, Kim *et al.* 2010; Kallifidas, Kang *et al.* 2012) (Figure 9). In addition to isolating type II polyketides, homology-based metagenomic screening has been successfully used to identify novel tryptophan dimer analogues from eDNA libraries (Chang *et al.* 2015, Chang and Brady 2014, Chang *et al.* 2013, Chang and Brady 2011).

1.2.2.2. Congeners of known metabolites

Homology-based metagenomic approaches have been used in a number of studies to identify new derivatives of known compounds (Donia, Hathaway *et al.* 2006; Schmidt and Donia 2009; Ziemert, Ishida *et al.* 2010). Banik *et al.* employed a PCR screen targeting OxyC, a glycopeptide oxidative coupling enzyme, to identify a number of glycopeptide biosynthetic gene clusters from soil eDNA libraries (Banik and Brady 2008; Banik, Craig *et al.* 2010) that produced novel mono, di- and trisulfated glycopeptide congeners from the teicoplanin aglycone (**11**) (Figure 10) (Banik and Brady 2008). In another homology-based metagenomic study, Kang *et al.* identified an eDNA-derived cluster encoding for arimetamycin A (**12**), an anthracycline found to be more potent than clinically used natural anthracyclines against multidrug-resistant cancer cells (Kang and Brady 2013).

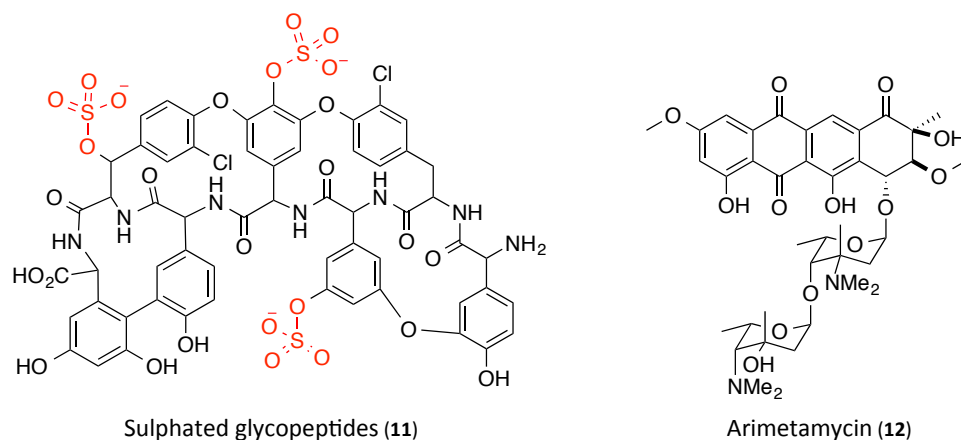


Figure 10. Derivatives of known natural products identified in homology-based metagenomic screening: Three eDNA-derived sulfotransferases were used to produce mono-, di- and trisulfated glycopeptide congeners from teicoplanin aglycone. An anthracycline analogue with high activity against cancer cells was isolated in a similar study.

1.3. Future of metagenomic mining

Recent development in a range of experimental techniques has made possible the identification, isolation and cloning for novel biosynthetic genes at a reasonable pace. Metagenomics is one such approach for engendering novel genes. The efforts described in this chapter clearly demonstrate the potential of uncultured bacteria as a rich source of biologically active novel small molecules and metagenomics as means for accessing such molecules from uncultured microbes from essentially any environment. The foundation of all metagenomic approaches is the isolation and subsequent examination of DNA extracted directly from naturally occurring microbial populations, which avoids the difficulties associated with culturing environmental bacteria. The contributions made by metagenomics to the already existing reservoir of prokaryotic genes are quite impressive. However, metagenomics is still in its infancy in terms of technical development. Advances in heterologous gene expression, library construction, vector design and screening techniques will lead to further improvement towards a robust approach for isolating natural products with expanded chemical space and unique biological activity. The

results discussed in the current doctoral thesis begin to explore these areas in an endeavor to discover novel microbial natural products.

CHAPTER 2

Target Identification of eDNA-Encoded Antibiotics Fasamycins A and B

2.1 Introduction

The potential clinical utility of any antibiotic can only be augmented by efforts to gain more information about their mechanisms of action. A significant increase in bacterial resistance has recently overcome almost every class of antimicrobials in clinical use (Levy and Marshall 2004). Methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococcus* (VRE) are two of the most widely spread antibiotic-resistant organisms that cause nosocomial infections. Rates of resistance among these organisms have increased considerably in the past decade. About 30% of Enterococcal and 70% of *S. aureus* infections are reported to be resistant to common antibiotics (Moran *et al.* 2006, Klein *et al.* 2007). New antibiotics with novel targets are needed to combat these common drug resistant phenotypes. Discovery of novel antibiotics and elucidation of their molecular targets is therefore key to success in our battle against multi-drug resistant microorganisms.

One approach for determining the mode of action of a small molecule involves the selection and full genome sequencing of mutants that acquire compound resistance (O'Neill and Chopra 2004). The current chapter describes our efforts to decipher the mode of action of the novel Type II PKS derived antibiotics Fasamycins A (1) and B (2) employing this strategy (Figure 11). These two antibiotics were initially identified in a metagenomic homology-based screening endeavor through heterologous expression of Type II (iterative, aromatic) polyketide (PK)

biosynthetic gene clusters cloned directly from soil samples. These metabolites show activity against methicillin-resistant *S. aureus* and vancomycin-resistant *E. faecalis* (Feng, Kallifidas *et al.* 2011) (Figure 11).

Fasamycins bear close structural resemblance to KB-3346-5 substances reported in patent literature (Satoshi *et al.* 2009). However, there is no report on the mode of action of the fasamycins or that of the KB-3346-5 substances. In the current study, we generated multiple fasamycin mutant strains and sequenced their genomes. Each strain contained mutations in the FabT gene of Type II fatty acid synthase (FASII) cluster. FabT is a transcriptional repressor that regulates expression of FASII genes in *E. faecalis*. Disruption of FabT is predicted to augment the expression of this gene cluster. We therefore employed the constitutive overexpression strategy to find out which gene from the FASII is responsible for the antibiotic activity of the fasamycins. Subsequent *in vitro* biochemical assays confirmed the identification FabF as the *in vivo* target of the fasamycins.

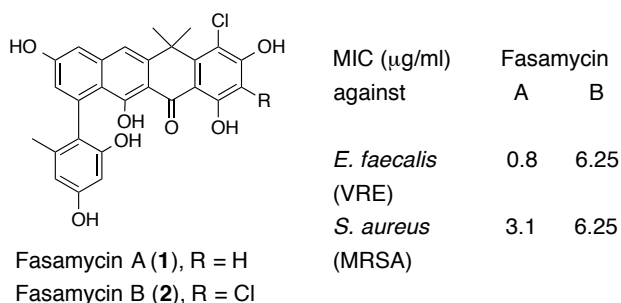


Figure 11. Antibiotics Fasamycins A (1) and B (2) were identified through the heterologous expression of type II polyketide gene cluster cloned directly from soil samples. Fasamycins showed activity against methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant *Enterococcus faecalis* (Feng, Kallifidas *et al.* 2011)

2.2 Results

Fasamycins A and B are environmental DNA (eDNA) derived antibiotics from a Type II PKS

biosynthetic gene cluster (Feng, Kallifidas *et al.* 2011). Both metabolites demonstrate Gram-positive specific antibiotic activity (Table 1).

Table 1. Biological Activity of Fasamycins A (1) and B (2)

Organism	1 ^a	2
Gram-positive bacteria		
<i>Bacillus subtilis</i> BR151	3.1	6.25
<i>Staphylococcus aureus</i> 6538P	3.1	6.25
<i>Staphylococcus aureus</i> USA300 (MRSA)	3.1	6.25
<i>Enterococcus faecalis</i> OG1RF	0.8	6.25
<i>Enterococcus faecalis</i> EF16 (VRA)	0.8	6.25
Gram-negative bacteria		
<i>Escherichia coli</i> EC100	>100	>100
<i>Escherichia coli</i> BAS849	12.5	25
<i>Ralstonia metallidurans</i> CH34	>100	>100
<i>Burkholderia thailandensis</i> E264	>100	>100
Eukaryote		
<i>Saccharomyces cerevisiae</i> w303	>100	>100

^aMinimum inhibitory concentrations are reported as µg/mL.

Monohalogenated Fasamycin A consistently showed higher activity than its dihalogenated counterpart, Fasamycin B against the Gram-positive bacteria that were examined (Table 1). Although the fasamycins are inactive against wild-type Gram-negative bacteria, they still exhibit activity against membrane permeabilized *E. coli* (BAS849). This data suggests that the lack of antibiotic activity of fasamycins towards Gram-negative bacteria is most likely due to their inability to access the right target. Neither metabolite showed cytotoxicity against yeast even at the highest concentrations tested.

Investigation of antibiotic resistance strains can provide useful insights into the mode of action of the particular antibiotic. To decipher the mechanistic principles of the fasamycins, a resistance

phenotype selection strategy was employed. A single wild-type colony of *E. faecalis* OG1RF was grown to confluence overnight and the resulting culture was plated on LB agar containing about 6 times the Minimum Inhibitory Concentration (MIC) for Fasamycin A (5 µg/mL) (Bourgogne *et al.* 2008, Dunny *et al.* 1978). For every 10^8 – 10^9 bacterial cells we plated, one grew at this elevated Fasamycin A concentration. The advantage of this is that by selecting mutants from a natural population of cells instead of from cultures treated with mutagenizing agents, the number of mutations per genome is minimized. With only a small number of mutations per genome, it should be much easier to establish a functional link between a resistance phenotype and the specific mutation leading to this phenotype.

Genomic DNA isolated from nine unique resistant colonies was sequenced using Illumina HiSeq2000 technology. The mutant genomes were then compared to *E. faecalis* OG1RF genome and queried for sequence differences (Figure 12). In all cases, the mutations were found in the *fabT* gene of the fasamycin resistant mutant strains. In total, nine mutant strains were sequenced and five unique *fabT* mutations (M1–M5) were found, each of which is predicted to result in truncation of the *fabT* gene product (Figure 13).

These mutations were further confirmed by PCR amplification and re-sequencing of the *fabT* locus from each mutant. The *fabT* gene is predicted to encode a MarR-like transcriptional repressor (Seoane and Levy 1995, Lu *et al.* 2004). In *E. faecalis*, *fabT* is found in the type II fatty acid biosynthesis (FASII) gene cluster (Figure 13) (Schujman and Mendoza, 2008). Therefore, if the FabT transcriptional repressor is unable to express functional protein product, it will lead to an increased expression of this fatty acid biosynthesis gene cluster. Based on this

WT	ATGGAACCTA	ATTTAGAAAC	AGTCAACGAT	TACTTAGTCA	GTGTCCTTAA	TGACATCTTA
M1	-----	-----	-----	-----	-----	-----
M2	ATGGAACCTA	ATTTAGAAAC	AGTCAACGAT	TACTTAGTCA	GTGTCCTTAA	TGACATCTTA
M3	ATGGAACCTA	ATTTAGAAAC	AGTCAACGAT	TACTTAGTCA	GTGTCCTTAA	TGACATCTTA
M4	ATGGAACCTA	ATTTAGAAAC	AGTCAACGAT	TACTTAGTCA	GTGTCCTTAA	TGACATCTTA
M5	ATGGAACCTA	ATTTAGAAAC	AGTCAACGAT	TACTTAGTCA	GTGTCCTTAA	TGACATCTTA
WT	ACGATTGAAG	AATCTGAATT	GAAAAAATCA	CAATTTAATG	ATTTATCCAT	TACGGAGATG
M1	-----	-----	-----	-----	-----	-----TG
M2	ACGATTGAAG	AATCTGAATT	GAAAAAATCA	CAATTTAATG	ATTTATCCAT	TACGGAGATG
M3	ACGATTGAAG	AATCTGAATT	GAAAAAATCA	CAATTTAATG	ATTTATCCAT	TACGGAGATG
M4	ACGATTGAAT	AATCTGAATT	GAAAAAATCA	CAATTTAATG	ATTTATCCAT	TACGGAGATG
M5	ACGATTGAAG	AATCTGAATT	GAAAAAATCA	CAATTTAATG	ATTTATCCAT	TACGGAGATG
WT	CACACAATCG	AAGCAATTGG	TATGTACAAG	AAAAAAA-CGT	CTTCAGAAGT	AGCCAAGGAA
M1	CACACAATCG	AAGCAATTGG	TATGTACAAG	AAAAAAA-CGT	CTTCAGAAGT	AGCCAAGGAA
M2	CACACAATCG	AAGCAATTGG	TATGTACAAG	AAAAAAA--CGT	CTTCAGAAGT	AGCCAAGGAA
M3	CACACAATCG	AAGCAATTGG	TATGTACAAG	AAAAAAA-CGT	CTTCAGAAGT	AGCCAAGGAA
M4	CACACAATCG	AAGCAATTGG	TATGTACAAG	AAAAAAA-CGT	CTTCAGAAGT	AGCCAAGGAA
M5	CACACAATCG	AAGCAATTGG	TATGTACAAG	AAAAAAAACGT	CTTCAGAAGT	AGCCAAGGAA
WT	CTGTCAATTA	CCGTCGGAAC	ACTTACTGTA	GCGATTAACA	ACTTAGTTAA	AAAAGGCTAT
M1	CTGTCAATTA	CCGTCGGAAC	ACTTACTGTA	GCGATTAACA	ACTTAGTTAA	AAAAGGCTAT
M2	CTGTCAATTA	CCGTCGGAAC	ACTTACTGTA	GCGATTAACA	ACTTAGTTAA	AAAAGGCTAT
M3	CTGTCAATTA	CCGTCGGAAC	ACTTAC-GTA	GCGATTAACA	ACTTAGTTAA	AAAAGGCTAT
M4	CTGTCAATTA	CCGTCGGAAC	ACTTACTGTA	GCGATTAACA	ACTTAGTTAA	AAAAGGCTAT
M5	CTGTCAATTA	CCGTCGGAAC	ACTTACTGTA	GCGATTAACA	ACTTAGTTAA	AAAAGGCTAT
WT	GTTGAACGTT	TGCGTAGTGA	AGATGATCGA	CGTGTAGTAA	AGCTTGGCTT	AACCAAAAAG
M1	GTTGAACGTT	TGCGTAGTGA	AGATGATCGA	CGTGTAGTAA	AGCTTGGCTT	AACCAAAAAG
M2	GTTGAACGTT	TGCGTAGTGA	AGATGATCGA	CGTGTAGTAA	AGCTTGGCTT	AACCAAAAAG
M3	GTTGAACGTT	TGCGTAGTGA	AGATGATCGA	CGTGTAGTAA	AGCTTGGCTT	AACCAAAAAG
M4	GTTGAACGTT	TGCGTAGTGA	AGATGATCGA	CGTGTAGTAA	AGCTTGGCTT	AACCAAAAAG
M5	GTTGAACGTT	TGCGTAGTGA	AGATGATCGA	CGTGTAGTAA	AGCTTGGCTT	AACCAAAAAG
WT	GGAAAATTAC	TTTTTAGAGT	CCATCAGCAT	TTTCATCGTG	AAATGGTTAA	AAACATTTTG
M1	GGAAAATTAC	TTTTTAGAGT	CCATCAGCAT	TTTCATCGTG	AAATGGTTAA	AAACATTTTG
M2	GGAAAATTAC	TTTTTAGAGT	CCATCAGCAT	TTTCATCGTG	AAATGGTTAA	AAACATTTTG
M3	GGAAAATTAC	TTTTTAGAGT	CCATCAGCAT	TTTCATCGTG	AAATGGTTAA	AAACATTTTG
M4	GGAAAATTAC	TTTTTAGAGT	CCATCAGCAT	TTTCATCGTG	AAATGGTTAA	AAACATTTTG
M5	GGAAAATTAC	TTTTTAGAGT	CCATCAGCAT	TTTCATCGTG	AAATGGTTAA	AAACATTTTG
WT	AAAGGCATGG	AGCAAGAAGA	AGAACAAGCG	CTGTTAAGAG	CACTAAAAAA	TCTTCATGAT
M1	AAAGGCATGG	AGCAAGAAGA	AGAACAAGCG	CTGTTAAGAG	CACTAAAAAA	TCTTCATGAT
M2	AAAGGCATGG	AGCAAGAAGA	AGAACAAGCG	CTGTTAAGAG	CACTAAAAAA	TCTTCATGAT
M3	AAAGGCATGG	AGCAAGAAGA	AGAACAAGCG	CTGTTAAGAG	CACTAAAAAA	TCTTCATGAT
M4	AAAGGCATGG	AGCAAGAAGA	AGAACAAGCG	CTGTTAAGAG	CACTAAAAAA	TCTTCATGAT
M5	AAAGGCATGG	AGCAAGAAGA	AGAACAAGCG	CTGTTAAGAG	CACTAAAAAA	TCTTCATGAT
WT	TTCTTGCAAG	AATACAAATA	A			
M1	TTCTTGCAAG	AATACAAATA	A			
M2	TTCTTGCAAG	AATACAAATA	A			
M3	TTCTTGCAAG	AATACAAATA	A			
M4	TTCTTGCAAG	AATACAAATA	A			
M5	TTCTTGCAAG	AATACAAATA	A			

Figure 12. Alignment of fabT mutations with reference genome sequence. Sequencing reads were aligned to the *E. faecalis* OG1RF genome assembly from Baylor University using the Stampy alignment package. Consensus fabT gene sequences (M1-M5) observed in fasamycin resistant mutants are shown. Mutations and stop codons are highlighted in red. (Adapted from Feng, Chakaraboty *et al.* 2012)

observation, we hypothesized that if the primary *in vivo* antibacterial target of the fasamycins is a FASII biosynthetic enzyme, constitutive overexpression of the FASII gene cluster could lead to an increased MIC for Fasamycin A.

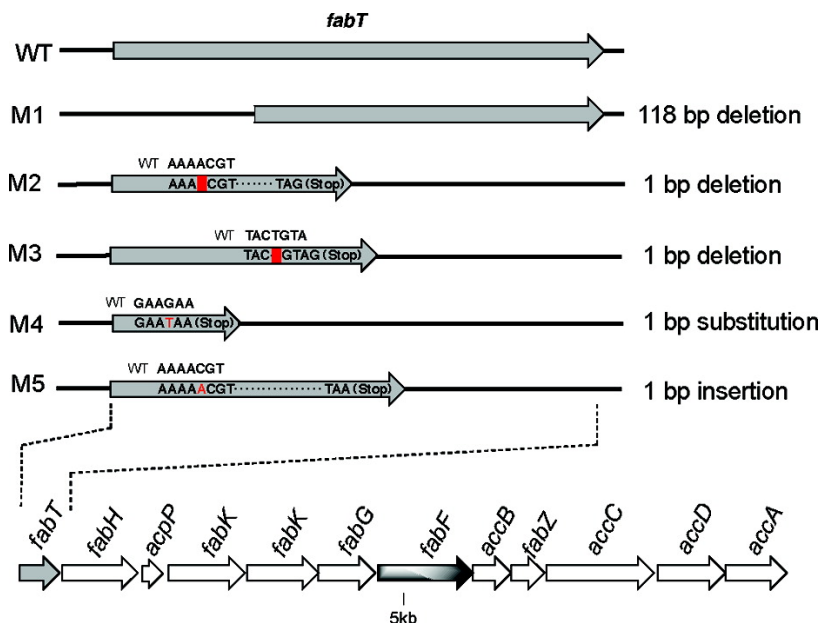


Figure 13. Genomic DNA from nine unique resistant strains were isolated and sequenced. Nine sequenced mutant strains contained five different mutations. All mutations were in the *fabT* gene and are predicted to result in truncated gene product. *FabT* is a transcriptional repressor that regulates the expression of the FASII gene cluster in *E. faecalis*. The five unique *fabT* mutations (M1–M5) that we observed are shown. (Adapted from Feng, Chakraborty *et al.* 2012)

To test our hypothesis, the fasamycins were tested in an *in vitro* fatty acid elongation assay. This assay employed crude FASII extract that contains all of the enzymes required for fatty acid elongation, long-chain acyl CoA (lauroyl-CoA) starter units, and [¹⁴C]-malonyl CoA that together are used to extend fatty acids on purified holo-ACP substrates. ACP-linked long-chain fatty acids are then captured using a phospholipid Flashplate, and ¹⁴C incorporation into newly generated fatty acids is read by a scintillation counter. In elongation assays with *S. aureus* FASII extract and recombinant *S. aureus* 6-His-ACP (6 His-SA-ACP), Fasamycins A and B both inhibit fatty acid elongation (IC₅₀ = 50 and 80 μg/mL, respectively) and show inhibition curves

similar to those of other known FASII inhibitors, BABX and Cerulinn (Figure 14).

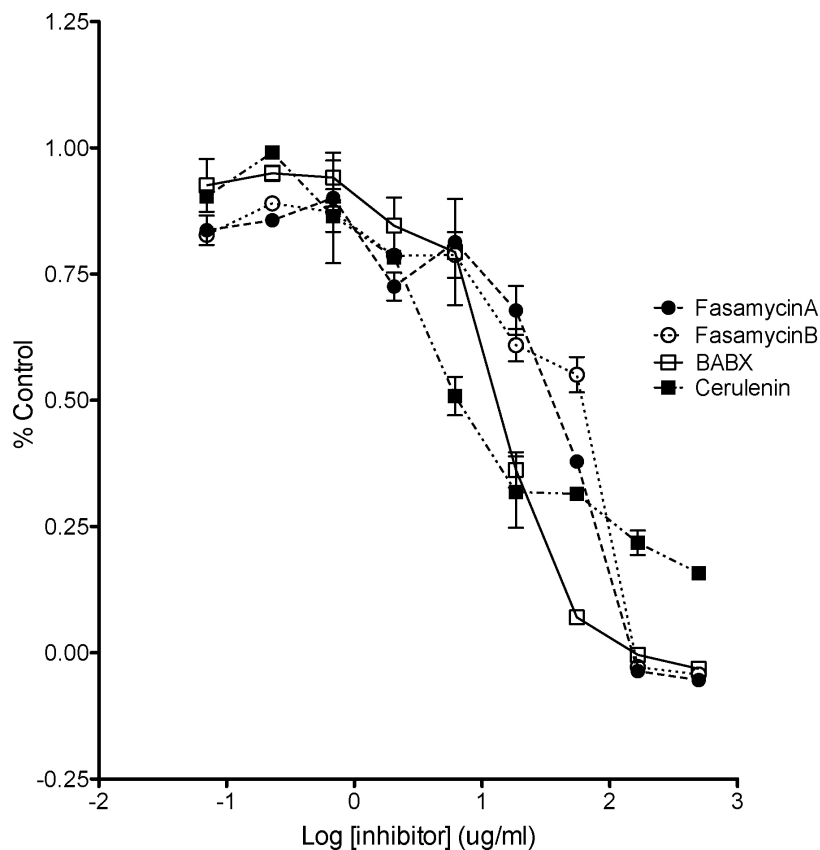


Figure 14. Activity of the Fasamycins A and B was tested in an *in vitro* fatty acid elongation assay. Crude FASII extract from *S. aureus* and recombinant *S. aureus* 6-His-ACP (6-His-SA-ACP) were used. Both Fasamycins A and B inhibit fatty acid elongation ($IC_{50} = 50$ and $80 \mu\text{g/mL}$, respectively) and show inhibition curves similar to those of other known FASII inhibitors. Reactions were run with an excess of $[^{14}\text{C}]$ -malonyl-CoA and lauroyl-CoA. Elongation assays were run in duplicate. (Adapted from Feng, Chakraborty *et al.* 2012)

Based on their ability to inhibit FASII elongation pathway in *in vitro* assay, only five FASII enzymes, FabD, F, G, K, and Z, could potentially be candidates for fasamycins' targets (Figure 15).

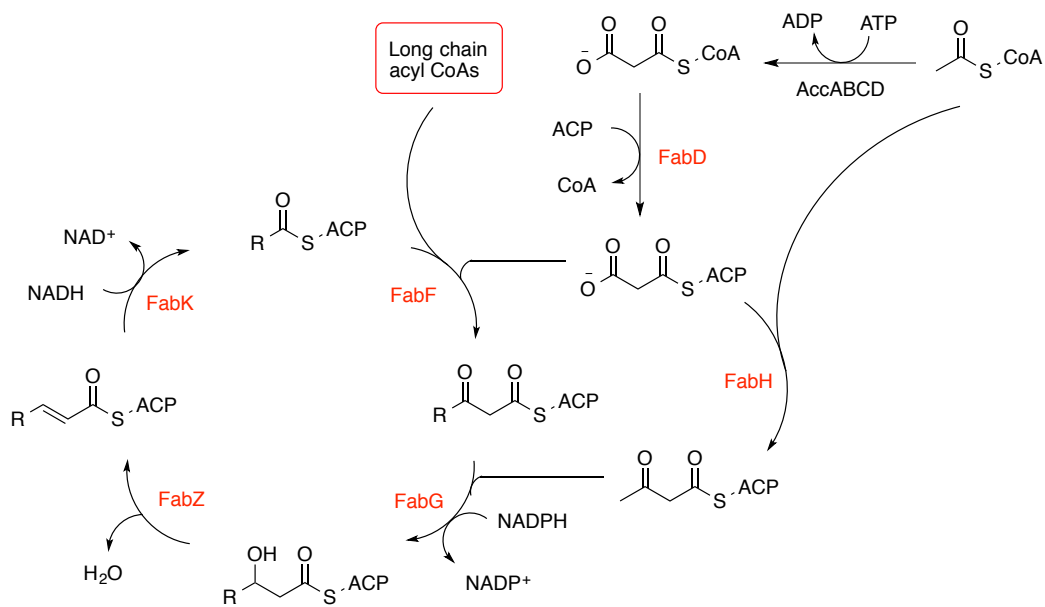


Figure 15. Schematic of FASII biosynthesis. Based on the data obtained from the *in vitro* FASII elongation assay, only five FASII enzymes, FabD, F, G, K, and Z, appear to be potential candidates as fasamycins' target.

To distinguish the inhibition of FabD or FabF from the reductive enzymes of the elongation cycle (FabG, Z, and K), FASII elongation assays were subjected to urea polyacrylamide gel electrophoresis. With urea gels it is possible to resolve malonyl-ACP, the product of FabD, from the longer chain acyl-ACPs generated in the FASII elongation cycle. It is apparent in Figure 16 that Fasamycin A does not inhibit FabD-dependent formation of malonyl-ACP (Figure 16, upper band) but does inhibit the formation of long-chain acyl-ACPs from the FabD product (Figure 16, lower bands).

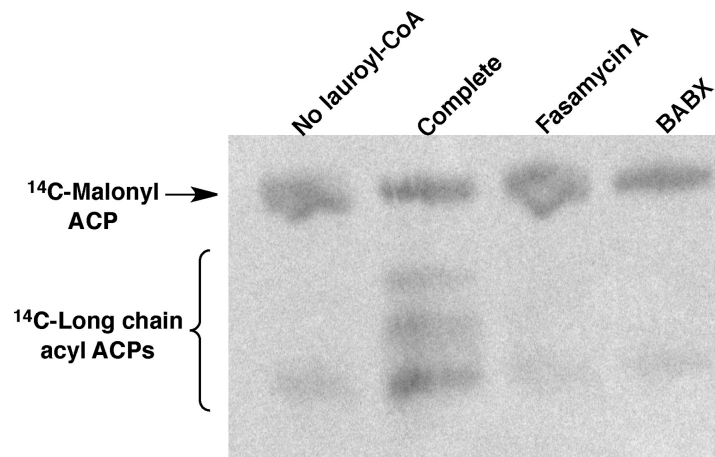


Figure 16. *In vitro* fatty acid elongation assays use a long-chain acyl CoA starter and the incorporation of [^{14}C]-malonyl-CoA to measure FASII activity. The reactions were run using lauroyl-CoA, recombinant 6His-SA-ACP, and crude *S. aureus* FASII extracts. Fasamycin A inhibits fatty acid elongation but does not inhibit FabD-dependent formation of malonyl-ACP. Reactions were run on a 16% polyacrylamide gel containing 4 M urea to resolve ^{14}C -labeled malonyl-ACP (upper band) from ^{14}C -labeled long-chain ACPs (lower bands). Lane 1, control reaction without lauroyl-CoA; lane 2, complete reaction; lane 3, 200 $\mu\text{g}/\text{mL}$ fasamycin A; lane 4, 200 $\mu\text{g}/\text{mL}$ BABX. (Adapted from Feng, Chakraborty *et al.* 2012)

The absence of any labeled long-chain acyl-ACPs in the Fasamycin A-containing reaction indicates that Fasamycin A inhibits FabF, the initial condensation step of the elongation cycle, and not simply a reductive step (FabG, K, and Z) in this cycle. FabF is one of two FASII-specific condensation enzymes. It is specifically used in the fatty acid chain elongation cycle, while the second condensing enzyme, FabH, is used in fatty acid initiation (Figure 15). Our biochemical studies do not rule out the possibility that the fasamycins might also inhibit FabH.

Both gene knockdown and overexpression strategies have been used to identify the protein targets of other FASII inhibitors (Wang *et al.* 2007). For this study, candidate gene overexpression was used to further confirm the specific target of fasamycins. In this strategy, each FASII gene was individually cloned, overexpressed and assayed for the ability to confer Fasamycin A resistance to *E. faecalis*. FASII genes were PCR-amplified from wild-type *E.*

faecalis genomic DNA and cloned into the *E. coli*–*E. faecalis* shuttle expression vector, pMGS100 (Fujimoto and Ike, 2001). Individual FASII gene expression constructs were then checked for the ability to confer Fasamycin A tolerance to wild-type *E. faecalis*. As would be expected for a FabF specific antibiotic, only *E. faecalis* transformed with the FabF overexpression construct grew robustly in the presence of lethal concentrations of Fasamycin A (Figure 17a).

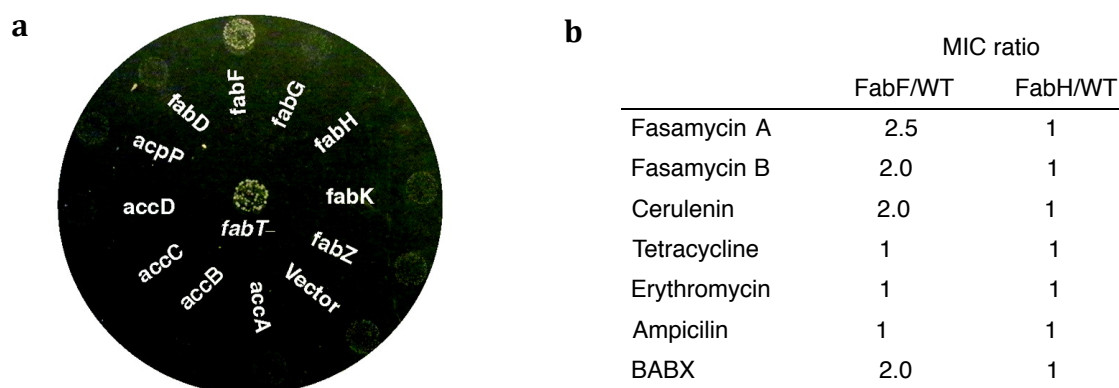


Figure 17. a) Candidate gene over expression assay. Individual FASII gene expression constructs were examined for the ability to confer Fasamycin A tolerance to wild-type *E. faecalis*. Only *E. faecalis* transformed with the FabF overexpression construct grew robustly in the presence of lethal concentrations of Fasamycin A. A FabT knockout mutant with the M1 mutation and wild-type *E. faecalis* transformed with either a vector control or individual FASII overexpression constructs were inoculated onto LB agar containing 5 µg/mL Fasamycin A. b) The MIC ratios for antibiotics measured against wild-type *E. faecalis* and *E. faecalis* transformed with either the fabF or fabH overexpression constructs are listed. (Adapted from Feng, Chakraborty *et al.* 2012)

To further corroborate our results, we used Cerulenin, a well-characterized FabF inhibitor as a control. Cerulenin showed a similar MIC profile against these same FASII overexpression strains, while antibiotics with different protein targets did not exhibit elevated MICs against FASII overexpression strains (Figure 17b).

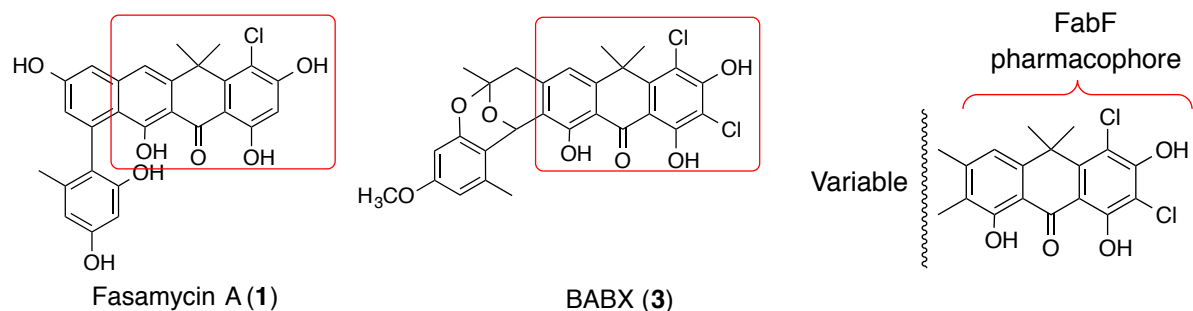


Figure 18. The gem-dimethyl-anthracenone substructure that is common to both the fasamycins and BABX appears to represent a natural FabF-specific antibacterial pharmacophore.

The fasamycins and BABX (3) share a chloro-gem-dimethyl-anthracenone substructure (Figure 18). BABX is a known FabF inhibitor (Kodali *et al.* 2005, Herath *et al.* 2005). BABX and Fasamycin A also show very similar *in vitro* FASII and FabF inhibitory activities (Figures 14 and 16). As with Fasamycin A, BABX shows reduced activity against the FabF overexpression strain but not against any other FASII gene overexpression strains (Figure 17b).

2.3 Discussion

In most bacteria, fatty acid biosynthesis is catalyzed by a group of highly conserved proteins known as Fatty acid synthesis type II (FASII) system enzymes. FASII system enzymes are essential for bacterial membrane lipid biosynthesis and represent increasingly promising targets for the discovery of antibacterial agents with new mechanisms of action (Zhang *et al.* 2006, Wright and Reynolds 2007). Compounds that target the bacterial FAS pathway are not toxic against humans because of radical differences in bacterial and human protein structures. Additionally, FAS is vital to cell physiology and aids in the design of highly effective genus-specific inhibitors against major pathogens (Parsons and Rock 2011). Therefore, FAS systems are one of the most attractive biochemical pathways to be used as targets for new antibacterial

agents. So far, there have been only a handful of reports regarding FASII inhibitors as therapeutic agents. Only two FASII inhibitors, Triclosan and Isoniazid, both of which target the enoyl-acyl carrier protein reductase FabI, are used commercially as antibiotics. FabF-specific antibiotics like the fasamycins show activity against therapeutically relevant antibiotic resistant bacteria. In the case of the fasamycins and BABX, the high convergence to the same pharmacophore on at least two separate occasions, suggests this may provide a promising lead scaffold for future drug development against resistant strains. Optimizing the chloro-gem-dimethyl-anthracenone pharmacophore and the variable region could yield additional FabF-specific antibiotics with increased efficacy and potentially differing spectra of activity. In fact, the specific chlorination pattern on the gem-dimethylanthracenone substructure is clearly important for the fasamycins' activity against *E. faecalis*, as the monohalogenated analogue **1** is almost an order of magnitude more active against this pathogen than the dihalogenated derivative **2**. The study presented in this chapter leads us to believe that we need to expand upon both available and yet unknown scaffolds to identify new FASII inhibitors that may have antibacterial potential against resistant bacterial strains.

2.4 Materials and Methods

2.4.1. Isolation of Fasamycins A (1) and B (2)

Streptomyces albus transformed with the eDNA-derived cosmid cosAZ154 was grown in ISP4 medium containing 5% HP-20 resin at 30 °C (200 rpm) for 15 days. The HP-20 resin was collected, washed with water, and then soaked with methanol. The methanol eluent was fractionated on a silica gel flash-column using a CHCl₃:MeOH (0.1% acetic acid) step gradient. Compounds **1** and **2** eluted from this column with the 95:5 CHCl₃:MeOH fraction. Each compound was then purified by reversed-phase (XBridge C18, 10 X 150 mm, 5 μm) HPLC (7 mL/ min) using a linear gradient from 10:90 H₂O:MeOH (containing 0.1% formic acid) to 100% MeOH (containing 0.1% formic acid) over 30 min. Fasamycins A and B are produced at approximately 0.5 mg/L (Feng, Kallifidas *et al.* 2011). BABX was purchased from Santa Cruz Biotechnology.

2.4.2. Selection of *Enterococcus faecalis* strains resistant to Fasamycin A

E. faecalis OG1RF was grown overnight in Lysogeny Broth (LB) (Bertani 2004) at 37 °C, and 100 μL of the overnight culture was spread on a LB agar plate containing 5 μg/mL Fasamycin A. The plate was incubated at 37 °C. Resistant colonies appeared after 18 h. Nine resistant colonies were picked and struck on selection plates, and then cultures of the mutants grown in the absence of fasamycin were used for sequencing. Selection and analysis of resistant mutants were performed by Dr. Z. Feng.

2.4.3. Genome sequencing and analysis methods

Library Preparation. Genomic DNA libraries were prepared using Illumina TruSeq library kits in accordance with the manufacturer's protocols. Briefly, genomic DNA was sheared using a Covaris S2 ultrasonicator, the resultant dsDNA fragments were end-repaired and A-tailed, and indexed adapters were ligated to the end-repaired DNA. PCR was performed for 10 cycles, and the libraries were quantitated using an Agilent Bioanalyzer 2100 and a High-Sensitivity DNA kit. Indexed adapters were used to allow for the multiplexing of samples at the sequencing stage.

Sequencing and Raw Data Analysis. Samples were sequenced on an Illumina HiSeq2000 for 101 cycles, with an additional 7 cycles for the indexing read. Sequencing was performed according to manufacturer's protocols using TruSeq chemistry. Image data were analyzed in real time by the onboard RTA software package. Bcl files produced by RTA were converted to qseq files by Illumina's OLB1 (Satoshi *et al.* 2009). Software package, and the qseq files were then converted to a fastq file for subsequent analysis.

Alignments and Variant Detection. Sequencing reads were aligned to the *E. faecalis* OG1RF genome assembly from Baylor University using the Stampy alignment package. The BWA aligner within Stampy was used to decrease processing time (Bourgogne *et al.* 2008, Albers *et al.* 2011). BWA and Stampy allow for gapped alignment to facilitate the alignment of sequencing reads in situations where there is significant heterogeneity due to sequence insertions, deletions, or other lesions with regard to the reference genome. The alignment files were processed by Dindel and the Samtools mpileup program to yield potential variant calls consisting of single nucleotide polymorphisms, insertions, and deletions (Lunter and Goodson, 2011, Li *et al.* 2009).

Confirmation of fabT Mutations. Sanger sequencing of *fabT* genes PCR amplified from resistant strains was used to verify each *fabT* mutation detected in the Illumina sequencing experiment (Supporting Information). Colony PCR was performed using whole cells, *fabT* flanking primers (*fabTF* TGCTTATTTACGATATAGTTTGTGCTTATTTACGATATAGTTTG, *fabTR* TGAAGATTTTTTAGTGCTC), Phusion High-Fidelity DNA polymerase (New England Biolabs), and the following cycling parameters: denaturation (98 °C, 2 min), 10 touchdown cycles (98 °C, 20 s; 55 °C (dt -1 °C/cycle), 30 s; 72 °C, 30 s), 30 standard cycles (98 °C, 20 s; 45 °C, 30 s; 72 °C, 30 s), and a final extension step (72 °C, 7 min). Gel-purified amplicons were sequenced from both ends using the PCR primers.

2.4.4. Cloning FASII genes from *E. faecalis* OG1RF

Each individual gene in the FASII gene cluster of *E. faecalis* was amplified by PCR using *E. faecalis* OG1RF genomic DNA as a template and Phusion High-Fidelity DNA polymerase. The primers used are listed in Table 2. PCR amplicons of the correct size were double-digested with *EagI* (or *NotI*) and *NruI* (or *ScaI*) and ligated with the pMGS100 *E. faecalis* expression vector that had also been double-digested with *EagI* and *NruI* (Fujimoto and Ike 2001). Each of the constructs was transferred into *E. faecalis* by electroporation transformation (Fujimoto, *et al.*, 1991).

2.4.5. Bioactivity: Growth on elevated concentrations of Fasamycin A

Each *E. faecalis* strains harboring individual FASII biosynthetic gene expression construct was grown overnight at 37 °C in LB containing 12.5 µg/ mL of chloramphenicol. Overnight cultures were diluted 104-fold and 5 µL aliquots of each diluted culture were then inoculated onto a LB

agar plate containing 12.5 µg/mL of chloramphenicol and 5 µg/mL of fasamycin A. Plates were incubated at 37 °C for 18 h, and examined for spots of growth or no growth. Wild-type *E. faecalis* OG1RF transformed with the empty pMGS100 expression vector was used as a negative control. Positive control was a strain containing the M1 mutation, transformed with the empty pMGS100 expression vector.

2.4.6. Bioactivity: *Ex vivo* studies

In these studies, *E. faecalis* strains harboring FASII gene expression constructs were examined for resistance to Fasamycin A as well as other antibiotics. Wild-type *E. faecalis* transformed with empty pMGS100 was used as control. Overnight cultures grown in LB with 12.5 µg/mL chloramphenicol were diluted 106-fold. 50 µL aliquot of the dilute cultures were added to each well of a 96-well microtiter plate. Fasamycin A and other compounds were dissolved in methanol at 5 mg/mL concentration and diluted 100-fold with LB medium. 150 µL of this solution was added to the first well of the microtiter plate and then serially diluted across the plate. The plates were incubated at 30 °C for 24–48 h. The lowest concentration at which no bacterial growth was observed after 24 h are reported as minimum inhibitory concentrations (MICs). MICs against other microorganisms were determined in the same manner. Yeast extract -peptone-dextrose (YPD) broth was used as media to conduct MIC studies for yeast.

Table 2. Primers Used for Cloning *E. faecalis* FASII Genes

Name	Sequence ^a
FabH_For	GCGCCGGCCGATGAAGAATTATGCACGAATT
FabH_Rev	GCGCAGTACTTTTTACAGCGTTAGGAGCAG
acpP_For	GCGCCGGCCGATGGTATTTGAAAAAATTCAA
acpP_Rev	GCGCTCGCGACAGCCCCCTATTTTCATGTAT
FabK_For	GCGCCGGCCGATGAAGTGTACTTATCTTAGA
FabK_Rev	GCGCTCGCGATCACTTAGCCCCAACGCTGAT
FabD_For	GCGCCGGCCGATGAGGTGTCGTATGAAAACA
FabD_Rev	GCGCTCGCGACCATTTTTTACCTCCCAGTAAG
FabG_For	GCGCCGGCCGATGGAATTAACAGGGAAAAACG
FabG_Rev	GCGCTCGCGACCTTTCGTTTATCCGTGCATG
FabF_For	GCGCCGGCCGATGAATCGAGTAGTTATTACCGG
FabF_Rev	GCGCTCGCGAACTGCATGTTAATCCTCCCAG
accB_For	GCGCCGGCCGATGCAGTTAGAAGAAGTAAAAGC
accB_Rev	GCGCTCGCGAGTTAATTTTCATGTTTTATTCTCC
FabZ_For	GCGCCGGCCGATGAAATTAACAATTACAGAAATTC
FabZ_Rev	GCGCAGTACTTCGAAAACATTTTTTCACCTATCC
accC_For	GCGCGCGGCCGCATGTTTTCGAAAGTATTAATCGC
accC_Rev	GCGCAGTACTCTTGCGGCGTTCTTTCTTTAATG
accD_For	GCGCGCGGCCGCATGGCATTATTTAAAAAGAAA
accD_Rev	GCGCTCGCGATTAGCGCCACCCTTCCAA
accA_For	GCGCCGGCCGATGGAAAAGAAAACAGCCAATGA
accA_Rev	GCGCTCGCGAGAGTAGTAACTTCTAATATTTGCG

^aRestriction sites added to each primer are underlined.

2.4.7. Bioactivity: *In Vitro* studies

Acyl Carrier Protein (ACP) Cloning. In order to clone Acyl Carrier Protein (ACP), the ACP gene from *S. aureus* was amplified from *S. aureus* genomic DNA using the following primers and PCR conditions. Primers: forward StaphACPFHis, GCGCGGATCCGATGGAAAATTTTCGATAAAGTAA; reverse StaphACPR, GCGCAAGCTTATTTTTCAAGACTGTTAATA (restriction sites added for cloning are shown

in bold). Cycling parameters: denaturation (98 °C, 2 min), 35 cycles (98 °C, 20 s; 55 °C, 30 s; 72 °C, 30 s), and a final extension step (72 °C, 7 min). 50 µL reaction conditions: 10 µL of 5X Phusion HF buffer, 1.5 µL of DMSO, 0.3 µL of each 100 mM oligonucleotide primer, 1 µL of 10 mM dNTPs mix, 1 unit of Phusion High-Fidelity DNA polymerase and water as needed. Amplicons were doubly digested with BamHI/HindIII, ligated into BamHI/HindIII digested pColaDuet-1 (Novagen), and then transformed into *E. coli* BAP1 cells to yield pColaDuet: 6His-SA-ACP. BAP1 is a BL21(DE3)-derived expression strain that contains an inducible phosphopantethienyl transferase (*sfp* gene), which should assist with generating the desired holo-ACP (Pfeifer, *et al.*, 2001). Cloning was performed by Dr. Z. Zheng.

Expression and Purification of 6His-SA-ACP. 1 L LB media was inoculated with overnight culture of BAP1/pColaDuet:6His-SA-ACP (1:1000 dilution). It was then grown at 37 °C to an OD₆₀₀ of 0.6. The incubation temperature was then reduced to 20 °C, and after 1 h, IPTG (0.5 mM) was added to induce protein expression. After an additional 16 h at 20 °C, the cells were collected by centrifugation (3200g, 15 min). The cell-pellet was resuspended in 40 mL of lysis buffer (50 mM Tris, pH 7.5, 0.3 M NaCl, 10 mM imidazole), and the cells were lysed by sonication. Crude cell lysates were centrifuged at 25000g for 30 min and the supernatant was incubated at 4 °C for 15 min with 1 mL of Ni-NTA resin (Qiagen). This mixture was washed with 20 mL of lysis buffer and then 20 mL of lysis buffer containing 50 mM imidazole. 6His-SA-ACP was subsequently eluted with 5 mL of lysis buffer containing 125 mM imidazole. 6His-SA-ACP was further purified by anion-exchange chromatography (Mono Q 10/100 GL) using a linear gradient from 0.1 M NaCl (10 mM Tris, pH 7.5) to 1 M NaCl (10 mM Tris, pH 7.5) over 120 min. Protein concentrations were determined with the Bradford assay. Aliquots of protein

were flash-frozen in liquid nitrogen and stored at $-80\text{ }^{\circ}\text{C}$.

FASII Enzyme Extract from S. aureus. For the preparation of FASII enzymes (Kodali *et al.*, 2005) 6 L of *S. aureus* was grown in LB media at $37\text{ }^{\circ}\text{C}$ until to stationary phase is reached. Cells were then harvested by centrifugation (2500g, 10 min, $4\text{ }^{\circ}\text{C}$) and the cell pellet washed twice with ice-cold FASII buffer (0.1 M sodium phosphate, pH 7, 1 mM EDTA, 5 mM β -mercaptoethanol). The pellet resuspended in 0.5 L of the same buffer and lysed using a French press. Cell debris was separated by centrifugation (20000 rpm, 15 min, $4\text{ }^{\circ}\text{C}$) and the supernatant was collected. In order to precipitate the protein, ammonium sulfate was slowly added to the supernatant to reach 45% saturation. Precipitated protein was removed by centrifugation (10000 rpm, 5 min, $4\text{ }^{\circ}\text{C}$), and ammonium sulfate was slowly added to the supernatant to reach 80% saturation. Precipitated protein was collected by centrifugation (10000 rpm, 5 min, $4\text{ }^{\circ}\text{C}$). The resulting pellet was resuspended in 20 mL of FASII buffer and dialyzed (10 kDa MW cutoff) at $4\text{ }^{\circ}\text{C}$. Buffer was changed four times during the process. This crude extract from *S. aureus* should contain all the factors necessary for reconstituting the FASII elongation process *in vitro*. Protein concentration measurements were done using the Bradford assay. Aliquots of protein were stored at $-80\text{ }^{\circ}\text{C}$ after flash freezing in liquid nitrogen.

FASII Elongation Assay. 96-well phospholipid Flashplates (Perkin-Elmer) were used to perform FASII inhibition assays (Kodali *et al.*, 2005). The buffer used in this assay was composed of 100 mM sodium phosphate (pH 7), 5 mM EDTA, 1 mM NADPH, 1 mM NADH, 150 μM DTT, 5 mM β -mercaptoethanol, 20 μM lauroyl-CoA, 4% Me_2SO , and 5 μM DTT pretreated 6xHis-SA-ACP. 0.4 μg of the FASII protein extract was added to 40 μL of this buffer and incubated with

inhibitor at room temperature for 20 min. Elongation reactions were initiated with the addition of 8 μ L of C-2 labeled [14 C]-malonyl-CoA (Perkin-Elmer 57.5 mCi/mmol) dissolved in water to a final concentration of 4 μ M. After 60 min at 37 $^{\circ}$ C, reactions were quenched by adding 80 μ L of 14% perchloric acid. The 96-well phospholipid Flashplates were then sealed and incubated at room temperature overnight. The reactions were assessed using a TopCount scintillation counter (Trilux Microbeta).

FASII Elongation and ACP Loading Assay. For this assay, 0.2 μ g of FASII extract was added to 50 μ L of the buffer mentioned above and reactions were run in microcentrifuge tubes. After a preincubation with inhibitor at room temperature for 20 min, fatty acid elongation was initiated by adding a 10 μ L aliquot of C-2 labeled [14 C]-malonyl-CoA (Perkin-Elmer 57.5 mCi/mmol, final concentration 4 μ M). After 30 min of incubation at 37 $^{\circ}$ C, 15 μ L of native sample buffer was added to each reaction. 16% polyacrylamide gels containing 4 M urea were used to resolve the reaction mixtures. Gels were blotted on to polyvinylidene difluoride membranes (Bio-Rad) and imaged using a PhosphorImager.

CHAPTER 3

Homology-Based Metagenomic Screening Efforts towards the Identification of Novel Type II Polyketide Synthase Gene Clusters

3.1 Introduction

Given the clinical and commercial success of a diverse collection of aromatic compounds that arise from iterative type II polyketide synthases, considerable interest has developed to gain access to these microbial secondary metabolites (O'Hagan 1991). The traditional protocol used to isolate and identify natural products from bacteria is to first culture an individual bacterium from the environment, extract bacterial cultures using organic solvents, and isolate pure compounds through activity guided fractionation. However, as mentioned earlier, the difficulties associated with culturing environmental bacteria prevent traditional culture-dependent methods from being applied to a significant fraction of the bacterial species found in nature. On the other hand, culture-independent metagenomic strategies involve direct cloning of DNA from environmental samples, which contain the genes encoding for small molecule, bypassing the need for isolation and culture of new bacterial species. PCR based studies as well as shotgun-sequencing efforts indicate that eDNA samples are rich in unique minimal PKS genes (Seow *et al.* 1997, Wawrik *et al.* 2007, Pang *et al.* 2008).

These types of culture-independent techniques are the basis of the work described in this chapter. There have been multiple reports in the literature describing the identification of type II PKS genes from eDNA samples. King, Brady *et al.* expressed a pathway encoding the biosynthesis of a type II polyketide and characterized a new molecule, erdacin (King *et al.* 2009) with previously

unknown pentacyclic skeleton, lending credence to the idea that metagenomics could be successfully implemented to tap into the chemical diversity of type II PKS-derived molecules from uncultured bacteria. Since this report, several other reports have been published detailing the discovery of additional novel type II PKS-derived compounds, such as bioactive pentangular polyphenols, calixanthomycin A, arenimycins C-D, arixanthomycins A-C (Kang and Brady 2014, Kang and Brady 2013), fasamycins (Feng, Kallifidas *et al.* 2011; Feng, Chakraborty *et al.* 2012), fluostatin F (Feng, Kim *et al.* 2010), and MRSA-active antibiotic tetarimycin A (Kallifidas, Kang *et al.* 2012) to mention a few. Taken together, all these reports offer a glimpse of the previously unseen chemistry of the type II PKS derived natural products that metagenomic screening has the potential to uncover.

Aromatic polyketide synthases (type II PKSs) are a family of multienzyme systems that catalyze the biosynthesis of a structurally diverse and pharmaceutically important class of natural products known as aromatic type II polyketides (Hopwood, 1997). Many of these molecules exhibit anticancer (e.g. doxorubicin), antibacterial (e.g. oxytetracycline), antifungal (e.g. pradimicin), antiviral (e.g. A-74528) and other related medicinally important activities (O'Hagan, 1991). These molecules are synthesized, in part, by multifunctional polyketide synthases (PKSs), which catalyze repeated decarboxylative condensations between acyl thioesters (acetyl, propionyl, malonyl, methylmalonyl, etc.). Type II PKS systems comprise of aggregates of small polypeptides, each having a single catalytic activity that is used iteratively during the biosynthesis. All type II PKS gene clusters encode a “minimal PKS” consisting of three functionally discrete proteins: a ketosynthase (KS_{α}), chain length factor (KS_{β}), and acyl carrier protein (ACP). This core of three enzymes is responsible for the construction of the polyketide backbone through iterative condensation of multiple acyl-CoA units (Figure 19). Additional PKS

subunits, ketoreductases, cyclases and aromatases determine the folding pattern of the nascent poly- β -keto intermediate and a set of tailoring enzymes, such as oxygenase, glycosyltransferase, etc. further functionalize the aromatic polyphenols to create the final polyketide natural products (Shen 2000, Rawlings 1999, Hertweck 2009)

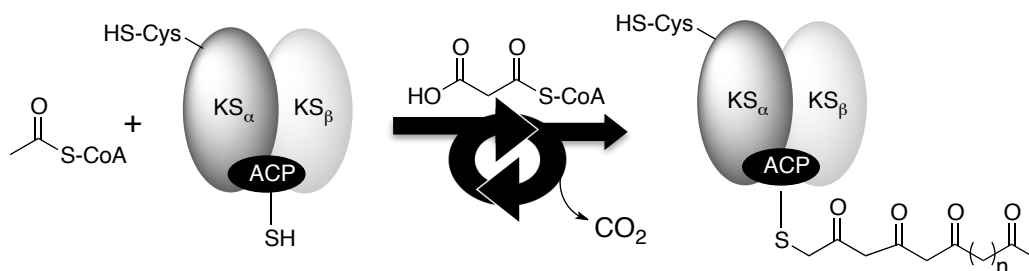


Figure 19. Iterative use of the minimal PKS for the assembly of aromatic type II polyketide backbone. Minimal PKS composed of three proteins: ketosynthase alpha (KS_{α}), ketosynthase beta/chain length factor (KS_{β}), and acyl carrier protein, ACP. The minimal PKS is responsible for the iterative condensation of malonyl CoAs into a nascent polyketide chain that is then cyclized, aromatized, reduced, oxidized, rearranged, and functionalized in pathway-specific ways to generate structural diversity that is known to arise from type II PKS systems

The cloning and analysis of DNA extracted directly from environmental samples (environmental DNA, eDNA) provides a means of exploring the biosynthetic capacity of natural bacterial populations. eDNA libraries are essentially large reservoirs of bacterial genetic diversity from which new secondary metabolite gene clusters can be systematically recovered and studied. In this chapter, I describe my efforts to gain access to the chemical diversity of the aromatic natural products synthesized by bacterial type II polyketide synthases (PKSs) through metagenomic approaches. Through homology-based screening of a large metagenomic library, I identified, for the first time, the gene cluster for the biosynthesis of the known Type II PKS derived angucycline antibiotic seitomycin and related natural products (Abdelfattah *et. al.* 2003).

3.2 Results

3.2.1 Screening for Type II PKS pathways

In our attempt to identify novel type II polyketide natural products, a multimillion membered metagenomic library constructed using sample collected from Texas was screened. The biosynthetic gene-clusters of these natural products contain, among other genes, a set of three highly conserved genes, a ketosynthase (KS_{α}), a chain length factor (CLF or KS_{β}) and an acyl carrier protein (ACP) (Figure 19, 20). For the present work, I used a set

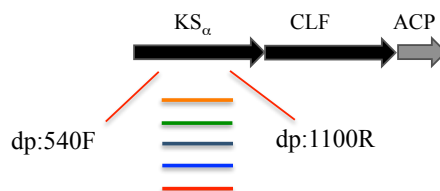


Figure 20. eDNA derived PKS probes: Unique KS_{α} genes were used as probes to recover type II PKS containing clones from the metagenomic libraries.

of degenerate primers targeting conserved regions on the conserved KS_{α} gene to perform homology-based screening of the environmental DNA (eDNA) library. It has been shown previously that KS_{α} genes group into clades that correlate well with the chain length and initial cyclization pattern of the polyketide precursor produced by a gene cluster (Wawrik *et al.* 2005). KS_{α} genes can therefore be used as bioinformatics markers for predicting both the length of the polyketide chain produced by a minimal PKS and the initial polyketide cyclization pattern of the polyketide chain. The primers (dp:540F and dp:1100R, Figure 20) were designed to amplify ~550 bp long DNA sequence from the Texas (Tx) library (Figure 21). In order to characterize the amplicons, BLAST search was performed on individual sequences, followed by a ClustalW-based phylogenetic analysis comparing the eDNA derived KS_{α} sequences to those from functionally characterized or known type II PKS clusters. It's been found that type II PKS clusters with KS_{α} gene sharing >85% identity, tend to encode for molecules with very similar chemical structures (Reddy *et al.* 2012). Clustering KS_{α} sequences at 85% identity revealed 91 unique sequences from the Tx library (Figure 22, 23). A summary of the results of the screen is presented in Figure 21.

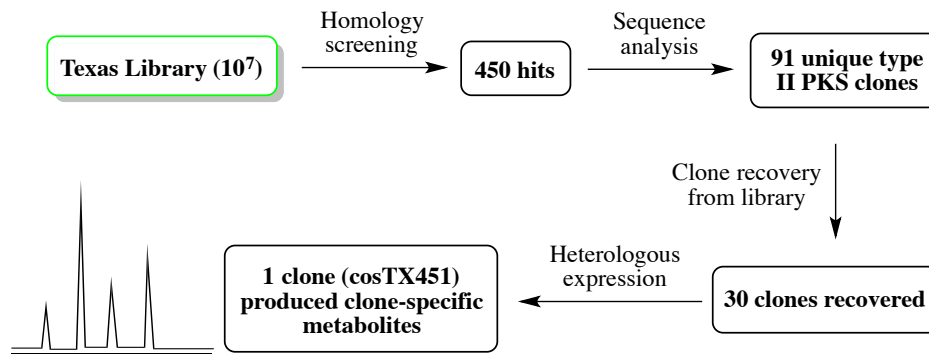


Figure 21. An overview of TX library screening results.

3.2.2 Recovery, retrofitting and heterologous expression of eDNA-derived Type II PKS pathways

Based on the above analyses, forty clones carrying unique KS_{α} sequences were picked for recovery. In order to recover the clones from eDNA library, the unique KS_{α} sequences were used as probes. A second set of specific primers was designed based on the particular amplicon sequence of that KS_{α} gene obtained during our initial screening. The type II PKS cosmid clones of interest were then recovered from library using successive rounds of serial dilution and PCR screening. In total thirty type II PKS gene containing clones were recovered and carried forward for functional analysis (Figure 21, 22).

Although, the metagenomic libraries are constructed in *E. coli*, it is not a desirable host for heterologous expression of small molecules. Instead *Streptomyces albus* was chosen as the heterologous host for our expression studies. The predominantly soil dwelling gram-positive *Streptomyces sp.* is well known for secondary metabolite production and is the source for about two-thirds of clinically useful antibiotic natural products (Kieser, Bibb *et al.* 2000). *Streptomyces sp.* has extensive precedent as a successful heterologous host (Pfeifer and Khosla 2001). Furthermore, nearly all of the identified KS_{α} genes in this work have closest homology with KS_{α} genes from *Streptomyces sp.*

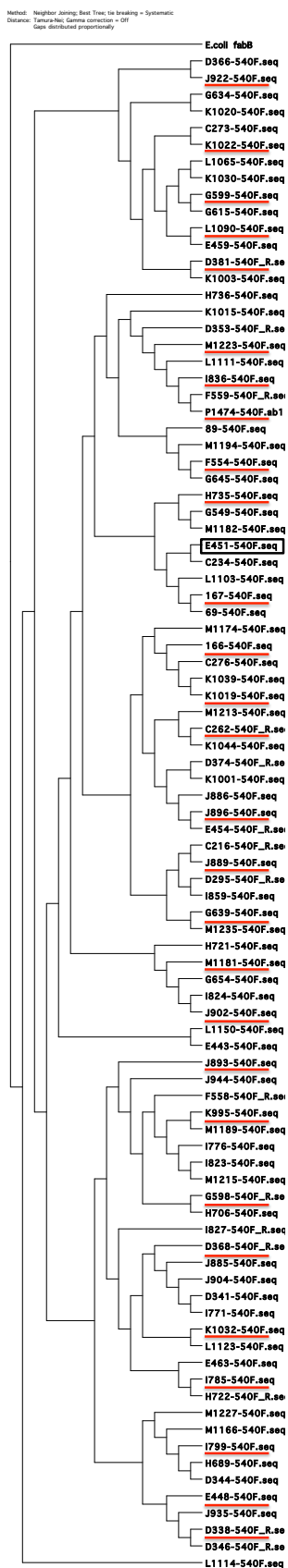


Figure 22. A ClustalW-based phylogenetic tree of the eDNA-derived KS_{α} genes is shown. Thirty clones (underlined) were selected randomly for recovery. The tree was constructed using the Neighbor-joining method. *E. Coli fabB* gene was used as an outgroup.

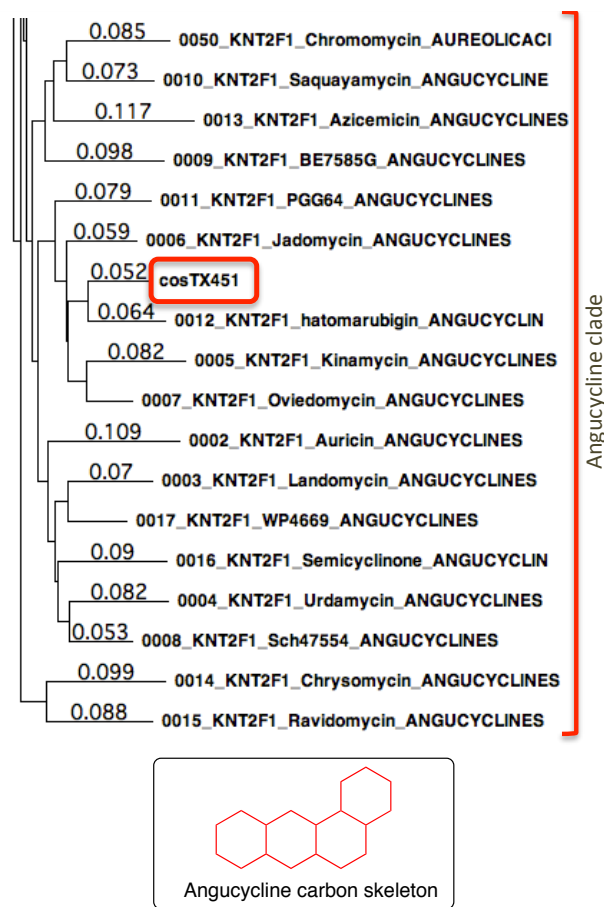


Figure 23. Comparison of the cosTX451 derived KS_{α} amplicon sequence with those from sequenced PKS gene clusters that encode known type II PKS derived secondary metabolites. ClustalW-based phylogenetic tree was constructed using Neighbor-joining method. Shown here is an angucycline carbon skeleton.

Streptomyces albus has proved to be an extremely efficient heterologous host because of its relatively fast growth rate, ample sporulation and ease of conjugation with *E. coli*. These factors helped facilitate both the conjugation of eDNA-derived Type-II PKS gene cluster into *S. albus* and the characterization of the molecules these pathways encode for.

Recovered clones from the Tx library containing type II PKS genes were retrofitted so that they could be conjugated into *S. albus* for heterologous expression studies. Recovered cosmids were digested with *PsiI*, which cuts in the cosmid vector but rarely cuts in eDNA inserts. Each linearized cosmid was then ligated with the 6.81 kb *DraI* fragment from pOJ436 (Feng *et al.* 2011). This fragment contains an origin of transfer (*oriT*) and elements of the *phiC31 attP-attB* integration system needed for integration into diverse *Streptomyces sp.* (Bierman *et al.* 1992). Retrofitted cosmids were transformed into *E. coli S17.1* and conjugated into *S. albus* using standard *Streptomyces* conjugation protocols (Practical *Streptomyces* Genetics by T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater, D. A. Hopwood).

Color production, TLC and HPLC-MS analysis were used to identify clone specific metabolite production in recombinant *Streptomyces* harboring PKS pathways. In the present screen, out of thirty PKS gene clusters from Tx library examined in this study, one clone, cosTX451, conferred the production of clone specific metabolites to *S. albus*, based on visual inspection, TLC and HPLC-MS of crude extracts (Figure 24). A number of clone-specific metabolites were observed upon HPLC-MS analyses of the fermentation broth extracts derived from *S. albus* transformed with cosTX451 (Figure 22). A ClustalW-based phylogenetic analysis using KS_{α} genes from the eDNA derived clone cosTX451 as well as those from functionally characterized type II gene clusters indicated that the pathway cosTX451 was likely to produce angucycline (Rohr *et al.*

1992) family of type II polyketide natural products (Figure 23). The closest homolog to this KS_α gene is from the hatomarubigin pathway from *Streptomyces* sp. strain 2238-SVT4 (Hayakawa *et al.* 1991).

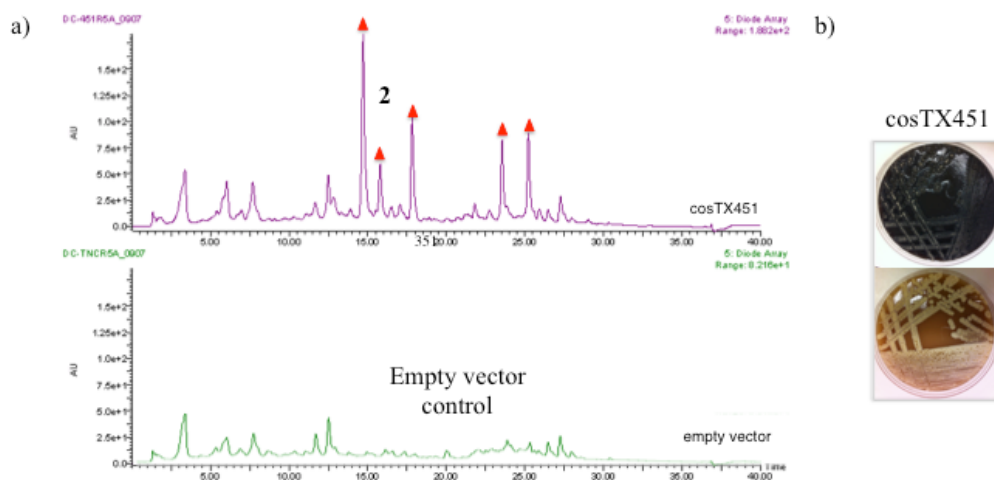
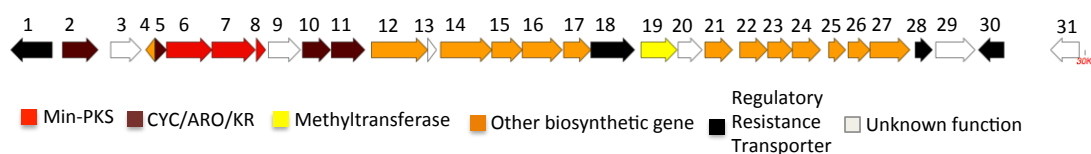


Figure 24. *S. albus* transformed with eDNA-derived PKS clone cosTX451 was screened for the production of clone specific metabolites. a) LC-MS analysis of crude ethyl acetate extract of *Streptomyces* culture broth showed multiple clone-specific peaks (red triangles). Clone-specific metabolite **2** was isolated for characterization. A negative control was carried out using *S. albus* transformed with empty vector. b) Recombinant *S. albus* on MS agar plates showed colored metabolite production for clone cosTX451. Empty vector control did not show strong color phenotype.

3.2.3 Sequencing of PKS containing eDNA-derived clones

In order to reveal the gene sequence and organization, the eDNA insert in cosTX451 clone has been fully sequenced. From the sequencing results, the polyketide biosynthetic pathway appears to be only partially captured in this clone. Two overlapping clones, cosTX578 and cosTX2512, were identified by rescreening the Tx library using primer pairs designed to recognize both the proximal and distal ends of cosTX451. These clones were successfully recovered and sequenced. Sequencing data analysis of cosTX578 revealed no relevant biosynthetic genes. Other overlapping clone, cosTX2512, however contained additional type II PKS biosynthetic genes that could reconstruct the original pathway. The completed sequencing efforts on cosTX451 and cosTX2512 revealed ~30 kb biosynthetic gene cluster containing 31 predicted open reading

frame (ORFs) (Figure 25). These 31 ORFs include the minimal PKS, ketoreductase, cyclase, and aromatase expected in a type II PKS gene cluster. Additionally, a number of post-PKS enzymes, regulatory proteins, a transporter and several hypothetical proteins are also encoded for by this gene cluster. Phylogenetic comparison of the minimal PKS gene sequences revealed they are most closely related to the angucyclines hatomarubigin (Kawasaki *et al.* 2010) and jadomycin (Han *et al.* 1994) pathways (Figure 23).



Gene	Clone	Size	Deduced function of protein	I/S%	Origin	accession no.
<i>orf1</i>	451	1194	MFS transporter	73/68	<i>Streptomyces sp.</i>	WP_009083509.1
<i>orf2</i>	451	993	aldo/keto reductase	82/88	<i>S. bingchenggensis</i>	YP_004958664.1
<i>orf3</i>	451	813	hypothetical protein	57/67	<i>Streptomyces sp.</i>	BAJ07840.1
<i>orf4</i>	451	225	putative oxygenase	41/75	<i>Streptomyces sp.</i>	YP_008012343.1
<i>orf5</i>	451	330	putative polyketide cyclase	82/91	<i>S. ambofaciens</i>	AAR30153.1
<i>orf6</i>	451	1272	putative ketoacyl synthase	83/92	<i>S. ambofaciens</i>	AAR30152.1
<i>orf7</i>	451	1215	Polyketide chain length factor	79/87	<i>S. venezuelae</i>	YP_006881526.1
<i>orf8</i>	451	261	putative acyl carrier protein	71/83	<i>Streptomyces sp.</i>	BAJ07852.1
<i>orf9</i>	451	891	thioesterase	58/70	<i>Streptomyces sp.</i>	YP_004806667.1
<i>orf10</i>	451	786	polyketide ketoreductase	89/95	<i>Streptomyces sp.</i>	AAK57528.1
<i>orf11</i>	451	936	aromatase	77/85	<i>S. antibioticus</i>	CAG14969.1
<i>orf12</i>	451	1572	decarboxylase JadN	84/91	<i>S. venezuelae</i>	AAK01934.1
<i>orf13</i>	451	246	hypothetical protein	53/63	<i>S. violaceusniger</i>	YP_004811904.1
<i>orf14</i>	451	1437	putative FAD-dependent monooxygenase	58/68	<i>S. albaduncus</i>	CBH32083.1
<i>orf15</i>	451	849	bifunctional oxygenase-reductase protein LndM2	53/62	<i>S. globisporus</i>	AAT64432.1
<i>orf16</i>	2512	1113	oxidoreductase Sim8	65/79	<i>S. antibioticus</i>	AID47011.1
<i>orf17</i>	2512	765	ketoacyl reductase	65/74	<i>Streptomyces sp.</i>	WP_007453033.1
<i>orf18</i>	2512	1224	putative transporter	65/76	<i>Streptomyces sp.</i>	BAJ07859.1
<i>orf19</i>	2512	1005	putative O-methyltransferase	77/86	<i>Streptomyces sp.</i>	BAJ07860.1
<i>orf20</i>	2512	672	hypothetical protein	31/46	<i>S. venezuelae</i>	YP_006877087.1
<i>orf21</i>	2512	768	putative N5,N10- methylenete tetrahydromethanopterin reductase	75/85	<i>Streptomyces sp.</i>	BAJ07867.1
<i>orf22</i>	2512	768	oxygenase-reductase	58/76	<i>S. fradiae</i>	AFU51427.1
<i>orf23</i>	2512	687	hypothetical protein	65/82	<i>Streptomyces sp.</i>	BAJ07849.1
<i>orf24</i>	2512	786	putative short-chain dehydrogenase/reductase	71/86	<i>Streptomyces sp.</i>	BAJ07861.1
<i>orf25</i>	2512	474	putative reductase	80/90	<i>Streptomyces sp.</i>	BAJ07862.1
<i>orf26</i>	2512	612	putative oxygenase	61/76	<i>Streptomyces sp.</i>	BAJ07850.1
<i>orf27</i>	2512	1116	putative hydroxylase	38/52	<i>S. vietnamensis</i>	ADO32775.1
<i>orf28</i>	2512	483	MarR family transcriptional regulator	65/74	<i>S. viridochromogenes</i>	WP_003990963.1
<i>orf29</i>	2512	1098	FAD-dependent pyridine	52/64	<i>Saccharopolyspora erythraea</i>	YP_001102692.1
<i>orf30</i>	2512	714	Phosphate regulon transcriptional regulator	69/79	<i>S. venezuelae</i>	YP_006881518.1
<i>orf31</i>	2512	1353	hypothetical protein	54/86	<i>Amycolatopsis orientalis</i>	YP_008012343.1

Figure 25. ORF diagram and gene table for TX451/2512 pathway. ORFs were predicted using MetaGeneMark and predicted ORFs were annotated by Blast search and Pfam domain analysis. MacVector was used for sequence manipulation.

3.2.4 Overlapping-clone recovery and transformation-associated recombination (TAR) of eDNA-derived Type II PKS pathways

In the present screen, I identified one clone (cosTX451) that conferred the production of clone specific metabolites to *S. albus*. A number of clone-specific metabolites were observed upon HPLC-MS analyses of the fermentation broth extracts derived from *S. albus* transformed with cosTX451.

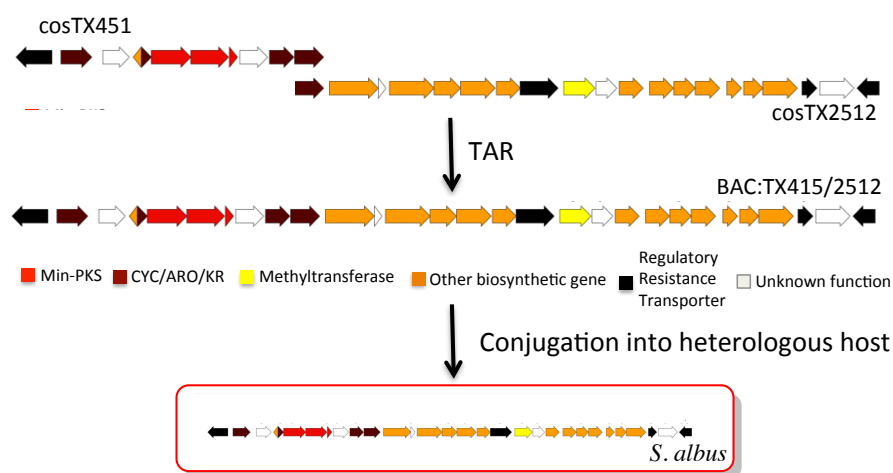


Figure 26. TAR cloning of two overlapping eDNA clones carrying the type II PKS pathway BAC:TX451/2512. The TAR reassembled construct was shuttled into *S. albus* for heterologous expression experiments.

Full sequencing of the eDNA insert in cosTX451 revealed that the biosynthetic pathway is only partially captured in this cosmid. In order to gain access to the complete pathway, one overlapping clones, cosTX2512, was identified and recovered by rescreening the Tx library using primer pairs designed to recognize both the proximal and distal ends of cosTX451 (Figure 26). Sequence analysis of cosTX2512 suggested that this clone captures a large part of the pathway and the reassembled pathway was likely to produce fully functionalized secondary metabolites. This sequencing data further corroborated our conclusion that the current type II

PKS gene-cluster is novel and had not been reported in literature previously. Therefore, cosTX451 was assembled with the overlapping clone cosTX2512 into a complete pathway through Transformation Associated Recombination (TAR) in *Saccharomyces cerevisiae* (Feng, Kim *et al.* 2010; Kim *et al.* 2010) into a larger bacterial artificial chromosome (BAC) clone (Figure 26). The construct BAC:TX451/2512 was confirmed with sequencing and conjugated into *S. albus* for heterologous expression (Figure 26).

3.2.5 Isolation and structural characterization of eDNA encoded natural products: Seitomycin, Rabelomycin, 8-Methylrabelomycin and 1-Deoxo-1-hydroxy-8-O-methylrabelomycin

To learn more about the molecules encoded by eDNA-derived Type-II PKS gene cluster, the major clone specific metabolites of the recombinant *S. albus* BAC:TX451/2512 clone was isolated and structurally characterized. The secondary metabolite profile obtained from heterologous expression of the TAR construct (BAC:TX451/2512) was different from what was observed for cosTX451 alone (Figure 27). Attempts were made to isolate and structurally characterize metabolites from the truncated cosTX451 pathway. However, only one metabolite (**4**) was successfully isolated from this clone. The rest of the compounds appeared to be unstable and attempts to isolate them were unsuccessful. Efforts were focused on the isolation and structural elucidation of the clone-specific metabolites produced by the fully assembled pathway. Cultures of *S. albus* transformed with BAC:TX451/2512 were grown in R5A media. After incubating at 30 °C for 10 days, the cultures were extracted with ethyl acetate. This crude extract was partitioned by reversed phase HPLC to isolate the major metabolites. An extensive 1- and 2-D NMR analysis utilizing ¹HNMR, ¹³CNMR, COSY, HMBC and HMQC were employed to deduce a potential structure of the metabolites **1-4**.

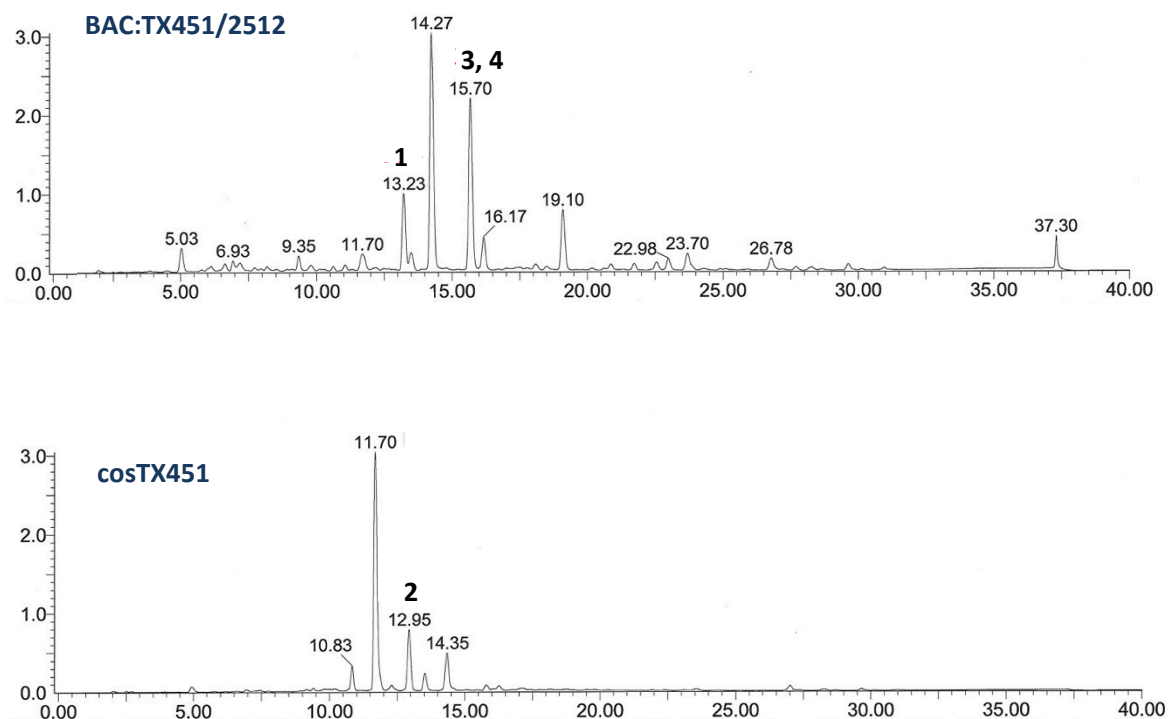


Figure 27. HPLC analysis of crude ethyl acetate extracts of *S. albus* harboring BAC:TC451/2512 and cosTX451 clones. The secondary metabolite profile obtained from heterologous expression of the TAR construct (BAC:TX451/2512) was different from what was observed for cosTX451 alone. Compounds **1-4** were isolated for further analysis.

Characterization of seitomycin (1)

The ^1H NMR spectrum showed the presence of H-bonded OH group at δ 12.3, and two pairs of aromatic protons at δ 8.0/7.6 (5-H, 6-H) and 7.64/7.4 (9-H, 10-H) that appeared as doublets with coupling constants 8.0 and 9.4 Hz, respectively. In the aliphatic region, two singlets at 3.9 (8-OCH₃) and 1.3 (3-CH₃) each for three hydrogens indicated the presence of methoxy and methyl groups. Presence of two AB at 2.8 and 3.0 (4-H) and ABX at 2.0 and 1.8 (2-H) signals, was due to two methylene groups. A triplet at 5.8 (1-H) suggests the presence of one H attached to an oxygen-attached carbon and next to the ABX methylene group.

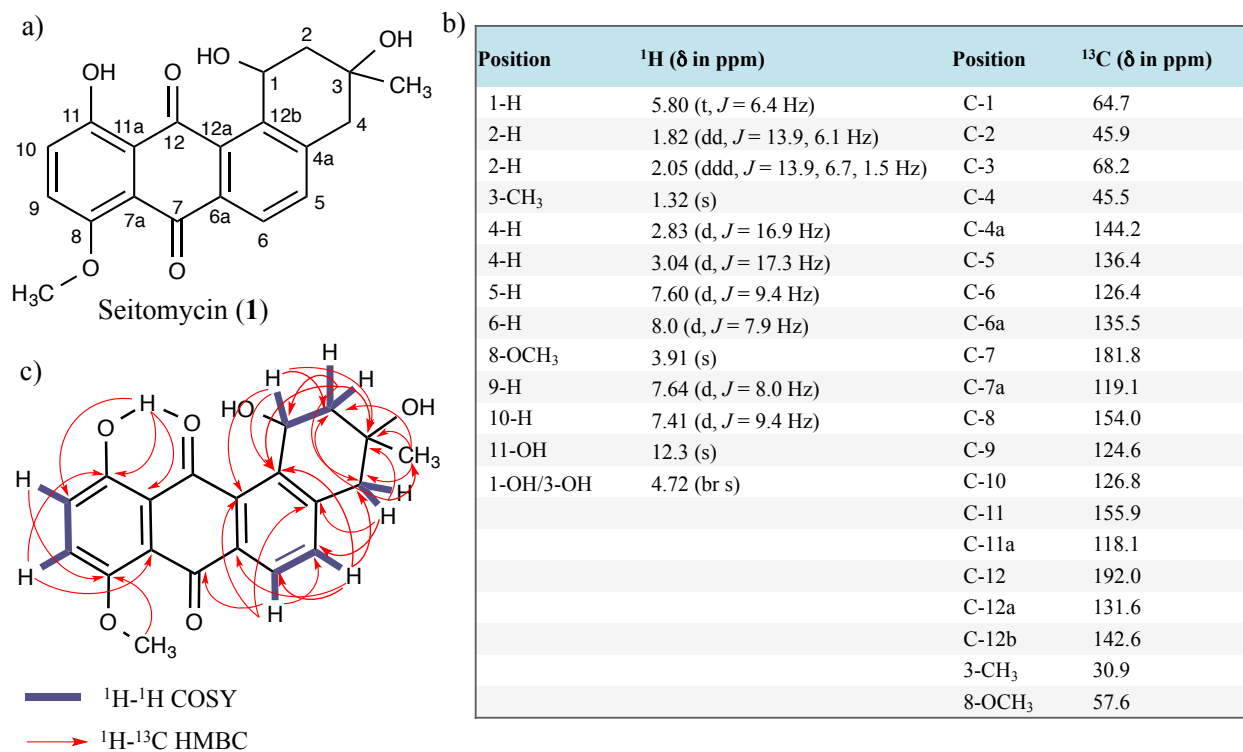


Figure 28. Structural characterization of Seitomycin (**1**). a) Chemical structure of Seitomycin with carbon numbering. b) Table showing chemical shifts and coupling constants in ¹H and ¹³C NMR spectra. c) COSY and HMBC correlations present in Seitomycin.

The ¹³C NMR spectrum consisted of 20 signals, of which two are carbonyl groups at δ 192 (C-12) and 181.8 (C-7), twelve aromatic and six aliphatic carbons. Two aromatic carbons at 155.9 (C-11) and 154 (C-8) and three aliphatic signals at 68.2 (C-3), 64.7 (C-1) and 57.6 (8-OCH₃) are likely oxygen attached. Given that there are only four aromatic hydrogens, it is rational to say that the remaining eight aromatic carbons are quaternary, two of which are oxygen attached. Among the aliphatic carbons, two methyl and two methylene groups are present. Existence of a methine group carrying a triplet hydrogen at δ 5.8 is highly likely.

In the HMBC spectrum, the first set of ortho-positioned aromatic protons at δ 7.64 and 7.4 coupled with carbon signals at δ 155.9 (C-11) and 119.1 (C-7a), and at δ 154.0 (C-8) and 118.1 (C-11a), respectively. The OH signal at δ 12.3 showed couplings to signals at δ 155.9 (C-11),

126.8 (C-10) and 118.1 (C-11a), and the methoxy protons correlated with a signal at δ 154.0 (C-8). Strong ^1H - ^1H COSY correlations were observed between the protons at δ 7.64 and 7.4. In the HMQC spectrum, these protons showed connectivity to carbons at 124.6 (C-9) and 126.8 (C-10), respectively. Methoxy protons are attached to a carbon at δ 57.6. These combined ^1H - ^1H COSY, HMBC and HMQC correlations allowed to construct a *p*-methoxy-phenol moiety present in seitomyacin. Similarly, the other pair of aromatic ortho-protons at δ 8.0 and 7.6, respectively, showed relevant three-bond couplings with signals at δ 181.8 (C-7), 142.2 (C-4a), 131.6 (C-12a) and at δ 142.6 (C-12b), 135.5 (C-6a) and 45.5 (C-4). In addition to having strong ^1H - ^1H COSY correlations, these protons at δ 8.0 and 7.6 are connected to carbons at 126.4 and 136.4 as revealed by HMQC. This and the COSY, HMBC and HMQC correlations of the aliphatic protons revealed the right half of the molecule. Due to the missing couplings of aliphatic ring protons with the quinone carbonyls, the position of the methoxy group at C8 or C11 could not be directly deduced from the HMBC spectrum. The ^{13}C signal of a chelated carbonyl group in a quinone is, however, usually by about $\Delta\delta$ 10 at deeper field as that of the non-chelated carbonyl. Since the H6 signal couples in the HMBC spectrum with the non-chelated quinone carbonyl at 181.8 (C7), the methoxy group must be located at C8 and the phenolic OH group therefore at C11, resulting in structure **1** for seitomyacin (Abdelfattah *et. al.* 2003).

Characterization of rabelomyacin (2)

The ^1H NMR spectrum showed the presence of two chelated OH groups at δ 11.6 (8-OH) and 12.3 (6-OH), and four aromatic protons: a singlet at δ 7.15 (5-H), two doublets at δ 7.49 (11-H) and 7.35 (9-H) each with coupling constants 8 Hz, and a triplet at δ 7.8 (10-H) with a coupling constant of 8 Hz. From the characteristic splitting pattern and the identical coupling constant

values, it appears that the three protons at δ 7.49, 7.35 and 7.8 are in the same aromatic rings. In the aliphatic region, a singlet at δ 1.3 (3-CH₃) for three hydrogens indicated the presence of a methyl group. Presence of a pair of AB systems at δ 3.0/3.1 (4-H) and 2.68/2.95 (2-H) were due to two methylene groups.

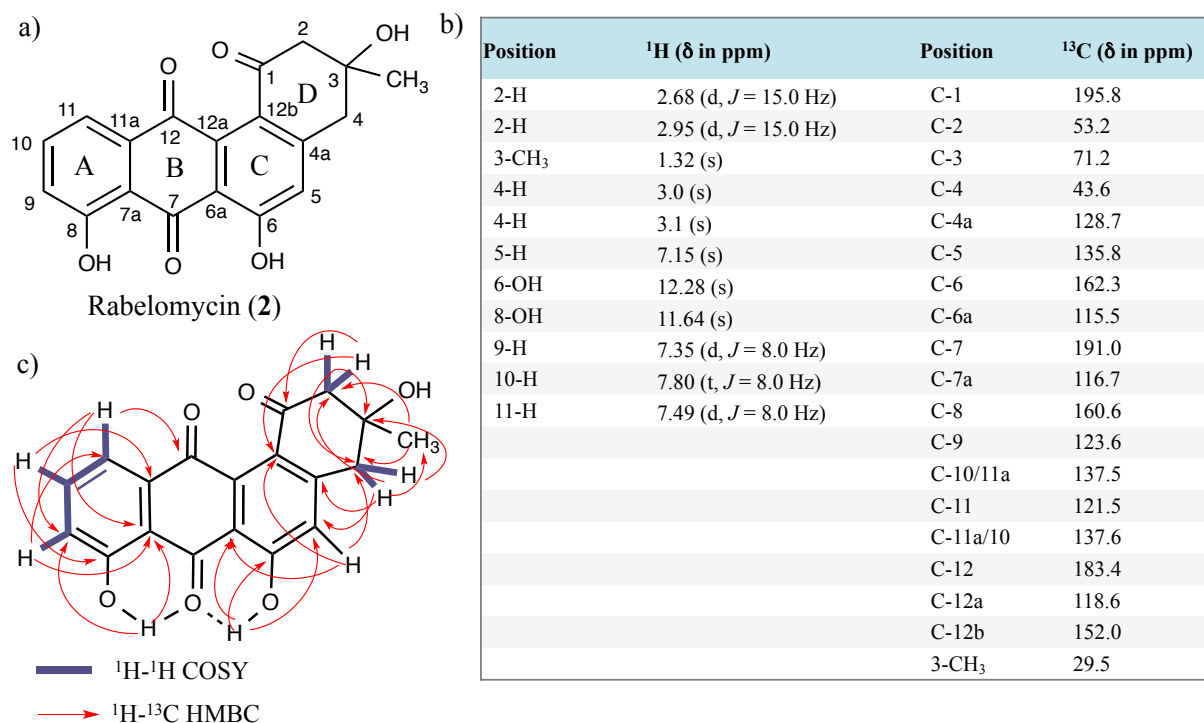


Figure 29. Structural characterization of Rabelomycin (2). a) Chemical structure of Rabelomycin with carbon numbering. b) Table showing chemical shifts and coupling constants in ¹H and ¹³C NMR spectra. c) COSY and HMBC correlations present in Rabelomycin.

The ¹³C NMR spectrum consisted of 19 signals, of which three are carbonyl groups at δ 195.8 (C-1), 191.0 (C-7) and 183.4 (C-12), twelve aromatic and four aliphatic carbons. Two aromatic carbons at 162.3 (C-6) and 160.6 (C-8) and one aliphatic carbon at 71.2 (C-3) are likely attached to oxygen. Given that there are only four aromatic hydrogens, it is likely that remaining eight aromatic carbons are quaternary, two of which are oxygen attached. Among four aliphatic carbons, one methyl and two methylene groups are present.

In the HMBC spectrum, three aromatic protons at δ 7.35, 7.49 and 7.8 made extensive three-bond couplings. These protons coupled with carbon signals at δ 116.7 (C-7a) and 121.5 (C-11), at δ 183.4 (C-12), 116.7 (C-7a) and 123.6 (C-9), and at δ 137.6 (C-10/11a) and 160.6 (C-8), respectively. The OH signal at δ 11.6 showed couplings to carbon signals at δ 116.7 (C-7a) and 123.6 (C-9). Strong ^1H - ^1H COSY correlations were observed among the three aromatic protons. In the HMQC spectrum, these protons showed connectivity to carbons at δ 123.6 (C-9), 121.5 (C-11) and 137.5 (C-10/11a), respectively. The combined ^1H - ^1H COSY, HMBC and HMQC correlations allowed constructing a phenol moiety containing A ring present in rabelomycin.

Similarly, in the HMBC spectrum, the other aromatic proton at δ 7.15 showed relevant three-bond couplings with signals at δ 115.5 (C-6a), 152.0 (C-12b) and 43.6 (C-4). The OH signal at δ 12.2 showed correlations with carbons at δ 135.8 (C-5), 115.5 (C-6a) and a two-bond coupling with carbon 162.3 (C-6). HMQC revealed that proton at δ 7.15 is connected to carbon at δ 135.8 (C-5). Extensive COSY, HMBC and HMQC correlations of the aliphatic protons established the C and D rings as well as the angular structure of the molecule. Although, there were no couplings among the protons in the C and D rings with the quinone carbonyls, the positions of the phenolic groups in A and C rings with respect to the quinone carbonyls C-7 and C-12 were easily established. Like seitomycin, the ^{13}C signal of a chelated carbonyl group (C-7) in a quinone appeared at deeper field by about $\Delta\delta$ 10 compared to that of the non-chelated carbonyl (C-12). Since the H11 signal couples in the HMBC spectrum with the non-chelated quinone carbonyl at δ 183.4 (C-12), the phenolic group in ring A must be located at C8. The position of the phenolic OH group in ring C was placed accordingly, resulting in structure **2** for rabelomycin (Liu *et. al.* 1970). This is also in accordance with the reported structure of rabelomycin.

Characterization of 8-*O*-methylrabelomycin (**3**)

Structure of 8-*O*-methylrabelomycin **3** was determined similarly. The numbers of aromatic and aliphatic protons as well the coupling patterns are identical to those present in rabelomycin. Obvious differences were the presence of a methoxy group at δ 3.96 (8-OCH₃) and the absence of phenolic OH at 11.6 with little or no changes in other values in the ¹H NMR spectrum.

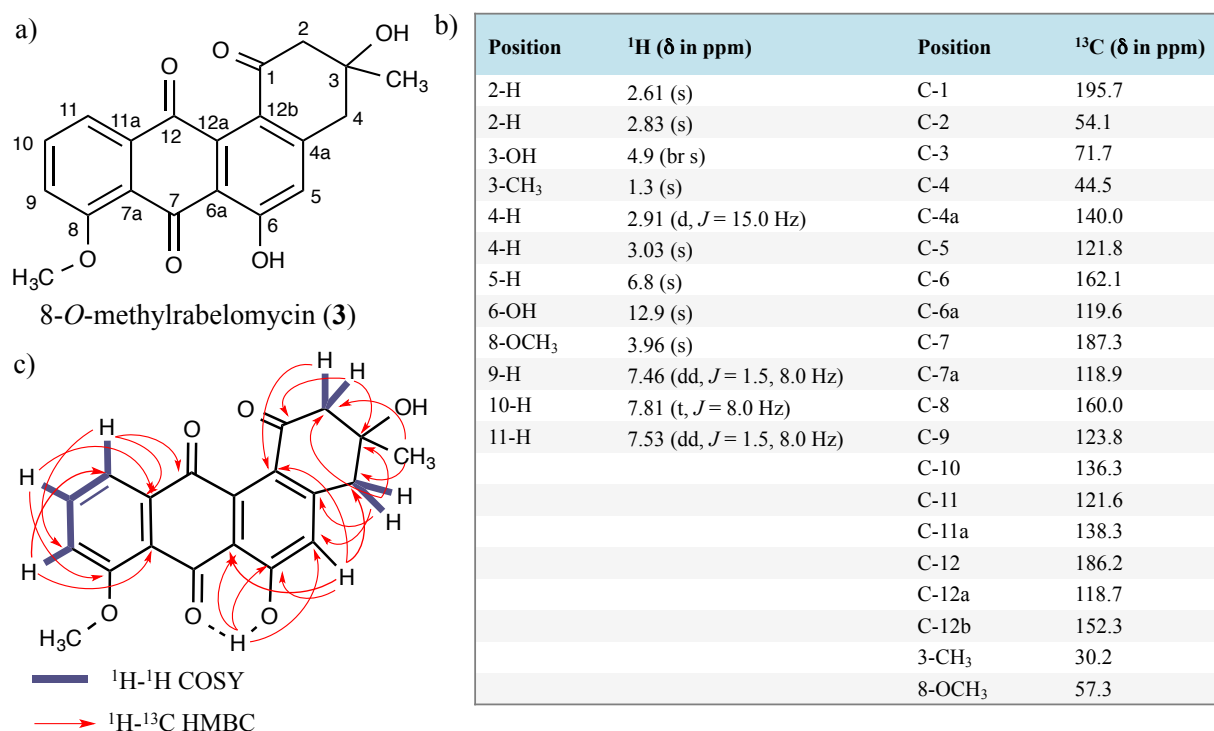


Figure 30. Structural characterization of 8-*O*-methylrabelomycin (**3**). a) Chemical structure of 8-*O*-methylrabelomycin with carbon numbering. b) Table showing chemical shifts and coupling constants in ¹H and ¹³C NMR spectra. c) COSY and HMBC correlations present in 8-*O*-methylrabelomycin.

In ¹³C spectrum, the appearance of a signal at 57.3 (8-OCH₃) corroborated with the presence of a methoxy group. Such observations indicated that the compound is a C8 methoxy derivative of rabelomycin. Further confirmation came from COSY, HMQC and HMBC correlations. Unlike rabelomycin, all the HMBC connectivity involving C8 phenolic was absent in the new compound. However, hydroxyl proton at δ 12.9 (6-OH) correlated well with carbon signals at δ

119.6 (C-6a), 162.1 (C-6) and 121.6 (C-11), thus confirming the presence of methoxy group at C8 position. The NMR spectra are identical to those reported for 8-*O*-methylrabelomycin (Fotso *et. al.* 2008).

Characterization of 1-Deoxo-1-hydroxy-8-*O*-methylrabelomycin (**4**)

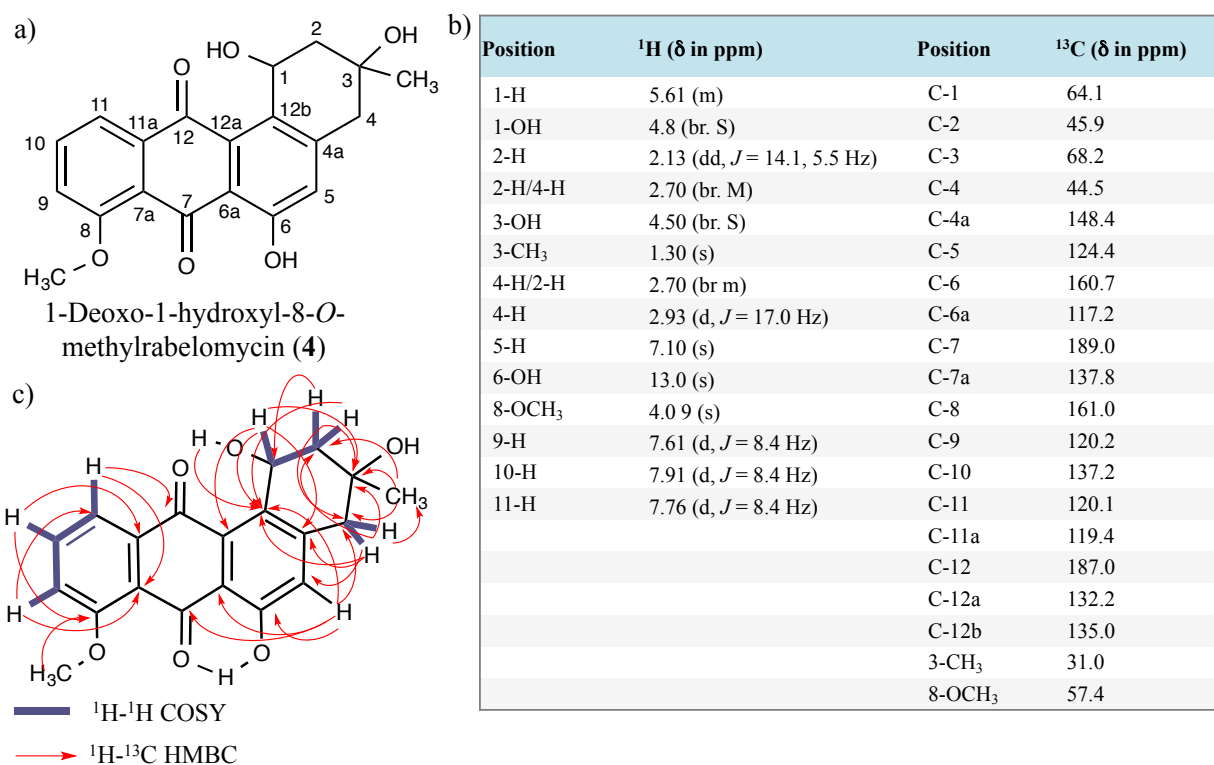


Figure 31. Structural characterization of 1-Deoxo-1-hydroxy-8-*O*-methylrabelomycin (**4**). a) Chemical structure of 1-Deoxo-1-hydroxy-8-*O*-methylrabelomycin with carbon numbering. b) Table showing chemical shifts and coupling constants in ¹H and ¹³C NMR spectra. c) COSY and HMBC correlations present in 1-Deoxo-1-hydroxy-8-*O*-methylrabelomycin.

The ¹H-NMR spectrum of **4** is very similar to that of 8-*O*-methylrabelomycin, with the major difference being the presence in **4** of a broad apparent quartet at δ 5.61 (1-H) and a doublet at δ 4.8 (1-OH). Presence of these protons each with integration one suggested that one of the carbonyls in 8-*O*-methylrabelomycin has been reduced. The ¹H-¹H COSY spectrum showed a

coupling between the methine at δ 5.6 (1-H) and the doublet of doublet methylene protons at δ 2.13 (2-H) indicating that the C-1 carbonyl has been reduced. The carbon spectrum supported this conclusion on the basis of only two carbonyl groups at δ 189.0 (C-7) and 187.0 (C-12) in **4** and the presence of an additional methine carbon at δ 64.1. HMQC correlation spectra showed the connectivity of H-1 to C-1 at δ 64.1. The reduction of the carbonyl at C-1 was further confirmed by the HMBC correlation of the H-2 protons to C-1 at δ 64.1. Presence of the other long-range couplings as revealed by HMBC remained identical to those present in 8-*O*-methylrabelomycin. The spectra are identical to those reported for 1-deoxy-1-hydroxy-8-*O*-methylrabelomycin (Fotso *et. al.* 2008).

3.2.6 Proposed biosynthesis of seitomycin

The biosynthesis of the type II PKS derived angucycline compound seitomycin can be rationalized on the basis of the predicted functions for the genes found within the BAC:TX451/2512 clone (Figure 25). The intermediate compounds isolated from this pathway, namely 8-*O*-methylrabelomycin **3** and 1-deoxy-1-hydroxy-8-*O*-methylrabelomycin **4** further support this hypothesis. Type II polyketide synthase (PKS) enzymes utilize a starter unit, most commonly, acetyl CoA and extender units (9 malonyl CoA) to establish a decaketide backbone (Figure 28). In our proposed biosynthetic scheme for seitomycin, the C20 angucline backbone (I) is predicted to be synthesized by enzymes encoded by the minimal PKS genes (6, 7, 8) in the cluster. The first ring closure is proposed to take place by the action of minimal PKS genes as well (Khosla 2009). The decaketide intermediate then undergoes further cyclization and aromatization to establish the tri-cyclic backbone (II) of C20 angucline. A ketoreductase (orf 10),

cyclase (orf 5) and aromatase (orf 11) in the cluster can be assigned to carry out this transformation.

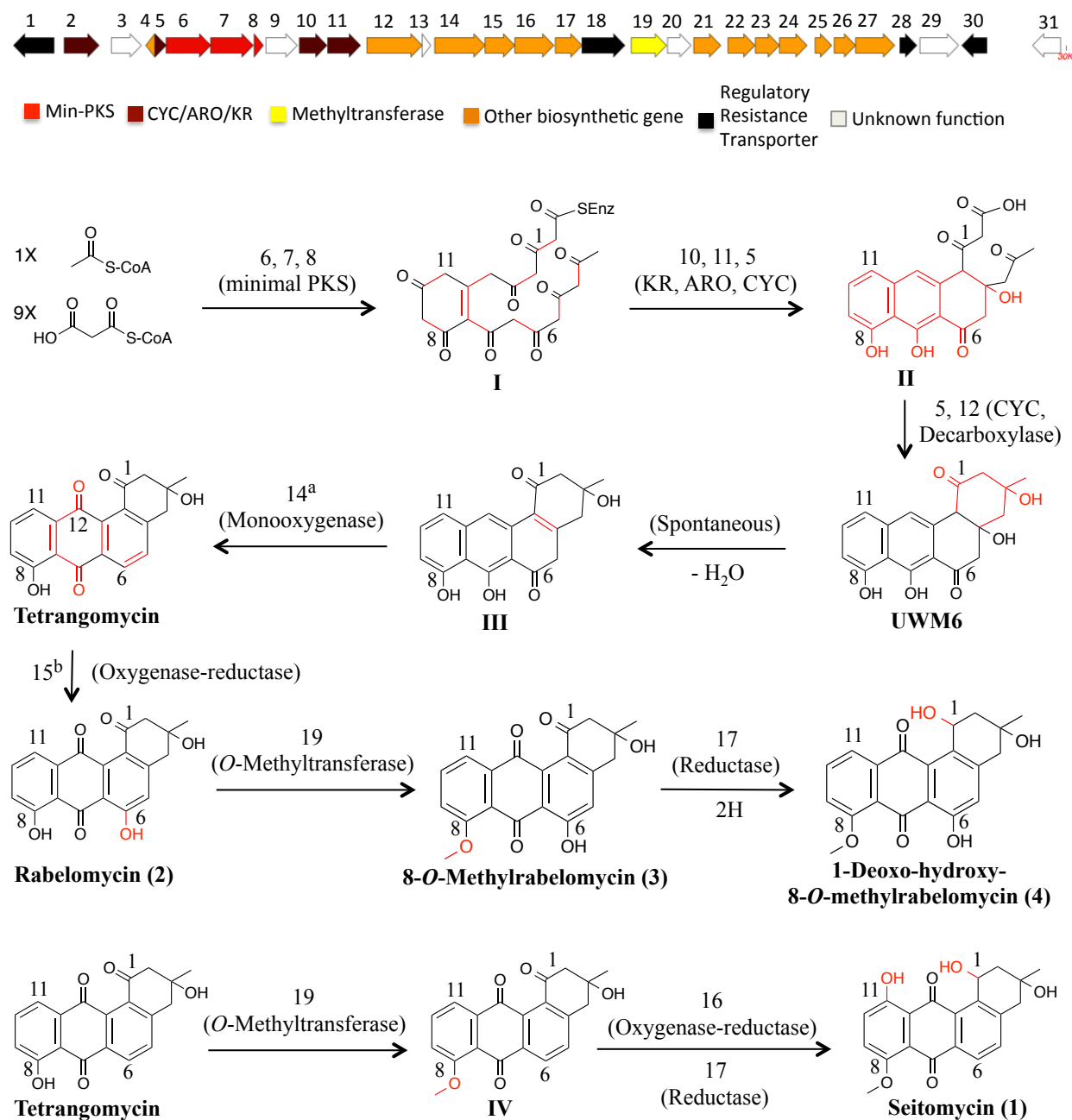


Figure 32. Proposed biosynthetic scheme for Seitomycin (I) based on gene function prediction defined by sequence homology and domain analysis. Red color indicates enzyme-mediated transformations. ^aThe monooxygenase (orf 14, homologue of LndE gene from landomycin E biosynthetic cluster) can catalyze oxidation at C-12 and removal of oxygen at C-6. ^bOrf 15 is a bifunctional oxygenase-reductase enzyme and a homologue of LndM2. In this case it is catalyzing an oxidation reaction at C-6 (Kharel *et al.* 2012).

Table 3: List of genes not used in biosynthetic hypothesis of seitomycin

Gene	Clone	Size	Deduced function of protein	I/S%	Origin	accession no.
<i>orf1</i>	451	1194	MFS transporter	73/68	<i>Streptomyces sp.</i>	WP_009083509.1
<i>orf2</i>	451	993	aldo/keto reductase	82/88	<i>S. bingchenggensis</i>	YP_004958664.1
<i>orf3</i>	451	813	hypothetical protein	57/67	<i>Streptomyces sp.</i>	BAJ07840.1
<i>orf4</i>	451	225	putative oxygenase	41/75	<i>Streptomyces sp.</i>	YP_008012343.1
<i>orf20</i>	2512	672	hypothetical protein	31/46	<i>S. venezuelae</i>	YP_006877087.1
<i>orf21</i>	2512	768	putative N5,N10- methylenete tetrahydromethanopterin reductase	75/85	<i>Streptomyces sp.</i>	BAJ07867.1
<i>orf22</i>	2512	768	oxygenase-reductase	58/76	<i>S. fradiae</i>	AFU51427.1
<i>orf23</i>	2512	687	hypothetical protein	65/82	<i>Streptomyces sp.</i>	BAJ07849.1
<i>orf24</i>	2512	786	putative short-chain dehydrogenase/reductase	71/86	<i>Streptomyces sp.</i>	BAJ07861.1
<i>orf25</i>	2512	474	putative reductase	80/90	<i>Streptomyces sp.</i>	BAJ07862.1
<i>orf26</i>	2512	612	putative oxygenase	61/76	<i>Streptomyces sp.</i>	BAJ07850.1
<i>orf27</i>	2512	1116	putative hydroxylase	38/52	<i>S. vietnamensis</i>	ADO32775.1
<i>orf28</i>	2512	483	MarR family transcriptional regulator	65/74	<i>S. viridochromogenes</i>	WP_003990963.1
<i>orf29</i>	2512	1098	FAD-dependent pyridine	52/64	<i>Saccharopolyspora erythraea</i>	YP_001102692.1
<i>orf30</i>	2512	714	Phosphate regulon transcriptional regulator	69/79	<i>S. venezuelae</i>	YP_006881518.1
<i>orf31</i>	2512	1353	hypothetical protein	54/86	<i>Amycolatopsis orientalis</i>	YP_008012343.1

The second and third ring closures take place through the action of cyclase (*orf 5*) and a decarboxylase (*orf 12*) to produce the well-known intermediate from angucycline pathways, UWM6. This intermediate then undergoes spontaneous elimination of one water molecule to generate intermediate III (Figure 32). This intermediate is then transformed into tetrangomycin, another well-characterized intermediate from angucycline biosynthesis, by a monooxygenase (*orf 14*). This monooxygenase is homologous to LndE enzyme from Landomycin E biosynthetic gene cluster and is known to oxygenate at C-12 of angucycline backbone (Kharel *et al.* 2012). In this biosynthetic proposal, intermediate tetrangomycin is predicted to enter into two distinct pathways resulting in the formation of differentially hydroxylated angucyclines. Hydroxylation at C-6 of tetrangomycin (*orf 15*) yields rabelomycin (**2**), a common shunt product/intermediate from angucycline pathways. *Orf 15* represents an oxygenase-reductase family enzyme that is homologous with LndM2, which is known to catalyze hydroxylation at C-6 position of tetrangomycin to produce rabelomycin (Kharel *et al.* 2012). Methylation of rabelomycin at C-8

hydroxyl group via an *O*-methyltransferase (orf 19) leads to 8-*O*-methyrrabelomycin (**3**), which through the action of a reductase, most likely orf 17 of our cluster, generates 1-deoxy-1-hydroxy-8-*O*-methyrrabelomycin (**4**). Methylation of tetrangomycin at the C-8 hydroxyl group (orf 19) followed by the reduction of C-1 carbonyl group (orf 17) and hydroxylation at C-11 by an oxidoreductase enzyme (orf 16) could lead to the formation seitomycin (**1**) (Figure 32).

3.3 Discussion

Uncultured soil bacteria is a prolific source of chemical diversity that is yet unexplored. eDNA libraries constructed using soil samples from different geographical regions could give us access to the genetic diversity of these uncultured microbes. To discover a broader range of biosynthetic pathways, an expression independent sequence-based strategy was applied to recover cosmids containing biosynthetic gene clusters. A number of studies have indicated that eDNA samples are rich in Type II PKS genes (Pang *et al.* 2008; Wawrik, Kutliev *et al.* 2007). I screened a multimillion membered metagenomic library constructed using sample collected from Texas using a set of degenerate primers based on the conserved KS_{α} gene from type II PKS cluster as probe. Preliminary sequence-based screens demonstrated the successful recovery of biosynthetic cosmids encoding a diverse range of type II polyketides. A phylogenetic analysis was carried out on the eDNA-derived KS_{α} genes. This analysis confirmed that the Tx eDNA library is rich in type II PKS biosynthetic pathways. Clustering KS_{α} sequences at 85% identity revealed 91 unique clones. Based on the above analysis, thirty clones carrying unique KS_{α} sequences were picked at random and were recovered from eDNA library. One clone, cosTX451, was identified that conferred the production of clone specific metabolites to *S. albus*. Full sequencing of the eDNA insert in cosTX451 revealed that the biosynthetic pathway is only partially captured in

this cosmid. In order to access the complete pathway, one overlapping clone, cosTX2512, was identified and recovered by rescreening the Tx library. TX451 was assembled with the overlapping clone cosTX2512 into a complete pathway through Transformation Associated Recombination (TAR). The secondary metabolite profile obtained from heterologous expression of the TAR construct (BAC:TX451/2512) was different from what was observed for cosTX451 alone. In total, four natural products were isolated and characterized from this pathway.

Extensive NMR studies revealed that the molecules encoded by this gene cluster indeed belonged to the angucycline class of type II PKS derived natural products. The angucyclines represent the largest group of type II PKS-engineered natural products (Rohr and Thiericke 1992). Seitomyacin, a Type II PKS derived angucycline antibiotic was isolated as the major metabolite of this pathway. Although no new metabolites were isolated in this study, using metagenomic tools, I successfully identified, for the first time, the biosynthetic gene cluster for the Type II PKS derived angucycline antibiotic Seitomyacin. The current study provides further support to the idea that sequence-based screening of large metagenomic libraries can lead to the discovery of biosynthetic machinery for new chemistry. These results, together with those of other small-molecule directed metagenomic studies, suggest that large-scale functional screening of eDNA derived clones should be a productive strategy for generating structurally previously uncharacterized gene clusters and chemical entities for use in future drug development efforts.

3.4 Materials and Methods

3.4.1 General experimental procedures

Streptomyces albus J1074 was used for heterologous expression studies. ¹H and 2D NMR data were obtained on a Bruker Avance-600 spectrometer. ¹³C NMR spectra were obtained on a Bruker Avance-600 spectrometer.

3.4.2 PCR screening of eDNA libraries for KS_α genes and clone recovery

Cosmid DNA isolated from the library served as the template to amplify KS_α sequences using previously described conditions and degenerate primers (dp:540F-GGITGCACSTCIGGIMTSGAC and dp:1100R-CCGATSGCICCSAGIGAGTG) (*Wawrik et. al.* 2005). Twenty-five microliter PCR reactions contained 50 ng of cosmid DNA, 2.5 μM of each primer, 2 mM dNTPs, 1× ThermoPol reaction buffer (New England Biolabs), 0.5 unit Taq DNA polymerase (New England Biolabs), and 5%DMSO. PCR was conducted using the following touchdown protocol: denaturation (95 °C, 2 min), 8 touchdown cycles [95 °C, 45 s; 65 °C (−1 °C per cycle), 1 min; 72 °C, 2 min], 35 standard cycles (95 °C, 45 s; 58 °C, 1 min; 72 °C, 2 min), and a final extension step (72 °C, 2 min). The PCR amplicons were gel-purified (Qiagen) and sequenced (Genewiz). A ClustalW-based phylogenetic analysis was carried out on the eDNA-derived KS_α amplicons using MacVector.

Sublibrary pools from which unique KS_α genes were amplified were used as the starting point for the recovery of unique KS_α containing eDNA clones. Overnight cultures of sublibraries containing KS_α genes of interest were plated into 96-well microtiter plates at a dilution of 10^{−6}.

After 18 h at 37 °C, the diluted cultures were screened by whole-cell PCR using KS_α specific primers. PCR positive wells were then plated directly on solid media to yield distinct colonies that were screened in a second round of whole-cell PCR. The PCR protocol used was as follows: denaturation (95 °C, 5 min), 36 standard cycles (95 °C, 30 s; 55 °C, 30 s; 72 °C, 40s), and a final extension step (72 °C, 7 min).

3.4.3 Cosmid retrofitting and integration into Streptomyces albus

Recovered cosmids were digested with the rare PstI to ensure that only the pWEB-TNC vector was cut and not the eDNA insert. Each linearized cosmid was then ligated with the 6.81 kb DraI fragment from pOJ436. This fragment contains an origin of transfer (oriT), an apramycin resistance marker, and elements of the phiC31 attP-attB integration system needed for integration into diverse Streptomyces sp. Retrofitted cosmids were transformed into *E. coli* S17.1 and conjugated into Streptomyces using published protocols.

3.4.4 eDNA clone sequencing

PKS containing cosmids recovered from the eDNA libraries were initially sequenced by 454 pyrosequencing (Roche) by Memorial Sloan Kettering Cancer Center Genomics Core Laboratory. Reads were processed with Newbler (Roche) and Velvet (Zerbino and Birney 2008). Sequence gaps between assembled contigs generated during 454 sequencing were completed by primer walking (GENEWIZ). The size and deduced role of the open reading frames of sequenced cosmids were sequenced by 454-pyrosequencing. ORFs (open reading frames) were predicted using MetaGeneMark (Zhu et al. 2010), and predicted ORFs were annotated by Blast search and Pfam domain analysis. MacVector was used for sequence manipulation.

3.4.5 Transformation-associated recombination (TAR) reassembly of overlapping clones

One overlapping clone for cosTX451 was subsequently identified in the Tx library using clone-specific primers designed to recognize the sequence at one end of cosTX451 (TX451R_FW: ATCACCTTGTCGGCGTCGTG, TX451R_RV: AGAACCACCTGGAGGTGGTG). Hits were identified in subpool TX2512. Overlapping eDNA cosmid (designated cosTX2512) was recovered from these pools using the serial dilution method outlined above. The PCR protocol used was as follows: denaturation (95, 5 min), 36 standard cycles (95 °C, 30 s; 55 °C, 30 s; 72 °C, 40s), and a final extension step (72 °C, 7 min).

The two overlapping eDNA cosmid clones that are predicted to contain the TX451 gene cluster (cosTX451 and cosTX2512) were assembled into a single bacterial artificial chromosome (BAC) clone using transformation-associated recombination (TAR) in yeast and the pTARa BAC shuttle vector. The pathway-specific TAR capture vector was constructed using InFusion cloning methodology (Clontech). ~500 bp upstream and downstream homology arms were amplified from the cos451 and cosTX2512 cosmids, respectively. For cosTX451 and cosTX2512, homology arms needed for TAR recombination were generated using the following primers:

TX451UPS_FW (5'- CTATCGATCTCGAGGGCTGCGCGACGCGGCCGACC-3'),

TX451UPS_RV (5'- GTGCTGCATGTTAACACGATGGCGATGGGGCCTCC-3'),

TX2512DWS_FW (5'-GTTAACATGCAGCACGGCGACGTGGT- 3'),

TX2512DWS_RV (5'- CCCTGCAGGAGCTCGGGCGCGGCGGAGGTCGACGT-3')

Primers TX451UPS_FW and TX2512DWS_RV include 15 bp sequences (underlined) that

overlap with BmtI/SphI linearized pTARa capture vector. Primer TX451UPS_RV was designed to contain a 15 bp overlap (underlined) with the TX2512DWS_FW primer and an HpaI site (bold), which was added to facilitate the subsequent linearization of the pathway-specific pTARa capture vector. Gel-purified amplicons (Qiagen) and BmtI/SphI linearized pTARa were combined in a standard InFusion cloning reaction (Clontech) to make the pathway-specific capture vector.

For TAR assembly, 200 ng of DraI-digested cosTX451 and cosTX2512 were mixed with 100 ng of the HpaI-cut pathway-specific capture vector and then transformed into 200 μ L of *Saccharomyces cerevisiae* CRY1-2 spheroplasts prepared according to published protocols. Transformed spheroplasts were mixed with synthetic complete (SC) top agar (1 M sorbitol, 1.92 g/L synthetic complete uracil dropout supplement, 6.7 g/L yeast nitrogen base, 2% glucose, and 2.5% agar) and overlaid onto SC uracil dropout plates. Plates were incubated at 30 °C for 72 h. DNA was prepared from 12 yeast colonies using a zymolyase lysis protocol (ZYMO RESEARCH) and screened by PCR with primers designed to recognize sequences from cosTX451 and cosTX2512. A BAC clone that was “PCR positive” with all the primer sets was electroporated into *E. coli* EPI300 (Epicentre). DNA isolated from the resulting *E. coli* was fully sequenced by 454 pyrosequencing and confirmed BAC:TX451/2512 to be faithful reassemblies of the overlapping cosmid clones.

3.4.6 Heterologous expression and analysis

cosTX451 and BAC:TX451/2512 transformed into *E. coli* S17.1 and conjugated into *S. albus*. Spore stocks of *S. albus* transformed with the retrofitted cosTX451 or BAC:TX451/2512 clone

was inoculated in 125 mL baffled flasks containing 50 mL of R5A media (Zhu *et al.* 2010) (for 1 L: 100 g sucrose, 0.25 g K₂SO₄, 10.12 g MgCl₂•6H₂O, 10 g glucose, 0.1 g casamino acids, 5 g yeast extract, 21 g MOPS, 2 g NaOH, 2 mL R2YE trace elements) and grown at 30 °C (200 rpm) for 7-10 d. Cultures were then extracted twice with an equal volume of ethyl acetate. The dried extracts were dissolved in methanol (2-4 mL/L of culture). The sample was then dried, redissolved in 1-2 mL of methanol. The resulting ethyl acetate extract was compared by LC-MS to the extract from a similarly grown culture of *S. albus* transformed with a pTARa empty vector control (HPLC conditions: 30 min gradient from 5% to 100% aqueous acetonitrile containing 0.1% formic acid, C18, 4.6 mm × 150 mm, 1 mL/min).

3.4.7 Compound extraction and purification of Seitomycin, Rabelomycin, 8-Methylrabelomycin, 1-Deoxy-1-hydroxy-8-O-methylrabelomycin

Compounds **1**, **3** and **4** were isolated from 2 L cultures of *S. albus* harboring BAC:TX451/2512 grown (200 rpm at 30 °C) in 125 mL baffled flasks containing 50 mL of R5A media. 7-10 day old cultures were extracted with ethyl acetate (1/3, v/v to the culture broth), and the ethyl acetate was dried in vacuo. Compounds were isolated from the resulting ethyl acetate extract using two rounds of reversed-phase HPLC (C column, 10 18mm × 250 mm, 3.5 mL/min). The first round of HPLC (isocratic, 45% acetonitrile containing 0.1% trifluoroacetic acid) yielded three crude compounds **1** (13.2 min), **3** and **4** (15.7 min), which were repurified by a second round of HPLC using 65%, 75%, and 77% aqueous methanol with 0.1% trifluoroacetic acid to yield compounds **1** (5 mg L⁻¹), **3** (1.5 mg L⁻¹), and **4** (0.5 mg L⁻¹), respectively. Compound **2** was isolated from 2 L cultures of *S. albus* harboring cosTX451 grown under the same conditions. 7-10 day old cultures were acidified with 2N HCL (pH 3) and extracted with ethyl acetate (1/3, v/v to the

culture broth) and the organic layer was dried in vacuo. Compound **2** was isolated from ethyl acetate extract using two rounds of reversed-phase HPLC (C column, 10 18mm × 250 mm, 3.5 mL/min). The first round of HPLC (isocratic, 50% acetonitrile containing 0.1% trifluoroacetic acid) was done to get crude compound **2** (12.9 min) which was repurified by a second round of HPLC using 70 % aqueous methanol with 0.1% trifluoroacetic acid to yield compounds **2** (1.0 mg L⁻¹). Fractions containing desired compounds were dried in vacuo overnight and analyzed by MS and NMR.

Seitomycin (1): ¹H NMR (DMSO-d₆, 600 MHz): δ 12.3 (s, 1H), 8.0 (d, *J* = 7.9 Hz, 1H), 7.64 (d, *J* = 8.0 Hz, 1H), 7.60 (d, *J* = 9.4 Hz, 1H), 7.41 (d, *J* = 9.4 Hz, 1H), 5.80 (t, *J* = 6.4 Hz, 1H), 4.72 (br s, 1H), 3.91 (s, 3H), 3.04 (d, *J* = 17.3 Hz, 1H), 2.83 (d, *J* = 16.9 Hz, 1H), 2.05 (ddd, *J* = 13.9, 6.7, 1.5 Hz, 1H), 1.82 (dd, *J* = 13.9, 6.1 Hz, 1H), 1.32 (s, 3H),

Seitomycin (1): ¹³C NMR (DMSO-d₆, 150 MHz): δ 192.5, 181.7, 155.9, 154.0, 144.2, 142.6, 136.4, 135.5, 131.6, 126.8, 126.4, 124.5, 119.1, 118.1, 68.2, 64.7, 57.6, 45.9, 45.5, 30.9.

Rabelomycin (2): ¹H NMR (DMSO-d₆, 600 MHz): δ 12.28 (s, 1H), 11.64 (s, 1H), 7.8 (t, *J* = 8.0 Hz, 1H), 7.49 (d, *J* = 8.0 Hz, 1H), 7.35 (d, *J* = 8.0 Hz, 1H), 7.15 (s, 1H), 3.47 (br s, 1H), 3.07 (s, 2H), 2.95 (d, *J* = 15.0 Hz, 1H), 2.68 (d, *J* = 15.0 Hz, 1H), 1.32 (s, 3H)

Rabelomycin (2): ¹³C (DMSO-d₆, 150 MHz): δ 195.8, 191.0, 183.4, 162.3, 160.6, 152.0, 137.6, 137.5, 135.8, 128.7, 123.6, 121.5, 118.6, 116.7, 115.5, 71.2, 53.2, 43.6, 29.5

8-O-Methylrabelomycin (3): ¹H NMR (DMSO-d₆, 600 MHz): δ 12.9 (s, 1H), 7.81 (t, *J* = 8.0 Hz, 1H), 7.53 (dd, *J* = 1.5, 8.0 Hz, 1H), 7.46 (dd, *J* = 1.5, 8.0 Hz, 1H), 6.8 (s, 1H), 4.9 (br s, 1H), 3.96 (s, 3H), 3.03 (s, 1H), 2.91 (d, *J* = 15.0 Hz, 1H), 2.83 (s, 1H), 2.61 (s, 1H), 1.3 (s, 3H)

8-O-Methylrabelomycin (3): ¹³C (DMSO-d₆, 150 MHz): δ 195.7, 187.3, 186.2, 162.1, 160.0, 150.3, 140.0, 138.3, 136.3, 123.8, 121.8, 121.6, 119.6, 118.9, 118.7, 71.7, 57.3, 54.1, 44.5, 30.2

1-Deoxo-1-hydroxy-8-O-Methylrabelomycin (4): ¹H NMR (DMSO-d₆, 600 MHz): δ 13.0 (s, 1H), 7.91 (d, *J* = 8.4 Hz, 1H), 7.76 (d, *J* = 8.4 Hz, 1H), 7.61 (d, *J* = 8.4 Hz, 1H), 7.10 (s, 1H), 5.61 (m, 1H), 4.8 (br s, 1H), 4.5 (br s, 1H), 4.09 (s, 3H), 2.93 (d, *J* = 17.0 Hz, 1H), 2.7 (br m, 2H), 2.13 (dd, *J* = 14.1, 5.5 Hz, 1H), 1.3 (s, 3H)

1-Deoxo-1-hydroxy-8-O-Methylrabelomycin (4): ¹³C (DMSO-d₆, 150 MHz): δ 189.0, 187.0, 161.0, 160.7, 148.4, 137.8, 137.2, 135.0, 132.2, 124.4, 120.2, 120.1, 119.4, 117.2, 68.2, 64.1, 57.4, 45.9, 45.5, 31.0

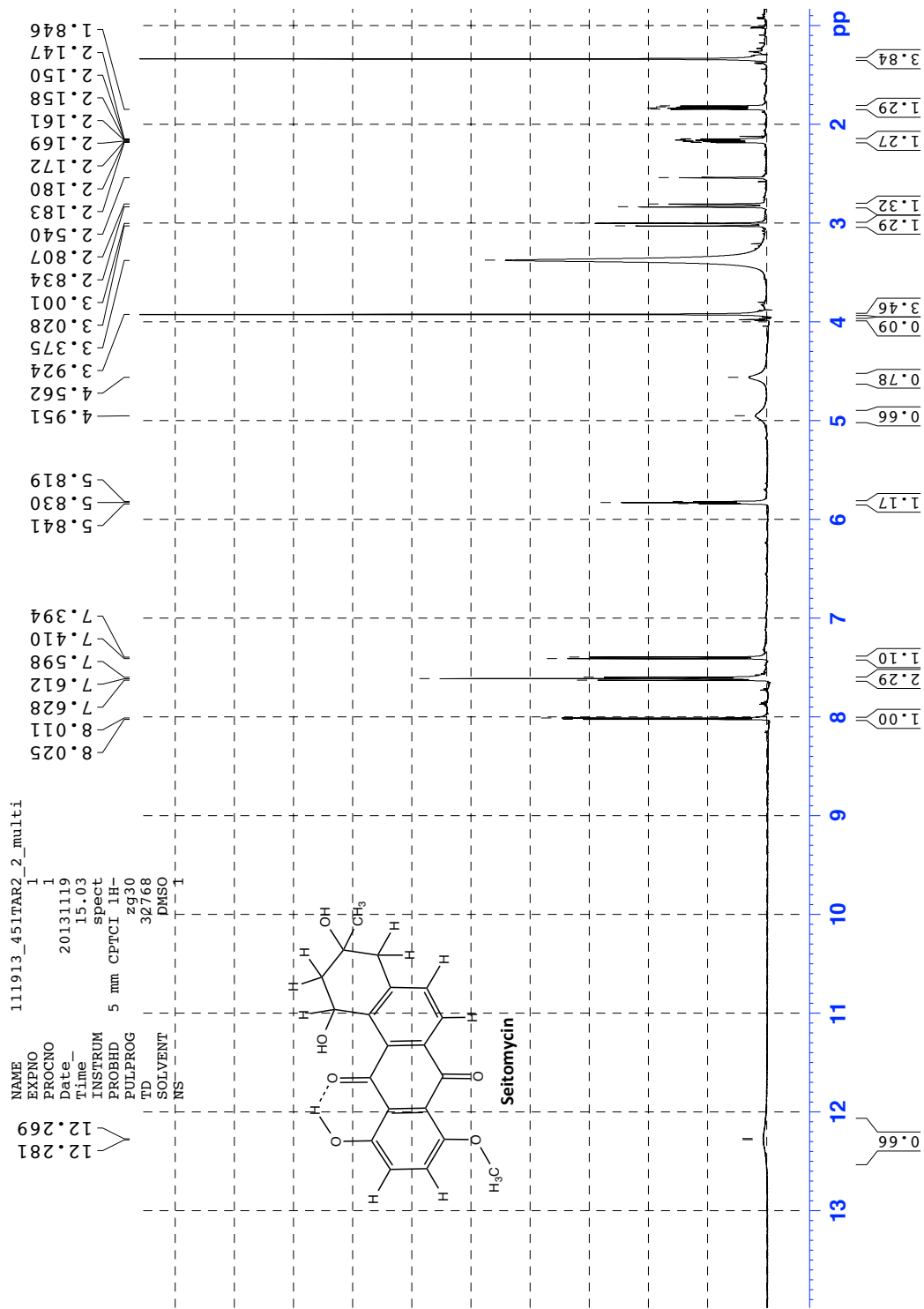


Figure 33. ^1H NMR spectrum of Seitoimycin (I)

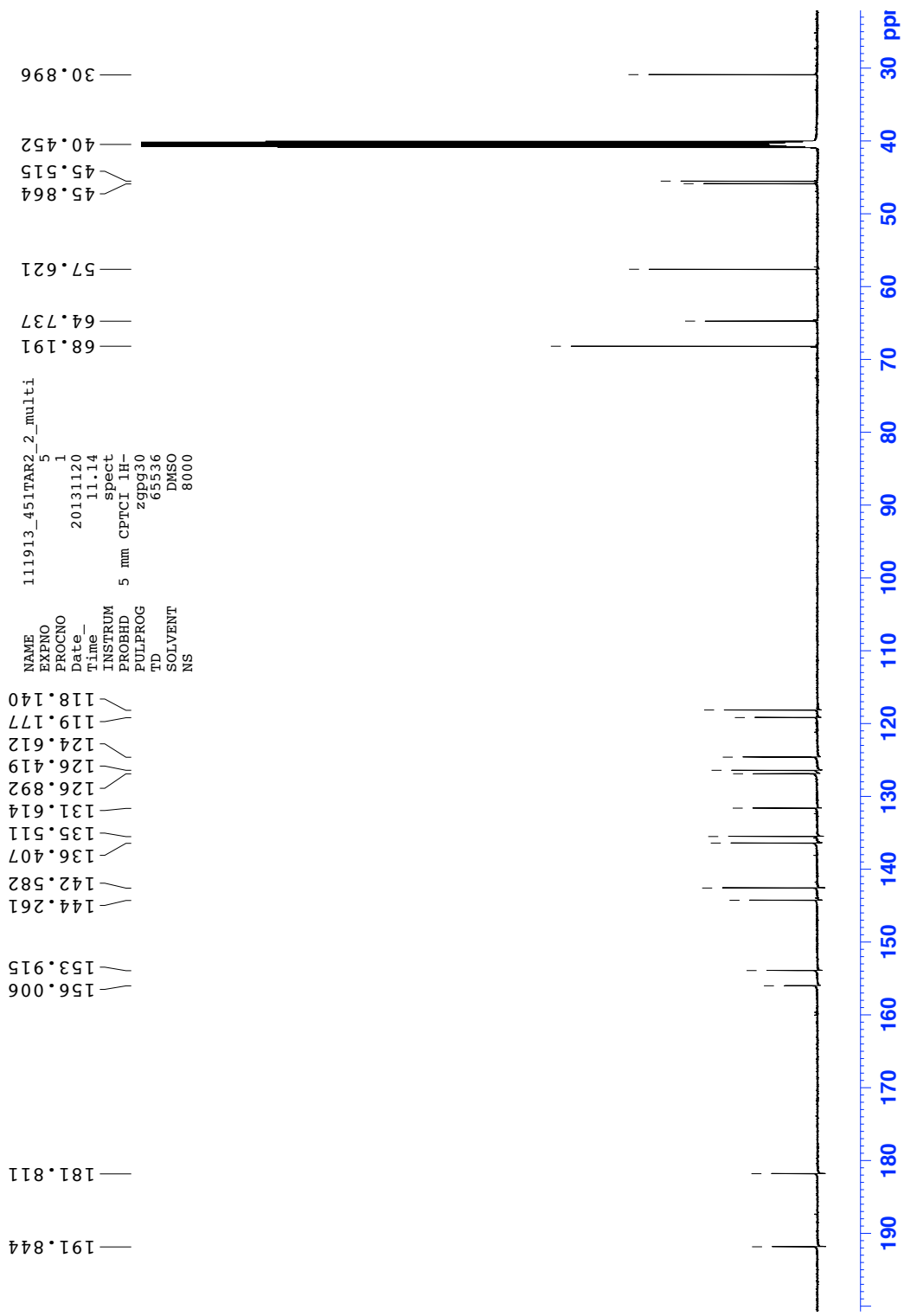


Figure 34. ¹³C NMR spectrum of Seitomycin (1)

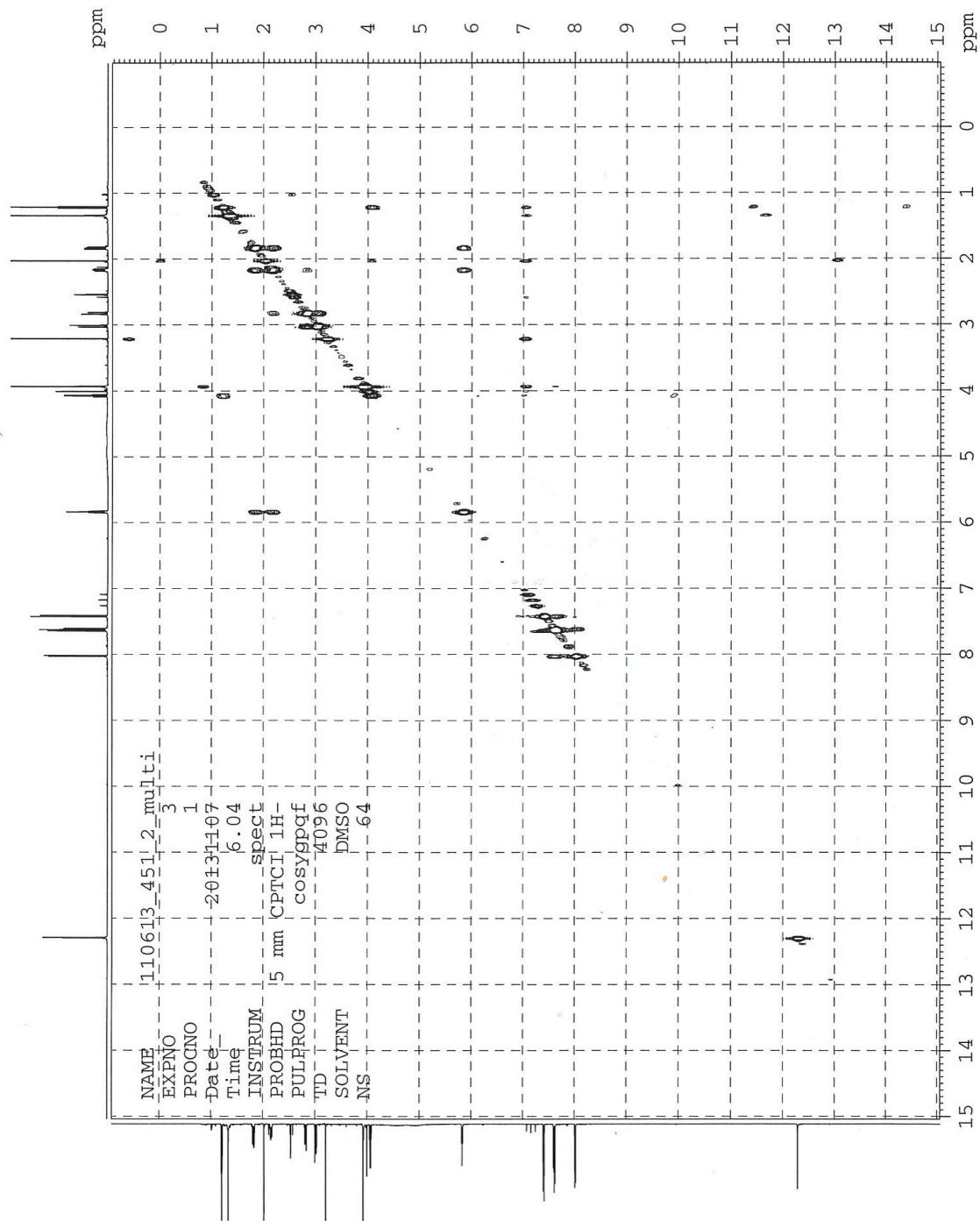


Figure 35. ^1H - ^1H COSY spectrum spectrum of Seitomycin (1)

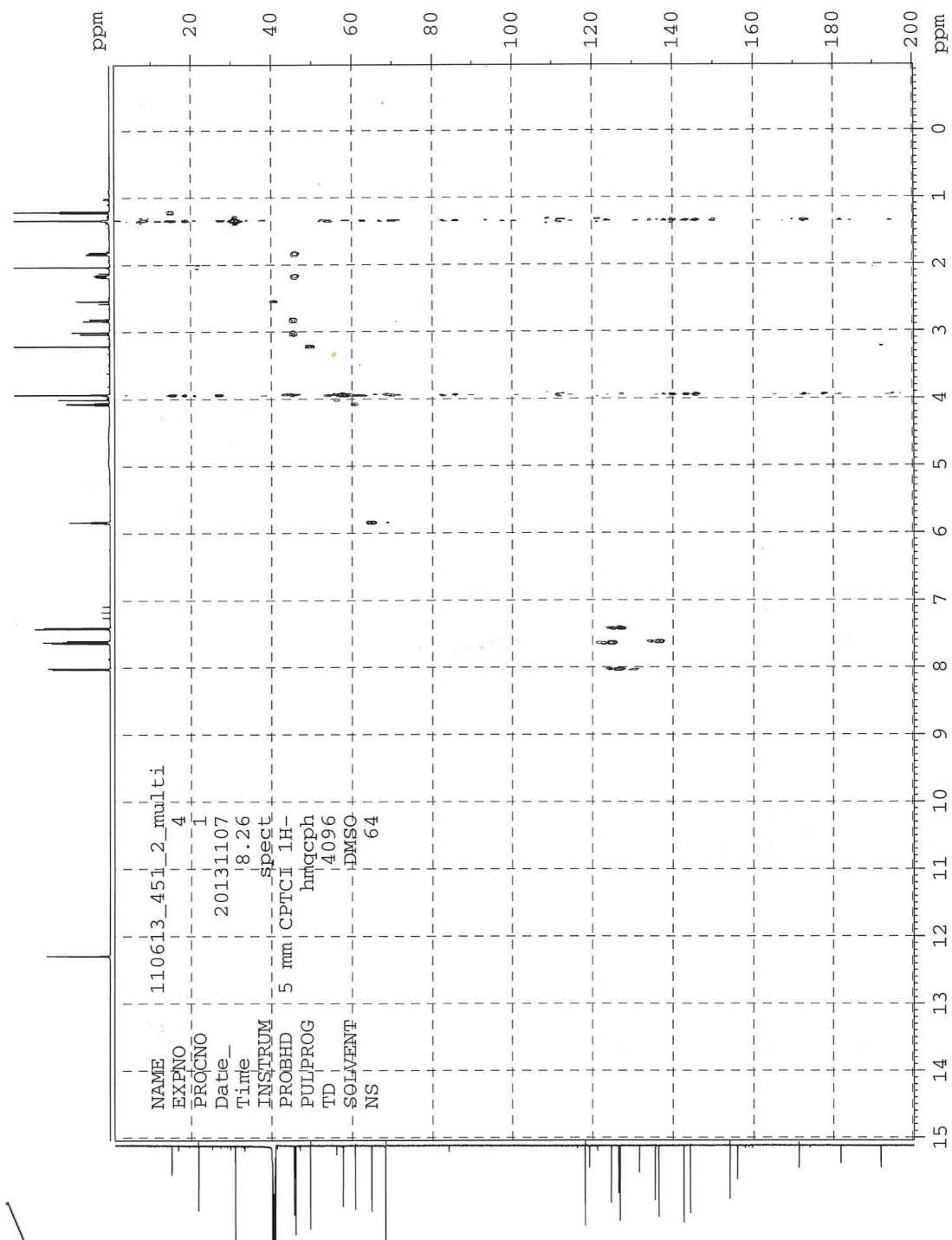


Figure 36. ^1H - ^{13}C HMQC spectrum of Setomycin (**1**)

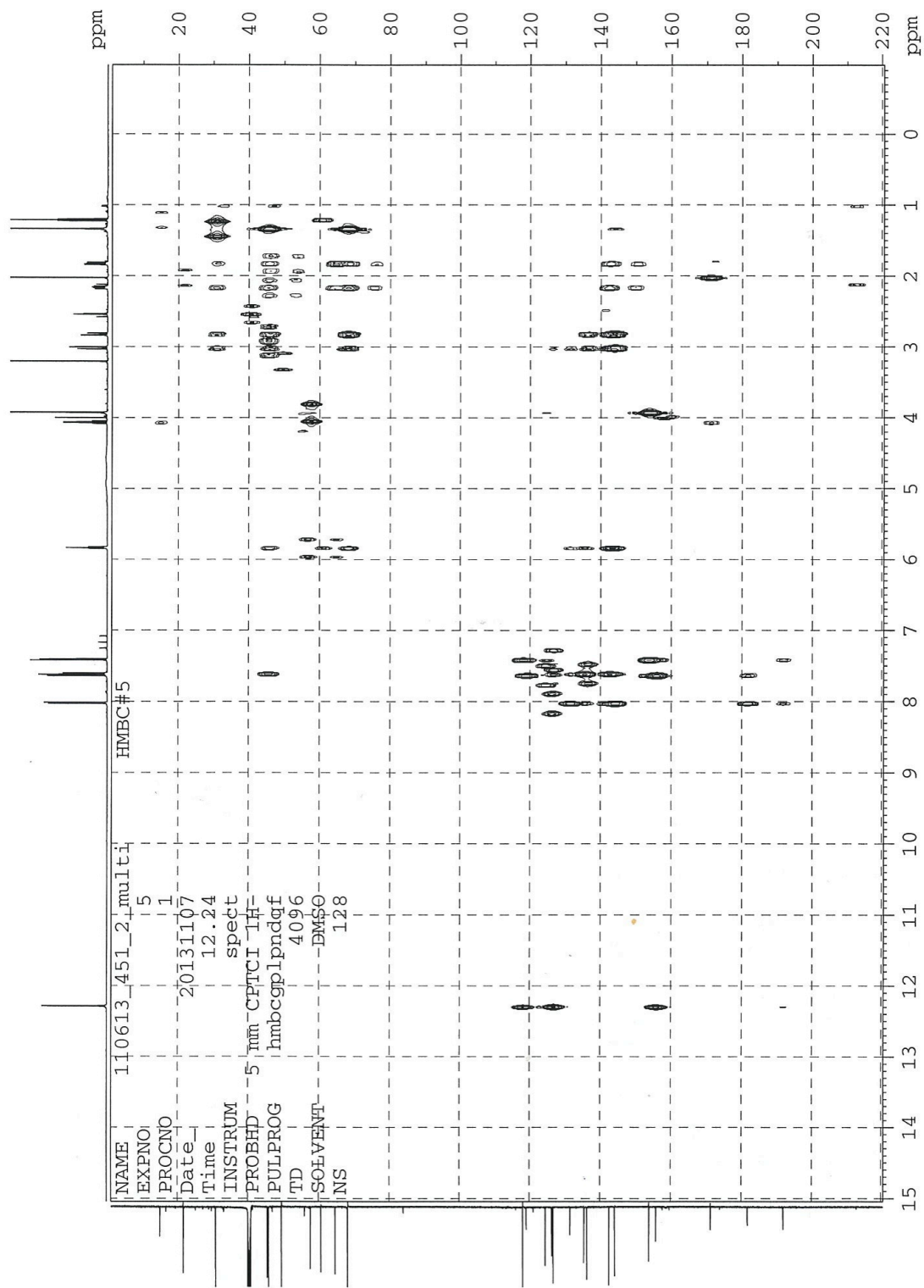


Figure 37. ^1H - ^{13}C HMBC spectrum of Seitomycin (1)

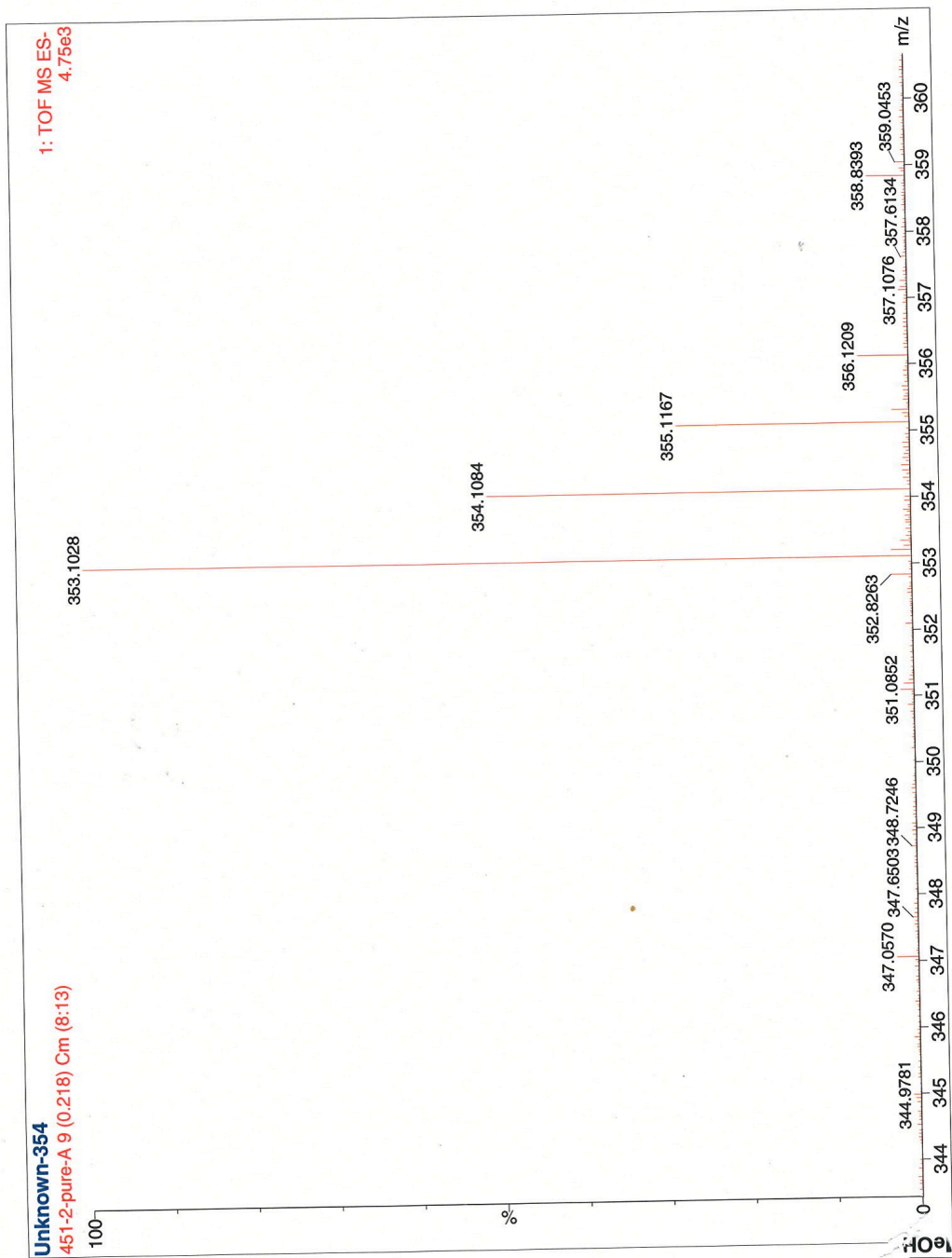


Figure 37a. High-resolution mass spectra (HRMS) of Seitomycin (1)

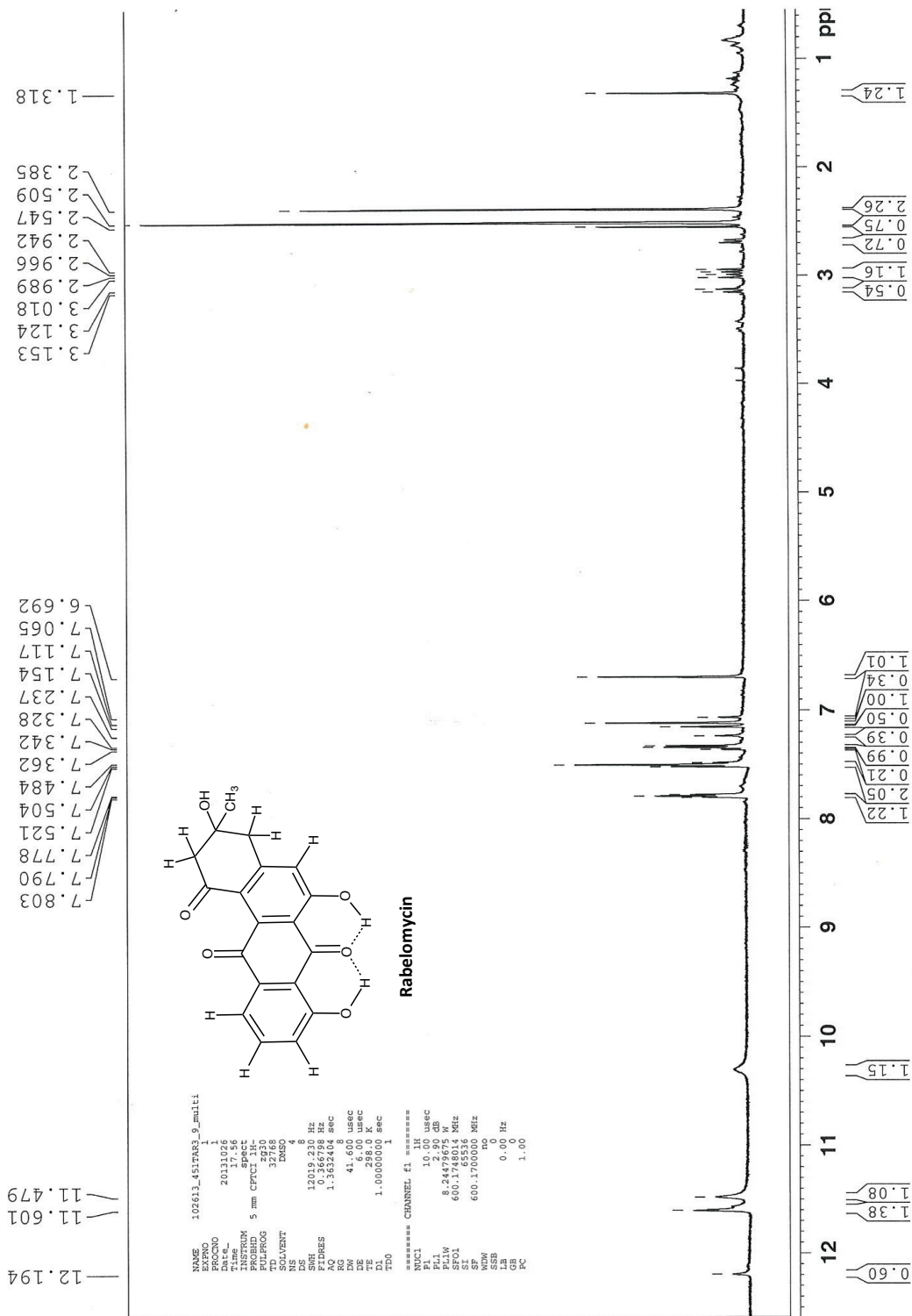


Figure 38. ¹H NMR spectrum of Rabelomycin (2)

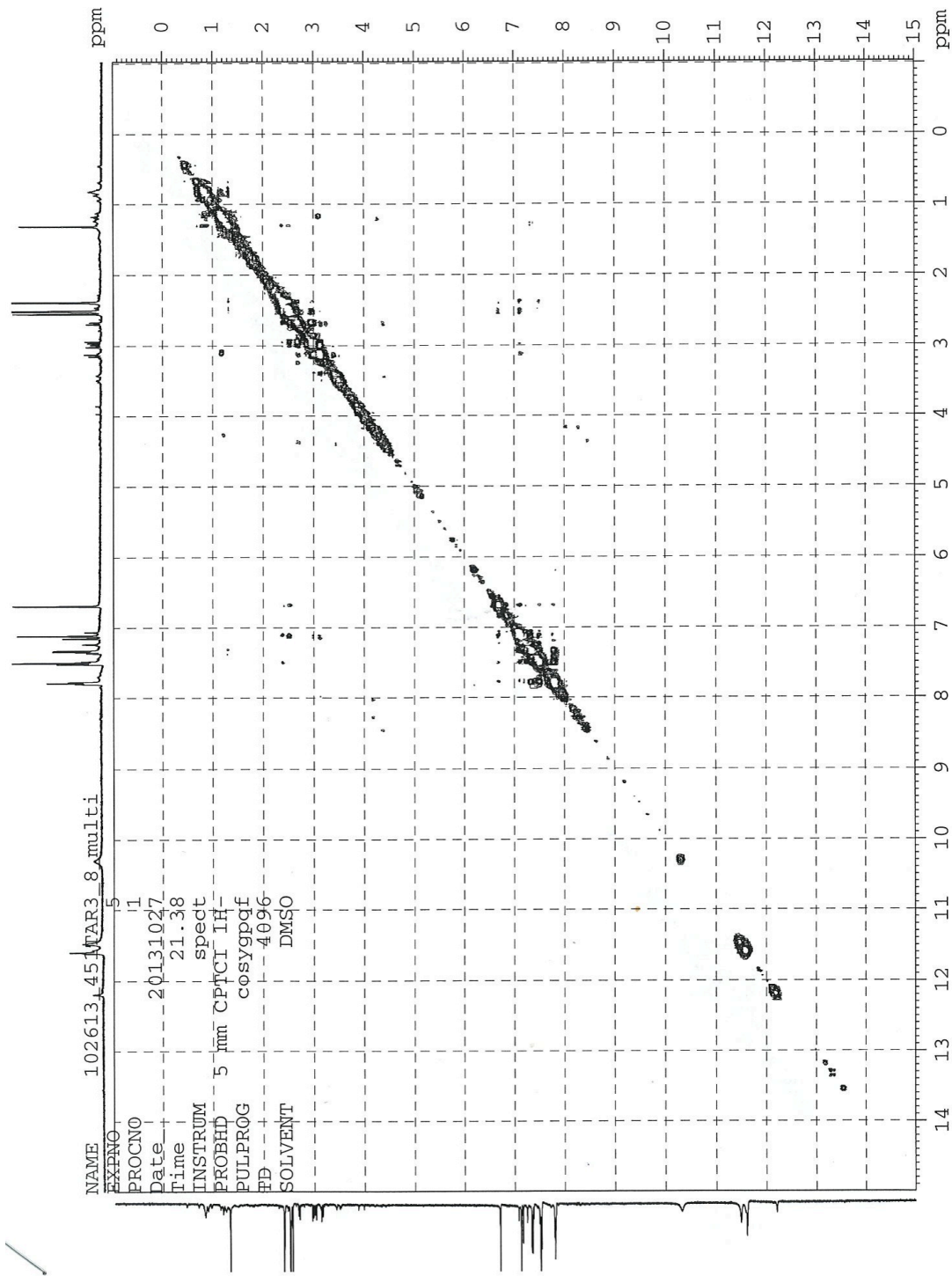


Figure 39. ^1H - ^1H COSY spectrum spectrum of Rabelomycin (2)

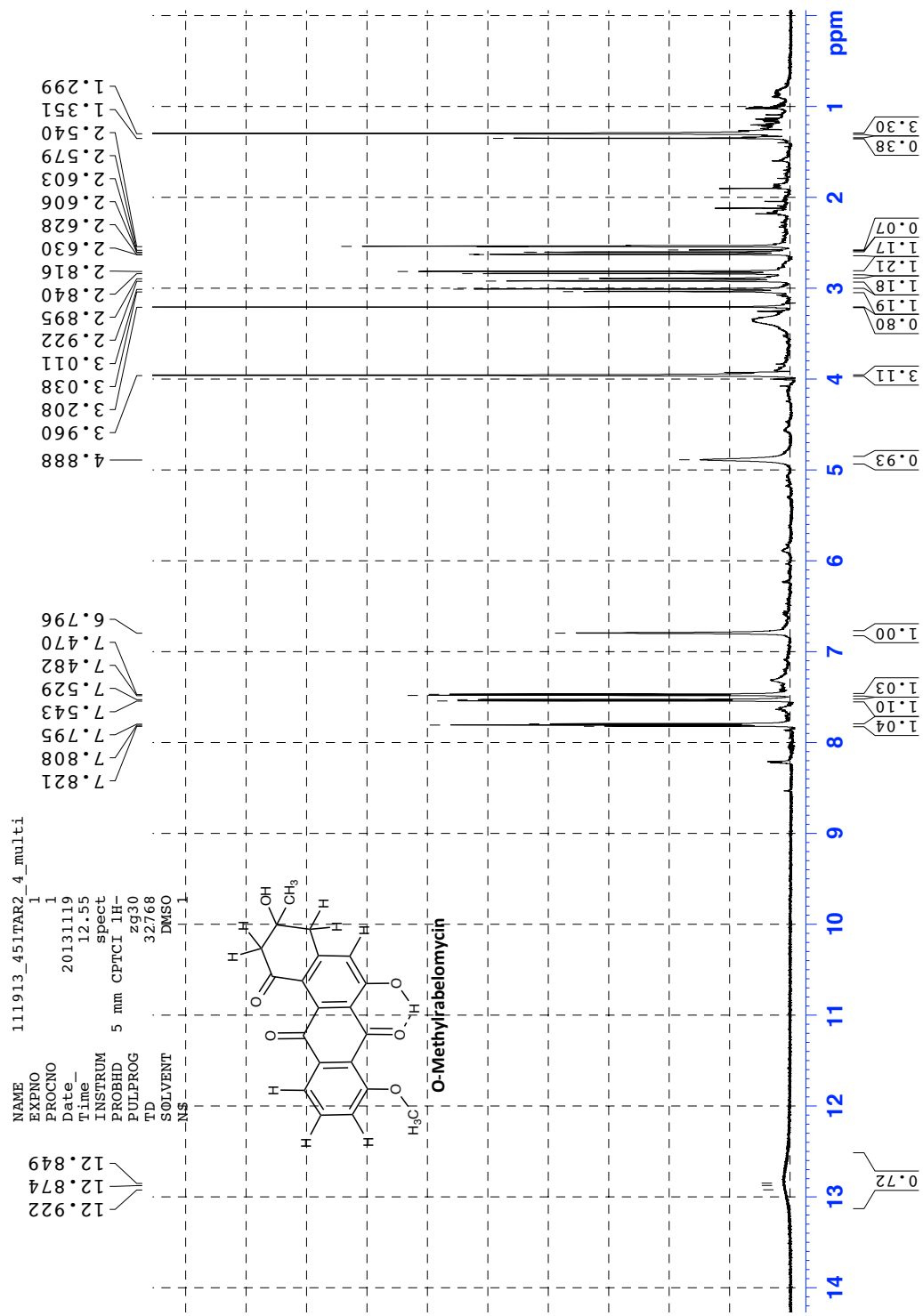


Figure 40. ¹H NMR spectrum of 8-O-Methylrababomycin (3)

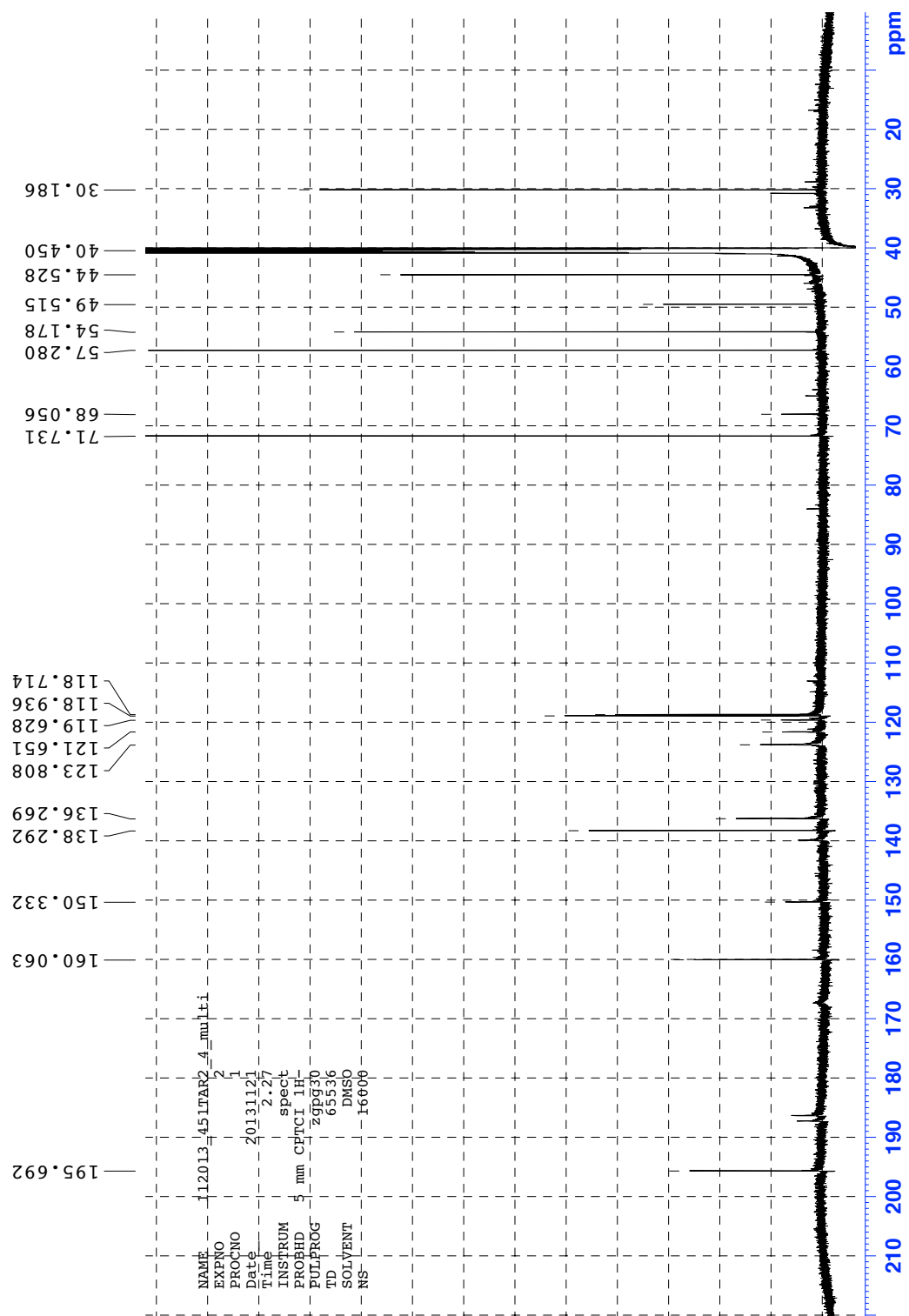


Figure 41. ¹³C NMR spectrum of 8-O-Methylrabelomycin (3)

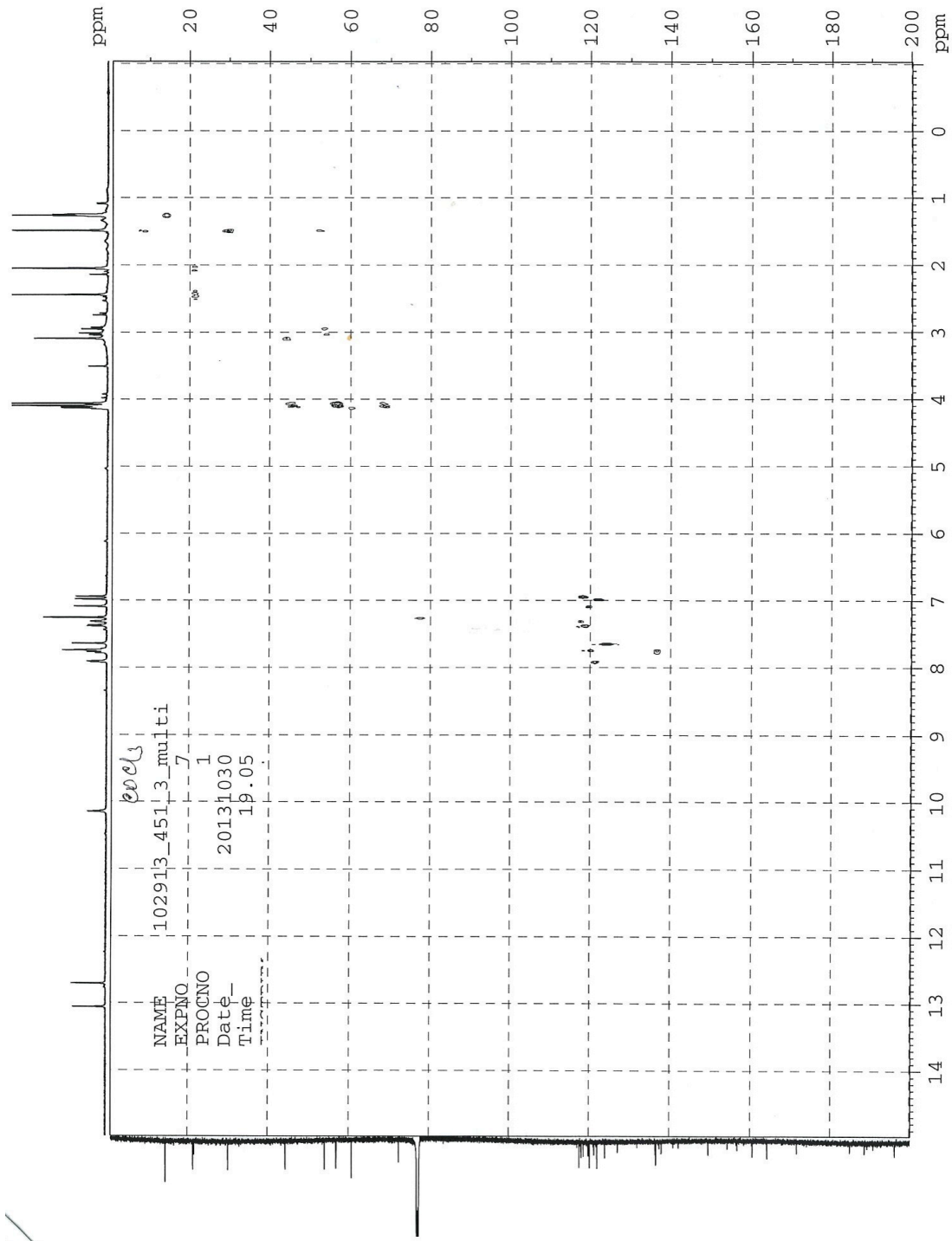


Figure 42. ^1H - ^{13}C HMQC spectrum of 8-*O*-Methylrabelomycin (**3**)

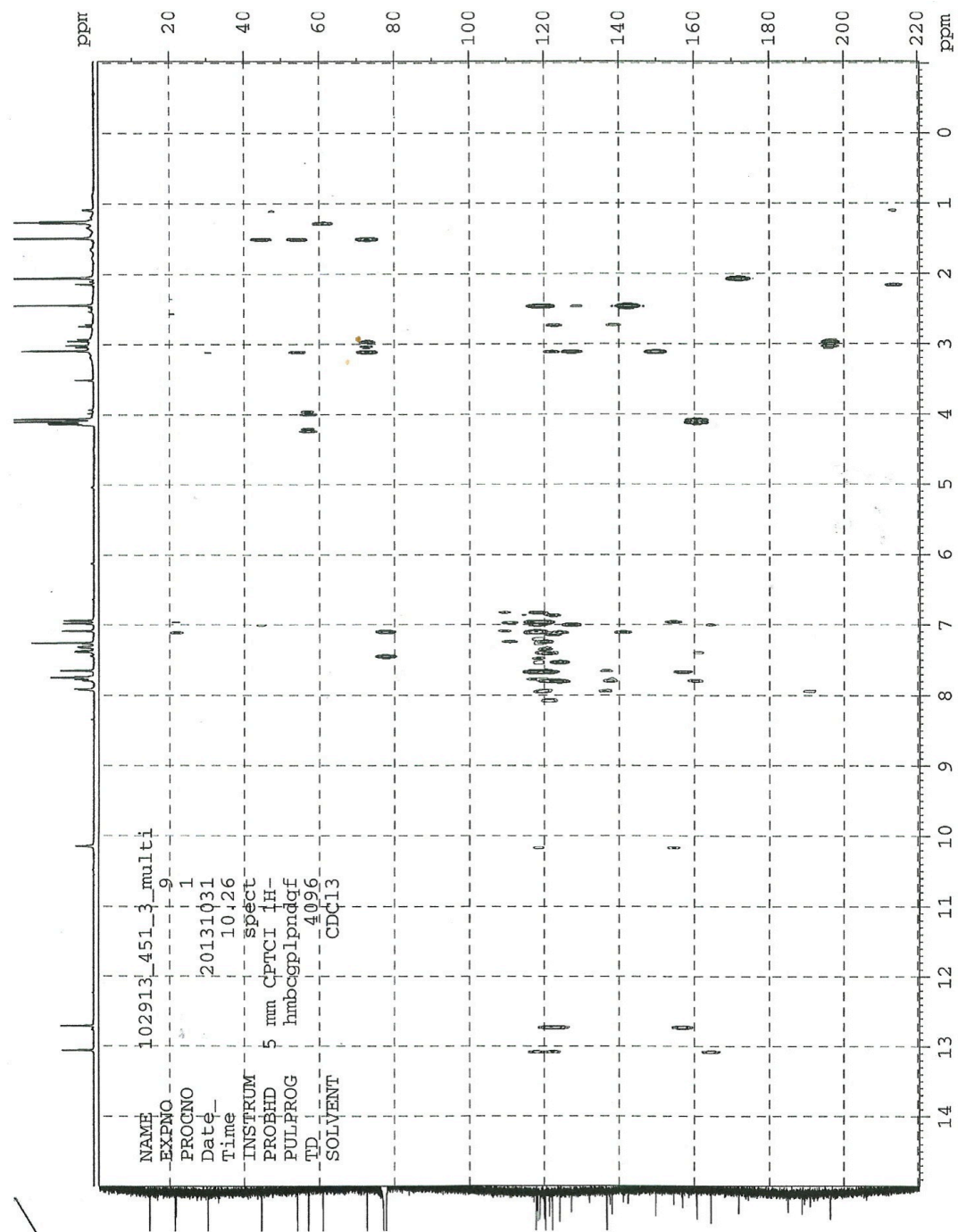


Figure 43. ^1H - ^{13}H HIMBC spectrum of 8-*O*-Methylrabelomycin (3)

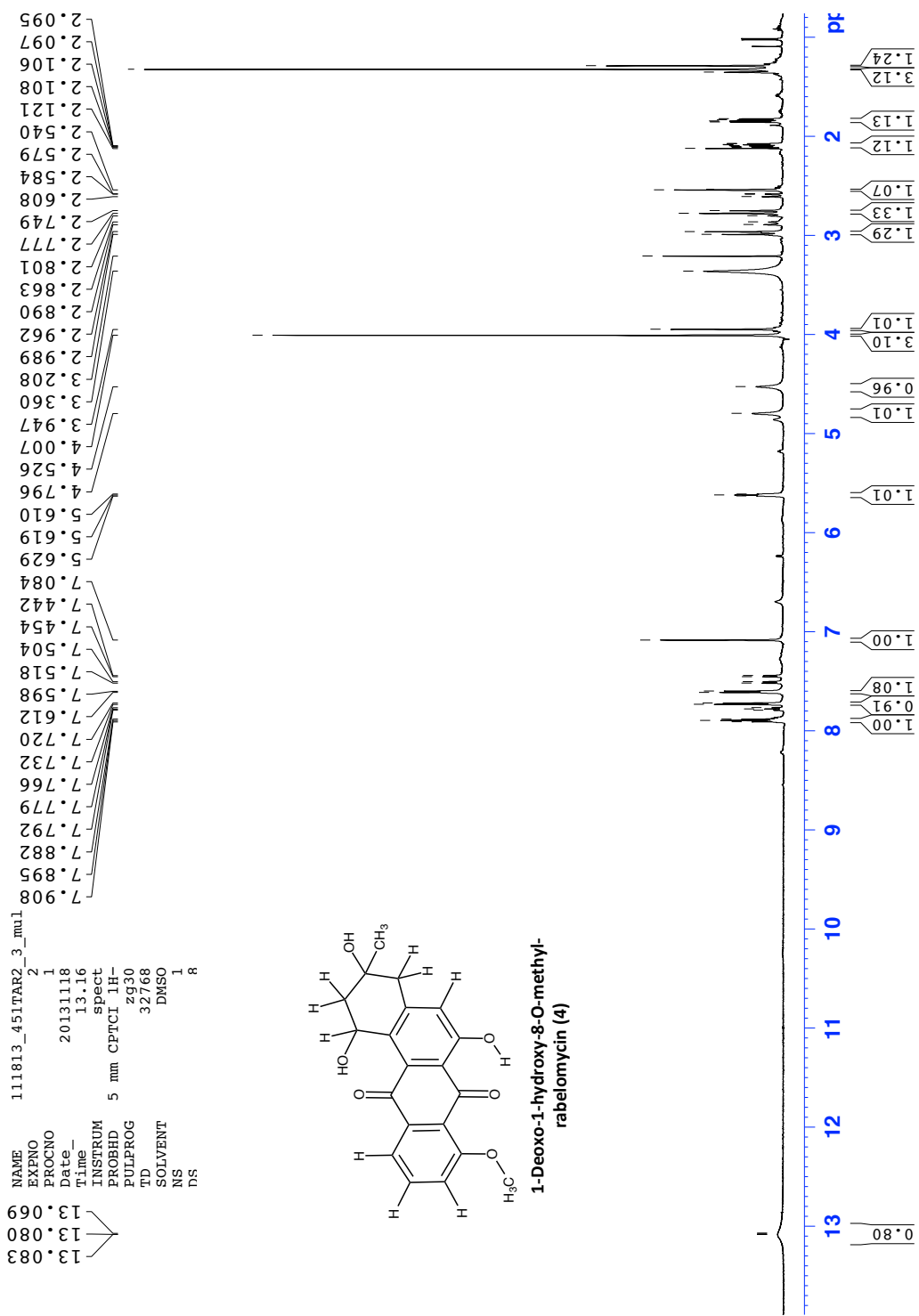


Figure 44. ¹H NMR spectrum of 1-deoxy-1-hydroxy-8-O-methylrabelomycin (4)

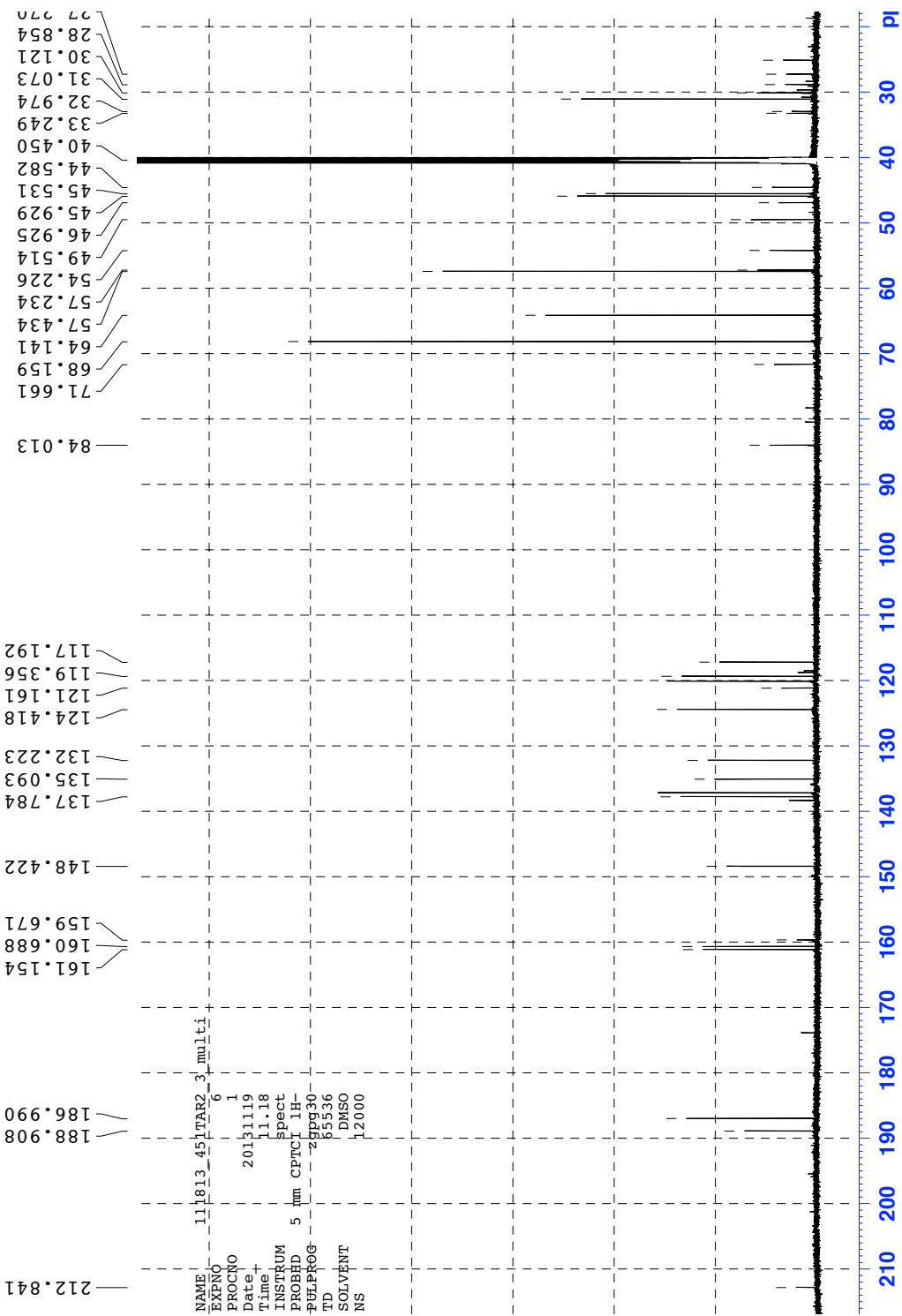


Figure 45. ¹³C NMR spectrum of 1-deoxy-1-hydroxy-8-O-methylrabelomycin (4)

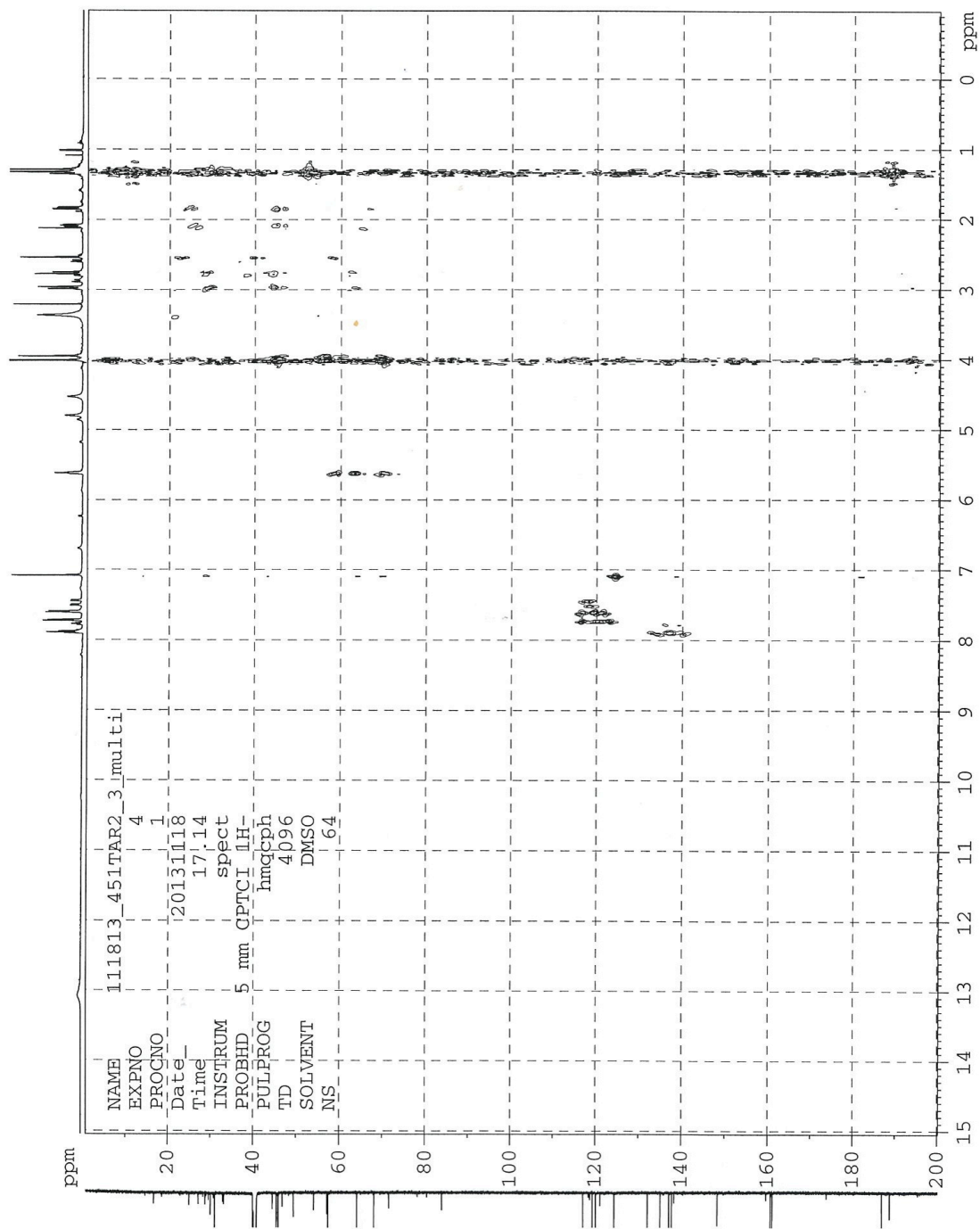


Figure 46. HMQC spectrum of 1-deoxy-1-hydroxy-8-O-methylrabelomycin (4)

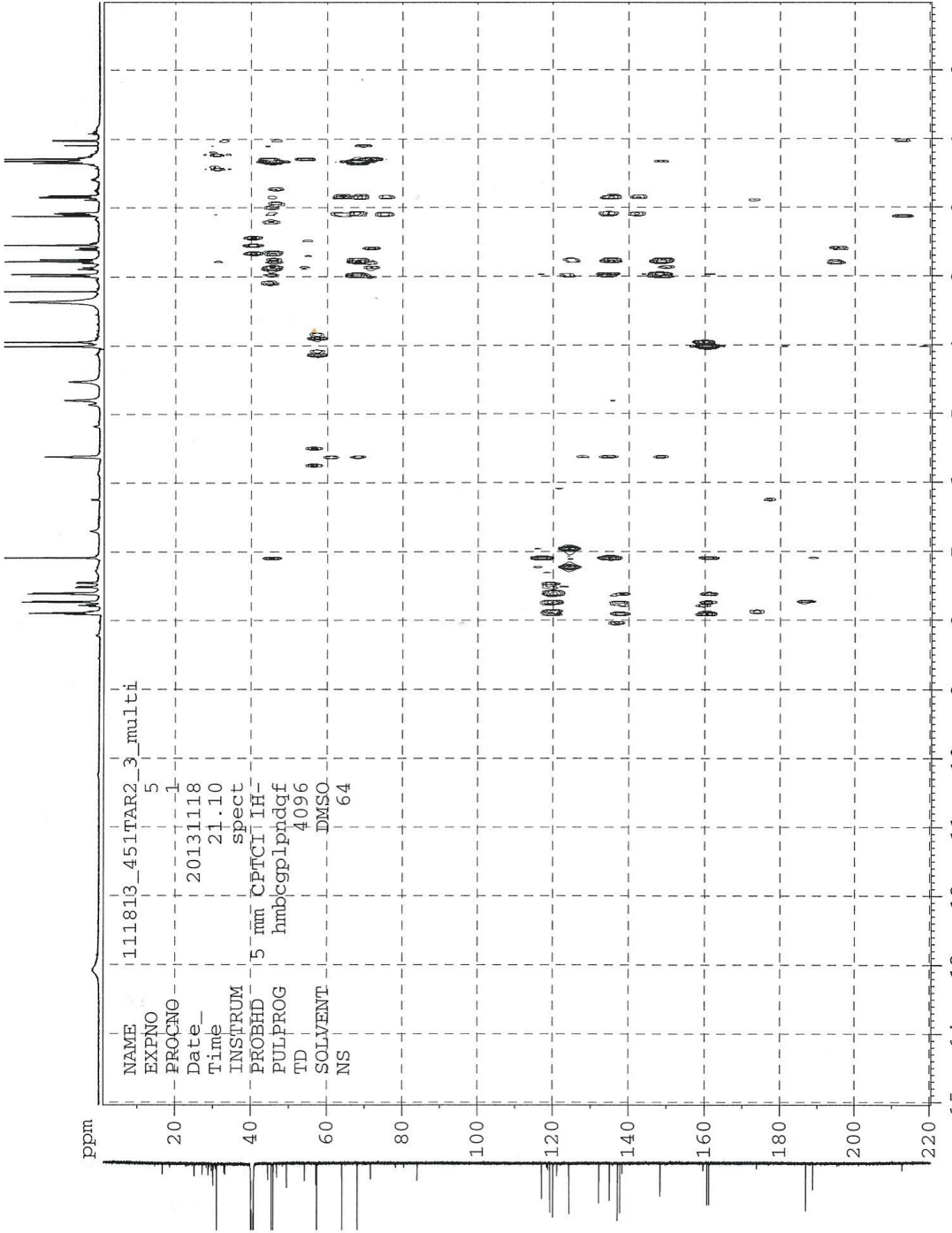


Figure 47. ^1H - ^{13}H HMBC spectrum of 1-deoxy-1-hydroxy-8-*O*-methylrabelomycin (**4**)

CHAPTER 4

Phylogeny-Guided Discovery and Heterologous Expression Studies of Novel Type II PKS Gene Clusters

4.1 Introduction

Soil microbiome is a vast reservoir of genomic diversity and metabolic pathways that are intimately associated with novel natural products. Natural product discovery programs have traditionally relied on random screening of microbial species to identify new chemistry in nature (Cragg and Newman 2013). However, microorganisms are difficult to culture using laboratory-based methods. It is now common knowledge that direct visualization of microorganisms in a natural sample by staining and microscopic studies hints toward a population count one to two orders of magnitude higher than that measured by culturing from that same sample. This has led to an anomaly: while some microorganisms have been extremely well studied, a great majority has not been explored at all. The data from 16S ribosomal RNA (or 16S-rRNA) gene-sequence studies provide probably the strongest support to the notion that our census of microbial populations is far from comprehensive. Therefore, harnessing the biosynthetic capacity of environmental microbial species to discover bioactive chemical compounds is contingent upon overcoming the “culture bias”.

Recent advancements in various next-generation sequencing techniques, bioinformatics analytical tools and DNA cloning methods have caused an

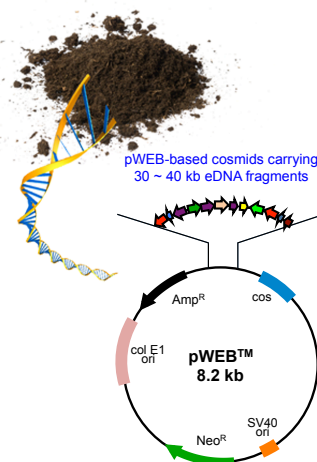


Figure 48. Metagenomic approach to natural product discovery. Microbial DNA is isolated from the environment and cloned into *E.coli* based vector that can be cultured under laboratory conditions.

explosion in the number of studies exploring species diversity of microbial environments (Figure 48). 16S rRNA sequence studies have confirmed that soil is an environment particularly rich in microbial diversity and hence a promising source of novel chemistry (Hugenholtz, Goebel *et al.* 1998). Even soil from environments expected to have low diversity, such as landfills or other contaminated sites, have turned out to be above-average (Nüsslein and Tiedje 1998; Lloyd-Jones and Lau 1998). Therefore, a systematic sequence-based screening of the soil metagenome should enable the functional characterization of diverse biosynthetic pathways that lay hidden in nature (Kallifidas and Brady 2012).

In this chapter, I report the use of a highly conserved biosynthetic gene as a phylogenetic marker to identify novel gene clusters. This approach can serve to offset new natural product discovery. I use short conserved biosynthetic gene sequences (sequence tags) amplified from soil microbial DNA as phylogenetic units to target the discovery of novel aromatic polyphenols (type II polyketides) from the environment. Aromatic polyketides are a structurally diverse collection of metabolites having important antimicrobial and anticancer properties (Hertweck, Luzhetskyy *et al.* 2007). Even though gene clusters encoding biosynthesis of these metabolites can differ substantially in gene content, they all contain a conserved set of three proteins: ketosynthase alpha (KS_{α}) ketosynthase beta/chain length factor (KS_{β}) and acyl carrier protein (ACP) forming a polyketide synthase (PKS) unit known as the minimal PKS. Shotgun-sequencing data as well as PCR-based studies indicate that environmental DNA (eDNA) samples are rich in unique minimal PKS genes (Kang and Brady, 2014; Feng, Kallifidas *et al.* 2011; Pang, Tan *et al.* 2008; Seow *et al.* 1997). Significant structural diversity, both in terms of carbon skeleton and functional groups, coupled with a wide variety of biological activities of this biosynthetic subgroup make aromatic polyphenols viable targets for the phylogeny-based natural product discovery approach.

Our efforts to identify novel polyketide natural products began by cataloguing eDNA-derived type II PKS gene clusters whose ketosynthase beta (KS_{β}) sequence-tag was phylogenetically distinct from any known KS_{β} sequence. In this study, I used polymerase chain reaction (PCR) amplicon sequences (sequence tags) of KS_{β} genes from archived soil eDNA libraries to explore the structural and biosynthetic diversities of aromatic polyphenols. The KS_{β} gene forms a subunit of the minimal PKS that is responsible for generating the non-reduced intermediate in aromatic polyketide biosynthesis (Figure 49). KS_{β} genes have proven to be robust phylogenetic markers that represent structural variations in the carbon skeleton of the natural products derived

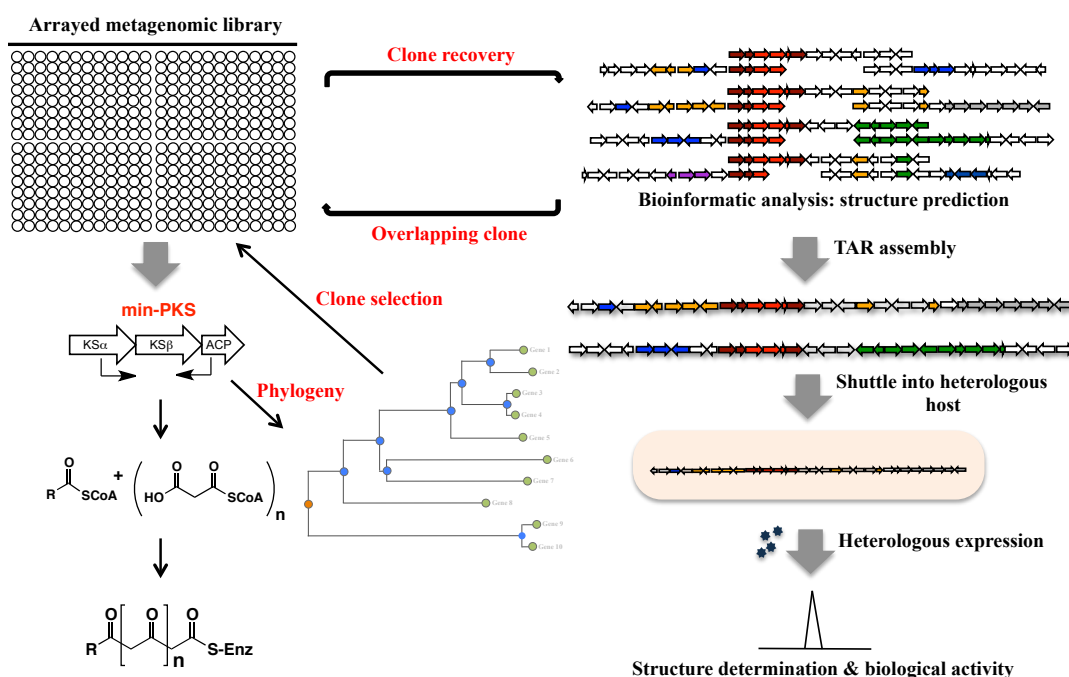


Figure 49. Overview of the phylogeny-guided strategy for the discovery of new natural products. eDNA from soil is used to construct cosmid megalibraries (> 10 000 000 clones), which are arrayed as pools of unique 5000-membered sublibraries. Aromatic polyketide min-PKS KS_{β} genes are amplified in PCR reactions with degenerate primers using DNA from each unique sublibrary within the megalibraries as template. The phylogenetic grouping of eDNA sequences with KS_{β} sequences from gene clusters that encode novel natural products is used to guide the recovery of clones containing gene clusters of interest from specific sublibraries. Heterologous expression of bioinformatically unique eDNA-derived complete gene clusters can be used to expand the structural diversity of natural products.

from type II PKS biosynthetic systems (Winter, Sloman *et al.* 2013; Feng *et al.* 2011; Sullivan 2005; Nakagawa, Omura *et al.* 1987). In other words, this single gene can predict the general class of a type II PKS gene cluster.

The eDNA-derived KS_{β} amplicons were mapped onto a phylogenetic tree constructed using KS_{β} genes from functionally characterized type II PKS clusters. A number of amplicons were found to fall into the well-defined type II PKS clades, such as anthracycline, tetracycline, angucycline, pentangular polyphenol etc. (Kang and Brady 2014; Hertweck 2009). Close inspection of the

KS_{β} gene phylogeny indicated that our soil eDNA libraries contain biosynthetic gene clusters that did not align with any known or functionally characterized type II PKS pathways (Figure 50). These clusters, therefore, could potentially encode for natural products with novel

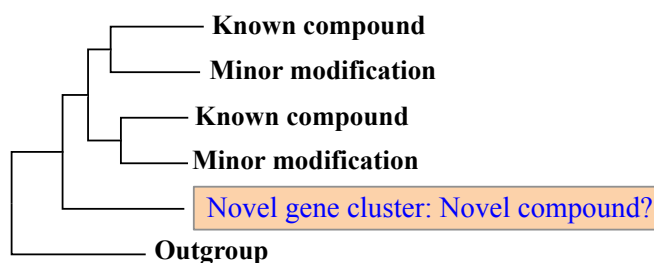


Figure 50. Biosynthetic marker phylogeny. Phylogenetic analysis to identify biosynthetic gene clusters that do not clade with functionally characterized pathways could lead to natural products with novel scaffold.

chemical scaffold and lead to new therapeutics. A selective few of these potentially novel type II PKS biosynthetic clusters were recovered from the archived libraries and carried forward for heterologous expression studies.

4.2 Results

4.2.1 KS_{β} sequence-guided screening of soil eDNA libraries.

To identify novel gene clusters encoding type II PKS, soil eDNA libraries were screened with degenerate primers targeting the KS_{β} gene of the min-PKS. KS_{β} genes clade into groups that

correlate strongly with the specific aromatic polyketide structural classes that are encoded by the gene clusters (Kang and Brady 2013; Feng *et al.* 2011; Ridley, Lee *et al.* 2008). Based on this observation, I surmised that it could be possible to systematically screen multimillion-member eDNA libraries for biosynthetic gene clusters encoding novel members of aromatic polyketide structural classes using PCR-amplified min-PKS gene fragments as sequence tags. Phylogenetic trees were constructed with the resulting eDNA KS β amplicon sequences and sequences from functionally characterized type II PKS gene clusters available from GenBank database. Through this analysis, from 78 sequenced type II PKS clones, I identified a group of KS β sequence tags from our Anza Borrego Desert (AB) and Arizona (AZ) eDNA libraries that were phylogenetically distant from KS β genes from any functionally characterized type II PKS gene clusters (Figures 51, 52). Gene clusters carrying these “unique” KS β genes could potentially encode for novel type II PKS derived metabolites. All these clones were, therefore, recovered from the archived libraries and fully sequenced. Based on phylogenetic comparison of the KS β genes and the gene content of each cluster, I identified three type II PKS pathways that were of particular interest: AB1215, AB1449 and AZ1237 (Figure 52).

4.2.2 Overlapping clone recovery

Sequencing data for all three cosmid clones, AB1215, AB1449 and AZ1237 indicated that the above-mentioned biosynthetic clusters were not completely captured in one cosmid clone. This sequence data was subsequently used to guide the recovery of additional overlapping cosmids in order to reconstruct the original pathways. Sequencing and bioinformatics annotation of each set of overlapping cosmid clones revealed biosynthetic gene clusters flanked by collections of genes predicted to encode primary metabolic enzymes, suggesting that all three gene clusters were recovered in their entirety on the overlapping cosmids.

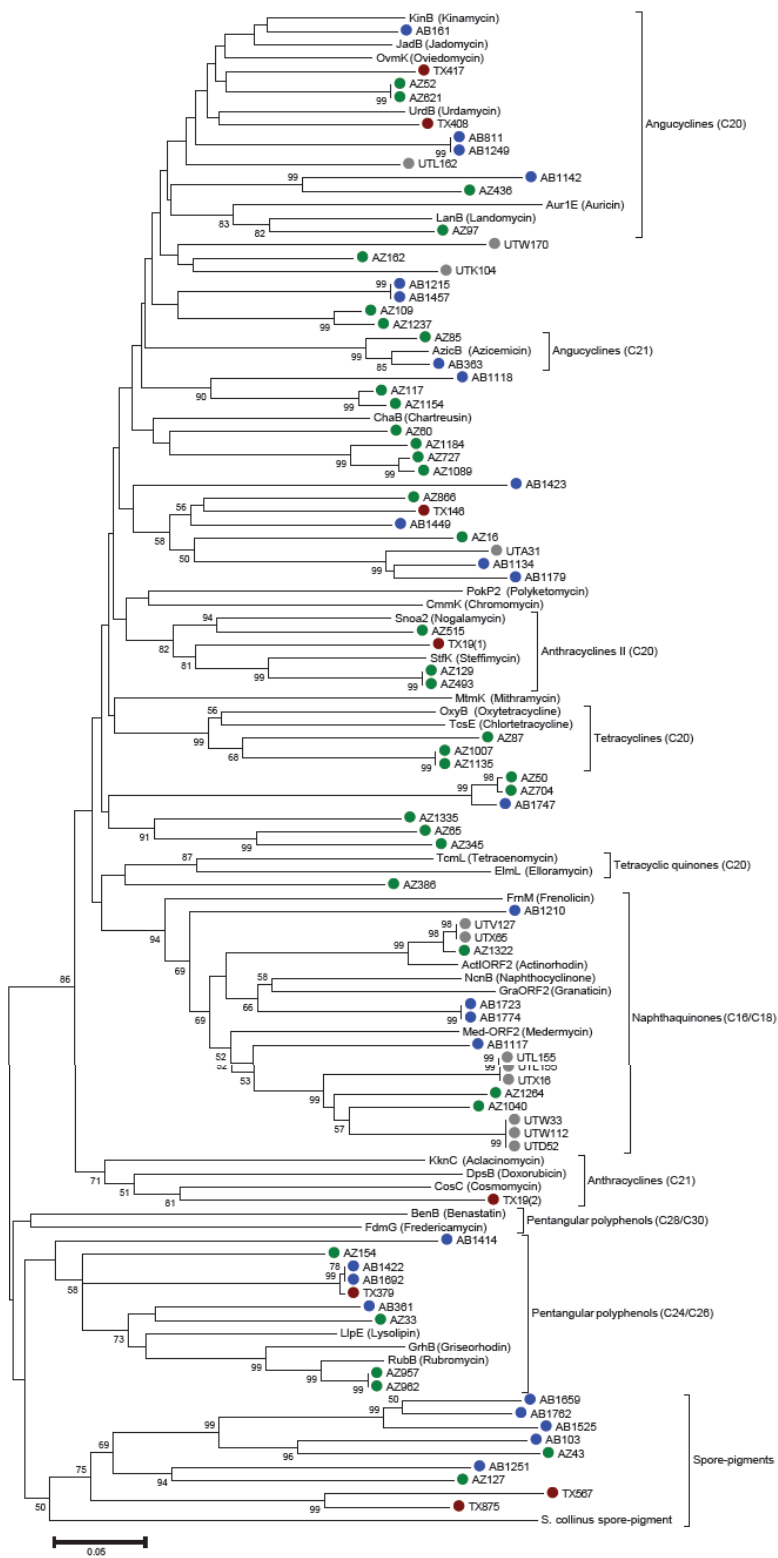


Figure 51. Maximum likelihood tree of Ksβ genes from type II PKS gene clusters. we hypothesized that phylogenetic comparison of eDNA-derived Ksβ sequence tags to Ksβ genes from functionally characterized gene clusters should enable us to select structurally distinct metabolites encoded by eDNA-derived gene clusters. AZ and AB denote amplicon sequences derived from Arizona and California desert soil eDNA libraries, respectively.

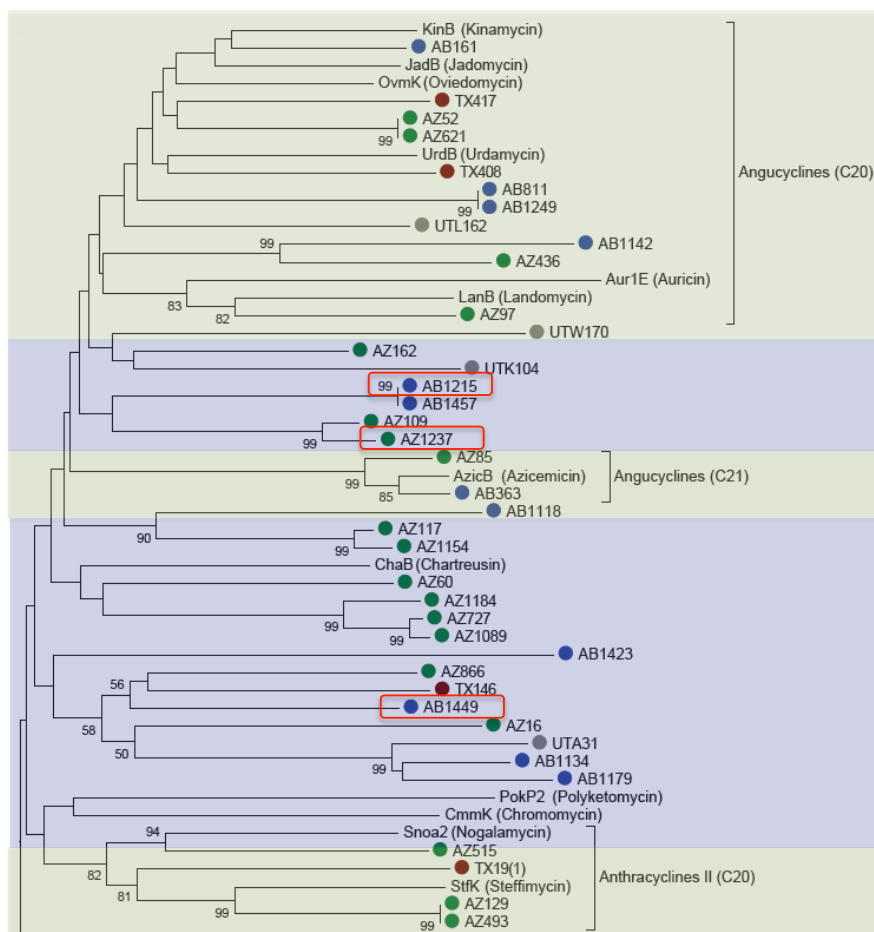


Figure 52. A part of the maximum likelihood tree of partial KS_{β} gene sequences is shown. AZ and AB denote amplicon sequences derived from Arizona and California desert soil eDNA libraries, respectively. The specific group of eDNA sequence tags used in this study are highlighted in red boxes.

4.2.3 Gene cluster analysis of complete AB1215, AB1449 and AZ1237 pathways

Biosynthetic highlights of AB1215/183 pathway.

Antibiotics containing sugar moieties are promising, especially regarding their bioactivity and the engineering of new/ novel analogues through glycodiversification (Kharel *et al.* 2012). Sequence analyses of this biosynthetic gene cluster revealed 8 ORFs that could potentially be involved in the deoxysugar biosynthetic pathways (ORFs 16, 17, 18, 19, 20, 22, 23 and 26) (Figure 53). The cluster also contains genes for sugar aminotransferase. The cluster, therefore, encodes for the biosynthesis of type II PKS derived small molecule with aminodeoxy sugars. In

addition, there is a gene for potential C-glycosylation, which is not very common (Rohr and Thiericke, 1992; Murenets 1990) and may lead to a structurally novel compound (Figure 53).

Due to the importance of the sugar moiety for bioactivity of aromatic polyphenols

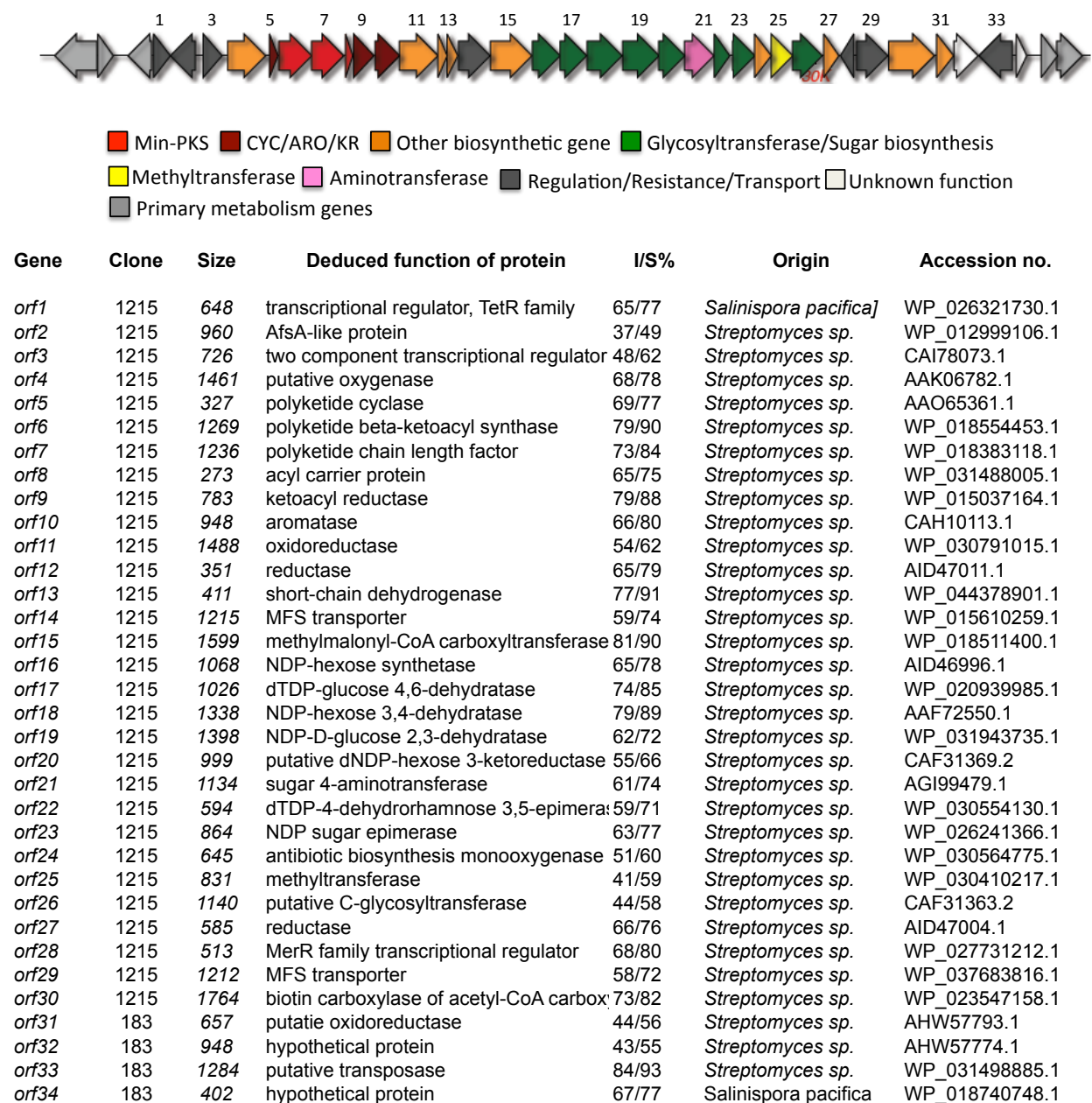
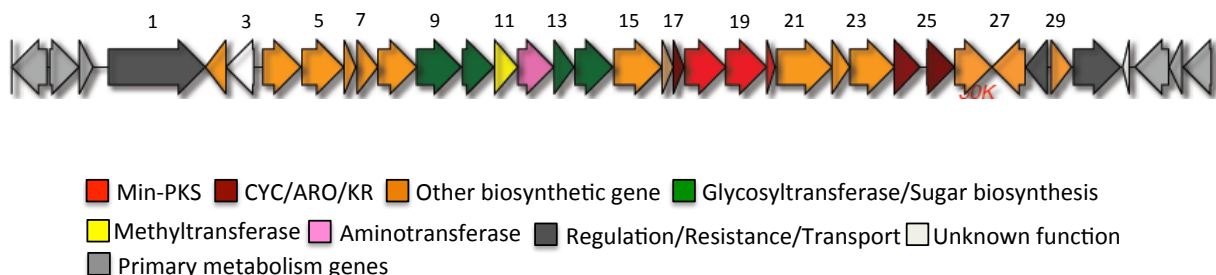


Figure 53. ORF diagram and gene table for Type II PKS pathway AB1215/183. ORFs were predicted using MetaGeneMark and predicted ORFs were annotated by Blast search and Pfam domain analysis. MacVector was used for sequence manipulation.

and the presence of a combination of post-PKS tailoring enzyme encoding genes such as C-glycosyltransferase the sugar aminotransferase found on clone AB1215/183, I selected to study this gene cluster using heterologous expression methods.

Biosynthetic highlights of AB1449/647 pathway.



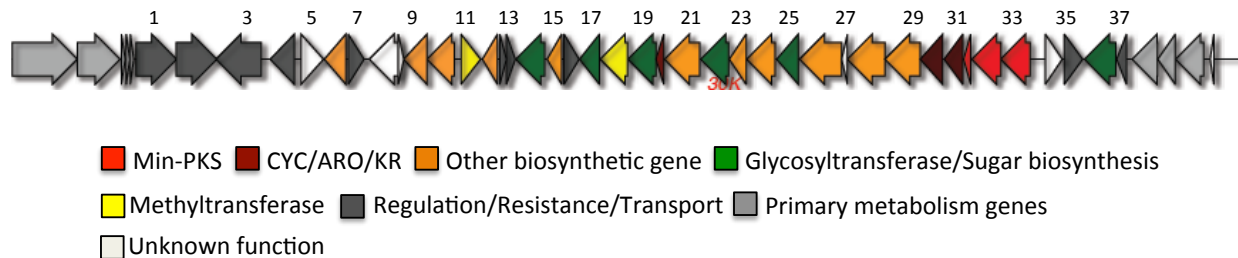
Gene	Clone	Size	Deduced function of protein	I/S%	Origin	Accession no.
<i>orf1</i>	1449	3003	SARP family transcriptional regulator	40/51	<i>Streptomyces</i> sp.	WP_030714058.1
<i>orf2</i>	1449	636	NADH-flavin reductase	54/66	<i>Streptomyces</i> sp.	WP_012891744.1
<i>orf3</i>	1449	867	hypothetical protein	38/47	<i>Streptomyces</i> sp.	AIA07021.1
<i>orf4</i>	1449	1182	cytochrome P450	70/81	<i>Streptomyces</i> sp.	WP_019890099.1
<i>orf5</i>	1449	1496	pyridine nucleotide-disulfide oxidoreductase	61/70	<i>Streptomyces</i> sp.	WP_045557816.1
<i>orf6</i>	1449	357	dehydrogenase	34/46	<i>Streptomyces</i> sp.	WP_028419658.1
<i>orf7</i>	1449	645	C-5 anthrone oxidase	66/84	<i>Streptomyces</i> sp.	AAC98518.1
<i>orf8</i>	1449	1209	cytochrome P450	27/45	<i>Streptomyces</i> sp.	WP_037831154.1
<i>orf9</i>	1449	1380	NDP-hexose 2,3-dehydratase	59/69	<i>Streptomyces</i> sp.	WP_037825067.1
<i>orf10</i>	1449	972	NDP-4-keto-6-deoxyhexose 4-ketoreductase	53/65	<i>Streptomyces</i> sp.	AAL14256.1
<i>orf11</i>	1449	720	methyltransferase	53/67	<i>Streptomyces</i> sp.	CAA05643.1
<i>orf12</i>	1449	1110	NDP-hexose aminotransferase	77/86	<i>Streptomyces</i> sp.	ABV49595.1
<i>orf13</i>	1449	612	dTDP-4-dehydrorhamnose 3,5-epimerase	71/83	<i>Streptomyces</i> sp.	WP_003963466.1
<i>orf14</i>	1449	1182	glycosyltransferase	61/72	<i>Streptomyces</i> sp.	ABO28818.1
<i>orf15</i>	1449	1494	FAD-dependent oxidoreductase	66/74	<i>Streptomyces</i> sp.	CCA59273.1
<i>orf16</i>	1449	348	putative oxygenase	54/56	<i>Streptomyces</i> sp.	CAA60567.1
<i>orf17</i>	1449	327	putative cyclase	79/83	<i>Streptomyces</i> sp.	CAH10118.1
<i>orf18</i>	1449	1269	putative ketosynthase	80/88	<i>Streptomyces</i> sp.	AAK06784.1
<i>orf19</i>	1449	1269	putative chain length factor	69/82	<i>Streptomyces</i> sp.	CBH32808.1
<i>orf20</i>	1449	249	acyl carrier protein	64/76	<i>Streptomyces</i> sp.	AGO50612.1
<i>orf21</i>	1449	1719	acetyl-CoA carboxyltransferase	69/77	<i>Streptomyces</i> sp.	WP_030283707.1
<i>orf22</i>	1449	558	biotin carboxylase carrier protein	67/80	<i>Streptomyces</i> sp.	AAG30192.1
<i>orf23</i>	1449	1389	pyruvate carboxylase subunit A	63/75	<i>Streptomyces</i> sp.	WP_046930293.1
<i>orf24</i>	1449	795	ketoacyl reductase	76/86	<i>Streptomyces</i> sp.	WP_031121806.1
<i>orf25</i>	647	852	putative aromatase	63/74	<i>Streptomyces</i> sp.	BAJ07855.1
<i>orf26</i>	647	1149	FAD-binding monooxygenase	62/72	<i>Streptomyces</i> sp.	WP_033205650.1
<i>orf27</i>	647	1026	aldo/keto reductase	69/80	<i>Streptomyces</i> sp.	WP_046709454.1
<i>orf28</i>	647	666	TetR family transcriptional regulator	61/72	<i>Streptomyces</i> sp.	WP_033175331.1
<i>orf29</i>	647	630	FMN reductase	67/78	<i>Streptomyces</i> sp.	WP_012895574.1
<i>orf30</i>	647	1458	MFS transporter	63/75	<i>Actinomodura madurae</i>	WP_033331410.1

Figure 54. ORF diagram and gene table for Type II PKS pathway AB1449/647. ORFs were predicted using MetaGeneMark and predicted ORFs were annotated by Blast search and Pfam domain analysis. MacVector was used for sequence manipulation.

Analysis of this biosynthetic gene cluster indicated that there are 4 ORFs that could potentially be involved in the biosynthesis of deoxysugar (ORFs 9, 10, 13 and 14) (Figure 54). In addition, there is putative hexose aminotransferase potentially leading to aminodeoxy sugar biosynthesis. There is a putative C-5 anthrone oxidase (ORF 7) gene in this cluster. This gene is often responsible for Baeyer-Villiger type oxidative cleavage of a ring system followed by rearrangements of the carbon scaffold of aromatic polyketides generating novel carbon motif. Similar genes are also found in certain angucycline gene clusters (GilOI, JadH and UrdM) (Li and Piel 2002). This cluster encodes for enzymes that can carry out additional post-PKS tailoring reactions, including the *O*-methylation, which is promising with regards to structural diversification (Figure 54). Based on the above analysis, I chose to carry forward this AB1449/647 type II PKS pathway for heterologous expression studies.

Biosynthetic highlights of AZ2250/1237/2224 pathway.

As mentioned earlier, glycosylation is one of the most important tailoring steps in aromatic polyketide biosynthesis. In this particular pathway, there are 5 ORFs (ORFs 14, 17, 19, 22 and 36) that could potentially lead to glycosylated natural product (Figure 55). Three of these genes are actually putative glycosyltransferases. Therefore, this pathway could produce an aromatic polyphenol with oligosaccharide attachment. Additionally, the presence of a sugar *O*-methyltransferase can impart further structural diversification (Figure 55). The above features made this AZ2250/1237/2224 cluster of substantial interest to us, and it was taken forward for heterologous expression analysis.



Gene	Clone	Size	Deduced function of protein	I/S%	Origin	Accession no.
<i>orf1</i>	2250	1698	ABC transporter ATP-binding protein	46/59	<i>Streptomyces</i> sp.	WP_030261462.1
<i>orf2</i>	2250	1695	ABC transporter ATP-binding protein	43/59	<i>Kutzneria albida</i>	WP_025354879.1
<i>orf3</i>	2250	1902	SARP family transcription regulator	49/62	<i>Kutzneria albida</i>	AHH93931.1
<i>orf4</i>	2250	1014	ArsR family transcriptional regulator	53/67	<i>Amycolatopsis decaplanina</i>	WP_007030015.1
<i>orf5</i>	2250	945	hypothetical protein	85/89	<i>Amycolatopsis rifamycinica</i>	WP_043782961.1
<i>orf6</i>	2250	903	short-chain dehydrogenase	94/96	<i>Streptomyces</i> sp.	WP_037683050.1
<i>orf7</i>	2250	666	TetR family transcriptional regulator	94/97	<i>Streptomyces</i> sp.	WP_044575537.1
<i>orf8</i>	2250	1254	hypothetical protein	43/60	<i>Streptomyces</i> sp.	WP_030677244.1
<i>orf9</i>	2250	984	alcohol dehydrogenase	34/51	<i>Ktedonobacter racemifer</i>	EFH82381.1
<i>orf10</i>	2250	1149	oxidoreductase	54/64	<i>Streptomyces</i> sp.	WP_014182277.1
<i>orf11</i>	2250	825	sugar O-methyltransferase	59/72	<i>Streptomyces</i> sp.	AAG42854.1
<i>orf12</i>	2250	633	polyketide synthase	41/57	<i>Streptomyces</i> sp.	WP_019662161.1
<i>orf13</i>	1237	651	TetR family transcriptional regulator	55/70	<i>Streptomyces</i> sp.	WP_033175331.1
<i>orf14</i>	1237	1209	glycosyltransferase	56/67	<i>Streptomyces</i> sp.	AAF00217.1
<i>orf15</i>	1237	612	FMN reductase	67/79	<i>Streptomyces</i> sp.	WP_012895574.1
<i>orf16</i>	1237	636	TetR family transcriptional regulator	54/68	<i>Streptomyces</i> sp.	WP_033175331.1
<i>orf17</i>	1237	840	NAD-dependent epimerase	55/67	<i>Streptomyces</i> sp.	WP_009714477.1
<i>orf18</i>	1237	1146	sugar O-methyltransferase	42/58	<i>Streptomyces</i> sp.	ACQ63636.1
<i>orf19</i>	1237	1179	glycosyltransferase	44/56	<i>Streptomyces</i> sp.	ABO28818.1
<i>orf20</i>	1237	327	putative polyketide cyclase	81/89	<i>Streptomyces</i> sp.	AAO65345.1
<i>orf21</i>	1237	1476	monooxygenase	71/78	<i>Streptomyces</i> sp.	WP_031488013.1
<i>orf22</i>	1237	1164	glycosyltransferase	63/71	<i>Streptomyces</i> sp.	ABO28818.1
<i>orf23</i>	1237	708	anthrone monooxygenase	66/78	<i>Streptomyces</i> sp.	WP_041993338.1
<i>orf24</i>	1237	1194	cytochrome P450	26/42	<i>Streptomyces</i> sp.	WP_030882035.1
<i>orf25</i>	1237	906	putative dTDP-1-glucose synthase	71/80	<i>Streptomyces</i> sp.	AAF73461.1
<i>orf26</i>	1237	1761	acetyl-CoA carboxylase	74/83	<i>Streptomyces</i> sp.	WP_028959689.1
<i>orf27</i>	1237	1584	methylmalonyl-CoA carboxyltransferase	81/90	<i>Streptomyces</i> sp.	WP_030624360.1
<i>orf28</i>	1237	1491	bifunctional oxygenase-reductase protein	66/72	<i>Streptomyces</i> sp.	AAT64432.1
<i>orf29</i>	1237	945	cyclase	68/78	<i>Streptomyces</i> sp.	WP_030683252.1
<i>orf30</i>	1237	786	polyketide ketoreductase	74/85	<i>Streptomyces</i> sp.	WP_015037164.1
<i>orf31</i>	1237	297	acyl carrier protein	62/77	<i>Streptomyces</i> sp.	WP_031943729.1
<i>orf32</i>	1237	1218	chain length factor	78/86	<i>Streptomyces</i> sp.	WP_018383118.1
<i>orf33</i>	1237	1233	ketoacyl synthase	81/88	<i>Streptomyces</i> sp.	WP_017237282.1
<i>orf34</i>	1237	711	hypothetical protein	49/57	<i>Streptomyces</i> sp.	AHW57770.1
<i>orf35</i>	1237	771	SARP-family transcriptional regulator	40/56	<i>Streptomyces</i> sp.	WP_020939953.1
<i>orf36</i>	1237	1311	UDP-glucose 4-epimerase	44/59	<i>Streptomyces</i> sp.	WP_037813783.1
<i>orf37</i>	1237	408	transposase	46/58	<i>Streptomyces</i> sp.	WP_045563702.1

Figure 55. ORF diagram and gene table for *Type II PKS pathway AZ2250/1237/2224*. ORFs were predicted using MetaGeneMark and predicted ORFs were annotated by Blast search and Pfam domain analysis. MacVector was used for sequence manipulation.

4.2.4 TAR cloning for reconstruction type II PKS pathways AB1215, AB1449 and AZ1237

It has been shown that transformation associated recombination (TAR) in *Saccharomyces*

cerevisiae could be used to reconstruct large continuous stretches of DNA captured on overlapping cosmid clones (Kim *et al.* 2010). To reassemble AB1215 and AB183, AB1449 and AB647, and AZ2225, AZ1237 and AZ2250 into single DNA fragments that could be used in heterologous expression studies, cosmids and their corresponding specific yeast capture vector were co-transformed into *S. cerevisiae* and allowed to recombine (Figure 56). The bacterial artificial chromosomes (BACs) constructed in this experiment were completely sequenced and found to be faithful reassemblies of the overlapping eDNA sequences. These newly assembled large insert clones were conjugated into *S. albus* and examined for the production of clone specific metabolites.

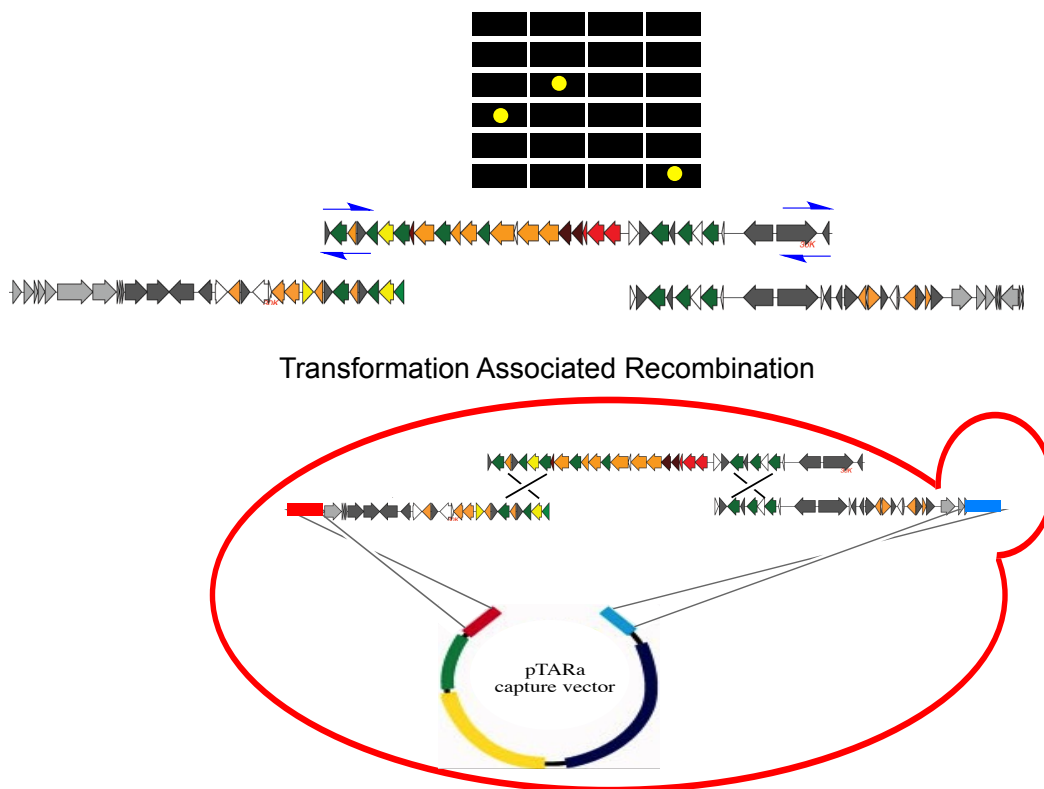


Figure 56. Transformation Associated Reassembly (TAR) of overlapping eDNA clones in yeast. Overlapping cosmid clones for the sequence of interest were screened for and recovered from eDNA library and assembled into one bacterial artificial chromosome (BAC) clone.

4.2.5 Heterologous expression of the eDNA-derived gene clusters

In order to examine the AB1215, AB1449 and AZ1237 pathways for production of metabolites encoded within the clusters, each set of three overlapping eDNA cosmid clones (AB1215, AB183; AB1449, AB647; AZ2250, AZ1237, AZ2224) was assembled into bacterial artificial chromosomes (BAC-AB1215/183, BAC-AB1449/647 and BAC-AZ22250/1237/2250) using pathway specific pTARa (*E. coli*:yeast:*Streptomyces*) shuttle capture vectors and transformation-associated recombination (TAR) in yeast (Kim *et al.* 2010) (Figure 56). These BAC constructs were sequenced to ensure correct assembly and subsequently conjugated into *Streptomyces albus* by intergenic conjugation and integrated into the *Streptomyces* genome using the Φ C31 integrase.

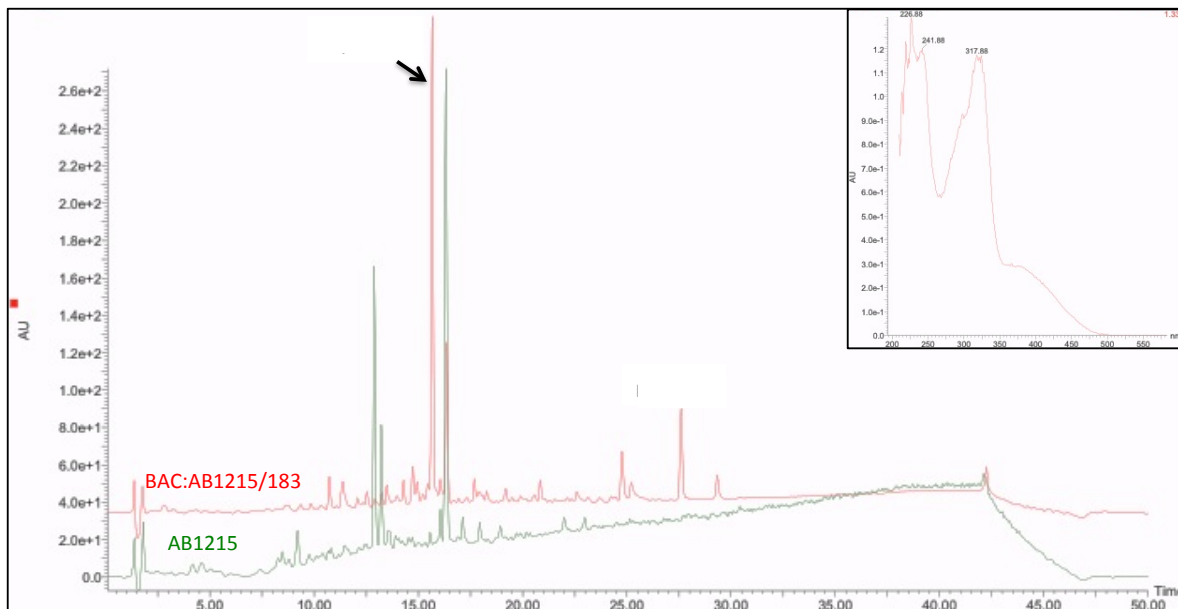


Figure 57. *S. albus* transformed with eDNA-derived type II PKS pathway BAC-AB1215/183 was screened for the production of clone specific metabolites. LC-MS analysis of crude ethyl acetate extract of *Streptomyces* culture broth showed one major clone specific peak with m/z 761. No clone specific peaks were observed on LCMS analysis with *S. albus* transformed with cosmid AB1215. UV signature is shown for the major clone specific metabolite.

For heterologous expression, the resulting strains, *S. albus* BAC-AB1215/183, *S. albus* BAC-AB1449/647 and *S. albus* BAC-AZ22250/1237/2250, were grown in R5A liquid media for 10 days at 30 °C. LC-MS analysis of ethylacetate extracts derived from these cultures showed the presence of one major clone-specific peak for BAC-AB1215/183 and BAC-AB1449/647 each (Figure 57 and 58, respectively). BAC-AZ22250/1237/2250 showed peaks corresponding to three clone specific metabolites (Figure 59).

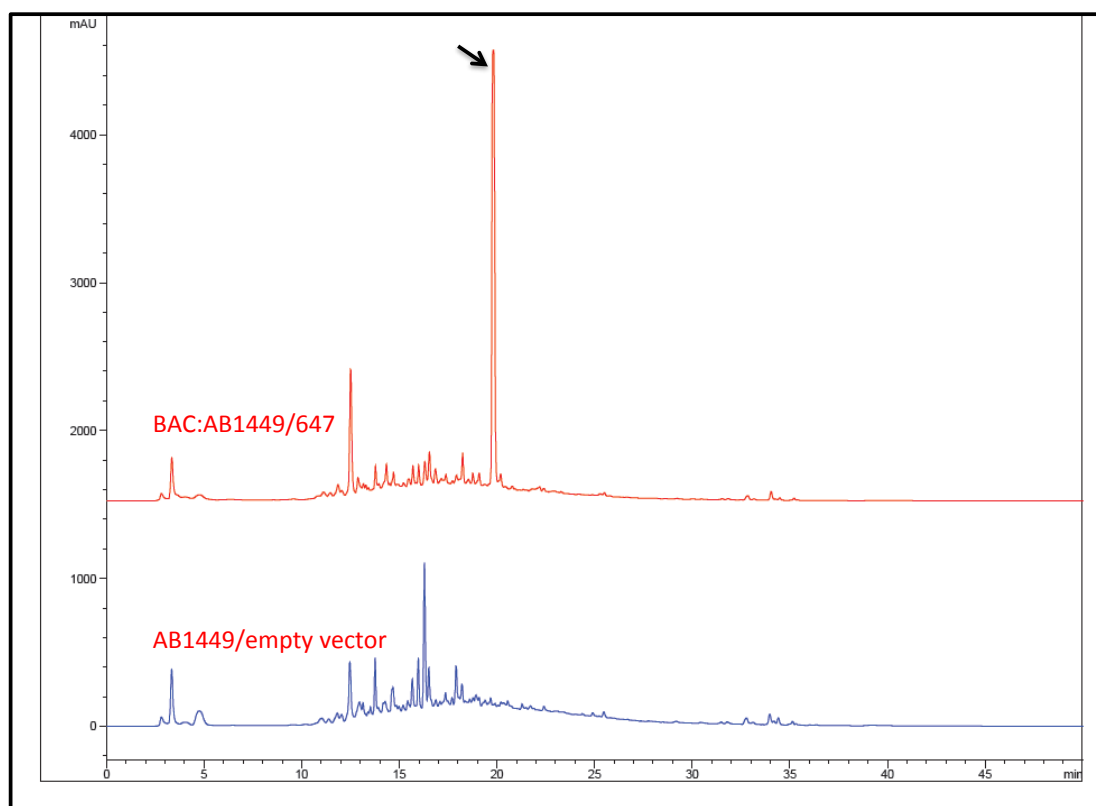


Figure 58. *S. albus* transformed with eDNA-derived type II PKS pathway BAC-AB1449/647 was screened for the production of clone specific metabolites. LC-MS analysis of crude ethyl acetate extract of *Streptomyces* culture broth showed one major clone specific peak. No clone specific peaks were observed on LCMS analysis with *S. albus* transformed with cosmid AB1449.

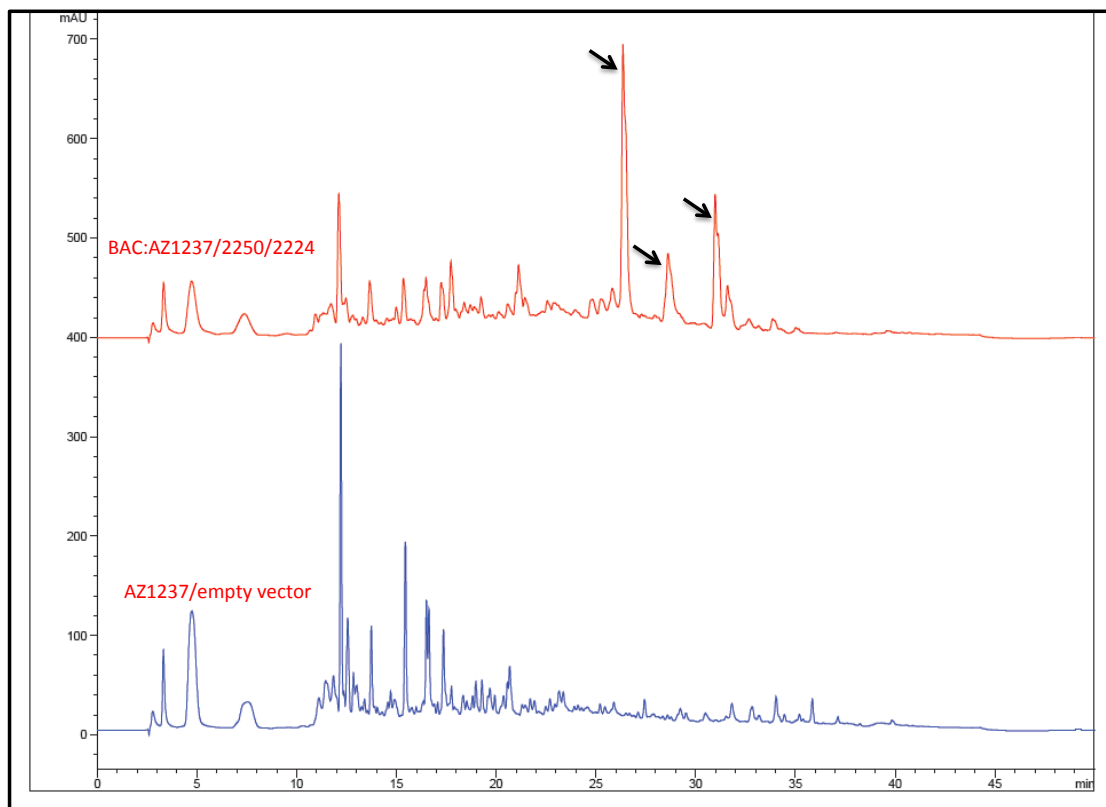


Figure 59. *S. albus* transformed with eDNA-derived type II PKS pathway BAC-AZ22250/1237/2250 was screened for the production of clone specific metabolites. LC-MS analysis of crude ethyl acetate extract of *Streptomyces* culture broth showed one major clone specific peak. A control was carried out using *S. albus* transformed with cosmid AZ1237 which did not show any clone specific peaks on LCMS analysis.

4.2.6 Complementation of deoxysugar biosynthesis

Extensive sequence analysis of the three type II PKS pathways revealed that AB449/647 and AZ2250/1237/2224 are missing key genes that are required to initiate sugar biosynthesis. The AZ2250/1237/2224 pathway has multiple glycosyltransferase genes, but it lacks two critical genes required for deoxysugar biosynthesis: dNDP-4-keto-6-deoxyglucose 3,5-epimerase and dNDP-4-ketohexulose reductase, while the AB449/647 pathway is missing the dNDP-D-glucose synthase gene. Therefore, in all likelihood, pathways AB449/647 and AZ2250/1237/2224 are producing just the aglycone moiety and the final glycosylated product encoded in the pathway.

These sugar biosynthesis genes appear in the primary metabolic background in some, but not all, *Streptomyces sp.* (Rodriguez, Menzella *et al.* 2009). The early genes required for deoxysugar biosynthesis were therefore provided *in trans* (pIJRham). The rhamnose sugar biosynthetic operon (oleL, oleS, oleE, and oleU) (Rodriguez, Oelkers *et al.* 2000) from oleandomycin biosynthesis was cloned under the control of the constitutive ermE* promoter within the conjugative integrative shuttle expression vector pIJ10257 (Hong, Hutchings *et al.* 2005). The pIJRham construct was generated by Dr. Haak-Soo Kang. The pIJRham construct was subsequently transformed into *E. coli S17.1* and conjugated into *S.albus/AB1449/647* and *S.albus/AZ2250/1237/22224* to yield *S.albus/AB1449/647/pIJRham* and *S.albus/AZ2250/1237/22224/pIJRham*, respectively.

4.3 Discussion

Soil metagenome is a rich resource for the discovery of novel natural products. In this chapter, I demonstrate that it is possible to identify collections of natural product gene clusters that encode novel metabolites using the sequence tag-based metagenomic screening approach. I show that sequence tags falling in between two well-defined clades associated with different structural classes can help identify novel biosynthetic gene clusters.

On the basis of the differences in min-PKS gene phylogeny, I shortlisted a collection of type II PKS gene clusters whose the KS β genes clade differently from any known existing biosynthetic clusters of the same family of natural products. The individual proteins that constitute the min-PKS (KS α , KS β , and ACP) are highly conserved across all type II PKS gene clusters. KS β phylogenetic tree correlates closely with differences in the core polyketide structure (carbon

chain length and cyclization pattern) encoded by the gene cluster from which the min-PKS arises (Khosla 2008). KS_{β} genes have thus proven to be useful phylogenetic markers for predicting differences in polyketide core structures encoded by type II PKS gene clusters (Wawrik, Kerkhof *et al.* 2005).

I carried out our phylogeny-based bioinformatics analysis on KS_{β} genes from 78 type II PKS gene clusters from our archived megalibraries constructed using Anza Borrego Desert soil from California (AB library) and Arizona (AZ library). Our phylogenetic analysis led to the identification of three unique eDNA encoded biosynthetic pathways, AB1215, AB1449 and AZ1237, containing Type II PKS gene clusters that do not clade with any known or functionally characterized pathways. They could therefore encode for and produce molecules with novel chemical scaffolds. These cosmid-based clones were fully sequenced and analysis of this sequencing data indicated that these gene clusters were only partially captured. Incomplete biosynthetic pathways either remain silent during heterologous expression studies, or they produce intermediate compounds or shunt products from the pathway. It was therefore essential that I recover each pathway in its entirety. It's been shown that an eDNA library of 10 to 15 million clones is likely to contain collections of clones encompassing complete PKS gene clusters and permit the recovery of overlapping clones from biosynthetic gene clusters found in the genomes of uncultured bacteria (Kim *et al.* 2010). I then identified, from the same archived libraries (consisting of $\sim 10^7$ clones) where these above mentioned pathways were originally found, multiple overlapping gene clusters and reconstructed these gene clusters on bacterial artificial chromosome (BAC) clones using Transformation Associated Recombination (TAR) in yeast (Kim *et al.* 2010).

Type II PKS pathway AB1215 and AB1449 were recovered on two overlapping cosmids, AB1215, AB183 and AB1449, AB647, respectively. AZ1237 was fully captured on three different cosmids, AZ2250, AZ1237, AZ2224. All the overlapping fragments were sequenced and the open reading frames (ORFs) of each pathway were assessed using bioinformatics tools. This analysis confirmed that these clusters were quite unique in their gene content. All three pathways encode for deoxysugar biosynthesis, which make them promising, both from structural and bioactivity perspective. AB1215/183 also contain genes for sugar aminotransferase and C-glycosylation, making it an interesting pathway to pursue. AB1449/647 again contains aminodeoxysugar biosynthesis genes. In addition, there is a putative C-5 anthrone oxidase gene in this cluster which can effect oxidative rearrangements and lead to a unique carbon skeleton. Pathway AZ2250/1237/2224 encodes for multiple glycosyltransferases, which can potentially lead to a unique aromatic polyphenol natural product with oligosaccharide side chain.

Based on both our phylogenetic analysis and ORF predictions, these three pathways were reassembled into large insert BAC clones using TAR cloning in yeast and subsequently examined for clone specific metabolite production in heterologous expression host, *S. albus*. BAC-AB1215/183 and BAC-AB1449/647 each conferred the production of one clone-specific metabolite in *Streptomyces*, whereas three clone specific metabolites were identified from heterologous expression of BAC-AZ22250/1237/2250 pathway. Characterization of these metabolites would lead to further information regarding the biosynthetic potential of these gene clusters.

Close inspection of the three type II PKS pathways indicated that pathways AB449/647 and

AZ2250/1237/2224 are missing key genes that are required to initiate the deoxysugar biosynthesis. Therefore, it is likely that pathways AB1449/647 and AZ2250/1237/2224 are producing just the aglycone moiety and not the final glycosylated product encoded in the pathway. The critical sugar biosynthesis genes were then provided *in trans* using pIJRham that contains the rhamnose sugar biosynthetic operon. Systematic comparative analysis of the metabolites produced by *S. albus* exconjugants harboring the BAC clones with and without the pIJRham construct should lead to definitive answers. Work is underway to better understand the metabolites produced by these pathways.

The work presented in this chapter demonstrates that one can rapidly access untapped chemical diversity of highly complex environmental microbiomes using single gene sequence tags. This methodology is scalable using high throughput sequencing technology and large eDNA libraries. The current study strengthens the idea that sequence-based eDNA mining approaches like those used in here can overcome the traditional culture-based natural product discovery paradigm.

4.4 Materials and Methods

4.4.1 Bioinformatic analysis of KS_{β} sequence tags

PCR amplicons were gel purified (Qiagen) and Sanger-sequenced (Genewiz). Approximately 600bp of KS_{β} gene fragments from each amplicon were aligned using ClustalW. Phylogenetic analysis and pairwise distant calculations were performed using MEGA5.1 (Tamura, Peterson *et al.* 2011). All PCR was performed using BioRad Tetrad thermocycler.

4.4.2 Clone recovery and bioinformatics analysis

eDNA clone AB1215, AB1449 and AZ1237 were recovered from the AB (Anza Borrego) and AZ eDNA libraries by serial dilution of the 5,000-member sub-pool from which the amplicons were originally amplified. For this, an overnight culture of each sub-pool was inoculated into 96-well microtiter plates at a dilution of 10^{-6} . After 18 h of shaking at 37 °C, the diluted cultures were screened by whole-cell PCR using the KS_{β} screening primers (dp:KS-TTCGGSGGITTCCAGWSIGCSATG and dp:ACP-TCSAKSAGSGCSAISGASTCGTAICC) (Seow *et al.* 1997) and touchdown PCR protocol. PCR positive wells were plated onto solid media (LB agar) to yield distinct colonies that were screened individually in a second round of whole-cell PCR that was conducted using the following touchdown protocol: denaturation (95 °C, 2 min), 8 touchdown cycles [95 °C, 45 s; 65 °C (−1 °C per cycle), 1 min; 72 °C, 1 min], 35 standard cycles (95 °C, 45 s; 58 °C, 1 min; 72 °C, 1 min) and a final extension step (72 °C, 2 min). Each 25 μ L PCR reaction contained 50 ng of cosmid DNA, 2.5 μ M of each primer, 2 mM dNTPs, 1X ThermoPol reaction buffer (New England Biolab), 0.5 units Taq DNA polymerase and 5% DMSO.

4.4.3 Cosmid retrofitting and integration into Streptomyces albus

Recovered cosmids were digested with the rare PstI to ensure that only the pWEB-TNC vector was cut and not the eDNA insert. Each linearized cosmid was then ligated with the 6.81 kb DraI fragment from pOJ436. This fragment contains an origin of transfer (oriT), an apramycin resistance marker, and elements of the phiC31 attP-attB integration system needed for integration into diverse Streptomyces sp. Retrofitted cosmids were transformed into *E. coli* S17.1 and conjugated into Streptomyces using published protocols (Feng, Khallifidas *et al.* 2011).

4.4.4 Overlapping clone recovery and bioinformatic analysis

Anza Borrego (California) and AZ (Arizona) desert soil eDNA libraries were screened for overlapping clones for AB1215, AB1449 and AZ1237. These were subsequently identified from the corresponding library using clone-specific primers designed to recognize the sequence at one or both ends of the original pathways. The PCR protocol used was as follows: denaturation (95 °C, 5 min), 36 standard cycles (95 °C, 30 s; 55 °C, 30 s; 72 °C, 40s), and a final extension step (72 °C, 7 min). Overlapping eDNA cosmids were recovered from these pools using the serial dilution method outlined above. The following primer sets were used:

AB1215 gene cluster

One overlapping clone for AB1215 pathway was identified in the AB library 183 sub-pool using the following clone-specific primers designed to recognize the sequence at one end of cosAB1215.

AB183_FW: GGTCGGTTCTGGTGTTGG

AB183_REV: CATGTAGTGGCGGTGGTTC

AB1449 gene cluster

One overlapping clone for AB1449 pathway was identified in the AB library 647 sub-pool using the following clone-specific primers designed to recognize the sequence at one end of cosAB1449.

AB647_FW: GATCAACGAGGCGAACATCG

AB647_REV: TACTGCGGCACCTTCCAGAA

AZ1237 gene cluster

Two overlapping clones for AZ1237 pathway was identified in the AZ library subpools 2224 and 2250 using the following clone-specific primers designed to recognize the sequence at both end of cosAZ1237.

AZ2224_FW: TCTGCGGGCATGGATGTAA

AZ2224_REV: GCTTGGCCCATCGAACTCAT

AZ2250_FW: CAGCTTCTCCTCCTCGAGTT

AZ2250_REV: GGACCACTGGAGGTGTTCAA

All cosmids were sequenced by 454-pyrosequencing. ORFs (open reading frames) were predicted using MetaGeneMark (Zhu, Lomsadze, and Borodovsky 2010) and predicted ORFs were annotated by Blast search and Pfam domain analysis.

4.4.5 eDNA clone sequencing

PKS containing cosmids recovered from the eDNA libraries were initially sequenced by 454 pyrosequencing (Roche) at the Memorial Sloan Kettering Cancer Center Genomics Core Laboratory. All overlapping cosmids and TAR constructs were sequenced using IonTorrent Personal Genome Machine (PGM). The size and deduced role of the open reading frames (ORFs) were predicted using MetaGeneMark (Zhu et al. 2010), and predicted ORFs were annotated by Blast search and Pfam domain analysis. MacVector was used for sequence manipulation.

4.4.6 Transformation-associated recombination (TAR) reassembly of three overlapping eDNA clones

Pathway-specific capture vector construction.

The overlapping eDNA cosmid clones were assembled into a single bacterial artificial chromosome (BAC) clone using transformation-associated recombination (TAR) in yeast and the pTARa BAC shuttle vector. The pathway-specific TAR capture vector was constructed using InFusion cloning methodology (Clontech). ~500 bp upstream and downstream homology arms were amplified from each cosmid. The homology arms needed for TAR recombination were generated using the following primers:

AB1215 gene cluster (AB1215 and AB189)

AB1215UPS_FW (5'- CTATCGATCTCGAGGGTGGTGGACATCGACCTGAT-3'),

AB1215UPS_RV (5'- ATGACCGATGTTAACCGTTAATTTCTCGCACCGG-3'),

AB183DWS_FW (5'-GTTAACATCGGTCATCGGTCGAATGC- 3'),

AB183DWS_RV (5'- CCCTGCAGGAGCTCGATTTCCAGGCTCAGCTCGTC-3')

AB1449 gene cluster (AB1449 and AB647)

AB1449UPS_FW (5'- CTATCGATCTCGAGGTTGGAAACCAGCCGGATGAC-3'),

AB1449UPS_RV (5'- GACCCGGATGTTAACCGTCCCTCAAGCCGTTCTAA-3'),

AB647DWS_FW (5'-GTTAACATCCGGGTCAACTCCTCGAT- 3'),

AB647DWS_RV (5'- CCCTGCAGGAGCTCGCATCTGTTACGCCAGCCAGA-3')

AZ1237 gene cluster (AZ2250, AZ1237 and AZ2224)

AZ2250UPS_FW (5'- CTATCGATCTCGAGGGCTGCGGCAATGAACATGAT-3'),

AZ2250UPS_RV (5'- CAGTCCTCTGTTAACGCCTGGACCAACAACCCAT-3'),

AZ2224DWS_FW (5'-GTTAACAGAGGACTGTTCGTGCTGAC- 3'),

AZ2224DWS_RV (5'- CCCTGCAGGAGCTCGATGGACAGTTCGGTGGTGTC-3')

TAR cloning and pathway assembly

TAR primers include 15 bp sequences (underlined) that overlap with BmtI/SphI linearized pTARa capture vector. Primer ending with UPS_RV was designed to contain a 15 bp overlap

(underlined) with the primer ending with DWS_FW primer and an HpaI site (bold), which was added to facilitate the subsequent linearization of the pathway-specific pTARa capture vector. Gel-purified amplicons (Qiagen) and BmtI/SphI linearized pTARa were combined in a standard InFusion cloning reaction (Clontech) to make the pathway-specific capture vector.

For TAR assembly, 200 ng of DraI-digested cosmids were mixed with 100 ng of the HpaI-cut pathway-specific capture vector and then transformed into 200 μ L of *Saccharomyces cerevisiae* CRY1-2 spheroplasts prepared according to published protocols (Feng, Khallifidas *et al.* 2011). Transformed spheroplasts were mixed with synthetic complete (SC) top agar (1 M sorbitol, 1.92 g/L synthetic complete uracil dropout supplement, 6.7 g/L yeast nitrogen base, 2% glucose, and 2.5% agar) and overlaid onto SC uracil dropout plates. Plates were incubated at 30 °C for 72 h. DNA was prepared from 12 yeast colonies using a zymolyase lysis protocol (ZYMO RESEARCH) and screened by PCR with primers designed to recognize sequences from individual cosmids. A BAC clone that was “PCR positive” with all the primer sets was electroporated into *E. coli* EPI300 (Epicentre). DNA isolated from the resulting *E. coli* was fully sequenced by IonTorrent Personal Genome Machine (PGM) and confirmed to be faithful reassemblies of the overlapping cosmid clones.

4.4.7 Heterologous expression and analysis

The cosmid clones (AB1215, AB1449 and AZ1237) as well as the BAC constructs (BAC-AB1215/183, BAC-AB1449/647 and BAC-AZ2250/1237/2224) were transformed into *E. coli* *SI7.1* and conjugated into *Streptomyces albus*. Spore stocks of *S. albus* transformed with the cosmid or BAC clone was inoculated in 125 mL baffled flasks containing 50 mL of R5A media

(Zhu *et al.* 2010) (for 1 L: 100 g sucrose, 0.25 g K₂SO₄, 10.12 g MgCl₂•6H₂O, 10 g glucose, 0.1 g casamino acids, 5 g yeast extract, 21 g MOPS, 2 g NaOH, 2 mL R2YE trace elements) and grown at 30 °C (200 rpm) for 7-10 d. Cultures were then extracted twice with an equal volume of ethyl acetate. The dried extracts were dissolved in methanol (2-4 mL/L of culture). The sample was then dried, redissolved in 1-2 mL of methanol. The resulting ethyl acetate extract was compared by LC-MS to the extract from a similarly grown culture of *S. albus* transformed with a pTARa empty vector control (HPLC conditions: 30 min gradient from 5% to 100% aqueous acetonitrile containing 0.1% formic acid, C18, 4.6 mm × 150 mm, 1 mL/min).

4.4.8 Cloning of the rhamnose biosynthetic genes

pIJRham expression vector containing the rhamnose biosynthetic operon (oleL: dNDP-4-keto-6-deoxyglucose 3,5-epimerase, oleS: dNDP-D-glucose synthase, oleE: dNDP-glucose 4,6-dehydratase and oleU: dNDP-4-ketohexulose reductase) was transformed into *E. coli* S17.1 and then shuttled into *S. albus*/BAC-AB1215/183, *S. albus*/BAC-AB1449/647 and *S. albus*/BAC-AZ22250/1237/2224 by intergenic conjugation. This expression vector utilizes an orthogonal phage integration system (Φ BT1) and resistance gene (hygromycin) to those present on the BAC clones thereby allowing the co-integration of both the biosynthetic gene cluster and the rhamnose biosynthetic operon into distinct chromosomal sites in *S. albus* to generate *S. albus*/BAC-AB1215/183/pIJRham, *S. albus*/BAC-AB1449/647/pIJRham and *S. albus*/BAC-AZ22250/1237/2224/pIJRham.

CHAPTER 5

Future Directions

5.1 Better bioinformatics

Considering the recent explosion in the number of studies taking advantage of next-generation sequencing to explore metagenomic or taxonomic diversity of environmental microbiome, we need better tools to best interpret the data. As of now, interpretation of sequencing data mostly relies on sequence similarity, on the premise that near identical or similar DNA sequences translate to identical or similar protein function. Although this general paradigm is valid, as many studies have demonstrated, our current knowledge and annotations of protein-coding genes is less than comprehensive. A large number of genes in newly sequenced microorganisms cannot even be identified or annotated (Roberts, 2004). This impedes our ability to use metagenomics to study microbial genome diversity as well as functional analysis thereof. Although efforts are being made to improve the coverage of known genome-derived proteins, we need new technology to link experimental data to large families of similar proteins. The bias in the genome and protein databases is one major hindrance to the use of existing annotations. For example, a large proportion of organisms being sequenced come from the medical community.

Another major concern is that different annotation pipelines potentially produce different results. Simple processes like comparative genomics routinely require re-analysis of all data involved in the comparison (Dinsdale *et al.*, 2008). Attempts to create a simple exchange language for microbial genome analysis have not been successful. With large volumes of data, re-analysis is

not feasible because data analysis cost will exceed the sequencing cost. Metagenomic analysis databases need to expand and be better curated to have a sustained annotation infrastructure.

5.2 Large insert metagenomic libraries

A major hurdle to natural product discovery through metagenomics is the recovery of complete biosynthetic gene clusters from the environment. Natural product biosynthetic gene clusters can vary in size from a few kilobases to more than 100 kilobases. We have successfully gained access to some of the natural products encoded in the genomes of uncultured bacteria using metagenomic techniques. However, the inability to construct very large eDNA libraries with inserts big enough to capture large biosynthetic pathways on individual clones, has been limiting. If individual clones could capture >100 kilobases of contiguous eDNA, we would have much better chance of finding novel chemistry in nature. An obvious solution to this issue is to construct libraries with larger insert clones that can be used to capture large natural product gene clusters. It's now routine to construct of 30–40 kb insert cosmid libraries from environmental DNA samples. However, making libraries of large insert clones has been challenging. Bacterial artificial chromosomes or BACs have been used to construct large insert DNA libraries. However, BAC-based metagenomic libraries are generally two to three orders of magnitude smaller than those constructed using cosmid-based cloning strategies (Liles, Williamson *et al.* 2008). It was therefore assumed that the overall coverage had to be compromised in order to have larger contiguous fragments of DNA. However, advancements in BAC cloning have led to the construction of libraries containing more than one million clones (Magbanua, Ozkan *et al.* 2011). Advent of such technology will inevitably revolutionize the field of metagenomics as a natural product discovery platform.

5.3 Activation of silent biosynthetic gene clusters

The ongoing exponential growth of genome sequencing data has led to the discovery of many natural product biosynthesis pathways for which no actual molecule has been characterized. The biosynthesis of natural products is highly regulated and these clusters often remain silent under laboratory conditions (Jensen, Chavarria *et al.* 2014). The regulation is conducted through dozens of pleiotropic regulatory genes and pathway-specific regulators (Bibb and Hesketh 2009). They form a complex network in response to a variety of physiological and environmental cues. The conditions under which a given gene cluster is naturally expressed are largely unknown. A number of technologies based on genetic engineering are available to induce silent genes. Most of the successful approaches have been based on generation of gene “knock outs”, promoter exchange, overexpression of transcription factors or other pleiotropic regulators (Kallifidas and Brady 2012; Scherlach and Hertweck 2009). Using engineered nucleases is another important new approach to regulate gene expression in prokaryotes through transcriptional control. Clustered, regularly interspaced, short palindromic repeat (CRISPR) technology is pioneering this approach in combination with RNA-guided nucleases, such as Cas9, to edit an organism’s genome with customizable specificities. Genome editing mediated by CRISPR-Cas9 systems has been successfully exploited to achieve programmable transcription activation and repression of bacterial gene expression (Bikard, Jiang *et al.* 2013). In a recent report from Brady lab, yeast homologous recombination have been successfully used to exchange all native promoters in silent gene clusters with constitutively active promoters (Montiel, Kang *et al.* 2015) which led to the induction of an otherwise silent pathway. Strategies based on epigenetics has opened new avenues for the elucidation of the regulation of secondary metabolite formation and will certainly continue to play a significant role for the elucidation of cryptic natural products (Brakhage

2013). Then there are simulation strategies that attempt to recreate natural habitat by co-cultivation of microorganisms from the same ecosystem (Bertrand, Bohni *et al.* 2014). Although a number of such strategies are being explored to activate silent gene clusters, we are in need of a robust technology that will allow detailed studies across multiple different types of biosynthetic gene clusters and provide access to the chemistry encoded therein.

REFERENCES

Abdelfattah M, Maskey RP, Asolkar RN, Grün-Wollny I, Laatsch H. (2003). "Seitomycin: isolation, structure elucidation and biological activity of a new angucycline antibiotic from a terrestrial Streptomycete." *J. Antibiot. (Tokyo)*. **56**(6):539-42.

Albers CA, Lunter G, MacArthur DG, McVean G, Ouwehand WH, Durbin R. (2011). "Dindel: accurate indel calls from short-read data." *Genome Res.* **21**(6):961-73.

Banik JJ, Brady SF. (2008). "Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary." *Proc. Natl. Acad. Sci. USA* 105(45):17273-7.

Banik JJ, Craig JW, Calle PY, Brady SF. (2010). "Tailoring enzyme-rich environmental DNA clones: a source of enzymes for generating libraries of unnatural natural products." *J. Am. Chem. Soc.* **132**(44):15661-70.

Bibb M, Hesketh A. (2009). "Analyzing the regulation of antibiotic production in streptomycetes." *Methods Enzymol.* **458**:93-116.

Bierman M, Logan R, O'Brien K, Seno ET, Rao RN, Schoner BE. (1992). "Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp." *Gene* **116**(1): 43-49.

Béjà O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP, Villacorta R, Amjadi M, Garrigues C, Jovanovich SB, Feldman RA, DeLong EF. (2000). "Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage." *Environ. Microbiol.* **2**(5):516-29.

Bertrand S, Bohni N, Schnee S, Schumpp O, Gindro K, Wolfender JL. (2014). "Metabolite induction via microorganism co-culture: a potential way to enhance chemical diversity for drug discovery." *J. Biotech. Adv.* **32**(6):1180-204.

Bourgogne A, Garsin DA, Qin X, Singh KV, Sillanpaa J, Yerrapragada S, Ding Y, Dugan-Rocha

S, Buhay C, Shen H, Chen G, Williams G, Muzny D, Maadani A, Fox KA, Gioia J, Chen L, Shang Y, Arias CA, Nallapareddy SR, Zhao M, Prakash VP, Chowdhury S, Jiang H, Gibbs RA, Murray BE, Highlander SK, Weinstock GM. (2008). "Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF." *Genome Biol.* **9**(7):R110.

Brady SF, Chao CJ, Handelsman J, Clardy J. (2001). "Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA." *Org. Lett.* **3**(13):1981-4.

Brady SF. (2007). "Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules." *Nat. Protoc.* **2**(5):1297-305.

Brady SF, Clardy J. (2000). "CR377, a new pentaketide antifungal agent isolated from an endophytic fungus." *J. Nat. Prod.* **63**(10):1447-8.

Brady SF, Clardy J. (2005). N-acyl derivatives of arginine and tryptophan isolated from environmental DNA expressed in *Escherichia coli*. *Org. Lett.* **7**(17):3613-6.

Brady SF, Simmons L, Kim JH, Schmidt EW. (2009). "Metagenomic approaches to natural products from free-living and symbiotic organisms." *Nat. Prod. Rep.* **26**(11):1488-503.

Brakhage, AA. (2013). "Regulation of fungal secondary metabolism." *Nat. Rev. Microbiol.* **11**(1):21-32.

Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F. (2002). "Genomic analysis of uncultured marine viral communities." *Proc. Natl. Acad. Sci. USA* **99**(22):14250-5.

Brulc JM, Antonopoulos DA, Miller ME, Wilson MK, Yannarell AC, Dinsdale EA, Edwards RE, Frank ED, Emerson JB, Wacklin P, Coutinho PM, Henrissat B, Nelson KE, White BA. (2009). "Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases." *Proc. Natl. Acad. Sci. USA* **106**(6):1948-53.

Chang FY, Ternei MA, Calle PY, Brady SF. (2015). "Targeted metagenomics: finding rare

tryptophan dimer natural products in the environment.” *J. Am. Chem. Soc.* **137**(18):6044-52.

Chang FY, Brady SF. (2014). “Characterization of an environmental DNA-derived gene cluster that encodes the bisindolylmaleimide methylarcyriarubin.” *ChemBioChem.* **15**(6):815-21.

Chang FY, Ternei MA, Calle PY, Brady SF. (2013). “Discovery and synthetic refactoring of tryptophan dimer gene clusters from the environment.” *J. Am. Chem. Soc.* **135**(47):17906-12.

Chang FY, Brady SF. (2011). “Cloning and characterization of an environmental DNA-derived gene cluster that encodes the biosynthesis of the antitumor substance BE-54017.” *J. Am. Chem. Soc.* **133**(26):9996-9.

Cragg GM, Newman DJ. (2001). “Natural product drug discovery in the next millennium.” *Pharm. Biol.* **39** Suppl 1:8-17.

Cragg GM, Grothaus PG, Newman DJ. (2009). “Impact of natural products on developing new anti-cancer agents.” *Chem. Rev.* **109**(7):3012-43.

Cragg GM, Katz F, Newman DJ, Rosenthal J. (2012). “The impact of the United Nations Convention on Biological Diversity on natural products research.” *Nat. Prod. Rep.* **29**(12):1407-23.

Cragg GM, Newman DJ. (2013). “Natural products: a continuing source of novel drug leads.” *Biochim. Biophys. Acta.* **1830**(6):3670-95.

Craig JW, Chang FY, Brady SF. (2009). “Natural products from environmental DNA hosted in *Ralstonia metallidurans*.” *ACS Chem. Biol.* **4**(1):23-8.

Craig JW, Chang FY, Kim JH, Obiajulu SC, Brady SF. (2010). “Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria.” *Appl. Environ. Microbiol.* **76**(5):1633-41.

Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G, Martinez A, Kolvek

SJ, Hopke J, Osburne MS, August PR, Nalin R, Guérineau M, Jeannin P, Simonet P, Pernodet JL. (2003). "Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products." *Appl. Environ. Microbiol.* **69**(1):49-55.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. (2006). "Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB." *Appl. Environ. Microbiol.* **72**(7): 5069-5072.

Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, Furlan M, Desnues C, Haynes M, Li L, McDaniel L, Moran MA, Nelson KE, Nilsson C, Olson R, Paul J, Brito BR, Ruan Y, Swan BK, Stevens R, Valentine DL, Thurber RV, Wegley L, White BA, Rohwer F. (2008). "Functional metagenomic profiling of nine biomes." *Nature* **452**(7187): 629-632.

Donia MS, Hathaway BJ, Sudek S, Haygood MG, Rosovitz MJ, Ravel J, Schmidt EW. (2006). "Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians." *Nat. Chem. Biol.* **2**(12):729-35.

Dunny GM, Brown BL, Clewell DB. (1978). "Induced cell aggregation and mating in *Streptococcus faecalis*: evidence for a bacterial sex pheromone." *Proc. Natl. Acad. Sci. USA* **75**(7):3479-83.

Feng Z, Kim JH, Brady SF. (2010). "Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type II PKS gene cluster." *J. Am. Chem. Soc.* **132**(34):11902-3

Feng Z, Kallifidas D, Brady SF. (2011). "Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites." *Proc. Natl. Acad. Sci. USA* **108**(31):12629-34.

Feng Z, Chakraborty D, Dewell SB, Reddy BV, Brady SF. (2012). "Environmental DNA-encoded antibiotics fasamycins A and B inhibit FabF in type II fatty acid biosynthesis." *J. Am. Chem. Soc.* **134**(6):2981-7.

Fernández-Moreno MA, Martínez E, Boto L, Hopwood DA, 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**(27):19278-90.

Fotso S, Mahmud T, Zabriskie TM, Santosa DA, Sulastri, Proteau PJ. (2008). "Angucyclinones from an Indonesian *Streptomyces* sp." *J. Nat. Prod.* **71**(1):61-5.

Fox JL. (2006). "The business of developing antibacterials." *Nat. Biotechnol.* **24**(12): 1521-1528.

Fujimoto, S. and Ike, Y. (2001). "pAM401-based shuttle vectors that enable overexpression of promoterless genes and one-step purification of tag fusion proteins directly from *Enterococcus faecalis*." *Appl. Environ. Microbiol.* **67**(3):1262-7.

Fujimoto S, Hashimoto H, Ike Y. (1991). "Low cost device for electrotransformation and its application to the highly efficient transformation of *Escherichia coli* and *Enterococcus faecalis*." *Plasmid* **26**(2):131-5.

Gabor EM, Alkema WB, Janssen DB. (2004). "Quantifying the accessibility of the metagenome by random expression cloning techniques." *Environ. Microbiol.* **6**(9):879-86.

Gans J, Wolinsky M, Dunbar J. (2005). "Computational improvements reveal great bacterial diversity and high metal toxicity in soil." *Science* **309**(5739): 1387-1390.

Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM, Handelsman J. (2002). "Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA." *Appl. Environ. Microbiol.* **68**(9):4301-6.

Guan C, Ju J, Borlee BR, Williamson LL, Shen B, Raffa KF, Handelsman J. (2007). "Signal mimics derived from a metagenomic analysis of the gypsy moth gut microbiota." *Appl. Environ. Microbiol.* **73**(11): 3669-3676.

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. (1998). "Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products." *Chem. Biol.* **5**(10):R245-9.

Hayakawa Y, Ha SC, Kim YJ, Furihata K, Seto H. (1991). "Studies on the isotetracenone antibiotics. IV. Hatomarubigins A, B, C and D, new isotetracenone antibiotics effective against multidrug-resistant tumor cells." *J. Antibiot.* **44**(11):1179-86.

Herath KB, Jayasuriya H, Guan Z, Schulman M, Ruby C, Sharma N, MacNaul K, Menke JG, Kodali S, Galgoci A, Wang J, Singh SB. (2005). "Anthrabenzoxocinones from *Streptomyces* sp. as liver X receptor ligands and antibacterial agents." *J. Nat. Prod.* **68**(9):1437-40.

Hertweck, C. (2009). "The biosynthetic logic of polyketide diversity." *Angew. Chem. Int. Ed. Engl.* **48**(26):4688-716.

Hertweck C, Luzhetskyy A, Rebets Y, Bechthold A. (2007). "Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork." *Nat. Prod. Rep.* **24**(1):162-90.

Hong HJ, Hutchings MI, Hill LM, Buttner MJ. (2005). "The role of the novel Fem protein VanK in vancomycin resistance in *Streptomyces coelicolor*." *J. Biol. Chem.* **280**(13):13055-61.

Hopwood, D. A. (1997). "Genetic Contributions to Understanding Polyketide Synthases." *Chem. Rev.* **97**(7):2465-2498.

Hugenholtz P, Goebel BM, Pace NR. (1998). "Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity." *J. Bacteriol.* **180**(18): 4765-4774.

Jensen PR, Chavarria KL, Fenical W, Moore BS, Ziemert N. (2014). "Challenges and triumphs to genomics-based natural product discovery." *J. Ind. Microbiol. Biotechnol.* **41**(2):203-9.

Kaeberlein T. (2002). "Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment." *Science* **296**(5570):1127-9.

Kallifidas D, Kang HS, Brady SF. (2012). "Tetarimycin A, an MRSA-active antibiotic identified through induced expression of environmental DNA gene clusters." *J. Am. Chem. Soc.* **134**(48):19552-5.

Kallifidas D, Brady SF. (2012). "Reassembly of functionally intact environmental DNA-derived biosynthetic gene clusters." *Methods Enzymol.* **517**:225-39.

Kang HS, Brady SF. (2014). "Mining soil metagenomes to better understand the evolution of natural product structural diversity: pentangular polyphenols as a case study." *J. Am. Chem. Soc.* **136**(52):18111-9.

Kang HS, Brady SF. (2013). "Arimetamycin A: improving clinically relevant families of natural products through sequence-guided screening of soil metagenomes." *Angew. Chem. Int. Ed. Engl.* **52**(42):11063-7.

Keller M, Zengler K. (2004). "Tapping into microbial diversity." *Nat. Rev. Microbiol.* **2**(2): 141-150.

Kharel MK, Pahari P, Shepherd MD, Tibrewal N, Nybo SE, Shaaban KA, Rohr J. (2012). "Angucyclines: Biosynthesis, mode-of-action, new natural products, and synthesis." *Nat. Prod. Rep.* **29**(2):264-325.

Kieser T. (2000). *Practical Streptomyces Genetics*, 2nd Ed., John Innes Foundation, Norwich, UK.

Kim JH, Feng Z, Bauer JD, Kallifidas D, Calle PY, Brady SF. (2010). "Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR." *Biopolymers* **93**(9):833-44.

King RW, Bauer JD, Brady SF. (2009). "An environmental DNA-derived type II polyketide biosynthetic pathway encodes the biosynthesis of the pentacyclic polyketide erdacin." *Angew. Chem. Int. Ed. Engl.* **48**(34):6257-61.

Klein E, Smith DL, Laxminarayan R. (2007). "Hospitalizations and deaths caused by methicillin-resistant *Staphylococcus aureus*." *Emerg. Infect. Dis.* **13**(12): 1840-6.

Kodali S, Galgoci A, Young K, Painter R, Silver LL, Herath KB, Singh SB, Cully D, Barrett JF, Schmatz D, Wang J. (2005). "Determination of selectivity and efficacy of fatty acid synthesis

inhibitors.” *J. Biol. Chem.* **280**(2):1669-77.

Levy SB, Marshall B. (2004). “Antibacterial resistance worldwide: causes, challenges and responses.” *Nat. Med.* **10**(12 Suppl):S122-9.

Li A, Piel J. (2002). “A gene cluster from a marine *Streptomyces* encoding the biosynthesis of the aromatic spiroketal polyketide griseorhodin A.” *Chem. Biol.* **9**(9):1017-26.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. (2009). “The Sequence Alignment/Map format and SAMtools.” *Bioinformatics* **25**(16):2078-9.

Liles MR, Williamson LL, Rodbumrer J, Torsvik V, Goodman RM, Handelsman J. (2008). “Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms.” *Appl. Environ. Microbiol.* **74**(10): 3302-3305.

Lim HK, Chung EJ, Kim JC, Choi GJ, Jang KS, Chung YR, Cho KY, Lee SW. (2005). “Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*.” *Appl. Environ. Microbiol.* **71**(12): 7768-7777.

Liu WC, Parker L, Slusarchyk S, Greenwood GL, Graham SF, Meyers E. (1970). “Isolation, characterization, and structure of rabelomycin, a new antibiotic.” *J. Antibiot.* **23**(9):437-41.

Li Y, Wexler M, Richardson DJ, Bond PL, Johnston AW. (2005). “Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned *trp* genes.” *Environ. Microbiol.* **7**(12):1927-36.

Lloyd-Jones G, Lau PC. (1998). “A molecular view of microbial diversity in a dynamic landfill in Québec.” *FEMS Microbiol. Lett.* **162**(2):219-26.

Lunter G, Goodson M. (2011). “Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads.” *Genome Res.* **21**(6):936-9.

Lu YJ1, Zhang YM, Rock CO. (2004). "Product diversity and regulation of type II fatty acid synthases." *Biochem. Cell Biol.* **82**(1):145-55.

MacNeil IA, Tiong CL, Minor C, August PR, Grossman TH, Loiacono KA, Lynch BA, Phillips T, Narula S, Sundaramoorthi R, Tyler A, Aldredge T, Long H, Gilman M, Holt D, Osburne MS. (2001). "Expression and isolation of antimicrobial small molecules from soil DNA libraries." *J. Mol. Microbiol. Biotechnol.* **3**(2):301-8.

Magbanua ZV, Ozkan S, Bartlett BD, Chouvarine P, Sasaki CA, Liston A, Cronn RC, Nelson CD, Peterson DG. (2011). "Adventures in the enormous: a 1.8 million clone BAC library for the 21.7 Gb genome of loblolly pine." *PLoS One.* **6**(1):e16214.

Malpartida F, Hopwood DA. (1984). "Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host." *Nature* **6**,309(5967):462-4.

Martinez A, Kolvek SJ, Yip CL, Hopke J, Brown KA, MacNeil IA, Osburne MS. (2004). "Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts." *Appl. Environ. Microbiol.* **70**(4):2452-63.

Montiel D, Kang HS, Chang FY, Charlop-Powers Z, Brady SF. (2015). "Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters." *Proc. Natl. Acad. Sci. USA* **112**(29):8953-8.

Moran GJ, Krishnadasan A, Gorwitz RJ, Fosheim GE, McDougal LK, Carey RB, Talan DA; EMERGENCY ID Net Study Group. (2006). "Methicillin-resistant *S. aureus* infections among patients in the emergency department." *N. Engl. J. Med.* **355**(7): 666-74.

Murenets NV. (1990). *Antibiotiki*, **35**:47.

Nakagawa A, Omura S, Kushida K, Shimizu H, Lukacs G. (1987). "Structure of cervinomycin, a novel xantone antibiotic active against anaerobe and mycoplasma." *J. Antibiot.* **40**(3):301-8.

Newman DJ, Cragg GM, Snader KM. (2000). "The influence of natural products upon drug discovery." *Nat. Prod. Rep.* **17**(3):215-34.

Newman DJ, Cragg GM. (2012). "Natural products as sources of new drugs over the 30 years from 1981 to 2010." *J. Nat. Prod.* **75**(3):311-35.

Newman DJ, Cragg GM. (2007). "Natural products as sources of new drugs over the last 25 years." *J. Nat. Prod.* **70**(3):461-77.

Newman DJ, Cragg GM, Snader KM. (2003). "Natural products as sources of new drugs over the period 1981-2002." *J. Nat. Prod.* **66**(7):1022-37.

Newman DJ, Cragg GM. (2004). "Marine natural products and related compounds in clinical and advanced preclinical trials." *J. Nat. Prod.* **67**(8):1216-38.

Nougayrède JP, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, Buchrieser C, Hacker J, Dobrindt U, Oswald E. (2006). "Escherichia coli induces DNA double-strand breaks in eukaryotic cells." *Science.* **313**(5788):848-51.

Nüsslein, K. and Tiedje, J. M. (1998). "Characterization of the dominant and rare members of a young Hawaiian soil bacterial community with small-subunit ribosomal DNA amplified from DNA fractionated on the basis of its guanine and cytosine composition." *Appl. Environ. Microbiol.* **64**(4):1283-9.

O'Neill AJ, Chopra I. (2004). "Preclinical evaluation of novel antibacterial agents by microbiological and molecular techniques." *Expert Opin. Investig. Drugs.* **13**(8):1045-63.

Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, Takahashi Y, Horikawa H, Nakazawa H, Osonoe T, Kikuchi H, Shiba T, Sakaki Y, Hattori M. (2001). "Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites." *Proc. Natl. Acad. Sci. USA* **98**(21):12215-20.

Mei-Fong Pang, Geok-Yuan Annie Tan, Noorlidah Abdullah, Choon-Weng Lee and Ching-Ching Ng. (2008). "Phylogenetic Analysis of Type I and Type II Polyketide Synthase from

Tropical Forest Soil.” *Biotechnology* 7(4): 660-668.

Parsons JB, Rock CO. (2011). “Is bacterial fatty acid synthesis a valid target for antibacterial drug discovery?” *Curr. Opin. Microbiol.* 14(5):544-9.

Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C. (2001). “Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*.” *Science* 291(5509):1790-2.

Rawlings BJ. (1999). “Biosynthesis of polyketides (other than actinomycete macrolides).” *Nat. Prod. Rep.* 16(4):425-84.

Roberts M, Hunt BR, Yorke JA, Bolanos RA, Delcher AL. “A preprocessor for shotgun assembly of large genomes.” *J. Comput. Biol.* 11(4):734-52.

Rodriguez E, Menzella HG, Gramajo H. (2009). “Heterologous production of polyketides in bacteria.” *Methods Enzymol.* 2009;459:339-65.

Rodriguez L, Oelkers C, Aguirrezabalaga I, Braña AF, Rohr J, Méndez C, Salas JA. (2000). “Generation of hybrid elloramycin analogs by combinatorial biosynthesis using genes from anthracycline-type and macrolide biosynthetic pathways.” *J. Mol. Microbiol. Biotechnol.* 2(3):271-6.

Rohr J, Thiericke R. (1992) “Angucycline group antibiotics.” *Nat. Prod. Rep.* 9(2):103-37.

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J, Goodman RM. (2000). “Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms.” *Appl. Environ. Microbiol.* 66(6): 2541-2547.

Omura S, Takahashi Y, Kim Yong-pil, Hanaki H, Koda H, Suzuki M, Shiomi K. (2009). “New KB-3346-5 substance and method for producing the same.” JP2009046404.

Shen B. (2000). “Biosynthesis of aromatic polyketides.” *Curr. Top. Chem.* 209:1-51.

Scherlach K, Hertweck, C. (2009). "Triggering cryptic natural product biosynthesis in microorganisms." *Org. Biomol. Chem.* **7**(9):1753-60.

Schipper C, Hornung C, Bijtenhoorn P, Quitschau M, Grond S, Streit WR. (2009). "Metagenome-derived clones encoding two novel lactonase family proteins involved in biofilm inhibition in *Pseudomonas aeruginosa*." *Appl. Environ. Microbiol.* **75**(1):224-33.

Schloss PD, Handelsman J. (2004). "Status of the microbial census." *Microbiol. Mol. Biol. Rev.* **68**(4): 686-691.

Schujman GE, de Mendoza D. (2008). "Regulation of type II fatty acid synthase in Gram-positive bacteria." *Curr. Opin. Microbiol.* **11**(2):148-52.

Seoane AS, Levy SB. (1995). "Characterization of MarR, the repressor of the multiple antibiotic resistance (mar) operon in *Escherichia coli*." *J. Bacteriol.* **177**(12):3414-9.

Seow KT, Meurer G, Gerlitz M, Wendt-Pienkowski E, Hutchinson CR, Davies J. (1997). "A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms." *J. Bacteriol.* **179**(23):7360-8.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. (2006). "Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. USA* **103**(32):12115-20.

Staunton J, Weissman KJ. (2001). "Polyketide biosynthesis: a millennium review." *Nat. Prod. Rep.* **18**(4):380-416.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods." *Mol. Biol. Evol.* **28**(10):2731-9.

Torsvik V, Goksøyr J, Daae FL. (1990). "High diversity in DNA of soil bacteria." *Appl. Environ. Microbiol.* **56**(3):782-7.

Torsvik V, Øvreås L. (2002). "Microbial diversity and function in soil: from genes to ecosystems." *Curr. Opin. Microbiol.* **5**(3):240-5.

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM. (2005). "Comparative metagenomics of microbial communities." *Science* **308**(5721): 554-557.

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF. (2004). "Community structure and metabolism through reconstruction of microbial genomes from the environment." *Nature* **428**(6978): 37-43.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO. (2004). "Environmental genome shotgun sequencing of the Sargasso Sea." *Science* **304**(5667):66-74.

Voget S, Leggewie C, Uesbeck A, Raasch C, Jaeger KE, Streit WR. (2003). "Prospecting for novel biocatalysts in a soil metagenome." *Appl. Environ. Microbiol.* **69**(10):6235-42.

Wang GY, Graziani E, Waters B, Pan W, Li X, McDermott J, Meurer G, Saxena G, Andersen RJ, Davies J. (2000). "Novel natural products from soil DNA libraries in a streptomycete host." *Org. Lett.* **2**(16):2401-4.

Wang J, Kodali S, Lee SH, Galgoci A, Painter R, Dorso K, Racine F, Motyl M, Hernandez L, Tinney E, Colletti SL, Herath K, Cummings R, Salazar O, González I, Basilio A, Vicente F, Genilloud O, Pelaez F, Jayasuriya H, Young K, Cully DF, Singh SB. (2007). "Discovery of platencin, a dual FabF and FabH inhibitor with *in vivo* antibiotic properties." *Proc. Natl. Acad. Sci. USA* **104**(18):7612-6.

Wawrik B, Kerkhof L, Zylstra GJ, Kukor JJ. (2005). "Identification of unique type II polyketide synthase genes in soil." *Appl. Environ. Microbiol.* **71**(5): 2232-8.

Wawrik B, Kutliev D, Abdivasievna UA, Kukor JJ, Zylstra GJ, Kerkhof L. (2007). "Biogeography of actinomycete communities and type II polyketide synthase genes in soils collected in New Jersey and Central Asia." *Appl. Environ. Microbiol.* **73**(9): 2982-9.

- Weissman KJ, Leadlay PF. (2005). "Combinatorial biosynthesis of reduced polyketides." *Nat. Rev. Microbiol.* **3**(12):925-36.
- Williamson LL, Borlee BR, Schloss PD, Guan C, Allen HK, Handelsman J. (2005). "Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor." *Appl. Environ. Microbiol.* **71**(10):6335-44.
- Wilkinson B, Micklefield J. (2007). "Mining and engineering natural-product biosynthetic pathways." *Nat. Chem. Biol.* **3**(7):379-86.
- Winter DK, Sloman DL, Porco JA Jr. (2013). "Polycyclic xanthone natural products: structure, biological activity and chemical synthesis." *Nat. Prod. Rep.* **30**(3):382-91.
- Wolfgang MC, Kulasekara BR, Liang X, Boyd D, Wu K, Yang Q, Miyada CG, Lory S. (2003). "Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*." *Proc. Natl. Acad. Sci. USA* **100**(14):8484-9.
- Wright HT, Reynolds KA. (2007). "Antibacterial targets in fatty acid biosynthesis." *Curr. Opin. Microbiol.* **10**(5):447-53.
- Wyatt MA, Wang W, Roux CM, Beasley FC, Heinrichs DE, Dunman PM, Magarvey NA. (2010). "Staphylococcus aureus nonribosomal peptide secondary metabolites regulate virulence." *Science*. **329**(5989):294-6.
- Zerbino DR, Birney E. (2008). "Velvet: algorithms for de novo short read assembly using de Bruijn graphs." *Genome Res.* **18**(5): 821-829.
- Zengler K. (2002). "Cultivating the uncultured." *Proc. Natl. Acad. Sci. USA* **99**(24):15681-6.
- Zhang YM, White SW, Rock CO. (2006). "Inhibiting bacterial fatty acid synthesis." *J. Biol. Chem.* **281**(26):17541-4.

Zhu W, Lomsadze A, Borodovsky M. (2010). “Ab initio gene identification in metagenomic sequences.” *Nucleic Acids Res.* **38**(12):e132.

Ziemert N, Ishida K, Weiz A, Hertweck C, Dittmann E. (2010). “Exploiting the natural diversity of microviridin gene clusters for discovery of novel tricyclic depsipeptides.” *Appl. Environ. Microbiol.* **76**(11):3568-74.