2013

Novel Molecules from Environmental DNA-Derived Type II Polyketide Biosynthetic Gene Clusters

Ryan W. King

NOVEL MOLECULES FROM ENVIRONMENTAL DNA-DERIVED TYPE II POLYKETIDE BIOSYNTHETIC GENE CLUSTERS

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Ryan W. King

June 2013

NOVEL MOLECULES FROM ENVIRONMENTAL DNA-DERIVED TYPE II POLYKETIDE BIOSYNTHETIC GENE CLUSTERS

Ryan W. King, Ph.D.

The Rockefeller University 2013

Natural products have been a significant source of new drugs over the past 25 years. In fact, the origins of nearly 70% of antibacterial, antifungal, antiparasitic and antiviral small molecule drugs are in natural products (Newman and Cragg 2007). A significant proportion of natural product drugs and leads have been isolated from microbes (Newman and Cragg 2007). However, the discovery of novel natural products from microbes faces the growing hurdle of rediscovering known metabolites. *Streptomyces* is a microbe that has been a particularly rich source of natural products; however, it has been predicted that the last easily accessible novel *Streptomyces* metabolite will be found within the next few years (Watve, Tickoo et al. 2001). Finding a new source rich in natural products would be very beneficial to the development of new drugs.

One potential new source of natural products is the plethora of uncultured bacteria found in environmental soil samples. It is estimated that greater than 99% of bacteria from soil samples are recalcitrant to culturing in a laboratory setting. A culture-independent approach of extracting

environmental DNA (eDNA) from soil and cloning it into a library was leveraged to take advantage of the genetic diversity of uncultured microbes.

Uncultured microbes have been found to be rich in type II polyketide synthase (PKS) gene clusters (Kieser, Bibb et al. 2000). PKS gene clusters produce a structurally diverse collection of aromatic natural products. The nascent polyketide in these pathways is produced by a conserved minimal PKS, composed of an α-ketosynthase, β-ketosynthase and acyl carrier protein. In order to access the structural diversity encoded by type II PKS gene clusters from uncultured bacteria, environmental DNA (eDNA) libraries were constructed from desert soil. These libraries were found to be rich in PKS genes and cosmids containing minimal PKS genes were recovered from the eDNA library using β-ketosynthase sequences as probes. The recovered clones containing minimal PKS genes from the eDNA library were screened for the ability to confer the production of clone-specific molecules to cultured *Streptomyces albus*.

One of the recombinant *Streptomyces* clones was found to produce erdacin, a 26-carbon polyketide with both a novel carbon backbone and a novel pentacyclic ring arrangement. The novel structure and pentacyclic ring arrangement of erdacin raised questions as to the biosynthetic origins of this metabolite encoded by the eDNA-derived V167 clone. A biosynthetic analysis of wild type and transposon mutant V167 cultures indicated that erdacin arises from the heterodimerization of two 13-carbon monomer subunits that are both derived from octaketide precursors. These two 13-carbon subunits arise from two different second cyclizations, but both have the same three carbons excised from their 16-carbon octaketide precursors.

In addition to erdacin, a number of other clone specific novel and known metabolites were characterized from cultures of *S. albus* conjugated with the wild-type V167 clone. These metabolites included utahmycin A, utahmycin B and 8-O-methylutahmycin A. The structure of Utahmycin A is novel and utahmycin B has not been previously observed in a biological system. The identification of the novel structure, erdacin, provides tangible evidence that eDNA-derived type II PKS gene clusters may encode the biosynthesis of molecules that are distinct from known metabolites identified by traditional culture-based strategies. The isolation of additional distinct novel metabolites from the same eDNA clone suggests these gene clusters are a rich source of novel natural products. In future studies, the heterologous expression of eDNA-derived type II PKS gene clusters is likely to be a rewarding avenue to access the structural diversity of natural products encoded by uncultured soil bacteria.

This thesis is dedicated to

My Family.

ACKNOWLEDGMENTS

I am exceptionally grateful to Dr. Vincent Fischetti for his patience, guidance, advice, honestly, fairness, the fantastic lab environment he has created and much more. I am… very grateful…

I would like to thank Dr. Sean Brady for both the time I spent in his lab and for graciously sitting on my committee. I want to thank Dr. Derek Tan for his insightful advice since we met during my interview weekend. I value his opinion and I am pleased to have had him on my committee over the years. Additionally, I would like to thank my committee chairman and program director Dr. Timothy Ryan for his advice and guidance during my most difficult times in graduate school. I would also like to thank Dr. Tarun Kapoor and Dr. Anthony Sauve for kindly joining my committee.

I would like to offer thanks to Dean Sidney Strickland for his advice and the opportunities he has provided me with during my time at Rockefeller.

I must thank past and present members of the Fischetti and Brady labs who have helped me over the years. I would like to thank Jacob Banik, Fang Chang, Jeff Craig and Chad Euler for not only helping me troubleshoot experiments, but for always making the lab a fun experience.

I want to thank all those who were forced to sit next to me in lab: Ben Winer, Paula Calle, Curtis O'Neal and Dimitris Kallifidas. I have enjoyed your sense of humor and companionship.

I would like to offer special thanks to Dr. John Bauer, who recovered the V167 clone from an eDNA library and started an initial characterization of some of the metabolites encoded by V167 before I took over the project. I would also like to thank Paula Calle for her extraordinary assistance in library screening and clone recovery.

Of course I am extremely grateful to my parents and my brother who have supported me through everything in life and academia.

Thank you all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACP	Acyl carrier protein
ATCC	American Type Culture Collection
AB	Anza-Borrego (eDNA-derived cosmid mega-library)
BAC	Bacterial artificial chromosome
BLAST	Basic local alignment search tool
COSY	COrrelation SpectroscopY
DNA	Deoxyribonucleic acid
DMSO	Dimethylsulfoxide
eDNA	Environmental DNA
ESI	Electrospray ionization
Erd	Erdacin
HPLC-MS	High pressure liquid chromatography - mass spectrometry
HRMS	High resolution mass spectrometry
kb	Kilobase
KS	Ketosynthase
KSα	α-Ketosynthase
KSβ	β-Ketosynthase
LB	Luria Bertani medium
LC-MS	Liquid cromatography – mass spectrometry
MS	Mass spectrometry
HMBC	Heteronuclear multiple bond correlation
NMR	Nuclear magnetic resonance spectroscopy
HMQC	Heteronuclear multiple-quantum correlation
NRPS	Non ribosomal peptide synthetase
ORF	Open reading frame
PCR	Polymerase chain reaction
PKS	Polyketide synthase
PPTase	4'-phosphopantetheinyltransferase
SMM	Supplemented minimal medium
TLC	Thin-layer chromatography
UT	Utah (eDNA-derived cosmid mega-library)

CHAPTER 1

Introduction and Background

Natural Products

Natural products have been a significant source of new drugs over the past three decades. In fact, the origins of nearly 70% of antibacterial, antifungal, antiparasitic and antiviral small molecule drugs are in natural products (Newman and Cragg 2007). In addition to antimicrobial properties, natural products have a wide range of therapeutic applications from anticancer to cholesterol-lowering agents. Natural products have been isolated from a wide range of sources including microbial, marine, plant and even animal sources. A significant proportion of natural product drugs and leads have been isolated from microbes (Newman and Cragg 2007). It has been estimated that almost two thirds of clinically useful antibiotics are derived from the primarily soil dwelling *Streptomyces* (Kieser 2000).

The majority of known bacterial natural products are not necessary for growth of the bacteria producing them. As a result, these natural products are referred to as secondary metabolites. Bacteria will expend the energy to produce secondary metabolites to gain a competitive advantage in their environment and utilize these metabolites for a number of applications,

including defense, cell-cell signaling and siderophore iron scavenging. The frequently complex structures of bacterial secondary metabolites are produced by groups of biosynthetic enzymes. The genes encoding for these biosynthetic enzymes are typically grouped in clusters in the genome of the microbe. This gene clustering is believed to have helped facilitate the evolution of natural product biosynthetic gene clusters (Maplestone, Stone et al. 1992).

Traditionally to isolate natural products from bacteria, the bacterium must first be cultured from a sample. Bacterial cultures are then extracted with organic solvents and the metabolites produced can be fractionated and assayed for desired activities. However, the number of new natural products discovered by this culture-dependant approach is declining (Fox 2006). While there are many reasons for this downward trend, one major hurdle has been the rediscovery of already known metabolites. In fact, while *Streptomyces* has been a particularly rich source of natural products, it has been predicted that the last easily accessible novel *Streptomyces* metabolite will be found within the next few years (Watve, Tickoo et al. 2001). Finding a new source rich in natural products would be beneficial to the development of new drugs.

Uncultured Microbes and Metagenomics

One potential new source of natural products is the large number of uncultured microbes found in the environment. It has been predicted that less than 1% of the bacteria in a given soil sample can be easily cultured (Torsvik, Salte et al. 1990; Torsvik, Goksoyr et al. 1990; Hugenholtz, Goebel et al. 1998; Torsvik, Daae et al. 1998; Torsvik and Ovreas 2002; Rappe and Giovannoni 2003). The large portion of uncultured bacteria in a given soil sample are likely to encode for the production of a diverse collection of natural products. Difficulties in growing these microbes from soil samples hamper traditional culture-dependant approaches from accessing these metabolites. In fact, traditional culture dependant approaches have most likely missed the majority of natural products in nature (Hugenholtz, Goebel et al. 1998; Rappe and Giovannoni 2003). Uncultured bacteria represent a large potential untapped source of novel natural products.

Fortunately, culture-independent metagenomic methods have been developed to study the genetic diversity of all the microbes in an environmental sample. Metagenomic methods avoid the limitations of traditional culture-dependant methods by directly extracting DNA representing the entire microbial genetic diversity from the environment. Environmental DNA (eDNA) extracted from nature contains an incredible amount of genetic material, including secondary metabolite biosynthetic gene clusters. This eDNA must then be cloned into a culturable host to enable

future analysis. To accomplish this eDNA can be ligated into a cosmid vector, packaged into lambda phage and transfected into *E. coli* (Brady 2007). This cosmid cloning method captures 30-50 kb pieces of eDNA in a single cosmid. Since the genes responsible for the biosynthesis of natural products are typically clustered together, it is possible to capture an entire biosynthetic gene cluster on a single cosmid (Handelsman, Rondon et al. 1998). Large libraries of eDNA clones, some containing natural product biosynthetic gene clusters, enable the exploration of eDNA-derived genes through subsequent screening or sequencing studies.

Functional metagenomic screening

One approach to access the structural diversity encoded in eDNA libraries is through a functional metagenomic screen (Figure 1). Functional metagenomic screening relies on the expression of eDNA clones in a heterologous host to identify natural products. In functional metagenomic screening, individual clones with phenotypes indicative of small molecule production are identified in eDNA libraries by high throughput assays. These phenotypes that indicate small molecule production include pigmentation, altered colony morphology and/or antibiosis. The pigmentation of clones has been used to identify a number of metabolites from soil eDNA including the pigments indigo (**1**), violacein (**2**) and the turbomycins (**3**) (Figure 2) (MacNeil, Tiong et al. 2001; Brady, Chao et al. 2001; Gillespie, Brady et al. 2002; Lim, Chung et al. 2005). Furthermore, eDNA libraries hosted in *E. coli*

can be screened for antibiosis in top agar overlay assays to detect zones of inhibition against test bacteria such as *B subtilis*. Clone specific metabolites from eDNA clone hits in these overlay assays have been characterized leading to the discovery of a novel isonitrile functionalized indole (**4**) and a collection of long-chain N-acylated amines (**5**) (Figure 2) (Brady and Clardy 2000; Brady and Clardy 2005).

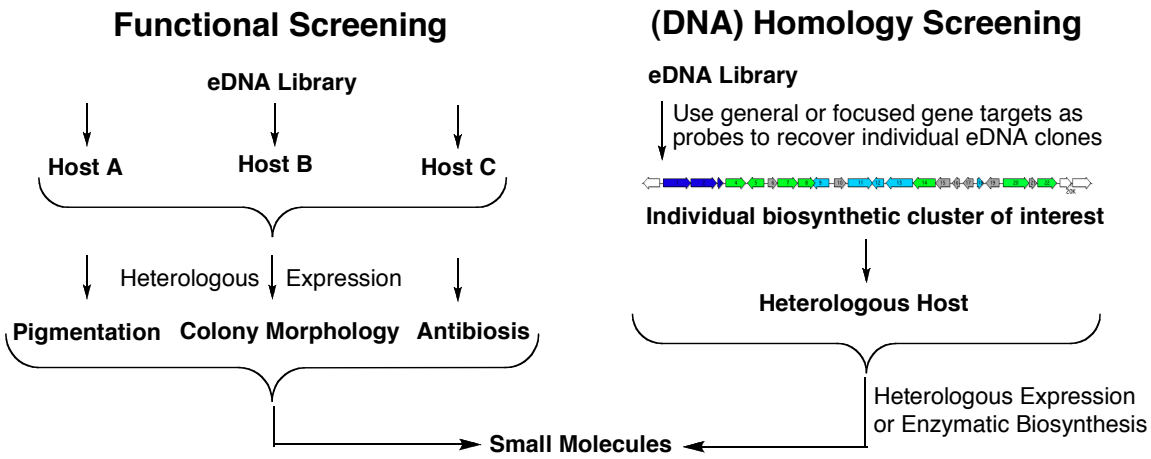


Figure 1: Functional and Homology-based metagenomic screening strategies. Functional screening relies on heterologous expression while the entire eDNA library is assayed in simple high throughput assay. Homology-based screening identifies individual clones of interest by PCR that can be subsequently recovered and heterologously expressed.

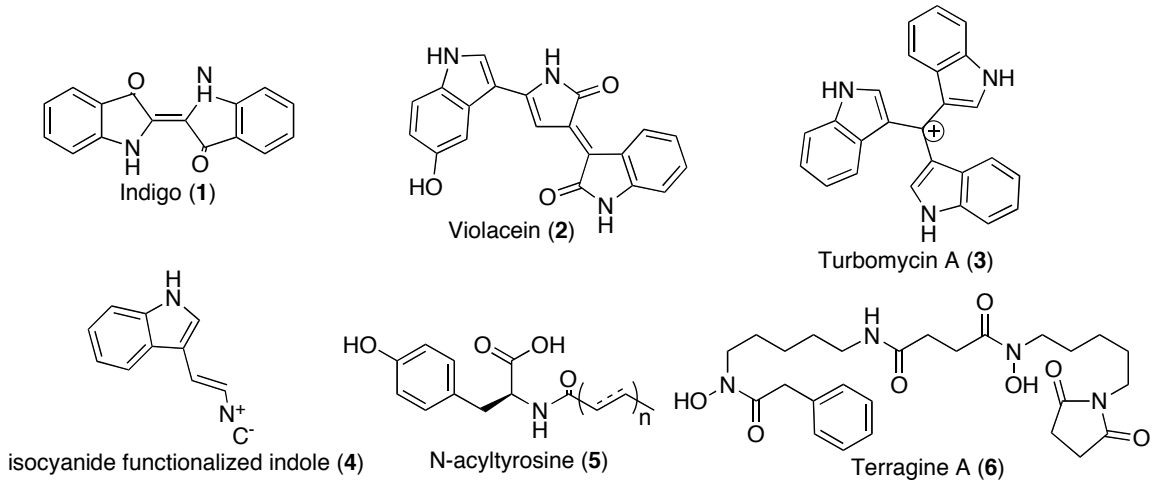


Figure 2: Compounds identified by functional metagenomic screening.

While the terragine (**6**) compounds have been isolated from the heterologous expression of soil eDNA in *S. lividans*, *E. coli* has been the host utilized in many functional metagenomic studies (Wang, Graziani et al. 2000; Banik and Brady 2010). However, it is predicted that *E. coli* is unable to successfully heterologously express much of the genetic information encoded by eDNA (Gabor, Alkema et al. 2004). In order to access more of the genetic diversity of eDNA, libraries have been constructed in broad-host-range vectors that could be moved into multiple phylogenetically distinct bacteria for functional screening (Courtois, Cappellano et al. 2003; Martinez, Kolvek et al. 2004; Aakvik, Degnes et al. 2009; Craig, Chang et al. 2009). One study utilized a broad-host-range vector to construct and screen eDNA libraries in six diverse hosts: *Agrobacterium tumefaciens*, *Burkholderia graminis*, *Caulobacter vibrioides*, *Escherichia coli*, *Pseudomonas putida* and Ralstonia *metallidurans*. This study demonstrated that eDNA clones will present

distinctive phenotypes in different hosts and that the same phenotype is rarely conferred to two different hosts (Craig, Chang et al. 2009; Craig, Chang et al. 2010).

In addition to small molecules, functional metagenomic screening has been used to find proteins of interest, including phage lysins. In fact, a functional metagenomic screen has been developed and used to express 29 phage lysins from environmental samples of animal feces and earthworm guts (Schmitz, Daniel et al. 2008; Schmitz, Schuch et al. 2010). Functional metagenomic screening is a powerful technique to gain access to the genetic diversity encoded by uncultured bacteria.

Homology-based metagenomic screening

Homology-based metagenomic screens take advantage of the sequence homology between previously sequenced biosynthetic genes and the unidentified biosynthetic genes of uncultured bacteria. In homology-based screening, degenerate polymerase chain reaction (PCR) primers are designed from previously sequenced conserved regions of a biosynthetic gene of interest. PCR screening with these degenerate primers enables the identification and recovery of individual clones, which contain a homologous gene, from an eDNA library (Figure 1). These degenerate primers can be used to find new derivatives of known metabolites by targeting relatively rare or pathway-specific biosynthetic genes. Alternatively, these primers can be designed to identify novel structures by targeting conserved general

biosynthetic genes found in a class of compounds. Once a clone with a gene of interested is recovered from an eDNA library, it can be moved into an appropriate heterolgous host for expression.

New derivatives of known metabolites

A number of studies have utilized homology-based metagenomic approaches to identify genes that can be used to develop new derivatives of known compounds (Donia, Hathaway et al. 2006; Schmidt and Donia 2009; Ziemert, Ishida et al. 2010). Banik et al. have used a PCR screen targeting OxyC, a glycopeptide oxidative coupling enzyme, to identify and recover a number of glycopeptide biosynthetic gene clusters from a soil eDNA library (Banik and Brady 2008; Banik, Craig et al. 2010). One of these recovered gene clusters was found to contain three sulfotransferases that were subsequently used to produce novel mono-, di- and trisulfated glycopeptide congeners from the teicoplanin aglycone (Figure 3) (Banik and Brady 2008). A follow up study found additional tailoring enzymes that were used to generate a collection of 15 new sulfated, glycosylated, halogenated and methylated glycopeptide congeners (Banik, Craig et al. 2010).

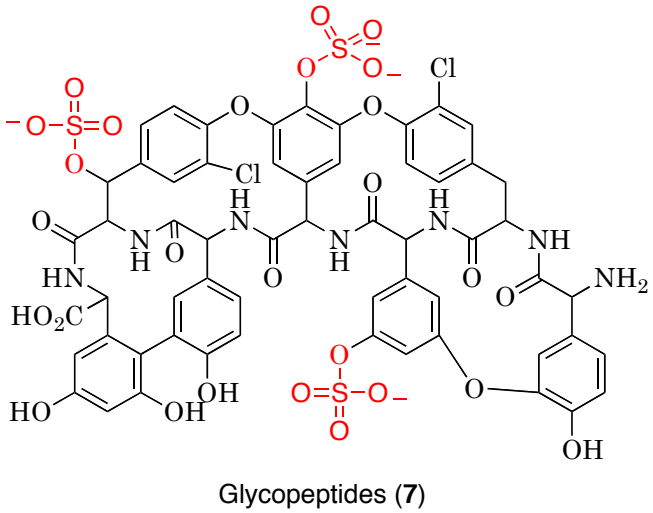


Figure 3: Sulfated glycopeptides.

Three eDNA-derived sulfotransferases were used to produce mono-, di- and trisulfated glycopeptide congeners from teicoplanin aglycone.

Novel Structures

Homology-based metagenomic screens can be designed to find novel structures by targeting conserved general biosynthetic genes found in a class of compounds. Type-II polyketide synthase (PKS) gene clusters contain a highly conserved minimal PKS composed of two ketosynthase genes (KSα and KSβ) and an acyl carrier protein (discussed in more detail below). Type II PKS genes first became targets of homology-based metagenomic screens when two KSβ genes were cloned from eDNA in 1997 (Seow, Meurer et al. 1997). High throughput sequencing studies have indicated that eDNA contains an abundance of novel PKS gene clusters (Wawrik, Kerkhof et al. 2005; Pang, Tan et al. 2008; Warwrik, Kutliev et al. 2007)). However, it has been a challenge to identify eDNA clones with intact gene clusters that can successfully heterologously express small molecules (Piel, Hui et al. 2004;

Schirmer, Gadkari et al. 2005; Courtois, Cappellano et al.2003). In fact, prior to the work presented in this thesis, no new metabolites had been isolated from eDNA-derived type II PKS gene clusters (Figure 4).

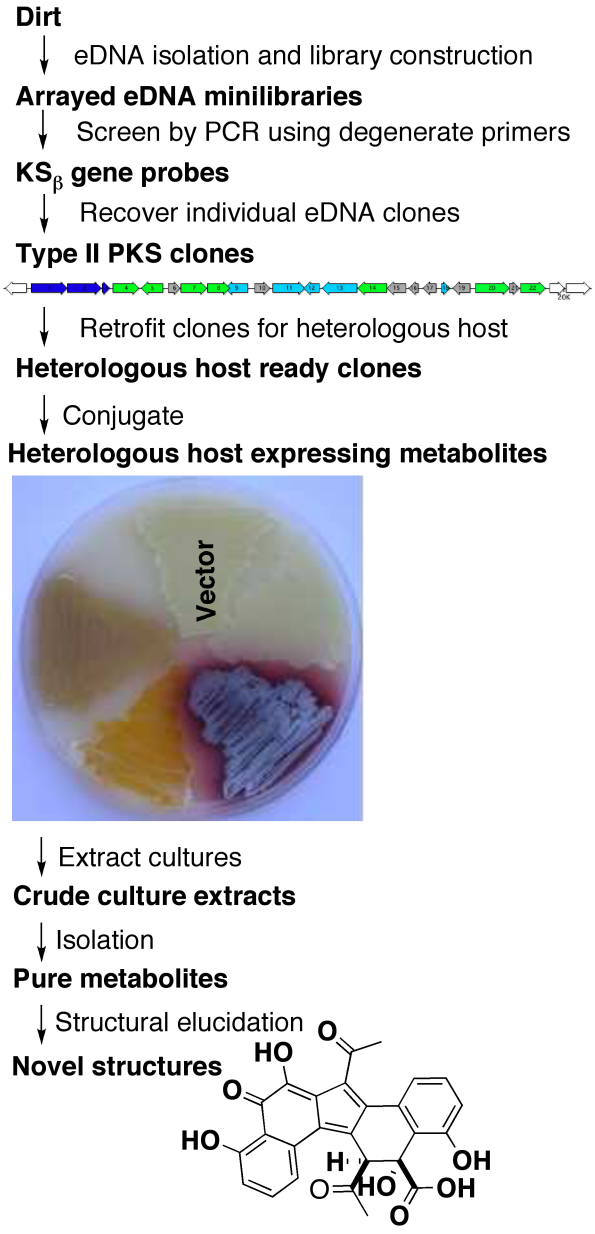


Figure 4: Homology-based metagenomics screen for Type II PKS gene clusters
Scheme for the recovery and heterologous expression of Type II PKS clones from an eDNA library.

After the discovery of erdacin, a novel polyketide with both a novel carbon skeleton and ring system, a number of other aromatic polyketides have been isolated from soil libraries (Figure 5) (King, Bauer et al. 2009). These metabolites include compound **8**, a second eDNA-derived type II polyketide found to have both a novel carbon skeleton and ring system (Feng, Kallifidas et al. 2011). The antibiotic fasamycins (**9**) 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). A collection of fluostatins (**10**) and the methicillin-resistant *Staphylococcus aureus* (MRSA)-active antibiotic tetarimycin A (**11**) have also been isolated from soil libraries (Feng, Kim et al. 2010; Kallifidas, Kang et al. 2012). In addition to isolating polyketides, a homology-based metagenomic screen targeting oxy-tryptophan dimerization has found an eDNA-derived gene cluster that encodes the biosynthesis of the antitumor substance BE-54017 (**12**) (Chang and Brady 2011). Homology-based metagenomic screening is proving to be a rewarding avenue to access the structural diversity encoded by uncultured bacteria.

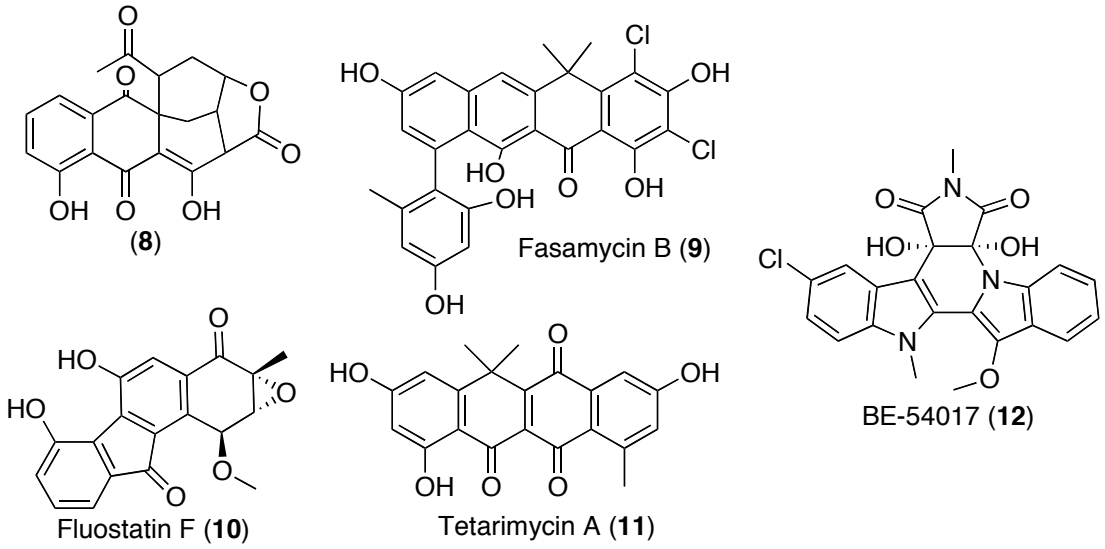


Figure 5: Metabolites isolated in homology-based metagenomic screening.
Four aromatic polyketides and the antitumor substance BE-54017 were all isolated from soil eDNA libraries.

Polyketides

Polyketides are an incredibly valuable class of natural products produced by plants, bacteria and fungi. The over 10,000 compounds in this class have therapeutic applications as antibiotic, immunosuppressant, anticancer, antifugal, antiparasitic and cholesterol-lowering drugs (Figure 6a) (Walla, Bearden et al. 2002; McMurry and Begley 2005; Lombó, Abdelfattah et al. 2009). Polyketides are subdivided into three classes based on their enzymatic architecture: type I, type II and type III (Staunton and Weissman 2001; Fujii 2008). Type I polyketide synthetases are made up of large multifunctional proteins that each contain multiple functional domains. These catalytic domains, including a ketosynthase (KS), acyltransferase, acyl carrier protein (ACP) and other modification domains, are organized into

modules. Type I PKSs are further divided into multi-modular (mPKS) and single modular iterative (iPKS) categories. Type II PKSs are composed of a number of separate individual enzymes that form a multi-enzyme complex. Type III PKSs have a simple architecture compared to other polyketides. They are composed of a homodimer of condensation subunits and can utilize acyl-CoA without the use of an ACP.

Most bacterial aromatic polyketides are encoded by type II PKS gene clusters (Fujii 2008). Type II polyketide gene clusters contain a conserved genetic arrangement known as the minimal PKS, which is composed of an α-ketosynthase (KSα), a β-ketosynthase (KSβ; aka. chain length factor) and an ACP. The minimal PKS is responsible for synthesizing the nascent polyketide by iteratively condensing malonyl-CoA units (Figure 6b). The minimal PKS is also responsible for catalyzing the first cyclization of the nascent polyketide it produces. Nature has derived an astonishing array of post-PKS and tailoring enzymes that cyclize, aromatize, reduce, rearrange and functionalize nascent polyketides into an incredibly diverse collection of natural products (Figure 6a) (McDaniel, Khosla et al. 1995; Matharu, Cox et al. 1998 Shen 2000; Hertweck, Luzhetskyy et al. 2007).

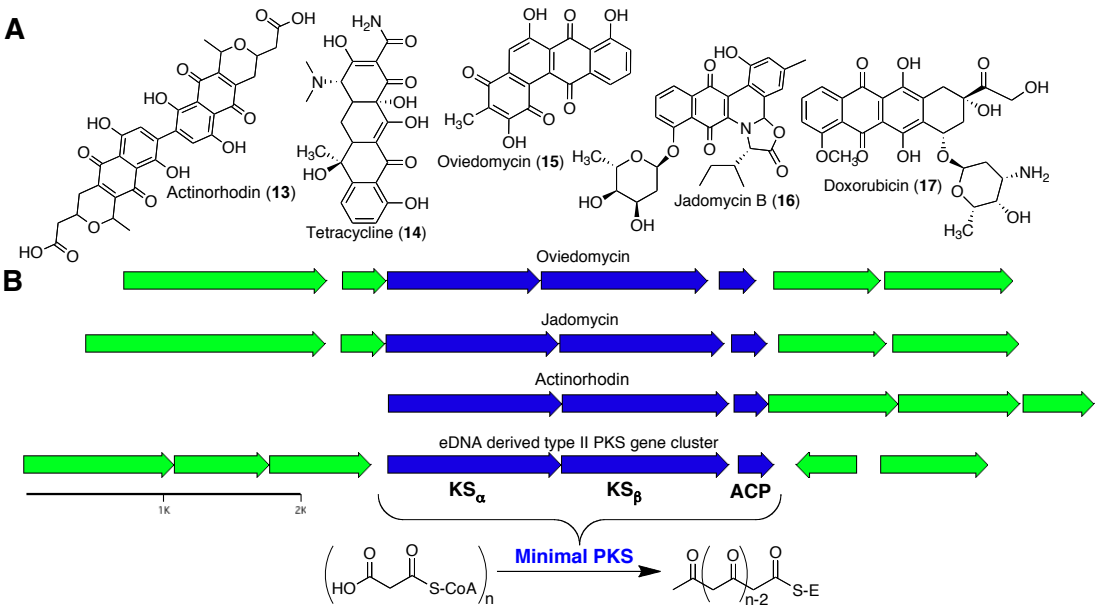


Figure 6: Type II Polyketides and the minimal PKS.

A) Examples of type II polyketides include the antibiotics actinorhodin (**13**) and tetracycline (**14**), the antitumor oviedomycin (**15**), the antifungal jadomycin B (**16**) and the anticancer drug doxorubicin (**17**). B) The highly conserved minimal PKS (KSα, KSβ and ACP) of type II PKS gene clusters iteratively condenses malonyl-CoA units into a nascent polyketide.

The iterative use of the minimal PKS typically makes type II PKS gene clusters smaller than other pathways encoding comparably complex molecules. This makes type II PKS pathways an attractive target for homology-based metagenomic screening, because entire type II PKS gene clusters are more easily captured on a single 30 to 50 kb eDNA cosmid insert. Furthermore, these biosynthetic systems only require universal substrates, which will be present in a heterologous host, to produce the nascent polyketide. The use of universal precursors by the biosynthetic machinery will help enable expression in a heterologous host. In fact, there is extensive precedent for the expression of type II PKS gene clusters in *Streptomyces*.

Furthermore, even a truncated gene cluster with an intact minimal PKS can be functional. As a result, it maybe possible to determine the novelty of a metabolite encoded by an incomplete truncated type II PKS gene cluster captured on a cosmid.

Type II PKS genes are an ideal target for homology-based metagenomic screening of eDNA libraries to find novel natural products because type II PKS pathways produce structurally diverse metabolites, are small enough to capture on a single cosmid, require universal substrates and have precedent of being heterologously expressed in *Streptomyces*. The exploration of type II PKS gene clusters from eDNA libraries is likely a rewarding avenue to gain access to an incredible amount of the structural diversity encoded by uncultured bacteria.

CHAPTER 2

Materials and Methods

General Experimental Procedures

Streptomyces albus J1074 was used for heterologous expression studies. ^{1}H and 2D NMR data were obtained on a Bruker Avance-600 spectrometer. ^{13}C NMR spectra were obtained on a Bruker Avance-400, Bruker Avance-800 spectrometer or Bruker Avance-900 spectrometer. Melting points were obtained on a Stanford Research Systems MPA100 OptiMelt melting point apparatus. A Varian Cary 100-Bio UV-vis spectrophotometer was used to collect UV data. And a Bruker Tensor-27 spectrometer was used to obtain IR data.

eDNA Library Construction and Formatting

Soil collected in Utah (UT) or the Anza-Borrego (AB) desert in California was used to construct a cosmid eDNA library following methods previously described (Brady 2007; Banik and Brady 2008). Soil was incubated for two hours at 70°C in lysis buffer (2% SDS (w/v), 100 mM Tris-HCl, 100 mM EDTA, 1.5 M NaCl and 1% CTAB (w/v)). Large particulates were removed from the sample by centrifugation at 4,000 × g for 30 min. DNA was

then precipitated by the addition of 0.6 volumes of isopropyl alcohol and pelleted by centrifugation at 4,000 × g for 30 min. The resulting pellet was washed with 70% ethanol and resuspended in TE. This DNA was gel purified on a 1% agarose gel for 16h at 20V. The resulting high molecular weight eDNA was blunt-ended (End-It, Epicentre), ligated into pWEB or pWEB-TN (Epicentre) and packaged into lambda phage. *E. coli* EC100 was transfected and 4,000 - 16,000 colonies per plate were plated on a LB plates with the appropriate antibiotic selection. Matching glycerol stock (15% glycerol) and DNA miniprep pairs were made from the cells scraped from each library plate. The miniprep DNA from each of these 4,000-16,000 member library pools were arrayed in 8 X 8 grids. To make screening of the library easier, DNA from the rows and columns of this 8 X 8 grid were pooled. These pooled rows and columns were further combined to yield master aliquots, each representing a single 8X8 grid of minipreps.

PCR Screening of eDNA Libraries for KSβ Genes

The constructed eDNA libraries were for screened for KSβ genes by PCR. Full-length KSβ genes were amplified from eDNA library pools using the following pair of degenerate PCR primers (Seow 1997): dp:KSα, 5' TTC GGS GGX TTC CAG WSX GCS ATG and dp:ACP, 5' TCS AKS AGS GCS AXS GAS TCG TAX CC (X=deoxyinosine). The 25 µl PCR reactions each contained 50 ng of eDNA template, 2.5 µM of each primer, 2 mM dNTPs, 1X ThermoPol Reaction Buffer (New England Biolabs), 0.5 U *Taq* DNA

polymerase (New England Biolabs) and 5% DMSO. A touchdown PCR protocol was used in the initial screening of the libraries: denaturation (95°C; 45s), annealing (touchdown of 65°C to 58°C over 8 cycles D-1°C, followed by 58°C for 35 cycles; 1min) and extension (72°C; 2 min). Amplicons of the correct predicted size (~1.5 kb) were identified by gel electrophoresis, gel purified and sequenced by Sanger sequencing (Biotic Solutions).

Clone Recovery from eDNA libraries

As described above, sub-library pools with PKS containing clones were initially identified by PCR screening of eDNA for KSβ sequences. Individual clones of interest were then recovered from the corresponding glycerol stock in the eDNA library using successive rounds of library dilution and PCR screening. Briefly, a 10^{-4} dilution of the corresponding library glycerol stock resuspended with LB to an OD_{600}=0.2. was prepared. This 10^{-4} dilution (20 μl) was then added to each of the individual wells of two to four 96-well microtiter plates prefilled with LB (with the appropriate antibiotic selection). These plates were incubated at 30°C overnight and 1 μl of this cell culture from each well was used as template in whole cell PCR. Individual colonies from PCR positive wells were then screened by colony PCR to identify individual clones containing KSβ sequences. In some cases a second round of dilution PCR screening of a single PCR positive well was required before individual colony PCR could be utilized.

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 filled by primer walking (Biotic Solutions). The size and deduced role of the open reading frames of sequenced cosmids were predicted using Glimmer (Salzberg, Delcher et al. 1998), SoftBerry (Tyson et al. 2004) and BLAST (Gish and States 1993). MacVector was used for sequence manipulation.

Clone Retrofitting and Integration into *S. albus*

Clones recovered from the eDNA libraries were retrofitted so that they could be integrated into *Streptomyces* for heterologous expression studies. The cassette used for retrofitting was constructed from the integrating vector pOJ436. pOJ436 was digested with DraI and ligated with SmaI cut pBC. The resulting vector was digested with XbaI and SpeI to create the desired retrofitting cassette. Recovered eDNA clones were digested with rare restriction enzymes (AvrII, NheI or PstI) to ensure that only the pWEB or pWEB-TNC vector was cut and not the eDNA insert. The XbaI/SpeI cut pOJ436 cassette was ligated directly into eDNA clones cut with AvrII or NheI. The V167 clone was one such clone digested with AvrII for ligation into the retrofitting cassette. Clones cut with PstI were blunt ended (End-It,

Epicentre) and ligated into a blunt ended retrofitting cassette. The retrofitted clones were then transformed into E. *coli* S17-1 and conjugated into S. *albus* using standard *Streptomyces* conjugation protocols (Kieser, Bibb et al 2000).

Expression, Isolation and Characterization of Erdacin and Juglomycin F from the V167 Clone

A spore stock of *S. albus* transformed with the retrofitted V167 clone (25 mL) was inoculated into 75 mL of modified supplemented minimal medium (SMM). Cultures larger than 75 mL yielded lower titers of erdacin. Modified SMM contained: 819 mL PEG 6000 (6.1%), 25 mL $MgSO_4 \cdot 7H_2O$ (24 g/L), 100 mL TES Buffer (0.25 M, pH 7.2), 2 mL Glucose (50% w/v), 1 mL SMM Trace Element Solution, 1 mL casamino acids (20% w/v) and 25 mL Glycine (20%). Uniformly labeled ^{13}C-erdacin was isolated from V167 cultures grown in modified SMM containing uniformly labeled ^{13}C-glucose and uniformly labeled ^{13}C-glycine. Erdacin labeled with [1,2-^{13}C]acetate was obtained by feeding V167 cultures grown in modified SMM three equal portions of a sodium acetate solution at 24, 48 and 72 hours to attain a final acetate concentration of 1.1 mM [1,2-^{13}C]acetate and 2.2 mM naturally abundant acetate.

V167 cultures were incubated for 4 days at 30°C and 250 rpm. The pH of these cultures was adjusted to 3 with HCl and they were extracted two times with an equal volume of ethyl acetate. The dried extracts were dissolved in methanol (2-4 mL/L of culture) and passed through a SPE

Previal™ C-18 plug. The sample was then dried, redissolved in 1-2 mL of methanol, and partitioned by reversed phase HPLC (XBridge™ C18, 10 × 150 mm, 5 mm, gradient (Table 1) of acetonitrileand water with 0.1% formic acid, 7 mL/min). Fractions containing erdacin (14.9 min) and juglomycin F (8.6 min) were dried down overnight in a speed vac and analyzed by MS and NMR. Plate-like crystals of erdacin were obtained by slow evaporation from a mixture of acetonitrile and water. X-ray crystallography was done by E. Lobkovsky at Cornell University and deposited in the Cambridge Crystallographic Data Centre (CDCC).

Table 1: HPLC gradient for isolating erdacin and juglomycin F

Time (min)	% Water with 0.1% formic acid	% Acetonitrile
Initial	80	20
2	80	20
26	50	50
27	0	100
32	0	100
35	80	20

CUPRAC Antioxidant assay of Erdacin

The antioxidant activity of erdacin was determined using the CUPRAC antioxidant assay (Apak, Kubilay, et al 2004). Cupric chloride dihydrate (10 mM in water, 25 μL), 25 μL 1M ammonium acetate and 25 μL 7.5 mM neocuproine (in ethanol) were added to the wells of a 96-well plate. Control wells of the reaction solution containing 25 μL of water, 25 μL ammonium acetate (1 M) and 25 μL of ethanol were prepared. Two fold serial dilutions (final concentration from 400 mM to 95 fM) of erdacin and ascorbic acid were prepared. Each of the serial dilutions was added (25 μL) to a well containing the 75 μL reaction or control mixtures.

After incubating at room temperature for one hour, the absorbance of each well at 450 nm was measured on a 96-well plate reader. Recorded absorbances of the ascorbic acid dilution series were blanked against a reagent blank. The erdacin dilution series was blanked against a reagent blank and further adjusted by subtracting the absorbance of a well containing 25 μL of the appropriate erdacin dilution, 25 μL of water, 25 μL of 1M ammonium acetate and 25 μL of ethanol. Absorbance vs. concentration was plotted and the molar absorptivity (ε) under the reaction conditions was calculated from the slope of the linear range for both erdacin (2.58×10^4 L $mol^{-1}cm^{-1}$) and ascorbic acid (1.29×10^4 L $mol^{-1}cm^{-1}$). The observed value for ascorbic acid was comparable to the previously reported value of 1.59×10^4 L

$mol^{-1}cm^{-1}$. The molar absorptivity (ε) calculation where A is absorbance, L is pathlength, ε is molar absorptivity, c is concentration and m is the slope of the line plotted in the experiment:

$$A = L \times \varepsilon \times c$$

$$m = L \times \varepsilon$$

$$\varepsilon_{erd} = (m_{erd})/(L) = (5157.1 \text{ L mol}^{-1})/(0.2\text{cm}) = 2.58 \times 10^4 \text{ L mol}^{-1}\text{cm}^{-1}$$

$$\varepsilon_{ascorbic\ acid} = (m_{ascorbic\ acid})/(L) = (2574.5 \text{ L mol}^{-1})/(0.2\text{cm}) = 1.29 \times 10^4 \text{ L mol}^{-1}\text{cm}^{-1}$$

Expression, Isolation and Characterization of Utahmycin A, Utahmycin B and 8-O-methylutahmycin A

A number of metabolites were observed under alternative culture conditions of the V167 clone. These metabolites were isolated and characterized. A spore stock of *S. albus* transformed with the retrofitted V167 clone (75 mL) was inoculated into two 1 L flasks of modified R2YE medium. Modified R2YE medium was prepared by adding 10 mL KH_2PO_4 (0.5%), 80 mL $CaCl_2 \bullet 2H_2O$ (3.68%), 15 mL L-proline (20%), 25 mL TES buffer (5.73%, pH 7.2), 2 mL R2 trace element solution, 5 mL NaOH (1M) and 50 mL yeast extract (10%) to a sterile 800 mL solution containing 103 g of sucrose, 250 mg K_2SO_4, 10.12 g $MgCl_2 \bullet 6H_2O$, 10 g glucose and 100 mg casaminoacids.

After incubating at 30°C with shaking (250 or 300 rpm) for 7-14 days, the pH of these cultures was adjusted to 3 with HCl and extracted with an

equal volume of ethyl acetate twice. The dried, crude extracts were then fractionated by silica flash chromatography using a chloroform/methanol gradient in which 100-150 mL fractions were collected for each of the following chloroform/methanol ratios: 100:0, 99:1, 97:3, 95:5, 90:10, 80:20 and 0:100. The 99:1 fraction was found to contain both utahmycin A and B by TLC. This fraction was further partitioned by reversed phase HPLC (XBridge™ C18, 10 × 150 mm, 5 mm, gradient (Table 2) of acetonitrile and water with 0.1% formic acid, 7 mL/min). The fractions containing utahmycin A (16.5 min, 1 mg/L) and B (12.0 min, 2.5 mg/L) were dried down and analyzed by HRMS and NMR. The structure of utahmycin A was confirmed by X-ray crystallography of crystals obtained by slow evaporation from dichloromethane. An alternative HPLC gradient (Table 3) was used to isolate 8-O-methylutahmycin A, observed in some R2YE cultures of V167. 8-O-methylutahmycin A (25.0 min, less than 0.5mg/L) was characterized by ^{1}HNMR and HRMS.

Table 2: HPLC method for isolating utahmycin A and B

Time (min)	% Water with 0.1% formic acid	% Acetonitrile
Initial	80	20
2	80	20
32	0	100
37	0	100
40	80	20

Table 3: HPLC method for isolating 8-O-methylutahmycin A

Time (min)	% Water with 0.1% formic acid	% Acetonitrile
Initial	80	20
2	80	20
22	50	50
27	0	100
30	80	20

Expression and Isolation of Prechrysophanol

Prechrysophanol was observed in cultures of the wild type V167 grown in R2YE media for 14 days and in cultures of Erd20 transposon mutants grown in SMM. Pure prechrysophanol was obtained by Dr. John Bauer from ethyl acetate extracts of wild type V167 grown in R2YE for 14 days at 30°C and 250 rpm. A linear HPLC gradient (Table 4) of acetonitrile and water with 0.1% formic acid was used to isolate prechrysophanol (14.5 min). Dr. John Bauer elucidated the structure of prechrysophanol by NMR and X-ray crystallography. NMR and MS was used to confirm the presence of prechrysophanol in cultures of Erd20 transposon mutants in this thesis.

Table 4: HPLC method for isolating prechrysophanol

Time (min)	% Water with 0.1% formic acid	% Acetonitrile
Initial	80	20
2	80	20
26	40	60
27	0	100
32	0	100
35	80	20

Transposon Mutagenesis of V167 and Anaylsis of the Transposon Mutants

Transposon mutagenesis of the retrofitted V167 cosmid was carried out using the Genome Priming System with GPS®-2.1 (NEB). Two independent 19 μL GPS reaction mixtures (1X GPS Buffer, 0.01 μg pGPS2.1 Donor DNA, 300 ng V167 cosmid, and 1U TnsABC* Transposase) were incubated for 10 min at 37°C. After adding 1 μL of Start Solution, the reactions were incubated at 37°C for an additional hour, heat inactivated for 10 min at 75°C and transformed into EC100 E. *coli* (Epicentre). Transformants that grew on chloramphenicol (12.5 μg/mL) plates were pooled, minipreped and transformed into electrocompetent S17-1 E. *coli*.

Using sequencing Primer S (ATA ATC CTT AAA AAC TCC ATT TCC ACC CCT; NEB) 150 unique transposon mutants were sequenced to determine the location of the individual transposon insertions. Unique transposon mutants were individually mated into *S. albus* and visually screened for phenotypic changes on mannitol soya plates. Spore stocks of individual transposon mutants, which either exhibited a phenotypic change or were found in genes of interest, were inoculated into 75 mL cultures of modified SMM. After shaking (250 rpm) at 30°C for 4 days, these cultures were extracted twice with an equal volume of ethyl acetate. The pH of the remaining aqueous layer was adjusted to 3 with HCl and extracted two more times with an equal volume of ethyl acetate. The ethyl acetate extracts were

dried, dissolved in methanol (200 μL) and analyzed by analytical LC-MS (XBridge™ C18 (4.6 × 150 mm; 5 mm), 1.5 mL/min, acetonitrile and water with 0.1% formic acid gradient).

Table 5: Analytical HPLC method used for transposon mutant analysis

Time (min)	% Water with 0.1% formic acid	% Acetonitrile
Initial	80	20
2	80	20
22	0	100
27	0	100
30	80	20

Table 6: HPLC method used to purify SEK34, SEK34b, mutactin and dehydromutactin

Time (min)	% Water with 0.1% formic acid	% Acetonitrile
Initial	80	20
20	65	35
25	50	50
26	0	100
30	0	100
33	80	20
35	80	20

Table 7: HPLC method used to purify SEK4b and AUR367

Time (min)	% Water with 0.1% formic acid	% Acetonitrile
Initial	85	15
25	80	20
26	0	100
30	0	100
33	85	15
35	85	15

Expression, Isolation and Characterization of SEK4b, AUR367, mutactin, dehydromutactin, SEK34 and SEK34b from V167 Mutants

The compounds present in extracts from *S. albus* transformed with transposon mutants in Erd4, 5 and 22 were purified and characterized by MS and NMR. Transposon mutants were grown and extracted in the same manner described for wild type V167 cultures. SEK34 (5.9 min) and SEK34b (9.8 min) were isolated from crude extracts of the ΔErd22 mutant with an the HPLC gradient (Table 6) of acetonitrile and water with 0.1% formic acid. Mutactin (6.9 min) and dehydromutactin (10.4 min) were purified from extracts of the ΔErd4 GPS mutant using the same HPLC conditions. SEK4b (8.1 min) and AUR367 (12.2 min) were isolated from cultures of the ΔErd5 mutant using an alternative HPLC gradient (Table 7) of acetonitrile and water with 0.1% formic acid. MS and ^{1}H NMR data were used to confirm the identity of each known compound. SEK4b (HR-ESI-MS *m/z* 285.0791, calcd. for $C_{16}H_{13}O_5$, 285.0763; SEK34b) was crystallized from acetonitrile:water and its structure was confirmed by X-ray crystallography.

CHAPTER 3

Results

Library Construction and Screening for Type II PKS Pathways

In order to access the genetic diversity of uncultured microbes, the lab constructed a ten million-member cosmid eDNA mega-library from soil collected from the desert in Utah (UT). Degenerate primers targeting conserved regions of the KSα and ACP of the minimal PKS were used by Dr. John Bauer to amplify full-length KSβ sequences from the UT eDNA library. Sequencing of these KSβ amplicons revealed that, of the 21 KSβ genes amplified from the UT eDNA library, only one showed greater than 80% identity to a previously sequenced KSβ gene (Figure 7). A rule of thumb based on BLAST searches of KSβ genes indicates that two KSβ genes with greater than 80% identity likely encode for the same metabolite. While two KSβ genes with less than 80% identity likely encode for different metabolites. The low identity these UT eDNA-derived KSβ genes had with previously sequenced KSβ genes suggests they are not associated with known PKS gene clusters and likely encode for novel metabolites.

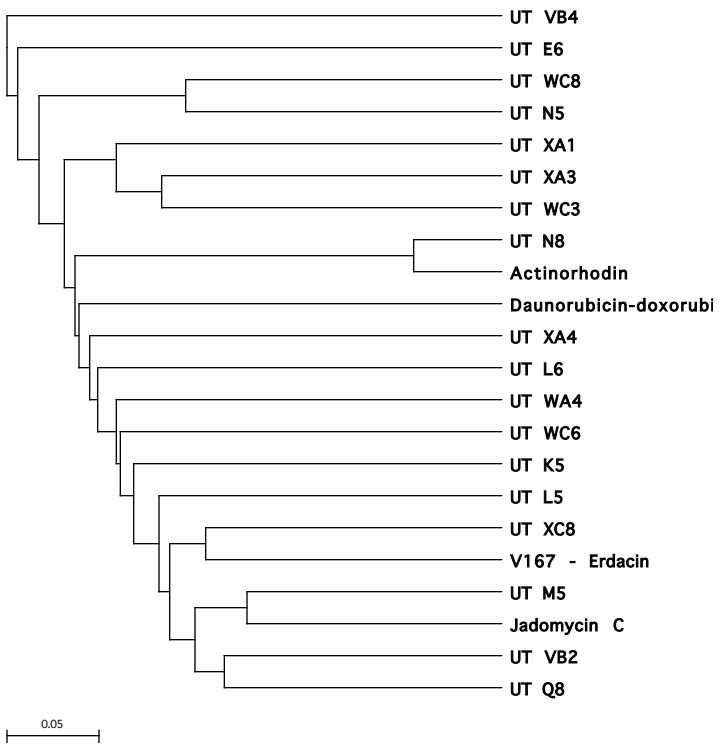


Figure 7: Phylogenic tree of sequenced KSβ amplicons from the UT library.
KSβ amplicons from the UT library sequenced by Dr. John Bauer. Only the UT-N8 KSβ amplicon showed greater than 80% identity to a previously sequenced actinorhodin KSβ gene. The low identity the other KSβ amplicons had with known genes indicates they are not associated with known PKS gene clusters and likely encode for novel metabolites.

Later, a second multimillion-member cosmid eDNA library was constructed by the lab from soil obtained from the Anza-Borrego (AB) desert in California. Screening of the AB library pools with degenerate primers indicated that the AB library is rich in type II PKS genes. Roughly one in every 12,000 clones in the AB library contains a KSβ gene. This high frequency of KSβ genes in the library may represent many unique gene clusters or a few profuse gene clusters. To distinguish these possibilities, KSβ gene amplicons, generated during the PCR screening of the AB library pools,

were sequenced to determine if the AB library was rich in unique KSβ genes. Of the 61 KSβ genes sequenced, 31 likely represent unique PKS gene clusters (Figure 8). Of these 31 unique KSβ genes identified in the AB library, 30 have less than 80% identity to previously sequenced KSβ genes and likely encode for novel metabolites. One KSβ gene makes up 38% (23/61) of the KSβ genes detected in the AB library. The closest homolog to this abundant KSβ gene is from the actinorhodin pathway of *S. coelicolor*. A second unique KSβ gene with homology to the actinorhodin gene cluster of *S. coelicolor* is represented 8 times. Two unique KSβ genes with closest homologs from *S. antibioticus* were both represented 3 times and the remaining 27 unique KSβ genes were each represented once. This sequencing data suggests the AB library is a rich source of unique novel type II PKS gene clusters.

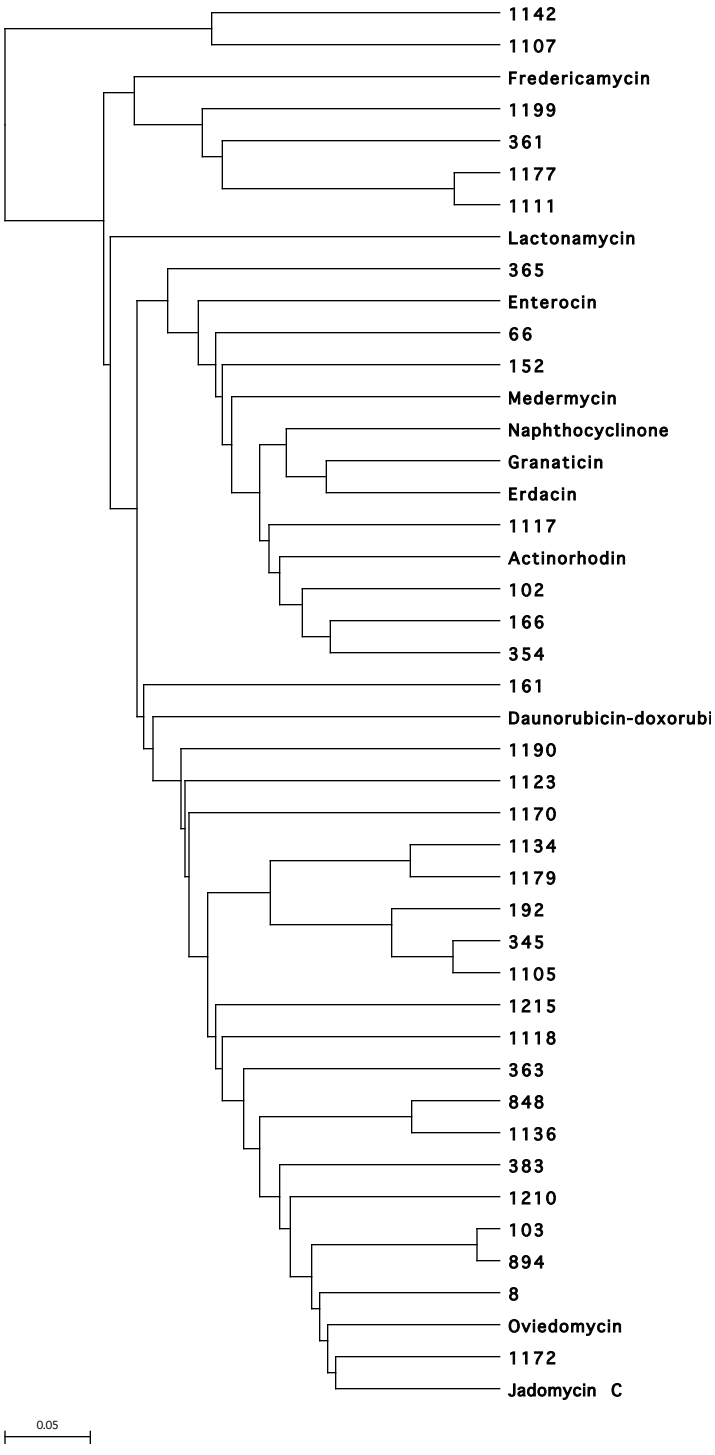


Figure 8: Phylogenic tree of sequenced KSβ genes amplified from the AB library. Of the 31 unique KSβ genes identified in the AB library, 30 had less than 80% identity to a previously sequenced KSβ gene and likely encode for novel metabolites.

Recovery and Heterologous expression of eDNA-derived Type II PKS pathways

PKS containing clones were recovered from the eDNA library and were heterologously expressed in order to access the structural diversity encoded by these gene clusters from uncultured bacteria. Clones containing PKS gene clusters were recovered from the libraries using KSβ genes as probes in successive rounds of library dilution and PCR screening. Dr. John Bauer recovered twelve unique eDNA clones containing type II PKS gene clusters from the UT eDNA library for heterologous expression studies. And with Paula Calle's assistance, nearly two dozen more clones containing PKS genes were recovered from the other AB library during the course of this work.

Knowing that any PKS gene clusters from the library that could be heterologously expressed in *E. coli* would likely have been identified in earlier phenotypic screens of the library, an alternative heterologous host, *Streptomyces*, was selected for expression studies. The predominantly soil dwelling gram-positive *Streptomyces* is synonymous with secondary metabolite production and produces two-thirds of clinically useful antibiotics derived from natural sources (Kieser, Bibb et al. 2000). *Streptomyces* have extensive precedent as a successful heterologous host. Furthermore, nearly all of the identified KSβ genes in this work have closest homology with KSβ genes from *Streptomyces*. In fact, only one of 31 KSβ genes identified in the

AB library had a closest homolog from outside *Streptomyces*. The closest homolog of this gene was with an *Actinomadura verrucosospora* KSβ gene. Recovered clones from the libraries containing type II PKS gene clusters were retrofitted so that they could be conjugated into *Streptomyces* for heterologous expression studies. Rare restriction enzymes (AvrII, NheI, PstI or PsiI) were used to remove PKS gene containing eDNA inserts from the library construction vector, without cutting the eDNA insert. The resulting intact linear eDNA was then ligated into the integrating *Streptomyces* shuttle vector pOJ436. These retrofitted clones were transformed into *E. coli* S17-1 donor strain and conjugated into *Streptomyces* using standard *Streptomyces* conjugation protocols. Dr. John Bauer successfully retrofitted and conjugated eight clones from the UT library into *S. albus*. Retrofitting and conjugation of the AB library clones was done by Dr. Sean Brady and myself, respectively, with the assistance of Paula Calle.

While other *Streptomyces* species were explored, *Streptomyces albus* proved to be an exceptional, laboratory friendly heterologous host because of its fast growth rate, ample sporulation and ease of mating. These factors helped facilitate both the conjugation of eDNA-derived Type-II PKS gene clusters into *S. albus* and the characterization of the molecules these pathways encode for. Eight of the clones recovered from the UT library were retrofitted with the genetic elements necessary for integration into *Streptomyces* and were successfully transformed into *S. albus*.

Recombinant *S. albus* clones from the UT and AB libraries were then screened for the production of clone specific metabolites. Antibacterial overlays, color production, TLC and HPLC-MS analysis were used to identified clone specific metabolites from cultures and crude extracts of recombinant clones (Figure 9). Five of the recombinant strains from the UT library were first observed by Dr. John Bauer to produce clone specific metabolites. An equivilant number of recombinant strains from the AB library were also observed to produce clone specific metabolites by visual inspection, TLC, or HPLC-MS of crude extracts. Two of these clones from the UT library (V167 and X26) and two clones from the AB library (848 and 1105) were examined in this work. However, the V167 clone was selected for a proof of principle study because, it produced large quantities (15-20 mg L^{-1}) of a purple clone specific metabolite (King, Bauer et al. 2009; Bauer, King et al. 2010).

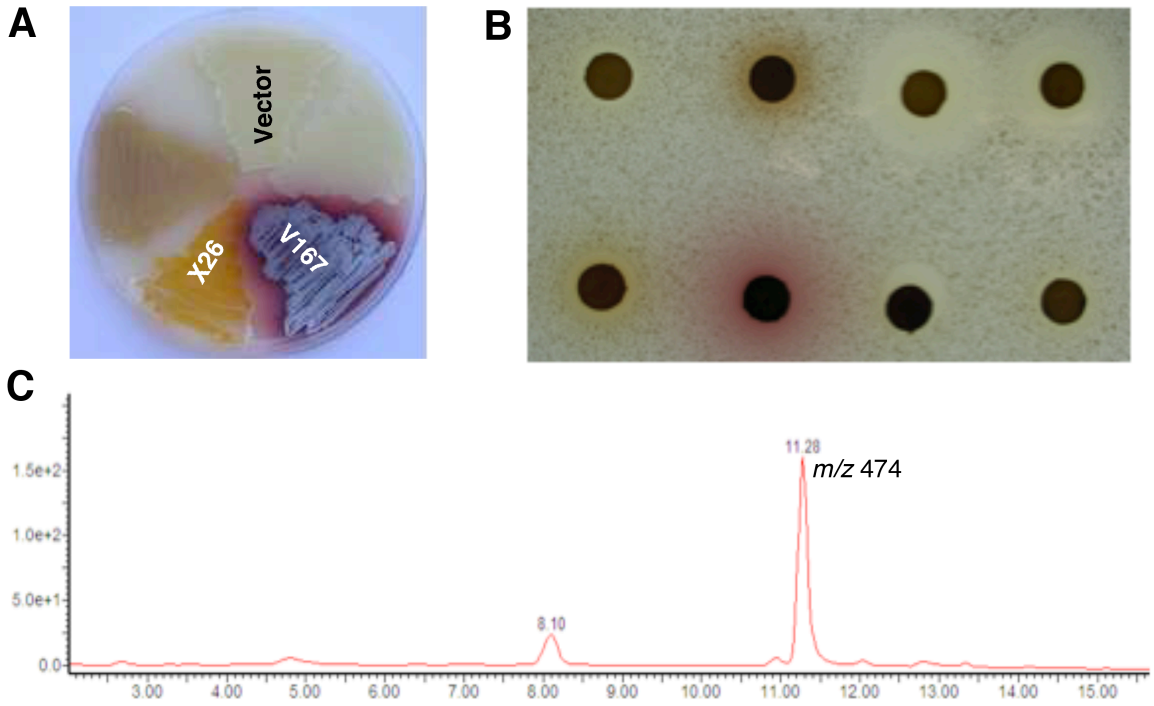


Figure 9: *S. albus* transformed with eDNA-derived PKS clones were screened for the production of clone specific metabolites.

Representative images of the screening process. A) Recombinant S. albus on MS agar plates to screen for colored metabolite production. The V167 and X26 clone are pictured. B) Crude neutral and acid ethyl acetate extracts of cultures spotted on a lawn of *B. anthrasis* in an antibacterial disk diffusion assay for four clones. C) Analytical LCMS trace of crude V167 culture extracts. V167 was selected for further studies because of the large quantities of clone specific metabolite it produced.

Isolation and Structural Characterization of eDNA Encoded Natural Products

To learn more about the molecules encoded by eDNA-derived Type-II PKS gene clusters, the major clone specific metabolite of the recombinant *S. albus* V167 clone was isolated and structurally characterized. Attempts were made to isolate and structurally characterize metabolites from other clones, including the X26 clone recovered from the UT library. However, the array of clone specific metabolites from the X26 clone appeared to be unstable and attempts to isolate them were unsuccessful. Efforts were focused on the isolation and structural elucidation of one clone specific metabolite produced in large quantities by the V167 clone.

Cultures of *S. albus* transformed with the retrofitted V167 clone were grown in modified supplemented minimal medium (SMM). After incubating at 30°C for 4 days, the cultures were acidified to pH 3 and extracted with ethyl acetate. This crude extract was passed through a SPE Previal™ C-18 plug and further partitioned by reversed phase HPLC to isolate the major metabolite, erdacin (HRMS (ESI) *m/z* 475.1037 $[M+H]^+$, calcd. for $C_{26}H_{19}O_9$, 475.1029). An extensive 1- and 2-D NMR analysis of naturally abundant and uniformly ^{13}C-labeled erdacin utilizing $^{1}HNMR$, $^{13}CNMR$, COSY, HMBC, HMQC and INADEQUATE was used to assemble a potential structure of the metabolite (Figure 10a).

Proton NMR revealed six aromatic protons, two methyls and one deshielded methanetriyl (4.29 ppm). The six aromatic protons included a triplet, three doublets and two protons in a multiplet. HMBC and COSY showed the triplet (7.28 ppm) and two of the doublets (6.92 and 7.16 ppm) formed a spin system. The remaining aromatic doublet (6.73 ppm) must couple with at least one of the protons in the multiplet (7.40 ppm). Given the complexity of the multiplet, a second d-t-d spin system can be assumed. This second spin system was confirmed in ^{1}HNMR and HMBC of chemically O-methylated erdacin where this multiplet was resolved into a distinct triplet and doublet (data not shown).

Carbon NMR was done on uniformly ^{13}C-labeled erdacin to increase carbon signal intensity. This carbon NMR data confirmed 26 carbon signals including the two methyl carbons, four carbonyls, two deshielded sp^3 hybridized carbons and 18 olefins. HMBC and HMQC further showed the two methyl groups were methyl ketones, where one was attached to an olefin (C-4) and the other to C-13 (64.6 ppm). Based on ^{13}C chemical shift, the two remaining carbonyls include an unsaturated ketone (C-1, 188.8 ppm) and a carboxylic acid (C-24, 176.6 ppm). Deshielding on three of the olefins suggested they were attached directly to heteroatoms. And HMQC revealed six olefins that were directly attached to the six aromatic protons.

The H-7, 8 to 9 spin system can be connected into an aromatic ring fragment with the olefins C-6, 10 and 11 based on HMBC correlations between these hydrogens and carbons. C-10 is one of the olefins attached directly to a heteroatom. HMQC determined the 4.29 ppm proton singlet was attached to a deshielded sp^3 hybridized carbon (C-13, 64.6 ppm). HMBC correlations from this H-13 singlet to C-5, 12, 14, 24, 25 and 26 enabled the assembly of a fragment containing the two deshielded sp^3 hybridized carbons, two olefins, the carboxylic acid and one of the methyl ketones. This fragment could then be connected to the H-7, 8 to 9 spin fragment to form partial structure **18-1** through two correlations between H-13 to C-11 and H-7 to C-5 (Figure 10a).

Based on HMBC correlations, the H-17, 18 to 19 spin system could also be assembled into an aromatic ring fragment with olefins C-16, 20 and 21. Deshielding of C-21 in carbon NMR suggests it is directly attached to a heteroatom. The placement of the unsaturated ketone (C-1) adjacent to C-21 to form fragment **18-2** is consistent with observed chemical shifts and 2D INADEQUATE NMR data.

These two partial structures (**18-1** and **18-2**) could be connected based on two key HMBC ^{1}H-^{13}C correlations to the olefin C-15 from H-17 and H-13. However, five carbons (C-2, C-3, C-4, C-22 and C-23) could not be assigned with certainty because of weak NMR signals. These remaining carbons included three olefins and required two ring closures to meet unsaturation

requirements. One of these olefins was attached to a methyl ketone and another was directly attached to a heteroatom (Figure 10b). There were a limited number of ways to attach these remaining pieces to partial structure **18-3** while closing two additional rings. Presented structure **18** is the possibility most consistent with all of the collected NMR data. However, it is difficult to distinguish by NMR this structure from the alternative structure where the positions of the C-2 hydroxyl and the C-4 methyl ketone are switched (Figure 10b).

Ultimately, the structure of erdacin was resolved by X-ray crystallography of crystals obtained by slow evaporation from acetonitrile and water (Figure 10c). The structure of erdacin revealed a potential C-2 hydroxy-assisted C-22 carbonyl tautomerization that could explain the weak and broad NMR signals observed in this region of the molecule. This tautomerization could also explain the disappearance of the C-23 methyl signal in deuterated, protic NMR solvents (Figure 11).

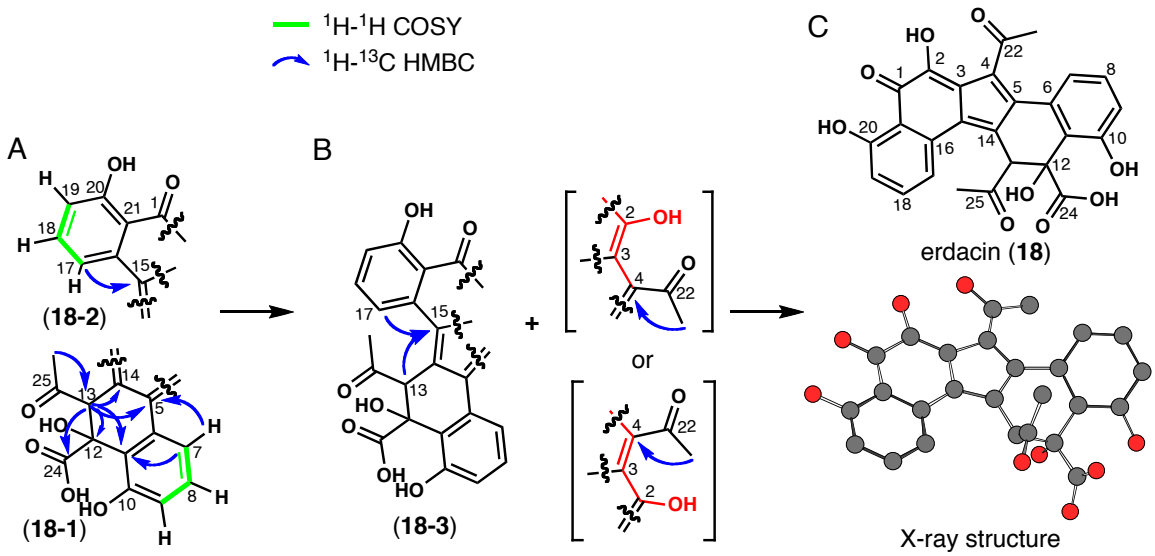


Figure 10: Structural elucidation of erdacin.

A) Key ^{1}H-^{1}H and ^{1}H-^{13}C NMR couplings used to assemble the two halves of erdacin are shown. B) Two key C-15 HMBC correlations from H-13 and H-17 allow the two halves to be connected. Extensive NMR analysis was unable to distinguish between two possible arrangements of the remaining atoms. C) Ultimately, the structure of erdacin was resolved by X-ray crystallography (grey = carbon and red = oxygen).

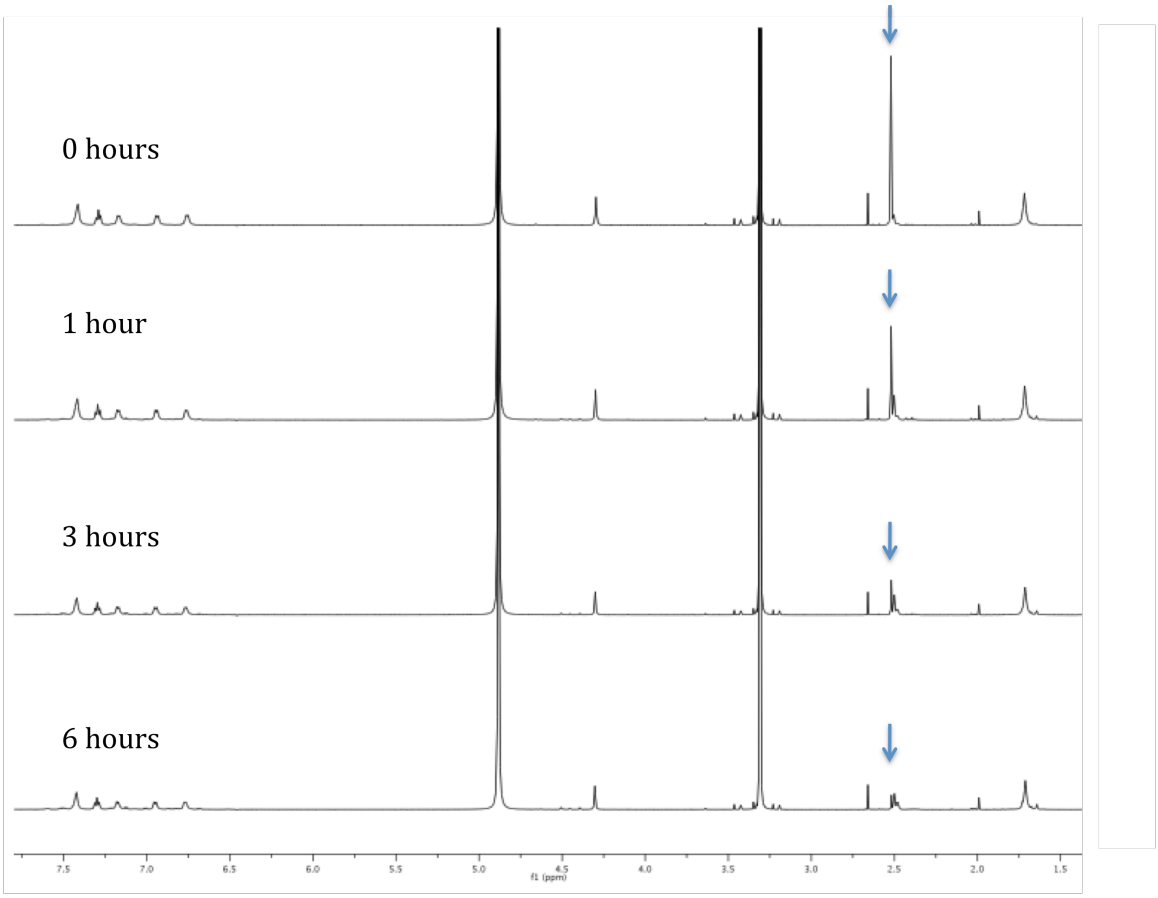


Figure 11: Hydroxy-assisted tautomerization of erdacin.

Time course charting the rapid loss of the 2.53 ppm C-23 methyl signal from ^{1}HNMR of erdacin in MeOD-d4. This is likely the result of a C-2 hydroxy-assisted C-22 carbonyl tautomerization.

The X-ray structure confirmed that erdacin is a novel metabolite. Neither the erdacin carbon skeleton nor the erdacin pentacyclic (6-6-5-6-6) ring system are substructures in known natural products (Figure 12). While pentacyclic ring systems with the same number of 5- and 6-membered rings are known, the particular arrangement of the erdacin ring system is novel. The pentacyclic backbone of erdacin is functionalized with acetyl groups, hydroxyls, a ketone and a carboxylate. The structural elucidation of erdacin provides evidence that eDNA-derived type II PKS gene clusters can encode for metabolites that are structurally distinct from those identified by traditional culture-dependent studies.

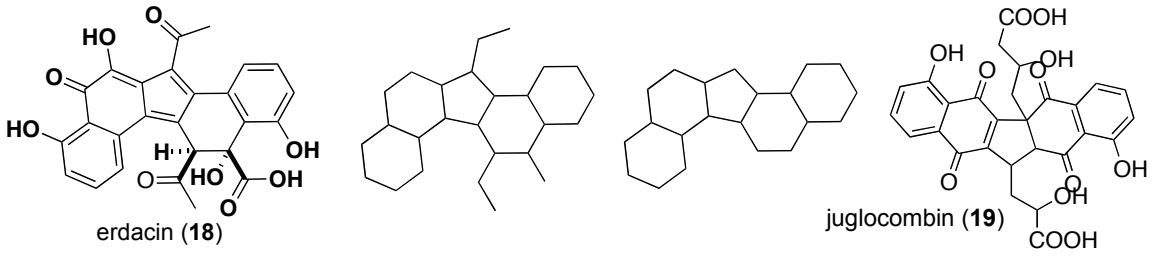


Figure 12: The novel structure of erdacin.

Both the carbon skeleton and the pentacyclic ring arrangement of erdacin are novel. Juglocombin is a previously reported metabolite with the same number and order of 5- and 6-membered rings, but the ring system is arranged differently.

Sequencing of PKS Containing eDNA-derived Clones

In order to develop a better understanding of the type-II PKS gene clusters recovered from the eDNA library, the V167 clone was sequenced (Figure 13a). Several contigs resulted from Dr. John Bauer's 454 sequencing

of the V167 clone. Primer walking, conducted for this thesis, to fill the gaps between these contigs proved not to be a trivial task due to the very high GC content of the eDNA insert. The completed sequencing efforts revealed the 20 kb erdacin gene cluster containing 22 predicted ORFs. These 22 ORFs include the minimal PKS, a ketoreductase, a cyclase, and an aromatase expected from a PKS gene cluster. Sequencing of the erdacin minimal PKS genes revealed they are most closely related to the minimal PKS genes of the naphthocyclinon, actinorhodin, granaticin and medermycin pathways (Figure 13b). Additionally, a collection of post-PKS enzymes, regulatory proteins, a transporter, and several hypothetical proteins are also encoded for by the erdacin gene cluster.

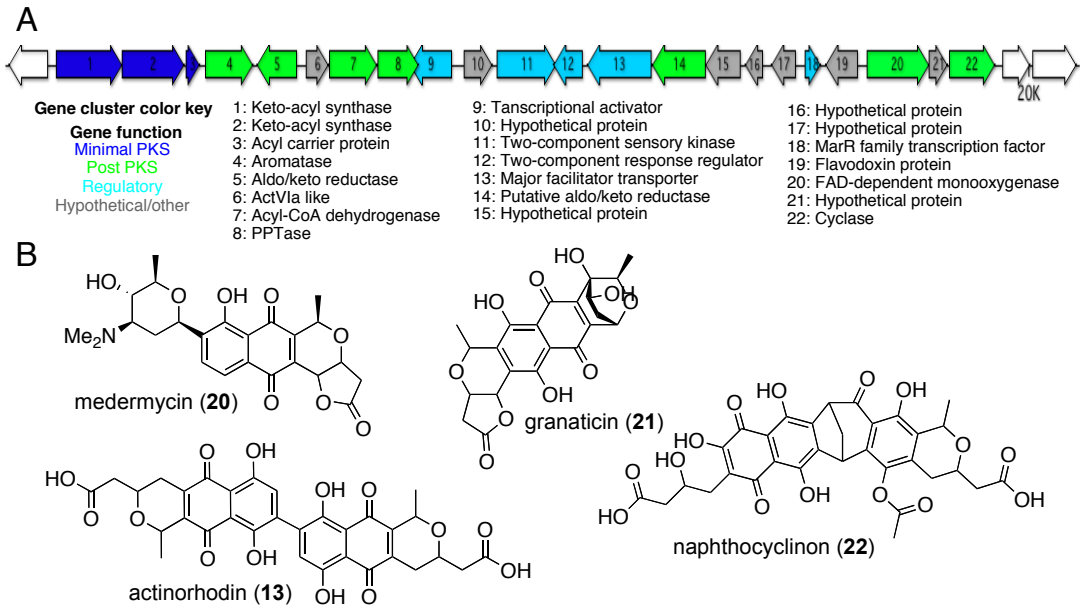


Figure 13: The erdacin gene cluster.

A) The erdacin gene cluster. B) Erdacin minimal PKS genes are most closely related to the minimal PKS genes of medermycin, actinorhodin, granaticin and naphthocyclinon gene clusters.

In addition to sequencing the V167 clone, a number of additional clones from the UT and AB libraries were sequenced by Dr. John Bauer and myself, respectively. Sequencing of the X26 clone from the UT library revealed a truncated PKS gene cluster with a minimal PKS (Figure 14). A cosmid that captured an incomplete PKS gene cluster from eDNA could explain the array of unstable clone specific metabolites observed in cultures of *S. albus* harboring the X26 clone.

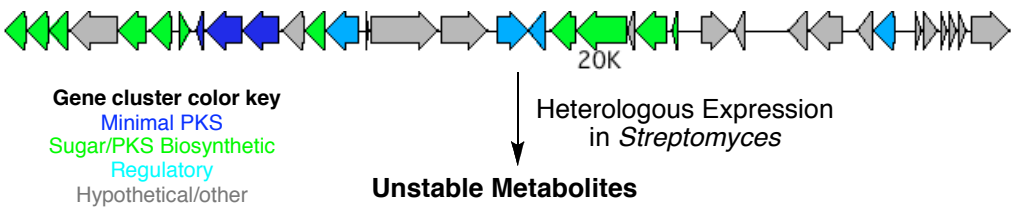


Figure 14: Sequencing of the X26 clone.
The X26 clone contains a minimal PKS and a truncated PKS gene cluster. Heterologous expression of the truncated X26 clone lead to the production of a number of uncharacterized unstable metabolites.

Erdacin Activity

With the structure of erdacin elucidated, several assays were conducted to explore the potential activity of erdacin. Erdacin (50-100µg/disk) was assayed against a number of Gram-positive and Gram-negative microbes in an antibacterial agar disk diffusion assay. However, no zone of inhibition was observed against the tested microbes. Assays of erdacin against mammalian cell lines also showed no activity. Erdacin did, however, show strong antioxidant activity. In fact, erdacin demonstrated twice the antioxidant activity as the known antioxidant ascorbic acid (vitamin C) as

determined by a standard antioxidant copper reduction assay (Apak, Kubilay, et al 2004). Since the bacteria that encodes for erdacin remains uncultured, it is difficult to know the function of erdacin in its native microbe.

Isolation of Juglomycin F

A number of clone specific novel and known metabolites, in addition to erdacin, were characterized from cultures of *S. albus* conjugated with the eDNA-derived V167 clone to develop a better understanding of this type-II PKS. In addition to the major metabolite erdacin, a second clone specific metabolite was identified in four-day-old SMM cultures of the V167 clone. *S. albus* cultures were extracted and this metabolite was isolated by reversed phase HPLC. NMR and MS analysis established the structure of this molecule was the known compound juglomycin F, confirming earlier observations by Dr. John Bauer of V167 in alternative culture conditions. The structure of juglomycin F maps onto half of the erdacin structure, providing a hint to the biosynthetic origin of erdacin. This observation raised the possibility that erdacin is derived from the heterodimerization of juglomycin F and a second 13-carbon subunit (Figure 15).

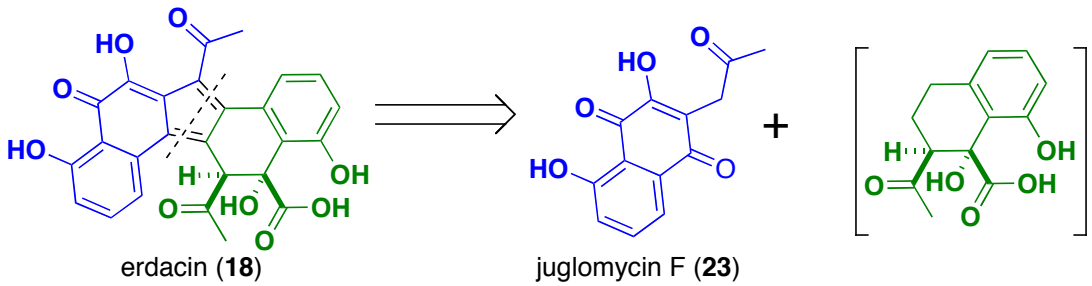


Figure 15: Retrobiosynthetic scheme for the production of erdacin.

Erdacin appears to arise from two 13-carbon monomers. Juglomycin F, isolated from V167 cultures, maps onto half of the erdacin structure. The hypothetical second 13-carbon polyketide contains the carbons needed to complete the structure of erdacin.

Isolation and Characterization of Utahmycins and Prechrysophanol

A number of additional compounds were observed in alternative culture conditions of *S. albus* conjugated with the V167 clone. In these alternative conditions, V167 cultures were incubated for a longer time (7-14 days) in R2YE media instead of SMM media. One of the metabolites produced was extracted, isolated using HPLC and identified by NMR, MS and crystallography as the known octaketide product prechrysophanol by Dr. John Bauer. The isolation of prechrysophanol provides another hint to the biosynthetic origin of erdacin suggesting that the erdacin minimal PKS may produce a nascent octaketide backbone.

Three of the remaining compounds observed in R2YE cultures had odd masses ($[M+H]^+$ *m/z* 228, 254, and 268) suggesting they contained a nitrogen and were distinct from erdacin and its predicted biosynthetic intermediates. Two of these metabolites were extracted and fractionated by

silica flash chromatography with a $CHCl_3$/MeOH step gradient. The fraction containing both of these compounds was further partitioned by reverse phase HPLC. HRMS data was collected on these two metabolites named utahmycin A (observed *m/z* 254.0829 $[M + H]^+$ (calcd for $C_{15}H_{12}NO_3$, 254.0817)) and utahmycin B (obsered *m/z* 228.0672 $[M + H]^+$ (calcd for $C_{13}H_{10}NO_3$, 228.0661)). The structure of utahmycin B was resolved by Dr. John Bauer using MS, ^{1}HNMR and HMBC NMR data. Missing publication quality ^{13}CNMR, COSY, HMBC, HMQC and other empirical data of utahmycin B was collected for this thesis after obtaining more material.

Work toward this thesis resolved the structure of utahmycin A using ^{1}H, ^{13}C, COSY, HMBC and HMQC NMR spectra (Figure 16). Proton NMR of utahmycin A revealed two methyl singlets, four aromatic protons and an alcohol. COSY and ^{1}HNMR showed three of the aromatic protons form a proton spin system (d-t-d) and the remaining aromatic proton is a singlet. Carbon NMR and HMQC revealed the two methyl carbons, two carbonyls, three olefins directly attached to heteroatoms, four olefins attached to a hydrogen and four additional olefins. The H-5, 6 and 7 proton spin system could be assembled into an aromatic ring with C-8, 11 and 12 based on HMBC correlations. The high deshielding of C-8 and HMBC correlations from the alcohol to C-7, 8 and 12 enabled the alcohol to be positioned on C-8. One of the two carbonyls (C-10) could be placed adjacent to C-11 based on an HMBC correlation between H-5 and C-10. C-10 also had an HMBC

correlation to the singlet H-4. Additional HMBC correlations from H-4 to C-3, 13, 14 and 16 were used to assemble the framework for the second half of utahmycin A. This was further supported by correlations from the methyl H-16 protons to C-3 and C-4. The second methyl group and the final olefin could both be connected to the structure based on correlations between H-15 protons and C-1 and 13. In this structure both of the methyl groups are attached to deshielded olefins (C-1 and C-3) that are attached directly to a heteroatom. With only a nitrogen and carbonyl left to place, the only way to satisfy the deshielded ^{13}C chemical shifts is to connect C-1 and C-2 through a nitrogen. The remaining carbonyl (C-9) could then only be placed one way to meet the unsaturation requirements of the chemical formula, between C-12 and C-13. The structure of utahmycin A was further confirmed by X-ray crystallography.

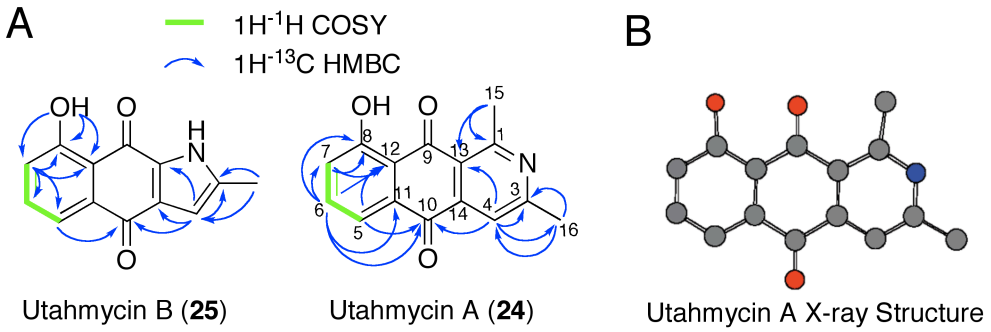


Figure 16: Structural elucidation of Utahmycin A and B.

A) Key ^{1}H-^{1}H and ^{1}H-^{13}C correlations used to elucidate the structures of Utahmycin A (**24**) and B (**25**) are shown. B) The structure of Utahmycin A was confirmed by X-ray crystallography (grey = carbon, red = oxygen and blue = nitrogen).

The third nitrogen-containing metabolite, obtained in much lower titers (< 0.5mg/L), was isolated. HRMS and ^{1}HNMR of this metabolite, suggested a compound related to the other two utahmycins. The absence of a hydroxyl signal and the appearance of a deshielded methyl group in ^{1}HNMR was consistent with HRMS data (HRMS (ESI) *m/z* 268.0953 $[M+H]^+$, calcd. for $C_{16}H_{14}NO_3$, 268.0974) and an O-methylation of utahmycin A.

The structure of Utahmycin A is novel and is the first 1,3-dimethyl-2-azaanthrqquinone reported from a biological system. While utahmycin B has been previously synthesized, it has not been observed in a biological system. 2-azaanthraquinones have primarily been isolated from fungi and lichens and present a wide range of activity from antibacterial to antimalarial activity (Figure 17b) (Zhang, Watanabe et al. 2006). In fungi 2-azaanthraquinones are typically found with structurally related metabolites, which contain an oxygen in place of the nitrogen of 2-azaanthraquinones (Medentsev, Akimenko et al. 1988; Parisot, Devys et al. 1990; Kurobane, Vining et al. 1980; Kurobane, Zaita et al. 1986). No antibiotic activity was observed when utahmycin A and B (50-100μg/disk) were assayed against a number of Gram-positive and Gram-negative microbes in a disk diffusion assay. The isolation of erdacin, utahmycin A, utahmycin B and 8-O-methylutahmycin A from a single eDNA-derived clone, indicates that eDNA-derived type-II PKS gene clusters are a rich source of novel natural products.

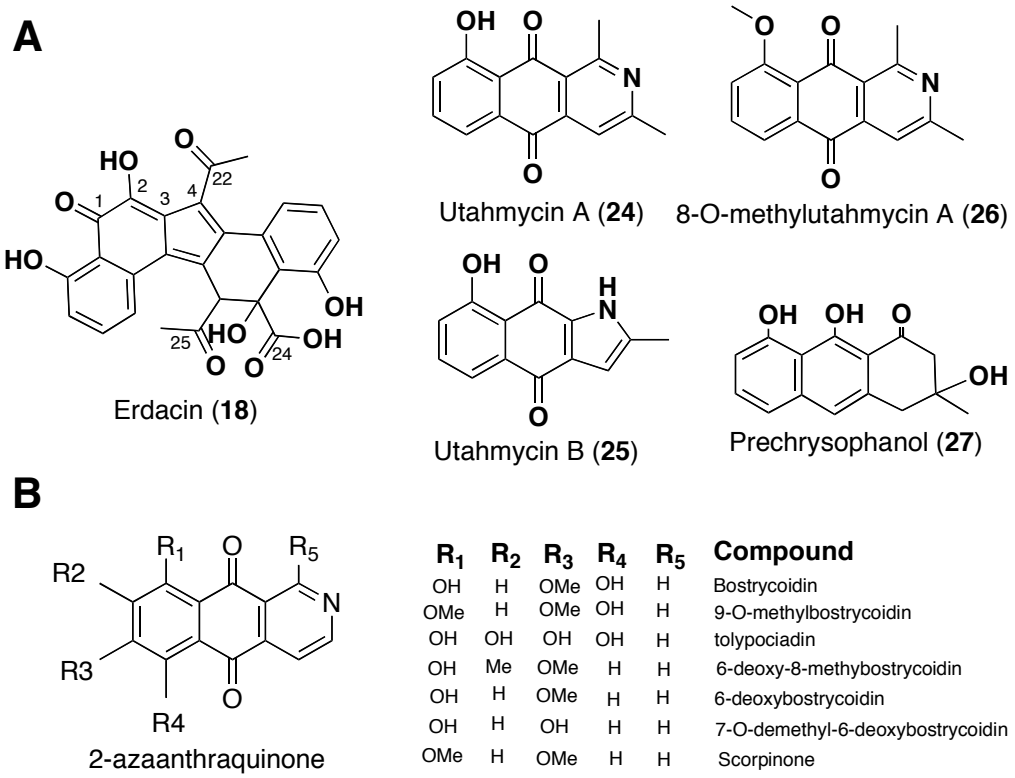


Figure 17: Metabolites of the eDNA-derived V167 clone.

A) Multiple metabolites were isolated from *S. albus* transformed with the V167 clone. These included the novel structure erdacin, utahmycin A, utahmycin B, 8-O-methylutahmycin A and the known polyketide prechrysophanol. B) Known 2-azaanthraquinones.

Biosynthesis of Erdacin

Knockouts and octaketide shunt products

A biosynthetic analysis of the eDNA-derived type-II PKS gene cluster that encodes for the novel structure and pentacyclic ring arrangement of erdacin was done. This biosynthetic analysis included transposon mutagenesis experiments to generate a number of single gene knockout erdacin gene clusters. Single gene knockouts of the retrofitted V167 clone were created by insertion of a chloramphenicol resistance gene using the

Genome Priming System with GPS®-2.1 (NEB, Ipswich, MA). The location of insertion was determined for 150 unique transposon mutants by Sanger sequencing. Due to the high GC content (>90%) of large regions of the Erd gene cluster, transposon insertions did not distribute evenly throughout the eDNA insert and, therefore, no transposon insertions were identified in several GC rich genes. A large number of unique transposon mutants were individually mated into *S. albus* and visually screened for phenotypic changes. The chemotypes of individual transposon mutants, which either exhibited a phenotypic change or were found in genes of interest by sequencing, were analyzed by analytical LC-MS and TLC of culture extracts.

Transposon mutants effecting the expression of erdacin all fell within a 20 kb region of the V167 clone, which was defined as the erdacin gene cluster (Figure 18b). The 20 kb erdacin gene cluster contains 22 predicted ORFs including the expected minimal PKS (Erd1, -2, and -3), ketoreductase (Erd5), cyclase (Erd22) and aromatase (Erd4) of PKS gene clusters. A collection of post-PKS enzymes, regulatory proteins, a transporter, a 4'-phosphopantetheinyl transferase and several hypothetical proteins are also located within the gene cluster. Transposon knockouts of the minimal PKS (Erd-1, -2, and -3) and a transcriptional activator (Erd-9) abolished clone specific secondary metabolite production. And clones with insertions in Erd4, -5, -20 and -22 showed a change in chemotype by visual inspection and LC-MS of extracts (Figure 18a)

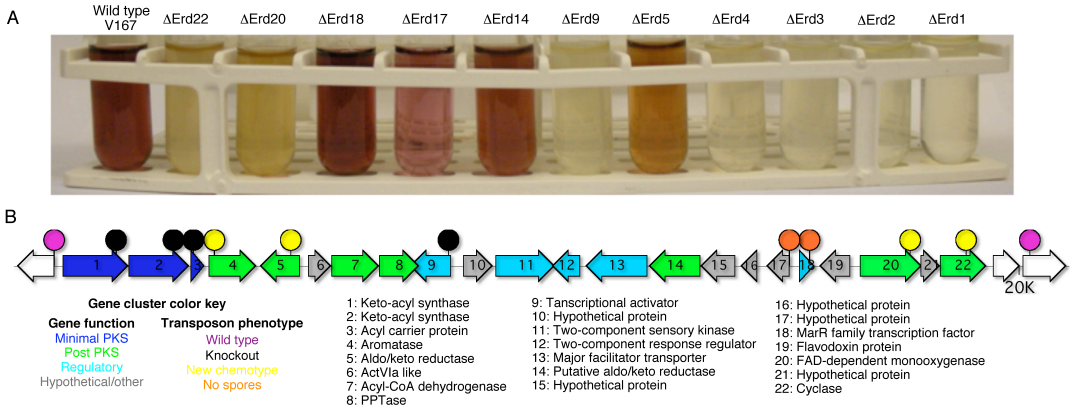


Figure 18: V167 transposon mutants.

A) Color changes observed in extracts of *S. albus* conjugated with mutant erdacin gene clusters. B) The locations of key transposon insertions in the erdacin gene cluster.

The compounds found in extracts of *S. albus* transposon mutants ΔErd4, -5, -20 and -22 were extracted, isolated and identified by MS and NMR to develop a better understanding of the erdacin gene cluster. Extracts of ΔErd5 ketoreductase mutant cultures were found to contain the known octaketide shunt products SEK4b and AUR367. These two compounds are derived from an alternative first cyclization that occurs when the nascent polyketide backbone partially falls off the minimal PKS (Fu, Hopwood et al. 1994; Seow, Meurer et al. 1997). This alternative first cyclization indicates a failure of the minimal PKS to cyclize the nascent polyketide backbone and is known to occur when the proper minimal PKS-ketoreductase complex does not form (other Ref on ketoreductase). Two more known octaketide shunt products, mutactin and dehydromutactin, were isolated from extracts of the ΔErd4 aromatase mutant. The presence of mutactin and dehydromutactin is consistent with an octaketide PKS gene cluster with an aromatase knockout

(Zhang, Adefarati et al. 1990; Fu, Khosla et al. 1994; Staunton and Weissman 2001; Meisner, Hintersteiner et al. 2007). Furthermore, SEK34 and SEK34b were isolated from crude extracts of the ΔErd22 cyclase mutant. X-ray crystallography was used to confirm the structure of SEK34b. The observation of SEK34 and SEK34b is consistent with a cyclase knockout (McDaniel, Khosla et al. 1994; Staunton and Weissman 2001). The deletion of Erd20, a predicted FAD-dependant monooxygenase, led to the production of the known octaketide product prechrysophanol (also identified in R2YE cultures of wild-type V167) and a number of other compounds with a range of molecular weights from m/z 240 to 315.

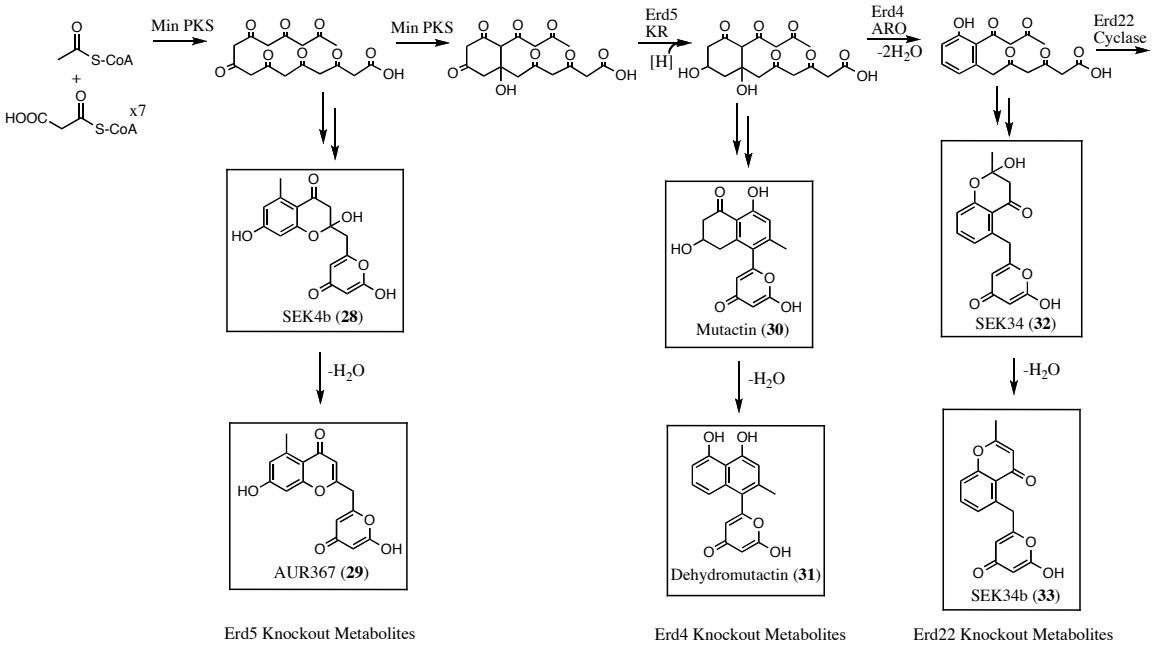


Figure 19: Biosynthetic scheme of metabolites produced by V167 early post-PKS gene knockouts.

SEK4b, AUR367, mutactin, dehydromutactin, SEK34 and SEK34b were isolated from cultures of early erdacin PKS gene knockouts. These metabolites isolated from cultures of ΔErd4, ΔErd5 or ΔErd22 mutants are known octaketide metabolites, indicating erdacin is derived from a 16-carbon backbone.

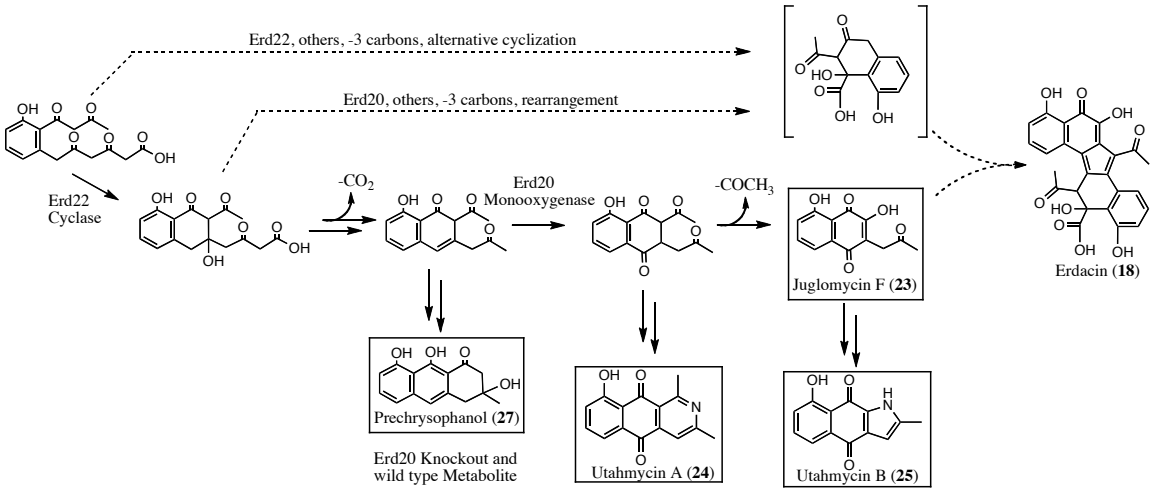


Figure 20: Biosynthetic scheme for the production of erdacin.

Evidence indicates that erdacin may be derived from the dimerization of two 13-carbon subunits, but the erdacin minimal PKS appears to produce an octaketide. A biosynthetic scheme for the production of these two 13-carbon monomers from an octaketide is shown. Utahmycin A, utahmycin B and prechrysophanol, isolated from wild type V167 cultures, are derived from intermediates in the proposed biosynthesis of erdacin.

Knockouts of early PKS genes lead to the production of known octaketide shunt products, suggesting that the minimal PKS of erdacin produces an octaketide backbone. However, the isolation of juglomycin F from wild type cultures hints that erdacin is derived from the heterodimerization of two distinct 13-carbon subunits (Figure 15). A biosynthetic scheme is proposed where two 13-carbon subunits of erdacin are derived from the excision of three carbons from octaketide precursors produced by the minimal PKS (Figure 19 and 20). While the isolation of juglomycin F from wild type cultures supports the existence of one 13-carbon intermediate, the second 13-carbon subunit was not identified. The two most likely possibilities of the origin of the second 13-carbon subunit share a common first cyclization with the juglomycin F monomer and are then derived from a rearrangement or an alternate second cyclization (Figure 20). This biosynthetic scheme is further supported by the isolation of prechrysophanol (**27**), utahmycin A (**24**) and utahmycin B (**25**), which are derived from various biosynthetic intermediates of erdacin production. Prechrysophanol, observed in wild type and ΔErd20 mutant cultures, is a known decarboxylated octaketide and its biosynthesis is understood. Utahmycin A and B both appear to be derived from the incorporation of nitrogen into erdacin biosynthetic intermediates (Figure 20).

Further evidence that the erdacin minimal PKS produces an octaketide comes from sequencing data that revealed Erd1, -2, and -3 are most closely related to the minimal PKS genes of the octaketide encoding

naphthocyclinon, actinorhodin, granaticin and medermycin pathways (Figure 13). Naphthocyclinon (**22**) and actinorhodin (**13**) are both octaketide dimers, which arise from two different dimerization strategies (McDaniel, Khosla et al. 1994; Ichinose, Bedford et al. 1998; Brucker, McKinney et al. 1999; Ichinose, Ozawa et al. 20003). Interestingly, erdacin appears to use an alternative octaketide heterodimerization strategy.

Cyclization Pattern of Erdacin

Knockout experiments suggest that octaketide precursors are converted into two 13-carbon monomer subunits that heterodimerized to produce erdacin. One of these 13-carbon monomers is represented by juglomycin F, isolated from wild type V167 cultures. A number of possibilities of the origin of the unidentified second 13-carbon subunit exist, including the two most likely possibilities that share a common first cyclization with the juglomycin F monomer and are then derived from a rearrangement or an alternate cyclization. An [1,2-^{13}C]acetate feeding experiment was conducted to distinguish the numerous folding pattern possibilities that could result in the second 13-carbon monomer.

Cultures of the V167 clone were feed one part labeled [1,2-^{13}C]acetate and two parts naturally abundant acetate. The labeled [1,2-^{13}C]acetate units were incorporated into erdacin by the bacteria during biosynthesis and the resulting labeled erdacin was isolated. On average, every incorporated labeled acetate unit was adjacent to two non-labeled acetate units. ^{13}C-^{13}C

coupling between the two carbons of labeled acetate units in 1- and 2D ^{13}CNMR of [1,2-^{13}C]acetate-labeled erdacin was then used to determine which carbons in erdacin were incorporated as intact acetate units. For example, C-23 appears as a singlet in ^{13}CNMR of [1,2-^{13}C]acetate-labeled erdacin, however, C-23 shows coupling to C-22 and appears as a doublet in ^{13}CNMR of erdacin that is uniformly labeled with ^{13}C (Figure 21a). The absence of coupling between C-23 and C-22 in the acetate labeled erdacin indicates that C-22 and C-23 arise from different acetate units during biosynthesis.

The experimentally observed folding pattern of erdacin is consistent with the octaketide (S′)-folding of the juglomycin F monomer in the proposed biosynthetic scheme (Figure 20). Specifically, observed coupling between C-20 to C-21, C-18 to C-19, and C-16 to C-17 of erdacin in these labeling experiments supports the proposed folding pattern of the juglomycin F monomer of erdacin (Figure 21, 22a). This proposed folding pattern is further supported by the absence of coupling between C-23 and C-22 in 1D ^{13}CNMR acetate labeled experiments, which can be observed by ^{13}CNMR if erdacin is uniformly ^{13}C-labeled (Figure 21a). This folding pattern of the juglomycin F monomer resembles the octaketide *Streptomyces* (S′)-folding of chrysophanol and actinorhodin (Bringmann, Irmer et al. 2009). Utahmycin A and B are also derived from this octaketide folding pattern. This differs from other known 2-azaanthraquinones, which are derived from a fungal (F)-folding of a septaketide (Thomas 2001; Van Wagoner, Mantle et al. 2008). Utahmycin A

and B are derived from both an alternative folding pattern and a larger polyketide precursor than previously known fungal 2-azaanthraquinones.

There are a number of possibilities for the biosynthetic origins of the hypothetical second 13-carbon subunit (Figure 22b). The two most likely possibilities share a common first cyclization with the juglomycin F monomer and are then derived from a rearrangement (***i***) or an alternate cyclization (***ii***). If the two 13-carbon monomers share common first and second cyclization patterns and the second monomer arises from a rearrangement (***i***), coupling between C-26 and C-25 is expected, but C-24 would not couple with the adjacent C-12 in erdacin. However, these couplings expected of this folding pattern were not observed in 2D INADEQUATE experiments (Figure 21a). In fact, coupling was observed between C-24 and C-12, but coupling was not observed for C-26. Furthermore, C-25 and C-13 appear to come from the same acetate unit, as coupling was observed between them. This 2D INADEQUATE data was supported by 1D ^{13}CNMR data comparing ^{13}C-^{13}C coupling of uniformly ^{13}C-labled and [1,2-^{13}C]acetate-labeled erdacin (Figure 21a). This observed coupling is consistent with the second monomer arising from a common first cyclization, but an alternative second cyclization to the juglomycin F monomer (***ii***). This folding pattern of an alternative second cyclization (***ii***) is the most likely biosynthetic possibility supported by the observed data.

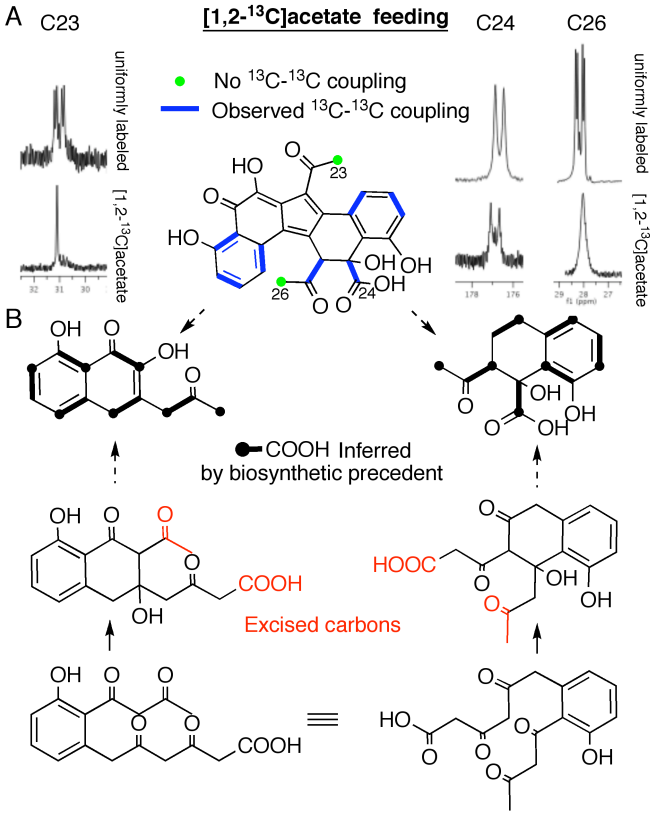
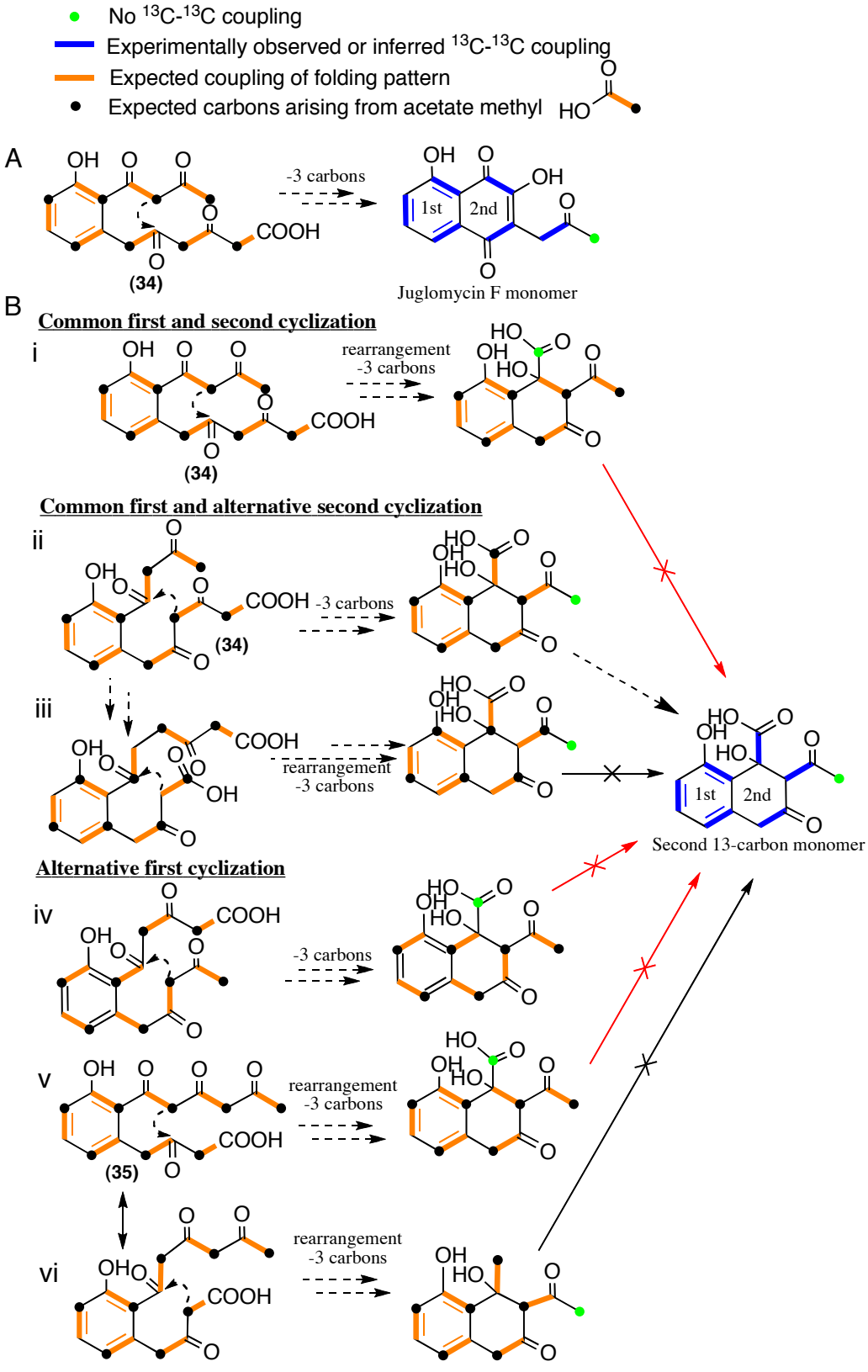


Figure 21: The cyclization pattern of erdacin and two 13-carbon precursors.

(A) Observed ^{13}C-^{13}C coupling of [1,2-^{13}C]acetate-labeled erdacin in $^{13}CNMR$ and INADEQUATE NMR spectra. (B) Coupling data suggests the two 13-carbon monomers that make up erdacin are derived from two different cyclization patterns.

Figure 22: Biosynthetic possibilities of the labeled acetate feeding experiment.
The expected carbon couplings of several possible folding patterns for the 13-carbon monomers that makeup erdacin and the experimentally determined folding patterns of the two monomers. (A) The observed folding pattern of erdacin is consistent with the proposed octaketide (S′)-folding of the Juglomycin F monomer similar to chrysophanol and actinorhodin (Figure 20). (B) Several possibilities for the biosynthetic origins of the second 13-carbon monomer are shown. (*i*) The second 13-carbon monomer could share a common first and second cyclization with the juglomycin F monomer and arise from a rearrangement of one of the carbons. The expected folding pattern of this possibility is not consistent with the observed folding pattern of the second ring. (*ii*) The second monomer could be derived from a common first cyclization, but an alternative second cyclization (Figure 20). The observed folding pattern is consistent with the expected folding pattern of this possibility and is the most likely biosynthetic possibility supported by the observed data. (*iii*) Another alternative second cyclization is also consistent with the observed data, but can be ruled out as it is inconsistent with known polyketide biosynthesis. (*iv*) An alternative first cyclization that resembles the fungi specific (F)-folding pattern is an unlikely possibility, as the erdacin gene cluster appears to be derived from *Streptomyces*. (*v*) The expected folding pattern of another alternative first cyclization resulting in (S)-folding similar to chrysophanol is not consistent with the observed folding pattern of the second ring. (*vi*) Another possibility could have the same first (S)-folding cyclization as ***v***, but have an alternative second cyclization pattern. The expected folding pattern of this possibility is also consistent with the observed folding pattern, however, this possibility is unlikely as it would require the minimal PKS of erdacin to catalyze two different first cyclization reactions.


No 13C-13C coupling
Experimentally observed or inferred 13C-13C coupling
Expected coupling of folding pattern
Expected carbons arising from acetate methyl
A
-3 carbons
1st
2nd
(34)
Juglomycin F monomer
B
Common first and second cyclization
i
rearrangement
-3 carbons
(34)
Common first and alternative second cyclization
ii
-3 carbons
(34)
iii
rearrangement
-3 carbons
1st
2nd
Second 13-carbon monomer
Alternative first cyclization
iv
-3 carbons
v
rearrangement
-3 carbons
(35)
vi
rearrangement
-3 carbons

Another biosynthetic possibility for the origin of the second 13-carbon monomer of erdacin is a second folding pattern that shares a common first cyclization with the juglomycin F monomer, but has an alternative second cyclization (***iii***). In this biosynthetic possibility, the second 13-carbon monomer would arise from a rearrangement and loss of three carbons from the (S`)-folded octaketide **34**. Alternatively, the same coupling pattern could be derived from the loss of three carbons from the (S)-folded octaketide **35**. In either case, the expected coupling of this folding pattern would be consistent with the observed data (Figure 22b iii), but inconsistent with polyketide biosynthesis. In fact, this possibility would require the addition of an oxygen functional group to four different carbon positions and the removal of oxygen from at least two additional carbons. Possibility ***iii*** is consistent with observed data, but can be ruled out because it is inconsistent with known polyketide biosynthesis.

Three other possibilities arise from alternative first cyclizations of the octaketide backbone. An alternative first cyclization that resembles the fungi specific (F)-folding pattern is inconsistent with observed data and is unlikely as the erdacin gene cluster appears to be derived from *Streptomyces* (***iv***). The expected folding pattern of an alternative first cyclization leading to (S)-folding is inconsistent with the observed folding pattern of the second ring (***v***). However, another possibility could share a first (S)-folding cyclization with possibility ***v***, but have an alternative second cyclization pattern (***vi***). The

expected folding pattern of this possibility is also consistent with the observed folding pattern. However, this possibility is unlikely as it would require the minimal PKS of erdacin to catalyze two different first cyclization reactions. Furthermore, a rearrangement would be required, making possibility ***vi*** far more complex than ***ii***.

The possibility most consistent with known biosynthetic precedent that was supported by 1- and 2D ^{13}CNMR data from the acetate feeding experiment, suggests the second 13-carbon monomer shares a first cyclization with the juglomycin F monomer and then arises from an alternate second cyclization. Interestingly, the biosynthetic scheme suggested by this data utilizes an alternate second cyclization, but results in the same three carbons being excised from the octaketide precursors of both monomers (Figure 21b).

CHAPTER 4

Discussion

Abundance of type II PKS gene clusters in eDNA from soil

Uncultured bacteria are thought to encode for a plethora of previously undiscovered small molecules. Libraries of DNA isolated from environmental samples were used to access the genetic diversity of these uncultured microbes. Previous studies have indicated that eDNA samples are rich in Type II PKS genes (Wawrik, Kerkhof et al. 2005; Pang, Tan et al. 2008; Wawrik, Kutliev et al. 2007). Sequencing of KSβ gene amplicons from the UT and AB libraries confirmed eDNA libraries are rich in PKS gene clusters. In fact, 21 unique KSβ genes were amplified from the UT library and another 31 were identified in the AB library. Only two of the genes identified in the AB and UT libraries showed greater than 80% identity to known KSβ genes. The remaining pathways all had less than 80% identity to previously sequenced KSβ genes and likely encode for novel small molecules. However, no new metabolites have been produced from eDNA-derived type II PKS gene clusters prior to this work.

Novel structures from uncultured bacteria

To explore the structural diversity encoded by uncultured bacteria, cosmids containing type II PKS gene clusters were recovered from an eDNA library using KSβ genes as a probe. The recovered clones were heterologously expressed in *S. albus* and screened for the production of clone specific metabolites. A number of the clones isolated from the AB and UT libraries were found to produce clone specific metabolites. One of these clones (V167) was selected for further characterization, because it produced large quantities of a stable clone specific metabolite. The molecule the V167 clone encodes for was structurally elucidated. This metabolite, erdacin, was found to have not only a novel carbon skeleton, but also a novel pentacyclic ring arrangement. The structural elucidation of the novel secondary metabolite erdacin provides tangible evidence that eDNA-derived type II PKS gene clusters have the potential to encode for metabolites that are structurally distinct from those identified by culture-dependent studies.

In addition to erdacin, a number of other clone specific novel and known metabolites were characterized from cultures of *S. albus* harboring the V167 clone. These metabolites included utahmycin A, utahmycin B and 8-O-methylutahmycin A. The structure of Utahmycin A is novel and utahmycin B has not been previously observed in a biological system. The utahmycins appear to arise from the incorporation of nitrogen into erdacin biosynthetic intermediates. Since there is no predicted amino transferase in the erdacin

gene cluster, the utahmycins are likely derived from erdacin biosynthetic intermediates independent of the erdacin gene cluster. Whether the utahmycins have a biological function in the uncultured microbe from which the erdacin gene cluster was derived or if they are unintended shunt products of erdacin biosynthesis is unknown. Regardless, the isolation of additional distinct novel metabolites from the same eDNA clone encoding the biosynthesis of erdacin, suggests that eDNA-derived type-II PKS gene clusters are a rich source of natural products.

Sequencing of KSβ amplicons has found the AB and UT eDNA libraries are rich in type II PKS gene clusters. The isolation of the novel metabolite erdacin provides tangible evidence that these eDNA-derived gene clusters may encode for the biosynthesis of small molecules that are distinct from compounds isolated by culture-dependant strategies. Isolation of the utahmycins suggests that heterologous expression of these PKS gene clusters may yield, in some cases, multiple metabolites from a single clone. There is now precedent that the recovery and screening of clones, containing eDNA-derived type II PKS gene clusters, for the ability to confer production of clone-specific metabolites to *Streptomyces* will be a rewarding avenue to access the structural diversity of natural products encoded by uncultured soil bacteria.

Biosynthesis of erdacin and the utahmycins

The novel structure and pentacyclic ring arrangement of erdacin raised questions as to its biosynthetic origins. To develop a better understanding of the origin of erdacin, a biosynthetic analysis of the V167 clone utilizing transposon mutagenesis and [1,2-^{13}C]acetate feeding experiments was conducted. The isolation of juglomycin F from wild type V167 cultures, provided an initial hint to the biosynthetic origin of erdacin. The structure of juglomycin F maps onto half of the erdacin structure; suggesting it may arise from the heterodimerization of two 13-carbon subunits. However, transposon knockouts of early PKS genes lead to the production of known octaketide shunt products, indicating that the minimal PKS of erdacin produces an octaketide backbone. A minimal PKS encoding an octaketide is further supported by the isolation of the octaketide product prechrysophanol from cultures of mutant and wild type V167. Additionally, the folding pattern leading to the novel ring system of erdacin was assessed in an [1,2-^{13}C]acetate feeding study to learn more about the origins of the two 13-carbon subunits that make up erdacin.

While the exact structure of late-stage polyketides used in erdacin biosysnthesis is unknown, a biosynthetic scheme where erdacin is derived from two 13-carbon subunits was proposed. In this biosynthetic scheme, the erdacin minimal PKS encodes for an octiketide and the two distinct 13-carbon monomers are derived from a common octaketide precursor. Two distinct

second cyclizations diverge the biosynthesis of the two 13-carbon intermediates. At some point, both of the 13-carbon intermediates have the same three carbons excised from their octaketide precursors. These two 13-carbon monomers would then heterodimerize to form erdacin.

The utahmycins appear to arise from the incorporation of nitrogen into erdacin biosynthetic intermediates. There is no predicted amino transferase in the erdacin gene cluster and the utahmycins are likely derived from biosynthetic intermediates independently of the erdacin gene cluster. Data from [1,2-^{13}C]acetate-label erdacin implies the folding pattern of the utahmycins resembles the octaketide *Streptomyces* (S′)-folding of chrysophanol (Bringmann, Irmer et al. 2009). This differs from other known 2-azaanthraquinones, which are derived from a fungal (F)-folding of a septaketide (Thomas 2001; Van Wagoner, Mantle et al. 2008). Utahmycins are derived from both an alternative folding pattern and a larger polyketide precursor than previously known fungal 2-azaanthraquinones.

Activity of eDNA-derived metabolites

Since the bacteria that eDNA is derived from remain uncultured, it can be difficult to determine the function of novel eDNA-derived metabolites in their native organism. No antibacterial activity was detected when erdacin and the utahmycins were assayed against a number of Gram-positive and Gram-negative microbes in an antibacterial agar disk diffusion assay. And

erdacin did not appear to have an effect when cultured with mammalian cell lines. Erdacin was, however, found to have twice the antioxidant activity as the known antioxidant ascorbic acid (vitamin C).

Cryptic eDNA-derived clones and future directions

In addition to the V167 clone, a number of other clones containing PKS genes were recovered from the UT and AB libraries. Screening of the recovered clones containing minimal PKS genes revealed a number of clones able to confer the production of clone-specific molecules to cultured *Streptomyces albus*. However, some of the clones recovered from the libraries produced a large amount of unstable clone specific metabolites and others were cryptic clones. The X26 clone is one such cosmid that produced an array of unstable metabolites. Dr. John Bauer's sequencing of X26 revealed a minimal PKS and truncated gene cluster. A cosmid that captured an incomplete PKS gene cluster from eDNA could explain the array of unstable clone specific metabolites observed in X26 cultures.

The ability to access the metabolites encoded by these truncated and cryptic pathways would greatly increase the number of novel compounds found from eDNA libraries. A number of strategies have been utilized in subsequent studies to access the structural diversity of these PKS gene clusters. In one approach, missing PKS biosynthetic enzymes not present in the truncated X26 clone were found by recovering a second overlapping eDNA clone representing the same PKS gene cluster. The reconstruction of these two overlapping clones into a single Bacterial

artificial chromosome (BAC) and the subsequent heterologous expression of the reconstructed PKS pathway resulted in the discovery of a second novel ring system from eDNA (Kim, Feng et al. 2010; Feng , Kallifidas et al. 2011; Kallifidas and Brady 2012).

While *S. albus* has been utilized with great success in this work and subsequent studies to heterologously express eDNA-derived PKS gene clusters, alternative *Streptomyces* heterologous hosts may more closely resemble some of the uncultured native organisms that these eDNA gene clusters are derived from. In fact, it has been reported that eDNA clones will present distinctive phenotypes in different hosts and that the same phenotype is rarely conferred to two different hosts (Craig, Chang et al. 2009; Craig, Chang et al. 2010). Expression of eDNA-derived PKS gene clusters in alternative *Streptomyces* heterologous hosts may lead to the production of metabolites from cryptic clones (Wang, Isaacs et al. 2009; Komatsu, Uchiyama et al. 2010). Another approach to accessing the metabolites encoded by cryptic eDNA clones is through the overexpression of transcription factors from eDNA-derived PKS gene clusters. A recent study has reported the successful use of a constitutively expressed pathway-specific *Streptomyces* antibiotic regulatory protein to promote the production of tetarimycin A from a previously cryptic eDNA clone (Kallifidas, Kang et al. 2012).

An alternative strategy to increase the number of natural products discovered from eDNA libraries is to build a library enriched with PKS gene clusters. Recently, the complementation of phosphopantetheine transferase (PPTase) deletion mutants with the PPTase genes found in some eDNA clones was successfully leveraged as a filter to construct an NRPS/PKS enriched eDNA library (Charlop-Powers, Banik et al. 2013). These advances and additional new approaches will continue to enhance our ability to access the incredible structural diversity of natural products that are encoded by uncultured bacteria.

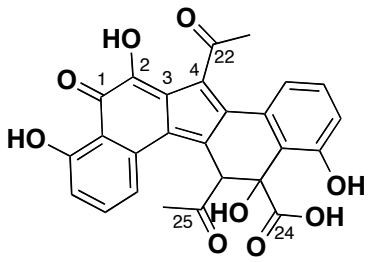
Sequencing advances and future directions

As sequencing technology continues to develop in both speed and cost effectiveness it will have a profound impact on the field of metagenomics (Chan, Hsu, et al. 2008; Morozova and Marra 2008; Yang, Peng et al. 2010). In fact, there are already efforts to sequence the first complete soil metagenome (Vogel, Simonet, et al. 2009). The data presented here and in other studies indicates that the genetic diversity encoded by uncultured bacteria is an affluent source of natural products that are distinct from those identified by traditional culture-dependant studies. Metagenomic methods will likely be a rewarding avenue for future research efforts.

APPENDIX

Table 8: Chemical shift table of Erdacin in MeOD with 0.1% TFA.

Atom #	Chemical δ	Proton δ (*J* Hz)	^{1}H-^{13}C HMBC
1	188.4		
2	153.7		
3	127.7		
4	131.2		
5	143.3		
6	133.4		
7	120.6	7.16 (d, 7.1, 1H)	5, 6, 8, 9, 11
8	131.3	7.28 (t, 7.7, 1H)	6, 7, 10, 11
9	118.7	6.92 (d, 7.7, 1H)	6, 7, 8, 10, 11, 12
10	156.7		
11	123.8		
12	75.5		
13	64.6	4.29 (s, 1H)	5, 11, 12, 14, 15, 24, 25, 26
14	134.6		
15	130.6		
16	135.6		
17	119.1	7.40 (m, 1H)	15, 16, 19, 20, 21
18	138.8	7.40 (m, 1H)	
19	118.4	6.73 (d, 7.7, 1H)	17, 20, 21
20	165.3		
21	114.1		
22	205.5		
23	30.3	2.53 (s, 3H)	4, 22
24	176.6		
25	203.5		
26	27.9	1.72 (s, 3H)	13, 25



Erdacin

Table 9: Genes of the Erdacin pathway and their deduced role.

Gene	Size	Deduced Role	Homologues				
	(aa)		Gene	Deduced role	% SM/ID	Origin	accession #
ORF 4	251		SSAG_02590	Acetyltransferase	77/72	Streptomyces sp.	EDX22799
Erd 1	430	Keto-acyl synthase α (KSα)	ORF1	Keto-acyl synthase α (KSα)	94/90	S. antibioticus	CAC05671
Erd 2	407	Keto-acyl synthase β (KSβ)	ORF2	Keto-acyl synthase β (KSβ)	83/72	S. antibioticus	CAC05672
Erd 3	89	Acyl carrier protein (ACP)	ORF3	Acyl carrier protein (ACP)	87/77	S. antibioticus	CAC05673
Erd 4	316	Aromatase	ORF4	putative aromatase	82/71	S. antibioticus	CAC05674
Erd 5	261	Keto reductase	ORF5	putative keto reductase	92/86	S. antibioticus	CAC05675
Erd 6	153	ActVI-A like	gra-orf31	unknown/aromatase	78/66	S. violaceoruber	CAA09658
Erd 7	317	Acyl-CoA dehydrogenase	YP_002192995	acyl-CoA dehydrogenase	39/28	S. clavuligerus	YP_002192995
Erd 8	266	4'-phosphopantetheinyl transferase	pSLA2-L_p056	4'-phosphopantetheinyl transferase	59/50	S. rochei	NP_851478
Erd 9	272	Pathway-specific transcriptional activator	gra-orf9	AlnR3 SARP family regulator	71/59	S. violaceoruber	CAA09630
Erd 10	192	HP	ABB04148	phenylalanine ammonia lyase	39/31	Rhodotorula glutinis	ABB04148
Erd 11	395	Two component sensory kinase	gra-orf11	two-component sensory kinase	54/40	S. violaceoruber	CAA09632
Erd 12	191	Two-component response regulator	gra-orf10	two-component response regulator	81/65	S. violaceoruber	CAA09631
Erd 13	413	Major facilitator transporter	RHA1_ro10147	major facilitator transporter	64/50	Rhodococcus sp.	YP_708498
Erd 14	344	Putative aldo/keto reductase	AZC_4378	Putative aldo/keto reductase	66/52	Azorhizobium caulinodans	YP_001527294
Erd 15	226	Membrane protein	RPC_3500	HP	69/56	Rhodopseudomonas palustris	YP_533359
Erd 16	110	HP	ZP_01548508	putative glyoxalase family protein	44/28	Labrenzia aggregata	ZP_01548508
Erd 17	157	HP	RTM1035_00205	HP	55/42	Roseovarius sp.	ZP_01881455
Erd 18	104	Transcriptional regulator	RSal33209_1249	MarR family transcriptional regulator	75/58	Renibacterium salmoninarum	YP_001624400
Erd 19	206	Flavodoxin	SGR_4078	putative flavodoxin	79/66	S. griseus	YP_001825590
Erd 20	406	FAD-dependent monooxygenase	remO	FAD-dependent monooxygenase	74/62	S. resistomycificus	CAE51184
Erd 21	125	HP/keto reductase Assesory protein	Aln5	HP/keto reductase Assesory protein	64/48	Streptomyces sp. CM020	ACI88872
Erd 22	302	Cyclase	gra-orf33	cyclase	74/62	S. violaceoruber	CAA09660
ORF 27	177		SCO0222	acetyltransferase	88/80	S. coelicolor	NP_624554
ORF 28	287			N5,N10-methylenetetra-hydromethanopterin reductase	87/79	S. tsusimaensis	ABA59536

Table 10: NMR Specroscopic Data (CD_2Cl_2) for Utahmycin A

position	δ_C, mult.	δ_H (*J* in Hz)	HMBC
1	162.1, qC		
2			
3	165.2, qC		
4	117.8, CH	7.85, s	3, 10, 13, 16
5	119.5, CH	7.77, dd (7.5, 1.1)	7, 10, 12
6	136.9, CH	7.69, t (7.9)	7, 8, 11
7	125.8, CH	7.35, dd (8.4, 1.1)	5, 8, 12
8	163.1, qC		
9	190.3, qC		
10	183.1, qC		
11	133.1, qC		
12	117.2, qC		
13	122.8, qC		
14	117.9, qC		
15	26.8, CH_3	3.05, s	1, 13
16	25.2, CH_3	2.70, s	3, 4
OH		12.78, s	7, 8, 12

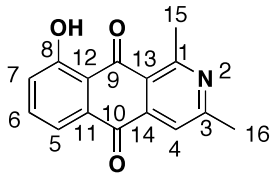


Table 11: NMR Specroscopic Data (acetone-d6) for Utahmycin B

position	δ_C, mult.	δ_H (*J* in Hz)	HMBC
NH (1)		11.77, br, s	
2	140.9, qC		
3	108.2, CH	6.49, s	2, 9, 12, 13
4	119.7, CH	7.62, m	6, 7, 9, 10, 11
5	136.5, CH	7.62, m	6, 7, 9, 10, 11
6	124.5, CH	7.18, dd (6.6, 2.8)	4, 5, 7, 11
7	163.0, qC		
8	180.8, qC		
9	180.5, qC		
10	135.5, qC		
11	116.5, qC		
12	131.9, qC		
13	130.0, qC		
14	13.1, CH_3	2.44, s	2, 3
OH		12.45, s	6, 7, 11

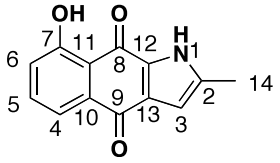


Figure 23: ^{1}HNMR of erdacin in MeOD-d4 with 0.1% TFA

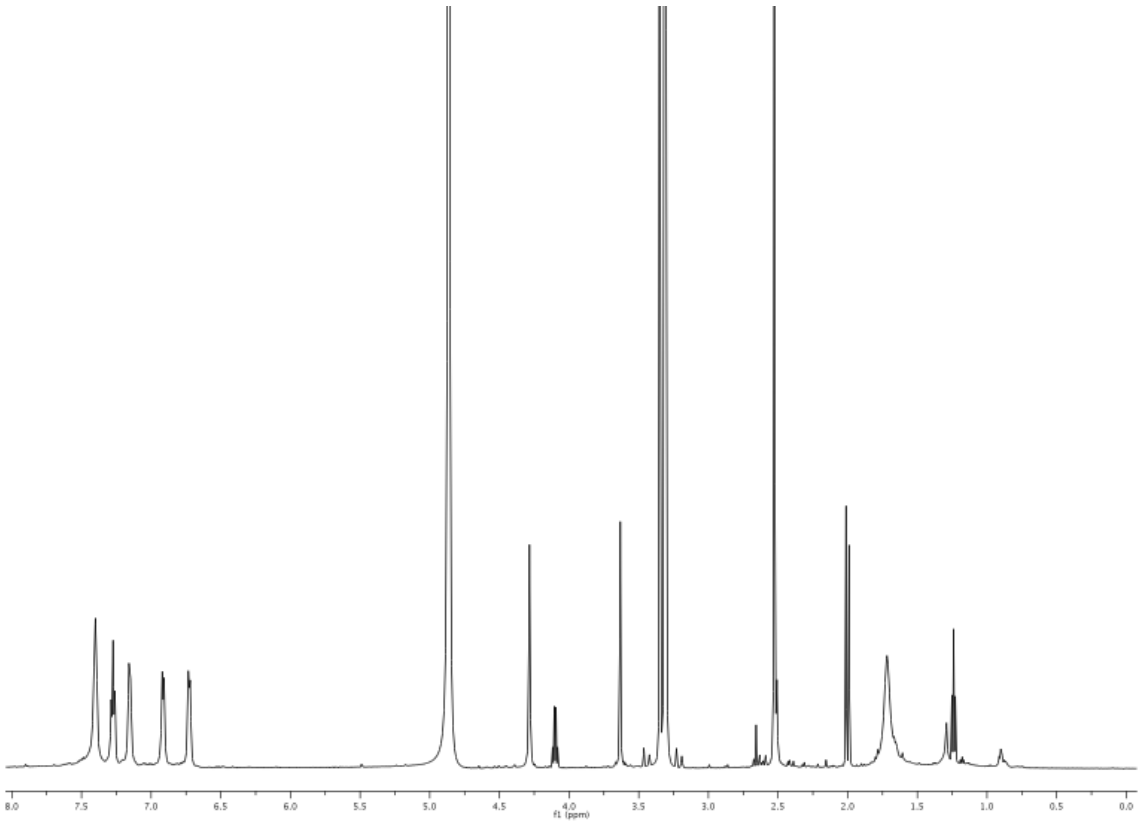


Figure 24: ^{13}CNMR of ^{13}C uniformly labeled erdacin in MeOD-d4 with 0.1% TFA

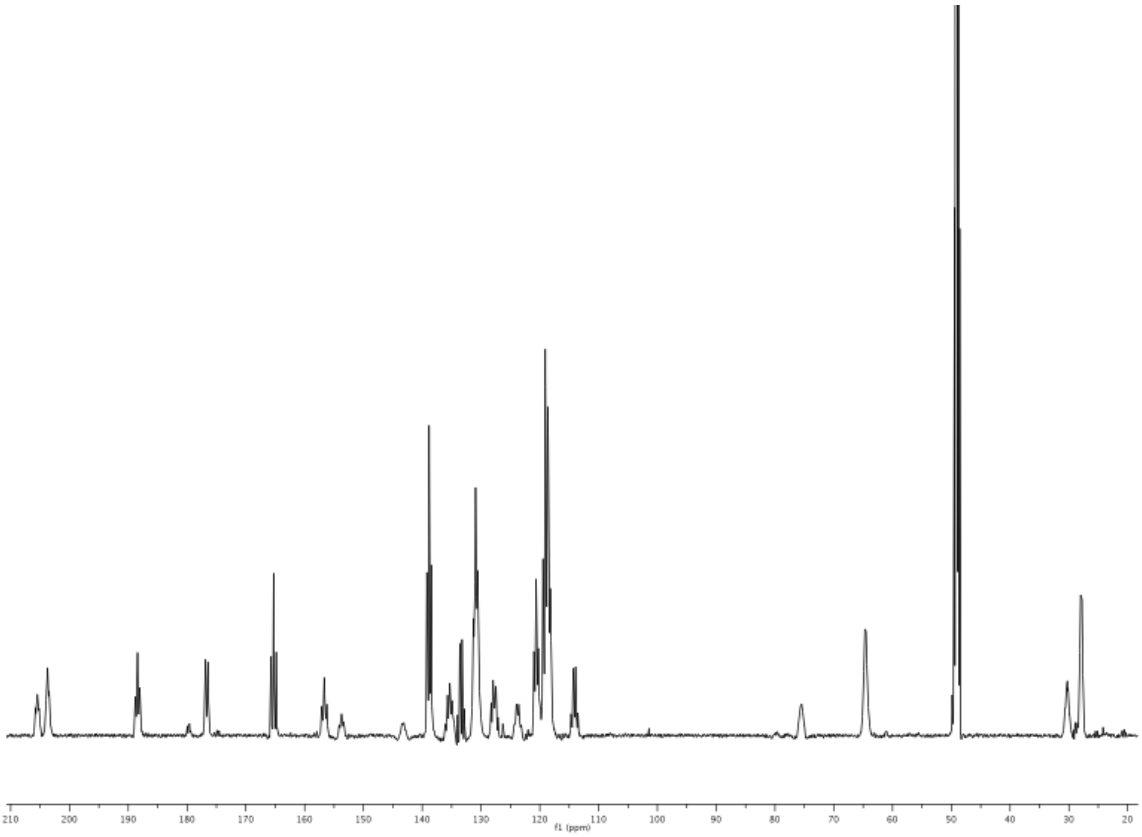


Figure 25: ^{1}HNMR of SEK34b in DMSO-d6

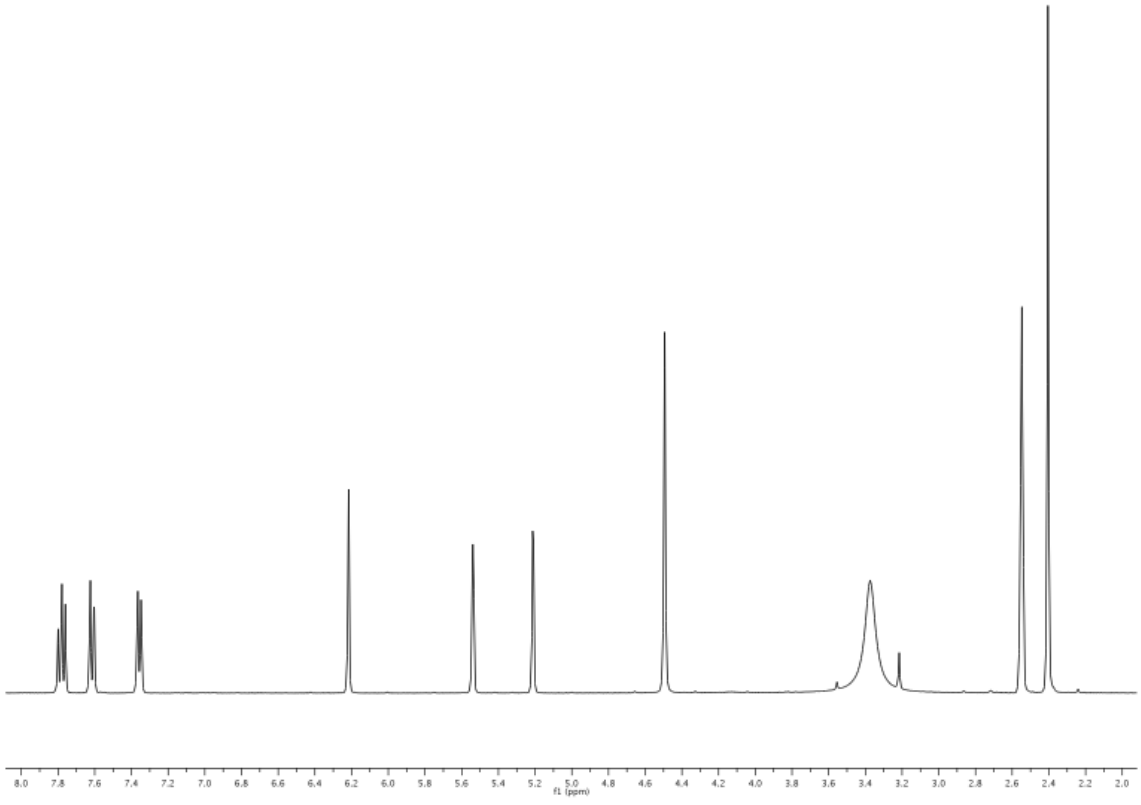


Figure 26: ^{1}HNMR of Juglomycin F in MeOD-d4

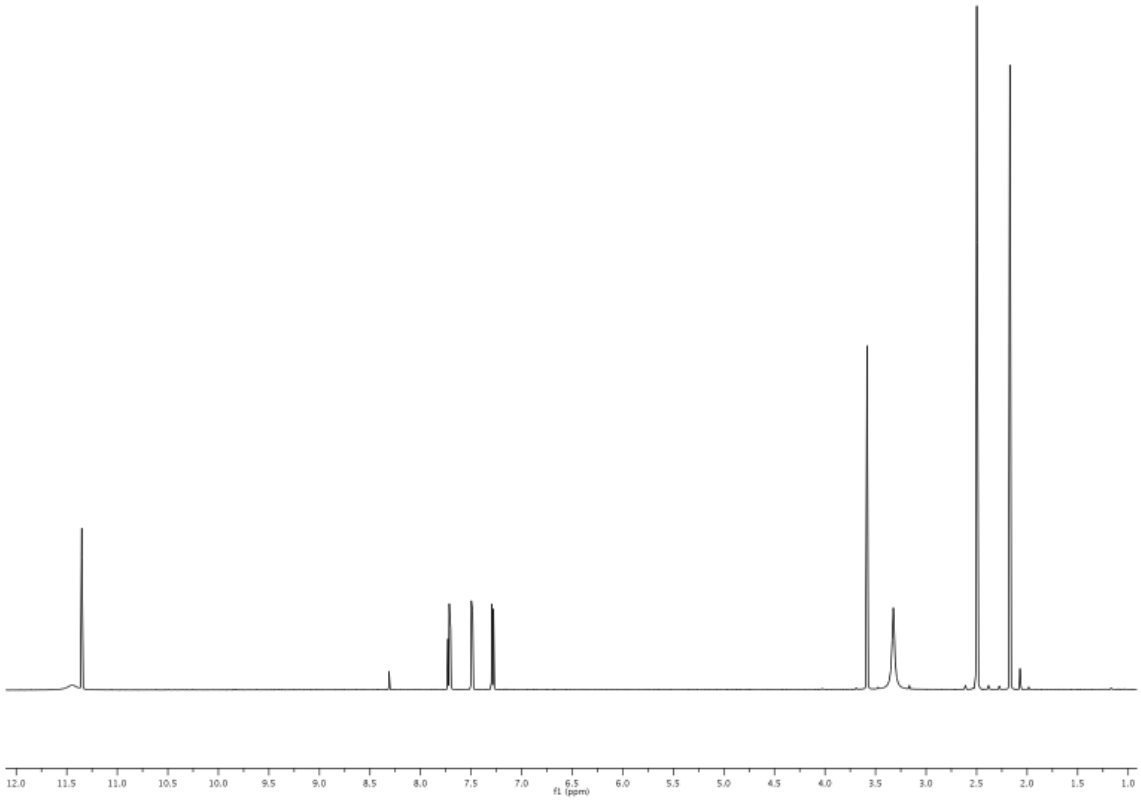


Figure 27: ^{1}HNMR of prechrysophanol in Acetone-d6

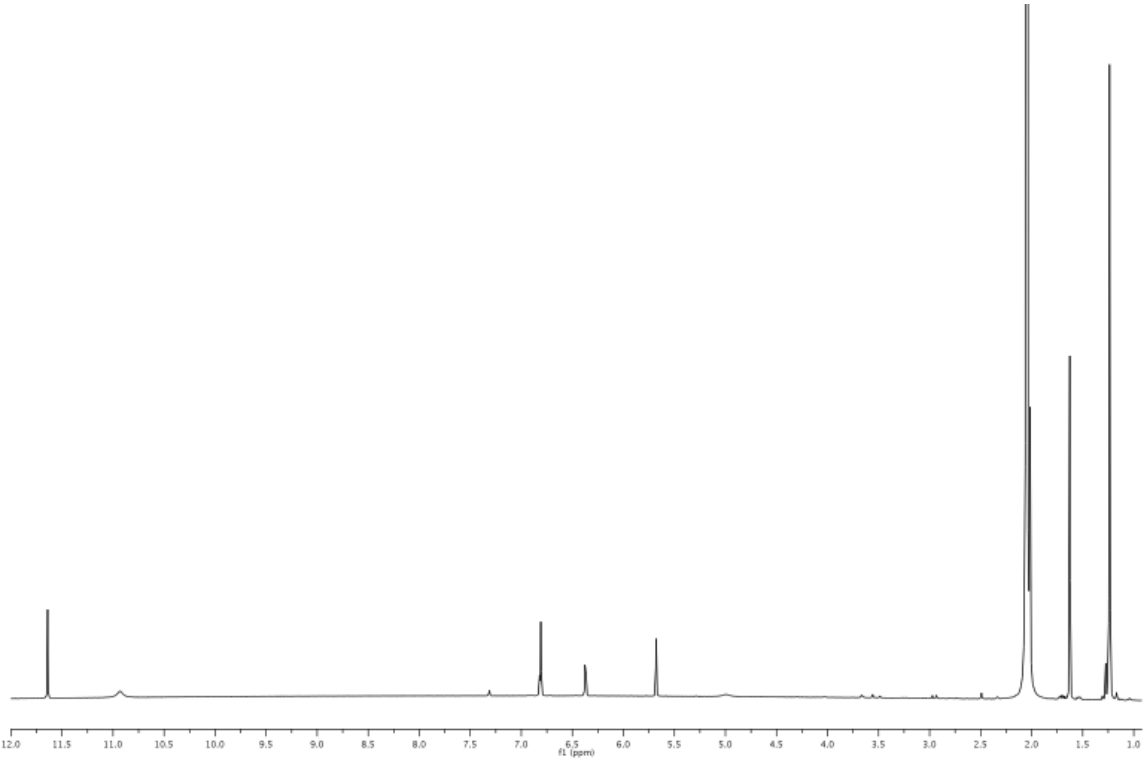


Figure 28: ^{1}HNMR of SEK4b in DMSO-d6

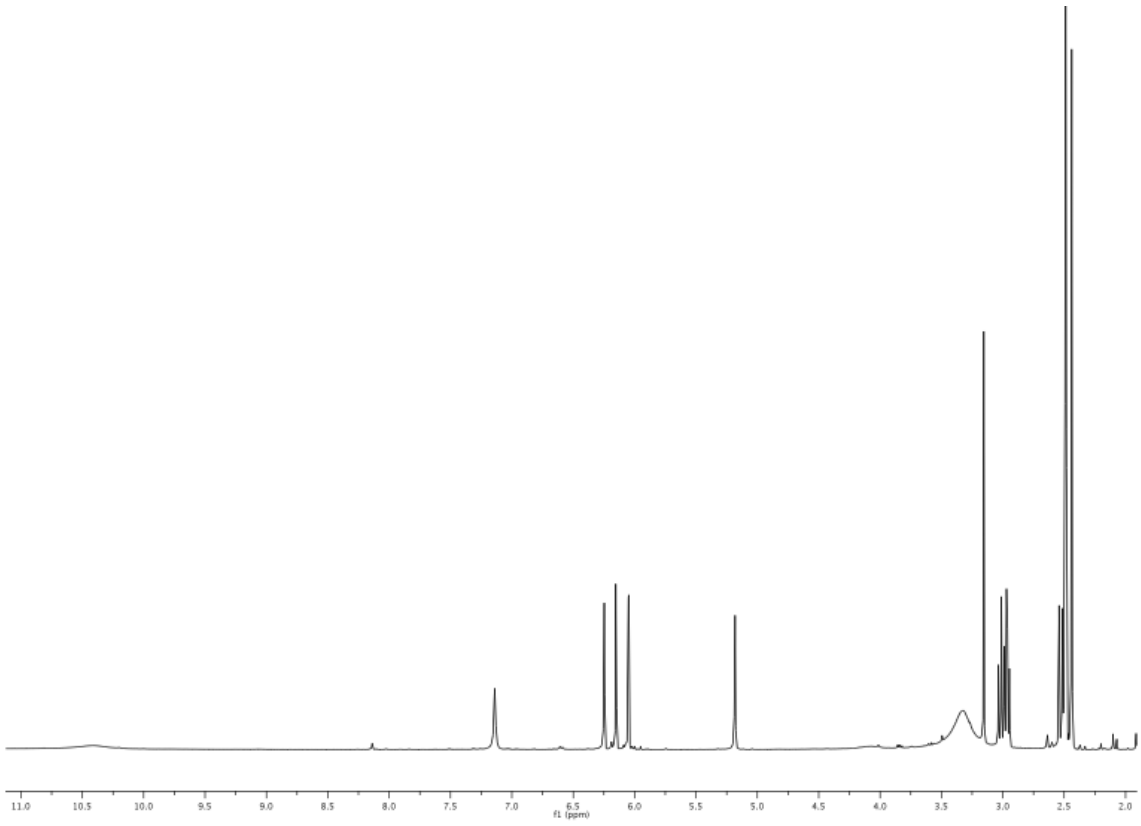


Figure 29: ^{1}HNMR of AUR367 in DMSO-d6

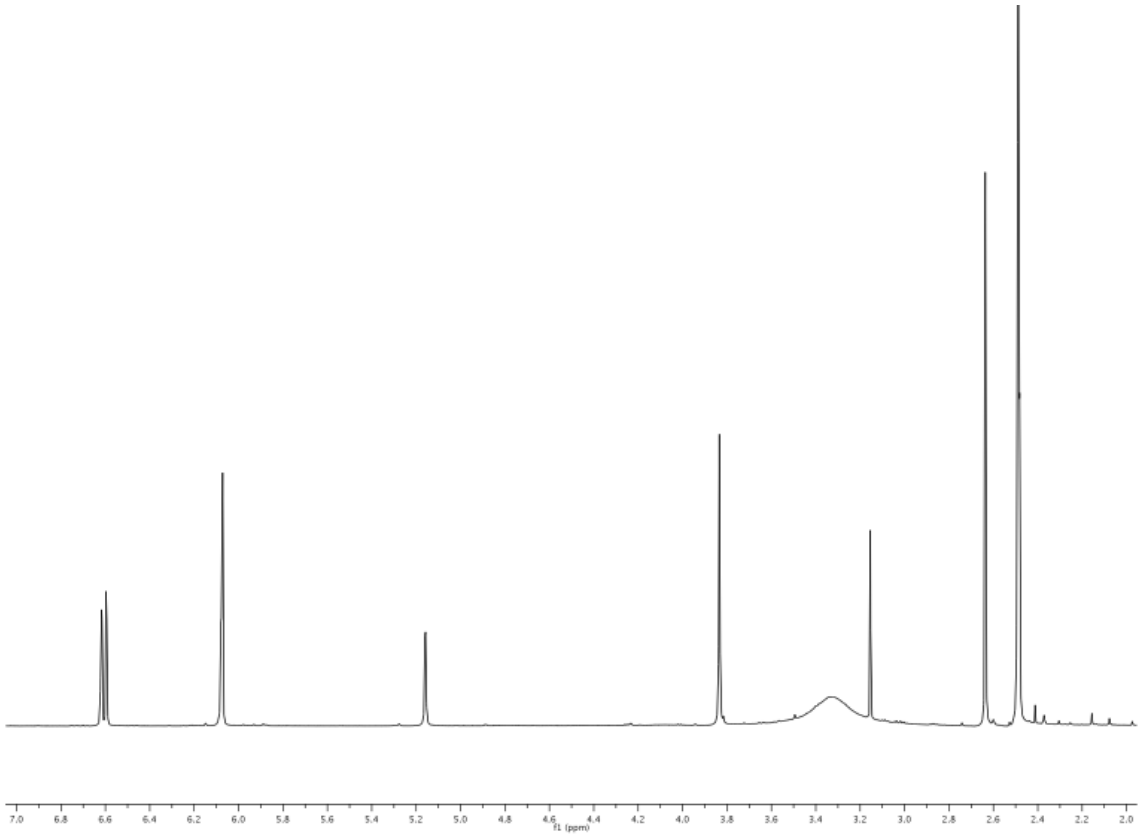


Figure 30: ^{1}HNMR of Mutactin in Acetone-d6

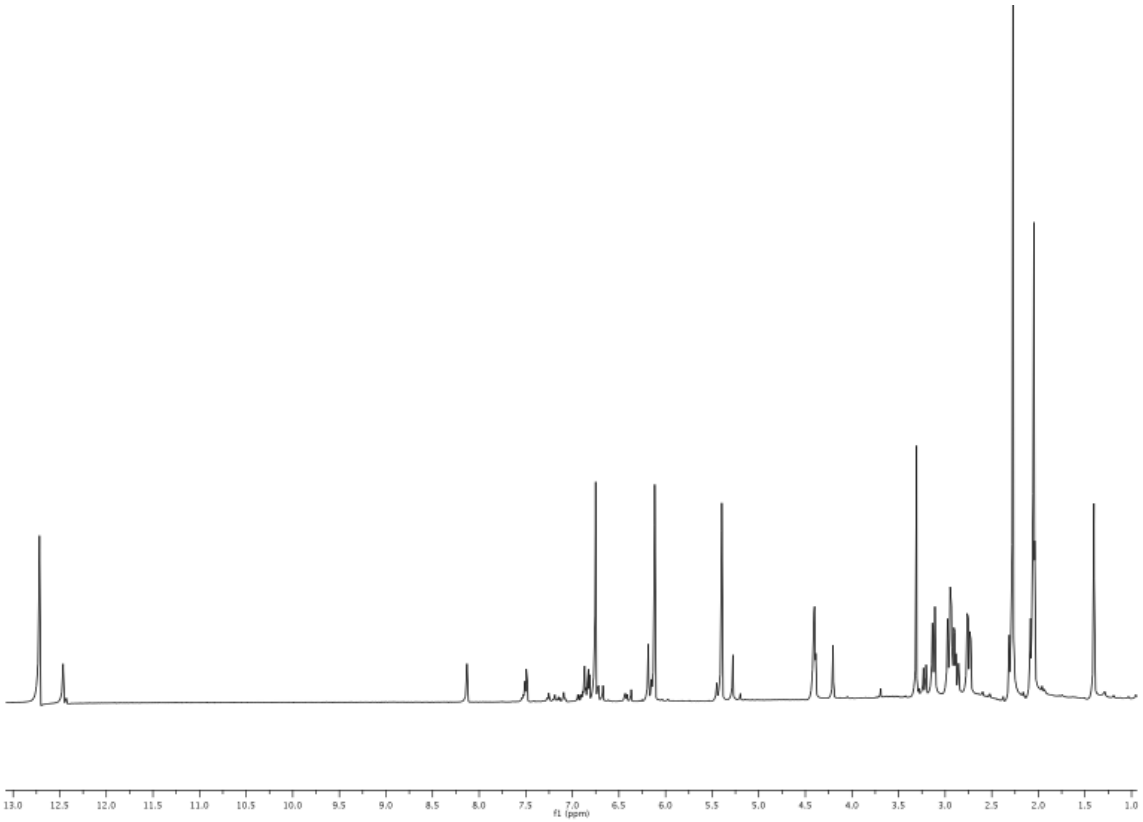


Figure 31: ^{1}HNMR of Dehydromutactin in DMSO-d6

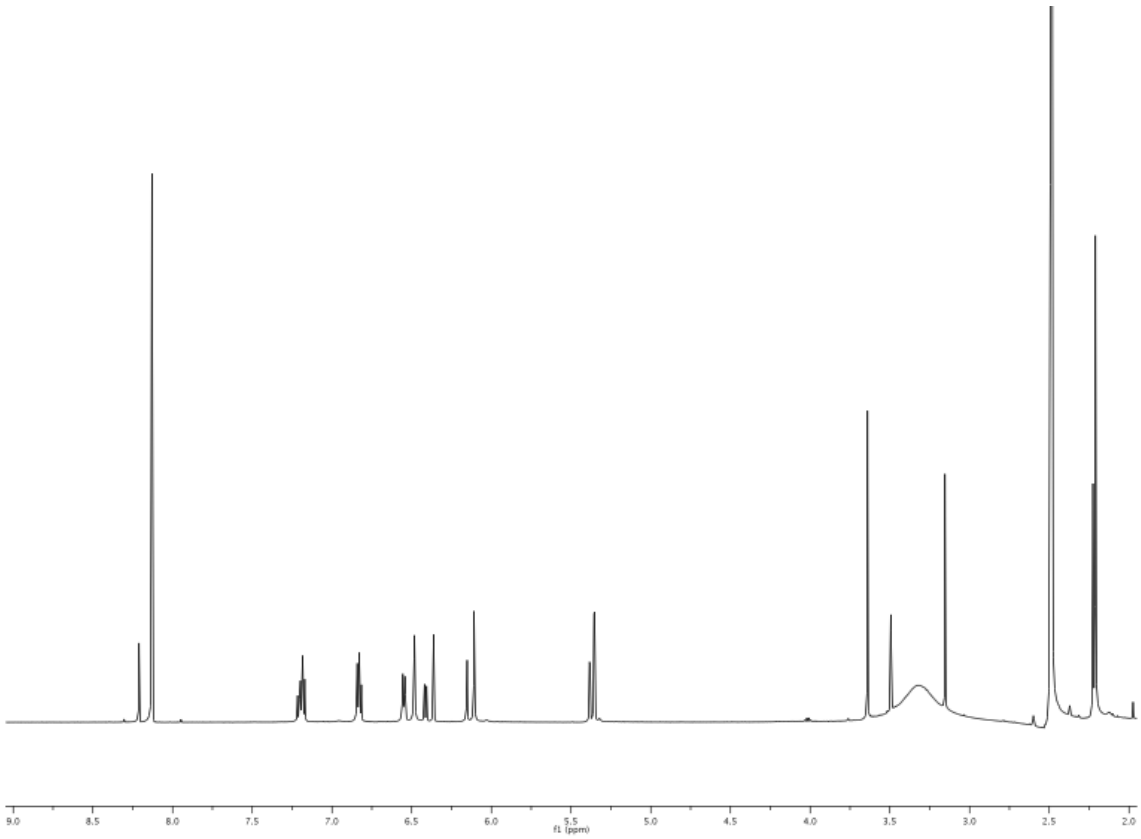


Figure 32: ^{1}HNMR of SEK34 in DMSO-d6

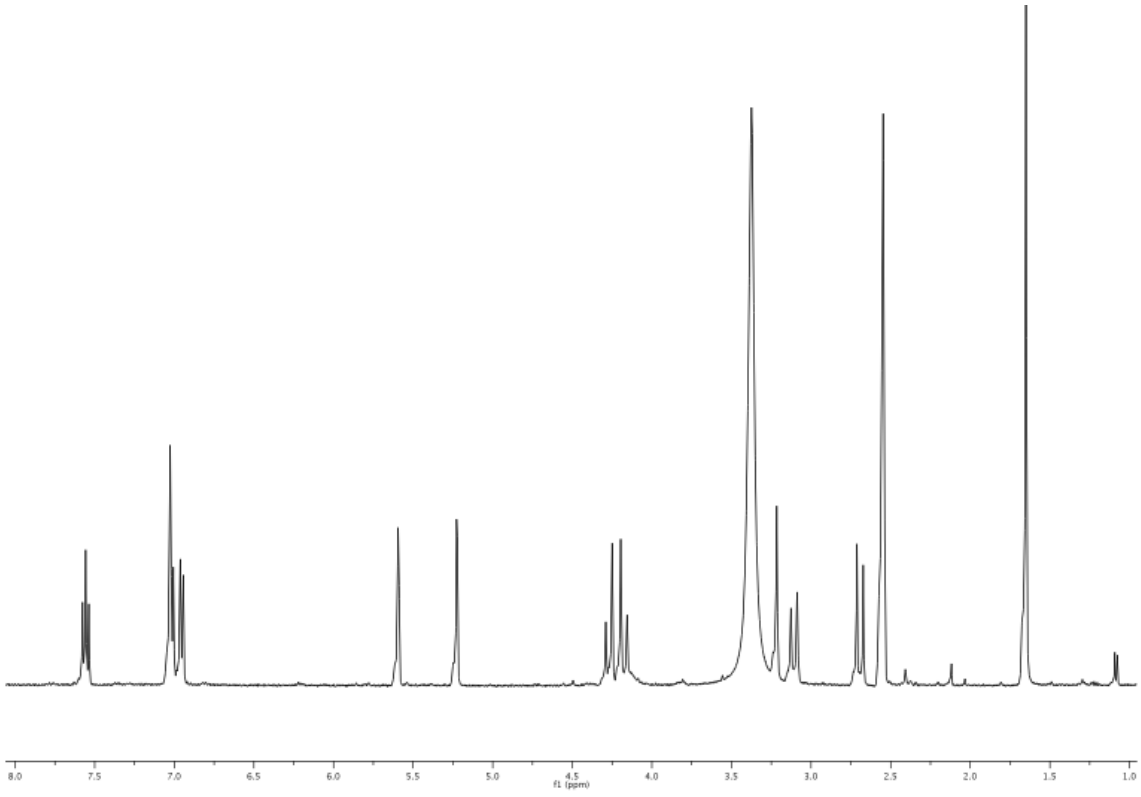


Figure 33: ^{1}HNMR of utahmycin A in CD_2Cl_2

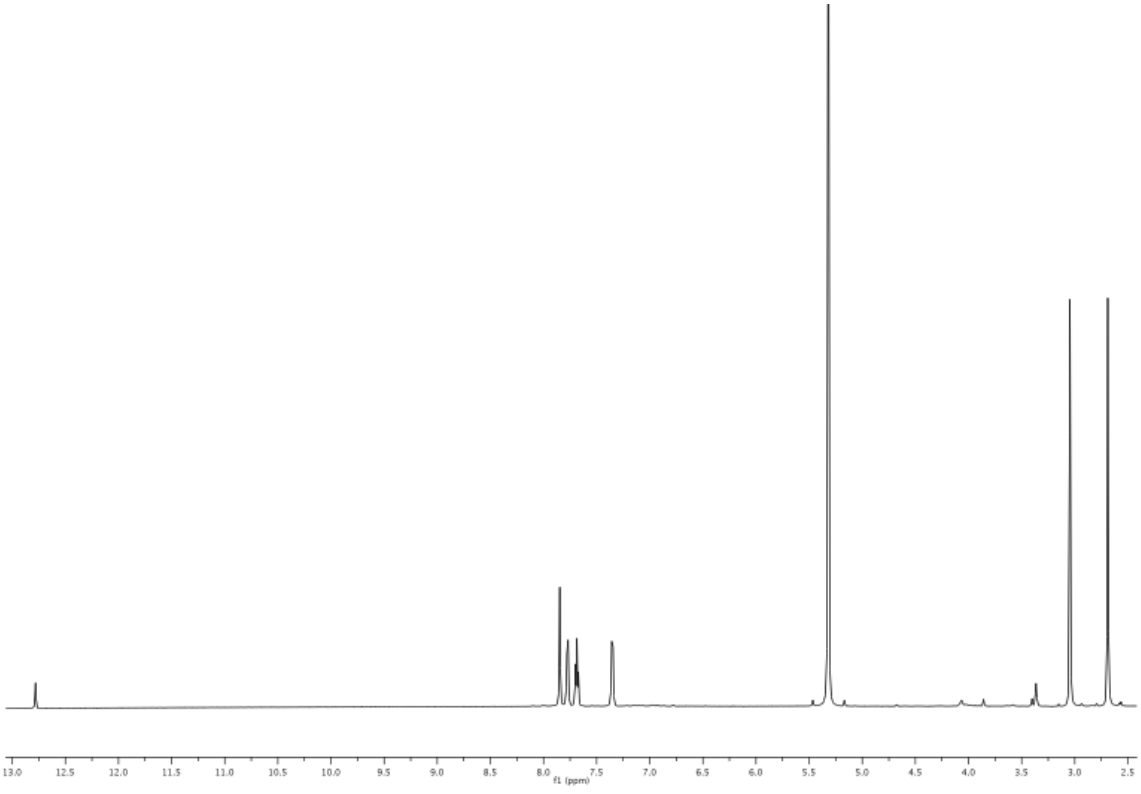


Figure 34: ^{13}CNMR of utahmycin A in CD_2Cl_2

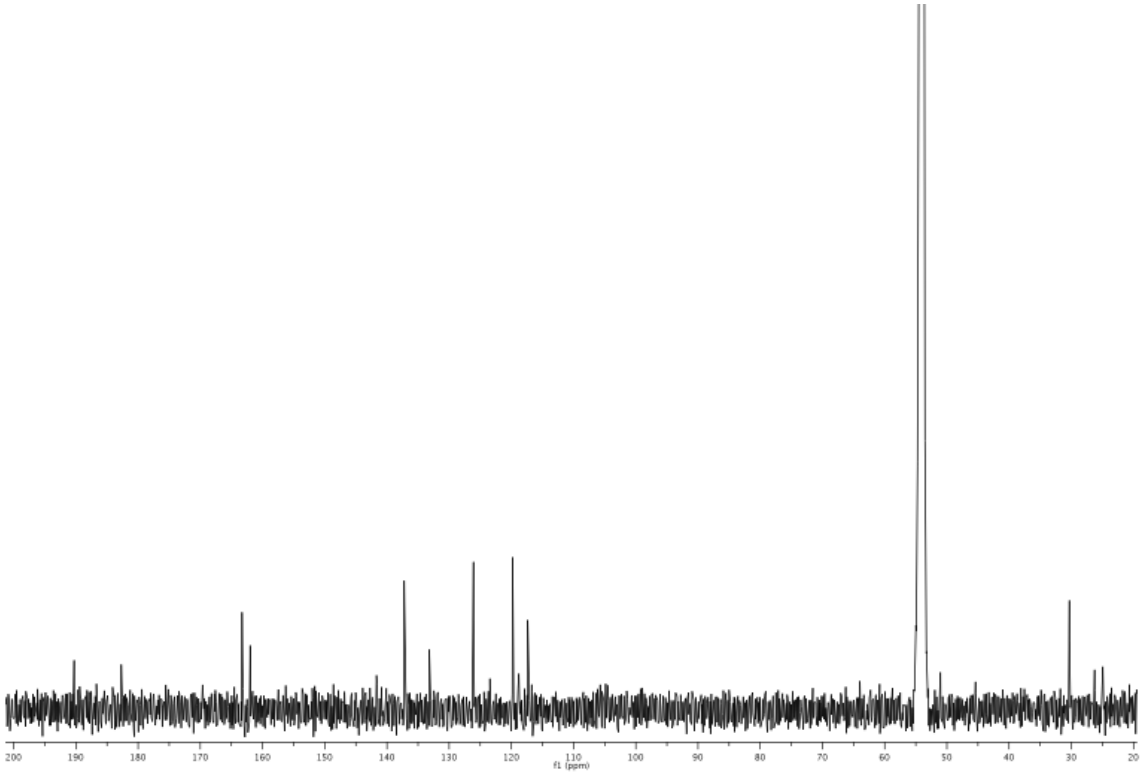


Figure 35: ^{1}H-^{13}C HMBC of utahmycin A in CD_2Cl_2

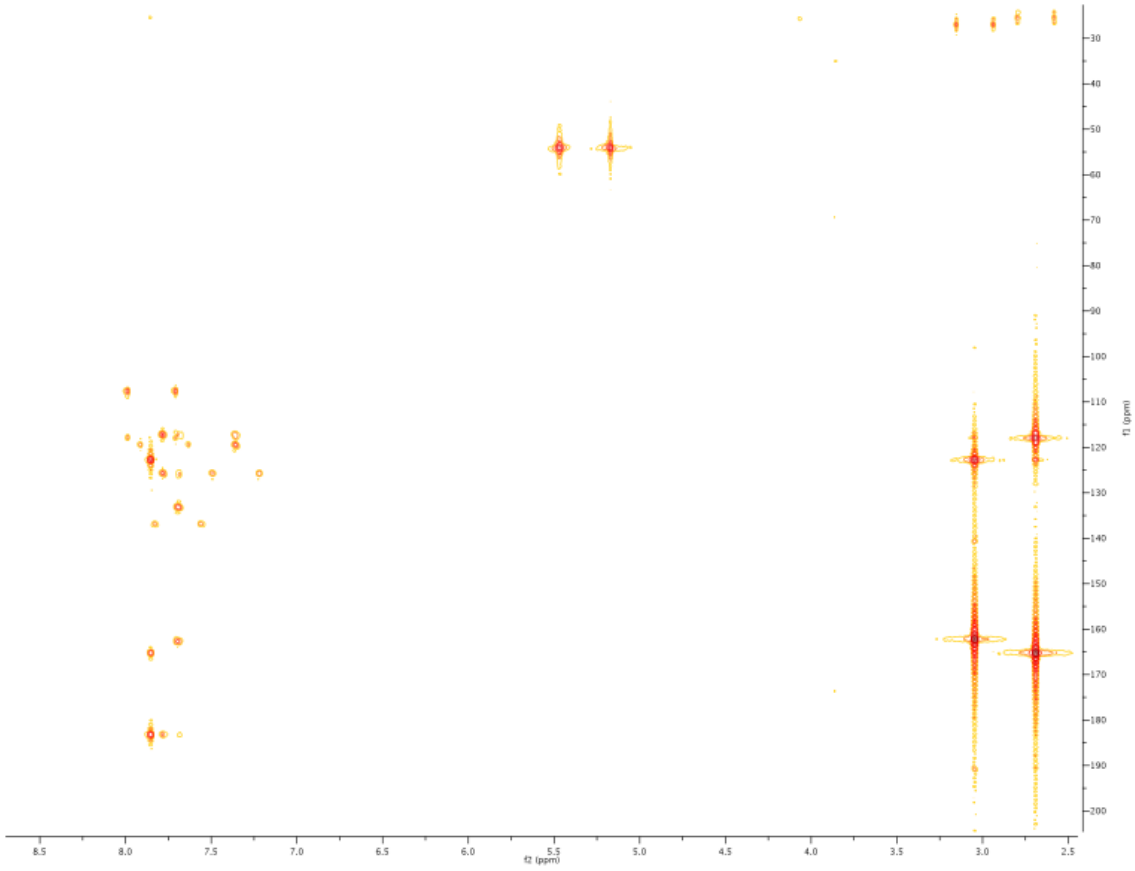


Figure 36: ^{1}H-^{13}C HMQC of utahmycin A in CD_2Cl_2

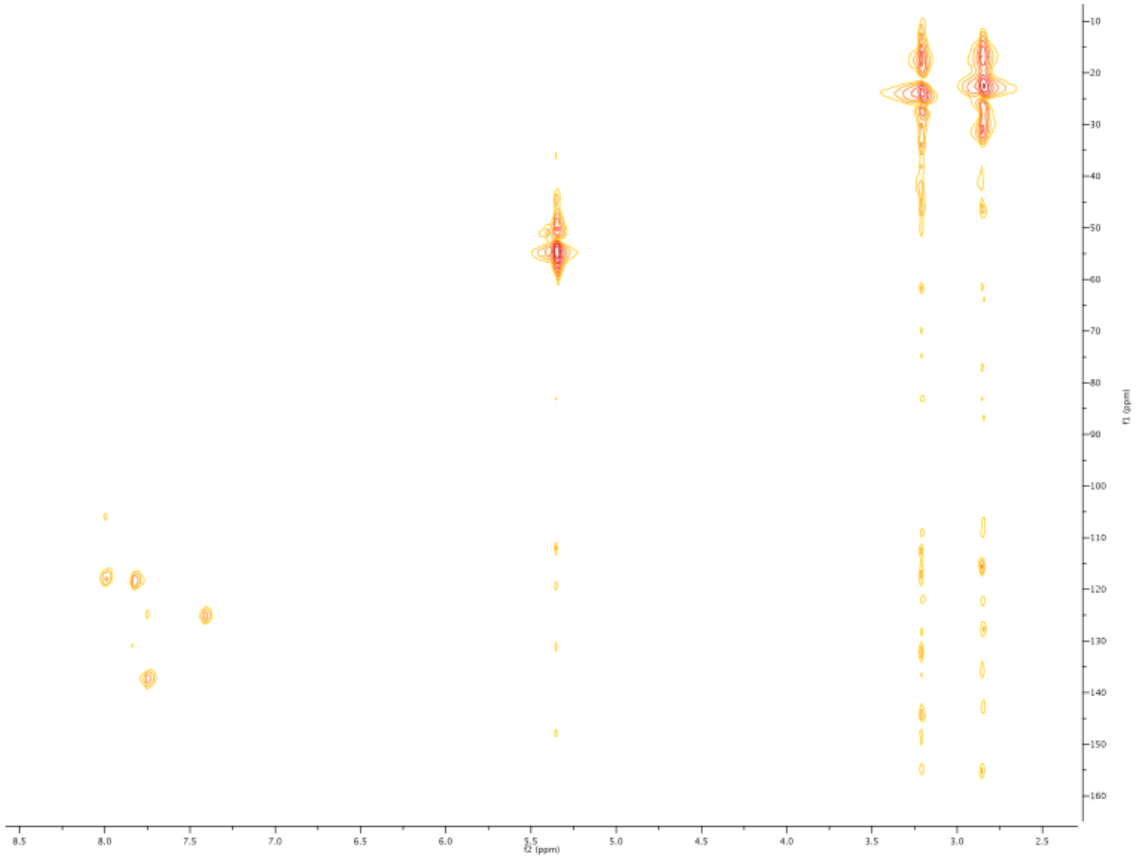


Figure 37: ^{1}HNMR of utahmycin B in acetone-d6

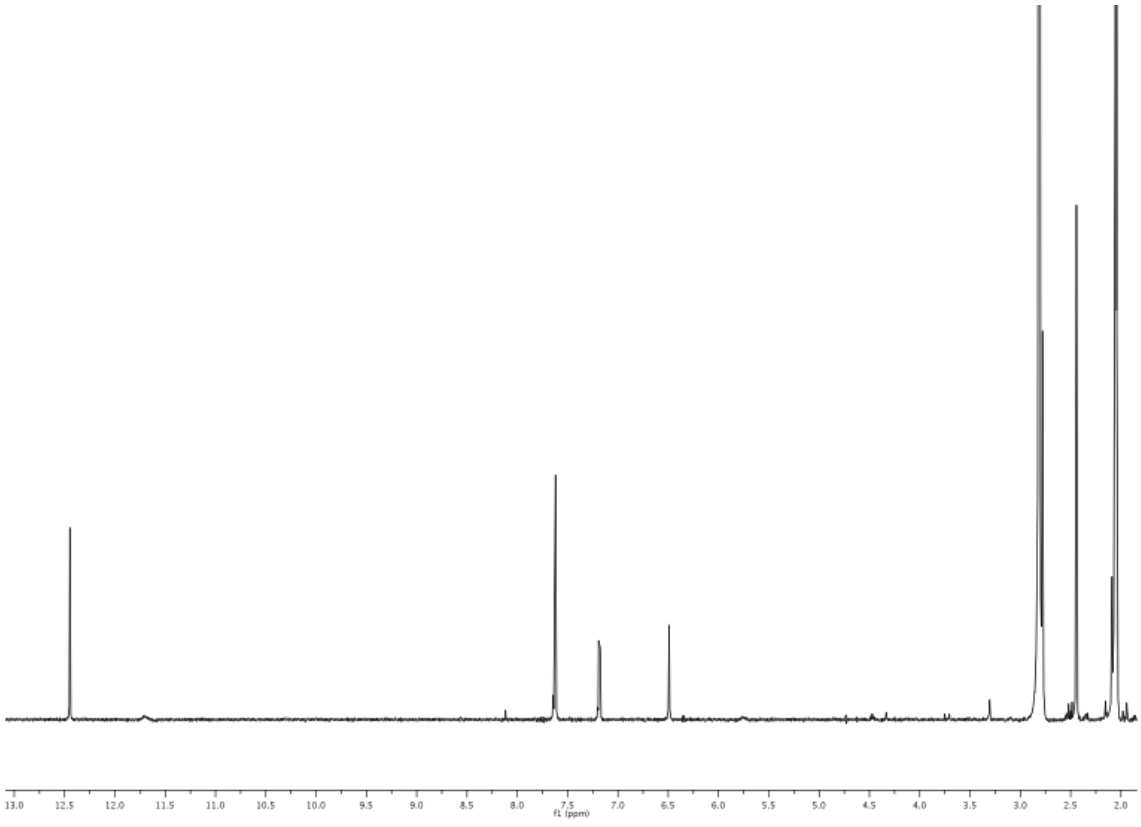


Figure 38: ^{13}CNMR of Utahmycin B in acetone-d6

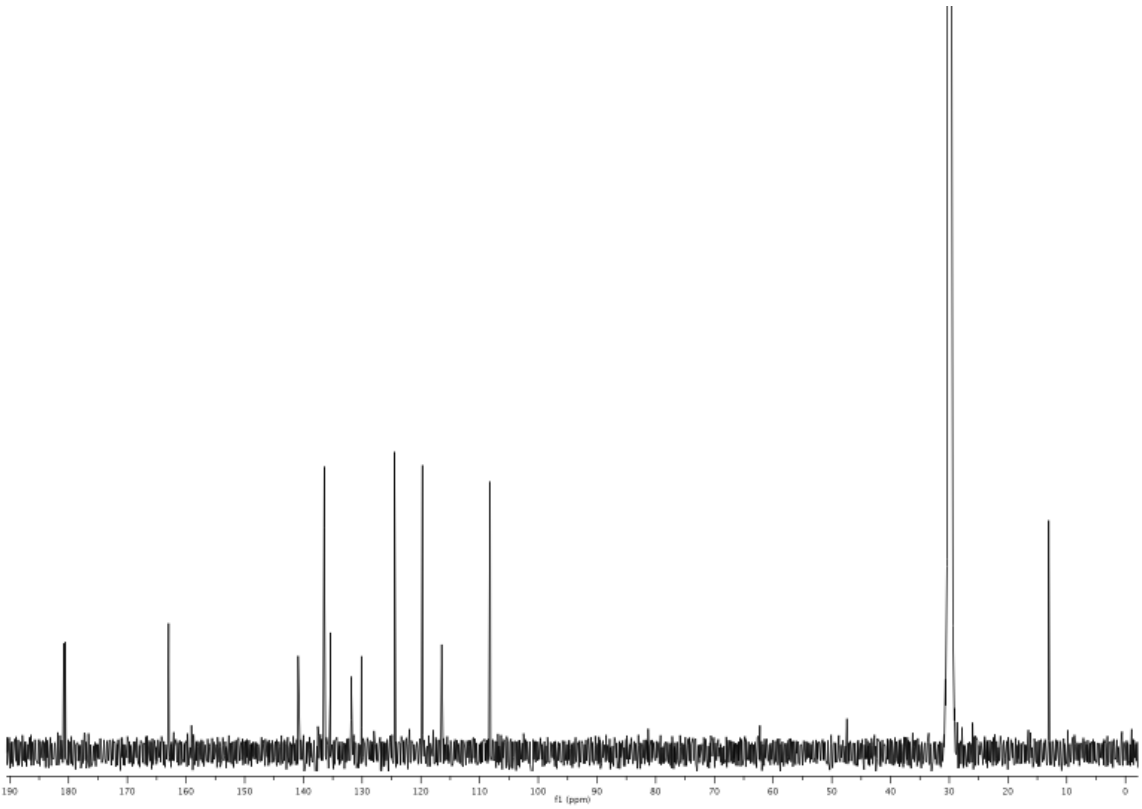


Figure 39: ^{1}H-^{13}C HMBC of utahmycin B in acetone-d6

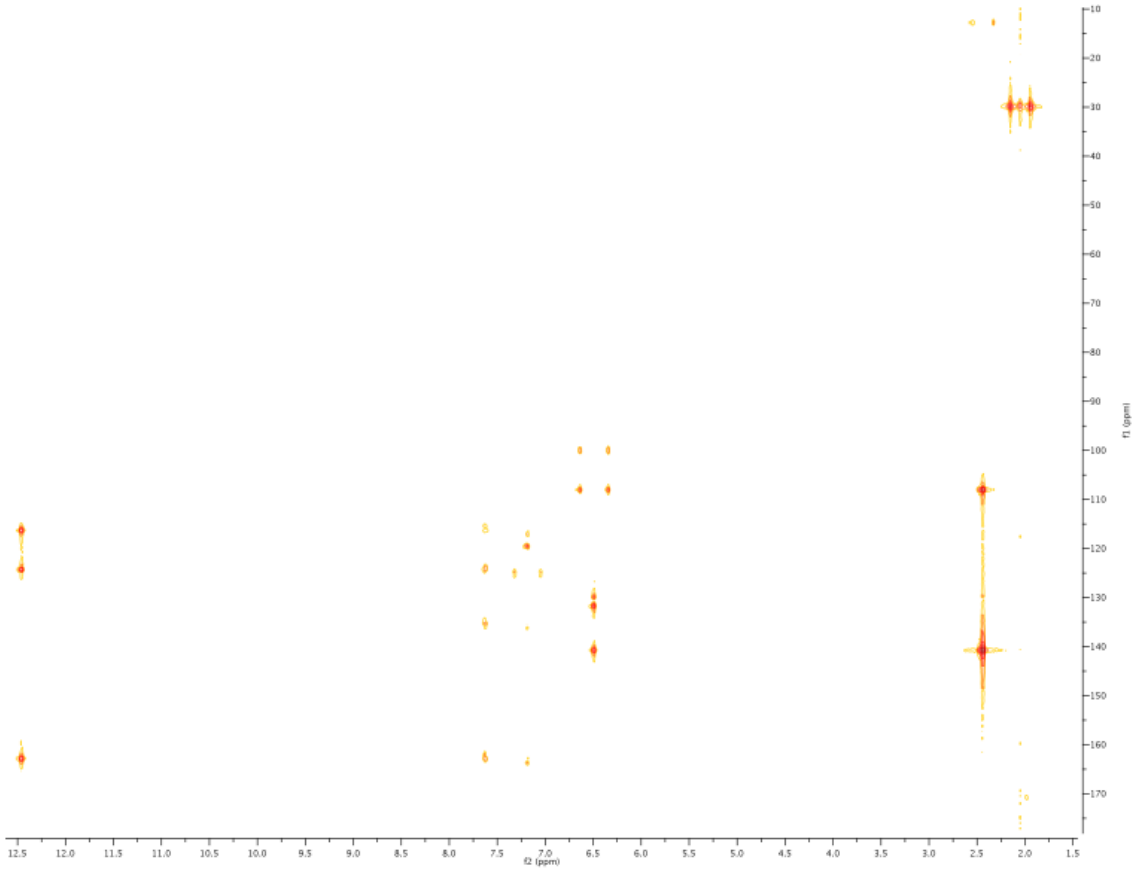
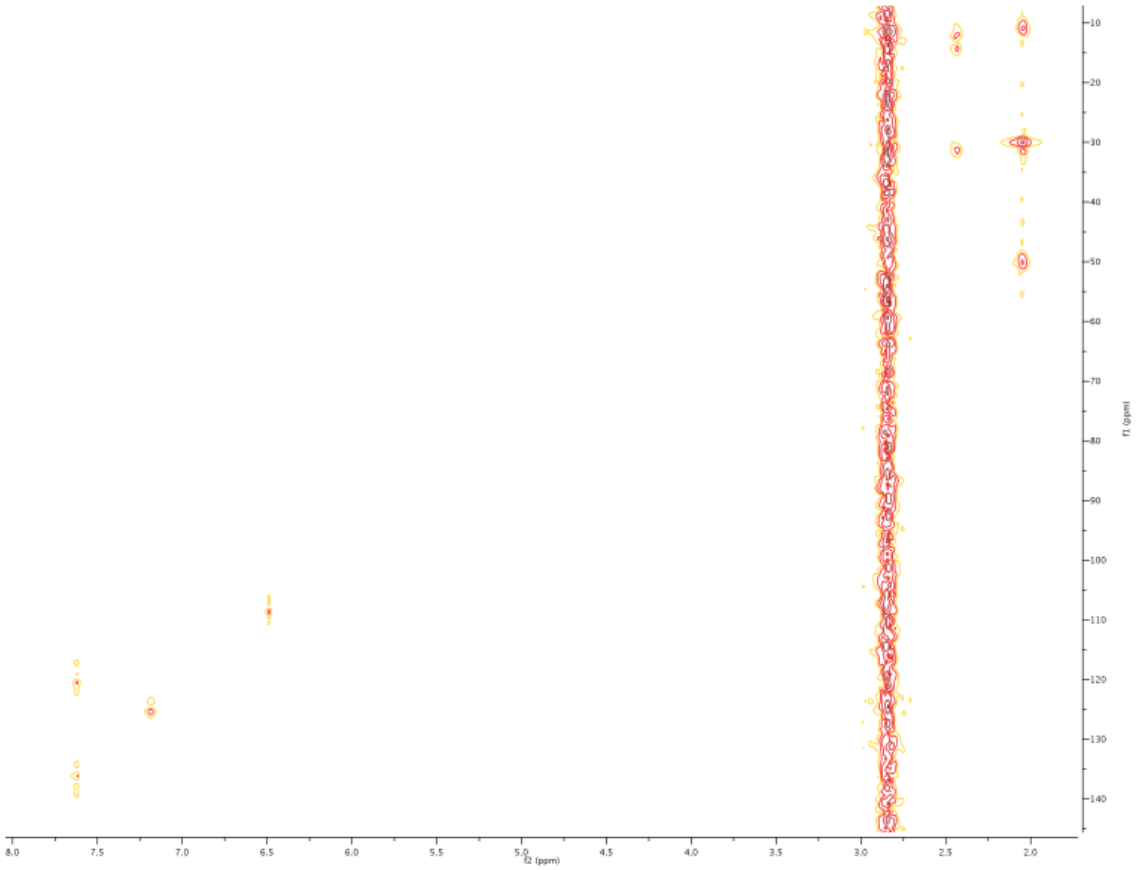


Figure 40: ^{1}H-^{13}C HMQC of utahmycin B in acetone-d6



REFERENCES

Apak,R., Kubilay, G., et al. (2004). Novel Total Antioxidant Capacity Index for Dietary Polyphenols and Vitamins C and E, Using Their Cupric Ion Reducing Capability in the Presence of Neocuproine: CUPRAC Method. *J. Agric.Food Chem.* **52**(26): 7970-7981.

Arsenault,G.P. (1965). *Tetrahedron Lett.* **45**:4033-4037.

Baker,R.A., Tatum, J.H., (1990). Standardized methods for the determination of antioxidant capacity and phenolics in foods and dietary supplements. *Mycopathologia.* **111**(1): 9-15.

Banik, J.J., Brady, S.F., (2008). Cloning and characterization of new glycopeptides gene clusters found in an environmental DNA megalibrary. *Proc Natl Acad Sci USA* **105**(45): 17273-17277.

Banik, J.J., Brady, S.F., (2010) Recent application of metagenomic approaches towards the discovery of antimicrobials and other bioactive small molecules. *Curr Opin Microbiol.* **13** (5): 603-609.

Banik, J.J., Craig, J.W.,et al. (2010). Tailoring enzyme-rich environmental DNA clones: a source of enzymes for generating libraries of unnatural natural products. *J Am Chem Soc.* **132**(44): 15661-15670.

Bauer, J. D., R. W. King, et al. (2010). "Utahmycins a and B, azaquinones produced by an environmental DNA clone." J Nat Prod **73**(5): 976-979.

Brady, S. F. (2007). Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nature Protocols* **2**: 1297-1305.

Brady,S.F., Bauer,J.D., et al. (2007). Natural products from isnA-containing biosynthetic gene clusters recovered from the genomes of cultured and uncultured bacteria. *J Am Chem Soc.* **129**(40): 12102-3.

Brady, S. F.,Chao, C.J., et al. (2001). Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Org Lett* **3**(13): 1981-1984.

Brady, S. F. and Clardy, J., (2000). Long-Chain N-Acyl Amino Acid Antibiotics Isolated from Heterologously Expressed Environmental

DNA. *Journal of the American Chemical Society* **122**(51): 12903-12904.

Brady, S. F. and Clardy,J., (2005). Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA. *Angew Chem Int Ed Engl* **44**(43): 7063-7065.

Brady, S.F., Simmons, L., et al. (2009). Metagenomic approaches to natural products from free-living and symbiotic organisms. *Nat Prod Rep.* **26**(11): 1488-503.

Bringmann, G., Irmer, A., et al.(2009). Convergence in the biosynthesis of acetogenic natural products from plants, fungi, and bacteria. *Phytochemistry.* **70**(15-16): 1776-86.

Brunker, P., McKinney, K., et al. (1999). Isolation and characterization of the naphthocyclinone gene cluster from Streptomyces arenae DSM 40737 and heterologous expression of the polyketide synthase genes. *Gene* **227**(2): 125-135.

Chan, C. K., Hsu, A. L. et al. (2008). Binning sequences using very sparse labels within a metagenome. *BMC Bioinformatics* **9**: 215.

Chang,F., Brady, S.F. (2011) Cloning and characterization of an eDNA derived gene cluster that encodes the biosynthesis of the antitumor substance BE-54017. *J Am Chem Soc.* **133**(26): 9996-9999.

Charlop-Powers, Z. Banik,J.J. et al. (2013). *ACS Chem Biol.* **8**(1): 138-43.

Courtois,S., Cappellano,C.M.,et al. (2003). Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol.* **69**(1): 49-55.

Craig, J. W., Chang, F.Y.,et al. (2009). Natural products from environmental DNA hosted in Ralstonia metallidurans. *ACS Chem Biol* **4**(1): 23-28.

Craig, J. W.,Chang, F.Y.,et al. (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): 1633-1641.

Curtis,T. P., Sloan, W. T. Scannell, J. W. (2002). Estimating prokaryotic diversity and its limits. *Proc. Natl. Acad. Sci. U. S. A.* ***99***(16): 10494.

Feng, Z., Chakraborty, D., et al. (2012). Environmental DNA-encoded antibiotics fasamycins A and B inhibit FabF in type II fatty acid biosynthesis. *J Am Chem Soc.* **134**(6): 2981-2987.

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

Feng, Z., Kim J.H., Brady, S.F. (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.

Fernandes, R., Chavan, V. (2008). A highly enantioselective synthesis of (-)- and (+)- juglomycin A through Dotz annulation and asymmetric dihydroxylation. *Tetrahedron Lett.***49**: 3899-3901.

Fox, J. L. (2006). The business of developing antibacterials. *Nat Biotechnol* ***2*4**(12): 1521-1528.

Fujii, I. (2008). Heterologous expression systems for polyketide syntheses. *Nat.Prod.Rep.* **26**: 155-169.

Gabor, E. M., Alkema, W. B.,et al. (2004). Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* **6**(9): 879-886.

Gillespie, D. E., S. F. Brady, et al. (2002). Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol* **68**(9): 4301-4306.

Ginolhac, A., Jarrin, C., et al. (2004). Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. *Appl Environ Microbiol.* **70**(9): 5522-7.

Gish,W., States, D.J. (1993). Identification of protein coding regions by database similarity search. *Nature Genet.* **3**: 266-272.

Grafe, U., Ihn,W.,(1990). *Biol.Met.***3**:39-44.

Grafe,U., Romer, W., (1991). 2-Aza-anthraquinones as inhibitors of the Ca(2+)- and calmodulin-dependent cyclic adenosine 3',5'-monophosphate phosphodiesterase. *Pharmazie.* **46**(4): 297-8.

Hamilton,M.A., Knorr,M.S., (1953). *Antibiot. Chemother.***1953**(3): 853-856.

Handelsman,J., Rondon M.R., (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5**(10): R245-9.

Hertweck, C. Luzhetskyy, A., et al. (2007). Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod.Rep.* **24**(1): 162.

Hugenholtz, P., Goebel, B. M.,et al. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180**(18): 4765-4774.

Ichinose, K., Bedford, D.J., et al. (1998). The granaticin biosynthetic gene cluster of Streptomyces violaceoruber Tu22: sequence analysis and expression in a heterologous host. *Chem. Biol.*, **5**(11): 647-59.

Ichinose, K., Ozawa, M., et al. (2003). Cloning, sequencing and heterologous expression of the medermycin biosynthetic gene cluster of Streptomyces sp. AM-7161: towards comparative analysis of the benzoisochromanequinone gene clusters. *Microbiology* **149**(Pt 7): 1633-45.

Iqbal, H.A., Feng, Z., et al. (2012). Biocatalysts and small molecule products from metagenomic studies. *Curr Opin Chem Biol.* **16** (1-2): 109-16.

Jung,W.S., Lee, S.K. et. al. (2006). Heterologous expression of tylosin polyketide synthase and production of a hybrid bioactive macrolide in Streptomyces venezuelae. *Appl Microbiol. Biotechnol.* **72**(4): 763-769.

Kallifidas,D., Brady,S.F., (2012). *Methods Enzymol* **517**: 225-39.

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

Khanapure,S.P.,Biehl, E.R. (1988). *Heterocycles.* **27**:2643-2650.

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000). Practical **Streptomyces Genetics**, 2nd Ed., John Innes Foundation, Norwich, UK.

Kim, J. H., Feng, Z.,et al. (2010). Cloning large natural product gene clusters from the environment: Piecing environmental DNA gene clusters back together with TAR. *Biopolymers* **93**(9): 833-844.

King, R. W., J. D. Bauer, et al. (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): 6257-6261.

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

Komatsu, M., Uchiyama T., et al. (2010). Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. *Proc Natl Acad Sci U S A* **107**(6): 2646-2651.

Koyama, J., Morita,I., et al. (2005). Structure-activity relations of azafluorenone and azaanthraquinone as antimicrobial compounds. *Bioorg. Med. Chem. Lett.***15**(4): 1079-1082.

Krapcho,A.P., Waterhous, D.J., (1999). *Heterocycles.* **51**: 737-749.

Kurobane, I., Vining, L.C., et al. (1980). Metabolites of Fusarium solani related to dihydrofusarubin. *J Antibiot(Tokyo).***33**(11): 1376-9.

Kurobane,I., Zaita, N., et al. (1986). New metabolites of Fusarium martii related to dihydrofusarubin. *J Antibiot(Tokyo).***39**(2): 205-14.

Lessmann, H., Krupa, J., et al. (1989). *Naturforsch.* **44b**, 353.

Lim, H. K., Chung, E.J. et al. (2005). Characterization of a forest soil metagenome clone that confers indirubin and indigo production on Escherichia coli. *Appl Environ Microbiol* **71**(12): 7768-7777.

Lombo, F., Abdelfattah, M., et al. (2009). Elucidation of Oxygenation Steps during Oviedomycin Biosynthesis and Generation of Derivatives with Increased Antitumor Activity. *ChemBioChem* **10**(2): 296-303.

MacNeil, I. A., Tiong,C.L. et al. (2001). Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J Mol Microbiol Biotechnol* **3**(2): 301-308.

Maplestone, R. A., Stone, M.J. et al. (1992). The evolutionary role of secondary metabolites -- a review. *Gene* **115**(1-2): 151-157.

Martinez, A., S. J. Kolvek, et al. (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): 2452-2463.

Matharu, A. L. , Cox, R. J., et al.(1998). MCAT is not required for in vitro polyketide synthesis in a minimal actinorhodin polyketide synthase from Streptomyces coelicolor , *Chem. Biol.* ***5***(12): 699-711.

McDaniel, R., Ebert-Khosla,S., et al. (1994). Engineered biosynthesis of novel polyketides: Influence of a downstream enzyme on the catalytic specificity of a minimal aromatic polyketide synthase. *Proc.Natl.Acad.Sci.USA* **91**: 11542-11546.

McDaniel, R., Ebert-Khosla,S., et al. (1994). *Am Chem Soc.* **116**:10855.

McDaniel, R., Ebert-Khosla, S., et al. (1995). Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. *Nature* **375**(6532): 549

McMurry,J., Begley, T.,(2005). The Organic Chemistry of Biological Pathways.

Medentsev, A.G., Akimenko, V.K., (1992). *Mikrobiologiya.* **61**: 824-829.

Medentsev, A.G., Arinbasarova, A.Y.,et al.(2005). *Appl.Biochem.Microbiol.***41**: 503-507.

Mendez,C., Kunzel, E., et al. (2002). Oviedomycin, an unusual angucyclinone encoded by genes of the oleandomycin-producer Streptomyces antibioticus ATCC11891. *J Nat Prod.* **65**(5): 779-82.

Miljkovic,A., Mantle, P.G.,(2001). *Nat.Prod.***64**:1251-1253.

Moran G.J., Krishnadasan A., et al. (2006). Methicillin-resistant S. aureus infections among patients in the emergency department. *N Engl J Med.* **355**(7): 666-74.

Moriyasu,Y., Miyagawa,H., et al. (2001). 5-Deoxy-7-methylbostrycoidin from cultured mycobionts from Haematomma sp. *Phytochemistry.* **58**(2): 239:41.

Morozova, O. and M. A. Marra (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**(5): 255-264.

Newman, D. J. and G. M. Cragg (2007). Natural products as sources of new drugs over the last 25 years. *J Nat Prod* **70**(3): 461-477.

Nok,A.J., (2002). *Cell Biochem. Function.* **20**:205-212.

Okunade,A.L., Clark,A.M., et al. (1999). Azaanthraquinone: an antimicrobial alkaloid from Mitracarpus scaber. *Planta Med.* **65**(5): 447-8.

Pang, M., Tan, G. A., Abdullah, N.,et al. (2008). Phylogenetic Analysis of Type I and Type II Polyketide Synthase from Tropical Forest Soil. *Biotechnology*. **7**(4): 660-668.

Park, S. R., Park, J.W., et al.(2008). Heterologous production of epothilones B and D in Streptomyces venezuelae. *Appl. Microbiol. Biotechnol.* **81**(1): 109-117.

Parisot,D., Devys,M., et al. (1989). *Chem.Sci.***44**: 1473-1474.

Parisot, D., Devys, M.,(1990). Naphthoquinone pigments related to Fusarubin from the fungus Fusarium solani (Mart.) Sacc. *Microbios.***64**(258): 31-47.

Piel,J., Hui, D.,et al. (2004). Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei. *Proc Natl Acad Sci USA.* **101**(46): 16222-7.

Prior,R.L., Wu,X., Schaich, K. (2005). Standardized methods for the determination of antioxidant capacity and phenolics in foods and dietary supplements. *J Agric Food Chem.* **53**(10): 4290-302.

Rappe, M. S. and S. J. Giovannoni (2003). The uncultured microbial majority. *Annu Rev Microbiol* **57**: 369-394.

Salzberg,S.L., Delcher, A.L., et al. (1998). Glimmer. *Nucleic Acids Research*. **26**: 544-548.

Schirmer, A., Gadkari, R., (2005). Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge Discodermia dissoluta. *Appl Environ Microbiol.* **71**(8): 4840-9.

Schmitz, J.E., Schuch, R., Fischetti,V.A. (2010). Identifying active phage lysins through functional viral metagenomics. *Appl.Environ.Microbiol.* **76**(21): 7181-7.

Schmitz,J.E., Daniel, A., et al. (2008). Rapid DNA library construction for functional genomic and metagenomic screening. *Appl Environ Microbiol.* **74**(5): 1649-52.

Schmitz, J.E., Schuch, R., et al. (2010). Identifying active phage lysins through functional viral metagenomics. *Appl Environ Microbiol.* **76**(21): 7181-7.

Seow, K. T., Meurer, G. ,et al. (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): 7360-7368.

Shen, B., (2000). *Curr Top Chem.**209:1.***

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

Thomas, R. (2001). A biosynthetic classification of fungal and striptomycete fused-ring aromatic polyketides. *Chembiochem.* **2**(9):612-27.

Torsvik, V., Daae, F.L.,et al. (1998). Novel techniques for analysing microbial diversity in natural and perturbed environments. *J Biotechnol* **64**(1): 53-62.

Torsvik, V.,Goksoyr, J.et al. (1990). High diversity in DNA of soil bacteria. *Appl Environ Microbiol* **56**(3): 782-787.

Torsvik,V., Ovreas, L. ,Thingstad, T. F. (2002). Prokaryotic diversity—magnitude, dynamics and controlling factors. *Science **296***(5570): 1064

Torsvik, V., Salte, K. et al. (1990). Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. *Appl Environ Microbiol* **56**(3): 776-781.

Tyson (2004) SoftBerry. *Nature* **428**: 37-43.

Van Wagoner, R.M., Mantle, P.G., Wright, J.L.C., (2008). Biosynthesis of Scorpinone, a 2-Azaanthraquinone from Amorosia littoralis, a Fungus from Marine Sediment. *J.Nat.Prod.***71** (3): 426-430.

Vogel, T. M.,Simonet P., et al. (2009). TerraGenome: a consortium for the sequencing of a soil metagenome. *Nat Rev Micro* **7**(4): 252-252.

Wang, G. Y., Graziani E., et al. (2000). Novel natural products from soil DNA libraries in a streptomycete host. *Org Lett* **2**(16): 2401-2404.

Wang, H. H.,Isaacs F.J., et al. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**(7257): 894-898.

Watve, M. G., Tickoo, R., Jog, M. M., Bhole B. D. (2001). How Many antibiotics are produced by the genus Streptomyces? *Arch.Microbiol.* **176**(5): 386-390.

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

Wawrik,B. ,Kutliev, D., et al.(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.

Wilkinson,S.P., Grove, A. (2006). Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. *Curr Issues Mol Biol.* **8**(1): 51-62.

Yang, B., Y. Peng, et al. (2010). Unsupervised binning of environmental genomic fragments based on an error robust selection of l-mers.*BMC Bioinformatics **11 Suppl 2**:* S5.

Yenesew, A., Ogur, J. A., Duddeck, H. (1993). *Phytochemistry* **34**, 1442.

Zerbino, D. R. and Birney,E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**(5): 821-829.

Zhang, W., Watanabe,K., et al. (2006). Heterologous biosynthesis of amidated polyketides with novel cyclization regioselectivity from oxytetracycline polyketide synthase. *J Nat Prod.* **69**(11): 1633-6.