Accessing Bioactive Natural Products from Cultured and Uncultured Microorganisms

by

Michael M. Schofield

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Microbiology and Immunology)
in the University of Michigan
2015

Doctoral Committee:

Professor David H. Sherman, Chair
Professor Gregory J. Dick
Professor Philip C. Hanna
Professor Patrick D. Schloss

This work is dedicated to my late grandfather, Richard (Dick) Lane.

Acknowledgements

Firstly, I would like to thank my adviser, Dr. David H. Sherman, for his guidance and support throughout my graduate career. The mentorship and laboratory environment he provided allowed me to develop research independence and the freedom to acquire the skillsets for a nonacademic career. I would also like to thank my committee members, Profs. Gregory Dick, Philip Hanna, and Patrick Schloss, for their continued support, insight, and guidance throughout the development of my thesis projects.

I am also grateful to my original mentors and former Sherman laboratory members Drs. Tyler Nusca (cultured bacteria) and Chris Rath (uncultured bacteria). They helped shape the very beginning of each of my dissertation projects. I owe a huge thank you as well to all past and current members of the Sherman laboratory. Working in such a collaborative and multidisciplinary laboratory environment has been a wonderful experience. I am grateful also for the constant efforts of Shamilya Williams and Pamela Shultz to help maintain this environment.

I would also like to thank all collaborators and co-authors that I have worked with throughout my graduate studies. I am particularly grateful to Dr. Ashu Tripathi from my own laboratory and Sunit Jain from the laboratory of Prof. Gregory Dick. Working with both of you exposed me to new branches of science and deepened my appreciation for the power of interdisciplinary work. Neither thesis project would have been as rewarding without your patience in teaching me new techniques, and the findings would not have been as exciting or impactful without your contributions.

I am also thankful to my undergraduate mentors, Cindy June and Drs. Bradley Wallar and Rachel Powers from Grand Valley State University. Your early guidance led me here, and your ongoing mentorship has helped shape me into the scientist I am today.

I would also like to thank all of my friends along with my two wonderful families for providing constant support and encouragement throughout the past five years. Whether it was being understanding of my frequent mental or physical absences, lending an ear, or just giving me an occasional word of encouragement, I could not have done this without all of you.

Finally, I would like to thank my husband and our ever-expanding, furry-footed family. Graduate school has been extremely tough at times, and I have always been grateful for your ability to keep me grounded. I am looking forward to starting the next chapter of our lives together in July.

Preface

This dissertation contains five chapters comprising the majority of my investigations into natural products produced by both cultured and uncultured bacteria. Chapter 1 begins with an introduction to natural products and the need for new antibiotics. A portion of this chapter was published in a review article during my graduate studies. (**Schofield MM** and Sherman DH. 2013. "Meta-omic characterization of bacterial gene clusters for natural product biosynthesis." *Curr. Opin. Biotechnol.* 2013, 24(6):1151-8).

Chapter 2 describes my efforts to identify a novel natural product from a cultured bacterium through the targeted screen of an extensive extract library. The work described in this chapter was previously published in a co-first author paper in the *Journal of the American Chemical Society*. (Tripathi A, **Schofield MM**, Chlipala GE, Schultz PJ, Yim I, Newmister SA, Nusca TD, Scaglione JB, Hanna PC, Tamayo-Castillo G, Sherman DH. "Baulamycins A and B, Broad-Spectrum Antibiotics Identified as Inhibitors of Siderophore Biosynthesis in Staphylococcus aureus and Bacillus anthracis." *J. Am. Chem. Soc*. 2014, 136(4):1579-86).

Chapter 3 delves into my metagenomic studies on the uncultured bacterium that makes the anti-cancer natural product ET-743. A manuscript encompassing portions of this chapter was prepared for submission to the *Proceedings of the National Academy of Sciences of the USA* for peer review during the completion of my thesis work.

Chapter 4 continues my ongoing investigation into ET-743. In this chapter, I describe our biochemical breakthroughs to characterize the enzymes involved in the biosynthesis of the ET-743 tetrahydroisoquinoline core.

Finally, I discuss ongoing and potential future directions for these research projects in Chapter 5.

TABLE OF CONTENTS

Dedication ii

Acknowledgements iii

Preface v

LIST OF FIGURES viii

LIST OF TABLES xi

Chapter 1 Background and Introduction 1
1.1 A Need for New Medicines 1
1.2 Natural products provide unique and desperately needed chemical scaffolds. 3
1.3 Microorganisms are an important source for natural products 4
1.4. Uncultivated microorganisms are an unexploited source of novel medicines 6
1.5. Summary and Chapter Outline 7
1.6 Notes 9

Chapter 2 Accessing natural products from cultured bacteria: Discovery of the broad-spectrum baulamycin antibiotics 10
2.1 Introduction 10
2.2 Results 12
2.2.1 High-Throughput Screening for Inhibitors of Siderophore Biosynthesis Derived from Natural Product Extracts 12
2.2.2. Isolation and Structural Elucidation of the Baulamycins (6−7) 13
2.2.3. Assessment of the *in vitro* Biological Activity of Compounds against Purified NIS Enzymes 19
2.2.4. Initial Enzyme Mechanism of Inhibition Studies. 21
2.2.5 Assessment of Biological Activity against Microbial Cultures. 22
2.3. Conclusion 24
2.4. Supplemental Information 26
2.5 Materials and Methods 50
2.6 Notes and Author Contributions 58

Chapter 3 Understanding natural products from uncultured bacteria: Metagenomic analysis of the endosymbiont producer of a chemotherapeutic natural product 59
3.1. Introduction 59
3.2 Results and Discussion 61
3.2.1 Overview of Samples and Dataset. 61
3.2.2 Genome Reduction in the Symbiont. 63
3.2.3 Phylogenetic analysis and novelty of *Ca.* E. frumentensis. 65
3.2.4 Primary Metabolism. 66
3.2.5 Secondary Metabolism. 70

3.3 Conclusions **73**
3.4 Supplemental Information **76**
3.5 Materials and Methods **88**
3.6 Notes and Author Contributions **90**

Chapter 4 Accessing natural products from uncultured bacteria: Initial Biochemical Investigations of ET-743 Biosynthetic Enzymes **91**
4.1 Introduction **91**
4.2 Results **96**
4.2.1 Codon optimization permits expression and purification of enzymes involved in ET-743 biosynthesis. 96
4.2.2 Biochemical assessment of EtuA1, an NRPS that selectively activates glycolic acid. 98
4.2.3 EtuA1 exhibits some promiscuity in substrate activation 100
4.3 Conclusions **102**
4.4 Supplemental Information **103**
4.5 Materials and Methods **105**
4.6 Notes and Author Contributions **106**

Chapter 5 Discussion and Future Work **107**
5.1 Overview **107**
5.2 Ongoing and future work for cultured microorganisms **110**
5.2.1 Identification and manipulation of the baulamycin gene cluster to increase antibiotic production and make more potent analogues 110
5.2.2 Identifying additional natural product inhibitors of siderophore biosynthesis 112
5.3 Future directions and ongoing work to access the natural products of uncultured microorganisms **115**
5.3.1 *In vitro* reconstitution of the ET-73 biosynthetic core (Figure 5-5A) 115
5.3.2 Using metagenomic information to cultivate the producer of ET-743 (Figure 5-5B) 116
5.3.3 Biosynthesis of ET-743 or new analogues through a heterologous host like *Streptomyces lavendulae* (Figure 5-5C) 117
5.4 Conclusions **118**

References **119**

LIST OF FIGURES

Figure 1-1. Most of today's antibiotics are created by synthetically tailoring a conserved chemical core. 2

Figure 1-2. Natural products are an old but effective source of medicine. 3

Figure 1-3. Examples of marketed natural products isolated from cultured microorganisms. 5

Figure 1-4. An outline of my dissertation 8

Figure 2-1. Biosynthesis of virulence-associated siderophores 11

Figure 2-2. Triage of hits derived during high throughput screening of SbnE and AsbA against the full microbial NPE library 13

Figure 2-3. Planar structure of baulamycins A (6) and B (7) showing COSY correlation with bold bonds and HMBC correlations with arrow. 14

Figure 2-4: H-H, C-H coupling values and relative configuration determined for C1'-C11. Arrows showing ROESY correlations. 17

Figure 2-5: H-H, C-H coupling values and relative configuration determined for C10-C4. Arrows showing ROESY correlations. 18

Figure 2-6. *In vitro* and live culture studies of baulamycin bioactivity. 20

Figure 2-7: Secondary HTS analysis on active NPEs from primary screening. 26

Figure 2-8. Development of an NIS synthetase assay for screening. 27

Figure 2-9. Overview of HTS and Initial Hits. 28

Figure 2-10: 16S rRNA analysis and alignment. 29

Figure 2-11: Small and large scale extraction schemes. 30

Figure 2-12. Bioactivity Guided De-replication of NPE. 31

Figure 2-13: HRAPCIMS spectrum for baulamycin A (6). 32

Figure 2-14: ^{1}H NMR spectrum of baulamycin A (6) recorded at 700 MHz (in CD_3OD). 33

Figure 2-15: ^{13}C NMR spectrum of baulamycin A (6) recorded at 176 MHz. (in CD_3OD). 34

Figure 2-16: HSQCAD NMR spectrum of baulamycin A (6) recorded at 500 MHz (in CD_3OD). 34

Figure 2-17: gCOSY NMR spectrum of baulamycin A (6) recorded at 500 MHz (in CD_3OD). 35

Figure 2-18: gHMBC NMR spectrum of baulamycin A (6) recorded at 500 MHz (in CD_3OD). 36

Figure 2-19: HOMO2DJ NMR spectrum of baulamycin A (6) recorded at 500 MHz (in CD_3OD). 37
Figure 2-20: HETLOC-gse NMR spectrum of baulamycin A (6) recorded at 500 MHz (in CD_3OD)..... 38
Figure 2-21: ROESYAD NMR spectrum of baulamycin A (6) recorded at 500 MHz (in CD_3OD)..... 39
Figure 2-22: ^{1}H NMR spectrum of baulamycin A (6) recorded at 500 MHz (in DMSO-d6). 40
Figure 2-23: ^{13}C NMR spectrum of baulamycin A (6) recorded at 125 MHz (in DMSO-d6). 41
Figure 2-24: HSQCAD NMR spectrum of baulamycin A (6) recorded at 500 MHz (in DMSO-d6). 41
Figure 2-25: PS-DQF-COSY NMR spectrum of baulamycin A (6) recorded at 500 MHz (in DMSO-d6). 42
Figure 2-26: HRAPCIMS spectra for baulamycin B (7). 43
Figure 2-27: ^{1}H NMR spectrum of baulamycin B (7) recorded at 700 MHz (in CD_3OD)..... 44
Figure 2-28: ^{13}C NMR spectrum of baulamycin B (7) recorded at 700 MHz (in CD_3OD)..... 44
Figure 2-29: gHSQCAD NMR spectrum of baulamycin B (7) recorded at 700 MHz (in CD_3OD). 45
Figure 2-30: COSY NMR spectrum of baulamycin B (7) recorded at 700 MHz (in CD_3OD). ... 46
Figure 2-31: HMBCAD NMR spectrum of baulamycin B (7) recorded at 700 MHz (in CD_3OD). 47
Figure 2-32. Kinetic analysis of the inhibition mechanism of BmcA. 48
Figure 2-33. Live culture bioactivity Studies 48
Figure 2-34. Phylogenetic trees 49
Figure 3-1. Mangrove tunicates the tetrahydroisoquinoline natural products. 60
Figure 3-2. A circular map of the closed genome of *Candidatus* E. frumentensis..... 62
Figure 3-3. Overview of the metabolism of *Ca.* E. frumentensis deduced from genomic analysis. 66
Figure 3-4. Completion of the ET-743 biosynthetic gene cluster..... 70
Figure 3-5. Origin of the four metagenomic DNA samples used to compile the consensus microbial genome of the ET-743 producer..... 76
Figure 3-6. An Emergent Self Organized Map of one of the four metagenomic DNA samples (IMG Submission ID 21664) from Ecteinascidia turbinata..... 77
Figure 3-7. The gene content of drastically reduced genomes. 78
Figure 3-8. Phylogenetic reconstruction using conserved markers. 81
Figure 3-9. Relatedness between the 16S rRNA gene and the average amino acid identity (AAI) for Ca. E. frumentensis and similar microorganisms..... 82

Figure 4-1. An outline of currently pursued routes to access the chemotherapeutic natural product ET-743 92
Figure 4-2. Four tetrahydroisoquinoline natural products and the conserved NRPS modules responsible for the biosynthesis of their cores. 93
Figure 4-3. Biosynthesis of the ET-743 biosynthetic core. 95
Figure 4-4. *Ca*. E. frumentensis utilizes different codons from many commonly used heterologous hosts. 97
Figure 4-5. The role of EtuA1 in the biosynthesis of ET-743 98
Figure 4-6. EtuA1 preferentially activates glycolic acid over other predicted substrates 100
Figure 4-7. Examination of the substrate flexibility of EtuA1 101
Figure 4-8. Expression of various ET-743 biosynthetic enzymes 104
Figure 4-9. Suboptimal purification of EtuA1-3 NRPS modules 104
Figure 4-10. Optimization of the malachite green assay for biochemical analysis of EtuA1 105
Figure 5-1. A review of my dissertation 109
Figure 5-2. Identification of the PKS gene cluster responsible for baulamycin biosynthesis. ... 110
Figure 5-3. Long-term goals of the baulamycin project. 111
Figure 5-4. Initial bioactivity of strain 44478-N5I against NIS synthetases SbnE and AsbA.... 114
Figure 5-5. An outline of current routes being taken to access the chemotherapeutic natural product ET-743 115
Figure 5-6. Culturing E. turbinata-derived bacteria in a microfluidic device..... 117

LIST OF TABLES

Table 1-1. Examples of natural products suspected to originate from uncultured microorganisms. 6

Table 2-1. NMR spectroscopic data for baulamycins A (6) and B (7) in CD_3OD at 700 MHz. .. 15

Table 2-2. *In vitro* bioactivity analysis of the baulamycins against purified enzymes. 21

Table 2-3. Bioactivity analysis of the baulamycins against Gram-positive and Gram-negative microbial cell cultures. 23

Table 2-4. Enzyme kinetics 50

Table 3-1. General features of the *Candidatus* E. frumentensis genome. 64

Table 3-2. An overview of the four metagenomic DNA sequence datasets isolated from Ecteinascidia turbinata. 83

Table 3-3. Coverage comparison between the complete Ca. E. frumentensis genome and the discarded contig. 84

Table 3-4. An overview of the differences in genomes of endosymbionts, intracellular pathogens, and free-living microorganisms. 85

Table 3-5. Pseudogenes identified in the noncoding regions of the *Ca*. E. frumentensis genome. 86

Table 3-6. New genes proposed to be involved in ET-743 biosynthesis in *Ca*. E. frumentensis. 87

Table 4-1. Cloned constructs for expression of EtuA1-3 103

Table 5-1. Top hits active against both AsbA and SbnE in the high throughput screen against the natural product extract library. 113

Chapter 1
Background and Introduction

1.1 A Need for New Medicines

Every year, at least two million people in the U.S. acquire infections that cannot be effectively treated with available antibiotics. These antibiotic-resistant illnesses directly kill 23,000 people annually, and even more die of associated medical complications (1).

Despite these frightening statistics, researchers have identified few new drugs to treat resistant infections. In fact, the vast majority of today's antibiotics belong to classes that were discovered before 1970 (2). Members from each of these of these drug classes possess a common core structure, or scaffold (**Figure 1-1**). When bacteria begin to develop resistance to an antibiotic, synthetic chemists modify the core scaffold with chemical groups that improve the drug's potency or target specificity. The end result is multiple generations of each antibiotic class (**Figure 1-1**).

This semi-synthetic method of drug development is an effective short-term strategy for fighting antibiotic resistance. However, it is not enough. Bacteria are acquiring resistance to new generations of antibiotics faster than scientists are able to synthetically tailor drugs and receive marketing approval (3). Many bacteria have even developed mechanisms that now provide resistance against an entire antibiotic class (2). For example, the generations of penicillin antibiotics each possess a β-lactam ring that inhibits cell wall biosynthesis (**Figure 1-1**). However, some bacteria possess β-lactamase enzymes that hydrolyze the core scaffold of these antibiotics and render them useless. Since the introduction of new generations of penicillin and other β-lactam antibiotic classes (*e.g.* cephalosporins, monobactams, and carbapenems), the number and diversity of β-lactamases has skyrocketed (4). Today, there are over 1,000

recognized β-lactamases found in a wide range of microbial species, severely limiting our ability to utilize many β-lactam antibiotics (5).

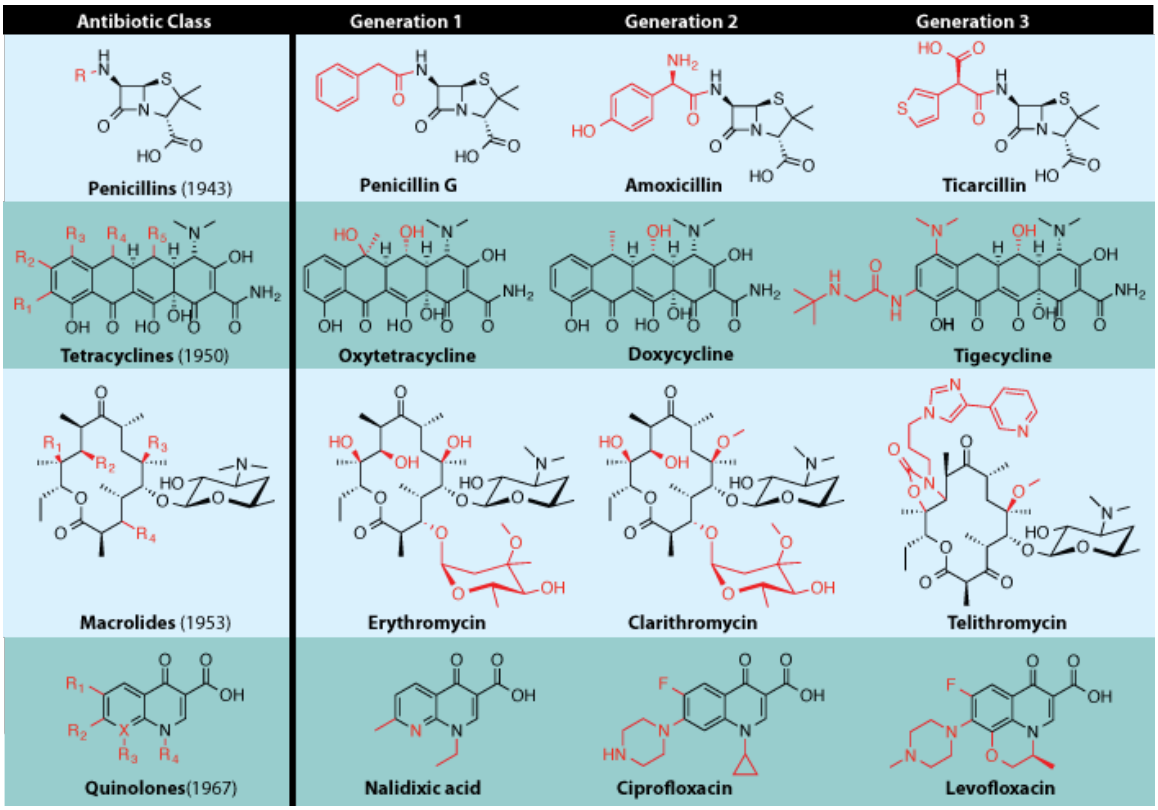


Figure 1-1. Most of today's antibiotics are created by synthetically tailoring a conserved chemical core.
The preserved scaffolds are depicted in black while added chemical groups are shown in red for the first three generations of each antibiotic class. The original year of discovery for each antibiotic class is also shown. Modified from Fischbach et. al. 2009 and Walsh et al. 2014 (6, 7).

Antibiotic resistance is a huge public health concern, but bacteria are not the only lifeforms that acquire resistance to currently marketed medicines. Viruses (8, 9), fungi (10), and even cancers (11, 12) are continually developing resistance to marketed antivirals, antifungals, and chemotherapeutic medicines. Finding new medicines is critical to ensure that we can continue to protect ourselves from drug resistant infections. New drugs that possess both a novel core scaffold and molecular target may help delay the development of resistance mechanisms (3). The biggest challenge is finding sources for new medicines.

1.2 Natural products provide unique and desperately needed chemical scaffolds.

Our surrounding environment provides one important source for new medicines. Mankind has relied on nature to treat different ailments for millennia. Early physicians collected plants, animals, and sediments from their environment to create elixirs and salves that were used as some of the world's first pharmaceuticals. Although not all of these unique concoctions were effective, the process of trial and error allowed early physicians to identify effective medicine sources (13). Even today, many cultures throughout the world rely on herbal remedies that have been passed down for generations to treat pain and disease (14).

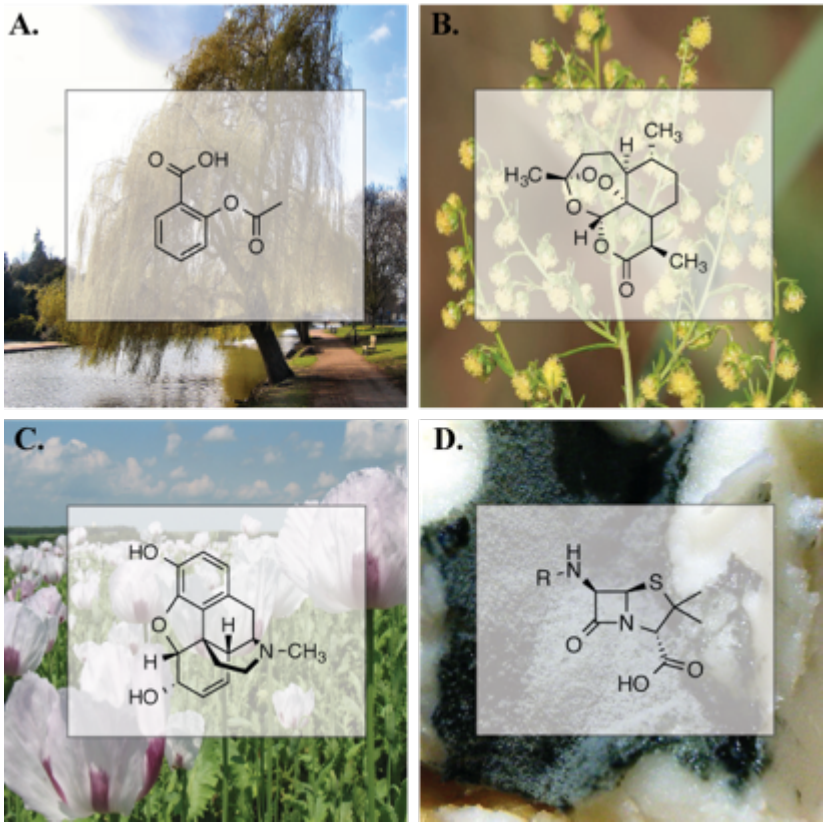


Figure 1-2. Natural products are an old but effective source of medicine.
A.) The natural product aspirin derived from willow bark. **B.)** The antimalarial artemsinin from the Chinese herb *Artemisia.* **C.)** The natural product morphine derived from opium poppies. **D.)** Penicillins are a class of antibiotics originally isolated from *Penicillium* fungi.

The most successful of these natural remedies have a strong scientific basis. Plants, animals, and bacteria are all capable of producing complex and bioactive small molecules called natural products. Today, we have the scientific ability to identify the natural product responsible for the

therapeutic effect in many of the most potent of these ancient remedies. For example, the compound salicin (aspirin, **Figure 1-2A**) is the active ingredient in dried willow bark concoctions that have been used in pain relief for the last 6,000 years (13). The antimalarial natural product artemsinin (**Figure 1-2B**) is responsible for the medicinal effects found in the Chinese herb *Artemisia annua* (14). Civilizations throughout Europe and Asia used poppies in medical treatments for thousands of years (**Figure 1-2C**). Today, the active ingredient morphine is common in pain relief and surgeries (15).

The continued prevalence of natural products in modern medicine is largely due to their potency and unique chemistry. Unlike manmade compounds, most natural products have elaborate structures with complex three-dimensional shapes (16). Evolution has also already tailored these unique compounds to penetrate membranes and interact with biological targets (17). This gives natural products a huge advantage over synthetic compounds, which often have to be tested and repeatedly fine-tuned before achieving any kind of biological relevance (17, 18).

With so many advantages, it is unsurprising that the majority of drugs currently marketed in the U.S. are natural products or natural product-inspired derivatives (19, 20). However, natural products do have some disadvantages. It is not always environmentally sustainable to harvest animals, plants, or other macroorganisms to acquire marketable amounts of the potent compounds. Bark from the evergreen tree *Taxus brevifolia*, for example, possess the powerful chemotherapeutic natural product taxol (21). But the evergreen tree is an endangered species, and large-scale harvest of the compound from the native source has never been an option. Fortunately, scientists are sometimes able to develop synthetic methods to produce these compounds in the laboratory. Although synthetic routes to natural products can be effective, they are also frequently time-consuming and costly. For environmental reasons, taxol is currently produced by a synthetic method, but its native evergreen producer remains the most reliable source for the powerful drug (22).

1.3 Microorganisms are an important source for natural products

Fortunately, there are alternative sources of natural products that are economically and environmentally sustainable to harvest on a large scale. Alexander Fleming's discovery of penicillin in the 1920s from a *Penicillium* fungus (**Figure 1-2D**) gave way to the beginning of

the antibiotic age and the widespread use of microbial natural products. Today, microbial natural products and their derivatives continue to constitute many currently marketed medicines (**Figure 1-3**).

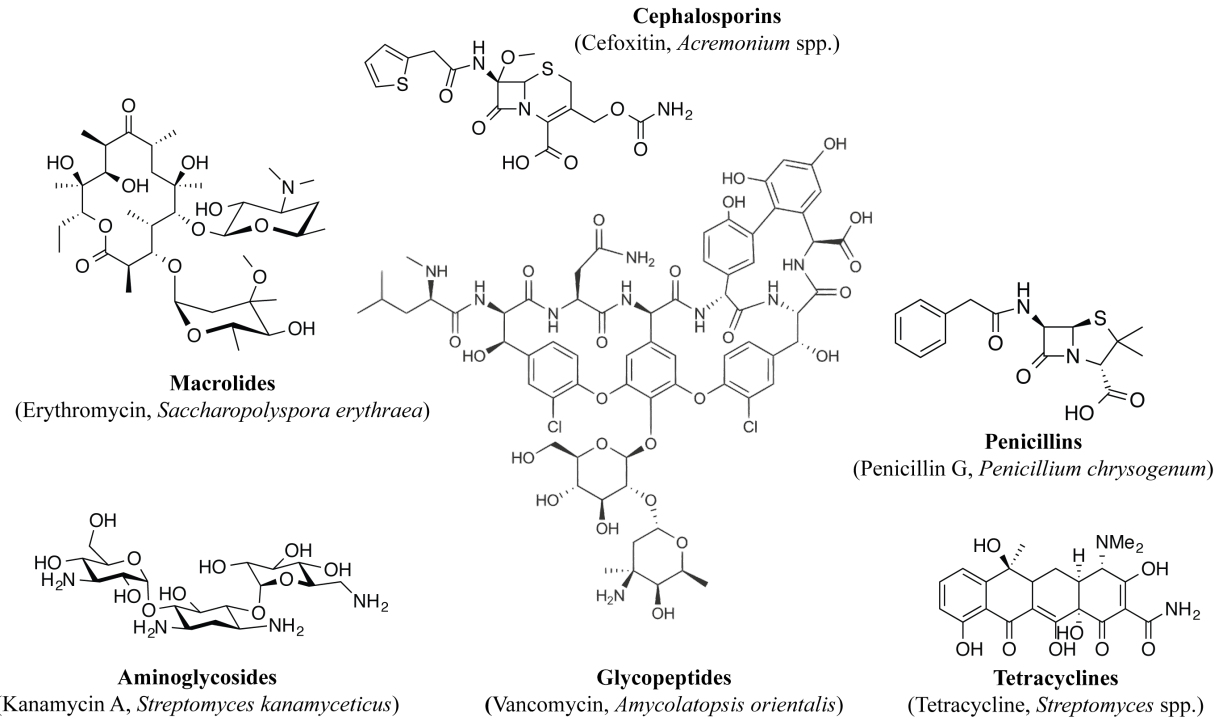


Figure 1-3. Examples of marketed natural products isolated from cultured microorganisms. Antibiotic classes are shown in bold. The depicted compound and the cultured organism that produces it are in parentheses.

Microorganisms are ideal sources for unique medicines (23, 24). They produce a wide variety of biologically intriguing compounds that play an important role in an ongoing war against their neighbors for environmental resources. Many microorganisms can also be grown in the laboratory using standard techniques. Since most species reproduce quickly, laboratory cultivation can lead to the rapid large-scale production of potent medicines without having a negative ecological impact. Furthermore, microorganisms have simple genomes in comparison to plants, animals, or insects. This permits quick and easy manipulation of the genes responsible for production of natural products, potentially helping researchers make more potent or selective medicines.

Given all of these benefits, it is unsurprising that researchers have identified more than 20,000 microbial natural products to date (25). But the sheer abundance of microbes on the planet suggests we have not even begun to exploit their true pharmaceutical potential.

1.4. Uncultivated microorganisms are an unexploited source of novel medicines

Cultured microorganisms have led to the discovery of many potent natural products used to treat a wide range of infections and diseases (**Figure 1-3**). However, less than 1% of the world's microorganisms can currently be cultivated in the laboratory using standard techniques (26-28).

Given the pharmaceutical success of natural products from cultured bacteria, it is very likely that the world's uncultured majority could also be hiding a wealth of bioactive compounds. In fact, many natural products—including some currently in clinical trials—are predicted to originate from uncultured bacteria (**Table 1-1**). These microorganisms may have complex metabolisms and require access to specific nutrients or precise environmental conditions to grow outside of their native environment. They are therefore currently considered uncultured.

Compound Name	Isolated Source	Therapeutic Use	Evidence for Microbial Origin
***Yondelis, ET-743**	Tunicate	Cancer	NRPS origin, FISH studies, meta-omics (see Chapter 3)
***Bryostatin 1**	Bryozoa	Cancer and Alzheimer's	PKS origin, FISH, PCR studies
***Hemiasterlin**	Sponge	Cancer	Found in diverse sponge species
Pederin	Beetles	Cancer	PKS origin, similar structure to onnamide, meta-omics
Onnamide	Sponge	Cancer	PKS origin, similar structure to pederin, meta-omics
Patellamides	Cyanobacteria, Tunicate	Cancer	Meta-omic studies

Table 1-1. Examples of natural products suspected to originate from uncultured microorganisms. Natural products that are currently approved or undergoing clinical trials are indicated with an asterisk.

Since researchers have failed to grow these microbes using standard techniques, synthetic methods are typically used to access their potent natural products. However, understanding more about their unique biology could one day pave the way for methodologies that permit the growth of these microbes in the laboratory (29, 30). Growth of these elusive microbes could facilitate large-scale production of medicines and access to more potent or selective analogues. Even if cultivation continues to prove impossible, a better understanding of the microorganism's biology could facilitate access to these medicines. Understanding their biosynthetic gene clusters could permit new routes to harness their natural products, either through heterologous expression of their complete gene clusters in cultured bacteria or through manipulation of key enzymes (31).

1.5. Summary and Chapter Outline

Natural products represent a very old but important and successful source of novel medicines for different diseases and ailments. Their ability to provide novel chemical scaffolds with unique molecular targets is increasingly important as bacteria, viruses, fungi, and cancers develop resistance to marketed medicines. Microorganisms in particular provide a critical source of natural products that can be both economically and environmentally sustainable to produce on a larger scale. To maximize the discovery of new medicines, researchers need to simultaneously explore the natural products of both cultured and currently uncultured microorganisms.

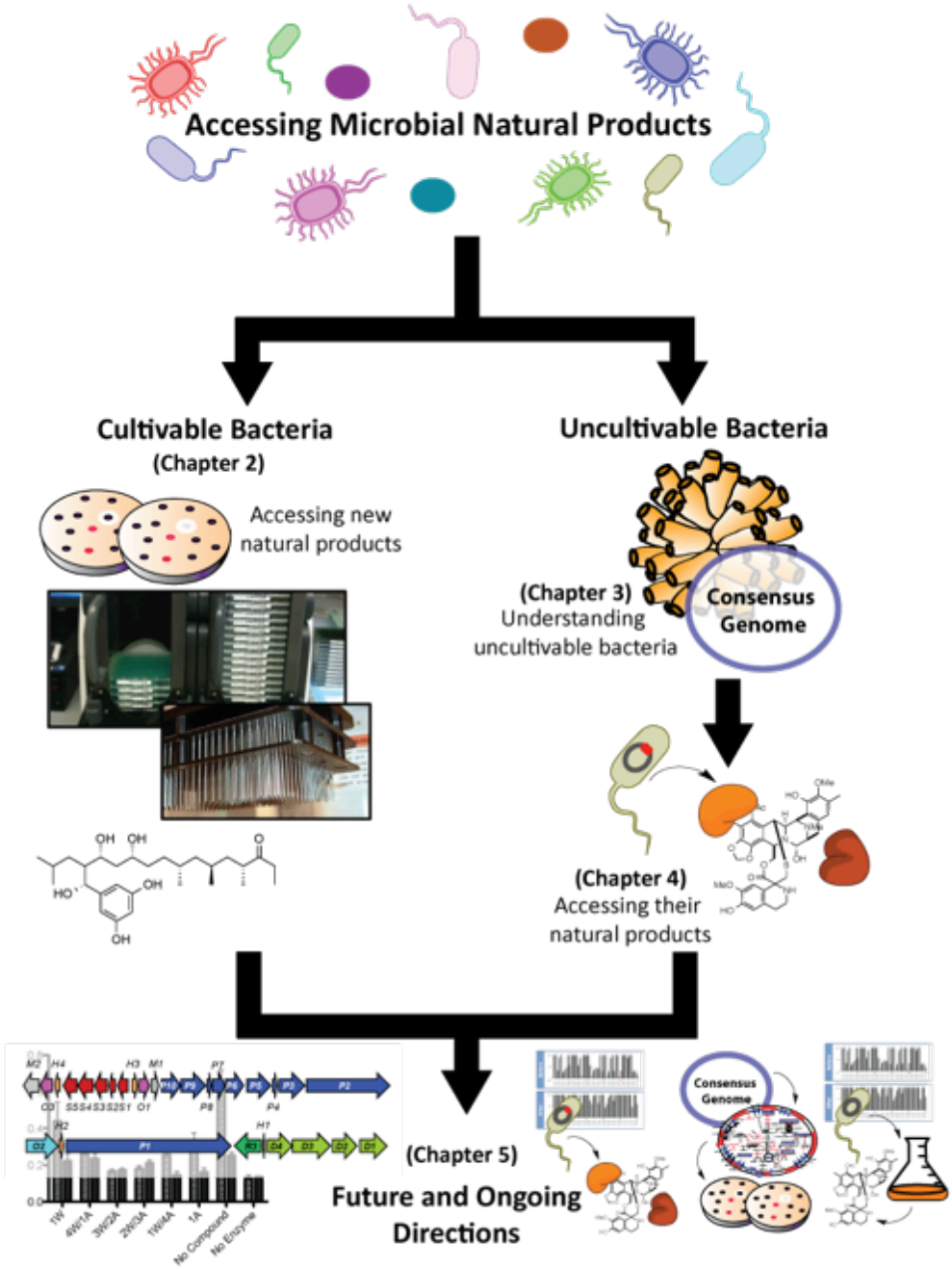


Figure 1-4. An outline of my dissertation

Chapter 2 describes a targeted high throughput screen from an extract library to identify novel natural products. Chapters 3 and 4 outline techniques to access the products of uncultured microorganisms. In chapter 3, we utilize metagenomic methods to understand an uncultured microorganism. In chapter 4, we use this new information to begin work that could one day enable us to access its natural product. Chapter 5 describes future directions and ongoing work for both of these projects.

My dissertation is equally split between two distinct thesis projects exploring new ways to access microbial natural products (**Figure 1-4**). The first project (chapter 2) utilizes a natural product extract library derived from cultured bacteria to identify medicines with both a novel chemical scaffold and a unique molecular target. Screening the expansive library for inhibitors of an enzyme that facilitates iron acquisition in pathogenic bacteria led to the discovery of a new class of antibiotics.

My second project (chapters 3 and 4) explores routes to understand and harness the natural products of bacteria that cannot currently be grown in the laboratory. In chapter 3, we used metagenomic methods to uncover the long elusive genome of an endosymbiotic bacterium capable of making a powerful and clinically approved chemotherapeutic drug called ET-743. We then analyzed the genome's primary and secondary metabolism, gaining significant insight into the microorganism's biology, the production of its bioactive natural product, and its long-term evolution with its animal host.

In chapter 4, we utilized the newfound knowledge from this uncultured genome to take the first steps toward more efficient access of its chemotherapeutic natural product and new analogues. We cloned the genes responsible for production of the unique core of the natural product and successfully expressed them in a cultured, heterologous host. We then provided biochemical evidence for one enzyme in the biosynthesis of the natural product and demonstrated that it can be manipulated to potentially allow for the creation of new analogues.

Chapter 5 includes a summary of results along with future and ongoing areas of study for each project.

1.6 Notes

All photographs in **Figure 1.1** are licensed under creative commons 2.0 or 3.0 licenses. The original photographs were altered to include an overlay of the chemical structures of the corresponding natural products. Credits are as follows: **A)** Photograph by Ronald Saunders. **B)** Photograph by Kristin Peters. **C)** Photograph by Wolfgang Horlacher. **D)** Photograph by Steve Jurvetson.

Chapter 2
Accessing natural products from cultured bacteria: Discovery of the broad-spectrum baulamycin antibiotics

2.1 Introduction

The rapid ability of pathogens to develop resistance to antibiotics is endangering the management of a multitude of serious infections (6, 32, 33). In alarming contrast to the increase of bacterial adaptation to currently marketed drugs, the discovery of novel classes of antimicrobials is on the decline (2). Although the majority of pharmaceutical efforts during the past six decades have focused on the synthetic enhancement of a limited set of unique core scaffolds, a more sustainable route to combat antibiotic resistance is the discovery of novel chemical structures possessing unique microbial targets (6, 34). Iron acquisition mechanisms in particular may represent effective antimicrobial targets that present substantial hurdles to bacterial antibiotic resistance (35). Iron is required for growth and survival of bacteria but remains tightly regulated in the mammalian host. Many pathogenic Gram-positive and Gram-negative bacteria utilize virulence-associated siderophores to scavenge iron from this restricted environment and return it to the microbial cell (36). Although previous studies have corroborated siderophore biosynthetic enzymes as effective drug targets through the tailoring of established chemical scaffolds (37-41), no novel chemical structures have been identified.

In our efforts to identify new structural antibiotic classes with inhibitory activity against siderophore biosynthetic enzymes, we selected the Gram-positive bacteria, methicillin-resistant *Staphylococcus aureus* (MRSA) and *Bacillus anthracis*, as model systems. The "superbug" MRSA is a major public health concern, attributed to more than 18,000 deaths a year in the United States (6, 41). In contrast, the spore-forming microorganism *B. anthracis* is the causative agent of anthrax. The ability of the bacterium to quickly achieve high concentrations within

infected hosts makes it a serious bioterrorism threat, with mortality rates for inhalational infection historically reaching as high as 94% (42).

Both pathogens are strongly associated with antimicrobial resistance, and their siderophore biosynthetic pathways have been extensively characterized (43, 44). The siderophores staphyloferrin B (**2**) of *S. aureus* (45-50) and petrobactin (**5**) of *B. anthracis* (51-58) in particular have been shown to be critical for survival in iron-limited environments.

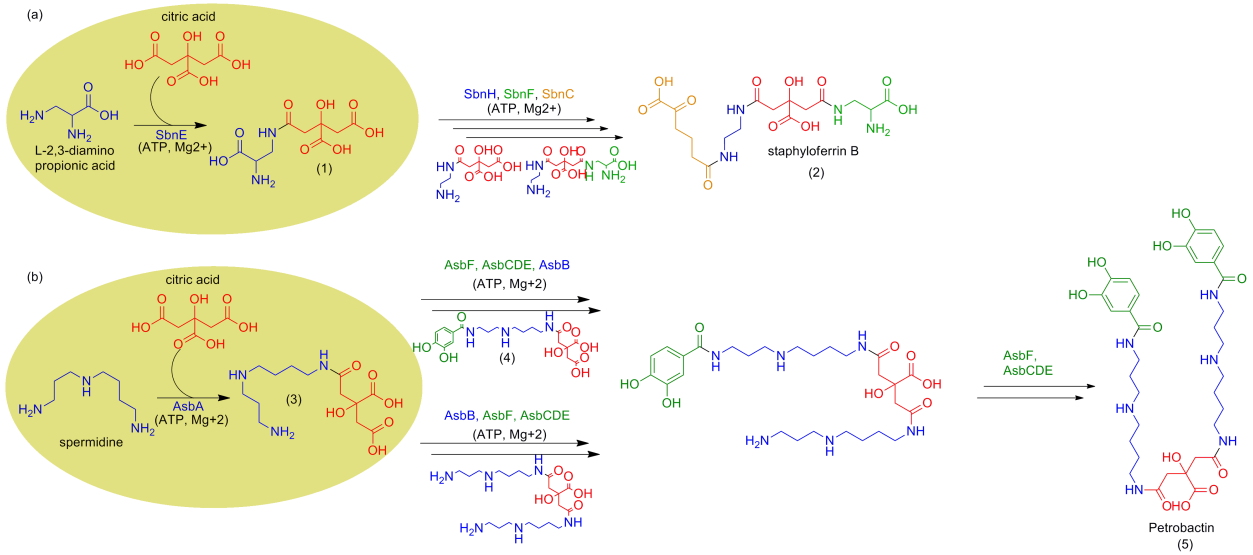


Figure 2-1. Biosynthesis of virulence-associated siderophores
(**A**) staphyloferrin B in *S. aureus* (**B**) petrobactin in *B. anthracis*. The target biosynthetic enzymes, SbnE and AsbA, are highlighted by the yellow circles

The biosynthetic pathways for both siderophores involve a type A nonribosomal peptide synthetase independent siderophore (NIS) synthetase, SbnE (**Figure 2-1A**) in staphyloferrin B, and AsbA (**Figure 2-1B**) in petrobactin. Type A NIS synthetases are a unique class of enzymes found within siderophore biosynthetic pathways of a number of pathogenic bacteria, including *Shigella flexneri*, *Escherichia coli*, and *Salmonella typhimurium* (59, 60). These enzymes catalyze the condensation of citric acid with either a polyamine or amino alcohol substrate in an ATP-dependent reaction (59, 60). Since type A NIS synthetases share similar catalytic mechanisms and substrate preferences, we sought to identify novel antibiotics against *S. aureus* or *B. anthracis* that could also serve as broad- spectrum antibiotics against other NIS synthetase-containing pathogens.

2.2 Results

2.2.1 High-Throughput Screening for Inhibitors of Siderophore Biosynthesis Derived from Natural Product Extracts

Fueled by the need to identify a new structural class of antibiotics, we selected a marine microbial-derived natural product extract (NPE) library to identify inhibitors of NIS synthetases in *S. aureus* and *B. anthracis*. The majority of currently marketed drugs are derived from natural products (19, 20), and the marine environment in particular is thought to be an underexplored source of novel chemical structures (23). We developed an enzymatic high-throughput assay specific to SbnE and AsbA and adapted it to a malachite green reporter system (61-63). The assay was then queried against a library of 19,855 marine microbial-derived NPEs to identify inhibitors of enzyme activity (**Figure 2-2**). The NPEs in this library are primarily from marine-derived microorganisms collected in Costa Rica, Panama, and Papua New Guinea. Screening SbnE and AsbA in parallel limited the number of false positives and provided a convenient method for prioritization of active extracts. Two rounds of screening yielded 33 strains with reater than 70% inhibition against SbnE at 7.5 μg/mL and 22 strains with extracts showing greater than 50% inhibition against AsbA at 75 μg/mL (**Figure 2-2**). The higher frequency and activity of extracts against SbnE is likely due to the lower enzyme concentration (25 nM) compared to the AsbA (100 nM) utilized in screening. These were the lowest enzyme concentrations for each respective assay that produced a response fit for screening (**Figure 2-8 in Supporting Information [SI**]). Low enzyme concentrations were desired to maximize the likelihood of detecting extracts possessing low-abundance bioactive components.

Among active extracts, *Streptomyces tempisquensis* (**Figure 2-10 in SI**) was of particular interest due to its high activity against both SbnE (95.9%) and AsbA (90.2%) (**Figure 2-9 in SI**). The strain was originally isolated from sediments collected in Playa Grande, Costa Rica (−85° 49’ 39.8’’, 10° 18’ 39.8’’) near Las Baulas National Marine Park.

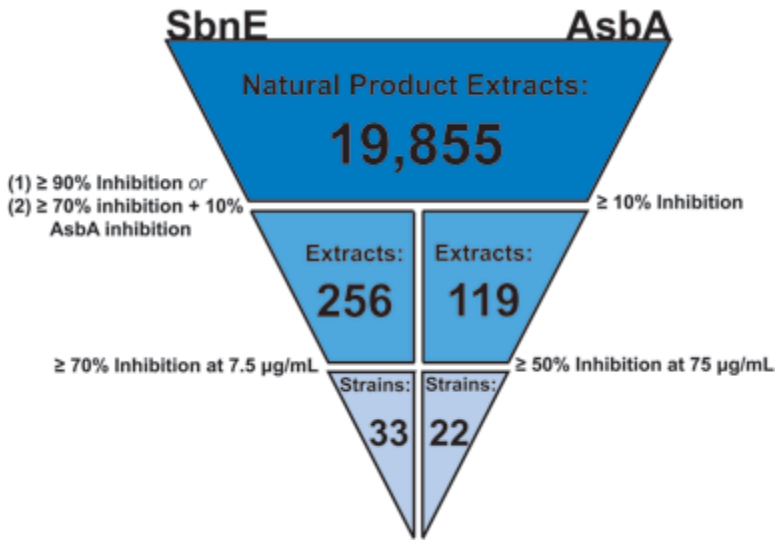


Figure 2-2. Triage of hits derived during high throughput screening of SbnE and AsbA against the full microbial NPE library.

2.2.2. Isolation and Structural Elucidation of the Baulamycins (6−7)

An iterative bioassay guided C18 fractionation (**Figure 2-11 in SI**), and subsequent RP-18 HPLC purification (**Figure 2-12 in SI**) yielded two novel bioactive molecules (**Figure 2-3**), baulamycins A (BmcA, compound **6**) and B (BmcB, compound **7**), whose structures are consistent with biogenesis from a type I polyketide synthase pathway. BmcA (compound **6**) was purified as a light-yellow amorphous solid and possesses a molecular formula of C H O as suggested by HRAPCIMS based on [M + H]+ 28 48 6 ion peak at m/z 481.3530 (**Figure 2-13 in SI**). The 1H (**Figure 2-14 in SI**) and 13C NMR (**Figure 2-15 in SI**) data, recorded in CD3OD indicated the polyketide nature of **6** and indicated the presence of at least three hydroxyl group bearing methines with chemical shifts at δ 4.47 (76.5), 4.01 (73.3), and 3.69 (72.5). Further analysis of the 1H NMR spectrum of 6 identified 12 aliphatic protons in the region of δH 0.95−2.77 and six methyl groups at δH 0.77 (d), 0.83 (d), 0.86 (d), 0.88 (d), 1.02 (t), and 1.06 (d). The 13C NMR and HSQCAD spectra (**Figure 2-16 in SI**) revealed the presence of four quaternary carbons at δC 148.6 (an aromatic carbon), 159.1 (two chemically equivalent aromatic carbons), and at δ 218.7 (a carbonyl carbon). The C gCOSY (**Figure 2-17 in SI**) correlations between two equivalent aromatic protons with a signal at δH 6.33, 6.15 and HMBCAD (**Figure 2-18 in SI**) correlation between δH 6.33, 6.15 to the equivalent carbons at δC 159.1, 105.9 and carbon at δC 148.6 clearly suggested the presence of resorcinol moiety (**Table 2-1**). Similarly, the connectivity from C-1 to C-17 was confirmed by an array of COSY and HMBC couplings in

both CD3OD and DMSO-d6 to construct a 17-carbon linear aliphatic chain with a characteristic carbonyl carbon at δC 218.7. In addition, COSY correlation was observed between H-1 and protons at C-2 along with their HMBC connection to C-3 suggesting an ethyl- ketone terminus for molecule **6**. Moreover, the COSY and HMBC correlations indicated the branching of an aliphatic chain through a methyl group at C-4 (δH 2.77, δC 44.6), C-6 (δH 1.42, δC 29.1), C-8 (δH 1.53, δC 30.9), and C-16 (δH 1.38, δC 26.6) positions. The position of the hydroxyl group at C-11 and C-13 were consistent with the distinctive 1H and 13C chemical shifts at δH 3.69, δC 72.5 and δH 4.01, δC 73.3, respectively. Furthermore, HMBC showed a correlation from H-1′ (δH 4.47) to C-2′ (δC 148.6) along with its contiguous COSY correlation to H-14 (δH 1.88) indicating a branch-point for the aliphatic chain at C-14 through a pendant attachment of 1′ (hydroxymethyl) resorcinol moiety to complete the structure of BmcA (**6**) (**Table 2-1**).

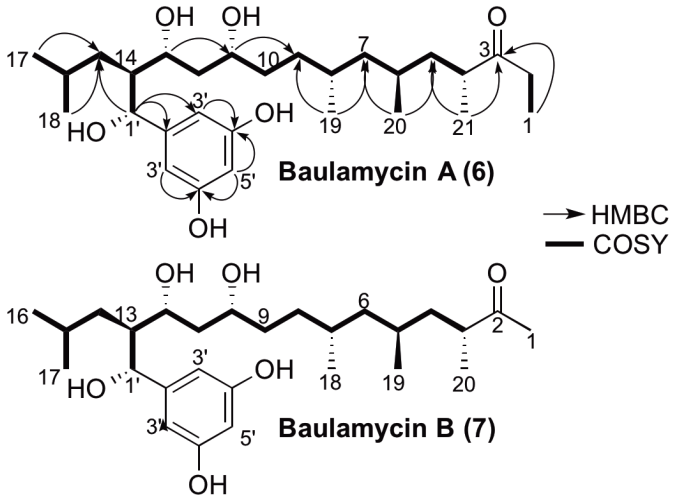


Figure 2-3. Planar structure of baulamycins A (6) and B (7) showing COSY correlation with bold bonds and HMBC correlations with arrow.

BmcA (**6**) consists of seven stereocenters with three hydroxyl-bearing carbons, three methyl-bearing carbons, and a 2-methylpropane-containing stereocenter. Initial attempts were made to obtain absolute stereochemical information through chemical manipulations for at least the hydroxyl-bearing chiral centers. However, the relatively low yield of baulamycins rendered this approach impractical. Therefore, a nondestructive J-based configuration analysis (JBCA)39 was employed to propose the relative configuration of BmcA (**6**).40−42 We calculated 3JH−H, 2JC−H, and 3JC−H values using a combination of phase-sensitive double quantum filtered (PS-

DQF)-COSY, homonuclear 2DJ spectroscopy (HOMO2DJ), gradient- and sensitivity-enhanced hetero (ω1) half-filtered TOCSY (HET- LOC-gse) and heteronuclear 2DJ (HET2DJ) spectral analyses (**Figures 2-19 to 2-21 in SI**).

	Baulamycin A (6)				Baulamycin B (7)	
	δ_C	δ_H, multi (*J* in Hz)	COSY	HMBC	δ_C	δ_H
1	35.1	2.57, dp	2"	1, 2", 2	27.4	2.15
		2.49, dp	2"	1, 2", 2		
2	7.8	1.02, t (6.9, 7.3)	1"	1, 1", 2		
3	218.7				214.6	
4	44.6	2.77, ddq (3.2, 6.7, 8.0)	3	1, 3, 19	43.9	2.70
5	41.7	1.73, ddd (3.2, 5.0, 11.0)	2	1, 2, 4	41.8	1.72,
		0.98, ddd (6.7, 8.0, 11.0)	4			0.97
6	29.1	1.42, ddddq (3.0, 3.2, 5.0, 13.0)	3, 5	6, 18	29.2	1.44
7	46.3	1.22, ddd (3.0, 11.0, 13.0)	6	17, 18	46.0	1.22
		0.95, ddd (3.2, 5.0, 11.0)	3	17, 18		0.95
8	30.9	1.53, ddddq (3.0, 5.0, 8.5, 13.0)	5, 7	5, 7, 17	31.0	1.55
9	33.0	1.33, dddd (3.0, 3.0, 10.6)	6, 8	9, 17	33.4	1.34
		1.19, dddd (8.5, 9.0, 10.6)	6, 8	7, 5		1.21
10	35.4	1.41, dddd (9.0, 11.0, 11.6)	7, 9	11, 17	35.1	1.40
		1.39, dddd (3.0, 4.0, 11.6)	7, 9			1.38
11	72.5	3.69, dddd (8.0, 11.0, 12.0)	8, 10	7, 8. 11	72.2	3.72
12	40.6	1.79, ddd (8.0, 7.2, 11.0)	9, 11		40.2	1.80,
		1.55, ddd (3.1, 3.2, 11.0)				1.50
13	73.3	4.01, dt (9.9, 3.4)	10, 12	1', 9, 13	73.0	3.99
14	48.5	1.88, ddq (7.7, 7.0)	11, 13	1', 11, 14	48.7	1.80
15	37.3	1.21, m	12, 14		37.0	1.20
16	26.6	1.38, m	13, 15		26.6	1.38
17	22.5	0.83, d (6.5)	14	13, 14	22.4	0.82
18	23.3	0.77, d (6.5)	14	13, 14	23.3	0.77
19	20.3	0.86, d (6.6)	6	5, 6, 7	20.1	0.85
20	20.5	0.88, d (6.5)	4	3, 4, 5	20.2	0.88
21	17.9	1.06, d (6.9)	2	1, 2, 3	18.0	1.01
1'	76.5	4.47, d (7.0)	12	2', 3', 8, 13, 11	76.4	4.50
2'	148.6				148.2	
3'	105.9	6.33, d (2.2)	5'	1', 2', 3', 4',	106.1	6.35
4'	159.1				158.5	
5'	101.9	6.15, t (2.2)	3'	3', 4'	101.8	6.16

Table 2-1. NMR spectroscopic data for baulamycins A (6) and B (7) in CD_3OD at 700 MHz.

The relative configurations of stereocenters (C11−C13− C14−C1′) in compound **6** established by JBCA are as follows. At the C1′ −C14 axis, large coupling constants 3JH1′ −H14 ≈ 7.2 Hz and 2JC′ −H14 ≈ 5.2 Hz inferred their anti and gauche orientations, respectively (64) (**Figure 2-4**). For the C14−C13 bond, the large 3JH14−H13 ≈ 7.7 Hz suggested anti orientation, which was also supported by the observed ROESY correlations (**Figure 2-21 in SI**). An additional large 2JC13−H14 ≈ 7.0 Hz was indicative of the gauche orientation between O-13 and H-1439 (**Figure 2-4**). Further, we sliced the HOMO2DJ spectrum at δH 4.01 (H-13), suggesting a large coupling constant between H13−H12a and a small value between H13−H12b, indicative of anti and gauche orientations, respectively (**Figure 2-4**). For the C12−C11 moiety, we sliced the HOMO2DJ spectrum at δH 3.69 (H-11) to yield 3JH12a−H11 ≈ 8.0 Hz and 3JH12a−H11 ≈ 3.2 Hz, suggesting anti and gauche orientation, respectively (**Figure 2-4**). Furthermore, large coupling constants were recorded for 2JC11−H12a ≈ 6.8 Hz and 2JC11−H12b ≈ 7.2 Hz, clearly indicating gauche orientations between O-11 and H-12a/H-12b39 (**Figure 2-4**). Moving down the axis over the C11−C10 moiety, large coupling constants between H-11/H-10a, H-11/H10b, and C11/H10a suggested gauche orientation between O-11 and H-10a. Hence, the relative configurations at C1′ , C14, C13, and C11 were proposed as 1′ R*, 14S*, 13R*, and 11R*, respectively (**Figure 2-3**).

Establishment of the relative orientation of protons associated with nonchiral C-10 and C-9 was necessary to measure the stereochemistry of subsequent chiral carbons (C8−C4). We sliced both HOMO2DJ and PS-DQF-COSY spectra at δH 1.41, 1.39 (H-10a, H-10b) and δH 1.19, 1.33 (H-10a, H-10b), respectively. The slicing provided large coupling constants between H-10a/H-9b and small between H-10a/H- 9a suggesting anti and gauche orientations, respectively (**Figure2-5**). Correspondingly, for the C9−C8 axis, calculations yielded 3JH9b−H8 ≈ 8.5 Hz (large) and 3JH9a−H8 ≈ 3.0 Hz (small) indicating anti and gauche conformations, respectively (**Figure 2-5**). Furthermore, the ROESY correlations and small coupling values obtained from 3JC7−H9a ≈ 2.0 Hz and 3JC7−H9b ≈ 2.1 Hz confirmed the gauche orientations between C7 and H-9a/H-9b facilitating assignment of the relative conformation of H-9a and H-9b at C9 (**Figure 2-5**).

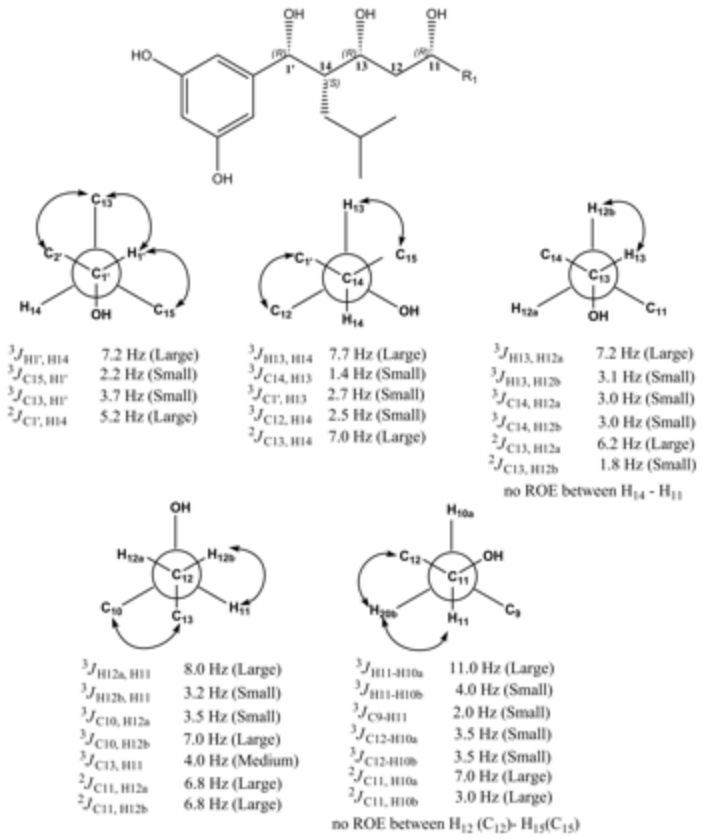


Figure 2-4: H-H, C-H coupling values and relative configuration determined for C1'-C11. Arrows showing ROESY correlations.

For the C9−C8 axis, the observation of large 3JH8−H9b (8.5 Hz) and 3JC19−H9a (8.0 Hz) values indicated that H-8 and H9b as well as C-19 and H-9a are in the anti orientation. The relative configuration at C8 was established to be R* according to the ROE between H-8 and H-9a (**Figure 2-5**). For the C8/C7 bond, the relationship of C19/H-7b was revealed as anti on the basis of a large 3JC19−H7b ≈ 9.0 Hz value. Another large coupling value (3JH7a - H8 ≈ 13.0 Hz) and corresponding small coupling value between C6 and H-8 (3JC6−H8 ≈ −2.4 Hz) led us to assign the relative conformation of H-7a and H-7b at C7 (**Figure 2-5**). For the C7/C6 bond, the observation of small coupling values (3JH7a - H6 ≈ 3.0 Hz and 3J7b - H6 ≈ 3.2 Hz) along with apparent ROE between H-6 and H-7a/H-7b suggested gauche orientation between H-6 and H-7a/H7b, establishing relative configuration at C6 as S* (**Figure 2-5**).

For C6 and C5, a large 3JH6−H5b value (~13.0 Hz) and small 3JH6−H5a value (~5.0 Hz) revealed that H-6/H-5b and H-6/H- 5a possess anti and gauche relationships, respectively. Additionally, small 3JC4−H6 ≈ 2.0 Hz and 3JC7−H5a ≈ −1.8 Hz values as well as observed ROE between H-5a and H-6 revealed the relative conformation of H-5a and H-5b at C5 (**Figure 2-5**). For the C5/C4 axis, large 3JH5b−H4 ≈ 8.0 Hz and small 3JH5a−H4 ≈ 3.2 Hz coupling values were obtained from HOMO2DJ and PS-DQF-COSY indicating anti and gauche conformations, respectively. In addition, an anti orientation of the C21 and H- 5a atoms was deduced from the large 3JC21−H15a ≈ 11.0 Hz value obtained from HETLOC-gse spectrum. Further, in the ROESY spectrum, ROEs between H-5a/H-4, H-4/H-6, and H-6/H-21 were observed. Taking these data into consideration, the relative configuration at C4 was proved to be R*. Therefore, the relative configuration for C10−C4 moiety was anticipated as 8R*, 6S* and 4R*, respectively, in BmcA (**6**).

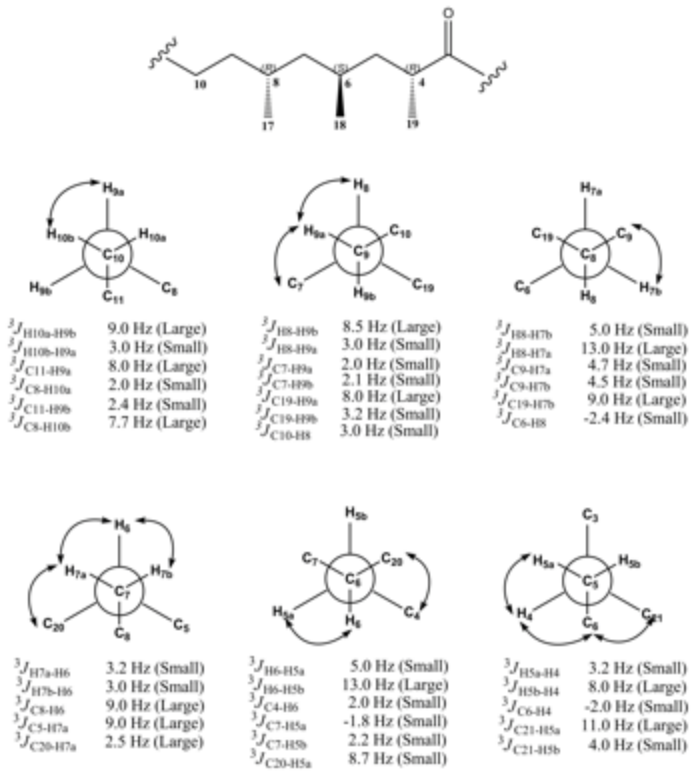


Figure 2-5: H-H, C-H coupling values and relative configuration determined for C10-C4. Arrows showing ROESY correlations.

BmcB (**7**) was isolated by RP-18-HPLC from the same C18 fraction containing compound **6**. The HRESIMS of the molecule provided a molecular formula of $C_{27}H_{46}O_6$ showing a [M + H]+ ion peak at 467.3171 (**Figure 2-26 in SI**). Moreover, baulamycin B (**7**) had high structural similarity to **6**, as evidenced by nearly identical 1H and 13C NMR chemical shifts when measured in CD3OD (**Table 2-1**). However, it displayed a subtle difference toward the carbonyl terminus, revealing the absence of terminal ethyl protons and substituting with a terminal methyl singlet at δH 2.18 in the 1H NMR spectrum of compound **7** (**Figure 2-27 in SI**). The relative stereochemistry from C1′ −C10 and C5−C3 were proposed to be identical to those of **6** on the basis of highly comparable NMR spectroscopy data (**Figures 2-27 to 2-31 in SI**).

2.2.3. Assessment of the *in vitro* Biological Activity of Compounds against Purified NIS Enzymes

The malachite green-based bioassay was next employed to assess the dose response of novel compounds BmcA (**6**) and BmcB (**7**) against purified SbnE and AsbA (**Table 2-2**). Both compounds exhibited *in vitro* bioactivity against SbnE, with IC_{50} values of 4.8 μM and 19 μM for BmcA and BmcB, respectively (**Figure 2-6A**). Inhibition was also observed against AsbA, with IC_{50} values of 180 μM and 200 μM for BmcA and BmcB, respectively (**Figure 2-6B**). The significant difference in apparent inhibition is likely due to the different enzyme concentrations used in the dose response assays, 25 nM for SbnE and 100 nM for AsbA (**Figure 2-8 in SI**). The assays were originally optimized at these enzyme concentrations to maximize the detection of extracts containing low levels of bioactive components during screening.

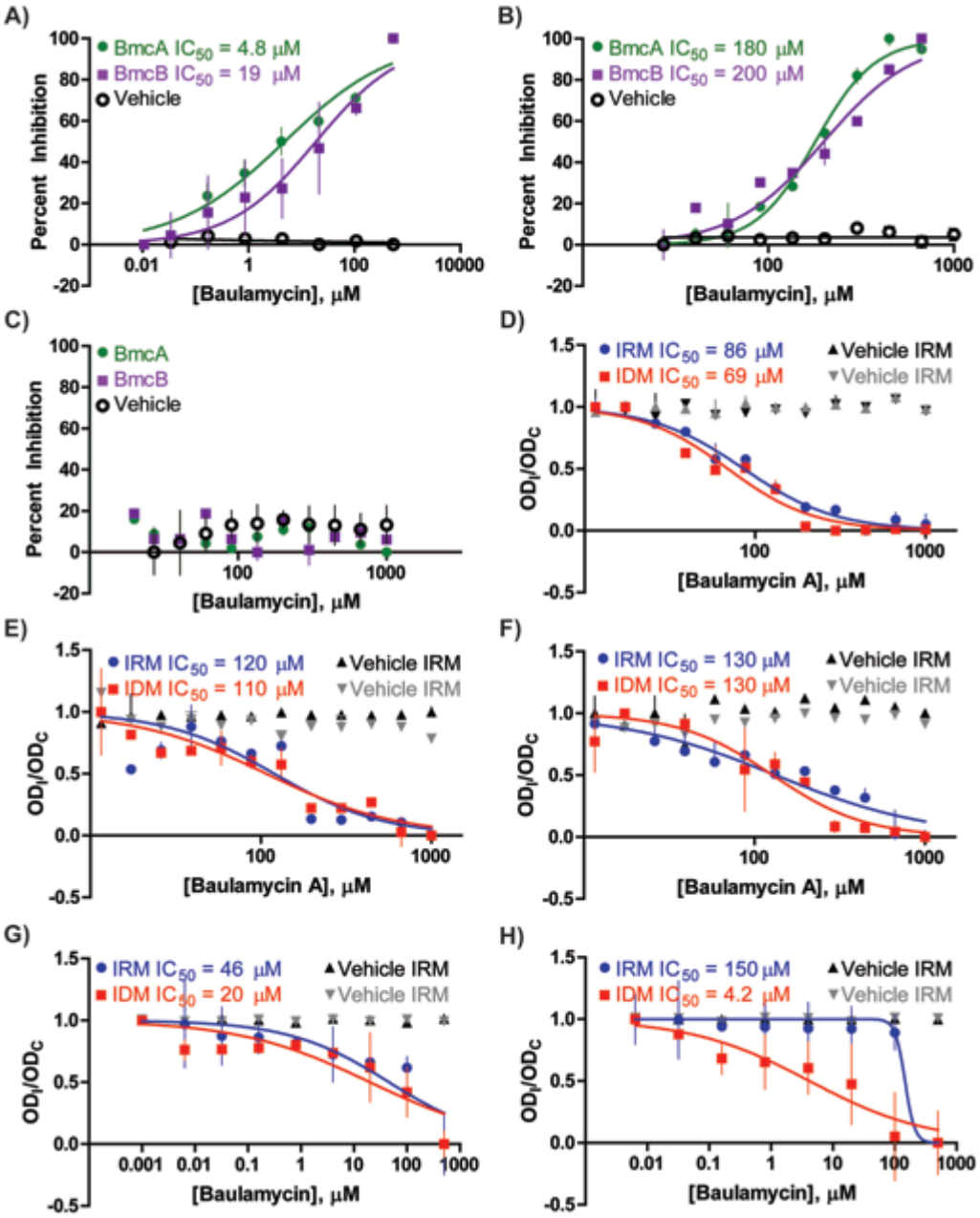


Figure 2-6. *In vitro* and live culture studies of baulamycin bioactivity.
(A−C) *In vitro* bioactivities against purified enzyme recorded for BmcA and BmcB against the NIS synthetase (A) SbnE, (B) AsbA, or (C) AsbB. (D−H) BmcA inhibition against live cultures of microbial strains (D) *S. aureus* (Newman), (E) B. *anthracis* (Sterne 34F2), (F) MRSA (USA 300), (G) *S. flexneri* (BS103), or (H) *E. coli* (MC 1061) in iron-depleted (IDM) or iron-rich (IRM) media. The y- axis represents a comparison of the optical densities of inhibitor- treated (ODI) and DMSO control-treated (ODC) cultures. Assays were conducted in duplicate due to the current limited availability of the baulamycins.

As a next step to address whether the baulamycins selectively inhibited type A over other NIS synthetase subfamilies, we explored the activity of both compounds against the petrobactin biosynthetic, type C NIS synthetase, AsbB. While type A NIS synthetases prefer citric acid as a substrate, type C enzymes generally utilize citric or succinic acid derivatives often found as complex intermediates within their respective siderophore biosynthetic pathways (59). In the petrobactin pathway, AsbB catalyzes condensation of a second molecule of spermidine with either N8-citryl-spermidine (compound **3**) or N1-(3,4- DHB)-N8-citryl spermidine (compound **4**) (58). Interestingly, recent work in our laboratory also demonstrates AsbB to be more flexible in substrate selection and capable of partial compensation for AsbA activity *in vitro* (58). Despite these findings, dose response studies revealed inhibition against AsbB to be negligible for both BmcA and BmcB (**Figure 2-6C**). These data indicate that both natural products are more potent *in vitro* against the type A NIS synthetases SbnE and AsbA than against the type C subfamily member AsbB.

Enzymatic Target	NIS Synthetase Classification	Associated Microbial Strain	Associated Siderophore	BmcA IC_{50} (μM)	BmcB IC_{50} (μM)
SbnE	Type A	*S. aureus*	Staphyloferrin B	4.8	19
AsbA	Type A	*B. anthracis*	Petrobactin	180	200
AsbB	Type C	*B. anthracis*	Petrobactin	> 1000	> 1000

Table 2-2. *In vitro* bioactivity analysis of the baulamycins against purified enzymes.
The IC_{50} values were determined through the NIS-tailored malachite green assay.

2.2.4. Initial Enzyme Mechanism of Inhibition Studies.

To investigate how the more potent of the two compounds, BmcA (**6**), inhibits SbnE and AsbA enzymatic activity, we monitored NIS synthetase ATP turnover using a previously described 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG) kinetic pyrophosphate detection assay (65, 66). These experiments were conducted by varying a single substrate and holding remaining reaction components at a constant saturating level in the presence and absence of inhibitor. The resulting double reciprocal (Lineweaver−Burk) plots confirmed that kinetic parameters closely matched previously reported values for both enzymes against varied citrate concentrations (**Table 2-4 in SI**) (48, 58). The plots suggested that BmcA inhibits both enzymes in a reversible, competitive manner with respect to citric acid, the corresponding polyamine, and ATP (**Figure 2-32, Table 2-4 in SI**). Although kinetic assays were only conducted with BmcA

due to its higher potency and availability (obtained in substantially higher yield from *S. tempisquensis*), it is reasonable to expect that BmcB also possesses these inhibition patterns.

A replot of slopes derived from the Lineweaver−Burk plot versus inhibitor concentration also enabled an initial investigation into the inhibition constants (Ki values) for BmcA against both enzymes (**Table 2-4 in SI**) (67). Interestingly, the inhibitor constants were lowest for citric acid (SbnE: 50 μM; AsbA: 110 μM) followed by the polyamine substrate (SbnE: 210 μM; AsbA: 170 μM) and ATP (SbnE: 680 μM; AsbA: 230 μM). This is surprising, given the polar nature of citric acid in comparison to that of the baulamycins. The explanation for this likely depends on the structure and substrate binding sites in the proteins. Although no structural studies have yet been conducted with SbnE or AsbA, the crystal structures of AsbB and the achromobactin type A NIS synthetase in *Pectobacterium chrysanthemi* have been determined (58, 61). Both reveal a single binding pocket that houses substrates and intermediates during catalysis. Elucidation of the differences in the structures of SbnE, AsbA, and these other NIS synthetases could shed light on the exact mechanism of inhibition exerted by the baulamycins and their complete lack of activity against AsbB. Efforts to acquire X-ray crystal structures of both SbnE and AsbA in the presence and absence of baulamycin compounds are expected in due course. Additionally, mutagenesis and selection of *S. tempisquensis* should improve the yield of the baulamycins, enabling a sufficient amount of these molecules for comprehensive kinetic analysis.

2.2.5 Assessment of Biological Activity against Microbial Cultures.

Since siderophore biosynthesis is required for bacterial survival in iron-limited environments, a selective siderophore synthetase inhibitor would be expected to limit growth only under these conditions. Therefore, we monitored bacterial growth in both iron-depleted (IDM) and iron-rich (IRM) media conditions in the presence of **6** and **7**. We also conducted live culture studies to demonstrate the ability of both natural products to penetrate bacteria, an established advantage of secondary metabolites over many synthetic chemicals (68). However, due to the extremely low yield of the baulamycins from *S. tempisquensis*, experiments could only be completed in duplicate, and the following studies represent an initial assessment of microbial culture bioactivity.

BmcA was found to inhibit growth of *S. aureus* (Newman) in both iron-depleted (IC_{50} = 69 μM) and iron-rich (IC_{50} = 86 μM) conditions (**Figure 2-6D, Table 2-3**). Similar inhibition was observed with B. anthracis (Sterne 34F2) under iron-depleted (IC_{50} = 110 μM) and iron-rich (IC_{50} = 120 μM) conditions (**Figure 2-6E, Table 2-3**), suggesting possible secondary targets in the cell. In agreement with *in vitro* findings against purified enzyme, BmcB was significantly less potent against *S. aureus* (Newman) and *B. anthracis* (Sterne 34F2) (**Figure 2-33 in SI**).

We next tested the potency of BmcA against a clinically isolated MRSA strain (USA 300). Similar to *S. aureus* (Newman) and *B. anthracis* (Sterne 34F2), BmcA also inhibited MRSA (USA 300) in iron-depleted (IC_{50} = 130 μM) and iron- rich (IC_{50} = 130 μM) conditions (**Figure 2-6F, Table 2-3**), again suggesting secondary cellular targets.

Microbial Strain	**Classification**	**Targeted Siderophore Pathway**	**Associated NIS Synthetase**	**BmcA IRM IC_{50} (μM)**	**BmcA IDM IC_{50} (μM)**
S. aureus (Newman)	Gram-positive	Staphyloferrin B	SbnE (A), SbnC (B), SbnF (C)	86	69
MRSA (USA 300)	Gram-positive	Staphyloferrin B	SbnE (A), SbnC (B), SbnF (C)	130	130
B. anthracis (Sterne $34F_2$)	Gram-positive	Petrobactin	AsbA (A), AsbB (C)	120	110
S. typhimurium	Gram-negative	Aerobactin	IucA (A), IucC (C)	> 1000	> 1000
E. coli (MC 1061)	Gram-negative	Aerobactin	IucA (A), IucC (C)	150	4.2
S. flexneri (BS103)	Gram-negative	Aerobactin	IucA (A), IucC (C)	46	20

Table 2-3. Bioactivity analysis of the baulamycins against Gram-positive and Gram-negative microbial cell cultures.
Dose response assays were conducted in iron-repleated (IRM) and iron-depleted (IDM) media.

Encouraged by the potency of these compounds on *S. aureus* and *B. anthracis* strains, we also tested inhibition of the more active BmcA on additional microorganisms possessing NIS synthetase siderophore biosynthetic pathways. Aerobactin was the first discovered siderophore to be assembled by an NIS synthetase system (59, 69). It contributes to the virulence of a multitude of bacteria, including *E. coli*, and multiple species of *Salmonella*, *Yersinia*, and *Shigella* (56, 70). The pathway involves the types A and C NIS synthetases IucA and IucC, respectively (71, 72), which share sequence similarity with NIS synthetases found in most other siderophore biosynthetic pathways (59, 60). Indeed, SbnE and AsbA appear to be relatives of IucA (59, 60). Furthermore, staphyloferrin B biosynthetic enzyme SbnF and petrobactin biosynthetic enzyme AsbB share sequence similarity with IucC (59, 60), collectively suggesting that the baulamycins

could also inhibit aerobactin-producing bacteria. Thus, we monitored the ability of BmcA to inhibit bacterial growth of the aerobactin-producing Gram-negative strains of *Shigella flexneri*, *E. coli*, and *Salmonella typhimurium* in iron- depleted and iron-rich conditions (**Figure 2-6G,H**). BmcA inhibited growth of *S. flexneri* and *E. coli* cultures in iron- depleted conditions (*S. flexneri*: IC_{50} = 20 μM; *E. coli*: IC_{50} = 4.2 μM) with lower potency in iron-rich conditions (*S. flexneri*: IC_{50} = 46 μM; *E. coli*: IC_{50} = 150 μM). Although increased potency of BmcA against *E. coli* in IDM suggests that this microbe could possess fewer nonspecific cellular drug targets, additional studies are required to draw any firm conclusions. While BmcA also appeared to inhibit growth of *S. typhimurium* cultures, the large standard deviations acquired during the assay led us to classify the IC_{50} as above 1 mM in both iron-depleted and -rich conditions (**Figure 2-33 in SI**). As improved access to baulamycins becomes possible, follow-up studies will enable more extensive analysis of its broad-spectrum antibiotic activity.

The current analysis provides an initial assessment of the ability of BmcA to inhibit growth and demonstrates that the molecule is cell permeable. However, inhibition of all target strains under both growth conditions suggests that BmcA is acting on some additional cellular targets. Indeed, SbnE and AsbA seem to share varying levels of sequence identity with a number of enzymes in *S. aureus* and *B. anthracis*, including aminoacyl-tRNA synthetases, CoA synthetases, and fatty-acid CoA ligases in addition to critical enzymes involved in cell envelope biogenesis, inhibition of apoptosis, and spore formation (**Figure 2-34 in SI**)

Finally, the ability of BmcA to inhibit bacterial growth of both Gram-positive and Gram-negative strains (**Table 2-3**) classifies it as a broad-spectrum antibiotic. Although BmcB was not included in these experiments due to low production by *Streptomyces tempisquensis*, the compound is of a comparable size and possesses similar functional groups, suggesting it also could be capable of inhibiting microbial growth and infiltrating the bacteria. On the other hand, the decreased chain length and potency of BmcB suggests that longer terminal carbon chains may be important in future structure−activity relationship (SAR) studies.

2.3. Conclusion

We have identified a new structural class of broad-spectrum antibiotics isolated from a marine microbial-derived NPE library. BmcA and BmcB inhibit siderophore biosynthesis, a crucial

virulence factor associated with iron sequestration in pathogenic bacteria. Although previous studies have validated siderophore biosynthesis as an effective synthetic drug target (37-40), the aforementioned study is the first to demonstrate its potential as a target for discovering novel chemical scaffolds. Structurally unique drugs like the baulamycins present hurdles for the bacterial development of antibiotic resistance. High- throughput screening for natural product inhibitors of virulence factors, including those for siderophore biosynthesis, could provide a fresh arsenal of chemical scaffolds to combat drug- resistant pathogens.

The baulamycin natural products represent promising lead structures that can be further manipulated to improve both potency and target selectivity if required. We have illustrated that the compounds are efficacious *in vitro* and are capable of penetrating bacterial barriers to inhibit growth of cultures, including *B. anthracis*, MRSA, *S. flexneri*, and *E. coli*. The ability of the drug to inhibit growth of both Gram-positive and Gram- negative bacteria suggests its potential use as a broad-spectrum antibiotic. These growth inhibition studies further suggest that the compounds possess multiple targets in the microbial cell. Given the simultaneous decline in antibiotic drug discovery and increase of multidrug resistant bacteria, the baulamycins may represent an auspicious starting-point for expanding discovery efforts against significant human pathogens.

2.4. Supplemental Information

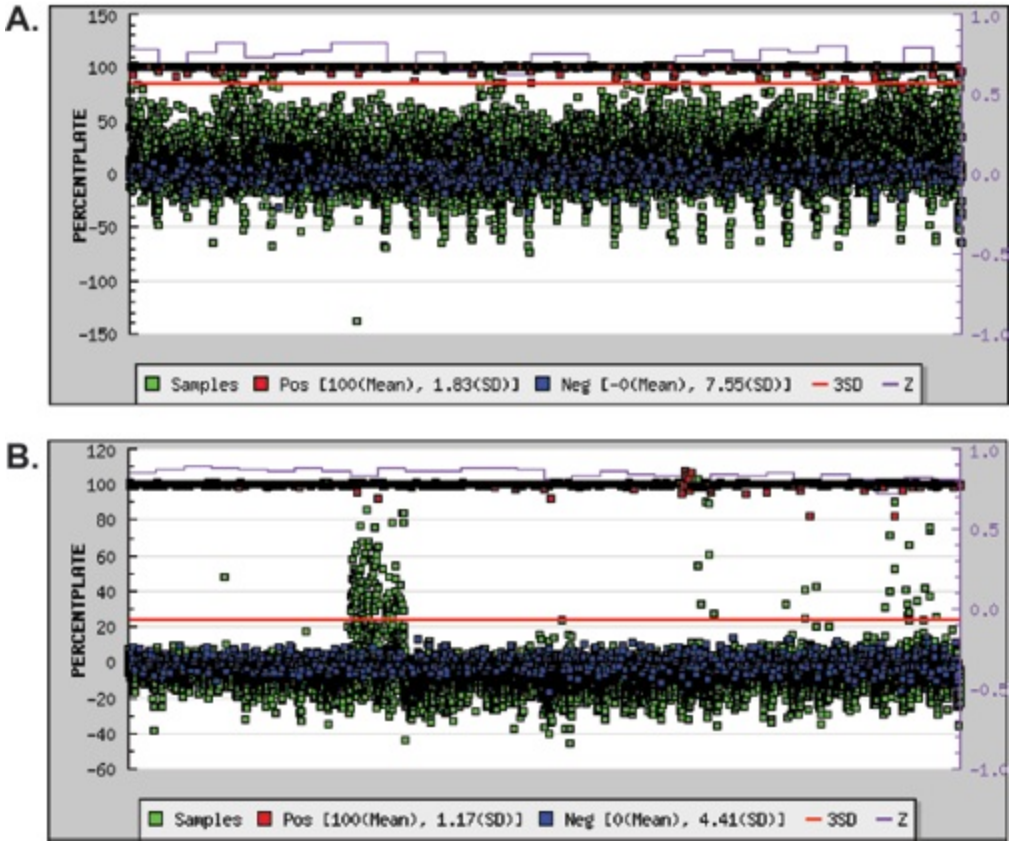


Figure 2-7: Secondary HTS analysis on active NPEs from primary screening.
(a) SbnE and (b) AsbA. Green squares represent percent activity in the presence of different NPEs, red squares percent activity in the absence of enzyme target, and blue squares percent activity with a DMSO vehicle.

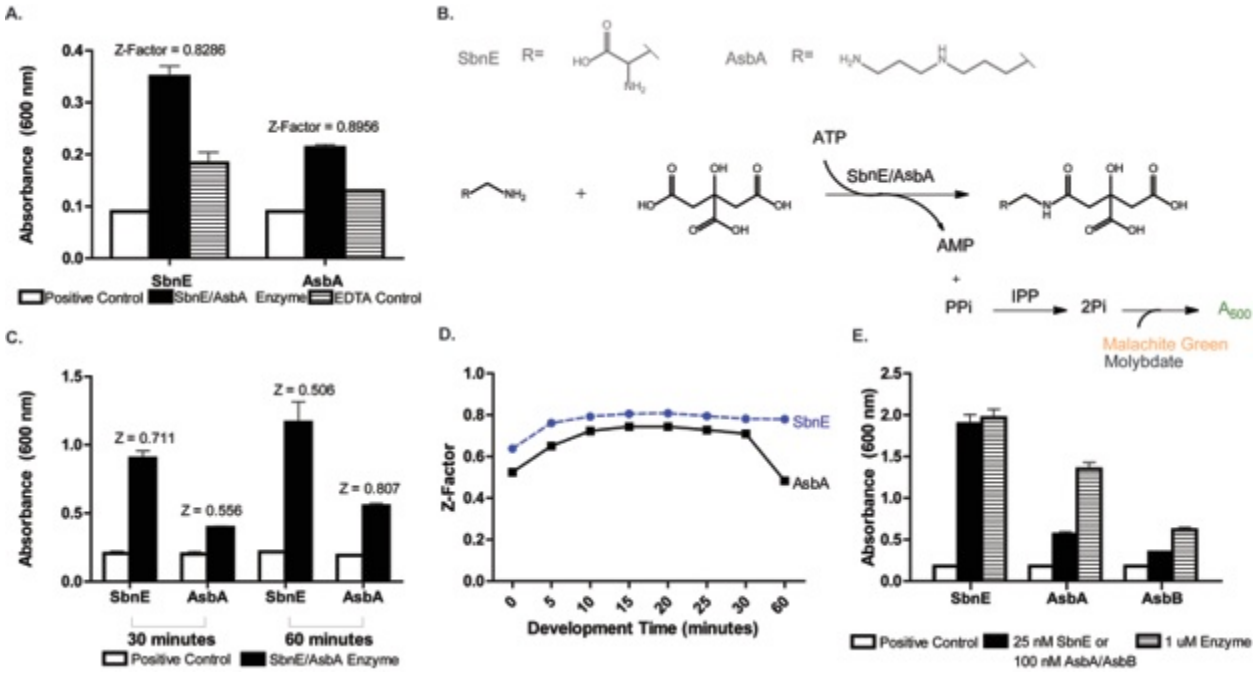


Figure 2-8. Development of an NIS synthetase assay for screening.
(a) Testing of the fitness of the NIS hydroxymate-formation assay for HTS. (b) The malachite green assay tailored to NIS synthetases SbnE and AsbA. (c) Optimization of the incubation time for SbnE and AsbA reactions. (d) Optimization of the development time for SbnE and AsbA after quenching with the malachite green solution. (e)Optimization of enzyme concentration for the high throughput assays.

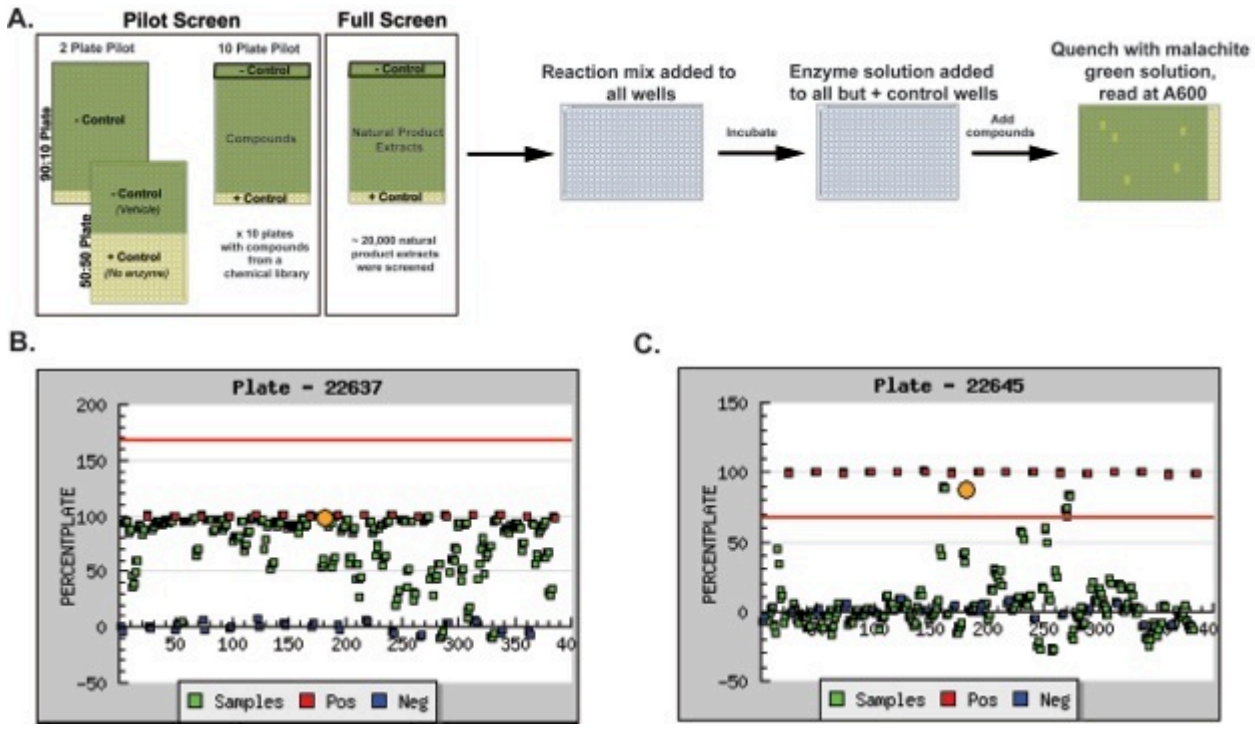


Figure 2-9. Overview of HTS and Initial Hits.
(a) Overview of the pilot and full HTS screen against the NPE library. (b) A scatterplot generated with MScreen depicting a 384 well plate from the full library screen against SbnE or (c) AsbA. The orange dot represents the well containing the active NPE derived from strain *Streptomyces tempisquensis*, the most intriguing hit from the full screen.

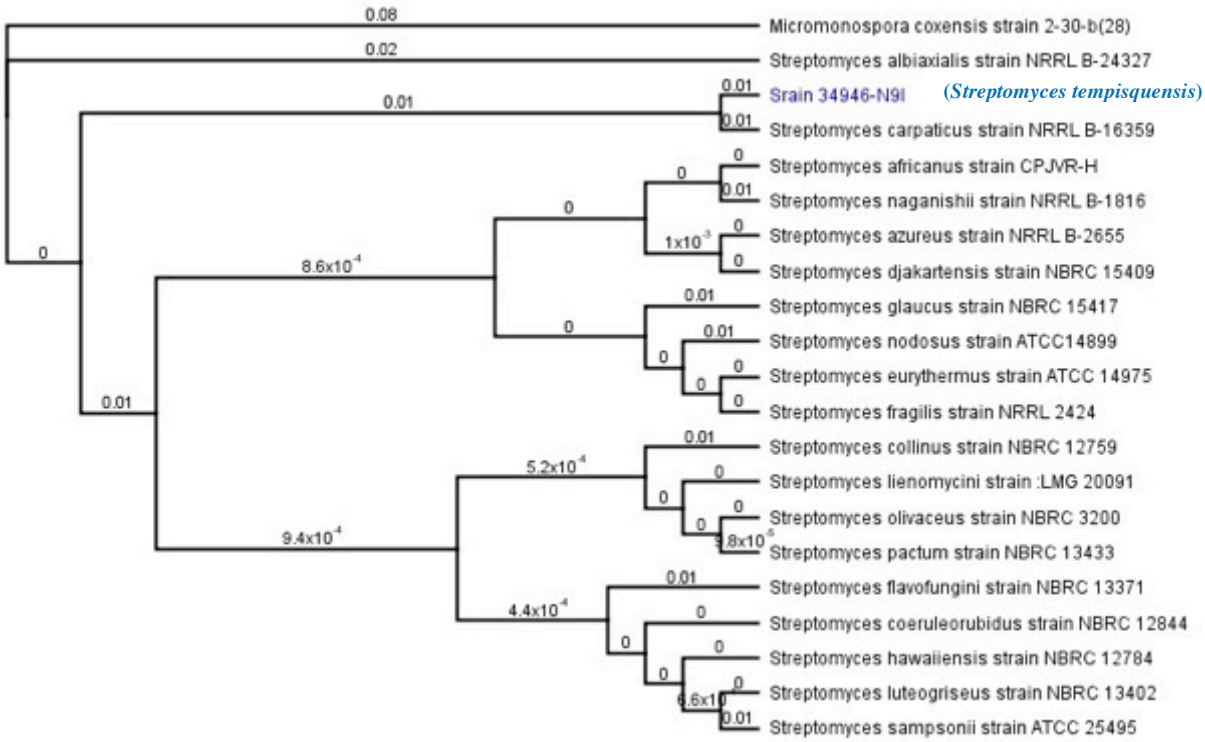
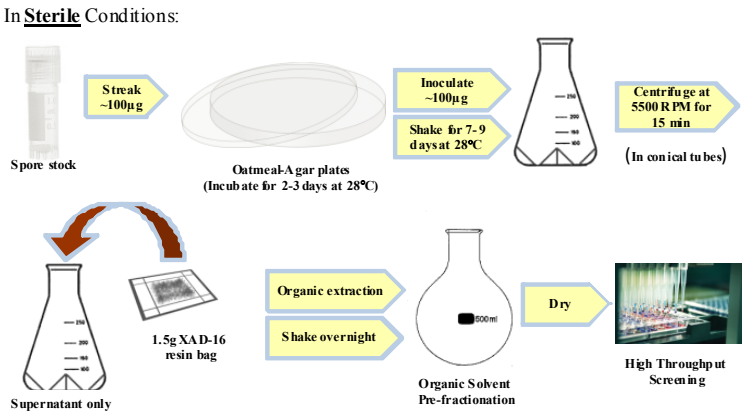


Figure 2-10: 16S rRNA analysis and alignment.
The 16S sequence has been given the GenBank accession number KF954543.

Small Scale Extraction Scheme



Large Scale Extraction Scheme

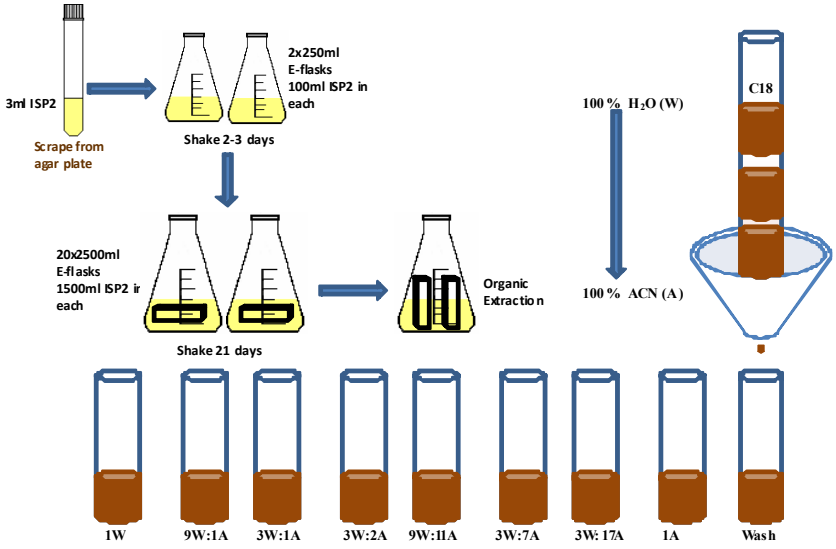


Figure 2-11: Small and large scale extraction schemes.

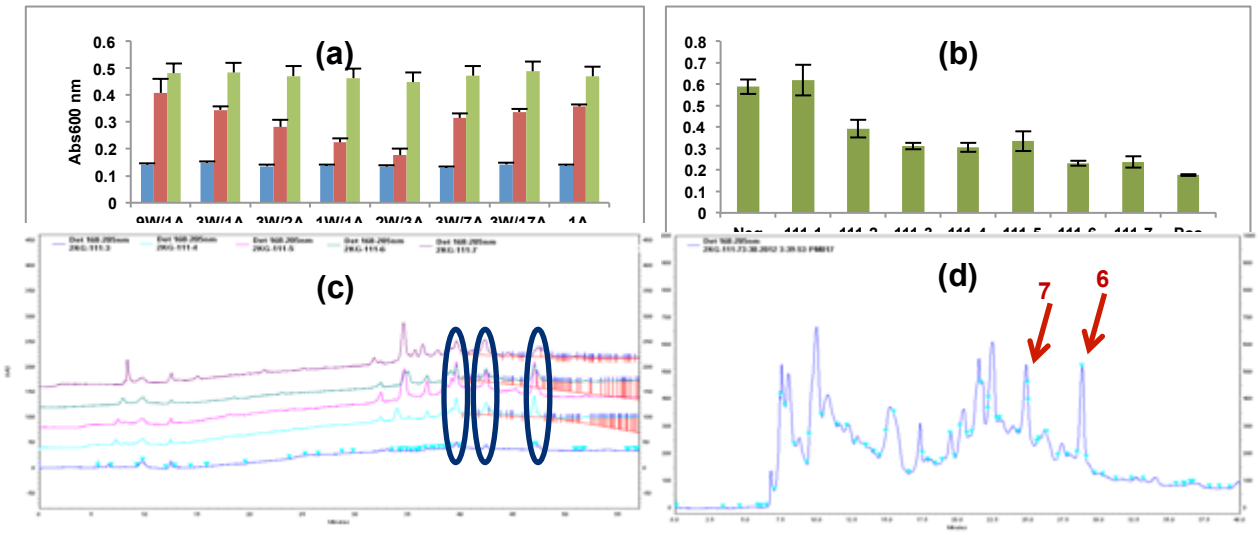


Figure 2-12. Bioactivity Guided De-replication of NPE.
Bioactivity Guided De-replication of NPE *Streptomyces tempisquensis In vitro* activities of C18 flash column fractions from and organic extract of XAD-16 bound culture extract from strain *Streptomyces tempisquensis* [Blue: positive control (no enzyme), Red: *Streptomyces tempisquensis*, Green: negative control (no inhibitor)]. **A.)** *In vitro* activities of size exclusion fractions originated from active 1W/1A and 2W/3A C18 fractions. **B.)** HPLC probe on active size exclusion (SE) fractions for possible bioactive molecules (shown under circle). **C.)** Reverse phase HPLC re-purification of pooled bioactive SE fractions (bioactive peaks shown by arrows)

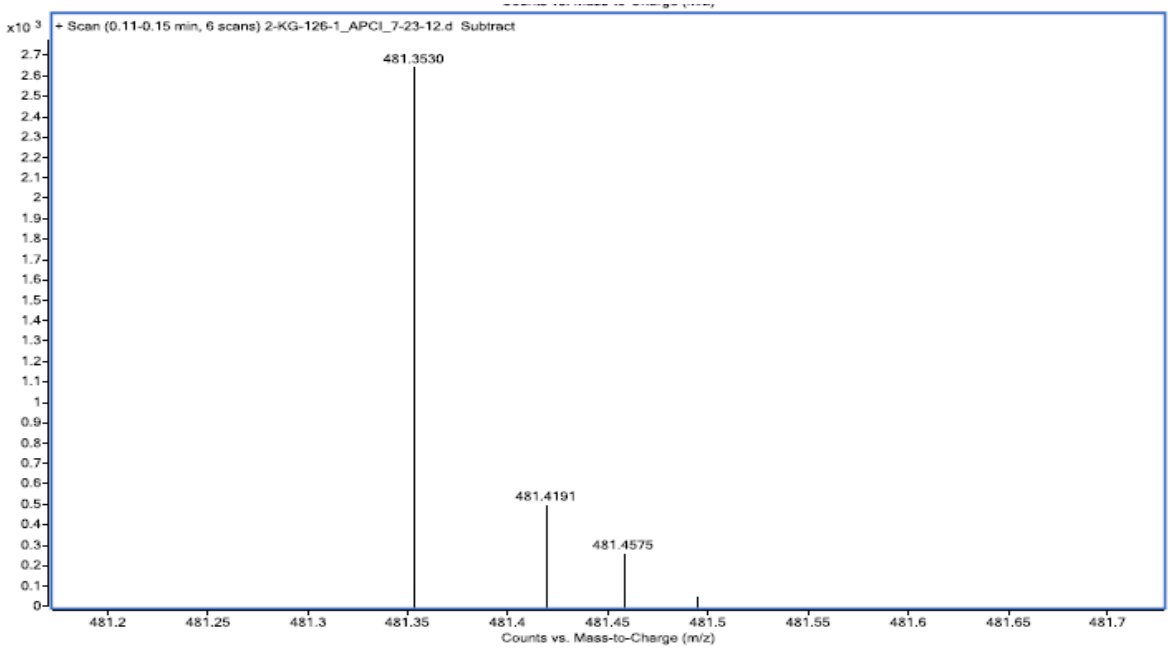


Figure 2-13: HRAPCIMS spectrum for baulamycin A (6).

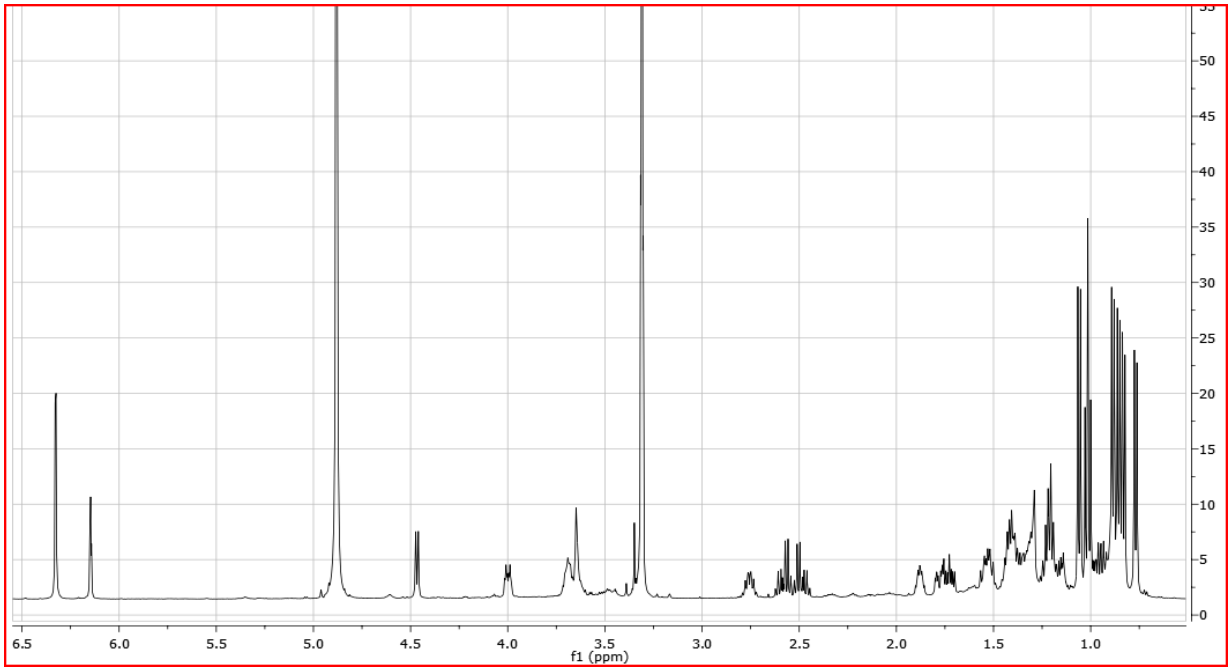


Figure 2-14: ^{1}H NMR spectrum of baulamycin A (6) recorded at 700 MHz (in CD_3OD).

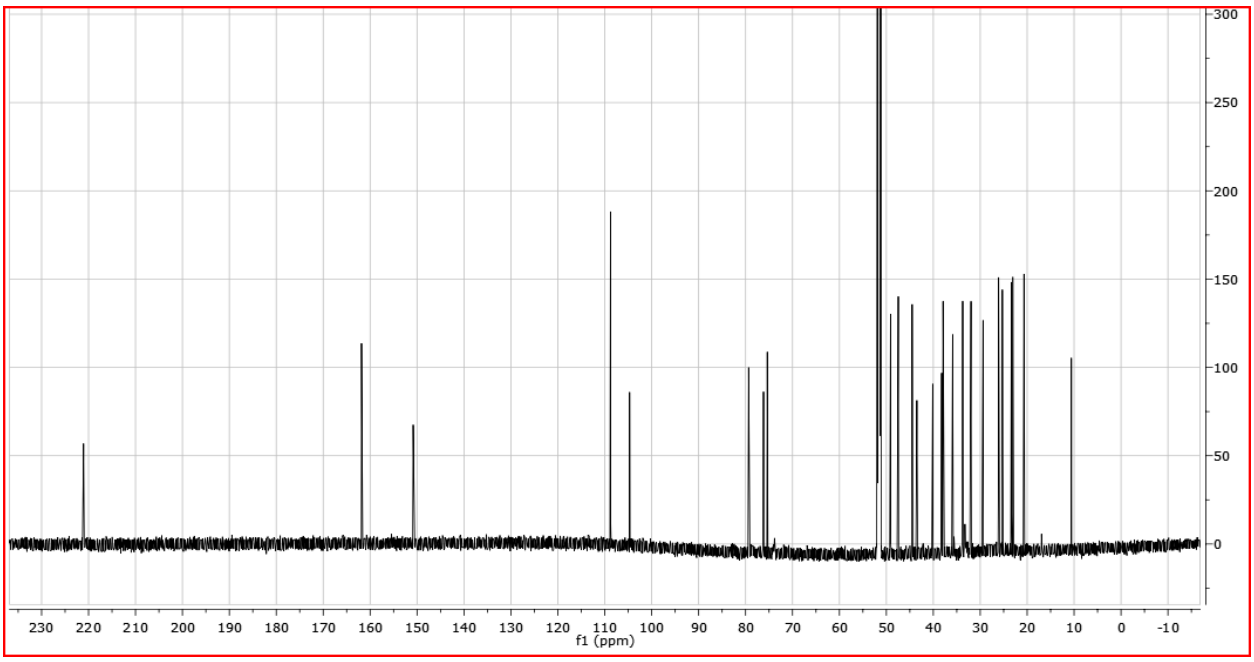


Figure 2-15: ^{13}C NMR spectrum of baulamycin A (6) recorded at 176 MHz. (in CD_3OD).

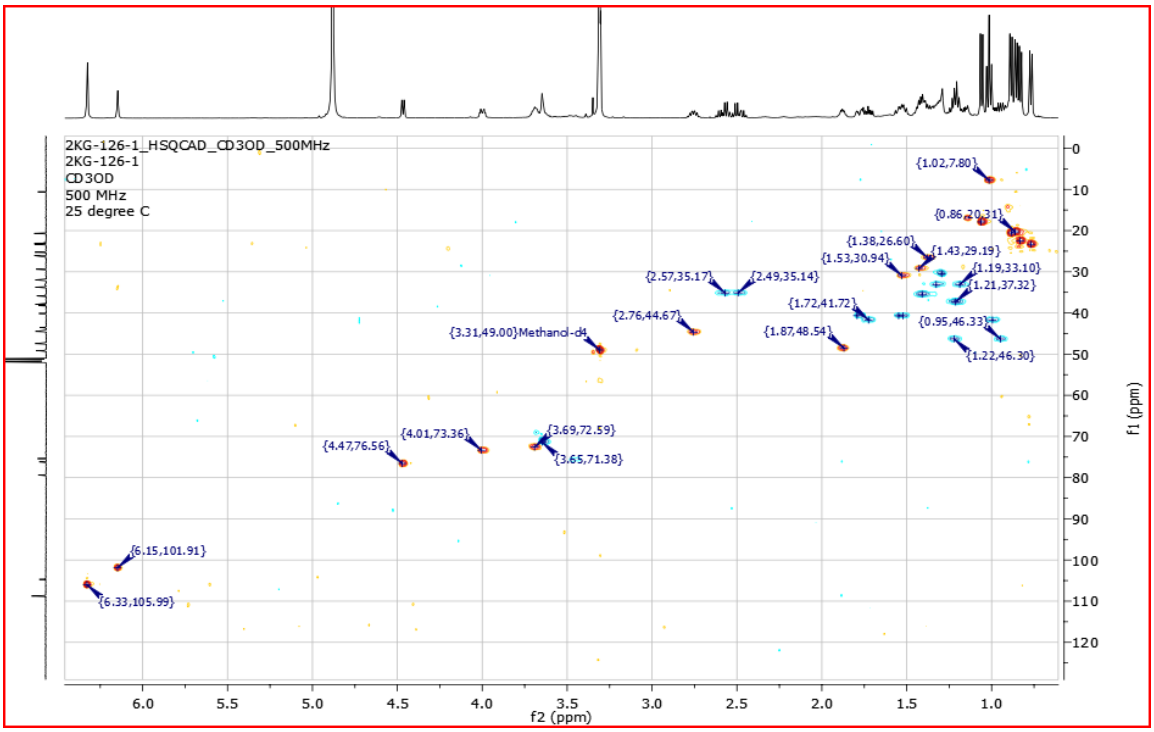


Figure 2-16: HSQCAD NMR spectrum of baulamycin A (6) recorded at 500 MHz (in CD_3OD).

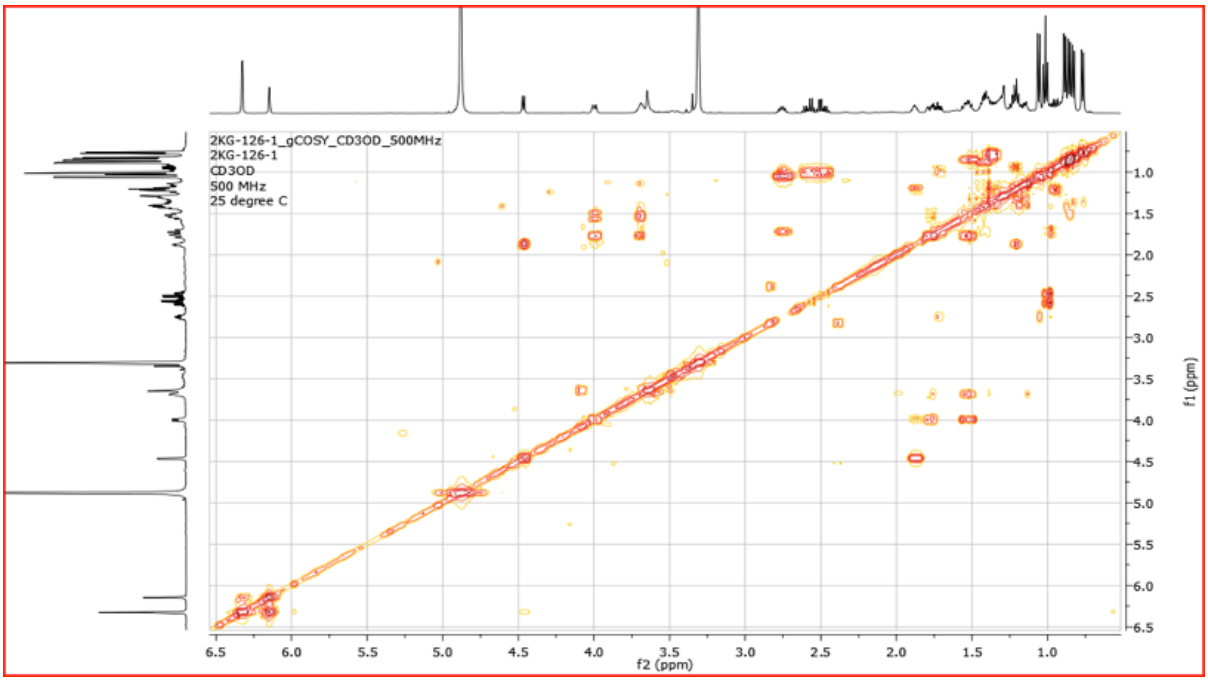


Figure 2-17: gCOSY NMR spectrum of baulamycin A (6) recorded at 500 MHz (in CD_3OD).

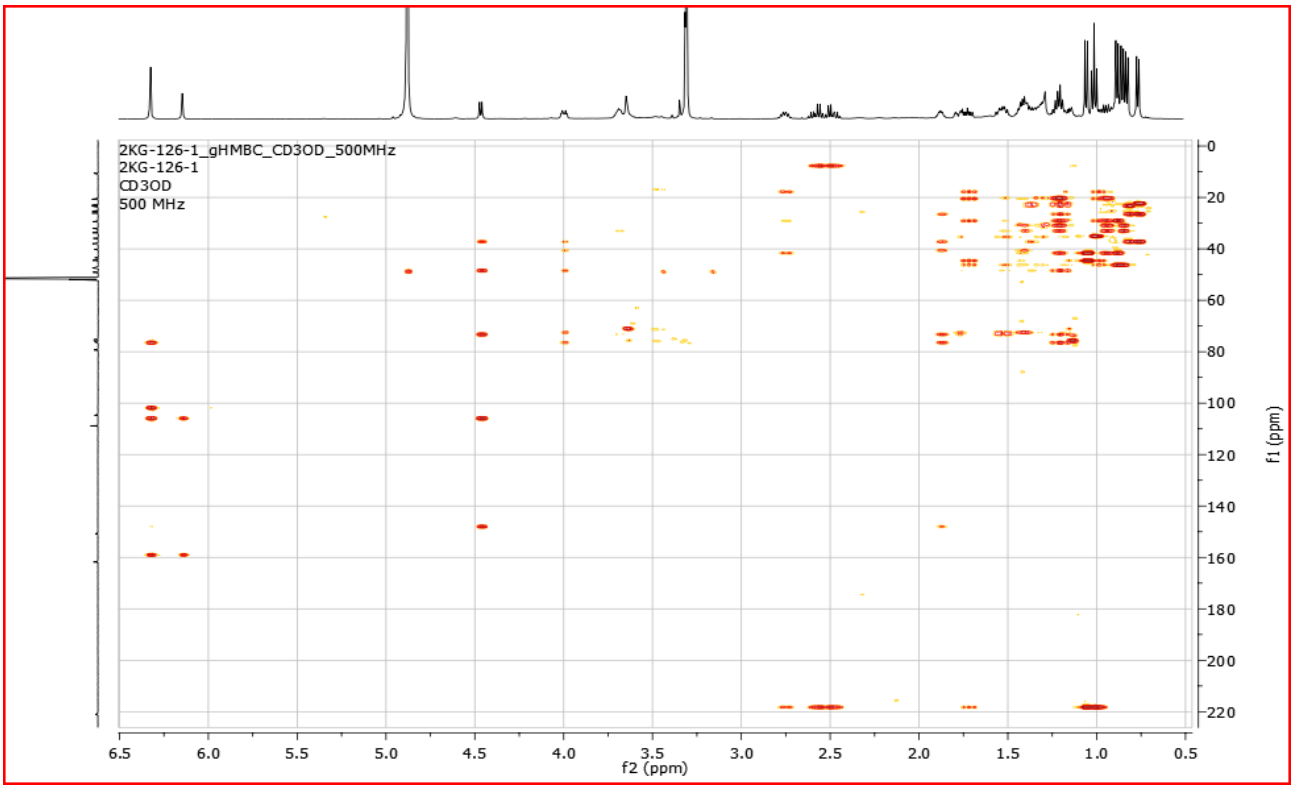


Figure 2-18: gHMBC NMR spectrum of baulamycin A (6) recorded at 500 MHz (in CD_3OD).

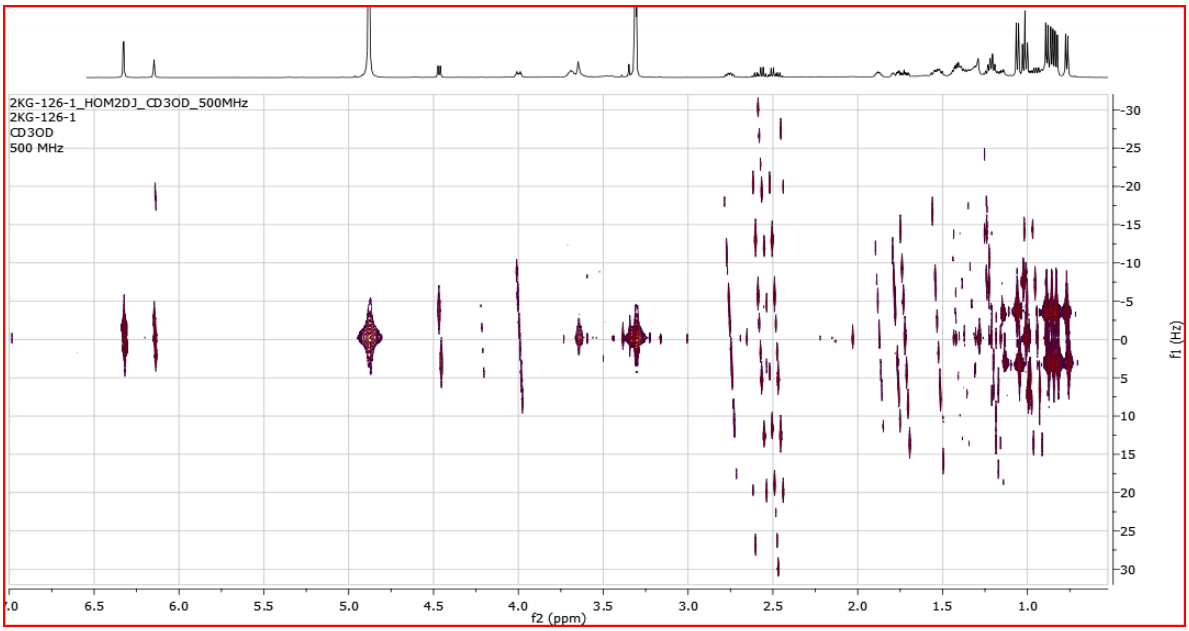


Figure 2-19: HOMO2DJ NMR spectrum of baulamycin A (6) recorded at 500 MHz (in CD_3OD).

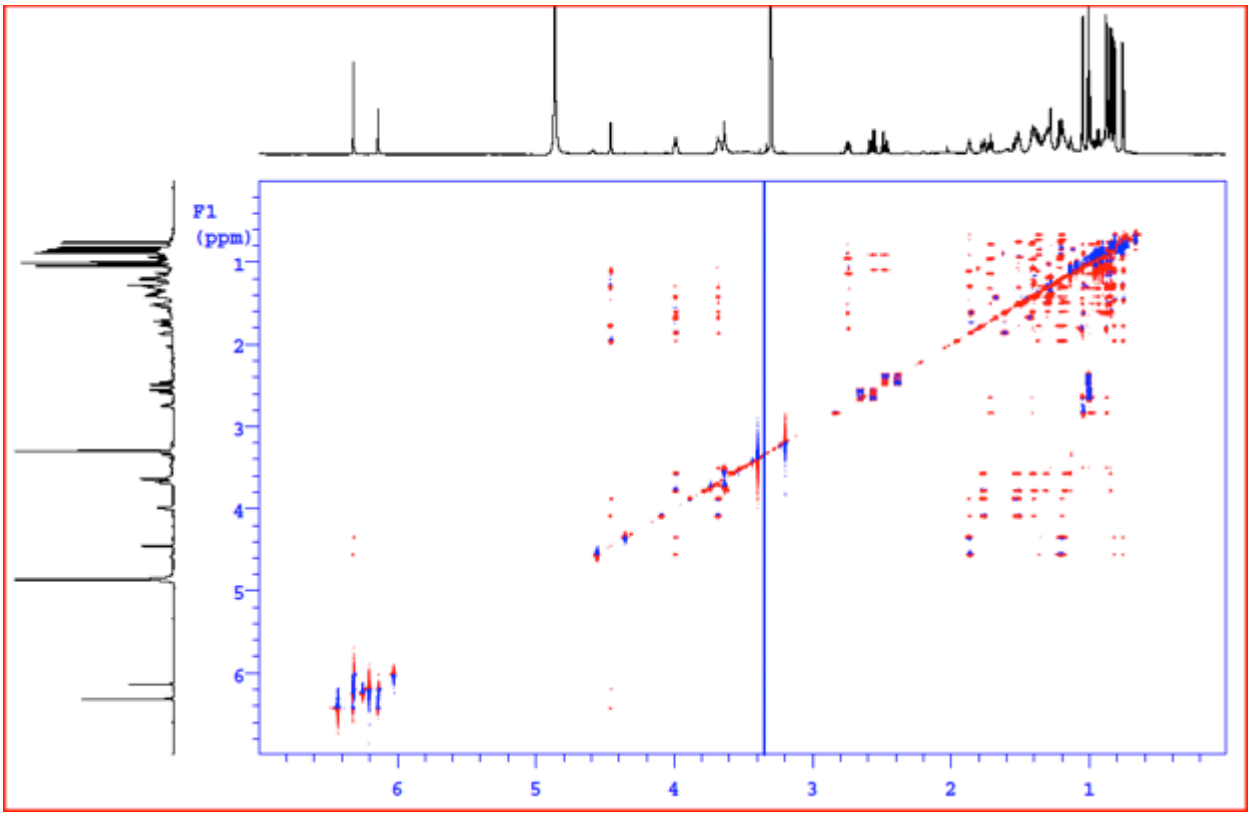


Figure 2-20: HETLOC-gse NMR spectrum of baulamycin A (6) recorded at 500 MHz (in CD_3OD).

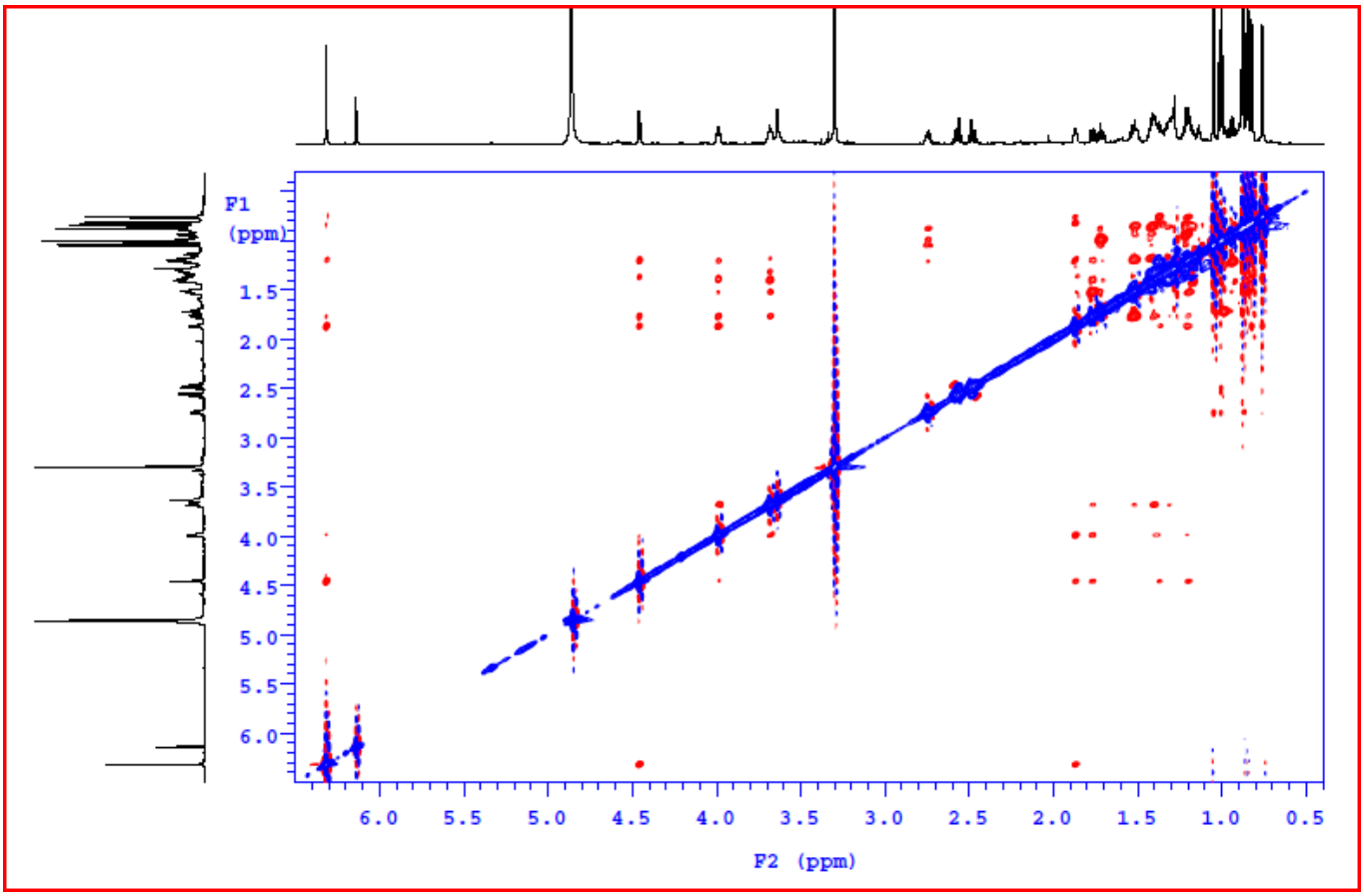


Figure 2-21: ROESYAD NMR spectrum of baulamycin A (6) recorded at 500 MHz (in CD_3OD).

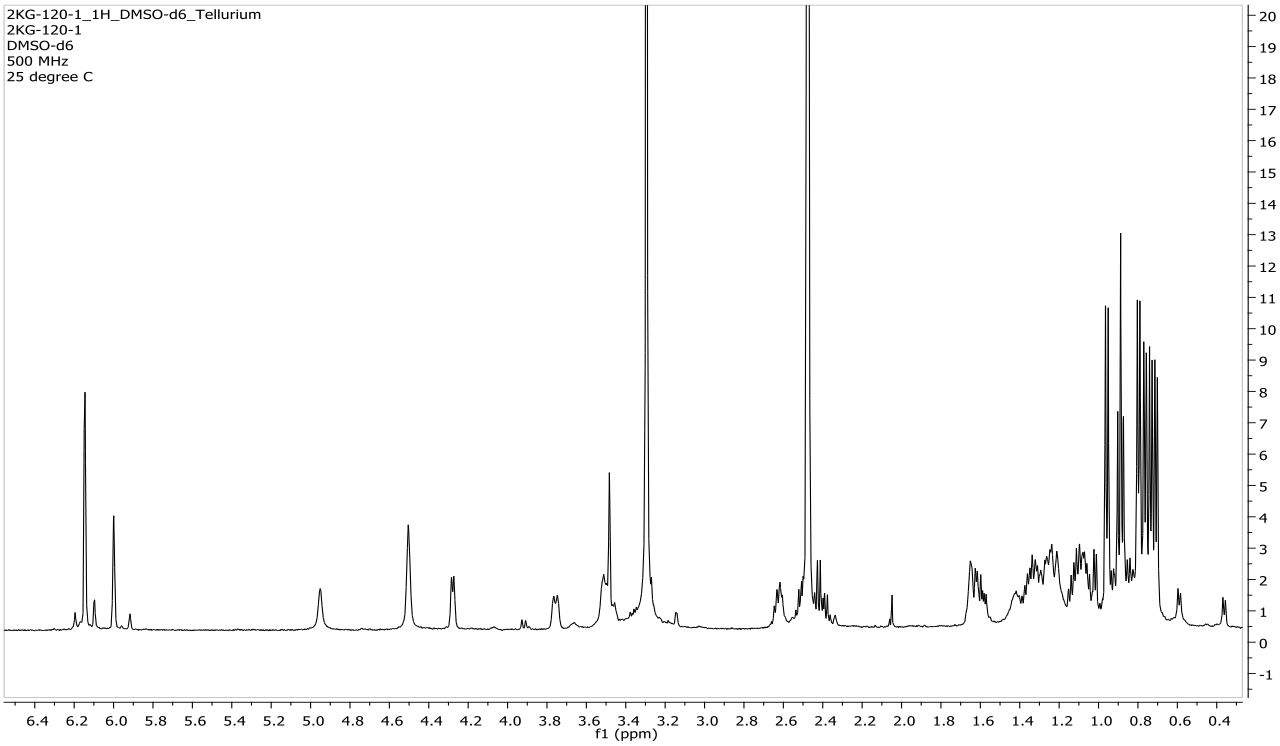


Figure 2-22: ^{1}H NMR spectrum of baulamycin A (6) recorded at 500 MHz (in DMSO-d6).

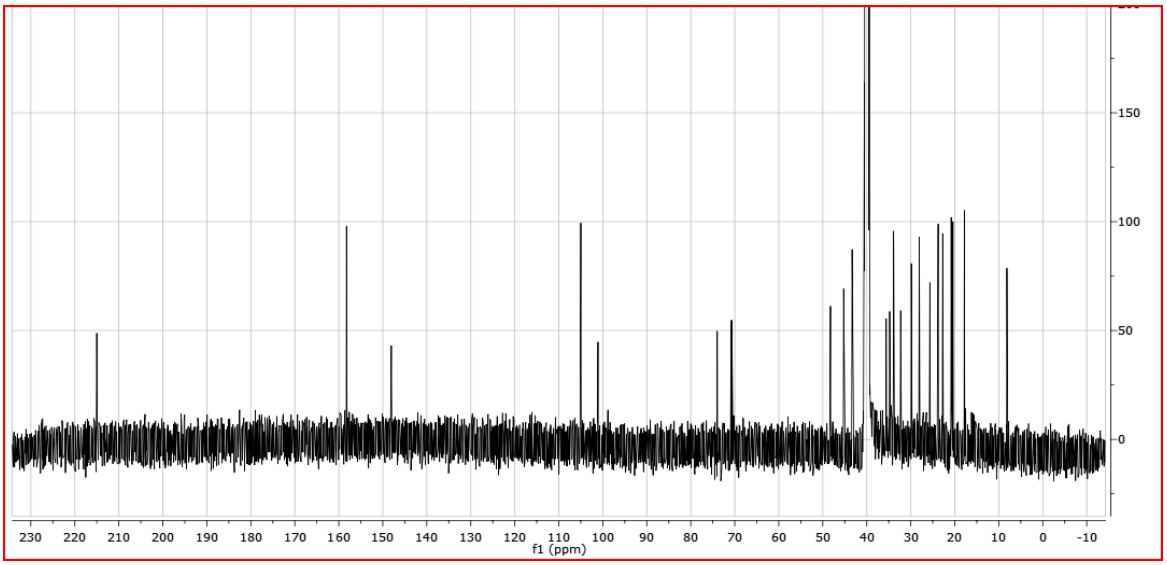


Figure 2-23: ^{13}C NMR spectrum of baulamycin A (6) recorded at 125 MHz (in DMSO-d6).

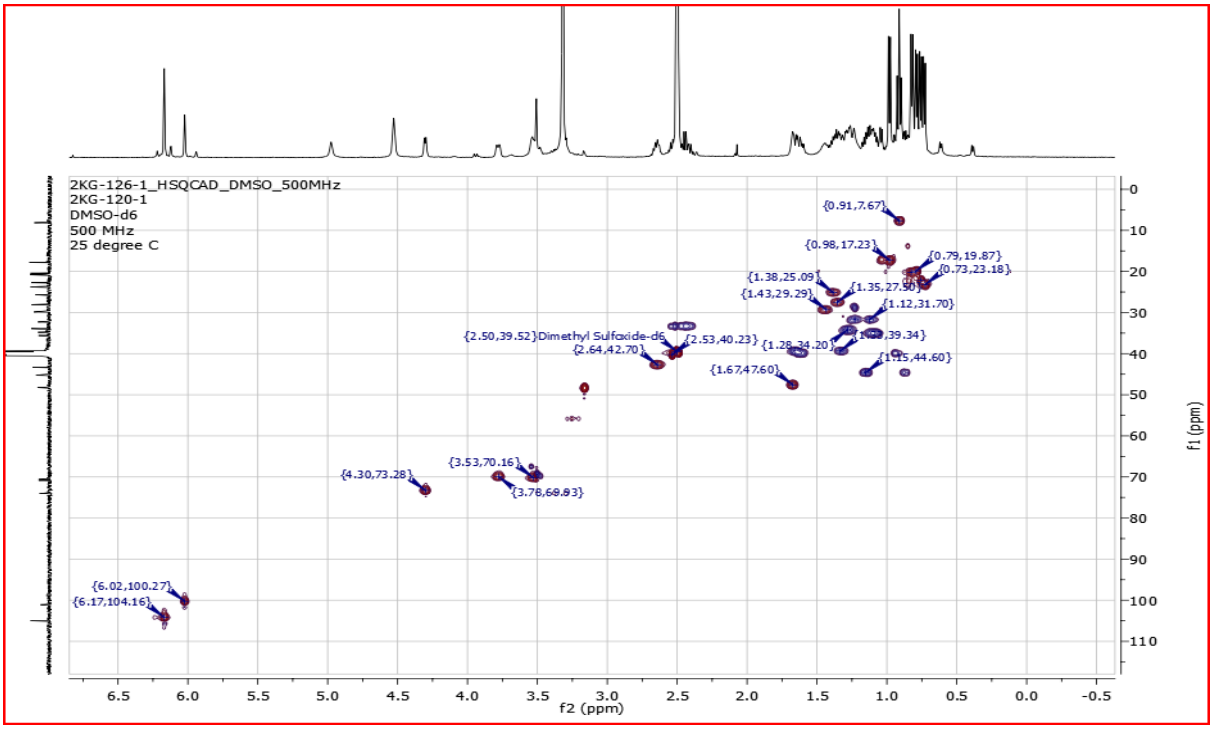


Figure 2-24: HSQCAD NMR spectrum of baulamycin A (6) recorded at 500 MHz (in DMSO-d6).

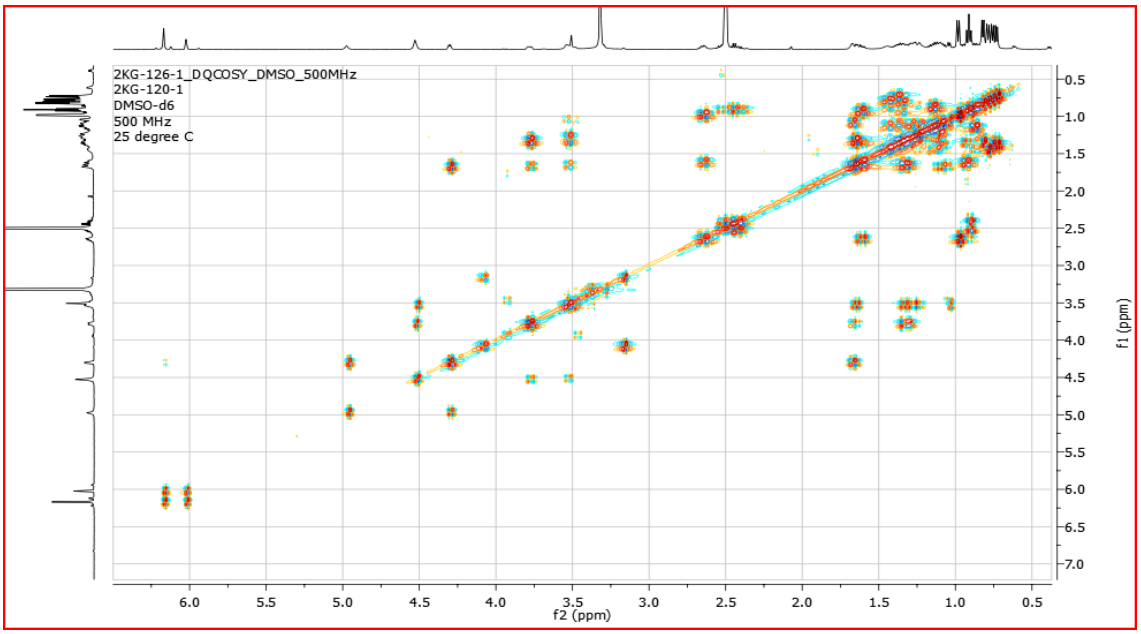


Figure 2-25: PS-DQF-COSY NMR spectrum of baulamycin A (6) recorded at 500 MHz (in DMSO-d6).

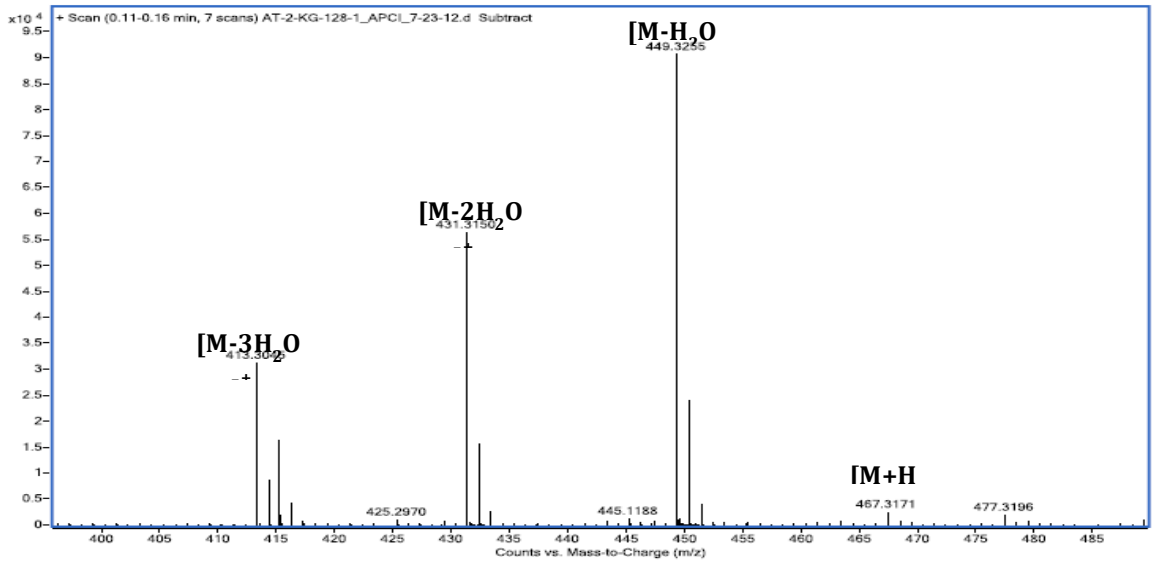


Figure 2-26: HRAPCIMS spectra for baulamycin B (7).

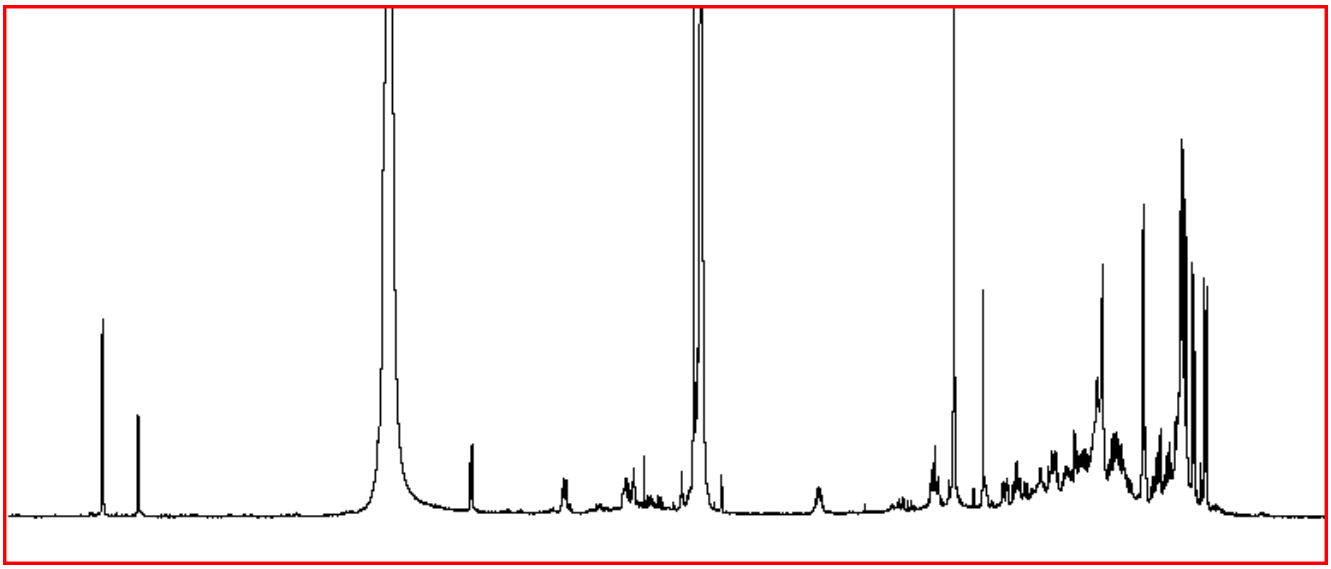


Figure 2-27: ^{1}H NMR spectrum of baulamycin B (7) recorded at 700 MHz (in CD_3OD).

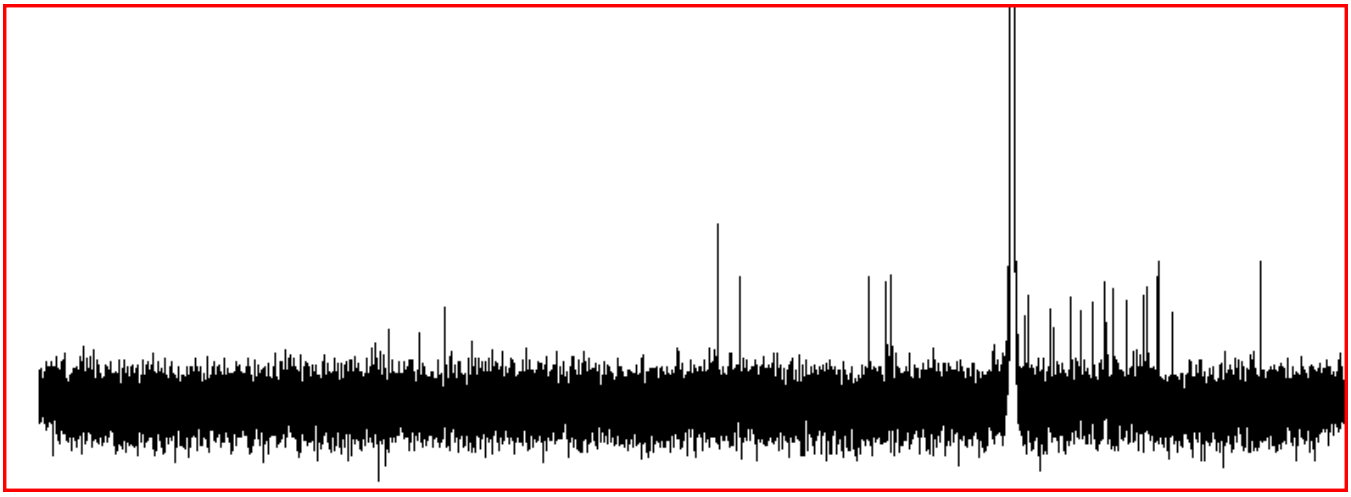


Figure 2-28: ^{13}C NMR spectrum of baulamycin B (7) recorded at 700 MHz (in CD_3OD).

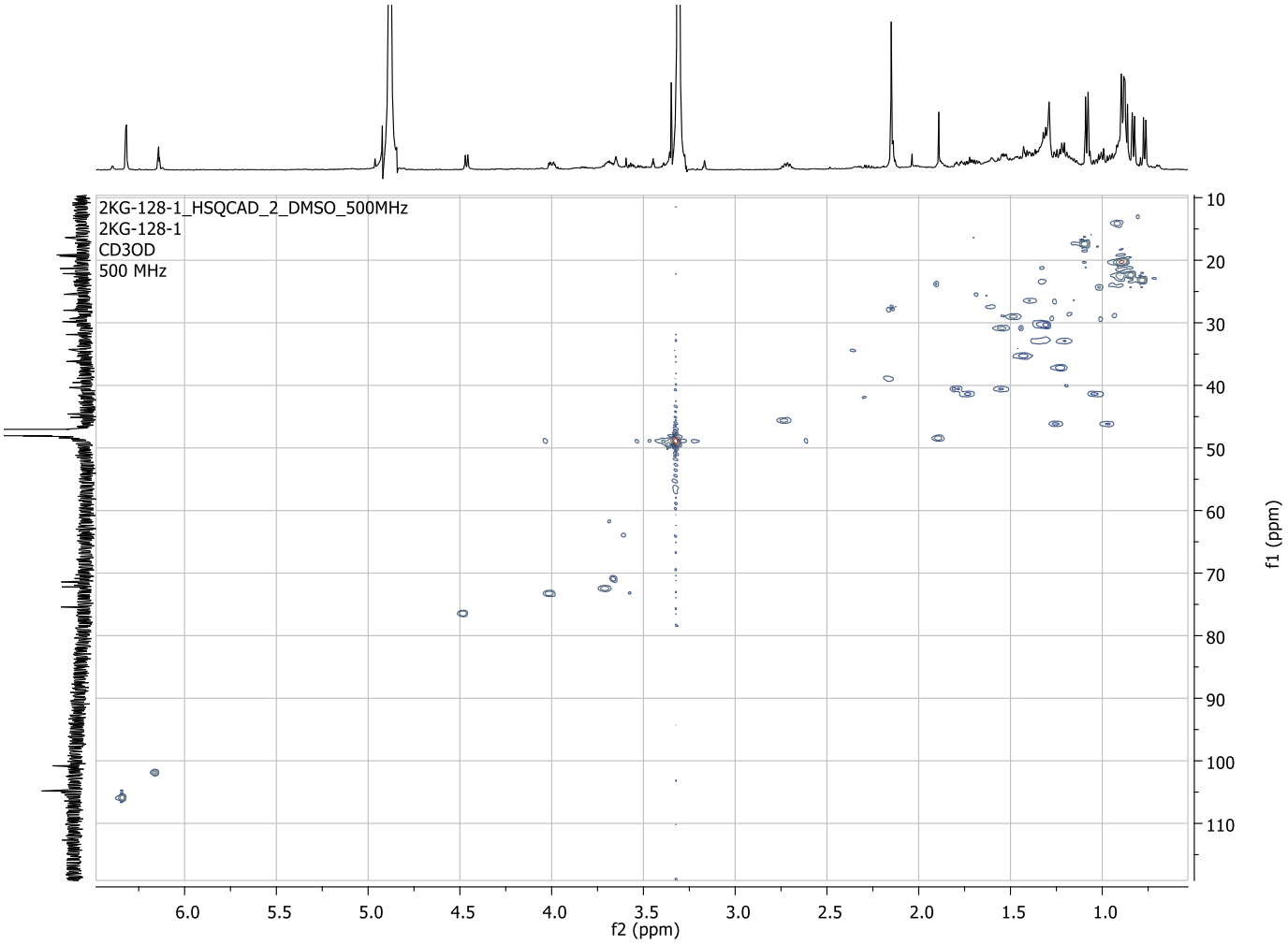


Figure 2-29: gHSQCAD NMR spectrum of baulamycin B (7) recorded at 700 MHz (in CD_3OD).

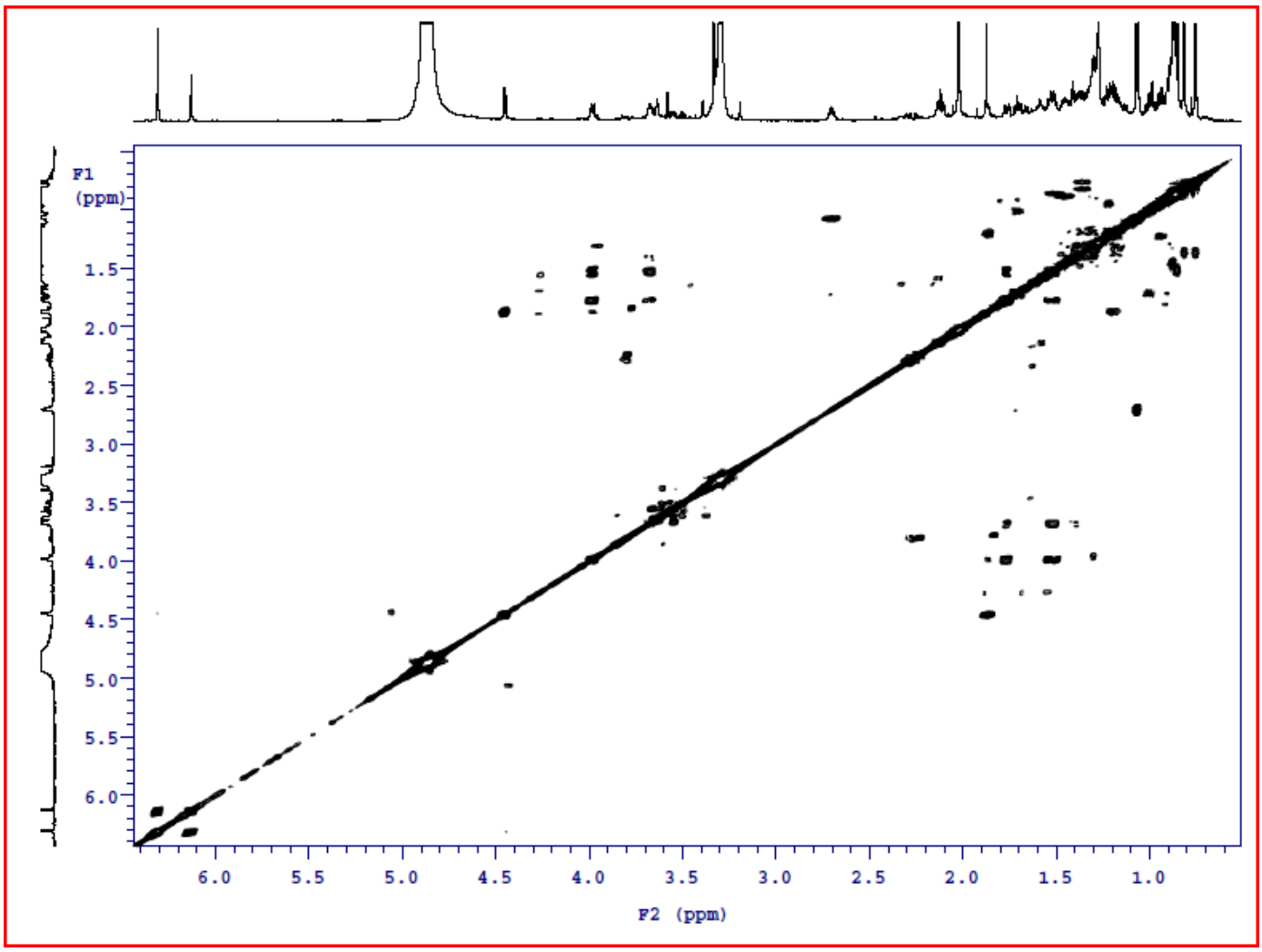


Figure 2-30: COSY NMR spectrum of baulamycin B (7) recorded at 700 MHz (in CD_3OD).

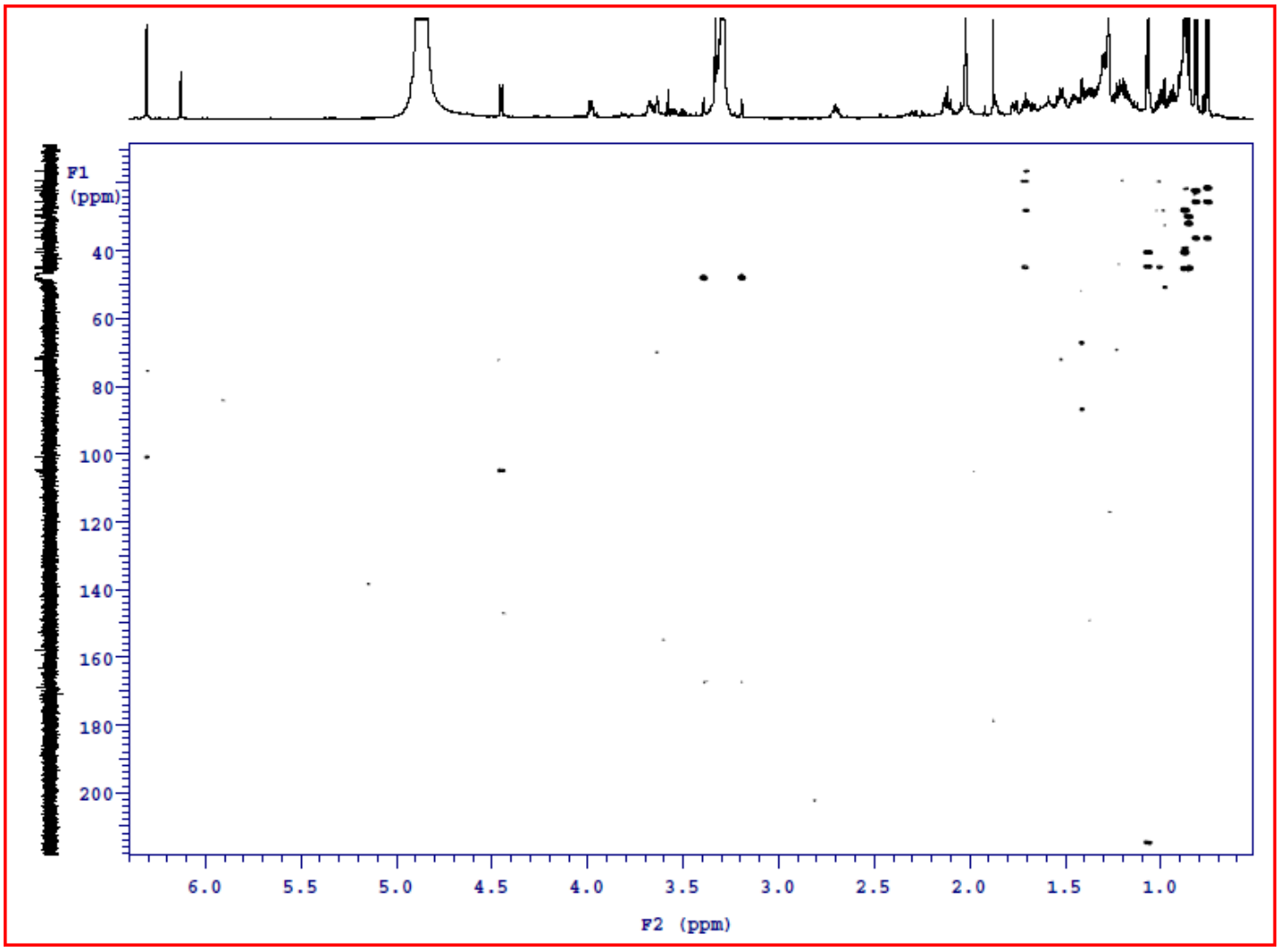


Figure 2-31: HMBCAD NMR spectrum of baulamycin B (7) recorded at 700 MHz (in CD_3OD).

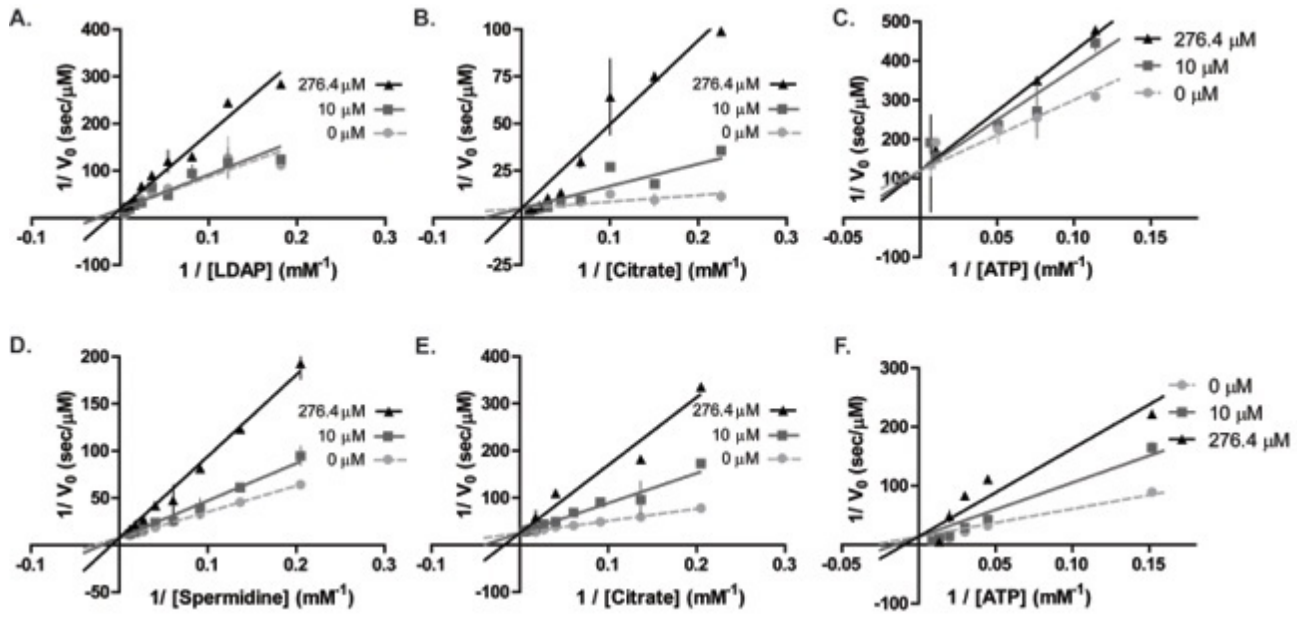


Figure 2-32. Kinetic analysis of the inhibition mechanism of BmcA.
Double reciprocal plots depicting BmcA inhibition of (a) SbnE as a function of L-DAP, (b) SbnE as a function of citrate, (c) SbnE as a function of ATP, (d) AsbA as a function of spermidine, or (e) AsbA as a function of citrate, (f) AsbA as a function of ATP.

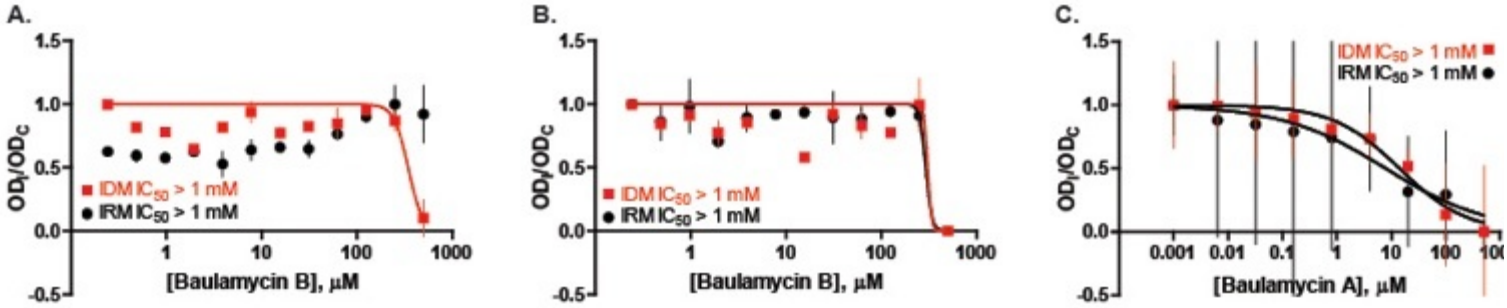


Figure 2-33. Live culture bioactivity Studies
(a) BmcB against *S. aureus* (Newman), (b) BmcB against *B. anthracis* (Sterne $34F_2$), and (c) BmcA against *S. typhimurium* in iron-depleted (IDM) and iron-rich (IRM) media.

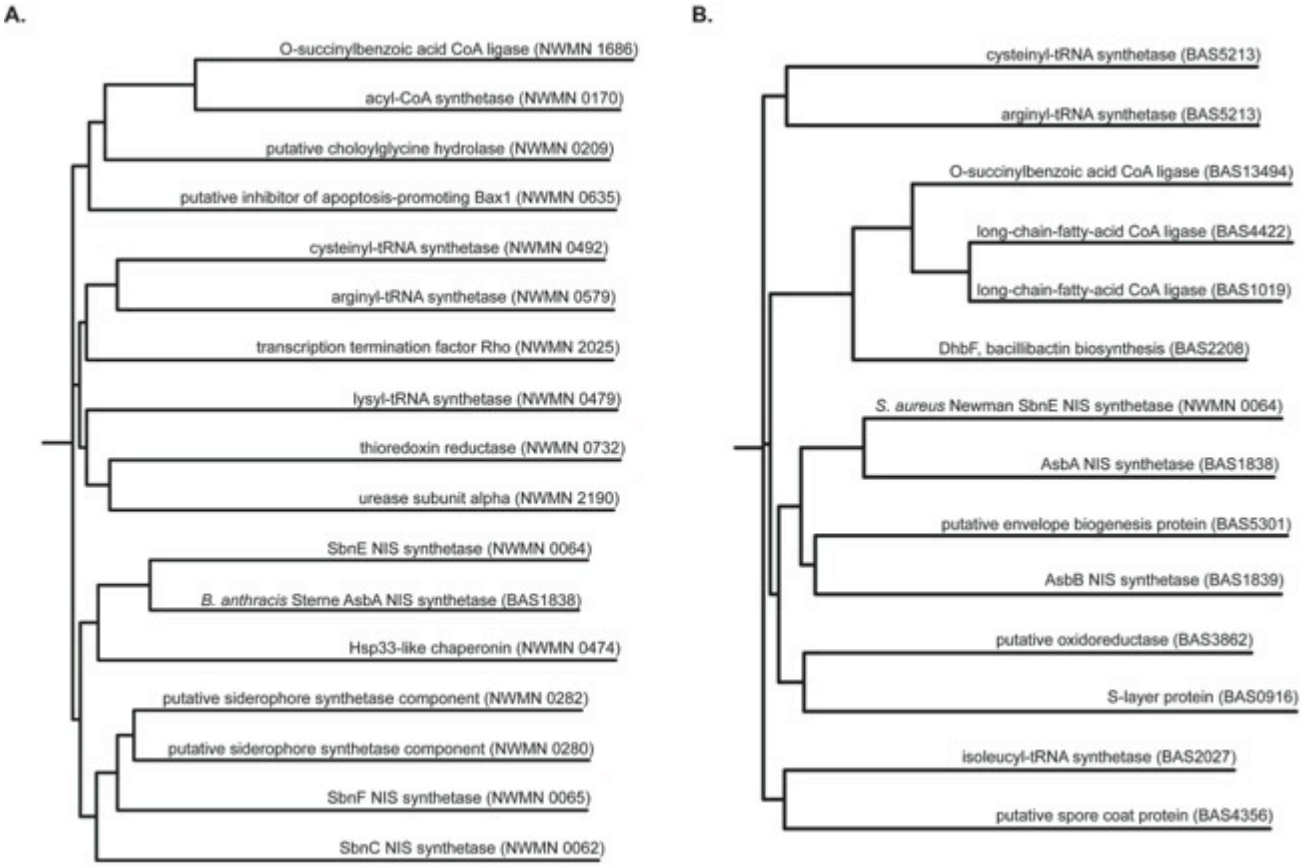


Figure 2-34. Phylogenetic trees
(a) SbnE and (b) AsbA depicting potential secondary targets within microbial cells. Both enzymes seem to share degrees of similarity to several enzymes, including adenylate-forming ones.

Substrate	V_{max} (μM/sec)	K_m (mM)	K_i (μM)	BmcA Inhibition Type
***SbnE in Staphyloferrin B Biosynthesis of* S. aureus**				
LDAP	0.054	37	210	Competitive
Citrate	0.20	6.9	50	Competitive
ATP	0.010	15	680	Competitive
***AsbA in Petrobactin Biosynthesis of* B. anthracis**				
Spermidine	0.13	36	170	Competitive
Citrate	0.039	10	110	Competitive
ATP	0.073	34	230	Competitive

Table 2-4. Enzyme kinetics

2.5 Materials and Methods

General Experimental Procedures. Optical rotation measurements were obtained on a Perkin-Elmer 241 Polarimeter calibrated using a Rudolph Quartz Control Plate Calibration Standard at sodium D line (at +11.502°). UV spectra were obtained on a UV-visible Molecular Devices SpectraMax M5 spectrophotometer using 1 mL cuvettes with 1.0 cm path lengths at room temperature in solvent MeOH. IR spectra were obtained on a Perkin Elmer Spectrum BX FT-IR spectrometer. Spectrophotometric assays were performed on Molecular Devices SpectraMax M5 384 variable wavelength spectrometer. All NMR spectra were acquired on a Varian INOVA 500 MHz and a Varian INOVA 700 MHz spectrometer at the NMR Facility, Department of Chemistry, University of Michigan. High-resolution APCIMS spectra were measured at the University of Michigan core facility in the Department of Chemistry using an Agilent 6520 Q-TOF mass spectrometer equipped with an Agilent 1290 HPLC system. RP-HPLC was performed using Waters Atlantis® Prep T3 OBD™ 5 μm 19 × 250 mm column and Luna 5 μm C8(2) 100 Å, AXIA packed column and a solvent system of MeCN and H_2O. The

LCMS analysis of HPLC fractions was performed on a Shimadzu 2010 EV APCI spectrometer.

Biological Material. *Streptomyces tempisquensis* (Strain # 34946-N9) was isolated from marine sediments collected at Playa Grande mangrove, Costa Rica (collection permit R-CM-INBio-30-2007). The procedure for the isolation of actinomycetes from these samples was previously described by Magarvey *et al.* (73). Maintenance and propagation of cultures were performed using standard media and protocols where 500 mg of wet sediment was diluted in 10 mL of sterile water and vortexed for 10 min. Then, 1 mL of this suspension was applied directly to the top of the discontinuous sucrose gradient and centrifuged for 30 minutes at 300 x g. 500 µL of the 20%, 30%, and 40% layers were then plated to HVA agar supplemented with 10 µg/mL chlortetracycline, 25 µg/ml cyclohexamide, and 25 µg/ml of nalidixic acid. The plates were then incubated at 28 °C for one month. The colony was picked off the plate and streaked onto ISP2 agar until pure. Seed cultures were grown in 17 mL dual position cap tubes containing 2 mL of ISP2 and grown for 4 days on a rotary shaker at 200 rpm. The seed culture was then poured into a 250 mL baffled flask containing 100 mL of ISP2 and grown for 18 days on a rotary shaker at 200 rpm. The culture was centrifuged at 4000 rpm for 10 min to remove the cells and 2 g of XAD16 resin (Sigma-Aldrich, St. Louis, Mo.) contained within a polypropylene mesh bag was added to the broth and incubated overnight on the rotary shaker. The resin bag was removed and placed into 10 ml of MeOH followed by 10 ml of acetone and 10 ml of ethyl acetate. Each of the three fractions was dried *in vacuo* and reconstituted to a final concentration of 15 mg/mL in DMSO.

Culture Maintenance and Fermentation. Seed cultures of 100 mL (×5) of ISP2 media (1% malt extract, 0.4% yeast extract, 0.4% dextrose, 3% NaCl) were inoculated with a loopful of vegetative cells from an oatmeal plate (6% oat meal, 1.25% agar, 3% NaCl) culture of *Streptomyces tempisquensis* and incubated with shaking (200 rpm) at 28 °C for 5 days. A 25 mL portion of the seed cultures were transferred to a 2.8 L Fernbach flask containing 1.5 L of the ISP2 medium, and the 39L fermentation was carried out on a rotary shaker (200 rpm) at 28 °C for 18 days. After 14-18 days of growth, the cultures

were harvested by centrifugation. The resulting cell free broth was subjected to solid phase extraction using 15 g of Amberlite XAD-16. The resin was then separated by filtration and subjected to organic extraction using MeOH: EtOAc (1:1).

Isolation and Purification of Baulamycins A-B (6-7). The organic extracts concentrated under vacuum to afford the crude extracts (~10 g) obtained from 39 L culture. The crude extracts were dissolved in 100 mL of H_2O and were subjected to a C18-silica gel column (20 × 2.6 cm, YMC Gel ODS-A, 12 nm, S-150 μm). The C-18 column was eluted with a stepwise gradient of H_2O/ACN (100:0 □ 0:100) to give nine fractions (Fr.1–Fr.8), which were concentrated *in vacuo* to yield fractionated organic materials, respectively. All eight fractions were assayed in the developed *in vitro* enzymatic assay at 10 and 1.0 ppm. The bio-active Fr.5 was further purified by RP-HPLC on a gradient of 50-90% ACN and was followed by UV/vis photodiode array detection at 210 nm to yield semi-pure compounds **6** (10.3 mg) and **11** (6.4 mg). Compounds were again subjected to re-purification over RP-HPLC on isocratic condition of 59% ACN using C-8 column to get compounds **6** (3.6 mg) and **7** (2.1 mg).

Baulamycin A (6): Isolated as light yellow amorphous solid: $[\alpha]^{20}_{D}$ -10.3 (c 0.20, MeOH); IR (film) 3324, 3012, 2952, 2867, 1699, 1602, 1456, 1377, 1143 cm^{-1}; UV_{max} (λ212 (log ε 3.24), 227 (log ε 2.93), and 280 (log ε 2.51).

Baulamycin B (7): Isolated as light yellow amorphous solid: $[\alpha]^{20}_{D}$ -10.1 (c 0.20, MeOH); IR (film) 3321, 3016, 2941, 2873, 1702, 1602, 1456, 1365, 1141 cm^{-1}; UV_{max} (λ) 212 (log ε 3.24), 227 (log ε 2.93), and 280 (log ε 2.51).

Gene Cloning, Expression, and Purification of Enzymes. The *sbnE* gene was PCR amplified from the genomic DNA of *S. aureus* (Newman). The resulting PCR fragment was cloned into the ligation independent cloning (LIC) vector pMCSG7, containing an N-terminal His_6 tag (74). Cloning of the construct containing the *asbA* gene into the C-terminal His_6 LIC vector pMCSG26 has been described previously (75). However, a mutation (L299P) was later discovered within the cloned *asbA* gene sequence. The gene was transferred into pET28a and site directed mutatgenesis was used to revert the

sequence to wild type. This reverted construct was utilized in dose response and kinetic studies. The cloning of the *asbB* gene into the vector pET28b has been previously described (58).

To obtain purified protein from constructs, recombinant *E. coli* BL21 (DE3)-Gold (Stratagene) cells containing the pRARE plasmid (Novagen) encoding tRNAs of rare codons were grown in terrific broth to an $A_{600\ nm}$ of ~0.8 at 37°C and then cooled to 18°C for an overnight induction with 0.2 mM isopropyl β-D-1-thiogalactopyranoside. Harvested cell pellets were lysed by sonication in a buffer containing 40 mM imidazole, 20 mM HEPES, 300 mM NaCl, 1 mM tris (2-carboxylethyl) phosphine (TCEP), 10% glycerol, a cocktail of protease inhibitors (EDTA-free Complete, Roche), and ~5 mg of lysing enzymes from *Trichoderma harzianum* (Sigma L1412). The cell-free extract was collected by ultracentrifugation at 45,000 x g for 45 minutes. The supernatant was subjected to Ni^{2+} affinity chromatography using a HiTrap™ Chelating HP column (GE healthcare) on a Fast-performance Liquid Chromatography system (FPLC system, Amersham Biosciences) with a linear gradient of 40-300 mM imidazole in 20 mM HEPES, 300 mM NaCl, 1 mM TCEP, and 10% glycerol. Purified protein was dialyzed against 2 L of storage buffer containing 40 mM HEPES, 150 mM NaCl, 1 mM TCEP, and 10% glycerol at pH 8. Protein was concentrated using Amicon Ultra centrifugal molecular weight cutoff filters (Millipore), flash frozen in liquid N_2, and stored at -80°C until analysis.

Natural Product Extract Library. At the time of screening, the natural product extract (NPE) library at the University of Michigan Center for Chemical Genomics contained ~20,000 extracts. Each extract in the library is derived from marine samples collected from all over the world, including Costa Rica, Panama, and Papua New Guinea. Some of these samples are from isolated microbes (n=19055) while others were derived from field-collected biomass samples ("macrosamples," n=800). Previous work describes in detail how these extracts are prepared for the library (76).

Assay Development and High-Throughput Screening. Fitness of assays for high throughput screening was determined using a Z factor statistical parameter (77), where *p* and *n* represent a positive control and negative control containing no enzyme and no inhibiting compound/extract, respectively:

$$Z = 1 - \frac{3(\sigma_p - \sigma_n)}{|\mu_p - \mu_n|}$$

The NIS hydroxymate-formation assays were conducted as previously described (78). The NIS-tailored malachite green assays were modified from previously described protocols (61-63). Assays were performed in a 384-well plate format with a final volume of 40 μL. Enzymatic activity of both enzymes was found to be most uniform across clear polystyrene microplates produced by Greiner Bio-One (781185). For both assays, we sought conditions that enabled the use of low enzyme concentrations to facilitate optimal detection of any extracts containing low abundance biologically active compounds. For the SbnE reactions, standard final conditions were 25 mM HEPES, pH 7.5, 5 mM $MgCl_2$, 100 μM ATP, 100 μM citrate, 100 μM L-DAP, 0.4 U/mL IPP, and 25 nM SbnE. For AsbA reactions, standard final conditions were 20 mM HEPES, pH 7.5, 1 mM $MgCl_2$, 150 μM ATP, 80 μM citrate, 800 μM spermidine, 0.2667 U/mL IPP, and 100 nM AsbA. Reaction mixtures were split into two solutions: a substrate solution and an enzyme solution to initiate reactions (Figure S3A). First, 30 μL of a substrate mix containing HEPES, $MgCl_2$, ATP, citrate, the corresponding polyamine, and IPP was dispensed into 384-well plates using a Multidrop (Thermo Fisher Scientific). Compounds, natural product extracts, or DMSO vehicle controls were then added to wells using a high density replication tool on a Biomek FX liquid handler (Beckman Coulter). To initiate the reactions, 10 μL of an enzyme solution containing HEPES, $MgCl_2$, and purified SbnE or AsbA enzyme was added to the reactions using a Multidrop. For positive controls, an equal volume of a solution only possessing HEPES and $MgCl_2$ was added. Following incubation at room temperature (SbnE: 30 minutes; AsbA: 1 hour), 10 μL of quenching solution of 10 parts malachite green, 2.5 parts 7.5% (NH4)2MoO4, and 0.2 parts 11% tween-20 was added to reactions with the Multidrop. Reactions were measured at A_{600} nm using the Pherastar multimode plate reader (BMG Labtech) after incubating 15 minutes at room temperature.

Dose-Response Assays against Purified Enzymes. A Mosquito X1 instrument (TTP Labtech) was utilized to dispense varying volumes of 50 mM isolated baulamycin compounds in DMSO onto clean 384 well microplates to make a final range of concentrations between 0 to 550 µM for SbnE and 0 to 1000 µM for AsbA and AsbB after a 40 µL reaction. Spotted plates were stored at -20°C until analysis.

Assays for SbnE and AsbA were carried out at the standard conditions used in screening (described above). For AsbB reactions, final conditions were 20 mM HEPES, pH 7.5, 1 mM $MgCl_2$, 150 µM ATP, 80 µM citrate, 800 µM *N*1-(3,4-DHB)-*N*8-citryl-spermidine (**6**), 0.2667 U/mL IPP, and 100 nM AsbB. Purification of **6** from *B. anthracis* Sterne $34F_2$ *ΔasbB* culture supernatants was performed as described previously (58). After incubation at room temperature (SbnE: 30 minutes; AsbA/AsbB: 1 hour), reactions were quenched and incubated for 15 minutes as before. Plates were read at A_{600} with a SpectraMax M5 microplate reader (Molecular Devices). To conserve baulamycin compound, reactions were conducted in duplicate.

Percent inhibition was calculated from raw absorbance data using a previously described formula (62) where *p* and *n* represent positive and negative controls respectively.

$$\%\ \text{Inhibition} = 100\left(\frac{\text{Abs}_{sample} - \text{Abs}_{\mu_n}}{\text{Abs}_{\mu_p} - \text{Abs}_{\mu_n}}\right)$$

Resulting data were fitted to the following standard log(inhibitor) *versus* normalized response model using GraphPad Prism version 5.0 (GraphPad Software). This model demonstrated adequate fit, with R^2 values for all SbnE and AsbA curves greater than 0.9.

$$y = \frac{100}{(1 + 10^{(\text{logIC}_{50} - X)\text{HillSlope}})}$$

Enzyme Kinetic Assays. Reactions were a modification of previously described protocols (58, 65, 66, 79). Experiments were conducted by varying a single substrate and

holding remaining reaction components at a constant saturating level in the presence and absence of inhibitor. The MesG continuous pyrophosphate detection assay was carried out in 384-well microplates (Greiner Bio-One) at 40 µL final volumes in triplicate. For both SbnE and AsbA reactions, constant standard conditions were 50 mM Tris, pH 8, 15 mM MgCl2, 0.5 mM reducing agent (dithiothreitol for SbnE and TCEP for AsbA), 0.001 units/µL PNP, 0.0004 units/µL IPP, 0.4 mM freshly prepared MesG, and 1 µM purified SbnE or AsbA enzyme. SbnE kinetic reactions were conducted at fixed saturating concentrations of L-DAP (40 mM), citrate (10 mM), or ATP (12 mM) while AsbA kinetic reactions were conducted at fixed saturating concentrations of spermidine (40 mM), citrate (20 mM), or ATP (12 mM). The SbnE kinetic reactions with varied concentrations of citrate, L-DAP, or ATP were incubated for 8, 5, or 3 hours, respectively, before the addition of enzyme. All AsbA reactions were incubated for 10 minutes before initiation of the reaction with enzyme. For both sets of reactions, assays were tested at fixed concentrations of BmcA (0, 10, or 276.4 µM in DMSO) spotted with a Mosquito X1 instrument at varied concentrations of one substrate. Measurements at $A_{360\ nm}$ were taken every 30s over the course of 20 minutes using a SpectraMax plate reader (Molecular Devices). The initial velocities under each set of reaction conditions were subsequently calculated using the slope function in Microsoft Excel. Double-reciprocal (Lineweaver-Bruk) plots of resulting kinetic data were then used to determine the mechanism of inhibition against each substrate. The K_i values were determined from a replot of the slopes from the double-reciprocal plots *versus* concentration of BmcA (80, 81).

Dose-Response Assays against Microbial Cell Cultures. Strains were grown in LB-Miller medium for iron-rich conditions and LB-Miller medium supplemented with the iron chelator 2,2'-bipyridyl (750 µM for *S. aureus* (Newman) and MRSA (USA 300) strains; 200 µM for the *B. anthracis* $34F_2$ strain) for iron-depleted conditions. Pelleted cells from cultures grown to mid log phase ($0.4 < OD_{600} < 0.6$) in iron-rich conditions were washed 3x with medium supplemented with 2, 2'-bipyridyl to remove excess iron. Cells re-suspended in iron-depleted media were then used to inoculate fresh 96-well plate cultures at an OD_{600} of 0.003 in either iron-rich or iron-limited conditions. Pure

baulamycins at concentrations designated in figures or vehicle controls containing an equal volume of DMSO were added to wells before incubation at 37°C, 200 rpm. For dose response analysis, optical densities after 8 hours of inhibitor-treated (OD_I) or DMSO-treated (OD_C) cultures were fit to the following previously described (39) sigmoidal equation using GraphPad Prism software:

$$\frac{OD_I}{OD_C} = b + \frac{(a-b)}{1+\left(\frac{[I]}{IC_{50}}\right)^s}$$

16S rDNA PCR Amplification, Cloning, and Sequencing. The genomic DNA of *Streptomyces tempisquensis* was extracted from 1 mL of culture using the Wizard® Genomic DNA Purification Kit (Promega A1120) per the manufacturer's protocol. The 16S rDNA was amplified from the isolated genomic DNA by PCR using the universal primers FC27 (5′-AGAGTTTGATCCTGGCTCAG-3′) and RC1492 (5′-TACGGCTACCTTGTTACGACTT-3′) (82). The PCR reaction mixtures were prepared according to the standard protocol described by GoTaq® Green Master Mix (Promega M7122). A Bio-Rad iCycler thermocycler was used for template amplification with the following PCR cycle conditions: an initial denaturation step at 98°C for 30 s, followed by 30 cycles of 95°C for 30 s, 50°C for 30 s, 72°C for 90 s, and a final extension of 72°C for 7 min. The PCR products were purified using the Wizard® SV Gel DNA Recovery Kit (Promega A9282). The purified PCR fragments were modified to add a 3' A overhang using Taq DNA polymerase. Ligation of the modified product into pGEM®-T Easy vector (Promega) was completed with T4 DNA Ligase (NEB M0202) according to the manufacturer's standard protocol. The resulting pGEM-16S rDNA construct was transformed into chemically competent *E. coli* DH-5α cells. The transformation reactions were mixed with 40 μL of X-gal (20 mg/mL in DMSO) and 40 μL IPTG (100 mM in sterile H2O) before plating on LB agar plates containing ampicillin (250 μg/mL). Plates were incubated overnight at 37ºC for α-complementation. After incubation, single white colonies were used to inoculate LB media in sterile culture tubes (×6). All the tubes were incubated at 37ºC, 250 rpm for overnight growth. Plasmid DNA was isolated from pelleted cells using the Wizard® SV Miniprep Kit (Promega A1460) and sequenced

using T7 and SP6 primers. The 16S sequence for *S. tempisquensis* has been given the GenBank accession number KF954543.

Phylogenetic Analysis. Phylogenetic analyses were conducted in GENEIOUS R6. Geneious version 1.8 (Biomatters). The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates was taken to represent the evolutionary history of the taxa analyzed.

2.6 Notes and Author Contributions

Portions of this chapter were published in the *Journal of the American Chemical Society*:

*Tripathi A, ***Schofield MM**, Chlipala GE, Schultz PJ, Yim I, Newmister SA, Nusca TD, Scaglione JB, Hanna PC, Tamayo-Castillo G, Sherman DH. "Baulamycins A and B, Broad-Spectrum Antibiotics Identified as Inhibitors of Siderophore Biosynthesis in *Staphylococcus aureus* and *Bacillus anthracis*." *J. Am. Chem. Soc*. 2014, 136(4):1579-86. Asterisk denotes that authors contributed equally to this work.

Author contributions are as follows: Ashootosh Tripathi, Michael M. Schofield, Philip C. Hanna, Giselle Tamayo-Castillo, and David H. Sherman contributed to the experimental design. Michael M. Schofield and George E. Chlipala developed the malachite green assay and carried out screening. Pam J. Schultz and Giselle Tamayo-Castillo managed strains and extracts for the Center of Chemical Genomics. Ashootosh Tripathi, Michael M. Schofield, and Pam J. Schultz regrew strains from top HTS hits. Ashootosh Tripathi and Isiah Yim carried out all fractionation while Michael M. Schofield tested fraction bioactivity. Ashootosh Tripathi carried out extensive structural elucidation. Michael M. Schofield developed assays for dose response, kinetic analysis, and live culture sturdies. Michael M. Schofield completed dose response analysis and live culture studies. Michael M. Schofield and Ashootosh Tripathi conducted kinetic assays and analysis. Sean A. Newmister, Tyler D. Nusca, and Jamie B. Scaglione carried out initial cloning experiments. Ashootosh Tripathi, Michael M. Schofield, Philip C. Hanna, Giselle Tamayo-Castillo, and David H. Sherman evaluated the data.

Chapter 3
Understanding natural products from uncultured bacteria: Metagenomic analysis of the endosymbiont producer of a chemotherapeutic natural product

3.1. Introduction

Natural products are a critical source of pharmaceuticals and lead compounds in drug discovery efforts (83, 84). Over the last several decades, scientists have isolated thousands of biologically active metabolites from terrestrial and marine macroorganisms, including plants, insects, and animals. Mounting evidence suggests that microbial symbionts may be the actual producers of most of these natural products (26, 27).

Currently, the vast majority of drug-producing symbiotic microbes remain uncharacterized. Most fall into the > 99% of bacterial species currently incapable of being cultured in the laboratory, hindering their study (26, 28). Identifying these symbionts and understanding their genetic, biochemical, and metabolic characteristics is critical for advancing fundamental knowledge and potential applications. Many symbiont-derived secondary metabolites can only be isolated in low yields from their hosts, making large-scale production for pharmaceutical purposes unsustainable from both an economic and environmental perspective. Although total synthesis can sometimes solve the supply problem, it can be costly and fails to address our understanding of the unique biosynthetic processes that are mediated by these elusive microbes (85). Sequencing and analysis of symbiont genomes could provide insight into the lifestyles of these elusive bacteria, illuminate possible host-free cultivation methods, and provide a route to economical and sustainable large-scale production with the opportunity for genetic manipulation to produce novel drug analogues.

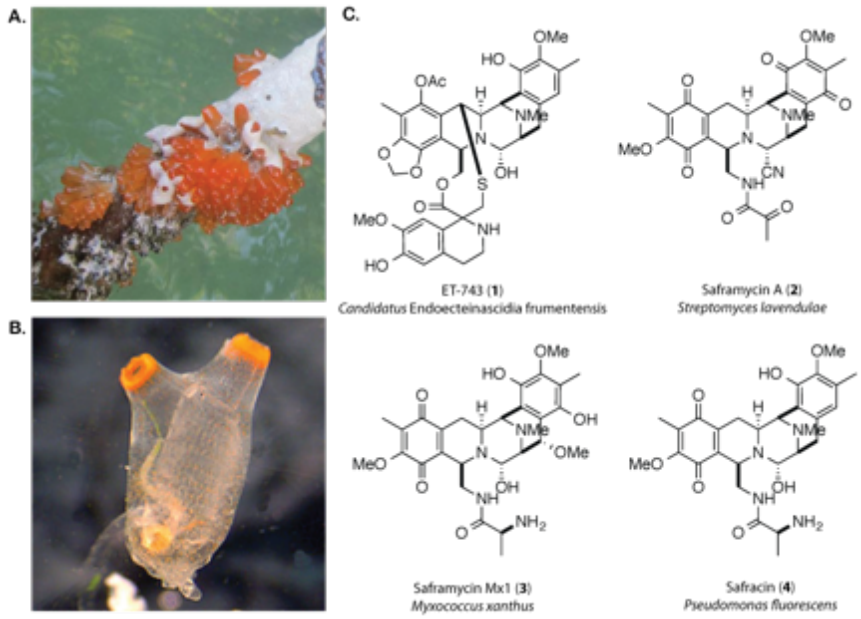


Figure 3-1. Mangrove tunicates the tetrahydroisoquinoline natural products.
A.) Tunicate colonies growing on the root of a mangrove tree in the Florida Keys. **B.)** The metagenomic DNA from four individual tunics (zooids) from two colonies was isolated and sequenced. **C.)** The chemotherapeutic compound ET-743 (**1**) and three natural products from cultured bacteria that share a similar tetrahydroisoquinoline core.

The chemotherapeutic compound ET-743 (**1**, Yondelis, Trabectedin) is one of the most important natural products suspected to be of symbiotic origin. Isolated directly from the mangrove tunicate *Ecteinascidia turbinata* (**Figure 3-1*AB***), the biological activity of the drug against cancer cells has inspired over 40 years of research (86, 87). Currently, ET-743 is clinically approved in Europe against soft tissue sarcoma and relapsed ovarian cancer and is currently in phase III trials as an anticancer therapeutic in the United States (88).

The tetrahydroisoquinoline alkaloid natural products saframycin A (**2**), saframycin Mx1 (**3**) and safracin (**4**) are derived from three distinct cultured bacteria and are structurally similar to ET-743, supporting a bacterial origin for the drug (**Figure 3-1C**). Studies of the mangrove tunicate over a decade ago identified the potential intracellular Gammaproteobacterium *Candidatus* Endoecteinascida frumentensis to be the most

prevalent member of the host microbial consortium (89, 90) and the only bacterial species consistently associated with tunicates in both the Mediterranean and Caribbean seas (89). A metagenomically-derived contig containing a partial ET-743 biosynthetic gene cluster was later indirectly linked to a separate contig bearing the 16S rRNA gene sequence for *Ca*. E. frumentensis through analysis of %G+C content and codon usage (91). Cultivation of the producing bacterium has so far been unsuccessful (89, 90), and aquaculture (92) of the host tunicate and total synthesis (93) have also failed to provide sustainable access to the drug for clinical applications. ET-743 is therefore currently generated by a lengthy semisynthetic process starting from fermentation-derived cyanosafracin B (94).

In this study, we utilized next generation sequencing technologies to expand and complete the biosynthetic gene cluster and uncover the complete genome of the microorganism responsible for ET-743 production. Analysis of phylogenetic markers and protein coding genes suggests that the microbe belongs to a novel family of Gammaproteobacteria. In-depth genomic analysis also provides initial insights into the endosymbiotic lifestyle of *Ca.* E. frumentensis, the ecological role of its sole secondary metabolic pathway, and key information that may provide access to host-cell free growth in the laboratory.

3.2 Results and Discussion

3.2.1 Overview of Samples and Dataset.

The colonies of *E. turbinata* consist of thick bundles of individual zooids connected by a network of stolons that enable adherence of the animal to a stable surface (92). Our laboratory previously isolated metagenomic DNA from individual zooids and uncovered a 35 kb gene cluster responsible for ET-743 biosynthesis using 454 pyrosequencing (91). In the present study, we isolated additional metagenomic DNA from four zooids obtained from two colonies (**Figure 3-5 in Supplemental Information [SI**]). We shotgun sequenced the resulting DNA samples using Illumina HiSeq technology and assembled

the data into contigs. The four zooids provided metagenome datasets each containing over 800 Mbp of sequence (**Table 3- 2 in SI**).

We assigned the assembled contigs to taxonomic bins using tetranucleotide frequency with emergent self-organizing maps (tetra-ESOM) as previously described (**Figure 3-6 in SI**) (95). Each of the four metagenomic samples possessed a single bin containing both the previously identified partial ET-743 biosynthetic gene cluster and the 16S rRNA gene for *Ca*. E. frumentensis (**Table 3-2 in SI**). The four bins containing the ET-743 producing microorganism were further assembled into a consensus genome containing three contigs. PCR amplification closed a 200 bp gap between two of the contigs to create a 630 kb scaffold. Additional PCR amplification closed a final 1.5 kb gap in the scaffold to create the closed genome for *Ca*. E. frumentensis (**Figure 3-2**, **Table 3-1**, and **Figure 3-5 in SI**).

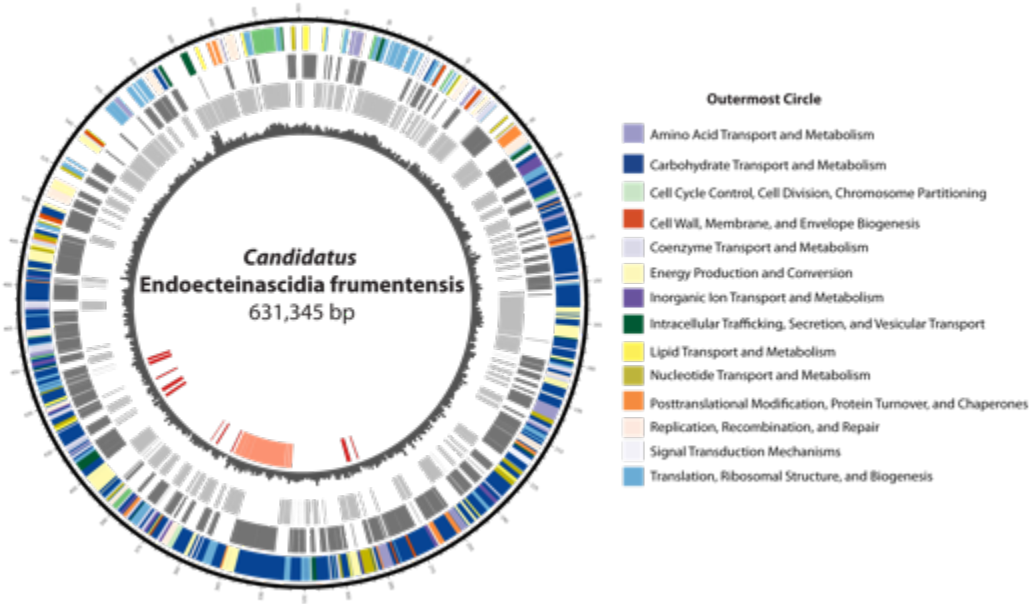


Figure 3-2. A circular map of the closed genome of *Candidatus* E. frumentensis.
The outermost circle displays protein-coding genes assigned to Pfam categories (see key). The dark grey and light grey circles display protein-coding genes on the plus strand and minus strands, respectively. The fourth circle depicts a histogram of G+C content throughout the genome. The innermost circle represents ET-743 biosynthetic genes. Genes previously identified are depicted in light red while putative new genes are shown in dark red.

The coverage depth for the endosymbiotic genome averaged 721x between the four samples (**Table 3-3 in SI**). However, one contig that consistently binned with ET-743 producer, and that was retrieved in all four samples, was not incorporated into the genome. This much smaller ~18 kb contig encodes a DNA primase and two protein-coding genes with ambiguous functions that repeat throughout the stretch of the sequence. Unlike the circular genome, the shorter contig has a coverage depth of only ~74x and reads could not be mapped to the sequence with confidence. The excluded contig may be an extrachromosomal element that is present in only a subset of the *Ca*. E. frumentensis population, or an artifact of the assembly and binning process. Given that the rest of the genome was closed and displayed even and deep coverage, we focused our analysis on the closed *Ca*. E. frumentensis genome in this study.

Very few other genomic bins were detected in the metagenomic datasets, despite prior evidence that the tunicate housed a complex microbial consortium (**Table 3-2 in SI**) (90, 91). However, previous studies indicated *Ca*. E. frumentensis was one of the most abundant species in the consortium (89-91) and the only microorganism found to be consistently associated with the tunicate host in both the Mediterranean and Caribbean marine habitats (89). Further, metagenomic assembly of the symbiont population was likely facilitated due to its low genomic diversity compared to populations that are non-specifically associated. Thus, it is likely that the eukaryotic host and *Ca*. E. frumentensis monopolized the sequencing data, especially the large assembled contigs, despite the presence of a complex but lower abundance microbial community. The only other notable bin after tetra-ESOM was a cyanobacterium from the order *Oscillatoriales* that was present in two of the four metagenomic DNA samples (**Table 3-2 and Figure 3-6 in SI**).

3.2.2 Genome Reduction in the Symbiont.

Previous *in situ* hybridization analysis provided an initial indication that *Ca*. E. frumentensis could be a bacterial endosymbiont (90). Assembly and analysis of the microbe's complete genome provides further convincing evidence of an intracellular lifestyle and long-term evolution with the tunicate host, *E. turbinata*. *Ca.* E. frumentensis

possesses many of the hallmarks of genome reduction, which is thought to be driven by a small bacterial population size and an inherent deletion bias (96-98). The circular genome for *Ca.* E. frumentensis is quite small, totaling only 631,345 bp (**Figure 3-2**). The small size of the genome rivals those of the model obligate endosymbionts *Buchnera aphidicola* in aphids and *Wigglesworthia glossinidia* in tsetse flies (**Table 3-4 in SI**). The functions maintained by *Ca*. E. frumentensis are also consistent with the minimal gene sets observed in these and other obligate symbionts (**Figure 3-7 in SI**). For example, *Ca.* E. frumentensis appears to have lost a number of genes involved in DNA replication and repair mechanisms (**Figure 3-7 in SI**). The loss of DNA repair mechanisms is thought to be a crucial turning point during the evolution of an endosymbiont (96, 99). Loss of these genes is frequently accompanied by increased mutation rates, an A+T DNA sequence bias, and the loss of additional nonessential genes.

Detail	*Candidatus* E. frumentensis
Genome Size (bp)	631,345
GC Content (%)	
Total	23.3
Coding Regions	24.2
Noncoding Regions	12.7
Coding Density (%)	90.7
Intergenic Pseudogenes	10
Protein-coding genes	585
With functional annotation	556 (95.0%)
With ambiguous function	29 (4.6%)
rRNA genes	3
tRNA genes	32

Table 3-1. General features of the *Candidatus* E. frumentensis genome.

Indeed, the exceptionally low total G+C content (23.3%) of *Ca.* E. frumentensis genomic DNA supports a mutational bias and an obligate endosymbiotic lifestyle. The G+C content disparity between the coding (24.2%) and noncoding (12.7%) regions of the genome (**Table 3-1**) further exemplifies this bias. Bacterial lineages that only recently became restricted to a host organism also often have higher numbers of pseudogenes within these noncoding regions and a consequently low overall coding density (100). However, as bacteria continue to co-evolve with their hosts, pseudogenes gradually shrink and become unrecognizable through deletions while genomes become more compact (97, 101, 102). The noncoding regions of the *Ca.* E. frumentensis genome have only 10 pseudogenes whose predicted translation products show amino acid sequence similarity to known proteins (**Table 3-5 in SI**). The genome also has a higher overall coding density of 90.7% (**Table 3-1**), similar to *B. aphidicola*, *W. glossinidia*, and other obligate endosymbionts that co-evolved with their hosts along the order of millions of years (99, 103, 104). Taken together, these data provide strong support that *Ca*. E. frumentensis is an obligate endosymbiont that has undergone long-term co-evolution with the tunicate host, *E. turbinata*.

3.2.3 Phylogenetic analysis and novelty of *Ca.* E. frumentensis.

The genome of *Ca*. E. frumentensis also appears to be remarkably distinct from other studied microorganisms. Analysis of conserved markers provided the first evidence that *Ca.* E. frumentensis may be phylogentically distant from characterized bacterial species. The closest homologues for genes encoding the 16S rRNA gene, *rpoB,* and *recA* had 86.1%, 69.0%, and 74.8% sequence identities respectively (**Figure 3-8 in SI**).

Phylogenetic markers can be useful for microorganisms that have many well-studied and cultured close relatives. However, in microorganisms with fewer obvious relatives, the average amino acid identity (AAI) of shared genes can be more revealing (105). To further explore the phylogenetic novelty of *Ca.* E. frumentensis, we compared the AAI and 16S rRNA gene of the microorganism to other bacterial species selected from a taxonomic profile of the *Ca*. E. frumentensis genome. This analysis confirmed that *Ca.* E. frumentensis is taxonomically distinct from many of its originally predicted relatives and likely represents a new family of Gammaproteobacteria (**Figure 3-9 in SI**) (106).

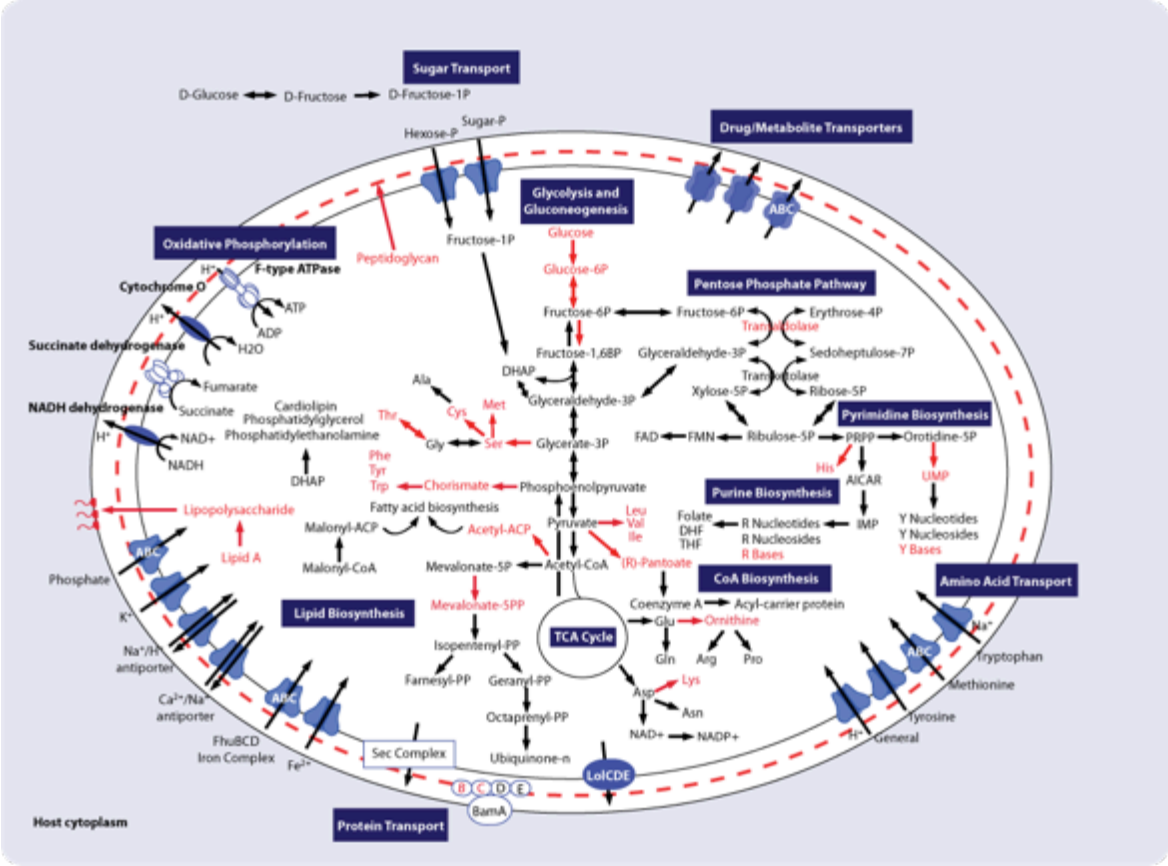


Figure 3-3. Overview of the metabolism of *Ca.* E. frumentensis deduced from genomic analysis.
Reaction products depicted in red have either missing or partially missing biosynthetic pathways. ACP, acyl carrier protein; AICAR, 5-Aminoimidazole carboxamide ribonucleotide; CoA, coenzyme A; DHAP, Dihydroxyacetone phosphate; DHF, dihydrofolate; DMAPP, Dimethylallyl pyrophosphate; FAD, flavin adenine dinucleotide; FMN, flavin mononucleotide; IMP, inosine monophosphate; NAD, Nicotinamide adenine dinucleotide; PRPP, Phosphoribosyl pyrophosphate; THF, tetrahydrofolate; UMP, uridine monophosphate.

3.2.4 Primary Metabolism.

Central metabolism and carbon sources. Analysis of primary metabolic pathways (**Figure 3-3**) provided further insight into the lifestyle of the taxonomically distinct endosymbiont *Ca*. E frumentenes and its relationship with the host tunicate. Despite the small size of the genome, *Ca.* E. frumentensis possesses portions of all three components of central metabolism. The tricarboxylic acid cycle (TCA cycle) is intact as is most of the non-oxidative branch of the pentose phosphate pathway. The symbiont also has the

majority of the genes involved in the glycolytic pathway and gluconeogenesis. Interestingly, the genome is devoid of any genes encoding enzymes involved in early glucose catabolism (**Figure 3-3**). This is surprising since glucose is the preferred carbon source for many bacteria (107). However, the symbiont may use alternative sugar sources from the host for energy. The genome encodes two sugar phosphate transporters and enzymes for the remainder of the glycolytic pathway (**Figure 3-3**). Sugar phosphates can be important carbon sources for intracellular pathogens and endosymbionts, especially those present in the host cytosol (108). The symbiont may import sugar phosphates from the host, including fructose-1-phosphate or more advanced sugar phosphate intermediates in the glycolytic or pentose phosphate pathways.

Electron transport chain. The symbiont genome encodes a complete electron transport chain and an F-type ATPase (**Figure 3-3**). The respiratory chain likely obtains electrons from succinate and NADH produced from the TCA cycle to generate an electrochemical gradient and produce ATP. Interestingly, cytochrome *bo3* (encoded by *cyoABCDE*) is the terminal cytochrome oxidase in the respiratory chain (**Figure 3-3**). This indicates that the symbiont is capable of growth under high oxygen tension, similar to endosymbionts in the genera *Buchnera* and *Wigglesworthia* (109, 110). Conversely, intracellular pathogens from the genera *Legionella*, *Brucella*, *Chlamydia*, *Rickettsia*, and *Coxiella* are thought to rely on microaerophilic metabolism during intracellular growth (111, 112).

Amino acid and cofactor metabolism. Similar to most obligate endosymbionts and many intracellular pathogens, *Ca.* E. frumentensis seems to lack a number of genes involved in the biosynthesis of key amino acids and cofactors (**Figure 3-3**). Genes encoding intact pathways for *de novo* biosynthesis of asparagine, aspartic acid, glutamine, and glutamic acid are present, but pathways for the remaining 16 amino acids are either partially or completely missing (**Figure 3-3**). The symbiont likely obtains many of these missing amino acids from the tunicate host. Indeed, the *Ca.* E. frumentensis genome encodes specific transporters for methionine, tryptophan, and tyrosine along with an additional putative amino acid transporter (**Figure 3-3**). The symbiont could also import pathway intermediates to complete production of some amino acids possessing partial biosynthetic

pathways (*e.g.* glycine, serine, alanine, proline, and arginine).

Ca. E. frumentensis also lacks many genes involved in coenzyme A (CoA) biosynthesis (**Figure 3-3**). Genes encoding enzymes responsible for production of CoA precursors are absent. However, later steps of the pathway are intact, suggesting the symbiont is capable of CoA biosynthesis from host-derived pantotheine, β-alanine, cysteine, or another CoA precursor. The inability to produce CoA *de novo* is relatively common in endosymbionts and bacterial pathogens (113). The obligate endosymbiont *Buchnera aphidcola*, for example, works collaboratively with its aphid host to biosynthesize CoA (109). Similarly, intracellular pathogens from the genera *Mycoplasma*, *Rickettsia*, and *Chlamydia* have incomplete pathways for the coenzyme and often need media supplements for host-cell free growth or pathogenicity (114-116). Only a selection of these species have transporters suspected to facilitate CoA and precursor environmental uptake (117). It is therefore not surprising that *Ca.* E. frumentensis lacks transporter candidates for this role.

Membranes and transport. The *Ca*. E. frumentensis genome contains gene sets for the biosynthesis of lipids commonly incorporated into bacterial membranes, including phosphatidylethanolamine, cardiolipin, and phosphatidylglycerol (**Figure 3-3**). However, the endosymbiont possess an incomplete gene set for peptidoglycan biosynthesis (**Figure 3-3** and **Figure 3-7 in SI**). The vast majority of bacteria incorporate some level of peptidoglycan into their cell walls, but peptidoglycan is absent in a few bacteria, including *Mycoplasma* species, *Planctomyces*, and *Rickettsia*. Like *Ca*. E. frumentensis, some insect endosymbionts may also lack the ability to make peptidoglycan; *B. aphidicola* BCc (118), *Carsonella ruddii* (119), and *Ca*. Sulcia muelleri (120) are missing the majority of genes involved in the biosynthesis of peptidoglycan. Only *dacA*, the gene responsible for the final processing step in peptidoglycan biosynthesis, is present in the *Ca*. E. frumentensis genome (**Figure 3-3** and **Figure 3-7 in SI**).

In addition to peptidoglycan, *Ca*. E. frumentensis also seems to lack the genes

responsible for biosynthesizing and incorporating lipopolysaccharides in its outer membrane (**Figure 3-3** and **Figure 3-7 in SI**). Lipid A biosynthetic genes are also absent from many endosymbionts, including *Baumannia cicadellinicola* and *Buchnera* species (98, 120). The loss of lipid A from these endosymbionts could be an evolutionary result of its high toxicity to most eukaryotic hosts.

Despite its unusual membrane composition, the endosymbiont still has considerable genomic potential for metabolite transport. The genome encodes 71 genes putatively linked to transporter function. Although the specificity of many of these transporters is difficult to assess, others are well annotated. In addition to amino acid and sugar phosphate transporters, the genome also encodes membrane proteins that may target inorganic phosphate, iron, and potassium along with sodium-calcium and sodium-hydrogen antiporters (**Figure 3-3**).

Ca. E. frumentensis also has genes encoding a near intact Sec protein translocation pathway and enzymes involved in the recognition of signal peptides (**Figure 3-3**). Only the nonessential components SecG and SecM appear to be absent. The ABC transporter responsible for localization of lipoproteins to the periplasmic surface of the outer membrane (LolCDE), as well as essential components involved in inserting β-barrel proteins into the outer membrane (BamADE), are also present.

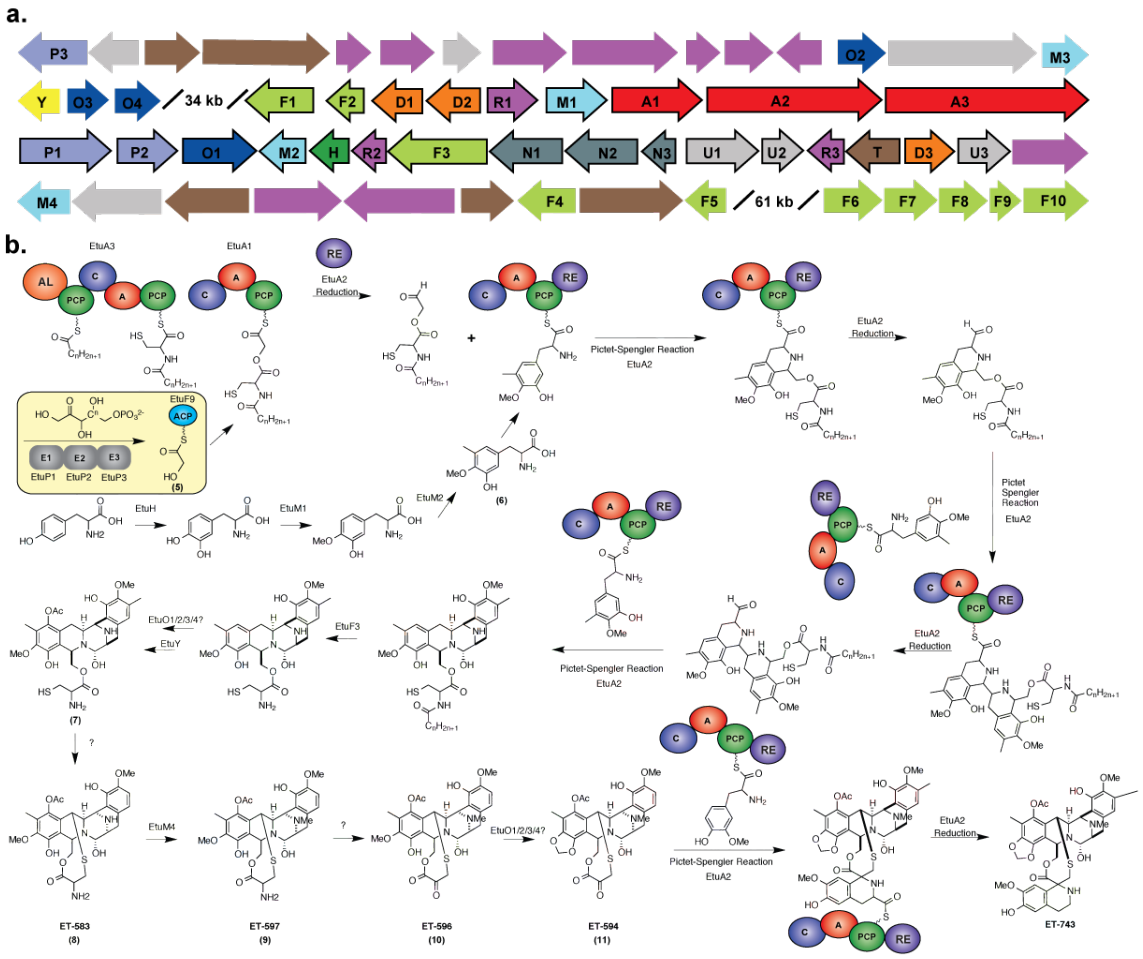


Figure 3-4. Completion of the ET-743 biosynthetic gene cluster.
A.) New ET-743 biosynthetic genes were identified upstream and downstream of the original ET-743 biosynthetic gene cluster (outlined in black). Gene products are classified as pyruvate dehydrogenases, enzymes of unknown function, transport proteins, regulatory and primary metabolism enzymes, flavoenzymes, methyltransferases, acetyltransferases, fatty acid biosynthetic enzymes, DNA processing enzymes, nonribosomal peptide synthetases, and hydroxylases. **B**.) An updated proposed biosynthetic pathway for ET-743.

3.2.5 Secondary Metabolism.

We previously identified a 35 kb contig containing many of the genes involved in the biosynthesis of the chemotherapeutic natural product ET-743 (91). However, close examination of ET-743, its previously isolated precursors (87), and other well-studied tetrahydroisoquinoline natural products (121-124) led us to suspect that we were still missing a number of key biosynthetic genes (91). Expanding the 35 kb gene cluster to a complete genome for *Ca*. E. frumentensis has enabled us to identify many of these

previously missing key genes and improved our understanding of ET-743 biosynthesis. Key genes involved in production of the chemotherapeutic drug span over 173 kb of the small 631 kb genome (**Figure 3-2**). Biosynthetic genes are split into three distinct clusters within this expansive genomic range (**Figure 3-4A**, **Table 3-6 in SI**). Newly detected gene products include the acetyltransferase EtuY and EtuM4, likely involved in acetylation and *N*-methylation to make **7** and ET-597 (**9**) respectively. We also identified three new flavoproteins in addition to the FAD-dependent monooxygenase (EtuO1) contained within the original ET-743 biosynthetic gene cluster (91).

We additionally identified a gene encoding the E3 component of the pyruvate dehydrogenase complex (EtuP3, **Figure 3-4**). The reactions catalyzed by this enzyme system typically provide the TCA cycle with acetyl-CoA (125). However, the primary metabolic enzymes were recently shown to also contribute to the biosynthesis of quniocarcin and naphtyridinomycin natural products (126). The enzyme complex can work with an acyl carrier protein (ACP) to provide a glycolicacyl-S-ACP extender unit (**5**) for a non-ribosomal peptide synthetase (NRPS). Both of these gene clusters in addition to SF-1739 (124) and the original ET-743 (91) biosynthetic gene cluster contain the E1 and E2 components for the enzyme complex. Although the E3 component has been absent in previously studied clusters, purified exogenous E3 does seem necessary for complete product conversion (126). The presence of the E3 component in *Ca*. E. frumentensis and its proximity to other ET-743 biosynthetic genes further exemplifies its importance in the biosynthesis of tetrahydroisoquinoline natural products.

Another genomic feature that may set the ET-743 cluster apart from other natural products is the placement of the ACP that operates with the pyruvate dehydrogenase complex. The ACP is located in the main biosynthetic gene clusters for quinocarcin, naphtyridinomycin, and SF-1739. However, the only ACP in the entire *Ca*. E. frumentensis genome is located within a fatty acid biosynthetic cluster 61 kb downstream of the original ET-743 gene cluster (EtuF9, **Figure 3-4A**). The location of the ACP and the presence of other fatty acid biosynthetic genes (EtuF1 and EtuF2) within the central ET-743 biosynthetic gene cluster further supports potential interaction between primary and secondary metabolism during ET-743 biosynthesis. This ACP most likely works in

concert with EtuP1, EtuP2, and EtuP3 to provide the glycolicacyl-S-ACP extender unit (**5**) to EtuA1 (**Figure 3-4B**).

Despite these new discoveries, we may still be missing some genes involved in ET-743 biosynthesis. For example, gene candidates for enzymes that catalyze formation of the thioether ring (**8**) and transamination to make ET-596 (**10**) remain to be identified. We cannot rule out that these genes may be located elsewhere in the *Ca*. E. frumentensis genome or that the microbe works together with its host to complete construction of the chemotherapeutic compound.

Although not included within the ET-743 gene cluster, the *Ca*. E. frumentensis genome also contains several genes found within the biosynthetic gene clusters of other tetrahydroisoquinoline natural products. For example, the gene encoding the excision nuclease subunit UvrA is found within the saframycin A, and SF-1739, and quinocarcin gene clusters, perhaps playing a role in repairing damage induced by these potent natural products. However, the gene in the *Ca*. E. frumentensis genome is located several hundred base pairs upstream from the original ET-743 gene cluster. The saframycin A gene cluster also contains a complete gene set for the recycling of S-adenosyl methionine (SAM), a coenzyme essential for methyltransferase activity during the biosynthesis of all tetrahydroisoquinoline natural products. The complete gene set for the recycling system is still present in the *Ca*. E. frumentensis genome, but the genes are located both upstream and downstream of the original gene cluster.

The semi-dispersed nature of ET-743 biosynthetic genes is notable as microbial secondary metabolite systems are typically tightly clustered in bacteria with clearly identifiable boundaries (127, 128). However, genes involved in ET-743 biosynthesis are located in different points throughout the genome, interspersed with genes involved in primary metabolism (**Figure 3-4**). The fragmented nature of *Ca*. E. frumentensis secondary metabolism could be a consequence of horizontal gene transfer (129) and co-regulation of gene expression within operons (130), which are two important forces thought to encourage selection and formation of gene clusters. However, the

endosymbiont lifestyle provides few opportunities for horizontal gene transfer, and regulatory mechanisms are often among the first genetic elements lost during genome reduction (96, 98). The lack of selective pressure to retain clusters is thought to contribute to fragmentation of biosynthetic genes in other endosymbionts (131), and likely also plays a role in the organization of genes involved in ET-743 production.

Analysis of the *Ca*. E. frumentensis genome has also improved our understanding of the importance of ET-743 biosynthesis in the relationship between the endosymbiont and the tunicate host, *E. turbinata*. In long-term co-evolution, bacterial genes that are useful to the host are retained despite ongoing genome erosion (96, 98). The survival of ET-743 biosynthetic genes despite clear evidence of extreme genome reduction is indicative of an important role for the secondary metabolite to the host. A query of the endosymbiont genome against the full complement of bioinformatics tools (*e.g.* antiSMASH 2.0 (132), NP.searcher (133), CLUSEAN (134), BAGEL3 (135), and 2metdb (136)) revealed that ET-743 was the only natural product gene cluster found within genome, further exemplifying its ecological value to the tunicate. Adult ascidians such as *E. turbinata* are sessile marine invertebrates with soft-bodies, making them particularly vulnerable to predation. Their large larvae are released during daylight hours, making them similarly susceptible to predators. The secondary metabolite ET-743 could serve as a defense mechanism for the host. Many other ascidians and sponges are thought to produce secondary metabolites and inorganic acids that make them unpalatable to predators (137-139). Indeed, ecological studies have already demonstrated that taste and orange coloring of larvae from *E. turbinata* protects the animal against predators (137, 140). If ET-743 is the chemical deterrent responsible for protecting the host, it provides a driving force to assure ET-743 gene cluster survival despite millions of years of genome reduction.

3.3 Conclusions

We have assembled a complete genome for *Ca*. E. frumentensis, an endosymbiont responsible for production of the chemotherapeutic drug ET-743. Microbial symbionts like *Ca*. E. frumentensis have long been thought to be the source of many natural products isolated from terrestrial and marine invertebrates. However, very little is known

about the majority of these microbes due to our current inability to culture them in the laboratory.

The complete genome of *Ca*. E. frumentensis has enriched our understanding of ET-743 biosynthesis. The expanded ET-743 gene cluster will enable future biochemical studies to confirm the roles of individual enzymes. A better understanding of its biosynthesis can facilitate future *in vitro* and heterologous expression efforts to engineer sustainable production of the drug and related analogues. Analysis of the complete genome has also highlighted the importance of ET-743 to the host-symbiont relationship. The lack of genomic evidence for other secondary metabolites, the survival of the gene cluster despite extreme genome reduction, and the cluster's dispersal across the small genome suggests the microbe has become specialized for production of the drug. The chemotherapeutic natural product is therefore likely crucial to the microorganism's relationship with the tunicate host and its continued survival. This is intriguing since secondary metabolites are traditionally thought to be nonessential for microbial life (141) despite their prevalence in microbial genomes (142) and ability to confer competitive advantages (143). However, improved sequencing technologies and metagenomic pipelines are now permit more detailed studies of genomes undergoing reduction. Full genome studies on the endosymbionts found in macroorganisms like bugs (144), tunicates (131, 145), or even fungi (146, 147) provide increasing evidence that natural products may sometimes play essential roles. When these secondary metabolites benefit a host organism, their preservation may ensure a microorganism's survival and even help guide coevolution with a host. The drastically reduced genome of *Ca*. E. frumentensis presented here further supports this theory.

The pharmaceutical significance of ET-743 also demonstrates the potential potency of endosymbiont natural products. A better understanding of symbiont genomes along with their primary and secondary metabolism could provide new routes to economical and sustainable large-scale production of bioactive natural products. Analysis of the drastically reduced genome of *Ca*. E. frumentensis provides unique insight into the microorganism's lifestyle and clues to possible host-free cultivation. Previous attempts

to grow the microorganism in the laboratory were unsuccessful. However, our ability to culture elusive microorganisms is continually improving. Recent advances in host-cell free growth of *Coxiella burnetii* (111) or the facultative symbionts *Burkholderia* spp., *Rhodococcus rhodnii*, and *Wolbachia* spp. (148) provide hope that the right growing conditions and techniques can provide access to the uncultured majority of bacteria. Genome analysis in particular has proven a powerful method to pinpoint nutrient and oxygen requirements for microbial growth (29, 111, 148). The loss of key primary metabolic pathways in *Ca*. E. frumentensis suggests that the microorganism could not live independently of the host using standard media and cultivation techniques. The loss of genes involved in amino acid, coenzyme A, and glucose biosynthesis indicates that media enhanced with nutrients, cofactors, and alternative carbon sources may be necessary. However, genomic evidence for aerobic respiration and transporters for key metabolites provide hope that the right environmental conditions could one day lead to host-cell free growth.

3.4 Supplemental Information

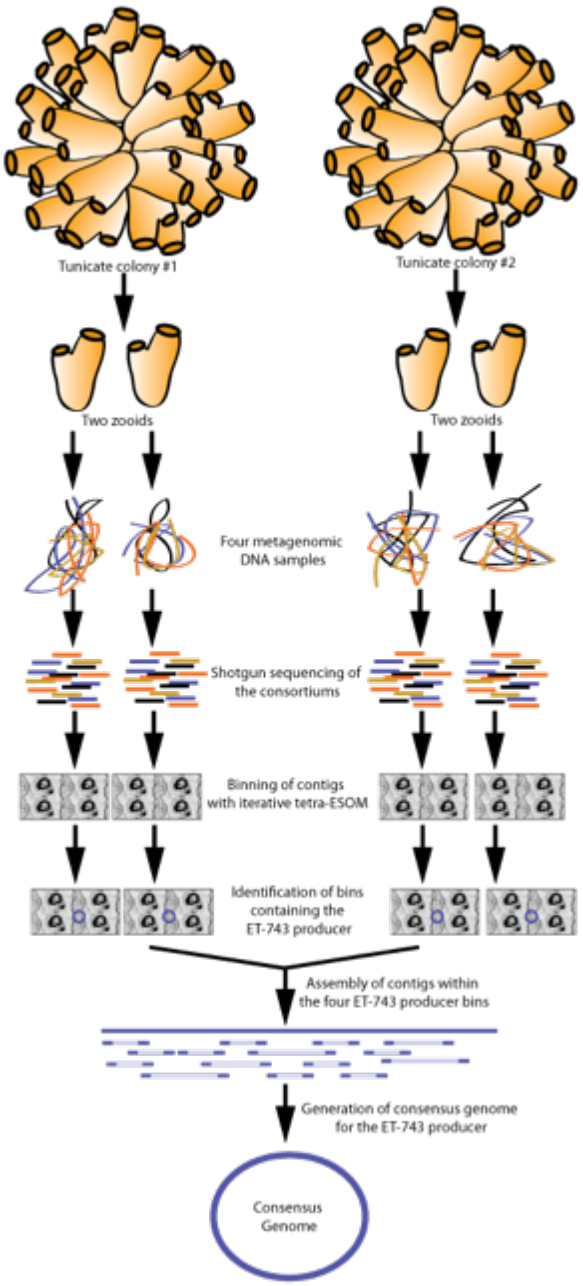


Figure 3-5. Origin of the four metagenomic DNA samples used to compile the consensus microbial genome of the ET-743 producer.
The metagenomic DNA from zooids isolated from two separate tunicate colonies was sequenced. Bins corresponding to the ET-743 producer in each metagenomic dataset were combined to create a consensus genome.

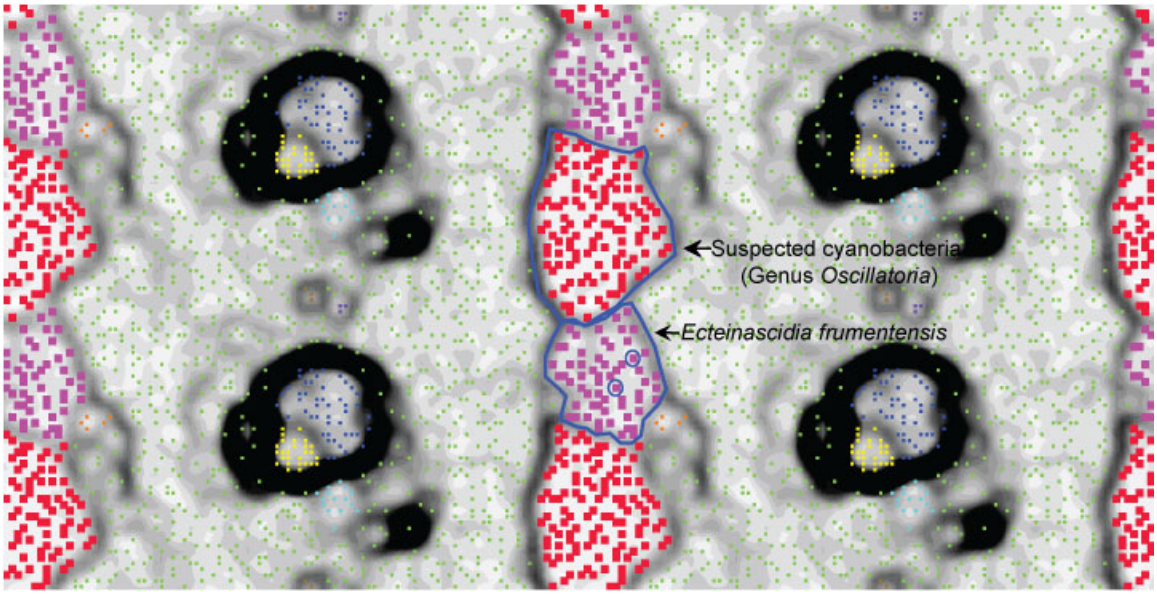


Figure 3-6. An Emergent Self Organized Map of one of the four metagenomic DNA samples (IMG Submission ID 21664) from Ecteinascidia turbinata.

Each data point represents a 5-kb sequence window, generated computationally from assembled contigs. Green data points are from unidentified contigs putatively assigned to the eukaryotic host. Data points shown in purple are from the *Candidatus* E. frumentensis bin, with circled data points contigs containing the 16S rRNA gene for E. frumentensis and the ET-743 biosynthetic gene cluster. Those in red are from a cyanobacterium present in only two of the four samples (IMG Submission IDs 21664 and 19872). The background represents Euclidean distance of tetranucleotide frequencies between data points; gray and dark colors indicate larger distances, which are used to visualize the borders between genomic bins. Borders defined for *Ca.* E. frumentensis and the Cyanobacterium are outlined in blue. Clustered data points in yellow and dark blue represent bins from unknown bacteria.

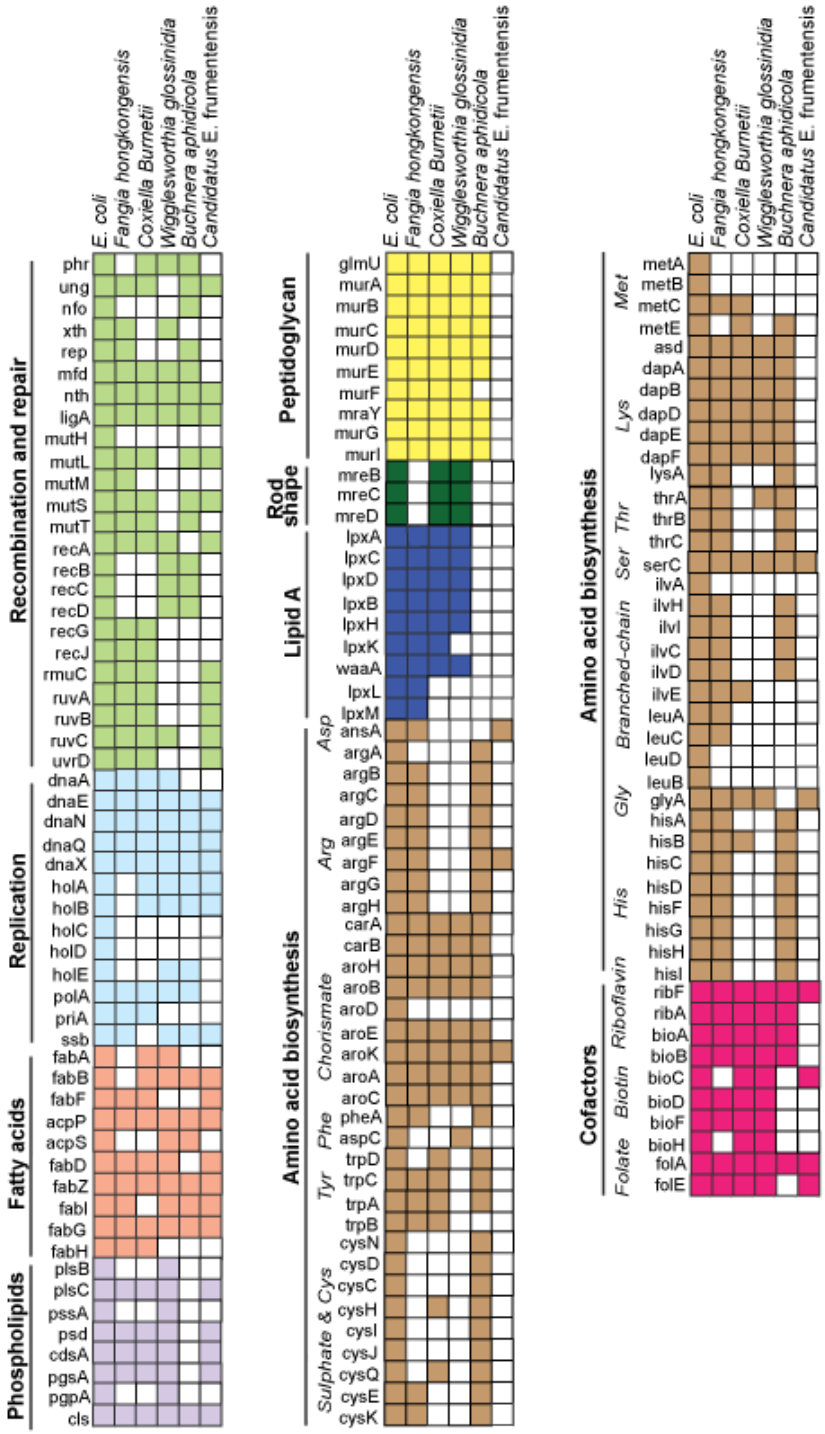
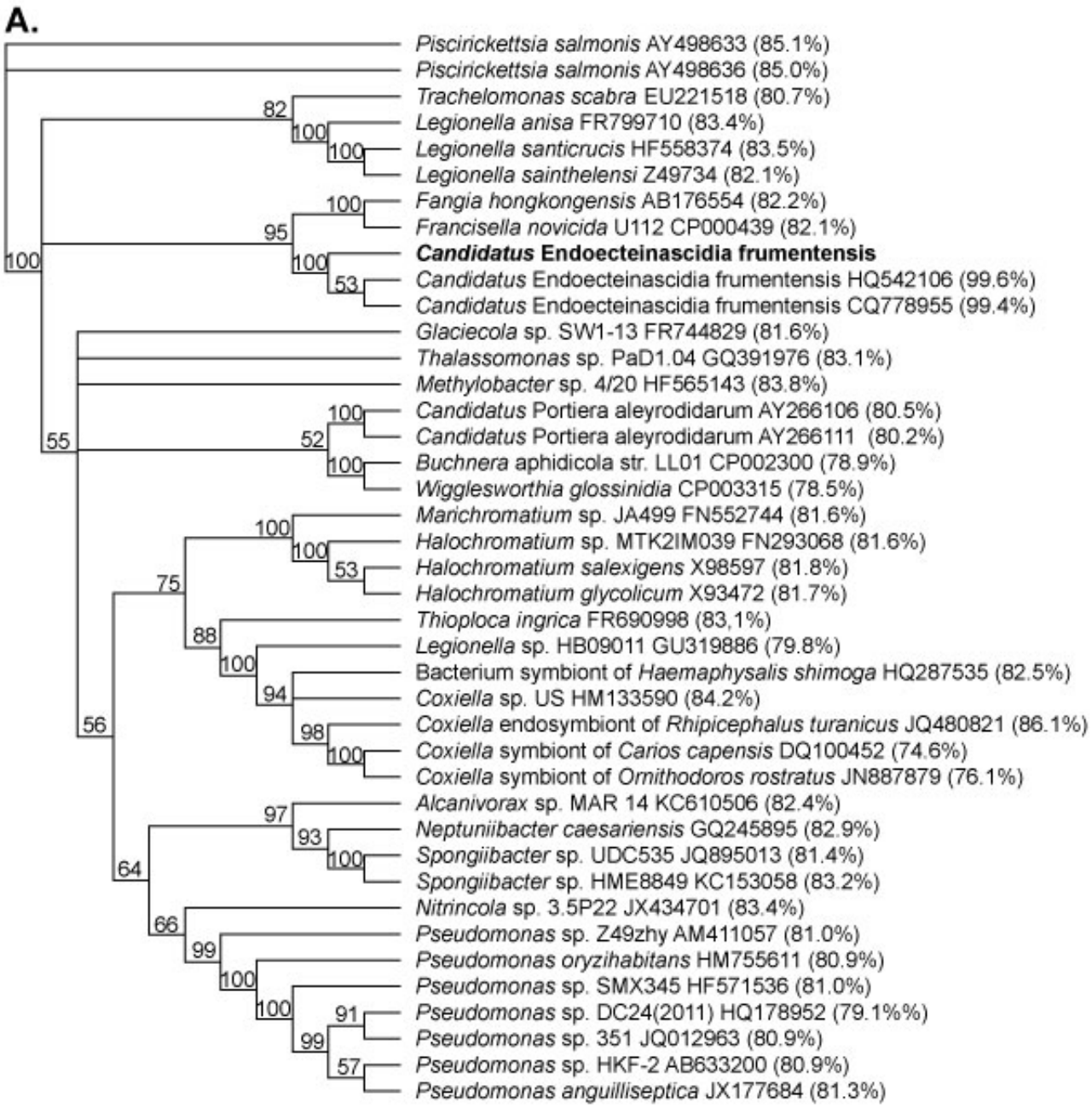
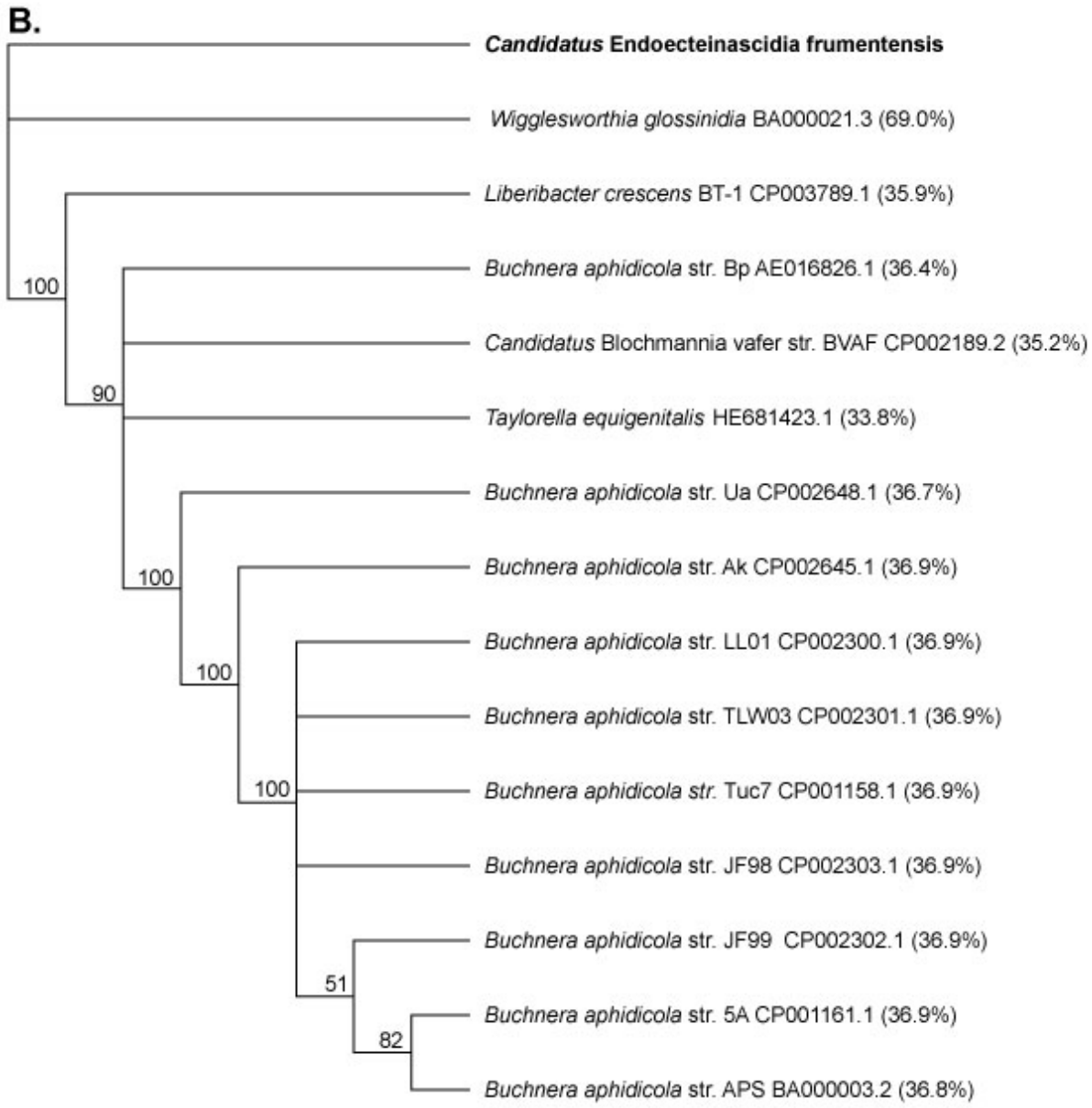


Figure 3-7. The gene content of drastically reduced genomes.
Shaded boxes represent the presence of a gene in the genome while white boxes represent its absence. The minimal gene content of *Ca*. E. frumentensis more closely resembles the reduced obligate symbiont genomes of *B. aphidicola* (NC_011834) and *W. glossinidia* (CP003315) than the intracellular pathogen *C. Burnetii* (NC_011528) or the free-living microorganisms *F. hongknongensis* (GCA_000379445.1) and *E. coli* (NC 000913).


A.
Piscirickettsia salmonis AY498633 (85.1%)
Piscirickettsia salmonis AY498636 (85.0%)
Trachelomonas scabra EU221518 (80.7%)
Legionella anisa FR799710 (83.4%)
Legionella santicrucis HF558374 (83.5%)
Legionella sainthelensi Z49734 (82.1%)
Fangia hongkongensis AB176554 (82.2%)
Francisella novicida U112 CP000439 (82.1%)
Candidatus Endoecteinascidia frumentensis
Candidatus Endoecteinascidia frumentensis HQ542106 (99.6%)
Candidatus Endoecteinascidia frumentensis CQ778955 (99.4%)
Glaciecola sp. SW1-13 FR744829 (81.6%)
Thalassomonas sp. PaD1.04 GQ391976 (83.1%)
Methylobacter sp. 4/20 HF565143 (83.8%)
Candidatus Portiera aleyrodidarum AY266106 (80.5%)
Candidatus Portiera aleyrodidarum AY266111 (80.2%)
Buchnera aphidicola str. LL01 CP002300 (78.9%)
Wigglesworthia glossinidia CP003315 (78.5%)
Marichromatium sp. JA499 FN552744 (81.6%)
Halochromatium sp. MTK2IM039 FN293068 (81.6%)
Halochromatium salexigens X98597 (81.8%)
Halochromatium glycolicum X93472 (81.7%)
Thioploca ingrica FR690998 (83,1%)
Legionella sp. HB09011 GU319886 (79.8%)
Bacterium symbiont of Haemaphysalis shimoga HQ287535 (82.5%)
Coxiella sp. US HM133590 (84.2%)
Coxiella endosymbiont of Rhipicephalus turanicus JQ480821 (86.1%)
Coxiella symbiont of Carios capensis DQ100452 (74.6%)
Coxiella symbiont of Ornithodoros rostratus JN887879 (76.1%)
Alcanivorax sp. MAR 14 KC610506 (82.4%)
Neptuniibacter caesariensis GQ245895 (82.9%)
Spongiibacter sp. UDC535 JQ895013 (81.4%)
Spongiibacter sp. HME8849 KC153058 (83.2%)
Nitrincola sp. 3.5P22 JX434701 (83.4%)
Pseudomonas sp. Z49zhy AM411057 (81.0%)
Pseudomonas oryzihabitans HM755611 (80.9%)
Pseudomonas sp. SMX345 HF571536 (81.0%)
Pseudomonas sp. DC24(2011) HQ178952 (79.1%%)
Pseudomonas sp. 351 JQ012963 (80.9%)
Pseudomonas sp. HKF-2 AB633200 (80.9%)
Pseudomonas anguilliseptica JX177684 (81.3%)


B.
Candidatus Endoecteinascidia frumentensis
Wigglesworthia glossinidia BA000021.3 (69.0%)
Liberibacter crescens BT-1 CP003789.1 (35.9%)
Buchnera aphidicola str. Bp AE016826.1 (36.4%)
Candidatus Blochmannia vafer str. BVAF CP002189.2 (35.2%)
Taylorella equigenitalis HE681423.1 (33.8%)
Buchnera aphidicola str. Ua CP002648.1 (36.7%)
Buchnera aphidicola str. Ak CP002645.1 (36.9%)
Buchnera aphidicola str. LL01 CP002300.1 (36.9%)
Buchnera aphidicola str. TLW03 CP002301.1 (36.9%)
Buchnera aphidicola str. Tuc7 CP001158.1 (36.9%)
Buchnera aphidicola str. JF98 CP002303.1 (36.9%)
Buchnera aphidicola str. JF99 CP002302.1 (36.9%)
Buchnera aphidicola str. 5A CP001161.1 (36.9%)
Buchnera aphidicola str. APS BA000003.2 (36.8%)
100
90
100
100
100
51
82

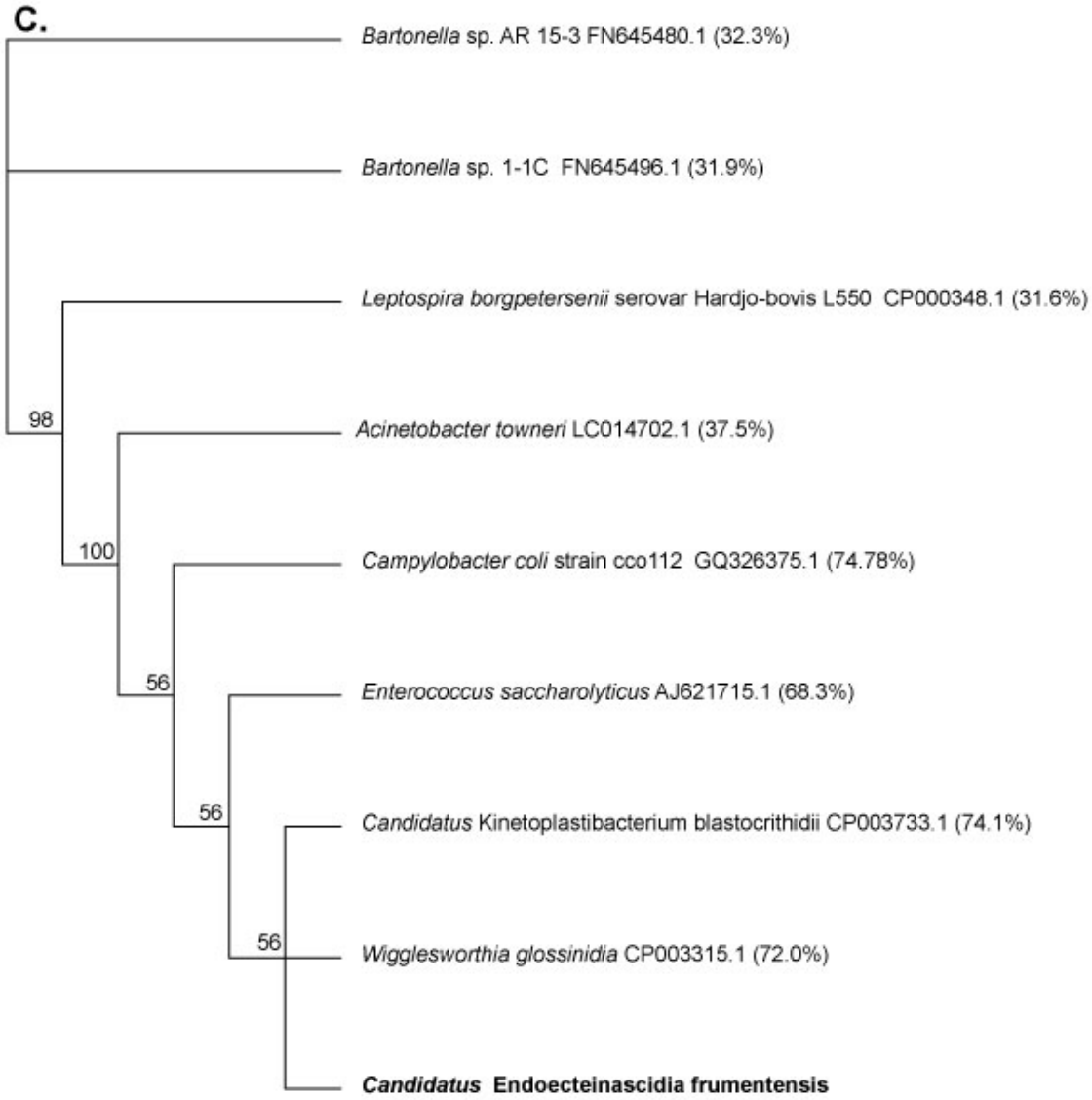


Figure 3-8. Phylogenetic reconstruction using conserved markers.
Reconstruction of A. 16S rRNA gene, B. RpoB, and C. RecA first suggested that Ca. E. frumentensis might have a novel taxonomic rank higher than the species level. Genes analyzed in this study for Ca. E. frumentensis are depicted in bold. The sequence identity between listed genes and the corresponding gene in Ca. E. frumentensis is also included. Branch labels on the bootstrapped trees represent consensus support (%).

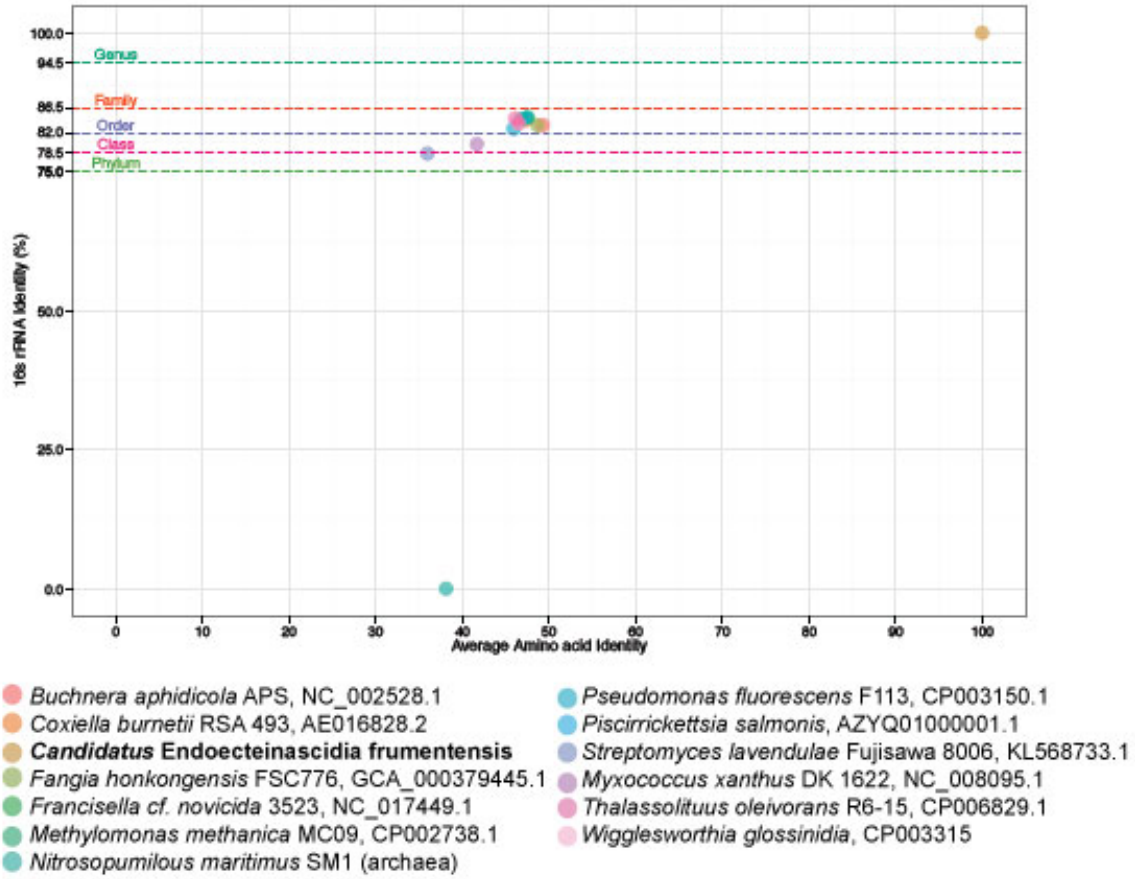


Figure 3-9. Relatedness between the 16S rRNA gene and the average amino acid identity (AAI) for Ca. E. frumentensis and similar microorganisms.
Previously described taxonomic thresholds for phylum, class, order, family, genus (106), and species are shown in dotted lines. Genomes were selected based on phylogenetic analysis of conserved genes and the taxonomic profile of the bin of the ET-743 producer. Genomes for the listed *P. fluorescens*, *S. lavendulae*, and *M. xanthus* strains were included because similar strains are associated with the tetrahydroisoquinoline natural products safracin, saframycin A, and saframycin Mx1 respectively.

Metagenome Detail	15233	15306	19872	21664
Total Assembled Bases	808,986,041	839,356,773	847,549,657	837,783,164
Longest Sequence	97,417	391,789	163,783	171,962
Shortest Sequence	200	200	200	200
Total Genes	3,184,772	3,930,810	4,966,920	2,910,719
16S rRNA genes	13	8	13	19
Bins with rRNA marker	1	1	2	3
***Candidatus* E. frumentensis Bin Detail**	**15233**	**15306**	**19872**	**21664**
Total Sequences	24	5	13	7
Total Bases	662,683	636,675	635,600	635,519
Longest Sequence	40,233	391,789	87,220	167,454
Mean Length	27,612	127,335	48,892	90.503
N50	40,233	391,789	87,220	167,454
Shortest Sequence	4,464	7,967	4,125	4,268
Estimated Completeness	100%	100%	100%	100%

Table 3-2. An overview of the four metagenomic DNA sequence datasets isolated from Ecteinascidia turbinata.
A single bin containing the ET-743 biosynthetic gene cluster and the 16S rRNA gene for *Ca.* E. frumentensis was present in every metagenomic sample. Sample 19872 and 21664 also contained a bin with an rRNA marker for an *Oscillatoriales* species. Sample 21664 contained an additional prokaryotic bin from an unknown microorganism.

Sample	Genome Coverage	Discarded Contig Coverage
Sample 15233	1619.4783	121.71451
Sample 15306	232.2210	34.82700
Sample 19872	700.9010	85.40445
Sample 21664	331.3278	54.26337
Average	720.9820	74.05233

Table 3-3. Coverage comparison between the complete Ca. E. frumentensis genome and the discarded contig.
The consistently lower coverage for the ~18 kb contig that binned with the ET-743 producer caused us to exclude it from our analysis.

Detail	*Buchnera aphidicola*	*Wigglesworthia glossinidia*	*Coxiella brunetii*	*Francisella novicida*	*Fangia hongkongensis*
Gold ID	Ga0028046	Gp0000669	Ga0028449	Ga0028838	Ga0025332
Genome Size (Mbp)	0.642	0.703	2.10	1.95	2.95
Scaffolds	1	1 + p	1 + p	1	37
GC Content	26.29	22.48	42.62	32.32	37.94
Coding Density (%)	86.53	88.50	77.71	90.96	91.42
Protein-coding genes	553	617	1,947	1,854	2,757
With functional annotation	529	131	1,231	1,339	1,971
With ambiguous function	24	486	716	515	786
rRNA genes	3	6	3	10	15
tRNA genes	32	34	42	38	38

Table 3-4. An overview of the differences in genomes of endosymbionts, intracellular pathogens, and free-living microorganisms.
The features of the draft genome of *Ca.* E. frumentensis correspond with those of obligate endosymbionts.

Possible Former Gene Product	Coordinates	
	Start	End
Preprotein translocase subunit secY	28,854	29,134
Peptide chain release factor 2	194,322	194,757
4Fe-4S ferredoxin	232,151	233,405
FAD-linked oxidoreducatse	233,604	234,101
tRNA pseudouridine synthase B	279,113	279,276
Aldehyde dehydrogenase	396,212	397,484
Dehydrogenase	397,758	398,668
Thymidylate kinase	439,535	439,798
Transcription-repair coupling factor	577,987	580,330

Table 3-5. Pseudogenes identified in the noncoding regions of the *Ca*. E. frumentensis genome.

Gene Product	Function	Putative Role	Coordinates (bp)	
			Start	End
EtuP3	Pyruvate dehydrogenase-E3 Component	Assists EtuP1, EtuP2, and EtuF9 in supplying the glycolicacyl-S-ACP extender unit (**5**) for NRPS	266,073	267,440
EtuO2	Flavodoxin reductase (ferredoxin NADPH reductase), phenol hydroxylase	Possible hydroxylation (**7**) or methylenedioxy bridge (**11**) formation	280,797	281,537
EtuM3	O-Methyltransferase	EtuA2 substrate formation (**6**) or later methylation	285,669	286,331
EtuY	Carbonic anhydrase/acetyltransferase	Acetylation to make **7**	286,333	286,884
EtuO3	Flavoprotein	Possible hydroxylation (**7**), methylenedioxy bridge (**11**), or involvement in CoA biosynthesis	286,966	287,511
EtuO4	Flavoprotein	Possible hydroxylation (**7**), methylenedioxy bridge (**11**), or involvement in CoA biosynthesis	287,508	288,164
EtuM4	N-Methyltransferase	N-methylation to make **9**	360,210	361,064
EtuF4	birA, biotin-[acetyl-CoA-carboxylase] ligase region	Fatty acid biosynthesis and/or substrate formation for the NRPS enzyme EtuA3	370,417	371,433
EtuF5	3-hydroxyacyl-[acyl-carrier-protein] dehydratase	Fatty acid biosynthesis and/or substrate formation for the NRPS enzyme EtuA3	374,717	375,166
EtuF6	phosphate:acyl-[acyl carrier protein] acyltransferase	Fatty acid biosynthesis and/or substrate formation for the NRPS enzyme EtuA3	435,106	436,137
EtuF7	malonyl CoA-acyl carrier protein transacylase	Fatty acid biosynthesis and/or substrate formation for the NRPS enzyme EtuA3	436,188	437,078
EtuF8	3-oxoacyl-[acyl-carrier-protein] reductase	Fatty acid biosynthesis and/or substrate formation for the NRPS enzyme EtuA3	437,153	437,905
EtuF9	Acyl Carrier Protein	Possible loading of glycolic acid for the NRPS enzyme EtuA1	437,952	438,188
EtuF10	beta-ketoacyl-acyl-carrier-protein synthase II	Fatty acid biosynthesis and/or substrate formation for the NRPS enzyme EtuA3	438,293	439,534

Table 3-6. New genes proposed to be involved in ET-743 biosynthesis in *Ca.* E. frumentensis.

3.5 Materials and Methods

Sample Collection and Isolation of Metagenomic DNA. Two tunicate colonies were collected off the coast of the Florida Keys. Animals were immediately frozen on dry ice after collection and stored at -80°C until processing. Metagenomic DNA was isolated from single zooids plucked from each colony (**Figure 3-5**) following the protocol outlined for mouse tails in the Wizard Genomic DNA Purification Kit (Promega). Metagenomic DNA samples were shipped on dry ice to the Joint Genome Institute (JGI) for immediate shotgun sequencing with Illumina HiSeq technology.

Genome Sequencing, Assembly, Binning and Annotation. The four metagenomic samples were sequenced and assembled by JGI. Gene calling and annotation of the assembled metagenome was then completed through JGI IMG/M (149). Individual contigs from each assembly were assigned to taxonomic groups through binning with tetranucleotide frequency with ESOM as described previously (95). Since the metagenomes had an excess of sequences belonging to the eukaryotic host tunicate, iterative rounds of ESOM were required to hone in on microbial communities present in the sample.

Genes from the previously identified ET-743 biosynthetic gene cluster (91) and the 16S gene for *E. frumentensis* (89-91) were used as BLAST queries to identify the bin containing the ET-743 producer in each of the four metagenomic samples. The four resulting bins were manually evaluated for completeness through the analysis of the distribution of conserved phylogenetic markers (150). Contigs from the four bins were assembled into a consensus genome with Geneious (v. 7.1.3).

Filing Genomic Gaps

We designed primers upstream of any suspected genomic gaps and carried out PCR using KOD Xtreme™ Hot Start DNA Polymerase (Novagen). Reactions contained 0.02 U/μL polymerase, 1X of the supplied buffer, 0.3 μM custom primers, 0.4 mM each dNTP, and 100 ng of metagenomic DNA. Reactions consisted of a hot start (94 °C, 2 min), followed by 35 cycles of denaturing (98 °C, 10 sec), annealing (variable temperatures for 30 sec),

and extension (68 °C for variable times). Since we were unsure about the size of genomic gaps, we began with a longer extension time of 5 minutes. If we saw a DNA band after running reactions on a 1% agarose gel, we repeated PCR and tailored the extension time to the size of the band (1 min/kbp) to limit any nonspecific amplification. Amplified DNA was then isolated from agarose gels using the standard protocol from the Wizard® SV Gel and PCR Clean-Up Kit (Promega).

Samples were submitted for Sanger sequencing with the primers used in the PCR reactions. Primer walking along the DNA strand then provided the missing sequence within both gaps. The complete consensus genome was submitted to JGI IMG (151) for gene calling and annotation. The final genome was reassessed for the completeness and accuracy through analysis of the distribution of conserved phylogenetic markers (150).

Genome Analysis

The common genes included in **Figure 3-7** were compiled from other studies examining genome reduction in endosymbionts and intracellular pathogens (98, 152). Analysis of primary metabolic pathways was completed using the KEGG and MetaCyc annotations provided through JGI/IMG. To confirm the absence of any missing genes, protein sequences from a model organism (typically from *E. coli* E12) were used as queries in a BLASTP search against the *Ca*. E. frumentensis annotated genome.

To detect pseudogenes, all intergenic regions larger than 100 bp were used as BLASTX queries against the entire NR database using default settings. Any hits with e-values lower than $1x10^{-3}$ against nonhypothetical proteins were considered pseudogenes.

Visualization of the complete genome (**Figure 3-2**) was constructed using Circos (153). Data for circles displaying Pfam categories for protein-coding genes, genes on the plus strands, and genes on the minus strands was provided directly through JGI IMG annotations and analysis.

Phylogenetic Analysis

The gene sequences for conserved phylogenetic markers (16S rRNA, *rpoB*, and *recA*) were used as BLASTN queries against the NT database. Trees were constructed with Geneious (v. 7.1.3) after ClustalW multiple alignments with an IUB cost matrix (default settings). Neighbor-joining trees were constructed with the Jukes-Cantor genetic distance model (default settings). Top hits for cultured or well-studied uncultured microorganisms were included in the phylogenetic tree for 16S rRNA gene sequences. All unique hits for *rpoB* and *recA* were used in respective genetic trees.

To further explore taxonomic uniqueness (**Figure 3-9**), the complete or draft genomes of the top hits from phylogenetic analysis were used in a two-way BLAST against *Ca*. E. frumentensis to acquire average amino acid identity (AAI) as previously described (105). Thresholds for unique taxonomic rankings were based on 16S rRNA gene sequence identity as previously described (106).

3.6 Notes and Author Contributions

A portion of this chapter was submitted to the *Proceedings of the National Academy of Sciences of the United States of America* during the preparation of this dissertation:

***Schofield MM**, *****Jain S, Porat D, Dick GJ, Sherman DH. "Identification and analysis of the bacterial endosymbiont specialized for production of the chemotherapeutic natural product ET-743." Submitted to *Proc Natl Acad Sci U S A.* Asterisks denote authors that contributed equally to this work.

Author contributions are as follows: Michael M. Schofield, Sunit Jain, Gregory J. Dick, and David H. Sherman contributed to the experimental design. Michael M. Schofield collected samples and isolated metagenomic DNA with assistance from Daphne Porat. Sunit Jain carried out tetra-ESOM, identified the ET-743 producing bin, and determined coverage. Michael M. Schofield filled genomic gaps and carried out secondary assemblies to acquire a consensus genome. Sunit Jain carried out analysis to determine taxonomic uniqueness. Michael M. Schofield carried out analysis of the genome, its primary metabolism, and its secondary metabolism. Michael M. Schofield, Sunit Jain, Gregory J. Dick, and David H. Sherman evaluated the data.

Chapter 4

Accessing natural products from uncultured bacteria: Initial Biochemical Investigations of ET-743 Biosynthetic Enzymes

4.1 Introduction

Accessing natural products from uncultured microorganisms can present a unique challenge. The inability to grow these organisms in the laboratory using standard techniques limits large-scale production for pharmaceutical applications. Many of these compounds are also isolated in extremely low yields from their natural environment. Therefore, harvesting plants, animals, sediments, or any environmental material the microorganisms inhabit is often economically and environmentally unsustainable. To combat these problems, the natural products of many uncultivated microorganisms are currently accessed synthetically.

However, many of these bioactive compounds have extremely complex structures, making total or even semi-synthesis a lengthy and costly process. The synthetic generation of extensive libraries to identify more potent and selective natural product analogues also presents a substantial challenge (154). It is critical that we continue to find new ways to access these medicines. Understanding the biology of the uncultured microorganisms responsible for the production of natural products provides a critical foundation for this goal.

The complete genome presented in chapter 3 elucidated the primary and secondary metabolism of *Ca*. E. frumentensis, the uncultured bacteria responsible for production of ET-743. Using this valuable information, we can begin to concurrently pursue different routes to access the chemotherapeutic drug ET-743 (**Figure 4-1**).

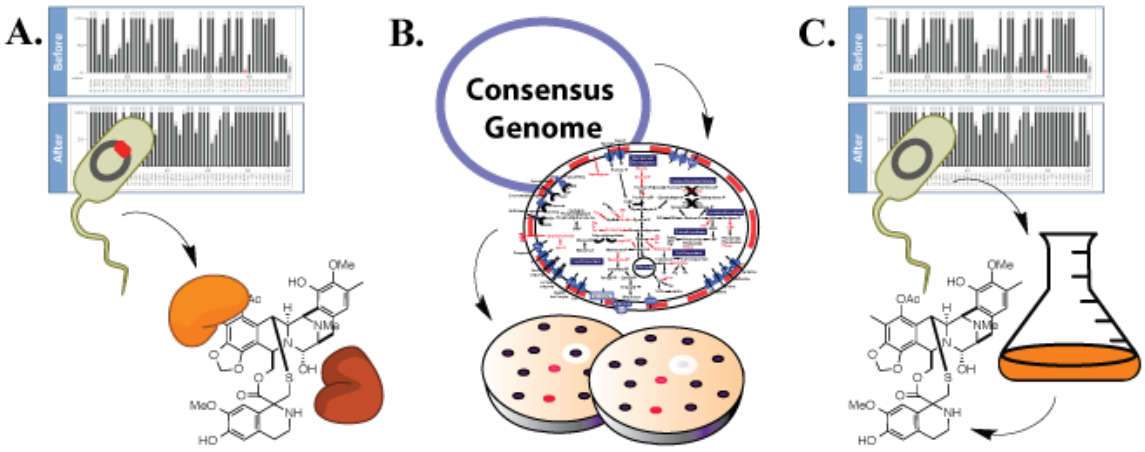


Figure 4-1. An outline of currently pursued routes to access the chemotherapeutic natural product ET-743
A.) Reconstitute the compound's core using purified NRPS enzymes *in vitro* (described in this chapter). **B.)** Utilize information gleaned from the genome of *Ca*. E. frumentensis to cultivate the microorganism in a laboratory setting (described in chapter 5). **C.)** Directly produce the natural product and new analogous in a cultured host organism through genetic engineering (described in chapter 5).

One particularly intriguing course to access natural products from uncultured microorganisms is through their *in vitro* reconstitution using key biosynthetic enzymes (**Figure 4-1A**). Enzymes called nonribosomal peptide synthetases (NRPS) are responsible for biosynthesizing the core structures of many natural products, including ET-743. Members of this fascinating class of enzymes form a complex assembly line to biosynthesize natural products that are then tailored with additional chemical groups by accessory enzymes. NRPS enzymes are organized into distinct modules that are each responsible for incorporating one amino acid or substrate into the core of the natural product. The simplest NRPS module contains 3 domains, including a condensation (C) domain, adenylation (A) domain, and a thiolation (T or PCP) domain. The A domain activates the substrate and loads it onto the T domain. The C domain then catalyzes the formation of a new peptide bond between the loaded substrate and growing natural product core. NRPS modules can additionally possess a wide variety of other domains that provide further structural diversity to the natural product.

As described in chapter 3, the core of the chemotherapeutic natural product ET-743 is likely biosynthesized by three NRPS enzymes encoded in the genome of *Ca*. E.

frumentensis (**Figure 4-2, Figure 4-3A**). These enzymes represent ideal first targets to begin work toward i*n vitro* reconstitution of ET-743 and new analogues. This is because NRPS enzyme complexes are prone to manipulation from several different routes. For example, they can frequently display substantial substrate flexibility. This allows us to feed the enzymes unnatural substrates that they can incorporate into the growing natural product core. Similar natural products also have fairly conserved NRPS modules. For example, the NRPS modules found in ET-743 are similar to those found in other tetrahydroisoquinoline natural products (**Figure 4-2**). This allows us to mix and match enzymes from different systems to create unique natural product hybrids. The natural product cores produced by NRPS enzymes can also be combined with semi-synthetic methods to create unique compounds through chemoenzymatic synthesis (154, 155).

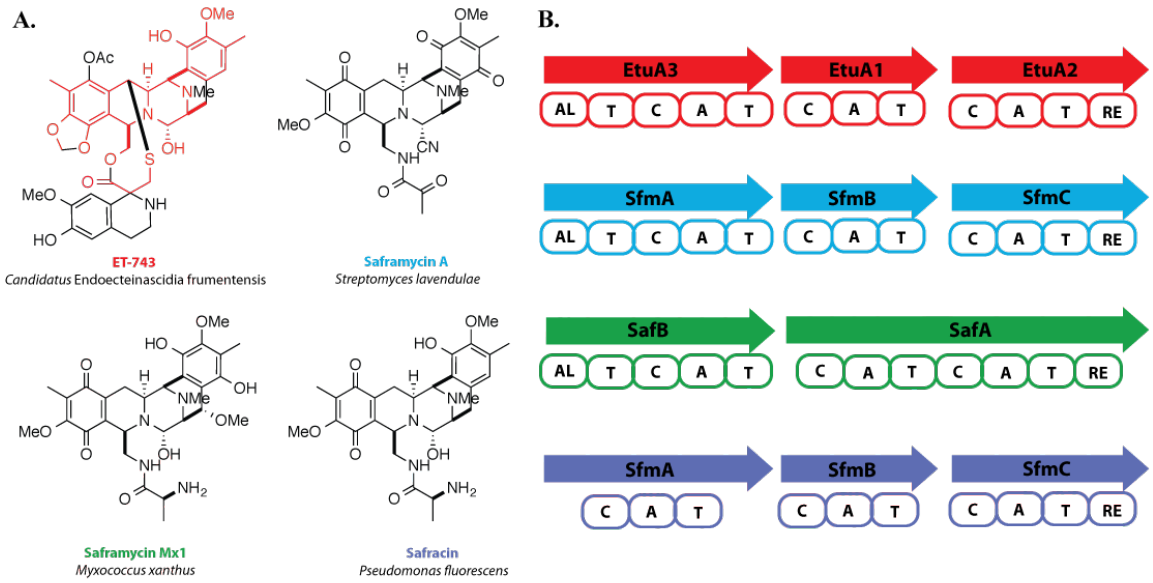


Figure 4-2. Four tetrahydroisoquinoline natural products and the conserved NRPS modules responsible for the biosynthesis of their cores.
A.) The structures of the four best-studied tetrahydroisoquinoline natural products, their names, and the microorganism responsible for their production. The NRPS enzymes EtuA1-3 are responsible for construction of the core (shown in red) in ET-743. **B.)** The NRPS modules responsible for production of tetrahydroisoquinoline cores are conserved between microorganisms. AL-acyl ligase; T-thiolation; C-condensation; A-adenylation; RE-reductase.

The complete genome of *Ca*. E. frumentensis and studies of other tetrahydroisoquinoline natural products (121-123) helped us understand how these three NRPS enzymes likely work together to biosynthesize ET-743 (**Figure 4-3**). We predict that the NRPS

dimodule EtuA3 (AL-T-C-A-T) initiates production of the ET-743 core by loading an acyl chain onto the adjacent T domain. The second module of the enzyme then activates and loads a cysteine onto the growing chain. The growing natural product core is then passed to the NRPS EtuA1which incorporates a glycolic acid, possibly with the help of the transketolases EtuP1-3 and a downstream acyl carrier protein (ACP). The reductase (RE) domain of EtuA2 is then thought to terminate the reaction and release the nonribosomal peptide. The remaining domains of EtuA2 then combine the product from EtuA3 and EtuA1 with a tyrosine derivative (3-hydroxy-4-O-methyl-5-methyl-tyrosine, 3H-4O-Me-5Me-Tyr) in multiple rounds of a cyclizing Pictet-Spengler reaction to create the final ET-743 natural product core (**Figure 4-3**).

In this investigation, we sought to take the first steps toward *in vitro* reconstitution and manipulation of the tetrahydroisoquinoline core of ET-743. Codon optimization enabled us to clone the NRPS genes and express the enzymes responsible for production of the molecule's core. We then selected the most promising of these enzymes (EtuA1) for analysis. Our biochemical study of the enzyme provided the first experimental evidence of its role in ET-743 biosynthesis. Analysis of its substrate flexibility also highlighted the potential to use the enzyme to generate new ET-743 analogues in future studies.

This investigation provides a firm foundation for continuing efforts to reconstitute the ET-743 biosynthetic core *in vitro* with these enzymes. This route could prove an important method to access ET-743 and new analogues. The beginning stages of accessing the natural product though other routes, including cultivation of the producing organism (**Figure 4-1B**) and engineering of genomes from *Streptomyces* species (**Figure 4-1C**) are described in chapter 5.

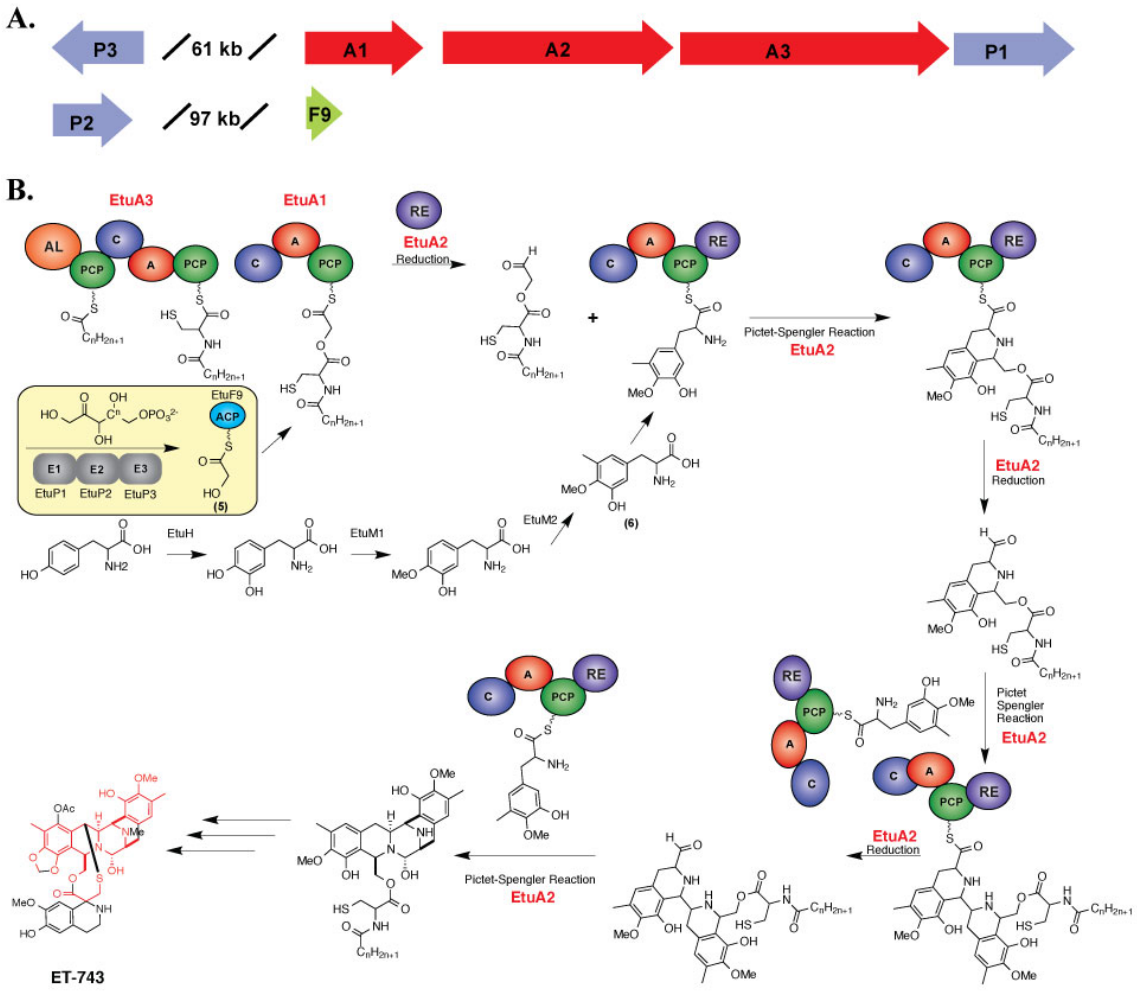


Figure 4-3. Biosynthesis of the ET-743 biosynthetic core.
A.) Putative genes involved in biosynthesis of the ET-743 biosynthetic core. **B.)** The corresponding gene products and their predicted roles in ET-743 biosynthesis. NRPS enzymes are shown in red. The pyruvate dehydrogenase enzymes EtuP1-P3 (purple) possibly acquire glycolic acid from a primary metabolic pathway and deliver it to EtuA1 using the ACP EtuF9 (green). Contributions by the NRPS enzymes are shown in red in the final structure for ET-743.

4.2 Results

4.2.1 Codon optimization permits expression and purification of enzymes involved in ET-743 biosynthesis.

Attempts to heterologously express ET-743 enzymes in *E. coli* and many other host organisms were largely unsuccessful for many years (91) despite trying multiple methodologies and exploring various host organisms. We never observed protein expression, although genes always amplified well from tunicate metagenomic DNA and were typically easy to clone into expression vectors.

Previous metagenomic analysis (91) and our work in chapter 3 suggested the likely cause of this is the extremely low %G+C content of *Ca*. E. frumentensis DNA. The genome possesses an overall 23.3% G+C content, with 24.2% found in the coding regions and 12.7% found in the noncoding regions. Every amino acid is capable of being encoded by more than one codon, and different organisms can have vastly differing codon preferences. One of the driving forces that determines these codon preferences is the G+C makeup of the genome (156).

We examined the codons used by the three NRPS genes and compared them to codon preferences of varying host organisms, including *E. coli*, yeast species, and varying species of *actinomycetes*. The results illustrated drastic codon differences and suggested that codon optimization could finally enable expression of ET-743 biosynthetic enzymes in heterologous hosts (**Figure 4-4**).

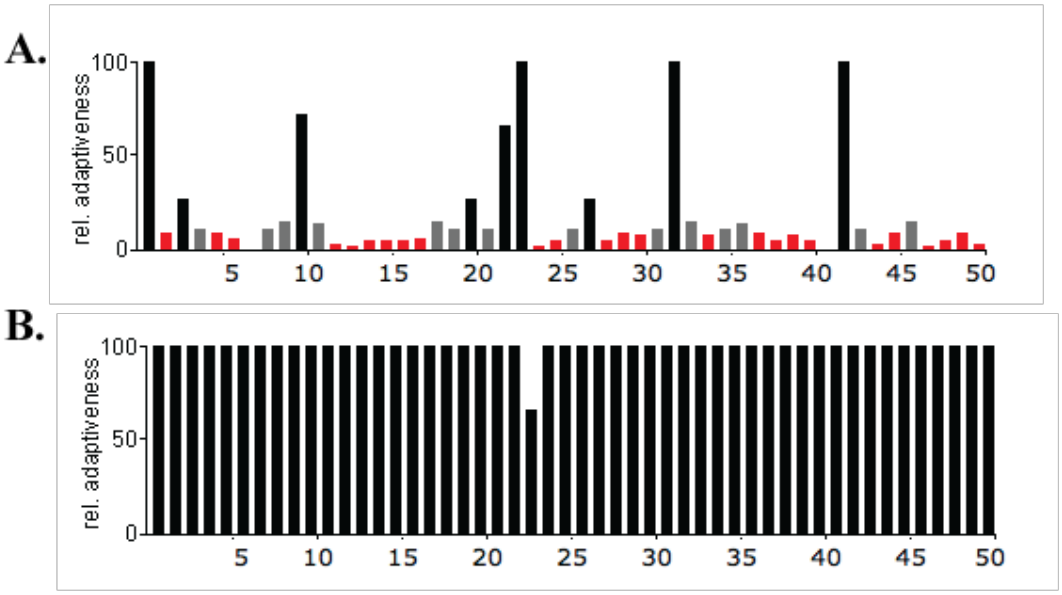


Figure 4-4. *Ca*. E. frumentensis utilizes different codons from many commonly used heterologous hosts.
The relative adaptiveness of codons for the first 50 amino acids in EtuA1 **A.)** before codon optimization of the gene and **B.)** after codon optimization of the gene for expression in *E. coli*. Red bars indicate poor codon agreement while black bars indicate good agreement.

We therefore purchased codon optimized fragments of the three ET-743 NRPS genes and utilized Gibson assembly to piece the fragments together and clone them into various *E. coli* protein expression vectors (**Table 4-1 in Supplemental Information [SI**]). This resulted in expression of each NRPS enzyme for the first time (**Figure 4-8 in SI**).

Although proteins expressed well, we next encountered solubility issues during the purification process. Utilizing an expression vector possessing the maltose binding protein (MBP) solubility tag, adding the chemical chaperone trimethylamine N-oxide (TMAO), and utilizing other solubilizing reagents partially corrected these solubility problems. However, purification of the enzymes by both nickel and amylose affinity chromatography was still suboptimal, (**Figure 4-9 in SI**) and yields were too low to attempt additional purification steps, such as size-exclusion chromatography.

Previously published (157) and unpublished work in our laboratory demonstrated that clean protein preps are not always necessary to conduct biochemical analysis or complete chemoenzymatic studies. We can still conduct simple biochemical analyses and generate

natural products using suboptimal protein preparations or even whole cell extracts. In fact, these methods can actually facilitate easier production of natural products on a larger scale.

We therefore decided to select the best behaving of the three ET-743 NRPS enzymes and continue our analysis using its suboptimal protein preparations. Of the three enzymes, EtuA1 expressed the best, was the most soluble, and was the easiest to purify (**Figure 4-9**). Our biochemical and substrate flexibility analysis with this enzyme represent a model that can be used for future work with EtuA2 and EtuA3.

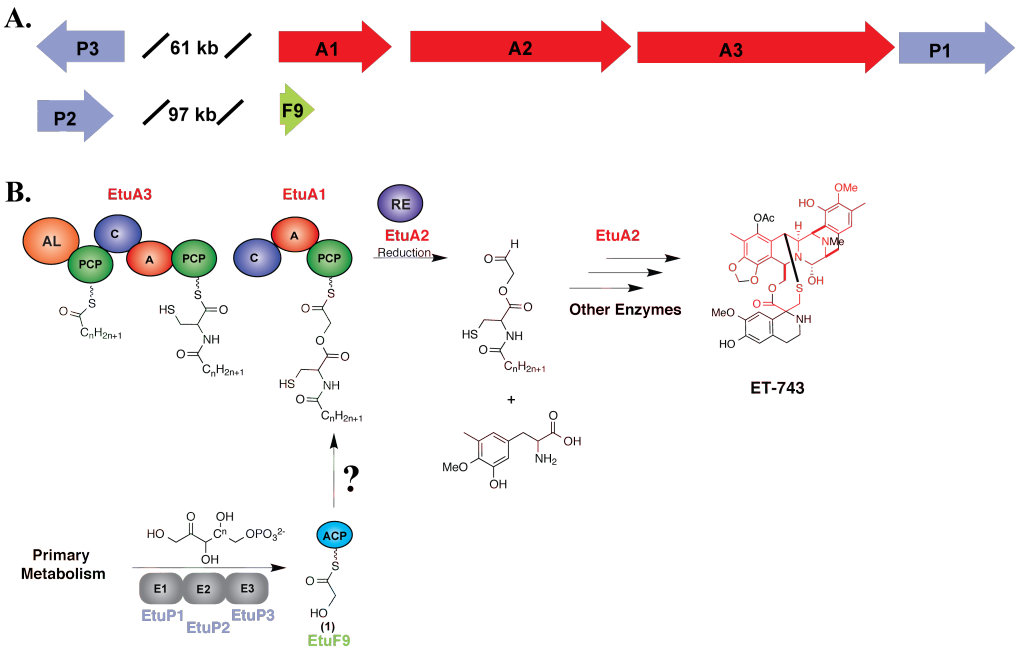


Figure 4-5. The role of EtuA1 in the biosynthesis of ET-743
A.) Putative genes involved in biosynthesis of the ET-743 biosynthetic core, including EtuA1. **B.)** The corresponding gene products and their predicted roles in ET-743 biosynthesis, with a focus on the role of EtuA1. We predict the enzyme incorporates glycolic acid into the growing natural product. However, the role of EtuP1-3 and EtuF9 is unknown.

4.2.2 Biochemical assessment of EtuA1, an NRPS that selectively activates glycolic acid.

Based on analysis of the biosynthetic gene cluster presented in chapter 3, we predicted that EtuA1 is likely responsible for incorporating glycolic acid into the core scaffold of ET-743. However, it was unclear what role the transketolases EtuP1-3 and the ACP

EtuF9 played in helping EtuA1 incorporate glycolic acid (**Figure 4-5**). Previous evidence in another tetrahydroisoquinoline system suggested that these enzymes may be necessary for successful NRPS incorporation of glycoaldehyde units into natural products (126). Similar enzymes in the biosynthesis of quinocarcin, naphthyridinomycin, and SF-1739 acquire the glycoaldehyde unit from a primary metabolic pathway, load it onto a separate ACP, and incorporate it directly onto the NRPS module (124, 126). If this was the case with EtuA1, we would expect the A domain would be incapable of activating glycolic acid on its own.

To test the ability of the A domain of EtuA1 to activate and load glycolic acid, we again utilized the malachite green assay outlined in chapter 2. Briefly, A domains recognize and activate a substrate using the coenzyme adenosine triphosphate (ATP). During the enzymatic reaction, they release pyrophosphate (PP_i). The malachite green assay utilizes inorganic pyrophosphatase to cleave PP_i which is then able to complex with a quenching solution to produce a colorimetric change. This permits us to visualize whether NRPS and related enzymes are capable of activating substrates. After significant optimization, the malachite green assay was found to be robust for use with EtuA1 with a Z-factor of 0.574 (**Figure 4-10 in SI**).'

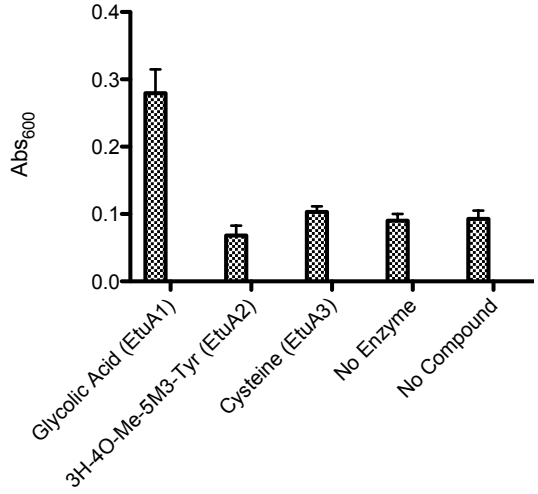


Figure 4-6. EtuA1 preferentially activates glycolic acid over other predicted substrates
The optimized malachite green assay was utilized with EtuA1 using glycolic acid, a tyrosine derivative (predicted substrate for EtuA2), and cysteine (predicted substrate for EtuA3). The enzyme preferentially activated glycolic acid over other substrates, indicated by the significantly higher absorbance readings.

The optimized assay was used to test the ability of the A domain of EtuA1 to accept glycolic acid over 3H-4O-Me-5M3-Tyr and cysteine, the predicted substrates for EtuA2 and EtuA3 respectively. The enzyme was shown to activate glycolic acid, but not the tyrosine derivative or cysteine (**Figure 4-6**). This provided the first experimental evidence that EtuA1 is indeed involved in incorporating glycolic acid into the final structure of ET-743. It also suggests that the role of EtuA1 may be unique from those of similar enzymes found in other tetrahydroisoquinoline biosynthetic pathways. The NRPS may not require direct interaction with the transketolase enzymes and an ACP to activate and load glycolic acid units.

4.2.3 EtuA1 exhibits some promiscuity in substrate activation

We were next interested in examining the substrate flexibility of EtuA1. Probing the ability of the A domain to activate unnatural substrates could help us predict the feasibility of using *in vitro* enzymatic reactions to create new ET-743 natural product

analogues. We therefore selected a variety of substrates (**Figure 4-7A**) to test against the enzyme.

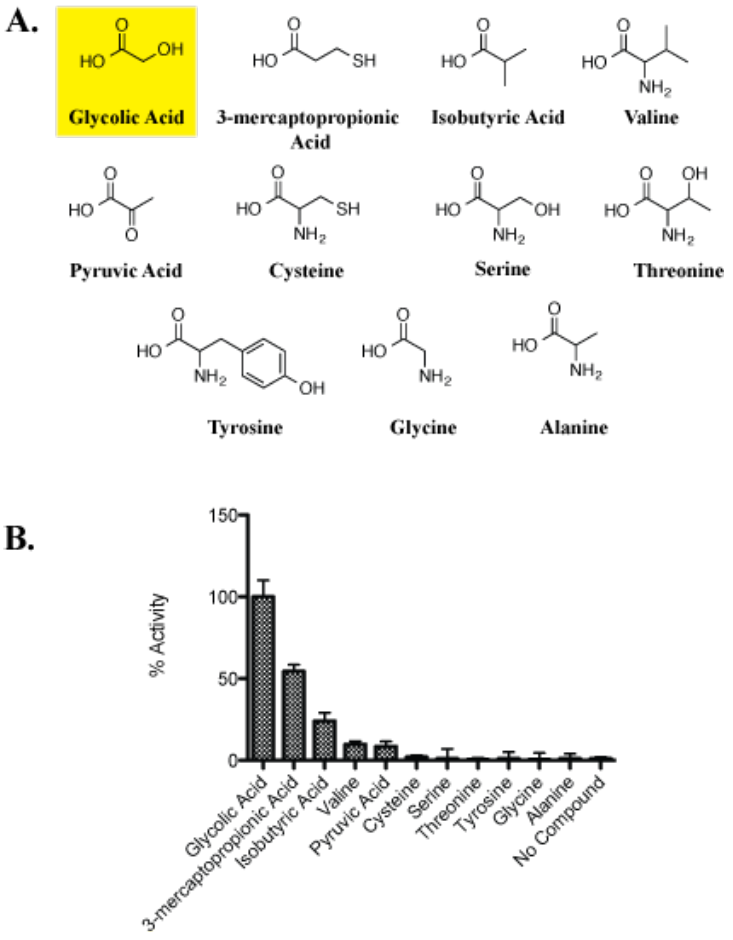


Figure 4-7. Examination of the substrate flexibility of EtuA1
A.) The structures of the various substrates tested against EtuA1. The natural substrate, glycolic acid, is highlighted in yellow. **B.)** The malachite green assay was used to assess the ability of the EtuA1 A domain to activate substrates. Data are normalized to the preferred substrate, glycolic acid.

EtuA1 preferentially activated glycolic acid over other tested substrates (**Figure 4-7B**). However, the A domain did seem capable of activating a few unnatural substrates, particularly 3-mercaptopropionic acid and isobutyric acid. This demonstrates that the enzyme is capable of some promiscuity and could potentially incorporate intriguing chemical groups into ET-743 to generate new analogues.

4.3 Conclusions

In vitro reconstitution of key biosynthetic enzymes can be an effective method to access natural products. Manipulation of individual enzymes with unnatural substrates also presents the potential to generate novel analogues with better potency or selectivity for molecular targets. In this investigation, we have taken the first steps toward *in vitro* reconstitution of the ET-743 core by examining one of the key enzymes involved in its biosynthesis.

We presented biochemical evidence that EtuA1 is directly capable of activating glycolic acid, unlike its relatives in the quinocarcin, naphthyridinomycin, and SF-1739 tetrahydroisoquinoline systems (124, 126). Although EtuA1 is obviously capable of activating a glycolic acid substrate, it is still unclear whether or not the enzyme can successfully load the substrate and condense it onto the growing peptide chain. We need to complete further studies with EtuA1 and the remaining NRPS enzymes to answer these questions.

Studies of other NRPS systems have illustrated that it is possible for some enzymes to directly load α-hydroxyacids onto nonribosomal peptides without assistance from transketolases (158). Alternatively, other systems load α-ketoacids that are then directly reduced to a α-hydroxyacid by a ketoreductase while tethered to the assembly line (159, 160). What method EtuA1 uses to incorporate glycolic acid remains unclear. However, the close proximity of the transketolase genes to the NRPS genes (**Figure 4-5**) suggests they likely play some important role, if only acquiring glycolic acid substrate from a primary metabolic pathway for direct use by the NRPS.

The substrate flexibility of EtuA1 also validates that *in vitro* enzymatic studies could provide an effective route for the generation of new ET-743 analogues. The ability of the enzyme's A domain to activate unnatural substrates suggests that the enzyme could be capable of incorporating unusual units into the final structure of ET-743. The substrates tested against EtuA1 in this investigation were by no means extensive, and it would be

interesting to explore what other types of chemical groups the enzyme is capable of recognizing.

In summary, this investigation provides a firm foundation for further efforts to reconstitute the ET-743 core scaffold *in vitro* using purified NRPS enzymes. The methodologies developed here with EtuA1 provide an effective roadmap for future efforts with EtuA2 and EtuA3. With further work, these techniques could provide direct access to ET-743 and new analogues, bypassing the costly and lengthy synthetic route currently used by pharmaceutical companies.

4.4 Supplemental Information

Gene Product and Predicted Function	Vector	Expression	Best Purification Conditions	Solubility
EtuA1, NRPS (Glycolic Acid)	pET30	Yes, Codon Optimized Only	TMAO, AI cells	Poor
	pMCSG7	Yes, Codon Optimized Only	TMAO, AI cells	Poor
	pTrc9	Yes, Codon Optimized Only	TMAO, Nickel over amylose affinity	Decent
EtuA2, NRPS (Tyr derivative)	pMCSG7	Yes, Codon Optimized Only	TMAO, AI cells	None
	pTrc9	Yes, Codon Optimized Only	TMAO, Nickel over amylose affinity	Poor
EtuA3, NRPS (Cysteine)	pMCSG7	Yes, Codon Optimized Only	TMAO, AI cells	None
	pTrc9	Yes, Codon Optimized Only	TMAO, Nickel over amylose affinity	Poor

Table 4-1. Cloned constructs for expression of EtuA1-3

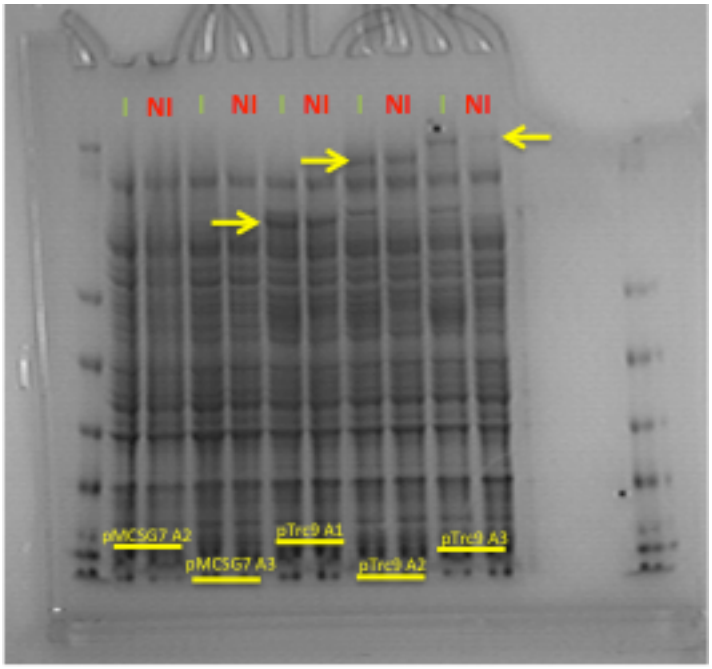


Figure 4-8. Expression of various ET-743 biosynthetic enzymes
The vector pTrc9 with an MBP fusion tag produced the best expression of EtuA1-3 in *E. coli* Bap1 cells. This gel is an expression test and depicts whole cell extracts. Overexpression of each protein is depicted by the yellow arrows in cultures that have been induced (I) and not induced (NI) with isopropyl β-D-1-thiogalactopyranoside (IPTG). The T7 promoter in pTrc9 is a little leaky, causing some expression in non-induced cultures.

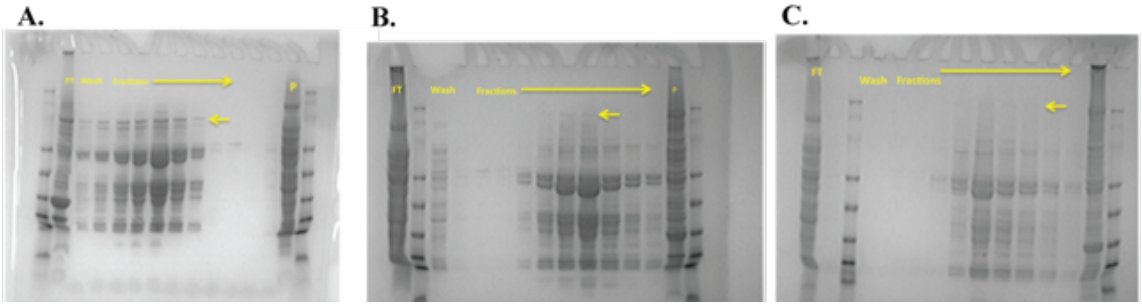


Figure 4-9. Suboptimal purification of EtuA1-3 NRPS modules
EtuA1-3 displayed significant solubility problems during the protein purification process. Although enzymes were far from pure, further purification techniques like size exclusion chromatography would have a negative impact on the yield of target proteins. We selected EtuA1 **(A)** over EtuA2 **(B)** and EtuA3 **(C)** because it provided the best yield and the enzyme was the most soluble to work with. Left facing arrows depict target enzymes, confirmed with western blots against the 6x poly-histidine tag (data not shown).

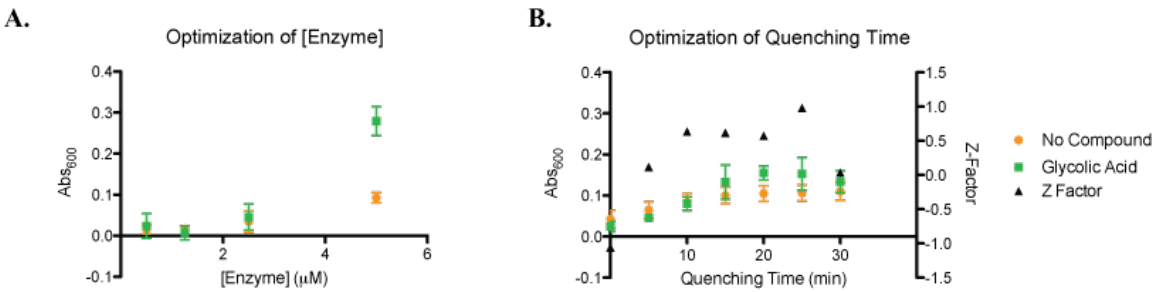


Figure 4-10. Optimization of the malachite green assay for biochemical analysis of EtuA1
A.) We tested various concentrations of enzyme during assay optimization. 5 μM produced the best Z factor and was used for final assay conditions. **B.)** We also tested varying incubation times and the optimal time after quenching to read the assays. We selected 20 minutes after quenching for reading, based on its excellent Z factor (0.574) and consistently lower standard error.

4.5 Materials and Methods

Cloning and protein purification.

The *etuA1*, *etuA2*, and *etuA3* genes were codon optimized for expression in E. coli using the Graphical Codon Usage Analyzer (GCUA) and ordered in fragments from Gene Strings (Invitrogen). The codon-optimized fragments were then assembled into various vectors with LIC overhangs using Gibson Assembly Master Mix (NEB).

To obtain purified protein from constructs, we attempted various expression and purification conditions with different vectors, cell lines, buffers, and reagents. Conditions that achieved the best results (**Figure 4-9 in SI**) included expression in the pTrc9 with an MBP fusion tag in Bap1 cells. Cells were grown in Terrific Broth (TB) at 37 °C with 200 mM TMAO until they reached a high OD (~2.0, ideal for toxic protein products). Cultures were then cooled to 18 °C, induced with 1 mM IPTG, and left to express for 6 hours.

Harvested cell pellets were lysed by sonication in a buffer containing 200 mM TMAO, 40 mM imidazole, 20 mM HEPES, 300 mM NaCl, 1 mM tris (2-carboxylethyl) phosphine (TCEP), 20% glycerol, a cocktail of protease inhibitors (EDTA-free Complete, Roche), 400 mg CellLytic Express (Sigma), and ~5 mg of lysing enzymes from *Trichoderma harzianum* (Sigma L1412). The cell-free extract was collected by ultracentrifugation at 45,000 x g for 45 minutes. The supernatant was subjected to flash

Ni^{2+} affinity chromatography after batch binding with nickel resin for 2 hours. Resin was rinsed with 100 mL of the lysis buffer before elution with lysis buffer including 300 mM imidazole. Protein preps were immediately frozen in the high imidazole buffer and stored at -80°C until analysis since any kind of buffer exchange system would result in protein yield loss to some degree.

Malachite green assays. Malachite green assays were based off of previous work (described in chapter 2). However, we optimized conditions to maximize the efficiency of the assay with EtuA1. Assays were performed in a 384-well plate format with a final volume of 40 µL. Enzymatic activity was found to be most uniform across clear polystyrene microplates produced by Greiner Bio-One (781185). Standard final conditions were 20 mM HEPES, pH 7.5, 5 mM $MgCl_2$, 2.25 mM ATP, 15 mM of various substrates, 0.2667 U/mL inorganic pyrophosphatase (IPP), and 5 µM EtuA1. Reaction mixtures were split into two solutions: a substrate solution and an enzyme solution to initiate reactions. First, 30 µL of a substrate mix containing HEPES, $MgCl_2$, ATP, substrate, and IPP was dispensed into the 384-well plate. To initiate the reactions, 10 µL of an enzyme solution containing HEPES and purified EtuA1 was added to the reactions. Reactions were incubated for 1.5 hours at room temperature before quenching with a malachite green solution as described in chapter 2. Quenched reactions were incubated for 20 minutes before reading at Abs_{600}.

4.6 Notes and Author Contributions

The work described in this chapter is ongoing. Michael M. Schofield and David H. Sherman contributed to experimental design. Michael M. Schofield conducted all cloning and experiments with some assistance from undergraduate students, including Daphne Porat. Chemist and Sherman laboratory member Jennifer Schmidt provided invaluable advice on initial substrates to test against EtuA1.

Chapter 5
Discussion and Future Work

5.1 Overview

Microbial natural products are a critical source for novel medicines in the ongoing war against drug resistant microorganisms, viruses, fungi, and cancers. To maximize the discovery of new chemical scaffolds, we need to concurrently explore the natural products of both cultured and currently uncultured microorganisms. This dissertation has outlined various methodologies that can facilitate access to medicines from both of these microbial sources (**Figure 5-1**).

In chapter 2, I described our work to find novel natural products from cultured bacteria. We first identified an ideal drug target—an enzyme that helps construct a molecule involved in microbial iron acquisition. We then used this drug target to screen an extensive library composed of extracts from thousands of cultured microorganisms for potential new natural product medicines. This led to the identification of a novel antibiotic scaffold that we demonstrated could be used as a broad-spectrum antibiotic. This targeted high throughput screening technique could be a powerful method to continue to identify new drugs to fight off resistant infections.

In chapters 3 and 4, I described our work to find new natural product medicines from uncultured bacteria. For both of these chapters, we selected the chemotherapeutic drug ET-743 as a model system. Researchers have long suspected that an uncultured microbe was responsible for the drug's production. Decades' worth of work into ET-743 and other tetrahydroisoquinoline natural products provided a perfect foundation and knowledgebase for our work.

In chapter 3, we first sought to identify and understand the uncultured microorganism responsible for ET-743 production. We used metagenomic methodologies to uncover the

complete genome of *Ca*. E. frumentensis. We provided strong evidence that the microbe is an endosymbiont inhabiting the cells of its mangrove tunicate host. We also conducted an in-depth analysis of its primary and secondary metabolism, providing a thorough look at the endosymbiont's lifestyle and its biosynthetic pathway. This understanding provided an essential knowledgebase that we could then use to begin to develop new routes to access the elusive natural product.

In chapter 4, we began to explore one of these routes. We heterologously expressed key ET-743 biosynthetic enzymes for the first time. We then biochemically analyzed one of these key enzymes (EtuA1) and provided evidence for its predicted role in ET-743 biosynthesis. We also provide data that suggests that its role may be different than similar enzymes in other tetrahydroisoquinoline systems. In addition, we explored the substrate flexibility of the enzyme. Our experiments in this chapter illustrate that EtuA1 and the rest of the NRPS assembly line could be used to reconstitute the ET-743 core *in vitro* and possibly to generate new analogues.

In this chapter, I will explore our ongoing and future work for the baulamycin and ET-743 projects.

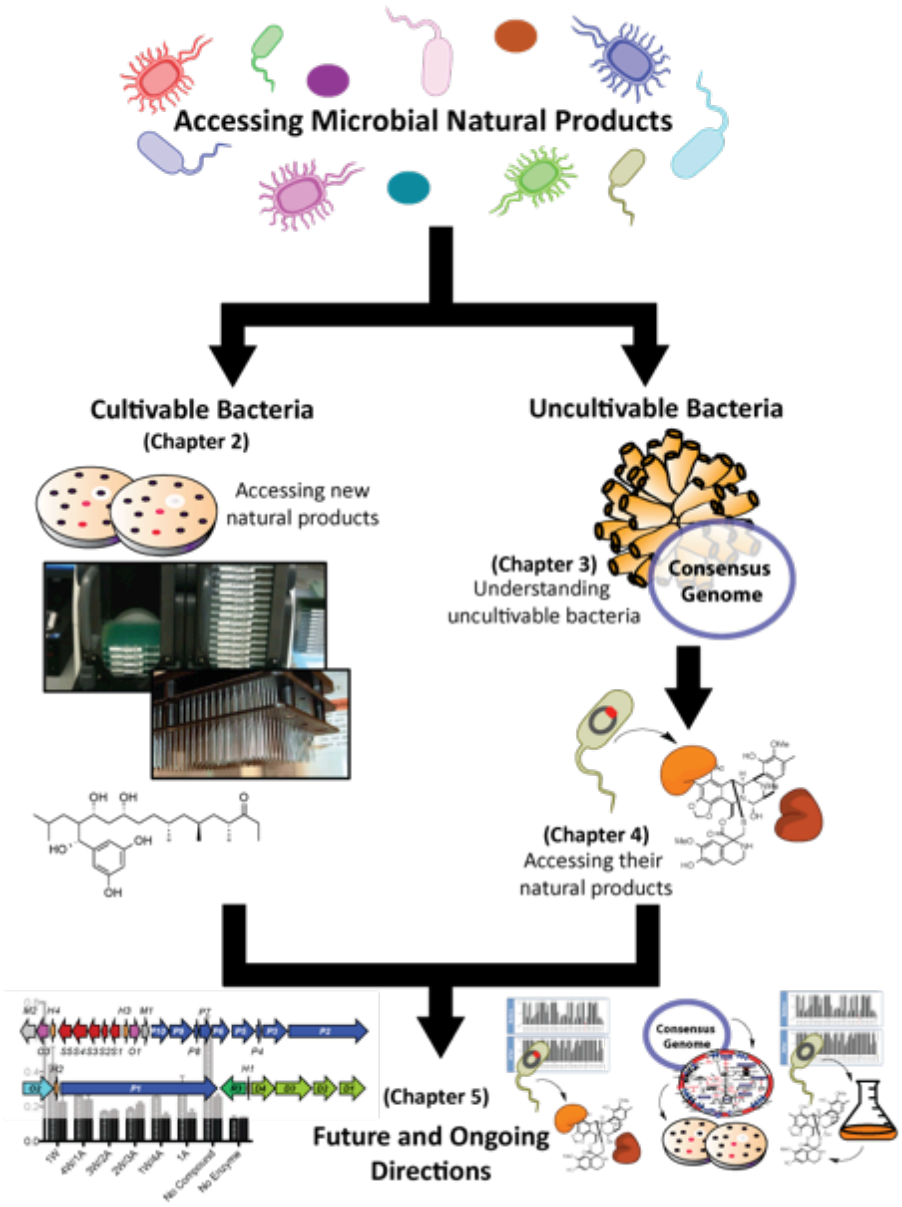


Figure 5-1. A review of my dissertation
Chapter 2 described efforts with a targeted high throughput screen from an extract library to identify novel natural products. Chapters 3 and 4 outlined techniques to access the products of uncultured microorganisms. In chapter 3, we utilized metagenomic methods to understand an uncultured microorganism. In chapter 4, we used this new information to begin work that could one day enable us to access its natural product. Here in chapter 5, we describe future directions and ongoing work for both of these projects.

5.2 Ongoing and future work for cultured microorganisms

5.2.1 Identification and manipulation of the baulamycin gene cluster to increase antibiotic production and make more potent analogues

The strain that produces the baulamycin antibiotics described in chapter 2 (**Figure 5-2A**) is cultured in the laboratory using standard techniques. The genes responsible for natural product biosynthesis are usually tightly clustered within genomes. The ability to cultivate the baulamycin-producing strain therefore provides direct access to the antibiotics and the ability to easily create more potent and selective analogues.

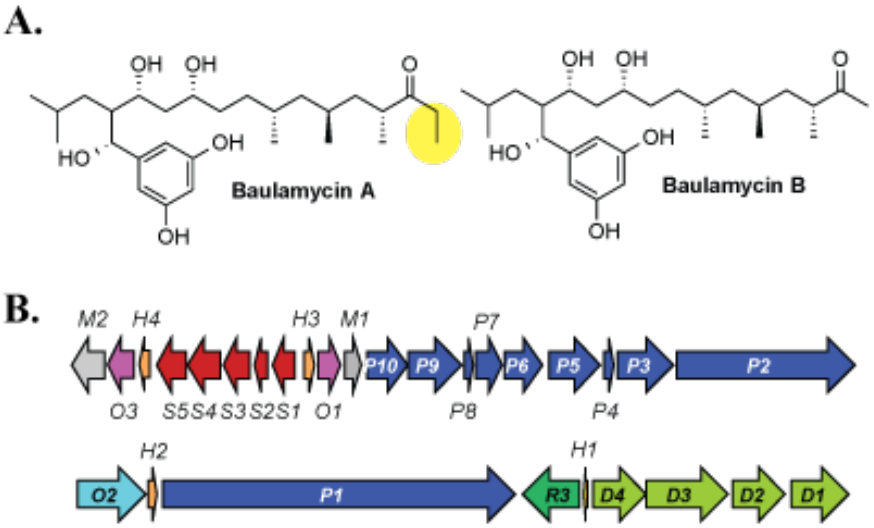


Figure 5-2. Identification of the PKS gene cluster responsible for baulamycin biosynthesis. **A.)** The structure of the baulamycin antibiotics described in chapter 2. The highlighted region depicts the key difference between BmcA and BmcB and the likely cause for observed differences in their efficacy. **B.)** The polyketide synthase (PKS) gene cluster responsible for the biosynthesis of the baulamycin antibiotics. Gene names relate to the proposed function for each protein: **M**, methyltransferase; **O**, monoxygenase; **H**, hypothetical proteins; **S**, enzymes responsible for biosynthesis of the starter unit; **M**, methyltransferase; **P**, polyketide synthase; **R**, regulatory proteins; **D**, enzymes responsible for biosynthesis of the first extender unit.

To identify the gene cluster responsible for baulamycin biosynthesis, we mined Illumina sequencing data of the producing microorganism, *Streptomyces tempisquensis*. The

structure of the baulamycins (**Figure 5-2A**) suggested that a type 1 polyketide synthase (PKS) enzyme was likely involved in the biosynthesis of the compounds. Mining sequencing data for type 1 PKS genes in conjunction with Drs. Fengan Yu and Ashootosh Tripathi from our laboratory led to the identification of the baulamycin biosynthetic gene cluster (**Figure 5-2B**).

Since my departure from this project, Drs. Ashootosh Tripathi and Sung Ryeol Park from our laboratory have begun pursuing a number of different methods with this gene cluster to understand and manipulate the biosynthesis of the baulamycin antibiotics (**Figure 5-3**).

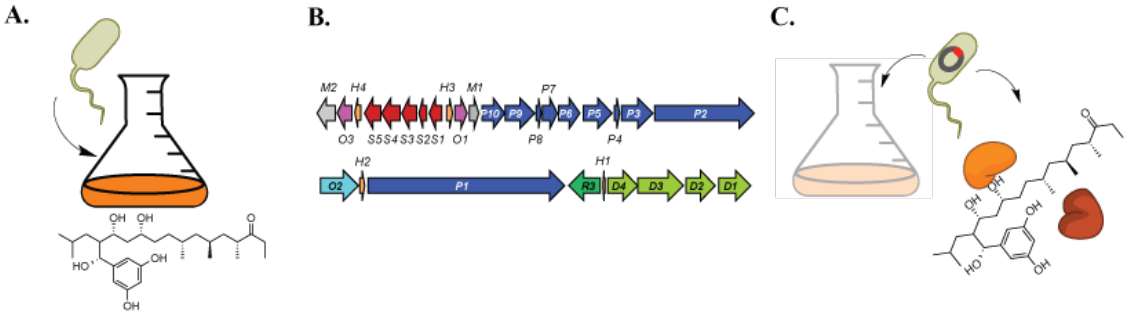


Figure 5-3. Long-term goals of the baulamycin project.
A.) Increase production of the baulamycins from the producing strain, *S. tempisquensis*. **B.)** Elucidate the biosynthetic pathway for baulamycin production and biochemically confirm key steps. **C.)** Generate new baulamycin analogues through gene cluster manipulation in the native strain and/or *in vitro* studies of key enzymes.

One of the major hurdles in chapter 2 was the extremely low yields of baulamycin production by *S. tempisquensis*. This limited our ability to explore the mechanism of action of these compounds and complete detailed studies of their efficacy against pathogenic microorganisms. Ongoing work by Drs. Tripathi and Park to increase the yield of the baulamycin compounds (**Figure 5-3A**) includes varying growth conditions of the producing strain and utilizing previously described ribosomal engineering methods (161, 162) to increase production of biosynthetic enzymes. This will hopefully increase our ability to isolate large amounts of the drug and eventually more potent analogues, facilitating pre-clinical studies.

Dr. Tripathi has also proposed a pathway for baulamycin production and begun *in vitro* analysis to biochemically confirm key steps of the antibiotics' biosynthesis (**Figure 5-3B**). Heterologous expression of key genes in *E. coli* has already confirmed one key biosynthetic step in the pathway. Confirmation of remaining steps with other cloned constructs is still in progress. Ongoing and future *in vitro* work with remaining enzymes and knockouts of corresponding genes in *S. tempisquensis* will help elucidate baulamycin biosynthesis (**Figure 5-3B**) and provide insight into the pathway's potential flexibility to incorporate unnatural substrates and generate new analogues (**Figure 5-3C**).

Generating more potent analogues is a particularly intriguing future direction for the baulamycin project (**Figure 5-3C**). The drastically increased efficacy of BmcA (4.9 μM SbnE/180 μM AsbA) over BmcB (19 μM SbnE/200 μM AsbA) shown in chapter 2 is likely due to the extended carbon chain present in the compound (**Figure 5-2A**). This suggests that tailoring the compound could have a drastic impact on its potency against pathogenic microorganisms. The incorporation of a longer carbon chain or additional methyl branching in particular could result in a more potent antibiotic. It would also be interesting to explore whether or not the compound could be cyclized and what effect this would have on its overall potency and selectivity for NIS synthetases.

5.2.2 Identifying additional natural product inhibitors of siderophore biosynthesis

In chapter 2, we pursued the best hit from our high throughput screen against an expansive natural product extract library. The cultured microorganism corresponding to this hit (*S. tempisquensis*) led to the discovery a new class of broad-spectrum antibiotics. There could be additional unique antibiotics hiding within other hits from the high throughput screen. To explore this possibility, we went back to examine the original screening results (**Table 5-1**).

Strain	AsbA Growth Variants	AsbA Top Hit (%)	AsbA Total Hits	SbnE Growth Variants	SbnE Top Hit (%)	SbnE Total Hits
34946	N9I	88.83	5	N12N, N23I, N9I, N9N	97.43	12
05545	I, N	72.77	6	I, N	86.53	6
44478	N5I, N5N	88.67	4	N5I, N5N	96.7	5
44341	A3N	83.47	5	A3N	98.5	4
58069	I	76.93	1	I	97.13	4
34362	A2N	90.2	2	A2N	98.7	2
5732	A2I	79.17	2	A2I	95.1	2
58195	I	76.03	2	I	96.2	2
5732	A2I	79.17	2	A2I	98.1	2

Table 5-1. Top hits active against both AsbA and SbnE in the high throughput screen against the natural product extract library.
The top hits from the original high throughput screen against both enzymes. Extracts in the library are derived from strain growth in multiple media ("growth variants") to maximize the types of biosynthesized natural products. The percent inhibition from the top hit is also shown. Strain 34946 (*S. tempisquensis*) was explored in chapter 2. We have begun exploring strain 44478 for potential novel natural products.

Several other hits from the NPE screen possessed bioactivity against AsbA and SbnE that rival the hit pursued in chapter 2 (34946, *S. tempisquensis*). Two hits of particular interest are 05545 and 44478. Unfortunately, we were unable to regrow strain 05545 from its freezer stock, a rare but possible problem associated with long-term storage of isolated environmental strains. However, strain 44478 grew exceptionally well and possessed multiple hits with high bioactivity against both targets during the original screen (**Table 5-1**).

I therefore carried out large-scale growth and extraction (10 L) of strain 44478-N5I and once again partnered with Dr. Ashootosh Tripathi to fractionate the extract using flash chromatography on a C18 cartridge. I then tested the bioactivity of the resultant fractions against both protein targets. Initial results suggest that multiple fractions have high bioactivity against both targets (**Figure 5-4**).

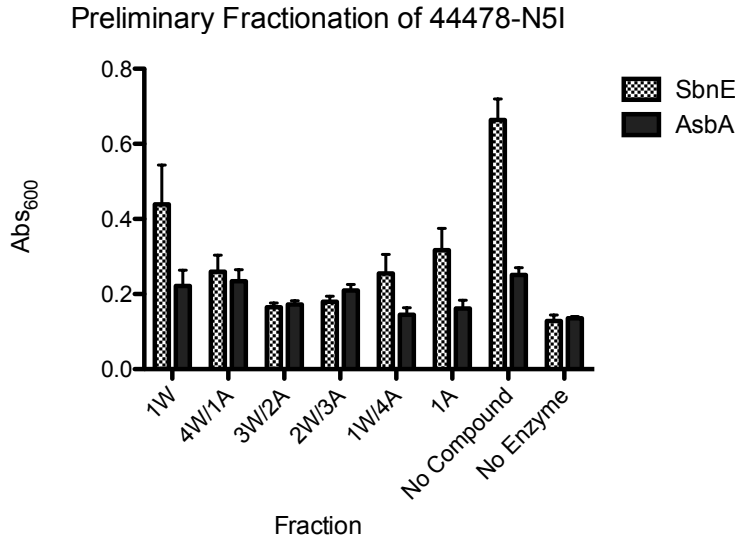


Figure 5-4. Initial bioactivity of strain 44478-N5I against NIS synthetases SbnE and AsbA. Initial fractionation was performed by flash chromatography on a C18 cartridge using varying ratios of water (W) to acetonitrile (A). Fractions were tested using the malachite green assay outlined in chapter 2.

Future work includes following the strategy outlined in chapter 2 to further fractionate this strain and isolate the natural product responsible for the observed bioactivity. Following structural elucidation of the compound, dose response assays with any purified natural products against both AsbA and SbnE will provide a better idea of potency. Testing the purified compounds against siderophore-producing microbial strains like *B. anthracis*, *S. aureus*, *E. coli*, *S. flexneri*, or *S. typimurium* in both iron-depleted and iron-rich medium could also provide clues into the specificity of any natural products. Ideally work with this strain will result in the discovery of another novel chemical scaffold with potent activity against NRPS-independent siderophore synthetases. An ideal antibiotic would selectivity bind these enzymes over other molecular targets present in the

microbial cell. This would limit selective pressure to iron-depleted conditions, slowing microbial acquisition of resistance to the antibiotic (35).

5.3 Future directions and ongoing work to access the natural products of uncultured microorganisms

Chapters 3 and 4 set the stage for multiple new routes to access the chemotherapeutic natural product ET-743. Although the drug is currently made semi-synthetically (94), the process is lengthy and costly. Synthetic methods also limit our ability to manipulate the drug and access more potent and selective analogues through cultivation or heterologous expression of key genes or the entire gene cluster. Future directions and ongoing work consists of pursuing multiple different routes to access the chemotherapeutic drug.

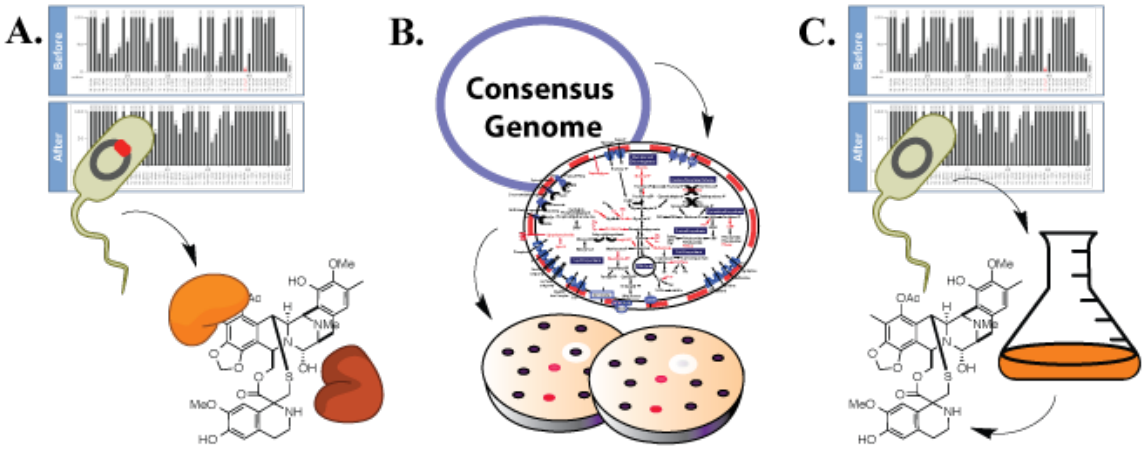


Figure 5-5. An outline of current routes being taken to access the chemotherapeutic natural product ET-743
A.) Reconstitute the compound's core using purified NRPS enzymes *in vitro* (described in chapter 4). **B.)** Utilize information gleaned from the genome of *Ca*. E. frumentensis to cultivate the microorganism in a laboratory setting (described here). **C.)** Directly produce the natural product and new analogous in a cultured host organism through genetic engineering (described here).

5.3.1 *In vitro* reconstitution of the ET-73 biosynthetic core (Figure 5-5A)

Chapter 4 described initial efforts toward biochemical analysis of NRPS enzymes involved in ET-743 biosynthesis and set the stage for future work toward *in vitro* reconstitution of the natural product's core.

Toward this goal, I have been continuing work to increase the solubility of EtuA2 and EtuA3 to improve the protein purification yield. Since the chemical chaperone TMAO showed a substantial increase in solubility, it is possible that enzymatic chaperones could also improve purification. I am currently coexpressing enzymatic chaperones and the NRPS enzymes to test for solubility improvements. Acquisition of the remaining NRPS modules will provide the essential components to complete *in vitro* reconstitution of the ET-743 biosynthetic core. We can also test the substrate flexibility of these enzymes as we did with EtuA1 and create new analogues through *in vitro* reactions or chemoenzyamtic synthesis.

There are also still some questions regarding the biochemical role of EtuA1, the transketolases EtuP1-3, and the ACP EtuF9. Future work to further analyze the enzymatic role of EtuA1 should include biochemical analysis of the enzyme with the rest of the NRPS assembly line along with EtuP1-3 and EtuF9. I have therefore successfully cloned codon-optimized versions of the transketolase pyruvate dehydrogenase enzymes EtuP1-3 into protein expression vectors. This will help determine what role—if any—the enzymes play in acquiring and/or loading glycolic acid for NRPS use. It will be interesting to see if the enzymes further improve the ability of EtuA1 to activate glycolic acid through the malachite green assay. Alternatively, *in vitro* enzymatic assays can be tested through mass spectrometry to examine whether or not activated glycolic acid is being successfully loaded onto EtuA1 in the presence or absence of these other enzymes.

5.3.2 Using metagenomic information to cultivate the producer of ET-743 (Figure 5-5B)

The genome of *Ca*. E. frumentensis presented in chapter 3 provided a detailed look at the microbe's primary metabolism. This analysis has strengthened our understanding of possible nutritional and environmental requirements for cultivation of the microorganism in the laboratory. An ongoing collaboration with the laboratory of Prof. Xiaoxia (Nina) Lin from the Chemical Engineering program at the University of Michigan could enable host-free cultivation of the microorganism in a laboratory setting (**Figure 5-5B**).

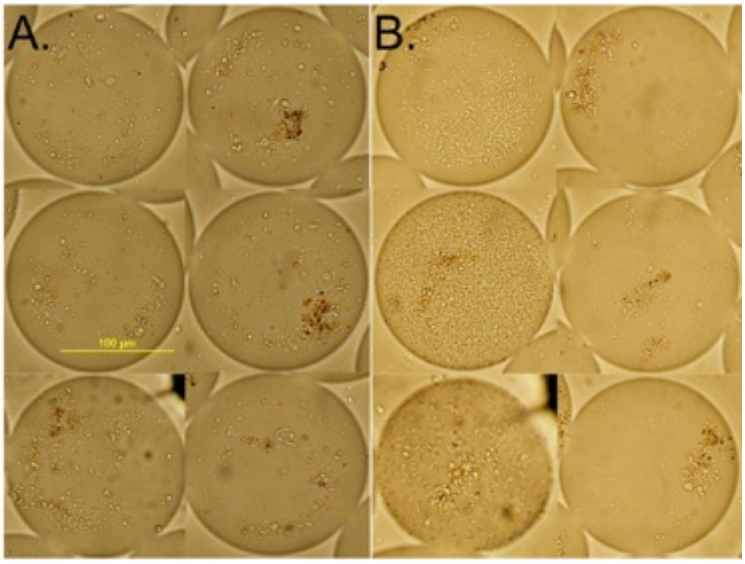


Figure 5-6. Culturing E. turbinata-derived bacteria in a microfluidic device
A.) 6 droplets at 0 hours. **B.)** 6 droplets at 26 hours. Data: (Steven Wang, Scott Mansfield, Dr. Jihynang Park, and Prof. Lin)

The collaboration with students Seven Wang and Scott Mansfield from the Lin laboratory utilizes a microfluidics device to attempt to cultivate the elusive microorganism. Bacteria isolated from fresh mangrove tunicates are subjected to a variety of different growth conditions on the microfluidic device (**Figure 5-6**) based on our analysis of the complete endosymbiont genome. For example, we now know that the microorganism is incapable of producing the majority of amino acids and many key cofactors. We also know that it likely uses an alternative carbon source. Supplementing growth media and tunicate extracts with ingredients like this could finally provide the conditions needed for host-free cultivation of the evasive microbe (**Figure 5-5B**).

5.3.3 Biosynthesis of ET-743 or new analogues through a heterologous host like *Streptomyces lavendulae* (Figure 5-5C)

An alternative route to expression and purification of individual enzymes is transferring key genes or the complete biosynthetic gene cluster directly into the genome of a

heterologous host (**Figure 5-5C**). *Streptomyces lavendulae* represents a particularly attractive host since it already biosynthesizes the tetrahydroisoquinoline natural product saframycin A. As we demonstrated in chapter 3, both the saframycin A and ET-743 gene clusters have genes predicted to play similar roles during the biosynthesis of the corresponding natural products. Some of the native genes in *S. lavendulae* may therefore be capable of tailoring the ET-743 tetrahydroisoquinoline core in place of the core of saframycin A. This could produce exciting new analogues on a larger scale.

The biggest obstacle with this route is that ET-743 biosynthetic genes—like the majority of the *Ca*. E. frumentensis genome—have an extremely low %G+C content. As shown in chapter 4, this contributes significantly to codon incompatibility observed with many heterologous hosts, *S. lavendulae* included. Codon optimized genes are the easiest way to combat this problem. To solve this dilemma, we have codon optimized the three NRPS genes responsible for the biosynthesis of the ET-743 core for expression in varying *Streptomyces* species. We have also begun collaborating with the laboratory of Dr. Eung-Soo Kim at Inha University in Korea. The laboratory is using their expertise in *actinomyces* genetics to attempt to biosynthesize ET-743 and analogues in different species, including the saframycin A producer, *S. lavendulae*.

5.4 Conclusions

The chapters of this dissertation have described significant efforts to explore and harness the natural products of both cultured and uncultured bacteria. The methodologies outlined here (targeted high throughput screening, metagenomic techniques to acquire and analyze complete genomes, *in vitro* reconstitution) can be applied to other natural product studies. Utilizing these and other available techniques will provide a better understanding of the world's microbial diversity and how we can harness that diversity to access novel medicines for our fight against drug resistant infections.

References

1. Center for Disease Control and Prevention (2013) *Antibiotic Resistance Threats in the United States, 2013*.

2. Livermore DM et al. (2011) Discovery research: the scientific challenge of finding new antibiotics. *J Antimicrob Chemother* 66:1941–1944.

3. Quadri LEN (2007) Strategic paradigm shifts in the antimicrobial drug discovery process of the 21st century. *Infect Disord Drug Targets* 7:230–237.

4. Medeiros AA (1997) Evolution and dissemination of beta-lactamases accelerated by generations of beta-lactam antibiotics. *Clin Infect Dis* 24 Suppl 1:S19–45.

5. Ehmann DE, Jahić H, Ross PL, Gu RF (2012) Avibactam is a covalent, reversible, non–β-lactam β-lactamase inhibitor. *Proc Natl Acad Sci USA* 109:11663–11668.

6. Fischbach MA, Walsh CT (2009) Antibiotics for emerging pathogens. *Science* 325:1089–1093.

7. Walsh CT, Wencewicz TA (2013) Prospects for new antibiotics: a molecule-centered perspective. *J Antibiot* 67:7–22.

8. Piret J, Boivin G (2014) Antiviral drug resistance in herpesviruses other than cytomegalovirus. *Rev Med Virol* 24:186–218.

9. BS JL, PhD KSGM (2014) Antiviral Resistance in Influenza Viruses. *Clin Lab Med* 34:387–408.

10. Khan R et al. (2009) Antimicrobial Activity of Five Herbal Extracts Against Multi Drug Resistant (MDR) Strains of Bacteria and Fungus of Clinical Origin. *Molecules* 14:586–597.

11. Cancer multidrug resistance (2000) Cancer multidrug resistance. *Nat Biotechnol* 18:IT18–IT20.

12. Broxterman HJ, Gotink KJ, Verheul HMW (2009) Understanding the causes of multidrug resistance in cancer: a comparison of doxorubicin and sunitinib. *Drug Resist Updat* 12:114–126.

13. Mahdi JG, Mahdi AJ, Mahdi AJ, Bowen ID (2006) The historical analysis of

aspirin discovery, its relation to the willow tree and antiproliferative and anticancer potential. *Cell Prolif* 39:147–155.

14. Gerwick WH, Marquez B, Milligan K, Tan LT, Williamson T (2004) *Encyclopedia of Biodiversity*.

15. Hamilton GR, Baskett TF (2001) History of AnesthesiaIn the arms of Morpheus: the development of morphine for postoperative pain relief. *J Antimicrob Chemother* 47:367–374.

16. Wright G (2014) Perspective: Synthetic biology revives antibiotics. *Nature* 509:S13.

17. Rosén J, Gottfries J, Muresan S, Backlund A, Oprea TI (2009) Novel chemical space exploration via natural products. *J Med Chem* 52:1953–1962.

18. Reymond JL, van Deursen R, Blum LC (2010) Chemical space as a source for new drugs. *MedChemComm* 1:30–38.

19. Newman DJ, Cragg GM (2012) Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *J Nat Prod* 75:311–335.

20. Newman DJ, Cragg GM (2007) Natural Products as Sources of New Drugs over the Last 25 Years. *J Nat Prod* 70:461–477.

21. Wani MC, Taylor HL, Wall ME, Coggon P, McPhail AT (1971) Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from Taxus brevifolia. *J Am Chem Soc* 93:2325–2327.

22. Zhou X, Zhu H, Liu L, Lin J, Tang K (2010) A review: recent advances and future prospects of taxol-producing endophytic fungi. *Appl Microbiol Biotechnol* 86:1707–1717.

23. Montaser R, Luesch H (2011) Marine natural products: a new wave of drugs? *Future Med Chem* 3:1475–1489.

24. Harvey AL (2008) Natural products in drug discovery. *Drug Discov Today* 13:894–901.

25. Knight V et al. (2003) Diversifying microbial natural products for drug discovery. *Appl Microbiol Biotechnol* 62:446–458.

26. Piel J (2009) Metabolites from symbiotic bacteria. *Nat Prod Rep* 26:338–362.

27. Piel J (2004) Metabolites from symbiotic bacteria. *Nat Prod Rep* 21:519–538.

28. Staley JT, Konopka A (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev*

Microbiol 39:321–346.

29. Stewart EJ (2012) Growing unculturable bacteria. *J Bacteriol* 194:4151–4160.

30. Puspita ID, Kamagata Y, Tanaka M, Asano K, Nakatsu CH (2012) Are uncultivated bacteria really uncultivable? *Microbes Environ* 27:356–366.

31. Brady SF, Chao CJ, Handelsman J, Clardy J (2001) Cloning and Heterologous Expression of a Natural Product Biosynthetic Gene Cluster from eDNA. *Org Lett* 3:1981–1984.

32. Andersson DI, Hughes D (2011) Persistence of antibiotic resistance in bacterial populations. *FEMS Microbiol Rev* 35:901–911.

33. Nathan C (2012) Fresh approaches to anti-infective therapies. *Sci Transl Med* 4:140sr2–140sr2.

34. Moir DT, Opperman TJ, Butler MM, Bowlin TL (2012) New classes of antibiotics. *Curr Opin Pharmacol* 12:535–544.

35. Quadri L (2007) Strategic paradigm shifts in the antimicrobial drug discovery process of the 21st century. *Infect Disord Drug Targets* 7:230–237.

36. Miethke M, Marahiel MA (2007) Siderophore-based iron acquisition and pathogen control. *Microbiol Mol Biol Rev* 71:413–451.

37. Somu RV et al. (2006) Rationally Designed Nucleoside Antibiotics That Inhibit Siderophore Biosynthesis of Mycobacteriumtuberculosis. *J Med Chem* 49:31–34.

38. Gupte A et al. (2008) Inhibition of Siderophore Biosynthesis by 2-Triazole Substituted Analogues of 5′- O-[N-(Salicyl)sulfamoyl]adenosine: Antibacterial Nucleosides Effective against Mycobacterium tuberculosis. *J Med Chem* 51:7495–7507.

39. Ferreras JA, Ryu J-S, Di Lello F, Tan DS, Quadri LEN (2005) Small-molecule inhibition of siderophore biosynthesis in Mycobacterium tuberculosis and Yersinia pestis. *Nat Chem Biol* 1:29–32.

40. Neres J et al. (2008) Inhibition of siderophore biosynthesis in Mycobacterium tuberculosis with nucleoside bisubstrate analogues: structure-activity relationships of the nucleobase domain of 5'-O-[N-(salicyl)sulfamoyl]adenosine. *J Med Chem* 51:5349–5370.

41. Klevens RM et al. (2007) Invasive methicillin-resistant Staphylococcus aureus infections in the United States. *JAMA* 298:1763–1771.

42. Sweeney DA, Hicks CW, Cui X, Li Y, Eichacker PQ (2011) Anthrax Infection. *Am J Respir Crit Care Med* 184:1333–1341.

43. Haley KP, Skaar EP (2012) A battle for iron: host sequestration and Staphylococcus aureus acquisition. *Microbes Infect* 14:217–227.

44. Hotta K, Kim C-Y, Fox DT, Koppisch AT (2010) Siderophore-mediated iron acquisition in Bacillus anthracis and related strains. *Microbiology* 156:1918–1925.

45. Haag H et al. (1994) Isolation and biological characterization of staphyloferrin B, a compound with siderophore activity from staphylococci. *FEMS Microbiol Lett* 115:125–130.

46. Drechsel H et al. (1993) Purification and chemical characterization of staphyloferrin B, a hydrophilic siderophore from staphylococci. *Biometals* 6:185–192.

47. Dale SE, Doherty-Kirby A, Lajoie G, Heinrichs DE (2003) Role of Siderophore Biosynthesis in Virulence of Staphylococcus aureus: Identification and Characterization of Genes Involved in Production of a Siderophore. *Infect Immun* 72:29–37.

48. Cheung J, Beasley FC, Liu S, Lajoie GA, Heinrichs DE (2009) Molecular characterization of staphyloferrin B biosynthesis in Staphylococcus aureus. *Mol Microbiol* 74:594–608.

49. Cheung J, Murphy MEP, Heinrichs DE (2012) Discovery of an Iron-Regulated Citrate Synthase in Staphylococcus aureus. *Chem Biol* 19:1568–1578.

50. Beasley FC, Cheung J, Heinrichs DE (2011) Mutation of L-2,3-diaminopropionic acid synthase genes blocks staphyloferrin B synthesis in Staphylococcus aureus. *BMC Microbiology* 11:199.

51. Cendrowski S, MacArthur W, Hanna P (2004) Bacillus anthracis requires siderophore biosynthesis for growth in macrophages and mouse virulence. *Mol Microbiol* 51:407–417.

52. Koppisch AT et al. (2005) Petrobactin is the Primary Siderophore Synthesized by Bacillus anthracis Str. Sterne under Conditions of Iron Starvation. *Biometals* 18:577–585.

53. Lee JY et al. (2007) Biosynthetic Analysis of the Petrobactin Siderophore Pathway from Bacillus anthracis. *J Bacteriol* 189:1698–1710.

54. Pfleger BF et al. (2007) Characterization and Analysis of Early Enzymes for Petrobactin Biosynthesis in Bacillus anthracis†. *Biochemistry* 46:4147–4157.

55. Pfleger BF et al. (2008) Structural and functional analysis of AsbF: origin of the stealth 3,4-dihydroxybenzoic acid subunit for petrobactin biosynthesis. *Proc Natl Acad Sci USA* 105:17133–17138.

56. Oves-Costales D et al. (2007) Enzymatic Logic of Anthrax Stealth Siderophore Biosynthesis: AsbA Catalyzes ATP-Dependent Condensation of Citric Acid and Spermidine. *J Am Chem Soc* 129:8416–8417.

57. Fox DT, Hotta K, Kim C-Y, Koppisch AT (2008) The missing link in petrobactin biosynthesis: asbF encodes a (-)-3-dehydroshikimate dehydratase. *Biochemistry* 47:12251–12253.

58. Nusca TD et al. (2012) Functional and Structural Analysis of the Siderophore Synthetase AsbB through Reconstitution of the Petrobactin Biosynthetic Pathway from Bacillus anthracis. *J Biol Chem* 287:16058–16072.

59. Oves-Costales D, Kadi N, Challis GL (2009) The long-overlooked enzymology of a nonribosomal peptide synthetase-independent pathway for virulence-conferring siderophore biosynthesis. *Chem Commun*:6530.

60. Challis GL (2005) A Widely Distributed Bacterial Pathway for Siderophore Biosynthesis Independent of Nonribosomal Peptide Synthetases. *ChemBioChem* 6:601–611.

61. Itaya K, Ui M (1966) A new micromethod for the colorimetric determination of inorganic phosphate. *Clin Chim Acta* 14:361–366.

62. Pegan SD, Tian Y, Sershon V, Mesecar AD (2010) A universal, fully automated high throughput screening assay for pyrophosphate and phosphate release from enzymatic reactions. *Comb Chem High Throughput Screen* 13:27–38.

63. McQuade TJ et al. (2009) A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. *Anal Biochem* 386:244–250.

64. Matsumori N, Kaneno D, Murata M, Nakamura H, Tachibana K (1999) Stereochemical Determination of Acyclic Structures Based on Carbon−Proton Spin-Coupling Constants. A Method of Configuration Analysis for Natural Products. *J Org Chem* 64:866–876.

65. Webb MR (1992) A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems. *Proc Natl Acad Sci USA* 89:4884–4887.

66. Wilson DJ, Aldrich CC (2010) A continuous kinetic assay for adenylation enzyme activity and inhibition. *Anal Biochem* 404:56–63.

67. Brandish PE et al. (1996) Modes of action of tunicamycin, liposidomycin B, and mureidomycin A: inhibition of phospho-N-acetylmuramyl-pentapeptide translocase from Escherichia coli. *Antimicrob Agents Chemother* 40:1640–1644.

68. Roemer T et al. (2011) Confronting the Challenges of Natural Product-Based

Antifungal Discovery. *Chem Biol* 18:148–164.

69. Gibson F, Magrath DI (1969) The isolation and characterization of a hydroxamic acid (aerobactin) formed by Aerobacter aerogenes 62-I. *Biochim Biophys Acta* 192:175–184.

70. Warner PJ, Williams PH, Bindereif A, Neilands JB (1981) ColV plasmid-specific aerobactin synthesis by invasive strains of Escherichia coli. *Infect Immun* 33:540–545.

71. de Lorenzo V, Bindereif A, Paw BH, Neilands JB (1986) Aerobactin biosynthesis and transport genes of plasmid ColV-K30 in Escherichia coli K-12. *J Bacteriol* 165:570–578.

72. de Lorenzo V, Neilands JB (1986) Characterization of iucA and iucC genes of the aerobactin system of plasmid ColV-K30 in Escherichia coli. *J Bacteriol* 167:350–355.

73. Magarvey NA, Keller JM, Bernan V, Dworkin M, Sherman DH (2004) Isolation and characterization of novel marine-derived actinomycete taxa rich in bioactive metabolites. *Appl Environ Microbiol* 70:7520–7529.

74. Stols L et al. (2002) A New Vector for High-Throughput, Ligation-Independent Cloning Encoding a Tobacco Etch Virus Protease Cleavage Site. *Protein Expr Purif* 25:8–15.

75. Eschenfeldt WH et al. (2010) Cleavable C-terminal His-tag vectors for structure determination. *J Struct Funct Genomics* 11:31–39.

76. Cruz PG et al. (2011) Titration-Based Screening for Evaluation of Natural Product Extracts: Identification of an Aspulvinone Family of Luciferase Inhibitors. *Chem Biol* 18:1442–1452.

77. Zhang JH, Chung T (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *Protein Expr Purif* 4:67–73.

78. Kadi N, Challis GL (2009) *Chapter 17 - Siderophore Biosynthesis: A Substrate Specificity Assay for Nonribosomal Peptide Synthetase-Independent Siderophore Synthetases Involving Trapping of Acyl-Adenylate Intermediates with Hydroxylamine* (Elsevier Inc.). 1st Ed.

79. Upson RH, Haugland RP, Malekzadeh MN (1996) A spectrophotometric method to measure enzymatic activity in reactions that generate inorganic pyrophosphate. *Anal Biochem* 243:41–45.

80. Campbell MK, Farrell O (2010) *Biochemistry* (Cengage Learning).

81. Kong LD, Cheng CHK, Tan RX (2004) Inhibition of MAO A and B by some

plant-derived alkaloids, phenols and anthraquinones. *Journal of Ethnopharmacology* 91:351–355.

82. Rainey FA, Ward-Rainey N, Kroppenstedt RM, Stackebrandt E (1996) The genus Nocardiopsis represents a phylogenetically coherent taxon and a distinct actinomycete lineage: proposal of Nocardiopsaceae fam. nov. *Int J Syst Bacteriol* 46:1088–1092.

83. Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod* 75:311–335.

84. Cragg GM, Grothaus PG, Newman DJ (2009) Impact of natural products on developing new anti-cancer agents. *Chem Rev* 109:3012–3043.

85. Wilson MC, Piel J (2013) Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem Biol* 20:636–647.

86. Lichter W, Lopez DM, Wellham L, Sigel MM (1975) Ecteinascidia turbinata extracts inhibit DNA synthesis in lymphocytes after mitogenic stimulation by lectins. *Exp Biol Med* 150.

87. Rinehart KL, Holt TG, Fregeau NL (1990) Ecteinascidins 729, 743, 745, 759A, 759B, and 770: potent antitumor agents from the Caribbean tunicate Ecteinascidia turbinata. *J Org Chem* 55.

88. McLaughlin K (2015) U.S. FDA Grants Priority Review for YONDELIS® (trabectedin) for the Treatment of Patients with Advan.

89. Pérez-Matos AE, Rosado W, Govind NS (2007) Bacterial diversity associated with the Caribbean tunicate Ecteinascidia turbinata. *Antonie Van Leeuwenhoek* 92:155–164.

90. Moss C, Green DH, Pérez B, Velasco A, Henríquez R (2003) Intracellular bacteria associated with the ascidian Ecteinascidia turbinata: phylogenetic and in situ hybridisation analysis. *Mar Biol* 143:99–110.

91. Rath CM et al. (2011) Meta-omic characterization of the marine invertebrate microbial consortium that produces the chemotherapeutic natural product ET-743. *ACS Chem Biol* 6:1244–1256.

92. Carballo JL, Naranjo S, Kukurtzü B, Calle F, Hernández Zanuy A (2000) Production of Ecteinascidia turbinata (Ascidiacea: Perophoridae) for Obtaining Anticancer Compounds. *J World Aquac Soc* 31:481–490.

93. Corey EJ, Gin DY, Kania RS (1996) Enantioselective total synthesis of ecteinascidin 743. *J Am Chem Soc* 118:9202–9203.

94. Cuevas C, Francesch A (2009) Development of Yondelis (trabectedin, ET-743). A semisynthetic process solves the supply problem. *Nat Prod Rep* 26:322–337.

95. Dick GJ et al. (2009) Community-wide analysis of microbial genome sequence signatures. *Genome Biol* 10:R85.

96. McCutcheon JP, Moran NA (2012) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10:13–26.

97. Moran NA (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci USA* 93:2873–2878.

98. Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 42:165–190.

99. Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 42:165–190.

100. Kuo C-H, Moran NA, Ochman H (2009) The consequences of genetic drift for bacterial genome complexity. *Genome Res* 19:1450–1454.

101. Mira A, Ochman H, Moran NA (2001) Deletional bias and the evolution of bacterial genomes. *Trends Genet* 17:589–596.

102. Kuo C-H, Ochman H (2009) Deletional bias across the three domains of life. *Genome Biol Evol* 1:145–152.

103. Moran NA, Munson MA (1993) A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc R Soc Lond B* 253:167–171.

104. Chen X, Li S, Aksoy S (1999) Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus Glossina and its bacteriome-associated endosymbiont, Wigglesworthia glossinidia. *J Mol Evol* 48:49–58.

105. Konstantinidis KT, Tiedje JM (2005) Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* 187:6258–6264.

106. Yarza P et al. (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 12:635–645.

107. Görke B, Stülke J (2008) Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* 6:613–624.

108. Munoz-Elias EJ, McKinney JD (2006) Carbon metabolism of intracellular bacteria. *Cell Microbiol* 8:10–22.

109. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome

sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. *Nature* 407:81–86.

110. Wernegreen JJ (2002) Genome evolution in bacterial endosymbionts of insects. *Nat Rev Genet* 3:850–861.

111. Omsland A et al. (2009) Host cell-free growth of the Q fever bacterium Coxiella burnetii. *Proc Natl Acad Sci USA* 106:4430–4434.

112. Omsland A, Heinzen RA (2011) Life on the outside: the rescue of Coxiella burnetii from its host cell. *Annu Rev Microbiol* 65:111–128.

113. Spry C, Kirk K, Saliba KJ (2008) Coenzyme A biosynthesis: an antimicrobial drug target. *FEMS Microbiol Rev* 32:56–106.

114. Trager W, Brohn FH (1975) Coezyme A requirement of malaria parasites: effects of coenzyme A precursors on extracellular development in vitro of Plasmodium lophurae. *Proc Natl Acad Sci USA* 72:1834–1837.

115. Bovarnick MR, Allen EG (1954) Reversible Inactivation of Typhus Rickettsiae. *J Gen Physiol* 38:169–179.

116. Tourtellotte ME, Morowitz HJ, Kasimer P (1964) Defined Medium for Mycoplasma Laidlawii. *J Bacteriol* 88:11–15.

117. Renesto P, Ogata H, Audic S, Claverie J-M, Raoult D (2005) Some lessons from Rickettsia genomics. *FEMS Microbiol Rev* 29:99–117.

118. Pérez-Brocal V et al. (2006) A small microbial genome: the end of a long symbiotic relationship? *Science* 314:312–313.

119. Nakabachi A et al. (2006) The 160-kilobase genome of the bacterial endosymbiont Carsonella. *Science* 314:267–267.

120. Wu D et al. (2006) Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol* 4:e188.

121. Lei L et al. (2008) Characterization of the saframycin A gene cluster from Streptomyces lavendulae NRRL 11002 revealing a nonribosomal peptide synthetase system for assembling the unusual tetrapeptidyl skeleton in an iterative manner. *J Bacteriol* 190:251–263.

122. Pospiech A, Cluzel B, Bietenhader J, Schupp T (1995) A new Myxococcus xanthus gene cluster for the biosynthesis of the antibiotic saframycin Mx1 encoding a peptide synthetase. *Microbiology* 141:1793–1803.

123. Velasco A et al. (2005) Molecular characterization of the safracin biosynthetic pathway from Pseudomonas fluorescens A2-2: designing new cytotoxic

compounds. *Mol Microbiol* 56:144–154.

124. Hiratsuka T et al. (2013) Core assembly mechanism of quinocarcin/SF-1739: bimodular complex nonribosomal peptide synthetases for sequential mannich-type reactions. *Chem Biol* 20:1523–1535.

125. Patel MS, Nemeria NS, Furey W, Jordan F (2014) The pyruvate dehydrogenase complexes: structure-based function and regulation. *J Biol Chem* 289:16615–16623.

126. Peng C et al. (2012) Hijacking a hydroxyethyl unit from a central metabolic ketose into a nonribosomal peptide assembly line. *Proc Natl Acad Sci USA* 109:8540–8545.

127. Walton JD (2000) Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: an hypothesis. *Fungal Genet Biol* 30:167–171.

128. Chu HY, Wegel E, Osbourn A (2011) From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. *Plant J* 66:66–79.

129. Lawrence JG, Roth JR (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* 143:1843–1860.

130. Price MN, Huang KH, Arkin AP, Alm EJ (2005) Operon formation is driven by co-regulation and not by horizontal gene transfer. *Genome Res* 15:809–819.

131. Kwan JC et al. (2012) Genome streamlining and chemical defense in a coral reef symbiosis. *Proc Natl Acad Sci USA* 109:20655–20660.

132. Blin K et al. (2013) antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res* 41:W204–12.

133. Li MH, Ung PMU, Zajkowski J, Garneau-Tsodikova S, Sherman DH (2009) Automated genome mining for natural products. *BMC Bioinformatics* 10:185.

134. Weber T, Rausch C, Lopez P, Hoof I, Gaykova V (2009) CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. *J Biotechnol* 140:13–17.

135. van Heel AJ, de Jong A, Montalbán-López M, Kok J, Kuipers OP (2013) BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res* 41:W448–53.

136. Bachmann BO, Ravel J (2009) Methods for In Silico Prediction of Microbial Polyketide and Nonribosomal Peptide Biosynthetic Pathways from DNA Sequence Data. *Methods Enzymol* 458:181–217.

137. Pisut DP, Pawlik JR (2002) Anti-predatory chemical defenses of ascidians: secondary metabolites or inorganic acids? *J Exp Mar Bio Ecol* 270:203–214.

138. Davis AR (1991) Alkaloids and ascidian chemical defense: evidence for the ecological role of natural products from Eudistoma olivaceum. *Mar Biol* 111:375–379.

139. Lindquist N, Hay ME, Fenical W (1992) Defense of ascidians and their conspicuous larvae: adult vs. larval chemical defenses. *Ecol Monogr* 62:547.

140. Young CM, Bingham BL (1987) Chemical defense and aposematic coloration in larvae of the ascidian Ecteinascidia turbinata. *Mar Biol* 96:539–544.

141. Williams DH, Stone MJ, Hauck PR, Rahman SK (1989) Why are secondary metabolites (natural-products) biosynthesized? *J Nat Prod* 52:1189–1208.

142. Nett M, Ikeda H, Moore BS (2009) Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep* 26:1362.

143. Stone MJ, Williams DH (1992) On the evolution of functional secondary metabolites (natural products). *Mol Microbiol* 6:29–34.

144. Nakabachi A et al. (2013) Defensive bacteriome symbiont with a drastically reduced genome. *Curr Biol* 23:1478–1484.

145. Kwan JC, Schmidt EW (2013) Bacterial endosymbiosis in a chordate host: long-term co-evolution and conservation of secondary metabolism. *PLoS ONE* 8:e80822.

146. Partida-Martinez LP, Hertweck C (2005) Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* 437:884–888.

147. Lackner G, Moebius N, Partida-Martinez LP, Boland S, Hertweck C (2011) Evolution of an endofungal lifestyle: Deductions from the Burkholderia rhizoxinica genome. *BMC Genomics* 12:210.

148. Kikuchi Y (2009) Endosymbiotic bacteria in insects: their diversity and culturability. *Microbes Environ* 24:195–204.

149. Markowitz VM et al. (2013) IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res* 42:D568–D573.

150. Ciccarelli FD (2006) Toward Automatic Reconstruction of a Highly Resolved Tree of Life. *Science* 311:1283–1287.

151. Markowitz VM et al. (2014) IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res* 42:D560–7.

152. Kwan JC et al. (2014) Host control of symbiont natural product chemistry in cryptic populations of the tunicate Lissoclinum patella. *PLoS ONE* 9:e95850.

153. Krzywinski M et al. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639–1645.

154. Mortison JD, Sherman DH (2010) Frontiers and Opportunities in Chemoenzymatic Synthesis. *J Org Chem* 75:7041–7051.

155. Kopp F, Marahiel MA (2007) Where chemistry meets biology: the chemoenzymatic synthesis of nonribosomal peptides and polyketides. *Curr Opin Biotechnol* 18:513–520.

156. Angov E (2011) Codon usage: Nature's roadmap to expression and folding of proteins. *Biotechnol J* 6:650–659.

157. Hansen DA et al. (2013) Biocatalytic Synthesis of Pikromycin, Methymycin, Neomethymycin, Novamethymycin, and Ketomethymycin. *J Am Chem Soc* 135:11232–11238.

158. Feifel SC et al. (2007) In Vitro Synthesis of New Enniatins: Probing the α-D-Hydroxy Carboxylic Acid Binding Pocket of the Multienzyme Enniatin Synthetase. *ChemBioChem* 8:1767–1770.

159. Magarvey NA, Ehling-Schulz M, Walsh CT (2006) Characterization of the Cereulide NRPS α-Hydroxy Acid Specifying Modules: Activation of α-Keto Acids and Chiral Reduction on the Assembly Line. *J Am Chem Soc* 128:10698–10699.

160. Calderone CT, Bumpus SB, Kelleher NL, Walsh CT, Magarvey NA (2008) A ketoreductase domain in the PksJ protein of the bacillaene assembly line carries out both alpha- and beta-ketone reduction during chain growth. *Proc Natl Acad Sci USA* 105:12809–12814.

161. Ochi K, Hosaka T (2012) New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. *Appl Microbiol Biotechnol* 97:87–98.

162. Ochi K et al. (2004) Ribosome engineering and secondary metabolite production. *Adv Appl Microbiol* 56:155–184.