Title
Biosynthesis and Heterologous Expression of Medicinally Active Natural Products

Author
Liu, Joyce

Publication Date
2016

Peer reviewed|Thesis/dissertation

Biosynthesis and Heterologous Expression of Medicinally Active Natural Products

by

Joyce Liu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy
with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Wenjun Zhang, Chair
Professor Michelle C. Y. Chang
Professor David F. Savage
Professor Patricia C. Babbitt

Fall 2016

Biosynthesis and Heterologous Expression of Medicinally Active Natural Products

by Joyce Liu

Abstract

Biosynthesis and Heterologous Expression of Medicinally Active Natural Products

by

Joyce Liu

Joint Doctor of Philosophy with University of California, San Francisco

in Bioengineering

University of California, Berkeley

Professor Wenjun Zhang, Chair

Natural products have long been appreciated for their potent biological activities, and their unique scaffolds and pharmacophores have served as inspiration for many pharmaceuticals. As the structural complexity of these compounds can make their production and diversification challenging for chemical synthesis, the study and engineering of natural product biosynthetic pathways can facilitate more efficient and sustainable production of natural products and their analogues. Here, we present our findings from the biosynthetic studies of several medicinally active natural products produced by *Streptomyces*, which reveal some unusual pathways and enzymes. In doing this work, we also demonstrate how heterologous expression can be used to expediently determine the necessity and function of individual biosynthetic enzymes, gain insight into biosynthetic mechanisms, and produce new natural product analogues.

The compounds of focus here include the cholinesterase inhibitor physostigmine, the peptidyl epoxyketone proteasome inhibitor eponemycin, and the family of antimycin-type depsipeptides. The *in vivo* and *in vitro* characterization of physostigmine biosynthesis revealed an unexpected pathway involving an acetylation-deacetylation-dependent reaction cascade and a unique indole methyltransferase. Heterologous expression-based investigations on the biosynthesis of eponemycin then revealed an unprecedented flavin-dependent enzyme to be necessary and sufficient for the formation of the terminal epoxyketone pharmacophore in addition to indicating the involvement of a decarboxylation step. Further efforts to study the biosynthesis of antimycins likewise resulted in the *in vivo* reconstitution of the 3-formamidosalicylate pharmacophore, and studies on the related neoantimycins led to advances in the biosynthetic understanding and heterologous expression of this ring-expanded class of antimycin-type depsipeptides. Collectively, these biosynthetic insights not only enrich our enzymatic knowledge but can also expand our ability to perform combinatorial biosynthesis.

Table of contents

List of figures

List of tables

Acknowledgments

I would first like to thank my advisor, Wenjun Zhang, for her mentorship over the past several years. I have learned an incredible amount about all aspects of science from her, and I am deeply appreciative of all of her patience and support.

I would also like to thank all of the current and former members of the Zhang Lab who have kept me motivated and made working in the lab enjoyable. I am grateful for the opportunity to have worked with graduate students: Xuejun Zhu and Will Skyrud; graduate rotation students: Omer Ad and Jackie Blake-Hedges; and undergraduate students: Tailun Ng, Joel Sanchez, and Vinh Tran, all of who have contributed to the work described here.

I also thank our collaborator, Dr. Ryan Seipke, for his assistance with the antimycin and neoantimycin work. Dr. Jeffrey Pelton's help with the acquisition of NMR data and Dr. Jeffrey Skerker's assistance with genome assembly have also been invaluable.

Finally, I would like to thank my family and especially my mother for all of their support and encouragement.

Chapter 1. Introduction

1.1 Significance of natural products

Natural products produced by plants and microorganisms are among some of the most important pharmaceutical compounds in existence. Collectively, they have played important roles in treating a vast variety human health conditions, including infectious, neurological, cardiovascular, metabolic, and oncological diseases.[1] Examples of particularly well-known natural products with clinical uses include the antibiotics, penicillin and vancomycin; the antimalarial agent, artemisinin; the pain medication, morphine; and the anticancer drugs, bleomycin and paclitaxel (Figure 1-1). Although not all natural products can directly be used as pharmaceuticals, they are nonetheless an abundant source of structurally diverse and biologically active compounds and can offer novel scaffolds and pharmacophores that serve as the basis for the development of new drugs.[2] It has been estimated that in the decades spanning 1981 - 2010, nearly 75% of all approved small molecule antibiotics and nearly 50% of all approved small molecule anticancer drugs were natural products or their derivatives.[1]

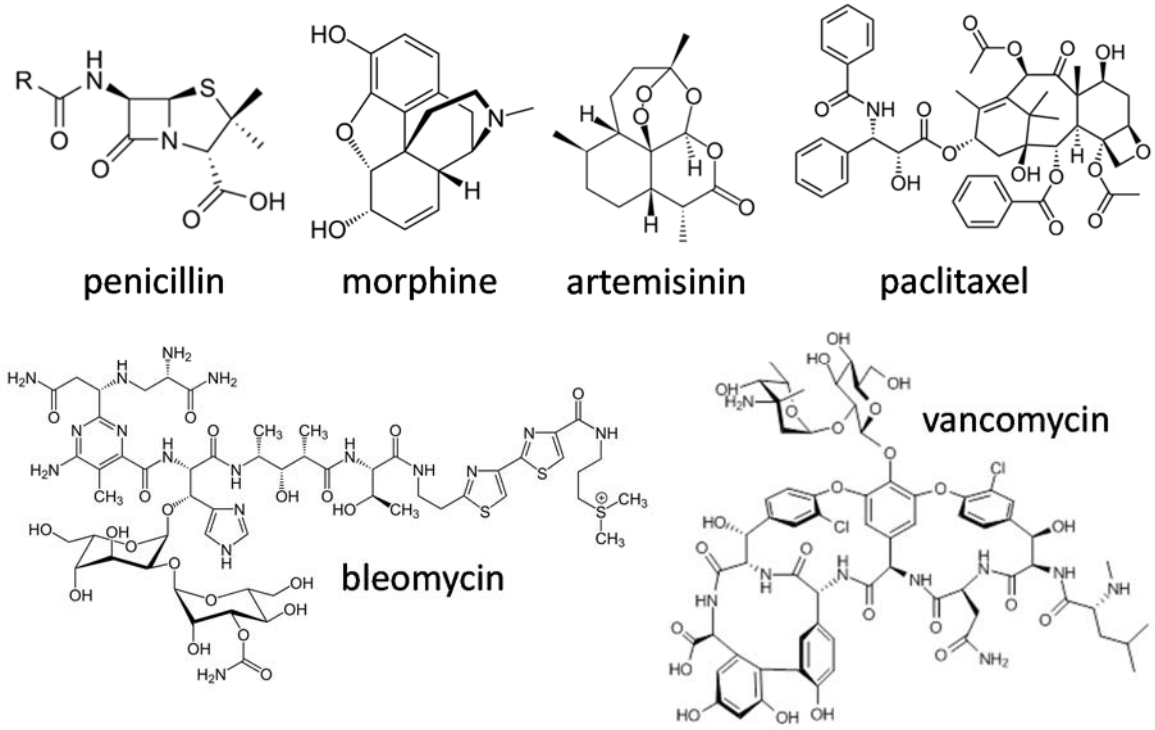


Figure 1-1. Examples of pharmaceutically relevant natural products.

Natural products are often produced by organisms as secondary metabolites, and they are generally believed to have evolved to promote the survival of the producing organism, be it through chemical defense, symbiosis, metal transport, or differentiation.[3] While the natural roles of these compounds confer them with some form of biological activity, the use of natural products in the context of treating human disease does not usually rely on the specific activity for which the metabolite was originally evolved. As such, efforts to alter these compounds through combinatorial chemistry and biosynthetic engineering approaches remain crucial for drug development.[2, 4] In addition to possessing

a wide range of biological activities, secondary metabolites also possess extremely varied structures, and they can be grouped into different classes based on their biosynthetic origins. Some of the major classes include polyketides, phenylpropanoids, terpenoids, alkaloids, and ribosomal and non-ribosomal peptides, and even within these classes, there is substantial scaffold and functional group diversity.[5] Chapter 2 will discuss the biosynthesis of a tryptophan-derived alkaloid known to be produced by a plant and a *Streptomyces* species, while Chapters 3-5 will discuss the biosynthesis of three different non-ribosomal peptide/polyketide hybrid compounds produced by *Streptomyces*.

1.2 Motivation for studying biosynthesis

As secondary metabolites are typically produced by their native producers in small quantities, the development of alternative methods that enable desired compounds to be produced at higher productivities and titers is extremely desirable. However, many natural products have complex structures with multiple stereocenters, diverse functional groups, and labile connectivities, which makes their total syntheses challenging.[6] Moreover, these chemical synthetic efforts often require costly and environmentally harmful reagents as well as numerous steps, which leads to reduced yields. Engineered biosynthesis is thus an attractive and more sustainable approach for obtaining greater quantities of natural products that can be used for further study, derivatization, or treatment. Toward this end, the understanding and complete characterization of the biosynthetic pathways of target compounds can facilitate the development and metabolic engineering of higher producing native or heterologous strains.[7] As an example, investigation into the biosynthesis of penicillin by *Aspergillus nidulans* revealed that the aminoadipyl-cysteinylvaline synthetase was the rate limiting enzyme in the pathway, and subsequent overexpression of the enzyme resulted in a 30-fold increase in penicillin production.[8] Similarly, glycosylation was determined to be the rate-limiting step in doxorubicin biosynthesis by *Streptomyces peucetius*, which prompted the overexpression of the glycosyltransferase and sugar biosynthesis genes for a six-fold increase in production.[9] An understanding of biosynthesis can also play a role in more efficient production of natural product analogues with different activities, which can be achieved through the deletion, insertion, or replacement of specific pathway genes as exemplified by the genetic engineering of the doxorubicin producer to produce epirubicin, a drug with fewer side-effects.[10]

Finally, investigating the biosynthesis of natural products with unique functionalities can shed light on new kinds of enzymes and their associated chemistries. Such discoveries can not only enrich our ability to perform desired enzymatic transformations but can also expand the potential to generate novel bioactive compounds through engineered biosynthetic pathways. For instance, the characterization of spinosyn A biosynthesis revealed the involvement of a novel "Diels-Alderase" enzyme capable of catalyzing [4+2] cycloaddition (Figure 1-2A).[11] Likewise, the elucidation of fosfomycin biosynthesis revealed an unusual epoxidase that catalyzes dehydrogenation of a secondary alcohol (Figure 1-2B),[12, 13] and the study of echinomycin biosynthesis led to the identification of a rare *S*-adenosylmethionine (SAM)-dependent methyltransferase that transforms a disulfide bridge into a thioacetal bond (Figure 1-2C).[14]

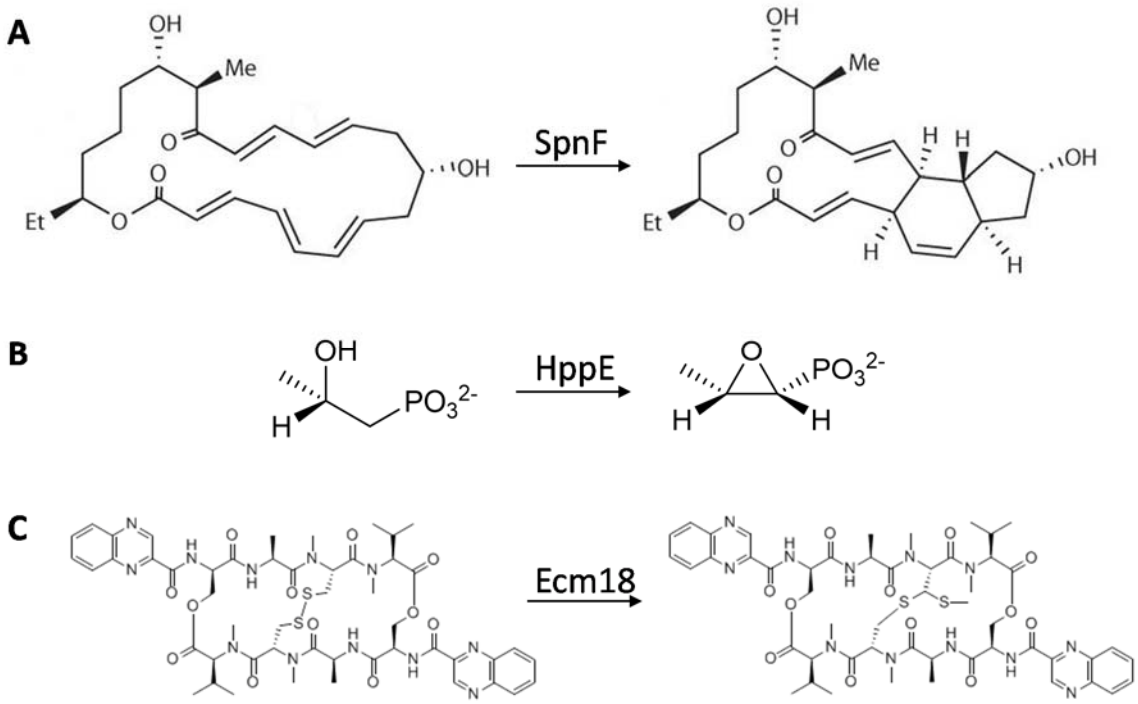


Figure 1-2. Examples of reactions catalyzed by novel biosynthetic enzymes. (A) SpnF catalyzes a [4 + 2] cycloaddition reaction in the biosynthesis of spinosyn A. (B) HppE catalyzes an unusual dehydrogenation reaction to form an epoxide in the biosynthesis of fosfomycin. (C) Ecm18 catalyzes the rare transformation of a disulfide bridge into a thioacetal bridge in the biosynthesis of echinomycin.

1.3 Applications of heterologous expression

In addition to biosynthetic understanding, another important aspect of engineered biosynthesis is the heterologous expression of biosynthetic pathways. Because the native producers of many natural products may be difficult to genetically manipulate, heterologous expression is becoming an increasingly popular way to both improve and modify the biosynthesis of natural products. One instance of the use of heterologous expression is the engineering of a *Saccaharomyces cerevisiae* strain for the production of 25 g/L of the artemisinin precursor, artemisinic acid, which represents a substantial improvement over extracting artemisinin from the naturally producing plant.[15-17] Notably, the heterologous expression efforts in this case involved both host engineering through the overexpression of all of the yeast mevalonate pathway enzymes and biosynthetic pathway engineering through the incorporation of an artemisinic aldehyde dehydrogenase and an artemisinic alcohol dehydrogenase.

Heterologous expression has also been demonstrated to be a valuable tool for engineering the biosynthesis of natural product analogues. In particular, the heterologous expression of the novobiocin biosynthetic gene cluster from *Streptomyces spheroides* in the more easily manipulated *Streptomyces coelicolor* host enabled the deletion of a methyltransferase and the introduction of a halogenase into the biosynthetic pathway to yield a novel chlorinated novobiocin analogue.[18, 19] A similar strategy was also employed to produce two new chlorinated coumermycin A_1 analogues.[20] Furthermore, the successful expression of the erythromycin biosynthetic gene cluster from

Saccharopolyspora erythraea in an *Escherichia coli* strain metabolically engineered to support polyketide production enabled the relatively facile production of new erythromycin analogues through modification of the polyketide synthase (PKS) modules and the tailoring pathway enzymes (Figure 1-3A).[21, 22] In one study, the loading module of 6-deoxyerythronolide B synthase (DEBS) was replaced with an alternative to enable the incorporation of a completely different substrate into the 6-dEB scaffold.[21] In another study, 16 different deoxysugar tailoring pathways were introduced into *E. coli* to produce an array of erythromycin analogues with different deoxysugar moieties, and some of these analogues even exhibited activity against an erythromycin-resistant *Bacillus subtilis* strain.[22] Impressively, the production of erythromycin by the engineered *E. coli* system involved the expression of over 50 kb of DNA and 26 genes. However, the titer of erythromycin from *E. coli* was approximately 1000-fold less than that from *S. erythraea* (~10 mg/L vs ~10 g/L),[21, 23] indicating that there may be some limitations to heterologous expression.

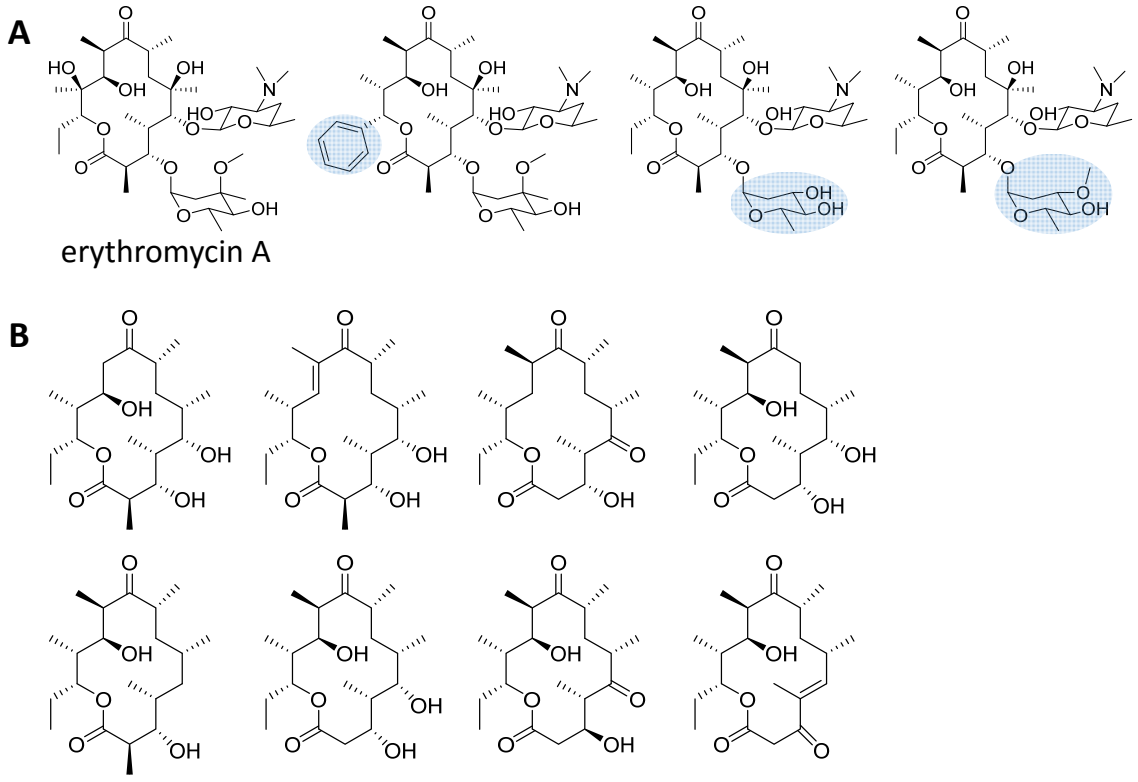


Figure 1-3. Erythromycin A and biosynthetic analogues. (A) Heterologously produced erythromycin analogues resulting from the modification of the loading module of 6-deoxyerythronolide B synthase (DEBS) and the addition of new deoxysugar tailoring pathways. (B) Small subset of the 6-dEB combinatorial library resulting from domain and module engineering of DEBS.

Nevertheless, heterologous expression also has promising applications for the expression and refactoring of silent gene clusters. As the gene clusters for some secondary metabolites may be highly regulated and not expressed under laboratory cultivation conditions, the transfer and refactoring of these gene clusters in heterologous hosts can provide the opportunity to reveal the production of new compounds.[24] This

approach has recently been demonstrated by the expression of the taromycin A and spectinabilin gene clusters in *Streptomyces coelicolor* and *Streptomyces lividans*, respectively.[25, 26] These studies relied on yeast-based homologous recombination to construct large shuttle vectors (some even upwards of 80 kb) containing biosynthetic gene clusters of interest that were modified for activation via the deletion of regulatory genes responsible for silencing the cluster or the addition of a strong promoter in front of each gene. This work thus underscores the importance of using heterologous hosts for which abundant genetic tools and regulatory elements are available such that large biosynthetic gene clusters can be both easily introduced and expressed in the host.

For natural products originating from *Actinobacteria*, which is the focus here, *E. coli* and model *Streptomyces* strains such as *S. coelicolor* and *S. lividans* are among some of the most promising heterologous hosts for engineered biosynthesis.[7, 27, 28] Examples of actinobacterial natural products that have been successfully heterologously expressed in *E. coli* include echinomycin,[14] erythromycin,[21] and valinomcycin;[29] and examples of compounds that have been heterologously expressed in *Streptomyces* hosts include daptomycin,[30] meridamycin,[31] and aureothin.[32] Although the titers resulting from heterologous expression are often reduced compared to that from the native producers,[27] there are some cases in which the titers are comparable (e.g. valinomycin and novobiocin) or even higher (e.g. tetracenomycins[33]) than that from the native host. Furthermore, a slight decrease in titers may be compensated if the heterologous host is easier to ferment,[27] and the choice of heterologous host can also have a significant effect on titers.[19, 33] While *Streptomyces* hosts are advantageous because of their similarity to native producers, the relevance of their precursor pools, and their ability to express extremely large enzymes, *E. coli* is also a particularly attractive host due to its fast growth rates, well-characterized metabolism, and the wide availability of genetic tools and parts.[28, 34-37] These traits make the engineering of both the host and imported biosynthetic pathways significantly easier, and to address some of the corresponding shortcomings of *Streptomyces* hosts, efforts have been made to engineer strains without competing secondary metabolite gene clusters,[38-40] to characterize new promoters,[41] and to develop genetic tools like CRISPR/Cas that facilitate the genetic engineering of *Streptomyces*.[42] Likewise, the engineering of *E. coli* to produce the precursor, methylmalonyl-CoA, and to express the post-translational modification enzyme, phosphopantetheinyl transferase, which resulted in the *E. coli* BAP1 strain, has greatly expanded its capability as a heterologous host.[35, 43] The more recent finding that the expression of *Streptomyces* chaperonins in *E. coli* could improve the solubility and activity of megaenzymes[44] also has promise for furthering the development of *E. coli* as a heterologous host for the expression of large biosynthetic gene clusters. Overall, advances in heterologous host engineering are improving our ability to achieve the heterologous expression of biosynthetic gene clusters and allowing us to exploit the many applications of heterologous expression. In Chapters 3-5, we show how heterologous expression in *E. coli* and *Streptomyces* can be used as a tool to probe the functions of individual biosynthetic enzymes and better characterize biosynthetic pathways when the native producer is not genetically tractable.

1.4 Outlook for combinatorial biosynthesis

With increased knowledge about the biosynthesis of natural products and the ability to produce them in organisms that can be easily engineered, the ultimate goal is to apply these tools to generate novel bioactive compounds through combinatorial biosynthesis. Combinatorial biosynthesis can broadly be thought of as the use of genetic engineering to modify biosynthetic pathways to produce altered compounds.[4, 36] As evidenced by some of the examples described in the previous sections, there has been considerable interest in applying combinatorial biosynthesis to natural products synthesized by type I polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs) due to the collinearity, modular nature, and diverse building block selectivity of these enzymes.[45-47] PKSs and NRPSs operate in an assembly line-like fashion to join together acyl coenzyme A (CoA) and amino acid building blocks, respectively.[45, 46, 48, 49] The core catalytic domains of type I PKSs include the ketosynthase (KS) domain, which is responsible for catalyzing decarboxylative Claisen condensations for chain extension; the acyltransferase (AT) domain, which is responsible for building monomer selection and loading; and the acyl carrier protein (ACP) domain, on which the polyketide chain is elongated (Figure 1-4A). Similarly, the core catalytic domains of NRPS include the condensation (C) domain, which is responsible for catalyzing peptide bond formation for chain extension; the adenylation (A) domain, which is responsible for amino acid activation and loading; and the thiolation (T) domain, which carries the growing peptide chain (Figure 1-4B).

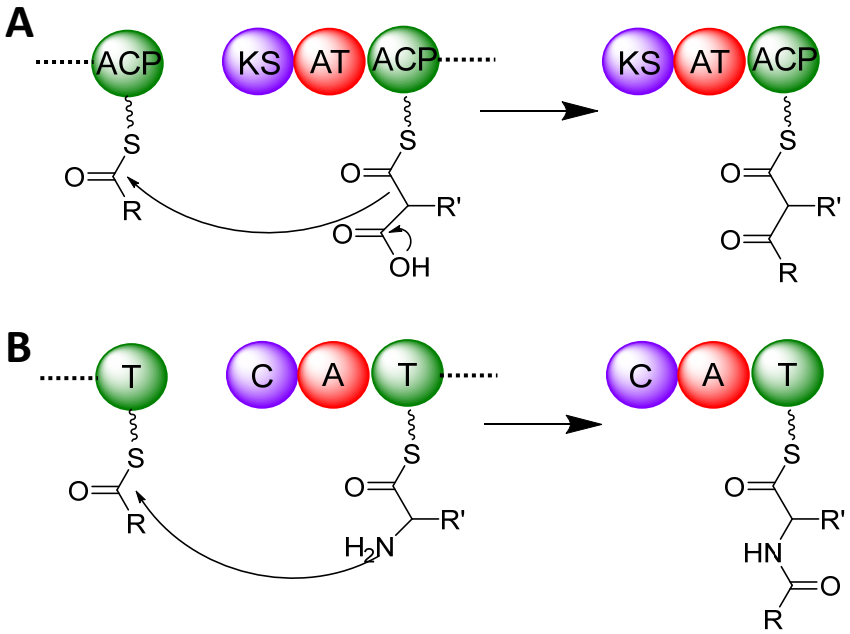


Figure 1-4. Scheme showing the core catalytic domains of PKSs (A) and NRPSs (B).

Additional enzymes that may either be part of the PKS/NRPS or standalone can also modify the nascent PK/NRP chain during or after assembly, further contributing to the diversity and complexity of these classes of natural products. Common optional domains for PKSs include the ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) domains, which can modify the β-ketone, and many PKS/NRPS assembly lines also have a terminal thioesterase (TE) domain that can catalyze macrocyclization.[48] Post-tailoring

enzymes, such as oxidoreductases, methyltransferases, and glycosyltransferases, are also found in many PKS/NRPS biosynthetic gene clusters, providing another source of structural variation. The diversity of the building blocks selected by the AT domains of PKSs and the A domains of NRPSs is another contributor to PK and NRP structural diversity. Recent studies have identified AT domains that select a number of different extender units aside from the usual malonyl- and methylmalonyl-CoA; for example, an AT domain from the FK506 PKS was shown to be capable of accepting ethylmalonyl-, propylmalonyl-, and allylmalonyl-CoA[50] while an AT domain from the antimycin PKS was able to accept linear, branched, cyclic, halogenated, and alkynoic alkylmalonyl-CoAs.[51-53] Likewise, many NRPS assembly lines have been shown to incorporate unusual modified amino acids such as phenylglycines, and halogenated and β-hydroxy amino acids.[54, 55] Given all of these aspects, the potential to create new molecules by engineering assembly lines composed of various modules and domains is vast. Furthermore, many PK/NRP natural products are produced as congeners, implying that these natural product assembly lines and tailoring enzymes have some degree of inherent promiscuity.

However, attempts to produce new bioactive PK and NRP compounds through combinatorial biosynthesis and specifically, module and domain swapping, have resulted in mixed success. While groups have successfully used such strategies to produce libraries of erythromycin[56] (Figure 1-3B) and daptomycin[57, 58] analogues, in many cases, engineered fusion assembly lines have resulted in non-functional recombinant lines, extremely low product titers, and undesired hybrid products.[4, 59] In fact, in a study that generated 154 bimodular combinations of PKS modules, less than half of the combinations resulted in the production of the expected triketide lactone, and less than 10% resulted in product titers exceeding 10 mg/L.[59] These simple mix-and-match experiments generally fail because they do not account for the extensive and orthogonal protein-protein interactions required for substrate channeling and catalysis in a complex pathway, though efforts to more strategically dissect and re-form these enzymatic assembly lines by studying and engineering interdomain and intermodular linkers,[60, 61] PKS docking domains,[61-63] and NRPS communication-mediating domains[64, 65] are increasing. Furthermore, the studies of most PKS and NRPS machineries have primarily focused on their functions in isolated, native assembly-line settings rather than on understanding the natural rules that promote successful assembly line modification. Thus, despite the promising functional and chemical diversity offered by PKS and NRPS assembly lines, combinatorial biosynthesis has not advanced as far as desired. Chapter 5 discusses the potential of using a set of naturally combinatorial NRPS/PKS-hybrid gene clusters to gain a better understanding of the general programming rules for recombining enzymatic domains and modules. Such studies along with advances in synthetic biology and improved structural knowledge of PKSs and NRPSs may facilitate future efforts to generate new "unnatural" natural products via combinatorial biosynthesis.[36, 66-70]

Chapter 2. Elucidating the Biosynthesis of Physostigmine

Parts of this chapter have been adapted from the following with permission:
Liu, J., Ng, T., Rui, Z., Ad, O., Zhang, W. "Unusual acetylation-dependent reaction cascade in the biosynthesis of the pyrroloindole drug physostigmine." *Angew. Chem. Int. Ed. Engl.* **53**, 136-139 (2014).

2.1 Introduction

Physostigmine is a tryptophan-derived pyrroloindole alkaloid that reversibly inhibits acetylcholinesterase through binding of the carbamate group to the active site serine of the enzyme.[5] It is used clinically as a parasympathomimetic drug to treat a wide variety of disorders, including Alzheimer's disease, glaucoma, delayed gastric emptying, and orthostatic hypertension. Furthermore, physostigmine can cross the blood-brain barrier and is used to counteract the effects on the central nervous system of overdoses of atropine, scopolamine, and other anticholinergic drugs.[71, 72] Because of its potent biological activity and unique structure (Figure 2-1), physostigmine (**1**) has attracted much interest from organic chemists.[73-75] However, the enzymatic machinery that directs the biosynthesis of physostigmine has not been elucidated until now.

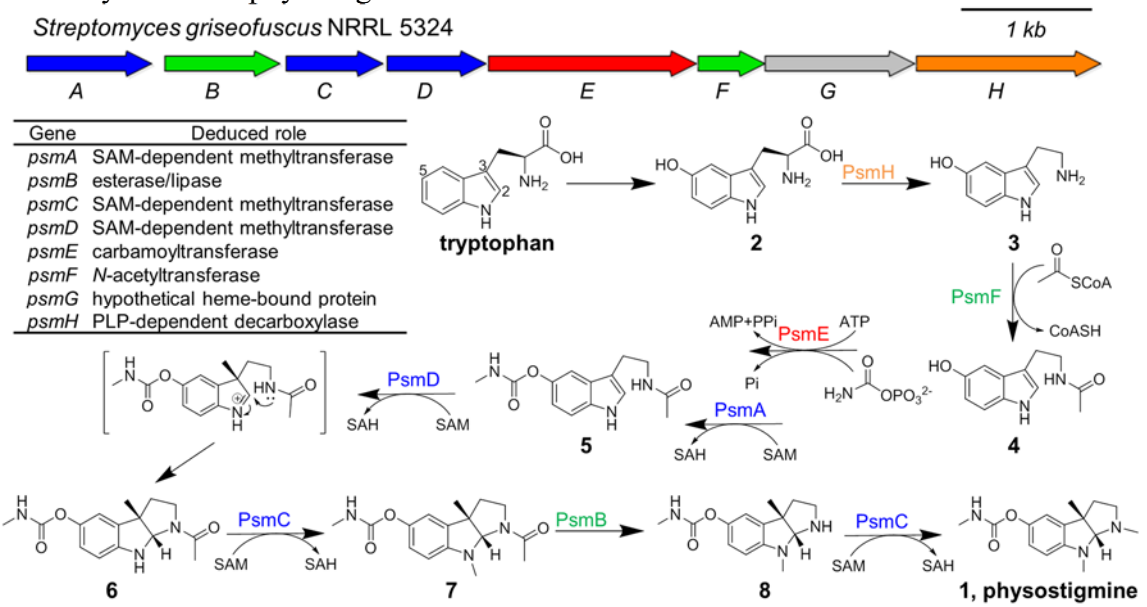


Figure 2-1. Physostigmine biosynthetic gene cluster and characterized biosynthetic pathway. SAH=*S*-adenosylhomocysteine.

Physostigmine was initially isolated from the seeds of *Physostigma venenosum*.[71] Several structurally related pyrroloindole alkaloids, such as eseramine and physovenine, have also been isolated from this source (Figure 2-2).[5, 76] In addition, many oligomeric pyrroloindole alkaloids, including chimonanthines, hodgkinsines, and psychotridines, have been isolated from other plants, and these compounds are of interest because of their potent analgesic activities.[77-79] Despite the prevalence of the pyrroloindole ring system in natural products, little is known about the enzymes responsible for the synthesis of the aforementioned alkaloids. Only two types of enzymes have previously been reported to

promote the formation of the pyrroloindole ring by exploiting the nucleophilicity of the indole C3 atom. These characterized enzymes include prenyltransferases and monooxygenases, which respectively prenylate and hydroxylate the indole C3 atom in the biosynthesis of pyrroloindole-containing peptide natural products, such as aszonalenin[80] and himastatin.[81] In contrast, physostigmine is methylated at the indole C3 atom, prompting us to initially hypothesize that its biosynthesis proceeds by the decarboxylation of tryptophan, followed by a rare C3-methylation of the indole ring and cyclization by attack of the primary-amine moiety onto the iminium ion. Further substitutions, including hydroxylation, carbamylation, and three *N*-methylation steps, then presumably complete the modification of the pyrroloindole skeleton.

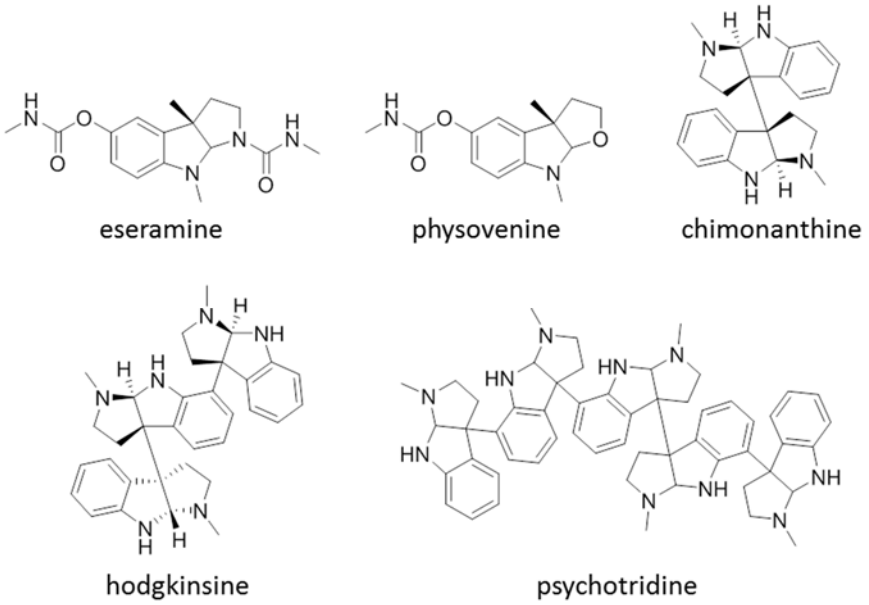


Figure 2-2. Molecular structures of related and oligomeric pyrroloindole alkaloids.

Here, we report the characterization of the gene cluster responsible for physostigmine biosynthesis. In addition to the identification of the first indole C3-methyltransferase, we show that the biosynthetic pathway to physostigmine involves an unusual reaction cascade consisting of highly coordinated methylation steps and acetylation/deacetylation reactions. We further demonstrate the ability to generate physostigmine analogues using different combinations of biosynthetic enzymes and the potential for producing physostigmine in a more genetically tractable heterologous host, *E. coli*.

2.2 Results

2.2.1 Biosynthetic gene cluster identification and confirmation

Because secondary metabolite biosynthetic genes are typically clustered in microbial hosts, the finding that physostigmine was also produced by the bacterium *Streptomyces griseofuscus* NRRL 5324[82] provided us with the opportunity to quickly identify the genes involved in physostigmine biosynthesis. The genome of *S. griseofuscus* was sequenced and subjected to BLASTP analysis with a carbamoyltransferase as a probe. The bioinformatics search identified one putative gene cluster (GenBank accession no.

KF201694), which spans 8.5 kb and encodes eight open reading frames designated *psmA-H* (Figure 2-1). The deduced roles of the gene cluster products did not align well with the originally proposed biosynthetic pathway: The functions of PsmB, F, and G are unknown; four methylation reactions are needed, but only three methyltransferases (PsmA, C, and D) are encoded; and no typical hydroxylase is encoded.

To confirm the involvement of this DNA region in physostigmine biosynthesis, we carried out heterologous expression of this cluster in non-physostigmine-producing *Streptomyces* hosts. A fosmid library of *S. griseofuscus* was constructed, and three different fosmids containing the putative gene cluster were identified by PCR screening. The fosmids were then subcloned into an *E. coli-Streptomyces* shuttle vector and introduced into *S. albus* R1 and *S. lividans* K4-114. Analysis of the culture extracts by liquid chromatography-high resolution mass spectrometry (LC-HRMS) and comparison with the physostigmine standard showed heterologous production of **1** with all three fosmids in both *Streptomyces* hosts, thus verifying that the identified gene cluster is directly involved in the biosynthesis of physostigmine (Figure 2-3). Interestingly, a homologous gene cluster was also found in the published genome of *Streptomyces albulus* CCRC 11814, a known poly-L-lysine producer,[83] although no physostigmine production has been reported from this organism.

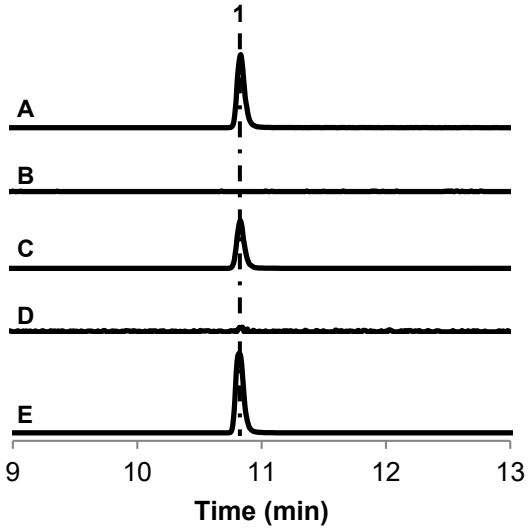


Figure 2-3. Heterologous production of physostigmine in *Streptomyces*. Extracted ion chromatograms showing (A) heterologous production of **1** in *S. lividans*, but not in (B) *S. lividans* with an empty vector; (C) heterologous production of **1** in *S. albus*, but not in (D) wild-type *S. albus*; (E) production of **1** by wild-type *S. griseofuscus*.

The functions of PsmA-H in physostigmine biosynthesis were further demonstrated by *in vitro* reconstitution. A one-pot assay with all eight of the purified enzymes (Figure 2-4) and their possible cosubstrates (*S*-adenosylmethionine (SAM), carbamoyl phosphate, adenosine 5-triphosphate (ATP), and acetyl coenzyme A (acetyl-CoA)) was performed. We propose that tryptophan is initially hydroxylated by a tryptophan 5-hydroxylase, which was confirmed to be encoded elsewhere on the genome by BLASTP analysis. Indeed, when 5-hydroxytryptophan (**2**) was used as a substrate, **1** was successfully produced in the one-pot assay, as demonstrated by LC-HRMS analysis. Individual

enzymes were then systematically removed from the assay to determine their necessity for physostigmine biosynthesis. Seven of the enzymes, PsmA-F, H, were determined to be essential for the formation of **1** from **2** (Figure 2-5A).

Protein	Role	Size (aa)	Size (kDa)
PsmA	methyltransferase	336	41
PsmB	deacetylase	310	38
PsmC	methyltransferase	262	34
PsmD	methyltransferase	268	35
PsmE	carbamoyltransferase	556	62
PsmF	*N*-acetyltransferase	183	25
PsmG	hypothetical protein	402	46
PsmH	decarboxylase	494	59

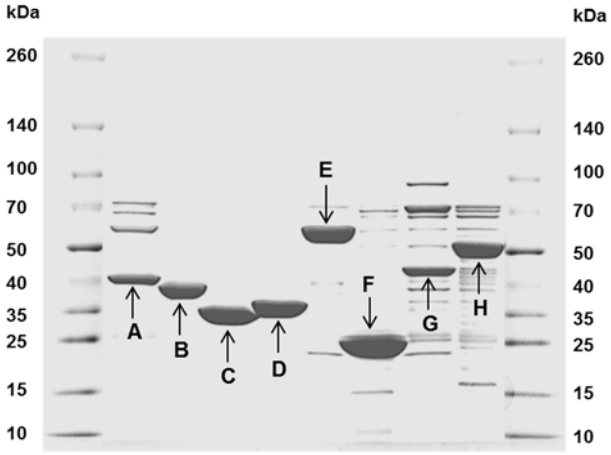


Figure 2-4. Physostigmine biosynthetic enzymes purified from *E. coli*.

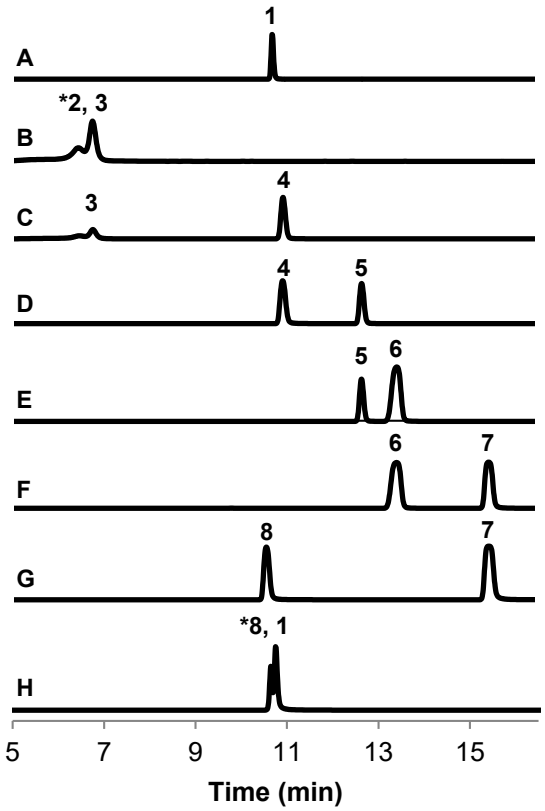


Figure 2-5. Extracted ion chromatograms for the characterization of PsmA-F, H. The calculated mass with a 10 ppm mass error tolerance was used. (A) Reaction of **2** under the catalysis of PsmA-F, H to generate **1**; (B) PsmH-catalyzed reaction of **2** to generate **3**; (C) PsmF-catalyzed reaction of **3** to generate **4**; (D) PsmE/PsmA-catalyzed reaction of **4** to generate **5**; (E) PsmD-catalyzed reaction of **5** to generate **6**; (F) PsmC-catalyzed reaction of **6** to generate **7**; (G) PsmB-catalyzed reaction of **7** to generate **8**; (H) PsmC-

catalyzed reaction of **8** to generate **1**. *Substrate and product have overlapping retention times.

2.2.2 Biosynthetic pathway characterization

The timing of each enzymatic reaction was delineated by a series of *in vitro* and *in vivo* studies. We first set out to determine the substrate (L-tryptophan or **2**) for the pyridoxal 5′-phosphate (PLP)-dependent decarboxylase, PsmH. Incubation of purified PsmH with tryptophan and **2** resulted in the formation of tryptamine and 5-hydroxytryptamine (**3**), respectively (Figure 2-5B). Comparison of the kinetic parameters of PsmH showed that it clearly prefers the hydroxylated substrate: the k_{cat}/K_m value for **2** ($8.6 \pm 0.9\ min^{-1}\ mM^{-1}$) was 650-fold higher than that for tryptophan ($0.013 \pm 0.002\ min^{-1}\ mM^{-1}$). This result is consistent with the one-pot assay in which **2** was efficiently utilized as a substrate. We propose that **3** then undergoes *N*-acetylation catalyzed by PsmF, an *N*-acetyltransferase homolog. The enzymatic reaction with PsmF, acetyl-CoA, and **3** yielded the expected product **4** (Figures 2-1 and 2-5C), thus confirming the function of PsmF as a 5-hydroxytryptamine *N*-acetyltransferase. These early pathway intermediates, **3** and **4**, also known as serotonin and normelatonin, respectively, are common secondary metabolites produced by many other organisms as well.[84-86]

The subsequent enzymatic reaction on **4** could be either carbamylation catalyzed by PsmE or C3-methylation catalyzed by one of the methyltransferases. The incubation of PsmE, carbamoyl phosphate, ATP, and **4** resulted in the formation of carbamylated **4**, which could be further methylated by PsmA to yield **5**, which showed a UV absorption spectrum similar to that of **4** (Figure 2-1 and Appendix A). The production of **5** from **4** confirmed the roles of PsmE as a normelatonin *O*-carbamoyltransferase and PsmA as a carbamoyl *N*-methyltransferase. Compound **5** also accumulated as one of the major metabolites in the culture extracts of the Δ*psmD* mutant of *S. griseofuscus* (Figure 2-6 and Table 2-1), thus indicating that carbamylation precedes the methylation catalyzed by PsmD. Although PsmE was also capable of carbamylating **2**, **3**, and eseroline (**9**), further kinetic characterization of the enzyme showed **4** to be the preferred substrate (Figure 2-7).

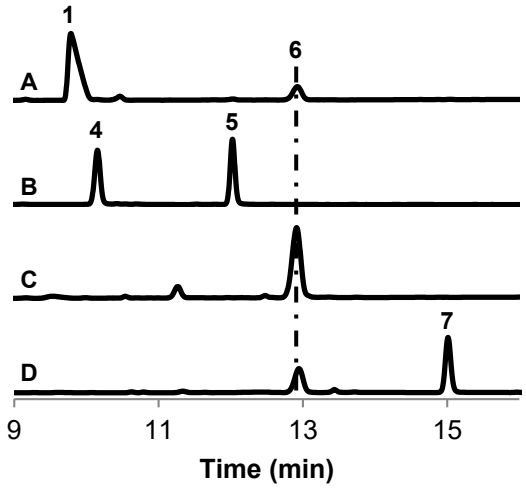
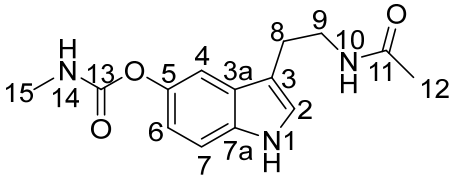


Figure 2-6. HPLC-UV analysis of metabolites produced by wild-type and mutant *S. griseofuscus* strains. (A) Wild-type *S. griseofuscus* (245 nm); (B) Δ*psmD* (280 nm); (C) Δ*psmC* (245 nm); (D) Δ*psmB* (245 nm).

Table 2-1. NMR data for **5** in $CDCl_3$.



Position	δ_c	δ_H (*J* in Hz)	COSY	HMBC
1-NH		8.16, s		2, 3, 3a, 7, 7a
2	123.64	7.02, s		3, 3a, 7a, 8
3	113.44			
3a	127.94			
4	111.39	7.29, d (1.8)		3, 5, 6, 7a
5	144.86			
6	117.06	6.95, dd (9.0, 1.8)	7	4, 5, 7, 7a
7	111.87	7.30, d (9.0)	6	3a, 4, 5
7a	134.37			
8	25.33	2.90, t (6.3)	9	2, 3, 3a, 9
9	40.23	3.54, t (6.3)	8	3, 8, 11
11	171.26			
12	23.39	1.94, s		11
13	156.68			
15	28.08	2.91, d (4.5)		13

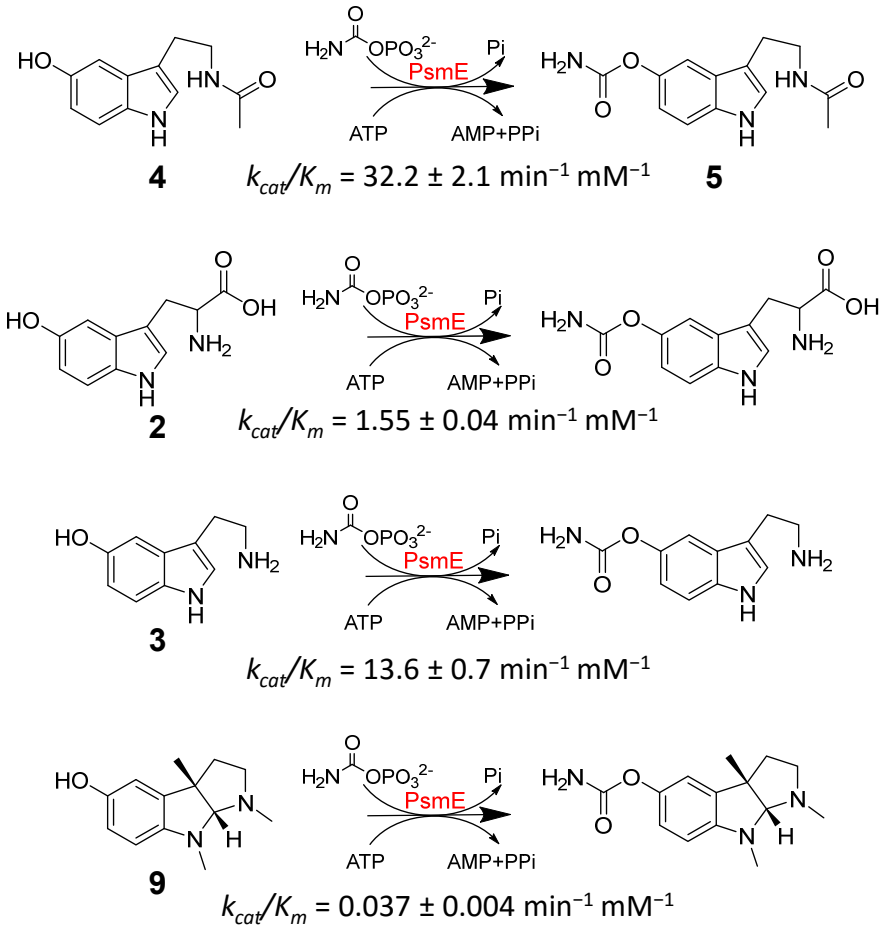
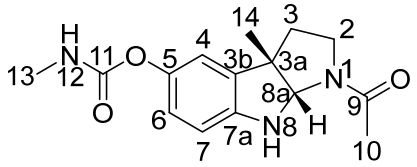


Figure 2-7. Substrate specificity of PsmE. PsmE is capable of carbamylating **2**, **3**, **4**, and **9** though kinetic parameters indicate a preference for **4**.

We next probed the putative function of PsmD as a C3-methyltransferase by using purified **5** as a substrate *in vitro*. C3-methylation would disrupt the conjugated indole ring and lead to a significant change in the UV absorption spectrum of the product. Accordingly, the enzymatic assay consisting of PsmD, SAM, and **5** resulted in the formation of **6**, which has a UV absorption spectrum distinct from that of **5** (Figures 2-1, 2-5E, and Appendix A. The molecular structure of **6** was confirmed by NMR spectroscopic analysis (Table 2-2), and the production of **6** confirmed the role of PsmD as a C3-methyltransferase. It is proposed that the C3-methylated product is highly unstable and undergoes spontaneous cyclization by nucleophilic attack of the amine onto the iminium ion to form the pyrroloindole skeleton (Figure 2-1). Similar cyclization mechanisms have been observed in the biosynthesis of aszonalenin and himastatin, whereby C3-prenylation and C3-hydroxylation, respectively, promote ring closure.[80, 81, 87] Notably, PsmD could also methylate **4**, but it did not exhibit any activity toward **3**, thus indicating that *N*-acetylation is required for *C*-methylation.

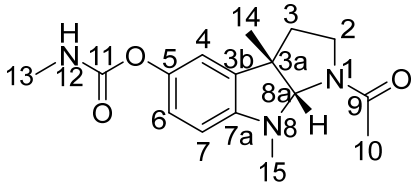
Table 2-2. NMR data for **6** in $CDCl_3$.



Position	δ_c	δ_H (*J* in Hz)	COSY	HMBC
2	47.70	3.36, dt (12.6, 7.2) 3.66, ddd (12.6, 7.2 1.8)	3	3, 3a, 8a, 9
3	36.96	2.17, dt (12.6, 7.2) 2.36, ddd (12.6, 7.2 1.8)	2	2, 3a, 3b, 8a, 14
3a	52.91			
3b	135.21			
4	116.56	6.92, d (1.8)		3a, 5, 6, 7, 7a
5	145.70			
6	122.01	6.87, dd (9.0, 1.8)	7	4, 5, 7a
7	112.91	6.75, d (9.0)	6	3a, 3b, 4, 5, 6, 7a
7a	142.71			
8a	82.67	5.32, s		2, 3, 3a, 3b, 7a, 9, 14
9	172.84			
10	21.55	2.13, s		2, 9
11	156.54			
13	27.81	2.90, d (4.5)		11
14	23.87	1.45, s		3, 3a, 3b, 8a

To complete the biosynthesis of **1** from **6**, one deacetylation and two *N*-methylation reactions on the pyrroloindole skeleton are necessary. These reactions are presumably catalyzed by PsmB, an esterase/lipase homolog, and PsmC, another methyltransferase homolog, respectively. To examine the timing of these tailoring reactions, *psmB* and *psmC* were individually deleted in frame by double cross-over in *S. griseofuscus*. Since **6** was observed to accumulate as the major metabolite in the Δ*psmC* culture extracts (Figure 2-6), we propose that methylation by PsmC directly follows methylation by PsmD. Analysis of the culture extracts of the Δ*psmB* mutant showed accumulation of a new major metabolite, **7**, which was revealed to be an *N*-methylated derivative of **6** by NMR spectroscopic analysis (Figure 2-6 and Table 2-3). It is thus proposed that PsmC first catalyzes one *N*-methylation on **6** to yield **7**, which is subsequently deacetylated by PsmB to yield **8**. Final *N*-methylation on **8** then completes the biosynthesis of **1** (Figure 2-1).

Table 2-3. NMR data for **7** in $CDCl_3$.



Position	δ_c	δ_H (*J* in Hz)	COSY	HMBC
2	47.88	3.36, dt (12.6, 7.2) 3.63, ddd (12.6, 7.2, 4.5)	3	3a, 3, 8a, 9
3	39.21	1.97, dt (12.6, 7.2) 2.18, ddd (12.6, 7.2, 4.5)	2	2, 3a, 3b, 8a, 14
3a	51.58			
3b	134.97			
4	116.32	6.79, d (1.8)		3a, 5, 6, 7, 7a
5	143.45			
6	121.48	6.83, dd (9.0, 1.8)	7	4, 5, 7a
7	106.59	6.34, d (9.0)	6	3a, 3b, 4, 5, 7a
7a	148.18			
8a	87.98	5.42, s		2, 3, 3a, 3b, 7a, 9, 14, 15
9	171.41			
10	22.89	2.12, s		2, 9
11	156.42			
13	28.02	2.88, d (5.4)		11
14	24.70	1.40, s		3, 3a, 3b, 8a
15	34.50	2.97, s		7a, 8a

To confirm this tailoring-reaction cascade, we biochemically reconstituted each transformation step (Figure 2-5). An enzymatic assay with PsmC, SAM, and **6** showed the formation of **7** as a major product, thus confirming the function of PsmC as an *N*-methyltransferase. Incubation of purified PsmB and **7** resulted in the formation of **8**, thus indicating that PsmB functions as a deacetylase on **7**. Finally, we incubated the three methyltransferases PsmA, C, and D individually with purified **8** and SAM to identify the enzyme responsible for the last *N*-methylation step. Only the assay with PsmC yielded **1**, thus demonstrating that PsmC catalyzes methylation reactions both before and after deacetylation. Although PsmB also exhibited activity toward **6** *in vitro*, the resulting deacetylated product could not be methylated twice by PsmC to form **1**. Instead, we only observed the formation of a product with the same mass as **8**. This result demonstrates

that the *N*-methylation and deacetylation reactions occur in a very precise order for physostigmine biosynthesis.

2.2.3 Evaluation of pathway flexibility and generation of analogues

We next explored the flexibility of the physostigmine biosynthetic pathway by producing several analogs using *in vitro* biosynthesis. These analogs were generated using a combination of PsmA-E with **4** as the starting substrate. The relaxed substrate specificities of PsmB-D resulted in the formation of products without a carbamoyl group (**10**) and its corresponding methyl group (**9**) as confirmed by LC-HRMS analysis and comparison with the eseroline standard (Figure 2-8). The production of **11** and **12** further showed that deacetylation by PsmB does not require *N*-methylations (Figure 2-8).

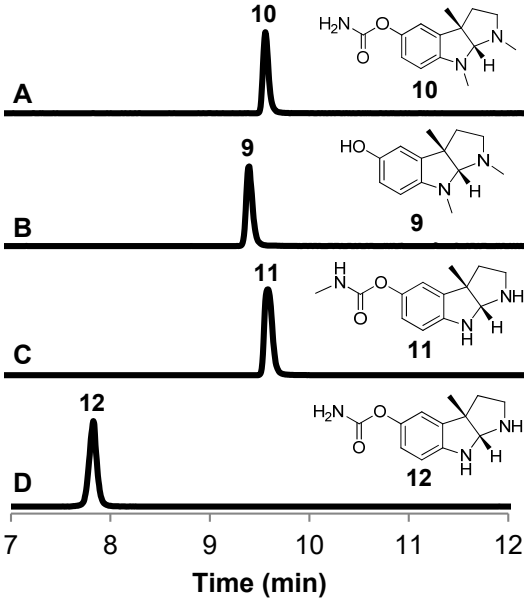


Figure 2-8. Extracted ion chromatograms showing in vitro production of physostigmine analogs from **4**. The calculated mass with 10-ppm mass error tolerance was used. Structures for **10**, **11**, and **12** are tentative. (A) Reaction with PsmB-E. A small amount of **10** was also observed to accumulate in the cultures of the Δ*psmA* mutant. (B) Reaction with PsmB-D; (C) Reaction with PsmA, B, D, and E. A small amount of **11** was also observed to accumulate in the cultures of the Δ*psmC* mutant. (D) Reaction with PsmB, D, and E.

2.2.4 Physostigmine production in *E. coli*

Given the versatility of *E. coli* as a heterologous host, we also endeavored to engineer a strain of *E. coli* capable of producing **1**. An initial attempt to introduce the entire physostigmine biosynthetic gene cluster as a single operon expressed under a T7 promoter failed to produce **1**, and further analysis of this strain's protein expression indicated that not all of the biosynthetic enzymes were being expressed. One possible reason for this may be the inability for *E. coli* to recognize the native ribosomal binding sites (RBSs) in the gene cluster. Consequently, a refactoring approach in which PsmA-E were individually expressed under T7 promoters across three different plasmids was undertaken. The introduction of these five genes into *E. coli* BL21 Gold (DE3) and

feeding of this strain with **4** then resulted in the production of **1** (Figure 2-9). Notably, quite a bit of the intermediates, **6** and **7**, was also observed to be produced by this strain, suggesting that the PsmC and PsmB may be the bottleneck enzymes in this pilot *E. coli* heterologous expression system.

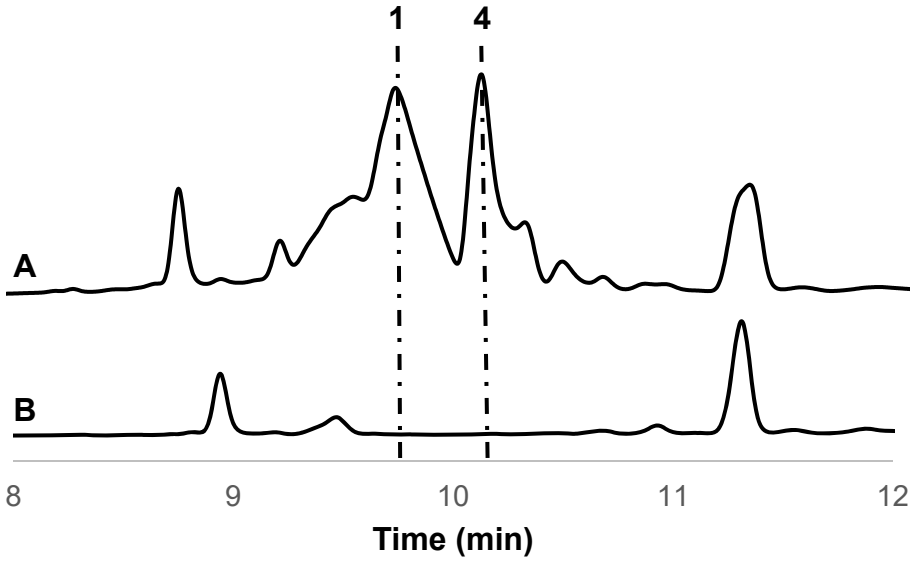


Figure 2-9. HPLC chromatograms showing production of **1** from **4** in *E. coli*. (A) *E. coli* expressing PsmA-E fed with **4** produces **1** (245 nm) while (B) no production of **1** is observed without the feeding of **4**.

2.3 Discussion

In summary, we have identified and dissected the seven enzymes and eight reaction steps involved in the biosynthesis of the pyrroloindole drug physostigmine from 5-hydroxytryptophan.[88] Initial decarboxylation of **2**, followed by acetylation of the amine, results in the well-known intermediate normelatonin, which is modified at the C5 hydroxy group by carbamylation and methylation. C3-methylation of the indole ring yields a highly reactive iminium ion, which is then attacked by the amine nucleophile to form a pyrroloindole skeleton. Further reactions, including tandem *N*-methylation, *N*-deacetylation, and *N*-methylation, complete the biosynthesis of physostigmine (Figure 2-1). This biosynthetic logic is rather unusual in that it involves an acetylation and subsequent deacetylation. We then showed that this unexpected acetylation-elimination is due to the highly coordinated nature of the methylation and acetylation/deacetylation reactions involved in physostigmine biosynthesis. Specifically, acetylation is necessary for indole C3-methylation, whereas the two *N*-methylation reactions of the pyrroloindole ring must occur in concert with deacetylation; no alternative route appears to result in physostigmine formation.

This biosynthetic gene cluster also features two unique methyltransferases. To our knowledge, PsmD is the first enzyme that has been characterized to catalyze indole C3-methylation, and PsmA is the first enzyme that has been characterized to methylate a carbamoyl group. Thus, this study not only provides some insight into the biosynthesis of other structurally related pyrroloindole alkaloids, but it also expands the repertoire of enzymes that can functionalize the indole C3 atom and the carbamoyl moiety.

Interestingly, an unusual hypothetical protein predicted to have a heme binding site was also identified in the gene cluster. While attempts to generate a Δ*psmG* mutant of *S. griseofuscus* and to probe the activity of PsmG *in vitro* have thus far been unsuccessful, we propose that PsmG may have a regulatory function in physostigmine biosynthesis.

Lastly, our *in vitro* studies using different combinations of the physostigmine biosynthetic enzymes demonstrate the feasibility of generating new pyrroloindole compounds related to physostigmine via biosynthesis. Such compounds could be further derivatized enzymatically or through traditional chemical synthesis to produce new analogues with different functional groups. Along these lines, our *E. coli* heterologous expression system may also prove to be advantageous for combinatorial biosynthesis approaches to generate new pyrroloindole compounds. Although the physostigmine biosynthetic pathway still needs to be optimized in *E. coli*, particularly since bottleneck enzymes have been identified, the relative ease with which precursor feeding experiments and gene manipulation can be performed in *E. coli* makes it an attractive host for combinatorial biosynthesis. Examples of approaches to extend the diversity of physostigmine analogues using *E. coli* include the feeding of tryptophan analogues and the introduction of tryptophan halogenases,[89, 90] prenyltransferases,[87, 91] or the recently reported cytochrome P450 enzymes capable of catalyzing the formation of dimeric pyrroloindole compounds.[92, 93]

2.4 Materials and Methods

Construction and screening of the S. griseofuscus genomic DNA library

The *S. griseofuscus* fosmid library was prepared using the CopyControl HTP Fosmid Library Production Kit (Epicentre) following standard protocols. The resulting library of 5,500 clones was screened by PCR using primers that bind to the end of *psmE* and the beginning of *psmH* (Table 2-4). Clones with this region of DNA were then induced to high-copy number following standard protocols and sequenced to ensure they contained the entire putative biosynthetic gene cluster.

Table 2-4. Primers used in this study.

Primer	**Sequence (5'→3')**	**Description**
Phy-CT-5	AACCACAAAGGCCGCGGTTT	Used for fosmid library screening and confirming the uptake of the heterologous expression vector
Decarb-R	CAGCACGATCCGTTCGAGGT	
pSET152L-phyfos-F	TTCCTGGTGGGATCCTCTAGAGTCGACCTGCAGGCATGCAGCGATATCGAATTCGTAATC	Used for cloning fosmid inserts into pSET152
pSET152L-phyfos-R	TATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCAGGATCCTCTAGAGTCGACCT	
Phy-MT1-F	GGTATTGAGGGTCGCATGGCCGCAAGTGCAACGCG	Used for PsmA purification
Phy-MT1-R	AGAGGAGAGTTAGAGCCTCAGGGGACCTGGGCCC	

	GGC	
Phy-TE-PET30F	GGTATTGAGGGTCGCGTGCTCGAACCCCGCATCA	Used for PsmB purification
Phy-TE-PET30R	AGAGGAGAGTTAGAGCCTCAGGCGAACGCGGTGGTCA	
Phy-MT2-F	GGTATTGAGGGTCGCATGGCAGACGACGCGCACGC	Used for PsmC purification
Phy-MT2-R	AGAGGAGAGTTAGAGCCTCAGGCGCGGGTCAGGACCC	
Phy-MT3-F	GGTATTGAGGGTCGCATGATGCAGGGACAGCCGCA	Used for PsmD purification
Phy-MT3-R	AGAGGAGAGTTAGAGCCTCAGCGCTCGTCTCGGGTGA	
CarbT-PET24F	AAACATATG CTGATCTGCGGACTGAA	Used for PsmE purification
CarbT-PET24R	AAACTCGAG TGCGCGGCGGTAGAGGCGGC	
Phy-NAT-F	GGTATTGAGGGTCGCATGAACACCTTCCGGACCGC	Used for PsmF purification
Phy-NAT-R	AGAGGAGAGTTAGAGCCTCAGTCGGAGTGGTCATGGATCG	
NdeI-TDO-F	AAA**CATATG**ACCACTCCGACTGAGCG	Used for PsmG purification
HindIII-TDO-R	TAT**AAGCTT**TCACAGGAGTGGCTCCTCAA	
Decarb-PET30F	GGTATTGAGGGTCGCGTGAAGCCCGCTGAGGCGAA	Used for PsmH purification
DeCarb-PET30R	AGAGGAGAGTTAGAGCCCTACCCGTTCAGCGCGTCAG	
specR-F	GGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTGTAGCTTGCAGTGGGCTTAC	Used for cloning *aadA* into fosmid
specR-R	CCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTTTATTTGCCGACTACCTTGG	
Phy MT1 KO-F	GAATTGCAACGCCTGTACTTCCACGGAAGGGACGGCATGATTCCGGGGATCCGTCGACC	Used for Δ*psmA* mutant construction
Phy MT1 KO-R	CTGGAGCCGGCCTCGTACGGCGTGCCGGCCTGGCGCTCATGTAGGCTGGAGCTGCTTC	
Phy TE KO-F	TCTCCTGCGGCTCCCAAGGAAAGGTCCTTCACCTCCGTGATTCCGGGGATCCGTCGACC	Used for Δ*psmB* mutant construction
Phy TE KO-R	CATGCCAGGTCCTCTCTGTCTGGAACAGGGTCTGGATCATGTAGGCTGGAGCTGCTTC	
Phy MT2 KO-F	TGATCCAGACCCTGTTCCAGACAGAGAGGACCTGGCATGATTCCGGGGATCCGTCGACC	Used for Δ*psmC* mutant construction
Phy MT2 KO-R	GTGCGGCTGTCCCTGCATCATGCTCGGTTTCCCTCCTCATGTAGGCTGGAGCTGCTTC	
MT1-Duet-F	AAA**CTGCAG**ATGGCCGCAAGTGCAACGCG	Used for PsmA expression
MT1-Duet-R	TAT**AAGCTT**TCAGGGGACCTGGGCCCGGC	
CarbT-Duet-R	AAACTCGAG TCATGCGCGGCGGTAGAGGC	Used with CarbT-PET24F

		for PsmE expression
MT2-Duet-F	AAA**CTGCAG**ATGGCAGACGACGCGCACGC	Used for PsmC expression
MT2-Duet-R	TAT**AAGCTT**TCAGGCGCGGGTCAGGACCC	
TE-Duet-F	AAA**CTGCAG**GTGCTCGAACCCCGCATCAA	Used for PsmB expression
TE-Duet-R	TAT**AAGCTT**TCAGGCGAACGCGGTGGTCA	
MT3-Duet-F	AAA**CATATG**ATGCAGGGACAGCCGCA	Used for PsmD expression
MT3-Duet-R	AAA**CTCGAG**TCAGCGCTCGTCTCGGGTGA	

*restriction endonuclease sites are bolded

Heterologous expression of the physostigmine gene cluster in Streptomyces

Fosmids containing the physostigmine gene cluster were digested, and the fragment with the gene cluster was purified and ligated to pSET152 following standard protocols. For heterologous expression in *S. albus* R1, the resulting construct was transformed into *E. coli* WM6026 (a diaminopimelic acid auxotroph) for conjugation with *S. albus* following standard protocols.[94] For heterologous expression in *S. lividans* K4-114, the construct was introduced into *S. lividans* by PEG-assisted protoplast transformation.[94] Successful transformants were selected by apramycin resistance and confirmed by PCR using the primers used for library screening (Table 2-4). The production, isolation, and analysis of physostigmine from these heterologous hosts is described below.

Gene disruption in S. griseofuscus

In vivo generation of targeted mutations in *S. griseofuscus* was achieved by conjugative transfer of disruption plasmids from *E. coli* WM6026 to *S. griseofuscus* according to standard protocols.[94] The knockout cassettes were constructed using the ReDirect technology.[95] In summary, the *aadA* gene from pIJ778 was cloned into fosmid pE12F containing the entire physostigmine gene cluster, and the resulting plasmid was introduced into *E. coli* BW25113/pIJ790 by electroporation. The *nptII-oriT* cassette amplified by PCR from pCR-Blunt and pIJ778 was then introduced to replace the entire gene of interest (*psmA*, *psmB*, *psmC*, or *psmD*) using PCR targeting and λ-red-mediated recombination. The resulting knockout cassette was transformed into *E. coli* WM6026 for conjugation with *S. griseofuscus*. The double-crossover strain was obtained from antibiotic selection (kan^R $spec^S$) and confirmed by PCR. Mutant analysis was carried out following the procedures detailed below.

Small-scale production and isolation of physostigmine and physostigmine biosynthetic intermediates

Wild-type *S. griseofuscus* was grown side by side as a control with the heterologous expression and mutant strains. To screen for physostigmine production, starter cultures in TSB medium with the appropriate antibiotic were inoculated with mycelia and grown at 30ºC, 200 rpm for 48 h. 1 mL of the starter culture was then inoculated into production medium (25 mL; 20 g/L soy flour and 20 g/L mannitol) with the appropriate antibiotic, and the cultures were grown at 30ºC, 200 rpm for 5 days. For *S. lividans* cultures, R5 medium[94] with apramycin (50 μg/mL) was used for both starter and production cultures. Physostigmine and its biosynthetic intermediates were extracted from the cell-free supernatant (25 mL) using ethyl acetate (two volumes). The solvent was removed by

rotary evaporation and the residue was redissolved in methanol (1 mL) and analyzed by LC-HRMS (2 µL injection) and LC-UV-MS (10 µL injection). For mutant cultures, 1 mL of supernatant was also lyophilized and redissolved in methanol (1 mL) for LC-UV-MS analysis (10 µL injection). LC-HRMS analysis was performed on an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS instrument, and LC-UV-MS analysis was performed on an Agilent Technologies 6120 Quadrupole LC-MS (with DAD) instrument with an Agilent Eclipse Plus C18 column (4.6 × 100 mm). A linear gradient of 2-50% CH_3CN (vol/vol) over 15 min in H_2O with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min was used.

Large-scale production, purification, and characterization of physostigmine biosynthetic intermediates

For large-scale production, 5 mL starter cultures of *S. griseofuscus* Δ*psmB*, Δ*psmC*, and Δ*psmD* in TSB medium with kanamycin (200 µg/mL) were inoculated with spores and incubated at 30ºC, 200 rpm for 48 h. 1 mL of the starter culture was inoculated into 25 mL seed culture in TSB with kanamycin and incubated at 30ºC, 200 rpm for 48 h. 1 mL of the seed culture was then inoculated into production medium (16 x 50 mL; 20 g/L soy flour and 20 g/L mannitol) with kanamycin, and the cultures were grown at 30ºC, 200 rpm for 5 days. Phyostigmine biosynthetic intermediates were extracted from the cell-free supernatant using two volumes of ethyl acetate. The solvent was removed by rotary evaporation and the residue was redissolved in methanol (5 mL).

Compound **4** from *S. griseofuscus* Δ*psmD*, **5** from *S. griseofuscus* Δ*psmC*, and **6** from *S. griseofuscus* Δ*psmB* were purified by reverse-phase high-performance liquid chromatography (RP-HPLC, Agilent 1260 HPLC with DAD) on a semi-preparative C18 column (10 × 250 mm, Vydac) with a linear gradient of 10-50% CH_3CN (vol/vol) over 13 min and 95% CH_3CN (vol/vol) for a further 5 min in H_2O with 0.025% (vol/vol) trifluoroacetic acid (TFA) at a flow rate of 3 mL/min. Fractions containing physostigmine biosynthetic intermediates were collected manually and concentrated under vacuum. These fractions were further purified by RP-HPLC with an Inertsil ODS-4 column (4.6 × 250 mm, GL Sciences Inc.) using an isocratic program of 35% CH_3CN (vol/vol) in H_2O with 0.025% TFA (vol/vol) at a flow rate of 1 mL/min. The resulting purified compounds **5**, **6**, and **7** (~20 mg/L culture) were dried and analyzed by LC-HRMS and NMR. NMR spectra (1D: ^{1}H, ^{13}C and 2D: HSQC, COSY, HMBC) were recorded on a Bruker Biospin 900 MHz spectrometer with a cryoprobe in chloroform-d ($CDCl_3$; Cambridge Isotope Laboratories).

Cloning, overexpression, and purification of proteins

The cloning, expression, and purification of PsmA-F, and H were carried out as follows: *psmA*, *B*, *C*, *D*, *E*, *F*, and *H* were PCR amplified from genomic DNA extracted from *S. griseofuscus* NRRL 5324. Purified PCR products were ligated to pET-24b (*psmE*) or cloned into pET-30 Xa/LIC (*psmA*, *B*, *C*, *D*, *F*, and *H*) (Novagen) following standard protocols and confirmed by DNA sequencing (Quintara Biosciences). The resulting expression constructs were transformed into *E. coli* BL21 Gold (DE3) for protein expression. Expression and purification for all proteins with a His_6-tag followed the same general procedure and is detailed as follows: in 0.7 L of liquid culture, cells were grown at 37°C in LB medium with kanamycin (50 µg/mL) to an OD_{600} of 0.4. The cells were

cooled on ice for 10 min and then induced with 0.12 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 16 h at 16°C. The cells were harvested by centrifugation (6371 × *g*, 15 min, 4°C), resuspended in 30 mL lysis buffer (25 mM HEPES, pH 7.5, 0.5 M NaCl, 5 mM imidazole), and lysed by homogenization on ice. Cellular debris was removed by centrifugation (27216 × *g*, 1 h, 4°C). Ni-NTA agarose resin was added to the supernatant (2 mL/L of culture) and the solution was nutated at 4°C for 1 h. The protein-resin mixture was loaded onto a gravity flow column, and proteins were eluted with increasing concentrations of imidazole in Buffer A1 (20 mM HEPES, pH 7.5, 1 mM EDTA). Purified proteins were concentrated and buffer exchanged into Buffer A1 + 10% glycerol using Amicon Ultra filters. The final proteins were flash-frozen in liquid nitrogen and stored at -80°C. The approximate protein yields were 3 mg/L for PsmA (41.1 kDa), 14.6 mg/L for PsmB (37.5 kDa), 37.0 mg/L PsmC (34.1 kDa), 46.2 mg/L PsmD (34.6 kDa), 44.5 mg/L PsmE (61.7 kDa), 8.0 mg/L PsmF (25.2 kDa), 7.4 mg/L PsmH (58.9 kDa).

The cloning, expression, and purification of PsmG was carried out similarly with the following differences: Purified *psmG* PCR product was ligated to pCWOri-HisN (M. Chang lab, UC Berkeley) to add a His_{10}-tag. The culture was induced with IPTG (0.5 mM), and δ-aminolevulinic acid (65 μg/mL) was also added at this time. The cells were resuspended in 30 mL lysis buffer (50 mM NaH_2PO_4, pH 7.5, 0.3 M NaCl, 10 mM imidazole, 4% glycerol), and proteins were eluted with increasing concentrations of imidazole in Buffer A2 (50 mM NaH_2PO_4, pH 7.5, 0.3 M NaCl, 4% glycerol). Purified proteins were concentrated and buffer exchanged into Buffer A2 + 10% glycerol. The approximate protein yield for PsmG (45.9 kDa) was 5.7 mg/L.

In vitro assays for pathway characterization and analogue generation

Assays were performed in 50 μL of 10 mM Tris (pH 8.0) containing 2 mM TCEP (for assays with PsmA, C-F, H), 0.1 mM PLP (for assays with PsmH), 2 mM ATP (for assays with PsmE), 2 mM $MgCl_2$ (for assays with PsmE), 2 mM carbamoyl phosphate (for assays with PsmE), 1 mM acetyl CoA (for assays with PsmB), 2 mM SAM (for assays with PsmA, C, and D), 2 mM substrate (**2**, **3**, **4**, **5**, **6**, **7**, or **8**), and 0-100 μM PsmA-H. Reactions were incubated for 2 h at 22ºC and quenched with two volumes of methanol. Precipitated proteins were removed by centrifugation, and the supernatant was analyzed by LC-HRMS and LC-UV-MS. Standard compounds **1**, **2**, **3**, **4**, and **9** were obtained from Sigma and used for comparison with the products of the enzymatic assays. LC-HRMS analysis was performed on an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS instrument, and LC-UV-MS analysis was performed on an Agilent Technologies 6120 Quadruopole LC-MS (with DAD) instrument with an Agilent Eclipse Plus C18 column (4.6 × 100 mm). A linear gradient of 2-50% CH_3CN (vol/vol) over 15 min in H_2O with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min was used.

Kinetic studies of PsmH and PsmE

For kinetic studies of PsmH, a typical assay contained, in a total volume of 250 μL, 50 mM HEPES (pH 7.5), 1 mM TCEP, 0.1 mM PLP, and 0.02 mM **2** or 0.2 mM L-tryptophan. Reactions were initiated by the addition of enzyme (2 μM PsmH). At regular time intervals (0, 5, 10, 15, 20 min for assays with **2** and 0, 30, 60, 90, 120 min for assays with L-tryptophan), 50 μL aliquots were quenched with 100 μL methanol. Precipitated

proteins were removed by centrifugation, and the supernatant was analyzed by LC-HRMS.

For kinetic studies of PsmE, a typical assay contained, in a total volume of 250 µL, 10 mM Tris (pH 8), 1 mM TCEP, 1 mM ATP, 1 mM $MgCl_2$, 1 mM carbamoyl phosphate, and 0.02 mM **2**, **3**, **4**, or **9**. Reactions were initiated by the addition of enzyme (2 µM PsmE). At regular time intervals (0, 5, 10, 20, 30 min), 50 µL aliquots were quenched with 100 µL methanol. Precipitated proteins were removed by centrifugation, and the supernatant was analyzed by LC-HRMS.

Values for k_{cat}/K_m were estimated by calculating the initial rate of product formation and fitting it to the Michaelis-Menten equation. Only data points within the linear range were used to calculate the initial product formation rates. Values reported are the average and standard deviation from at least three independently performed experiments.

Plasmid construction for physostigmine production in E. coli

Primers used in this study are listed in Table 2-4. *S. griseofuscus* genomic DNA was used to PCR amplify *psmA*, *B*, *C*, *D*, and *E*. The *psmA*, *psmB*, and *psmC* genes were digested with PstI/HindIII (Thermo Scientific) and ligated into multiple cloning site 1 (MCS1) of pETDuet-1, pCDFDuet-1, and pACYCDuet-1 (Novagen), respectively, using Quick T4 DNA ligase (New England Biolabs). The resulting plasmids with *psmA* and *psmB* were then digested with NdeI/XhoI for the ligation of *psmE* and *psmD* into MCS2, respectively. Plasmids were isolated using a QIAprep Spin Miniprep Kit (Qiagen) and confirmed by DNA sequencing (UC Berkeley DNA Sequencing Facility).

Production of physostigmine in E. coli

The plasmids, pETDuet-*psmA*-*psmE*, pCDFDuet-*psmB*-*psmD*, and pACYCDuet-*psmC*, were electroporated into *E. coli* BL21 Gold (DE3), and transformants were selected on LB agar plates with carbenicillin (100 µg/mL), spectinomycin (50 µg/mL), and chloramphenicol (34 µg/mL). Single colonies were inoculated into 3 mL of LB with antibiotics and grown overnight at 37°C as a seed culture, of which 0.25 mL was used to inoculate 25 mL of fresh LB medium with antibiotics. The cultures were then grown at 37°C to $OD_{600} \approx 0.4 – 0.6$ before induction with 0.16 mM IPTG and the addition of **4** at a final concentration of 0.1 mM. After induction, the temperature was dropped to 20 °C, and compound production was allowed to proceed for approximately 2 days. Compound extraction and analysis was performed as described above.

Chapter 3. Understanding the Biosynthesis of Epoxyketone Proteasome Inhibitors

Parts of this chapter have been adapted from the following with permission:
Liu, J., Zhu, X., Zhang, W. "Identifying the minimal enzymes required for biosynthesis of epoxyketone proteasome inhibitors." *Chembiochem* **16**, 2585-2589 (2014).

3.1 Introduction

Peptidyl epoxyketones are a class of proteasome inhibitors consisting of a short peptide core and a terminal epoxyketone moiety. Over the past few decades, a variety of these natural products, including eponemycin (**13**), epoxomicin (**14**), and other related compounds, have been isolated from a wide range of bacterial sources (Figure 3-1).[96-99] More recently, several additional epoxyketone proteasome inhibitors, specifically the clarepoxins, landepoxins, and macyranone A, were discovered through metagenome mining[100] and NMR screening approaches.[101] Although these naturally-occurring peptidyl epoxyketones exhibit great structural diversity in the length and composition of their peptide core and can also differ in the substitution on the epoxide ring, all of them possess a common terminal epoxyketone pharmacophore, which is responsible for their high efficacy and specificity as proteasome inhibitors. In particular, the strongly electrophilic carbonyl and epoxide groups are highly susceptible to a double nucleophilic attack by the *N*-terminal threonine of the 20S proteasome, leading to the irreversible formation of a morpholino ring.[102-104] Consequently, these natural products have inspired the current development of anticancer, anti-inflammatory, and antiparasitic drugs with a terminal epoxyketone warhead.[105-107] One such example is the synthetic epoxomicin analogue, carfilzomib (Figure 3-1), which has been used as a treatment for relapsed and refractory multiple myeloma.[108, 109]

Given the potency and unique structure of these peptidyl epoxyketone natural products, there has been great interest in the biosynthetic origins of these compounds. In particular, the genetic basis for the biosynthesis of epoxomicin and eponemycin was recently identified from *Goodfellowiella coeruleoviolacea* and *Streptomyces hygroscopicus*, respectively (Figure 3-2).[110] Moreover, gene clusters responsible for the biosynthesis of clarepoxins and landepoxins were recently identified in soil metagenomic libraries,[100] and the gene cluster responsible for the biosynthesis of macyranones was identified in the genome of *Cystobacter fuscus* MCy9118.[101] However, the specific enzymes and chemical logic governing the formation of the terminal epoxyketone pharmacophore have not yet been fully elucidated. Nonetheless, bioinformatics analysis has shown four genes to be conserved across the epoxomicin and eponemycin gene clusters, and these genes encode a non-ribosomal peptide synthetase (NRPS): EpnG/EpxD; a polyketide synthase (PKS): EpnH/EpxE; an acyl-CoA dehydrogenase (ACAD) homolog: EpnF/EpxF; and a cytochrome P450 monooxygenase: EpnI/EpxC. Based on this analysis, a biosynthetic pathway in which a hybrid NRPS-PKS assembly line generates a carboxylic acid that is subsequently modified by the ACAD and cytochrome P450 to form the terminal epoxyketone moiety was previously proposed.[110] Notably, however, the recently identified gene cluster for macyranone biosynthesis lacks a gene encoding a cytochrome P450, which led to an alternative acyl-carrier-bound

proposed mechanism for epoxyketone formation involving the ACAD homolog, MynC, and possibly the type II thioesterase, MynH.[101]

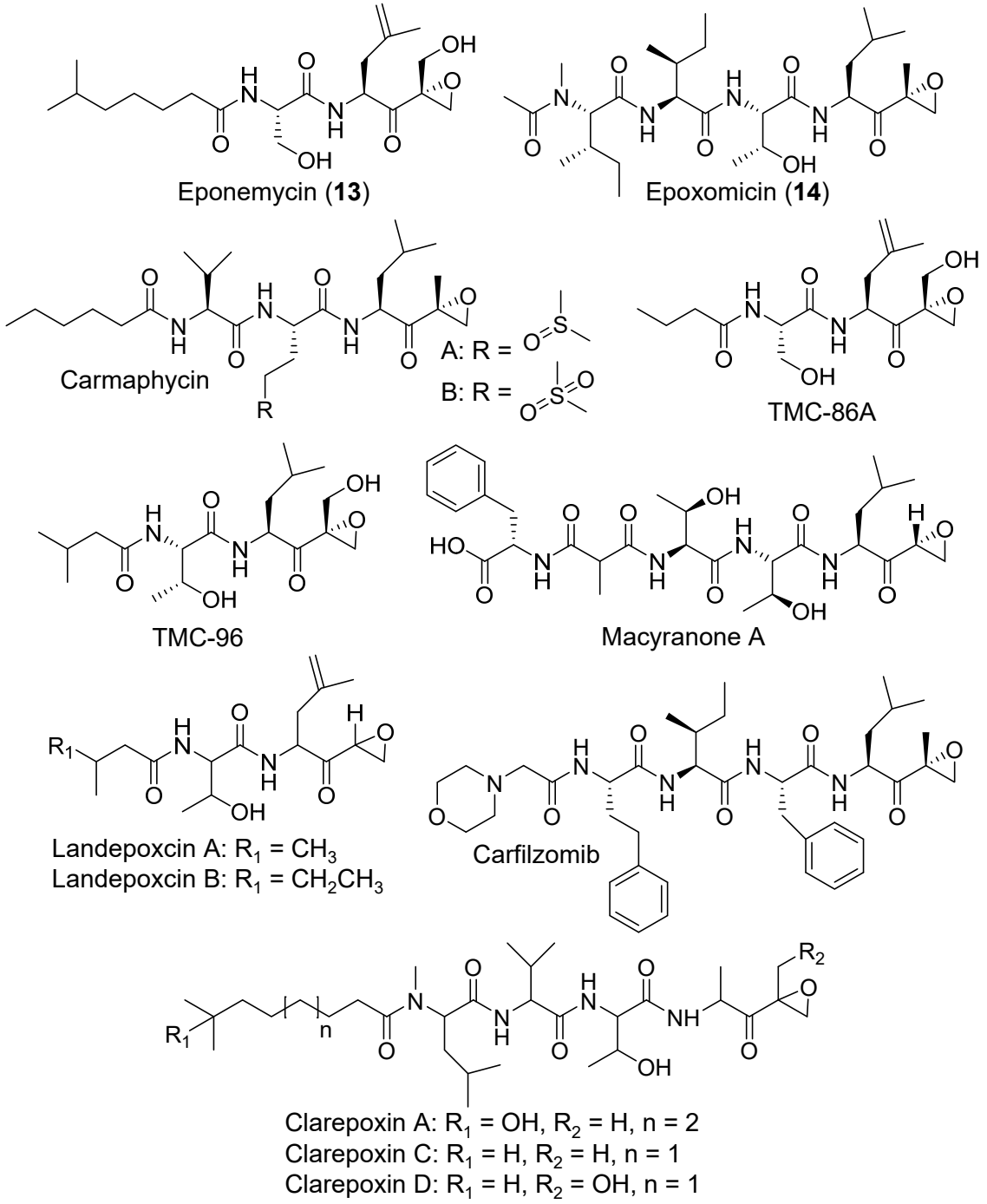


Figure 3-1. Structures of peptidyl epoxyketone proteasome inhibitors.

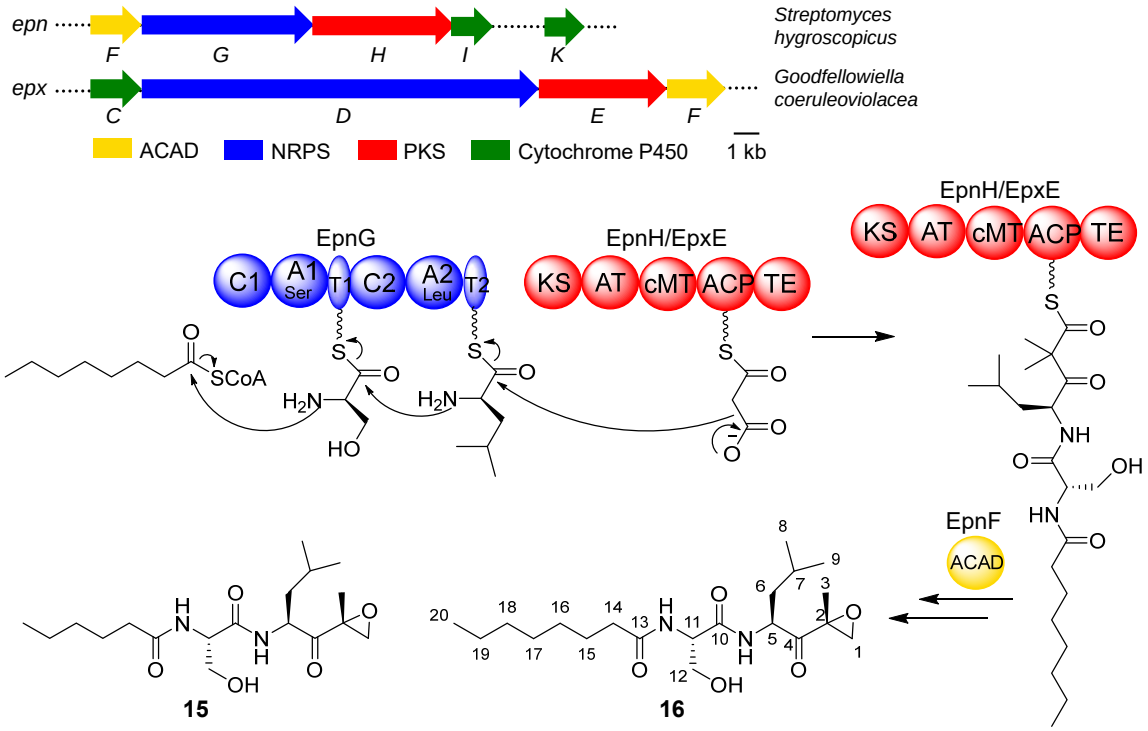


Figure 3-2. Schematic of eponemycin (*epn*) and epoxomicin (*epx*) gene clusters and the proposed biosynthetic pathway in *E. coli*.

In this work, we build upon the previous study of the biosynthesis of eponemycin and epoxomicin by turning to reconstitution in *E. coli* as a means to quickly and systematically identify the enzymes necessary for the generation of the terminal epoxyketone moiety. Due to its genetic tractability, fast growth rate, and well-characterized metabolism, *E. coli* is a particularly attractive host for studying the biosynthesis of natural products.[29, 34, 111] Furthermore, the reconstitution of NRP-PKS activity in *E. coli* provides a superior platform for the overproduction and diversification of their products.[112, 113] The production of peptidyl epoxyketones in *E. coli* described here thus not only provides us with additional insight into the mechanism of terminal epoxyketone biosynthesis but also advances the engineered biosynthesis of novel epoxyketone proteasome inhibitors in *E. coli*.

3.2 Results

3.2.1 Biosynthesis of peptidyl epoxyketones in *E. coli*

To determine the enzymes required for the biosynthesis of the terminal epoxyketone pharmacophore, we co-expressed the four genes from the eponemycin gene cluster (*epnFGHI*) proposed to be involved in *E. coli* on three plasmids. The plasmids were co-transformed into *E. coli* BAP1, which contains a chromosomal copy of the phosphopantetheinyl transferase gene, *sfp*,[43] to generate the strain JL6. A negative control strain transformed with empty vectors was also constructed. Although the natural fatty acyl moiety in eponemycin is derived from 6-methyl heptanoic acid, which is uncommon in bacterial primary metabolism[114] and is presumably generated by dedicated enzymes encoded in the eponemycin biosynthetic gene cluster,[110] we reasoned that EpnG-activated

serine could possibly be condensed with hexanoyl- or octanoyl-CoA produced endogenously by *E. coli* (Figure 3-2). Consequently, cultures of JL6 were supplemented with 1 mm hexanoic acid or octanoic acid at the time of induction to promote the incorporation of these alternative fatty acyl groups into the assembly line. The culture extracts of JL6 and the control strain were analyzed by LC-HRMS-based comparative metabolomics,[115] and two new compounds (**15** and **16**) with masses consistent with molecular formula of $C_{18}H_{32}N_2O_5$ and $C_{20}H_{36}N_2O_5$, respectively, were detected in high abundance in the extracts of JL6 but were absent from the extracts of the control strain (Figure 3-3). Further HRMS/MS analysis confirmed **15** and **16** to be consistent with a C_6 or C_8 fatty acyl group condensed with serine, leucine, and malonate, but it did not give insight into the structure of the terminal moiety resulting from modifications on the malonate (Appendix B).

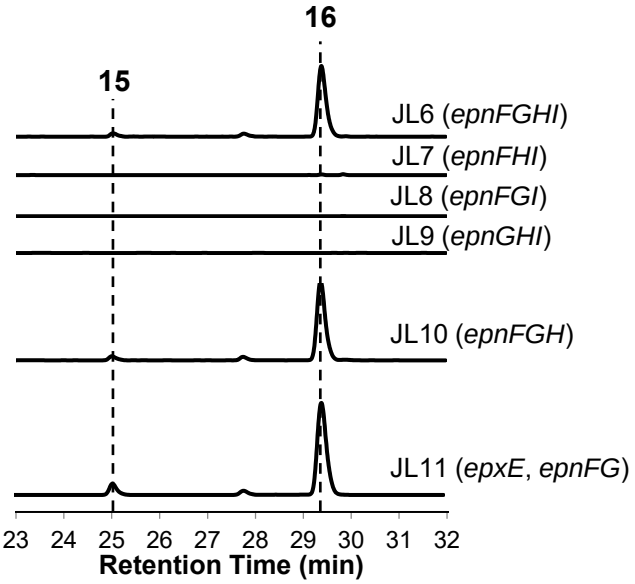
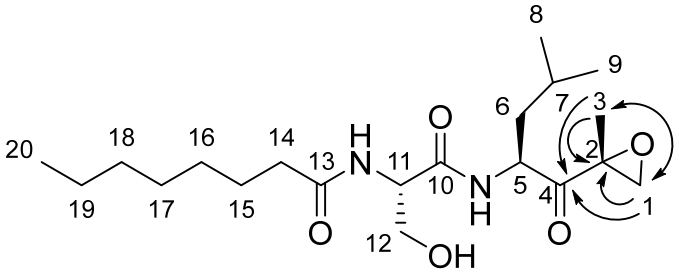


Figure 3-3. Biosynthesis of peptidyl epoxyketones in *E. coli*. Extracted ion chromatograms (**15**: *m/z* 357.2384 $[M+H]^+$, **16**: *m/z* 385.2697 $[M+H]^+$) showing the presence and absence of **15** and **16**, resulting from the heterologous expression of different combinations of EpnF, EpnG, EpnH/EpxE, and EpnI in *E. coli*. The calculated mass with a 10 ppm mass error tolerance was used.

To reveal the exact molecular structure of the new compounds, we then scaled up the production cultures and purified ~2 mg of the major product, **16**, for NMR spectroscopic analysis. Compound **16** was purified by following the fractions using LC-MS to give a yield of ~0.2 mg L^{-1}. Analysis of the 1 D (^1H, ^{13}C) and 2 D (HSQC, COSY, HMBC) NMR spectra of purified **16** confirmed the presence of an acylated peptide composed of an eight-carbon fatty acyl chain, serine, and leucine connected to a methyl-substituted epoxyketone (Figure 3-2 and Table 3-1). Notably, the proton and carbon shifts of the methyl-substituted epoxyketone and leucine residue are in strong agreement with those reported for a standard of **2**, which has these same functionalities.[97] Moreover, the methyl-substituted quaternary epoxide ring was supported by HMBC correlations from the C3 methyl protons ($\delta_H = 1.52$) to both the quaternary carbon C2 ($\delta_C = 59.5$) and the oxygenated methylene carbon C1 ($\delta_C = 52.7$) and by HMBC correlations from the C1

protons (δ_H = 2.91 and 3.28) to C2 and C3 (δ_C = 16.9; Table 3-1). HMBC correlations from the C1 and C3 protons to the ketone carbon C4 (δ_C=208.8) also confirmed the connectivity of the epoxide (Table 3-1). Thus, based on our HRMS and NMR analyses, we determined **16** to be a new terminal epoxyketone compound.

Table 3-1. NMR data for **16** in $CDCl_3$.



Position	δ_H (*J* in Hz)	δ_c	COSY	HMBC
1	2.91, d (4.5) 3.28, d (4.5)	52.7	1	2, 3, 4
2		59.5		
3	1.52, s	16.9		1, 2, 4
4		208.8		
5	4.50, m	51.3	5-NH, 6	4, 6, 7, 10
5-NH	6.87, d (5.4)		5	5, 6, 10
6	1.29, m 1.57, m	39.4	5, 6, 7	5, 7, 8, 9
7	1.64, m	25.4	6, 8, 9	5, 6, 9
8	0.93, d (7.2)*	21.1*	7	6, 7, 9
9	0.94, d (7.2)*	23.5*	7	6, 7, 8
10		171.6		
11	4.45, m	53.4	11-NH, 12	10, 12, 13
11-NH	6.46, d (4.5)		11	10, 11, 12, 13
12	3.54, m 4.05, m	62.8	11, 12	10, 11
13		174.0		
14	2.21, t (6.3)	36.6	15	13, 15, 16
15	1.60, m	25.8	14, 16	13, 14, 16, 17
16	1.28, m	29.1	15, 17	17, 18
17	1.25, m	29.3	16, 18	16, 19
18	1.24, m	31.8	17, 19	16, 17, 19, 20
19	1.27, m	22.7	18, 20	17, 18, 20
20	0.87, t (7.2)	14.2	19	18, 19

*may be interchanged

3.2.2 Identification of the enzymes necessary for epoxyketone biosynthesis

To probe the necessity of individual enzymes in the formation of **15** and **16**, we systematically removed each of the four genes from JL6 and compared the product profiles of the resulting strains. Analysis of the culture extracts of these strains by LC-HRMS showed that the production of **15** and **16** was abolished in JL7, JL8, and JL9, which lack *epnG*, *epnH*, and *epnF*, respectively, but was maintained in JL10, which lacks *epnI* (Figure 3-3). Moreover, LC-HRMS-based comparative metabolomics did not show any differences between the extracts of JL6 and JL10, and the titers of **15** and **16** were comparable in both strains, confirming that EpnI is not necessary for the formation of **16** in *E. coli*, though its function remains unclear. These results demonstrate that the NRPS, PKS, and ACAD are sufficient for the biosynthesis of peptidyl epoxyketones in *E. coli*, and epoxide formation does not necessarily require a cytochrome P450 monooxygenase. In addition, the titers of **15** and **16** from JL11 in which *epnH* is replaced by *epxE* appeared to be slightly greater than those from JL10, demonstrating the interchangeability of EpnH and its homolog, EpxE (similarity/identity: 55 %/44 %), which is encoded in the epoxomicin biosynthetic gene cluster. Interestingly, no apparent carboxylic acid intermediates could be detected by LC-HRMS analysis of the culture extracts of a strain expressing only the NRPS and PKS together, suggesting that the ACAD might act on the assembly line.

3.2.3 Biochemical studies of eponemycin biosynthetic enzymes

To complement the *in vivo* studies of EpnF, G, H, and I, we also attempted to perform *in vitro* characterization of these enzymes, beginning with their purification. EpnG, the dimodule NRPS that has a domain organization of C1-A1-T1-C2-A2-T2 (C: condensation, A: adenylation, T: thiolation), was purified from *E. coli* BAP1 though SDS-PAGE analysis indicated some possible protein degradation (Figures 3-2 and 3-4). The A domain substrate specificities of EpnG were then tested by using the well-established ATP-[^{32}P]PP$_i$ exchange assay.[116] As expected, EpnG demonstrated a strong preference for L-serine and L-leucine, though it was also able to activate L-cysteine, L-alanine, and L-threonine to some extent (Figure 3-5). We next attempted to purify EpnH, a single module PKS with a domain organization of KS-AT-cMT-ACP-TE (KS: ketosynthase; AT: acyltransferase, cMT: *C*-methyltransferase; ACP: acyl carrier protein, TE: thioesterase), from *E. coli* BAP1. However, despite numerous tries using different tags as well as different *Streptomyces* hosts, we could not obtain a sufficient amount of soluble and active EpnH (Figure 3-4). We likewise tested the EpnH homolog, EpxE, but we were also unable to obtain a sufficient amount of soluble and active EpxE (Figure 3-4). We suspect that this PKS enzyme may have a short half-life in *E. coli* and degrades quickly upon expression, since attempts to reconstitute the production of **15** and **16** using the lysate of JL10 or JL11 and to radioactively label the PKS using a sensitive [2-^{14}C]malonyl-CoA or *S*-[methyl-^{14}C]adenosylmethionine assay were both unsuccessful. Nevertheless, the overexpression and purification of EpnF and EpnI from *E. coli* BL21 Gold (DE3) both yielded soluble proteins that visually appeared to be bound to the correct cofactors, and purified EpnF was confirmed to contain FAD (Figure 3-6). However, without the ability to reconstitute the PKS activity *in vitro*, we were unable to further investigate the biochemical functions of these modification enzymes. Future efforts to obtain PKS enzymes that are functional *in vitro* may include the screening of

additional PKS homologs such as those encoded by the clarepoxin and landepoxin biosynthetic gene clusters or the utilization of dissected EpnH/EpxE domains to overcome the problem of PKS megasynthase instability.

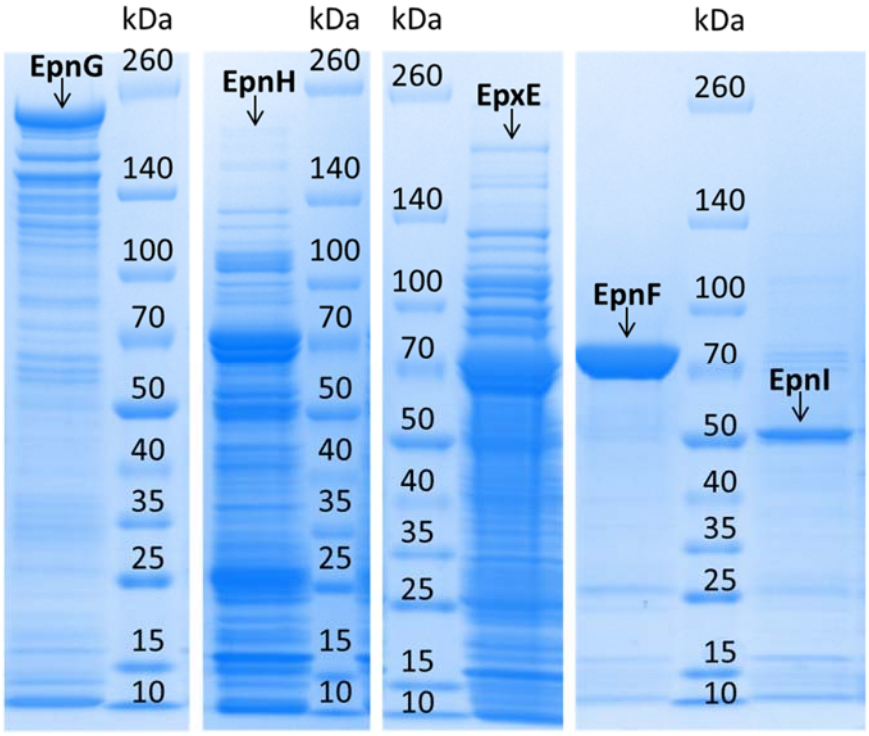


Figure 3-4. Eponemycin and epoxomicin biosynthetic enzymes purified from *E. coli*. Only a very small amount of the PKS megasynthases, EpnH and EpxE, could be purified from *E. coli* BAP1 as shown here.

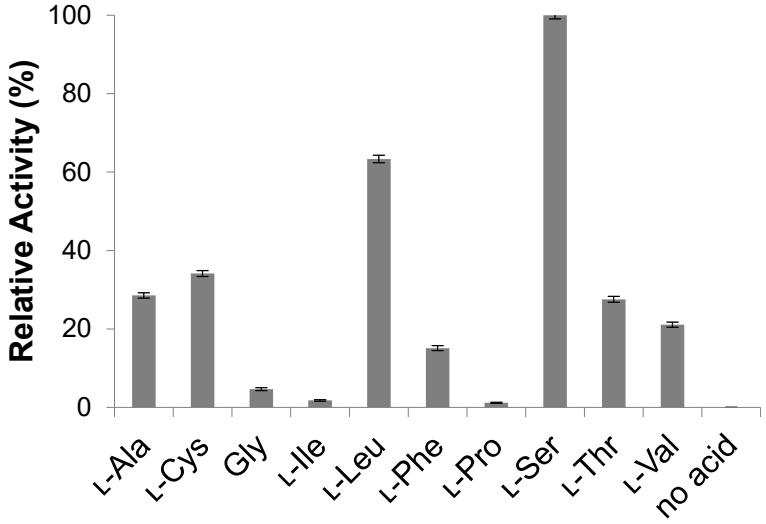


Figure 3-5. Characterization of the A domain activities of EpnG. Error bars represent standard deviations from three replicates. A relative activity of 100 % corresponds to 39 k cpm.

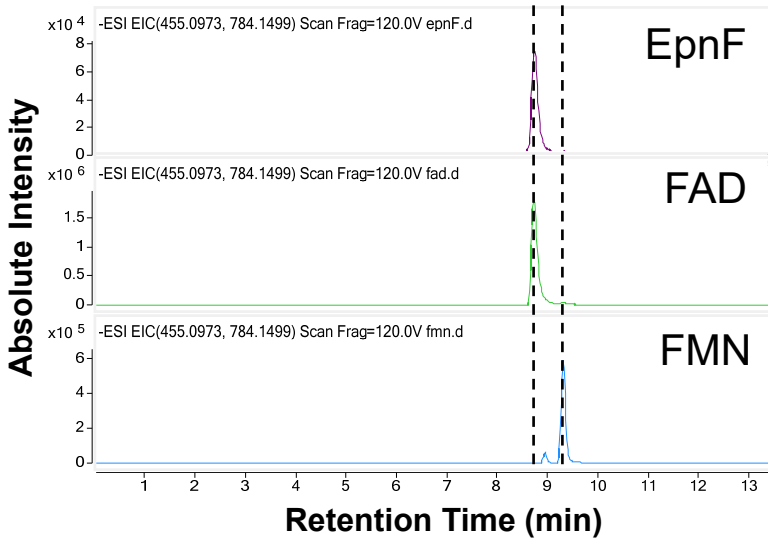


Figure 3-6. Cofactor analysis of EpnF. Extracted ion chromatograms (FAD: *m/z* 784.1499 [*M*-H]$^-$, FMN: *m/z* 455.0973 [*M*-H]$^-$) showing that the cofactor released from EpnF is FAD. The calculated mass with a 10 ppm mass error tolerance was used.

3.2.4 Isotope-labeled precursor incorporation into peptidyl epoxyketones produced by *E. coli*

To better elucidate the biosynthetic mechanism for terminal epoxyketone formation, we took advantage of our *E. coli* heterologous expression system and used it to perform ^{13}C-labeled precursor feeding studies. Two possible mechanisms (pathways A and B) for terminal epoxyketone formation have been previously proposed (Figure 3-7).[110] Pathway A postulates the introduction of a single methyl group at C2, followed by the reduction of the carboxylic acid to an olefin that can then undergo epoxidation, whereas pathway B proposes dimethylation at C2, followed by decarboxylation of the carboxylic acid and subsequent epoxide installation.[110] To distinguish between these two possibilities, we fed [1,2-^{13}C]sodium acetate to cultures of JL11, the highest producing strain, and compared the resulting ^{13}C incorporation to that from unlabeled and [1-^{13}C]sodium acetate feeding (Figures 3-7, 3-8, and 3-9). Although the titers of labeled **15** and **16** were insufficient for NMR analysis, LC-HRMS analysis of **15** and **16** from the extracts of cultures fed with [1,2-^{13}C]sodium acetate showed a significant increase in the ratio of the *M*+1 peak intensity to the *M*+ peak intensity compared to that of cultures fed with unlabeled and [1-^{13}C]sodium acetate (Figures 3-8 and 3-9). Further HRMS/MS analyses and examination of the isotope distribution of fragments with and without the epoxyketone group indicated that this increase is the result of a single ^{13}C atom from doubly labeled acetate being incorporated as the C2 carbon of **15** and **16** after decarboxylation (Figures 3-7, 3-8, and 3-9).

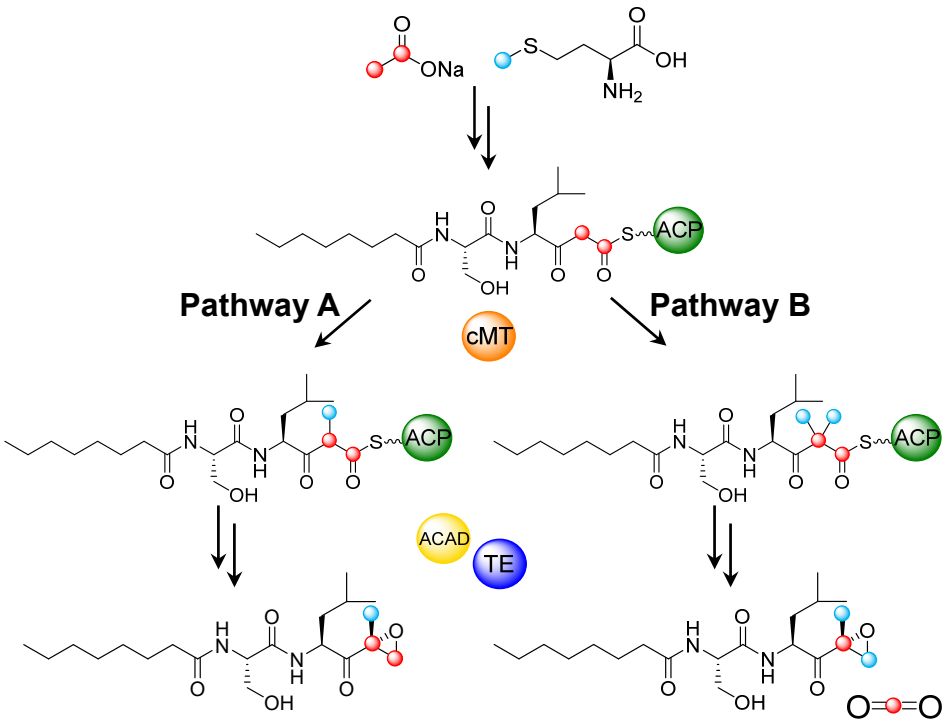


Figure 3-7. Two proposed pathways for terminal epoxyketone formation. Our ^{13}C-labeled precursor feeding results support pathway B.

To evaluate the occurrence of dimethylation, we then separately fed [methyl-^{13}C] L-methionine to cultures of JL11, as methionine is the precursor of the predicted methyl group donor, *S*-adenosylmethionine (SAM). This feeding experiment resulted in a significant increase in the ratio of the *M*+2 peak to the *M*+ peak for both **15** and **16**, which is indicative of the incorporation of two ^{13}C atoms, and additional HRMS/MS analyses showed that these ^{13}C atoms were incorporated into the epoxyketone-containing fragment of **15** and **16** (Figures 3-7, 3-8, and 3-9). These feeding results thus support pathway B, in which C2 is dimethylated by the cMT domain of the PKS and, either following or in concert with the activity of the ACAD, decarboxylation of a terminal carboxylic acid takes place.

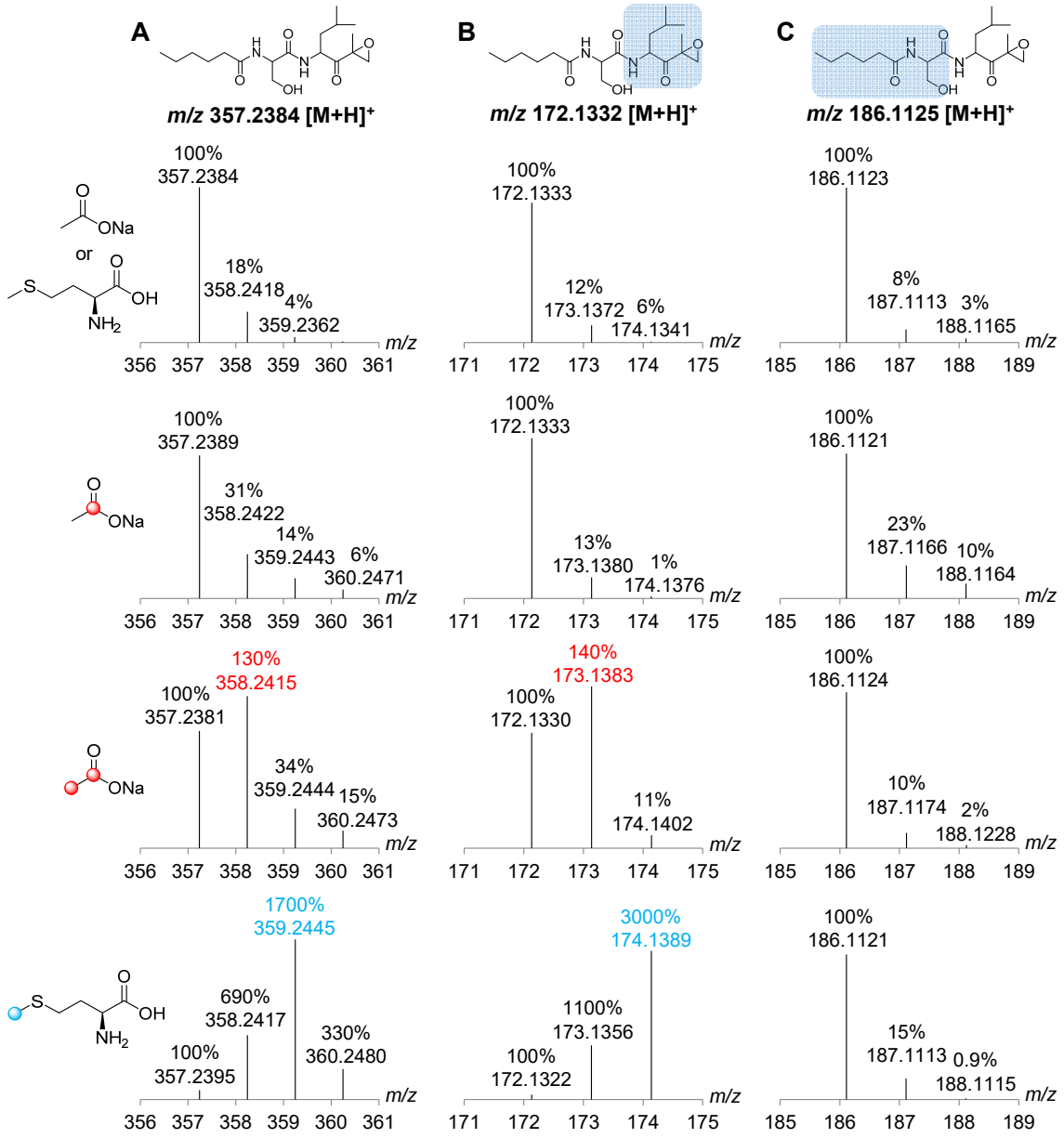


Figure 3-8. Isotopic peak analysis of **15** from ^{13}C-labeled precursor feeding studies. Isotopic peak patterns of **15** (A), a fragment of **15** with the terminal epoxyketone group (B), and a fragment of **15** without the epoxyketone group (C) from cultures fed with unlabeled sodium acetate or unlabeled L-methionine (row 1), [1-^{13}C] sodium acetate (row 2), [1,2-^{13}C] sodium acetate (row 3), and [methyl-^{13}C] L-methionine (row 4) are shown.

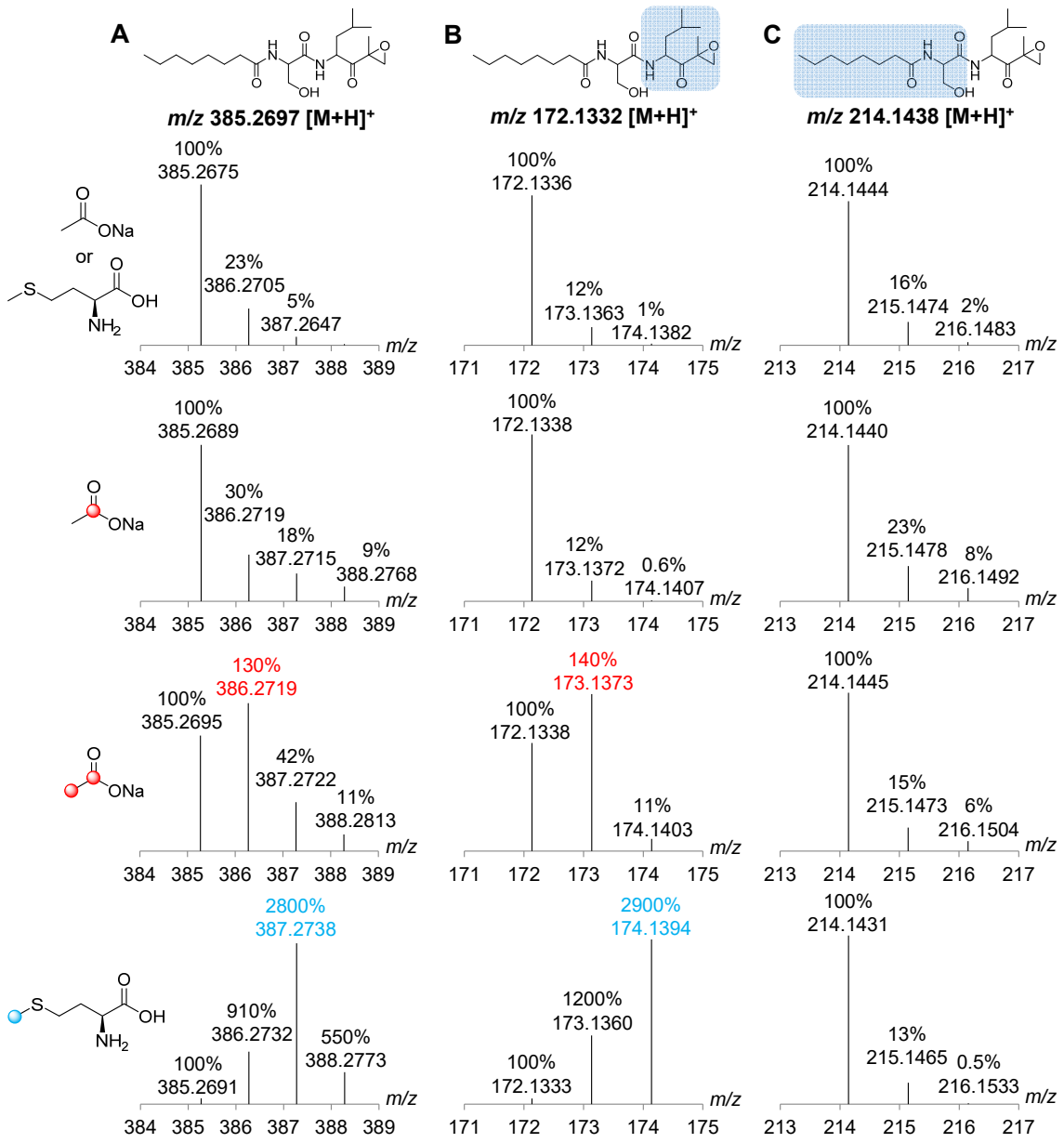


Figure 3-9. Isotopic peak analysis of **16** from ^{13}C-labeled precursor feeding studies. Isotopic peak patterns of **16** (A), a fragment of **16** with the terminal epoxyketone group (B), and a fragment of **16** without the epoxyketone group (C) from cultures fed with unlabeled sodium acetate or unlabeled L-methionine (row 1), [1-^{13}C] sodium acetate (row 2), [1,2-^{13}C] sodium acetate (row 3), and [methyl-^{13}C] L-methionine (row 4) are shown.

3.3 Discussion

Epoxyketone proteasome inhibitors have emerged as promising compounds for the treatment of cancer, and the unique structure of their terminal epoxyketone pharmacophore has raised intriguing questions about the biosynthesis of peptidyl epoxyketone natural products. To advance the understanding of the biosynthesis of these compounds, we took advantage of the convenience and flexibility offered by *E. coli* as a

heterologous host and reconstituted the activities of enzymes from the eponemycin biosynthetic gene cluster that were necessary and sufficient for the formation of two novel peptidyl epoxyketone products (Figure 3-2). Through the systematic expression of different combinations of eponemycin biosynthetic genes in *E. coli*, we established that this minimal set of enzymes includes an NRPS, a PKS, and an ACAD homolog. Interestingly, the cytochrome P450 that is conserved between the epoxomicin and eponemycin gene clusters (EpnI/EpxC; similarity/identity: 62%/47%) appeared to be unnecessary for terminal epoxyketone formation, and this result is consistent with the recent study on macyranone biosynthesis in which the reported gene cluster did not contain any genes encoding for a cytochrome P450 nor could a close homolog of EpnI/EpxC be identified elsewhere on the *Cystobacter fuscus* MCy9118 genome.[101] However, as we did not identify any new metabolites unique to the culture extracts of JL6, we were unable to confirm the precise function of EpnI/EpxC.

In addition to identifying the enzymes required for terminal epoxyketone biosynthesis, our ^{13}C-labeled sodium acetate and methionine feeding studies provided evidence that formation of the epoxyketone pharmacophore involves a dimethylation and subsequent decarboxylation reaction. Shortly after the publication of our work,[117] two other groups also published similar findings about epoxyketone biosynthesis.[118, 119] In particular, Zettler et al. fed [methyl-^{13}C] L-methionine to the native producers of eponemycin and epoxomicin and likewise observed the incorporation of two ^{13}C atoms into the epoxyketone warhead by MS and ^{13}C NMR.[118] In accordance with our conclusions that only an ACAD and not a cytochrome P450 is needed for epoxyketone formation, they also found that the deletion of the *epnI* homolog, *epxC*, in a *Streptomyces* strain heterologously expressing the epoxomicin gene cluster did not affect the production of **14**.[118]

More recently, Zabala et al. also fed [methyl-^{13}C] L-methionine to the native producer of TMC-86A (Figure 3-1) and arrived at the conclusion that the cMT domain catalyzes dimethylation.[119] Importantly, this group was also able to provide the first biochemical evidence for the function of EpnF by chemically synthesizing an isopropyl ketone and an unstable α-dimethyl-β-keto acid to be tested as substrates.[119] Although much of the synthesized acid was converted to the ketone through spontaneous decarboxylation, EpnF was shown to catalyze epoxyketone formation from the acid but not from the ketone. Accordingly, we were able to observe some of the putative isopropyl ketone compound but not any of the carboxylic acid intermediate in the culture extracts of *E. coli* expressing only EpnH/EpxE and EpnG. Unlike typical ACADs, which catalyze the FAD-dependent oxidation of fatty acid acyl-CoA substrates,[120, 121] the function of EpnF in epoxyketone biosynthesis appears to be quite unusual in its catalysis of a FAD-dependent decarboxylation-dehydrogenation-monoxygenation cascade. A proposed mechanism for this enzyme is shown in Figure 3-10, and it involves decarboxylation of the acid and hydride transfer to FAD, resulting in an α,β-unsaturated ketone that can be converted to an epoxyketone by a flavin hydroperoxide species.[119] Zabala et al. further showed *in vitro* that the EpnI homolog, TmcI, is responsible for hydroxylating the methyl group on the epoxyketone moiety of TMC-86A.[119] As we did not observe a hydroxylated version of **15** or **16** in the culture extracts of JL6, we reason that EpnI may be inactive in *E. coli* despite its apparent solubility (Figure 3-4). However, the function of the homologous cytochrome P450, EpxC, still remains somewhat unclear since the

epoxyketone group of **14** lacks hydroxylation. Based on further comparison of the gene clusters encoding peptidyl epoxyketone compounds, it was also proposed that the second cytochrome P450 encoded in the eponemycin biosynthetic gene cluster (EpnK; Figure 3-2) may be involved in the formation of the dehydroleucine moiety found in **13** though this has yet to be experimentally confirmed.[119]

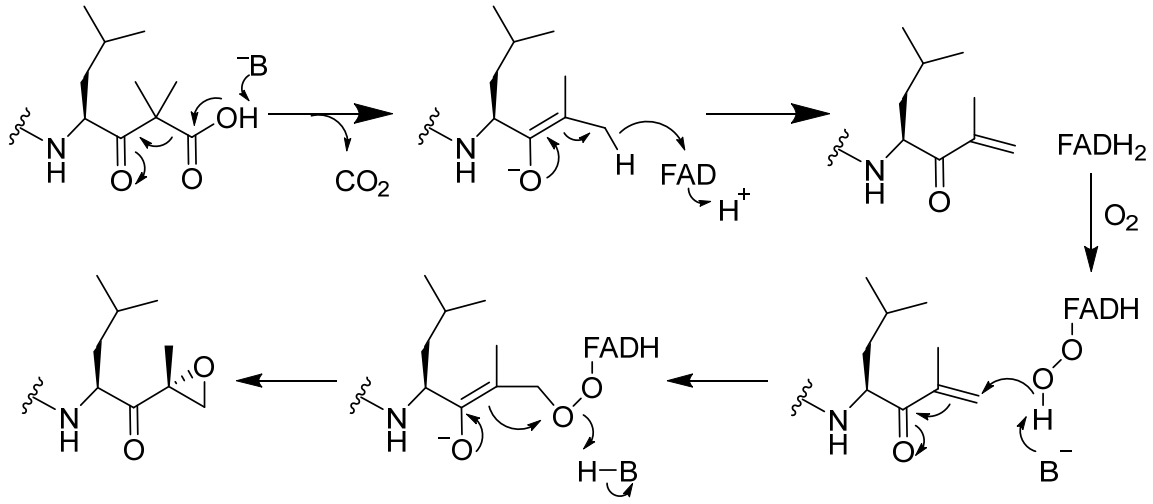


Figure 3-10. Proposed mechanism for the reaction catalyzed by EpnF.

Notably, our expression of the peptidyl epoxyketone biosynthetic machinery in *E. coli* also demonstrates the potential for using engineered biosynthesis to generate additional new epoxyketone proteasome inhibitors. The production of two novel compounds with C6 and C8 acyl groups as an alternative to the natural 6-methyl heptanoic acyl moiety and the ability of EpxE to complement EpnH show some degree of flexibility in the biosynthetic machinery. Furthermore, ATP-[^{32}P]PPi exchange assays showed that the A domains of EpnG were able to adenylate a few other amino acids besides the expected L-serine and L-leucine (Figure 3-5). These findings coupled with the recent identification of many related biosynthetic gene clusters provide the unique opportunity to expediently perform combinatorial biosynthesis in *E. coli* to produce novel peptidyl epoxyketone analogues.

3.4 Materials and Methods

Bacterial strains and growth conditions

Genomic DNA from *S. hygroscopicus* subsp. *hygroscopicus* ATCC 53709 was used to PCR amplify *epnF*, *epnG*, *epnH*, and *epnI*, and genomic DNA from *G. coeruleoviolacea* ATCC 53904 was used to PCR amplify *epxE*. *E. coli* XL1-Blue was used for all cloning procedures and *E. coli* BAP1 and *E. coli* BL21 Gold (DE3) were used for protein expression. LB was used for the propagation of *E. coli* at 37ºC unless specified otherwise. Growth media was supplemented with antibiotics as required at the following concentrations: kanamycin (50 μg/mL), chloramphenicol (25 μg/mL), and spectinomycin (50 μg/mL).

Construction of plasmids for protein expression in E. coli

Plasmids and primers used in this study are listed in Table 3-2 and 3-3, respectively. Individual genes were PCR amplified from genomic DNA and cloned into pET24b, pACYCDuet-1, or pCDFDuet-1 by restriction enzyme digestion (Thermo Scientific) and ligation with Quick T4 DNA ligase (New England Biolabs). To construct pCDFDuet-*epnG*-*epnI*, pCDFDuet-*epnG* was digested with NdeI/KpnI and *epnI* was cloned into the second MCS. For the overexpression and purification of EpnI, *epnI* was cloned into pET30 using the pET30 Xa/LIC Vector Kit (Novagen) to introduce an *N*-terminal His_6-tag. Plasmids were isolated using a QIAprep Spin Miniprep Kit (Qiagen) and confirmed by DNA Sequencing (UC Berkeley DNA Sequencing Facility).

Table 3-2. Plasmids used in this study.

Plasmid	**Derived from**	**Function**
pACYCDuet-*epnF*	pACYCDuet-1	Expression of EpnF in *E. coli*
pET24b-*epnH*	pET24b	Expression of EpnH in *E. coli*
pCDFDuet-*epnG*	pCDFDuet-1	Expression of EpnG in *E. coli*
pCDFDuet-*epnG*-*epnI*	pCDFDuet-1	Co-expression of EpnG and EpnI in *E. coli*
pCDFDuet-*epnI*	pCDFDuet-1	Expression of EpnI in *E. coli*
pET24b-*epxE*	pET24b	Expression of EpxE in *E. coli*
pET30-*epnI*	pET30-Xa/LIC	Expression of EpnI in *E. coli*

Table 3-3. Primers used in this study.

Primer	**Sequence (5' -> 3')**	**Description**
epnF-Duet-F	aaa**CTGCAG**gtgagtgacagcaaatcggt	Expression of EpnF
epnF-Duet-R	tat**AAGCTT**tcatcgcttccccgggtgag	
NdeI-epnH-F	aaa**CATATG**acgtcgaatcaacagat	Expression of EpnH
XhoI-epnH-R	aaa**CTCGAG**gtgattctctttctgcgata	
epnG-Duet-F	aaa**CTGCAG**atgaacaaggaaaaggaccg	Expression of EpnG
epnG-Duet-R	tat**AAGCTT**tcatgagttgcggttcctcc	
epnI-Duet-F	aaa**CATATG**gtgacgatcgacccgaa	Expression of EpnI
epnI-Duet-R	aaa**GGTACC**tcaggcatggacgttctctc	
epxE-Duet-F	aaa**CATATG**gttgacgagaccgccgt	Expression of EpxE
HindIII-epxE-R	tat**AAGCTT**ctgggtaaacctccgccttt	
pET30-epnI-F	GGTATTGAGGGTCGCatggtgacgatcgacccgaa	Expression of EpnI
pET30-epnI-R	AGAGGAGAGTTAGAGCCtcaggcatggacgttctctc	

*restriction endonuclease sites are bolded

Biosynthesis of peptidyl epoxyketones in E. coli

The plasmids, pACYCDuet-*epnF*, pCDFDuet-*epnG*-*epnI*, and pET24b-*epnH*, were electroporated into *E. coli* BAP1 to form strain JL6, and transformants were selected on LB agar plates supplemented with the appropriate antibiotics. Single colonies were inoculated into 3 mL of LB with antibiotics and grown overnight at 37ºC as a seed culture, of which 0.25 mL was used to inoculate 25 mL of fresh LB with antibiotics. The cultures were then grown at 37ºC to $OD_{600} \approx 0.4 – 0.6$ before induction with 0.5 mM IPTG and the addition of hexanoic or octanoic acid at a final concentration of 1 mM.

After induction, the temperature was dropped to 20ºC, and compound production was allowed to proceed for approximately 2 days. For the generation of strains JL7, 8, 9, 10, and 11, different combinations of the plasmids listed in Table 3-2 were electroporated into *E. coli* BAP1.

LC-HRMS analysis of peptidyl epoxyketones produced by E. coli

E. coli cultures were pelleted by centrifugation (4000 × *g* for 10 min), and the supernatant was extracted with two volumes of ethyl acetate. The solvent was removed by rotary evaporation, and the residue was redissolved in methanol (500 µL) for LC-HRMS and HRMS/MS analysis (10 µL injection). LC-HRMS analysis was performed on an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS instrument with an Agilent Eclipse Plus C18 column (4.6 × 100 mm). A linear gradient of 2-95% CH_3CN (vol/vol) over 40 min in H_2O with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min was used. For HRMS/MS experiments with **15** and **16**, a linear gradient of 40-80% CH_3CN (vol/vol) over 14 min in H_2O with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min and a collision energy of 10 V was used.

*Large-scale production, purification, and characterization of **16***

A total of 12-L (12 × 1 L) of strain JL11 (*E. coli* BAP1 co-transformed with pET24b-*epxE*, pACYCDuet-*epnF*, and pCDFDuet-*epnG*) was cultured in LB with antibiotics. The cultures were inoculated with 10 mL of a seed culture and grown at 37ºC to $OD_{600} \approx 0.4 – 0.6$ before induction with 0.5 mM IPTG and the addition of octanoic acid to a final concentration of 1 mM. After induction, the temperature was dropped to 20ºC, and compound production was allowed to proceed for approximately 2 days. Compound **16** was extracted from the cell-free supernatant using two volumes of ethyl acetate. The solvent was removed by rotary evaporation, and the combined residue was redissolved in dichloromethane (3 mL). The residue was chromatographed on a silica gel column and eluted with dichloromethane-methanol (95:5, vol/vol). Fractions containing **16** were determined by LC-MS (a linear gradient of 60-70% CH_3CN (vol/vol) over 10 min in H_2O with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min was used) and were combined for further purification by RP-HPLC (Agilent 1260 HPLC) with an Inertsil ODS-4 column (4.6 × 250 mm, GL Sciences Inc.) using an isocratic program of 50% CH_3CN (vol/vol) in H_2O at a flow rate of 0.5 mL/min. The resulting 2 mg of purified **16** was dried and analyzed by LC-HRMS and NMR. NMR spectra (1D: ^{1}H, ^{13}C and 2D: HSQC, COSY, HMBC) were recorded on a Bruker Biospin 900 MHz spectrometer with a cryoprobe in chloroform-d ($CDCl_3$; Cambridge Isotope Laboratories).

Overexpression and purification of proteins

The plasmids, pET24b-*epnH*, pET24b-*epxE*, and pCDFDuet-*epnG*, were transformed into *E. coli* BAP1, and the plasmids, pACYCDuet-*epnF* and pET30-*epnI*, were transformed into *E. coli* BL21 Gold (DE3) for protein expression. Expression and purification for all proteins with a His_6-tag followed the same general procedure and is detailed as follows: cells were grown at 37°C in 700 mL of LB with the appropriate antibiotic to an OD_{600} of 0.5. The cells were then cooled on ice for 10 min and induced with 0.12 mM IPTG for 16 h at 16°C. Subsequently, cells were harvested by centrifugation (6371 × *g*, 15 min, 4°C), resuspended in 30 mL lysis buffer (25 mM

HEPES, pH 8, 0.5 M NaCl, 5 mM imidazole), and lysed by homogenization on ice. Cellular debris was removed by centrifugation (27216 × *g*, 1 h, 4°C). Ni-NTA agarose resin was added to the supernatant (3 mL/L of culture), and the solution was nutated at 4°C for 1 h. The protein-resin mixture was loaded onto a gravity flow column, and proteins were eluted with increasing concentrations of imidazole in Buffer A (20 mM HEPES, pH 8.0, 1 mM DTT). Purified proteins were concentrated and buffer exchanged into Buffer A + 10% glycerol using Vivaspin centrifugal concentrators.

ATP-[^{32}P]PP$_i$ exchange assays

Assays were performed in 100 µL of reaction buffer (50 mM Tris-HCl, pH 8, 2 mM $MgCl_2$) containing 5 mM ATP, 1 mM Na_4[^{32}P]-pyrophosphate (PP$_i$) (~3 x 10^6 cpm/mL), 1 mM TCEP, 5 mM substrate, and 5 µM enzyme. Reactions were incubated at 25°C for 2 h, then quenched by the addition of a charcoal suspension (1.6% w/v activated charcoal, 0.1 M Na_4PP_i, 3.5% $HClO_4$). Free [^{32}P]PP$_i$ was removed by centrifugation of the sample followed by washing twice with wash solution (0.1 M Na_4PP_i and 3.5% $HClO_4$). Charcoal-bound radioactivity was measured on a Beckman LS 6500 scintillation counter.

LC-HRMS analysis of FAD-bound EpnF

Ten nanomoles of EpnF were denatured by the addition of methanol to release any bound cofactor. After centrifugation, the supernatant was analyzed with LC-HRMS using a linear gradient from 5-95% CH_3CN (vol/vol) over 20 min in H_2O with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min. The released flavin was detected in negative mode, and 200 pmol of authentic FAD and FMN were used as standards.

^{13}C-labeled precursor feeding experiments

Strain JL11 was cultured for the production of **15** and **16** as described above, and at the time of IPTG induction and the addition of octanoic acid, either unlabeled sodium acetate, [1-^{13}C] sodium acetate, [1,2-^{13}C] sodium acetate, unlabeled L-methionine, or [methyl-^{13}C] L-methionine was added to cultures to a final concentration of 1 g/L. Compound extraction and LC-HRMS and HRMS/MS analyses were performed as described above.

Chapter 4. Reconstituting the Biosynthesis of the 3-Formamidosalicylate Pharmacophore of Antimycins

Parts of this chapter have been adapted from the following with permission:

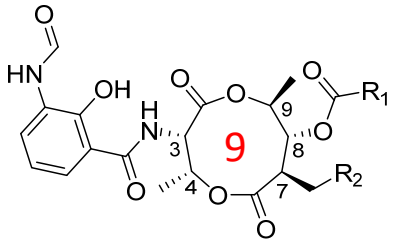
1. Liu, J., Zhu, X., Seipke, R. F., Zhang, W. "Biosynthesis of antimycins with a reconstituted 3-formamdosalicylate pharmacophore in *Escherichia coli*" *ACS Synth. Biol*. **4**, 559-565 (2014).
2. Liu, J., Zhu, X., Kim, S., Zhang, W. "Antimycin-type depsipeptides: discovery, biosynthesis, chemical synthesis, and bioactivties." *Nat. Prod. Rep.* DOI: 10.1039/c6np00004e (2016).

4.1 Introduction

Antimycins are a family of depsipeptides consisting of a nine-membered dilactone ring substituted with one alkyl (C-7), one acyloxy (C-8), two methyl moieties (C-4 and C-9), and an amide linkage (C-3) connecting to a 3-formamidosalicylic acid (Figure 4-1). They are produced by various *Streptomyces* species, and over the past few years, a growing number of natural (>40) and modified (>350) antimycin-type compounds varying in the alkyl and acyl chains have been reported.[52, 122-128] Additionally, several related antimycin-type depsipeptides with 12-, 15-, and 18-membered macrolactone rings were also recently isolated and characterized (Figure 4-1).[129-132] Notably, all of these compounds possess a common 3-formamidosalicylate unit, which previous studies have demonstrated to be essential for the antifungal, insecticidal, nematocidal, and piscicidal properties of antimycins.[123, 124, 126, 133, 134] While these bioactivities arise from the ability of antimycins to inhibit cytochrome c oxidoreductase in the mitochondrial electron transport chain,[135-137] other work has also shown antimycin-type compounds to be promising candidates for treating a variety of diseases, including cancer,[132, 138-140] asthma,[125] and neurodegenerative disorders.[130, 141-143] Understanding the enzymatic machinery and engineering efficient systems for antimycin biosynthesis can therefore expand the production of antimycin analogues with improved pharmaceutical properties.

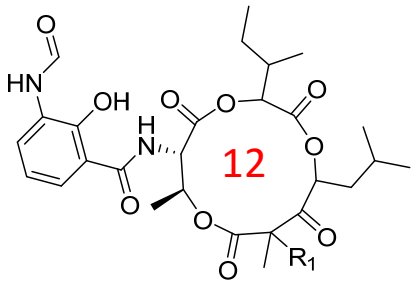
The antimycin biosynthetic gene cluster was initially identified in several *Streptomyces* species by Seipke et al.,[144, 145] and it was recently dissected by both genetic and enzymatic studies.[51] Specifically, Sandy et al. identified the minimum set of enzymes (AntCDEFGM) needed for the generation of the antimycin dilactone scaffold, which is built through a hybrid NRPS-PKS assembly line-based mechanism from four distinct monomers: an aminobenzoate, a natural amino acid, an α-keto acid, and an acylmalonyl moiety (Figure 4-2).[51] The biosynthesis of antimycins is proposed to begin with the catabolism of tryptophan to anthranilate, which proceeds through three enzymatic steps: (1) AntN, a tryptophan-2,3-dioxygenase, opens the indole ring of L-tryptophan to generate *N*-formyl-L-kynurenine; (2) AntO, a lipase homolog, deformylates *N*-formyl-L-kynurenine to L-kynurenine; and (3) AntP, which is present in some but not all antimycin gene clusters, or another kynureninase converts L-kynurenine to anthranilate (Figure 4-2).[146, 147] Anthranilate was demonstrated to be activated by the acyl-CoA ligase homolog, AntF, and loaded onto the acyl carrier protein (ACP), AntG, and the resulting anthraniloyl-*S*-AntG is proposed to be converted to 3-aminosalicyloyl-*S*-AntG by the multicomponent oxygenase, AntHIJKL, through epoxidation to a 6-amino-1,2-

epoxycyclohexa-3,5-dienecarboxylic acid moiety followed by a 1,2-shift of the thioester group (Figure 4-2).[51, 148] We hypothesize that AntO may subsequently catalyze *N*-formylation of the 3-aminosalicylate moiety to yield 3-formamidosalicyloyl-*S*-AntG, which is presented to the dimodule NRPS, AntC.

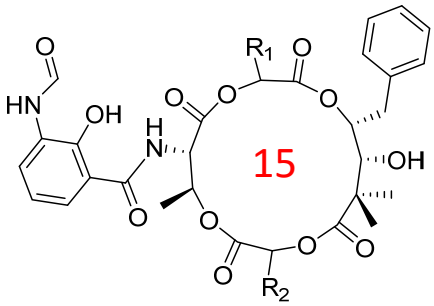


antimycin

A$_1$: $R_1 = CH(CH_3)CH_2CH_3$, $R_2 = (CH_2)_4CH_3$
A$_2$: $R_1 = CH(CH_3)_2$, $R_2 = (CH_2)_4CH_3$
A$_3$: $R_1 = CH(CH_3)CH_2CH_3$, $R_2 = (CH_2)_2CH_3$
A$_4$: $R_1 = CH(CH_3)_2$, $R_2 = (CH_2)_2CH_3$

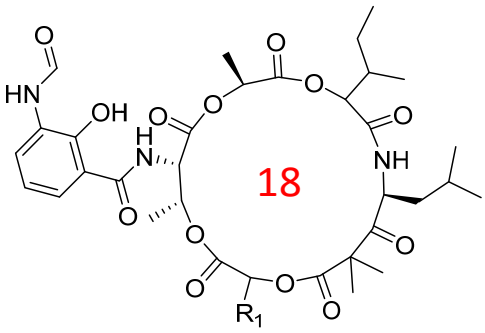


JBIR-06: $R_1 = CH_3$
JBIR-52: $R_1 = H$



neoantimycin

A: $R_1 = CH(CH_3)_2$, $R_2 = CH(CH_3)(CH_2CH_3)$
F: $R_1 = CH(CH_3)_2$, $R_2 = CH(CH_3)_2$
SW163-A: $R_1 = CH_3$, $R_2 = CH(CH_3)(CH_2CH_3)$
SW163-B: $R_1 = CH_3$, $R_2 = CH(CH_3)_2$



respirantin

$R_1 = CH_2CH(CH_3)_2$
$R_1 = CH(CH_3)_2$

Figure 4-1. Structures of antimycin-type depsipeptides. Only a few of the naturally-occuring antimycin and neoantimycin variants are shown here.

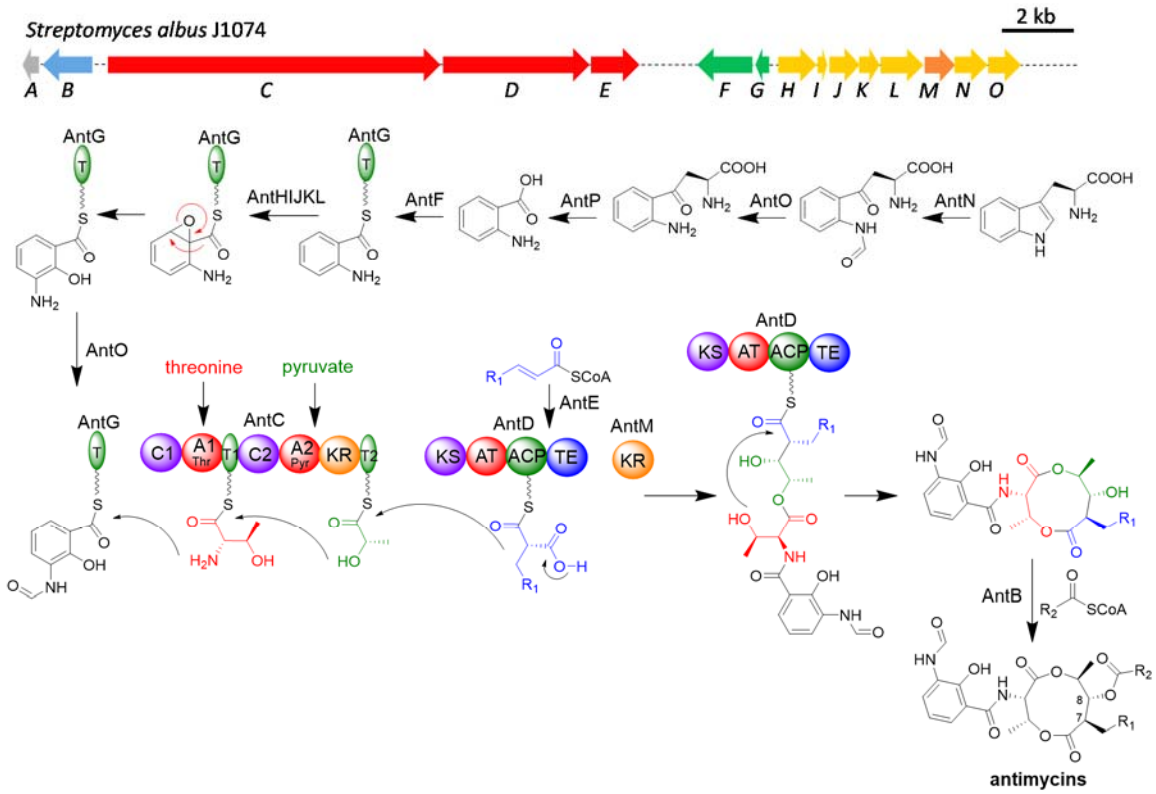


Figure 4-2. Biosynthetic gene cluster and proposed biosynthetic pathway of antimycins.

AntC has a domain organization of C1-A1-T1-C2-A2-KR-T2 (C: condensation, A: adenylation, T: thiolation, KR: ketoreduction). The A1 domain activates and loads threonine onto T1, which is condensed with 3-formamidosalicylate by C1, and the A2 domain activates and loads pyruvate onto T2, which is stereoselectively reduced by the KR domain and then condensed with threonine by C2. The downstream PKS, AntD, has a domain organization of KS-AT-ACP-TE (KS: ketosynthase, AT: acyltransferase, TE: thioesterase). The AT domain transfers a 2-carboxylated acyl moiety from CoA to the ACP, and it has been demonstrated to select atypical extender units such as butylmalonyl-CoA, which is generated from reductive carboxylation of 2*E*-hexenoyl-CoA by AntE, a crotonyl-CoA reductase (CCR) enzyme.[51, 52, 149, 150] The KS domain then catalyzes the decarboxylative condensation between the aminoacyl-*S*-T2 domain of AntC and the 2-carboxy-acyl-*S*-ACP domain of AntD, and following stereoselective reduction of the β-keto functionality by AntM, a 3-oxoacyl-ACP reductase, AntD-TE catalyzes regiospecific macrolactone cyclization and release of the nine-membered dilactone product. Finally, AntB, a promiscuous tailoring acyltransferase, catalyzes transesterification on the C8 hydroxyl group, and it has been shown to be capable of using a wide variety of acyl substrates.[52, 151]

While the enzymatic synthesis of the dilactone scaffold of antimycin has been achieved,[51] the biosynthetic pathway toward 3-formamidosalicylic acid, the putative starter unit conserved in all antimycin-type depsipeptides but not found in any other natural products (Figure 4-1), was not fully elucidated. In addition, the production of antimycins has thus far been limited to slow-growing *Streptomyces* organisms that can be challenging to genetically manipulate. As an alternative, *E. coli* has emerged as a

powerful microorganism for studying the biosynthesis of some natural products because of its fast growth rate, genetic tractability, and well-studied primary metabolism.[29, 34, 111] The reconstitution of NRPS-PKS activity in *E. coli* also provides immense opportunities for the biochemical understanding and protein engineering of these modular assembly lines in natural product synthesis and diversification.

In this work, we reconstituted the complete biosynthesis of antimycins in the heterologous host, *E. coli*. Using this heterologous expression system in combination with performing a mutagenesis study in *Streptomyces*, we systematically confirmed the roles of two previously proposed modification enzymes, AntHIJKL and AntO, in the generation of the 3-formamidosalicylate moiety of antimycins from an anthranilate precursor (Figure 4-2).[51, 128, 148] Down the line, our flexible and robust *E. coli* expression system can be used to facilitate the rapid generation of novel antimycin analogues through combinatorial biosynthesis as well as to increase the yield of target antimycin analogues through metabolic engineering.

4.2 Results

4.2.1 Production of the antimycin dilactone scaffold in *E. coli*

To produce the dilactone scaffold of antimycins in *E. coli*, we coexpressed six enzymes, AntCDEFGM, from either the *S. albus* J1074 or *S. ambofaciens* antimycin gene clusters[51, 144] on three Duet vectors. Specifically, the dimodule NRPS-encoding *antC* and the crotonyl-CoA carboxylase/reductase homolog-encoding *antE* were cloned into pETDuet-1; the PKS-encoding *antD* and an operon consisting of *antGF*, which encodes a carrier protein and an acyl-ACP ligase homolog, respectively, were cloned into pCDFDuet-1; and the 3-oxoacyl-ACP reductase homolog-encoding *antM* was cloned into pCOLADuet-1. The three plasmids were cotransformed into the BAP1 strain of *E. coli*, which contains a chromosomal copy of the phosphopantetheinyl transferase gene, *sfp*, to ensure the posttranslational modification of AntC, D, and G to their pantetheinylated forms.[43] Upon coexpression of *antCDEFGM* in this *E. coli* strain (JL1), trace amounts of antimycin-type compounds could be detected in the culture extracts by LC-HRMS and HRMS/MS analysis (Figures 4-3, 4-4, and Appendix C). Despite the relaxed substrate specificities of AntE and AntD,[51, 145] **17**, which is derived from the substrate octenoyl-CoA, was the major observed product, indicating that *E. coli* has a different endogenous pool of fatty acyl-CoA intermediates compared to *Streptomyces* by which a suite of antimycins varying at the C-7 position are typically produced.[52] As expected, the production of **17** was not observed in the extracts of cultures missing one or more of the six genes such as those of the *E. coli* strain JL5 (Figure 4-4).

When anthranilic acid and octanoic acid were both fed at a final concentration of 1 mM to JL1 cultures at the time of protein induction, the titer of **17** was increased by approximately 140-fold to ~0.2 mg/L. These precursors were thus included in all subsequent experiments. Since previous mutagenesis studies in *Streptomyces* showed that a starter unit other than 3-formamidosalicylate decreased the efficiency of NRPS-PKS assembly line by more than 1000-fold (deoxy-isoantimycins were produced at ~10 μg/L),[51] the current production titer of **17** by JL1 suggests that this heterologous expression system is more robust than the native *Streptomyces* system in generating the active modular assembly line for antimycin synthesis.

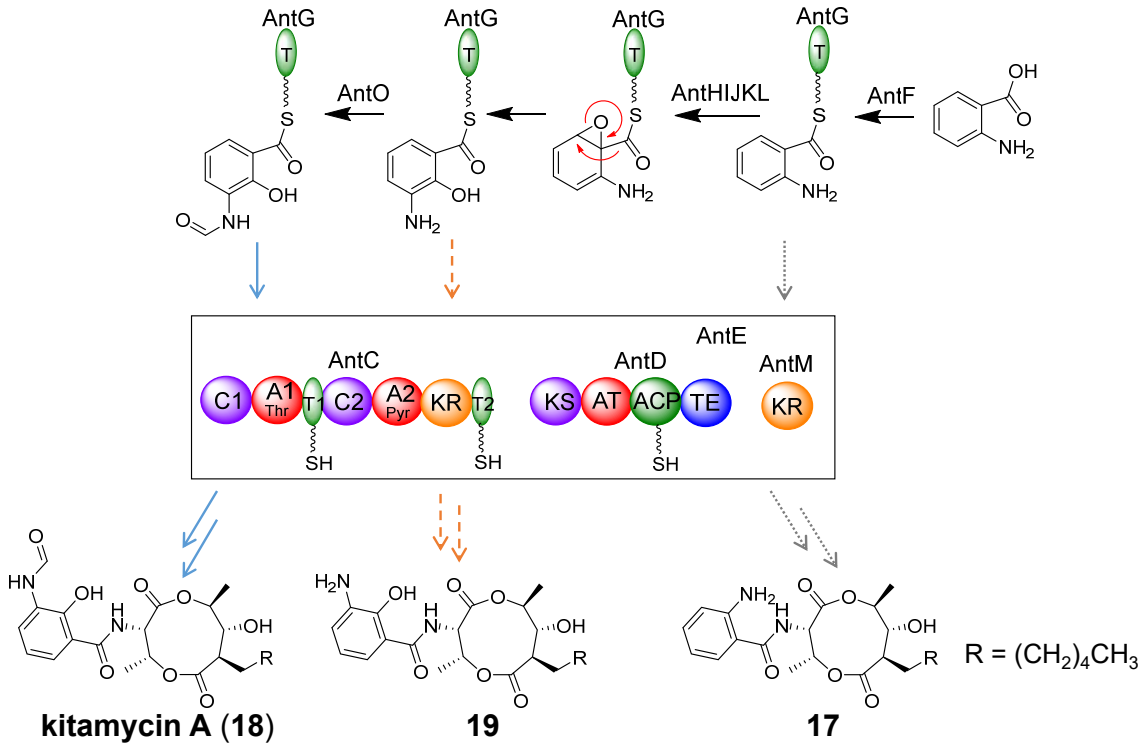


Figure 4-3. Proposed biosynthetic and shunt pathways for antimycin-type compounds.

Strain	Genes	Major products	Minor products
JL1	*antCDEFGM*	**17**	-
JL2	*antCDEFGHIJKLMNO*	**17**, **18**	-
JL3	*antCDEFGMO*	**17**	-
JL4	*antCDEFGHIJKLM*	**17**	**19**
JL5	*antCDEFGHIJKLNO*	-	-

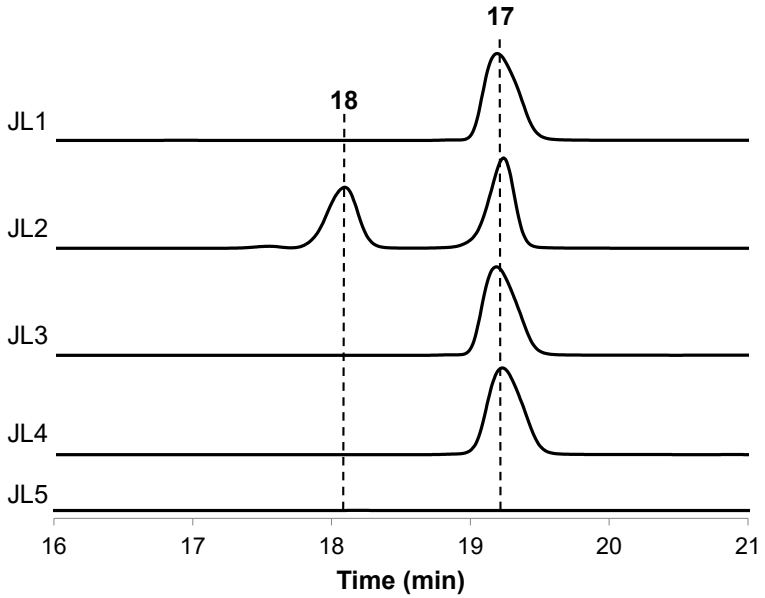


Figure 4-4. Biosynthesis of antimycins in *E. coli*. Table and extracted ion chromatograms (**17**: *m/z* 421.2333 $[M+H]^+$, **18**: *m/z* 465.2231 $[M+H]^+$, **19**: *m/z* 437.2282

[*M*+H]$^+$) showing compounds resulting from the heterologous expression of different combinations of antimycin biosynthetic genes in *E. coli*. Compound **19** was produced at ~100-fold less than **17** and could thus not be detected at this scale. The calculated mass with a 10 ppm mass error tolerance was used.

4.2.2 Elucidation of 3-formamidosalicylate biosynthesis

With the ability to generate the antimycin core scaffold in *E. coli*, we next sought to use this highly tractable system to dissect the formation of the conserved 3-formamidosalicylate pharmacophore. To this end, we introduced the remaining genes from the antimycin biosynthetic gene cluster presumed to be important for the synthesis of this moiety into the strain JL1 to yield the new strain, JL2 (Figure 4-4). The newly coexpressed genes included *antHIJKL*, which encodes a multicomponent oxygenase homologue; *antN*, which encodes a putative tryptophan 2,3-dioxygenase; and *antO*, which encodes a lipase homologue (Figure 4-2). Although previous gene disruption experiments have demonstrated *antHIJKL* to be important for the biosynthesis of antimycins,[51, 148] the possible roles of *antN* and *antO* in antimycin biosynthesis have never been studied before. Analysis of the culture extracts of JL2 by HPLC and LC-HRMS showed that the coexpression of *antCDEFGHIJKLMNO* not only resulted in the formation of **17**, but it also resulted in the formation of a product, **18**, in comparable yield (Figure 4-4 and Appendix D). HRMS/MS analysis suggested **18** to be the antimycin-type compound, kitamycin A, with a completely reconstituted 3-formamidosalicylate moiety (Figure 4-4 and Appendix C).[152]

To further confirm the identity of this compound, we purified **18** from *E. coli* culture extracts and incubated it with isobutyryl-CoA and AntB, the acyltransferase that catalyzes the C-8 acyloxy formation of antimycins.[151] This enzymatic assay resulted in the conversion of **18** into antimycin A_2 with a retention time, UV spectrum, and mass fragmentation patterns precisely matching those of the standard, thus supporting the successful reconstitution of 3-formamidosalicylate biosynthesis in *E. coli* (Figure 4-5 and Appendices C and D).

After fully reconstituting the biosynthesis of antimycins in *E. coli*, we set out to delineate the functions of the individual enzymes in the generation of the 3-formamidosalicylate group by systematically expressing different combinations of *antHIJKL*, *antN*, and *antO* in JL1. A strain containing all of the genes, except *antN*, was first constructed. LC-HRMS analysis of the extracts of this culture showed a profile very similar to that of JL2 with the production of both **17** and **18** (Figure 4-6), demonstrating that AntN is not required for the formation of the 3-formamidosalicylate moiety from anthranilate. Rather, it is known that anthranilate can be derived from tryptophan under the tandem action of a tryptophan 2,3-dioxygenase, kynureninase, and deformylase, and an operon containing these three genes is widespread in *Streptomyces*, including those that do not produce antimycins.[147] An additional copy of a tryptophan 2,3-dioxygenase encoded by *antN* in the antimycin biosynthetic gene cluster might therefore increase the flux in generating anthranilate from tryptophan, and similar strategies to increase substrate availability have been observed in biosynthesis of other natural products.[153, 154]

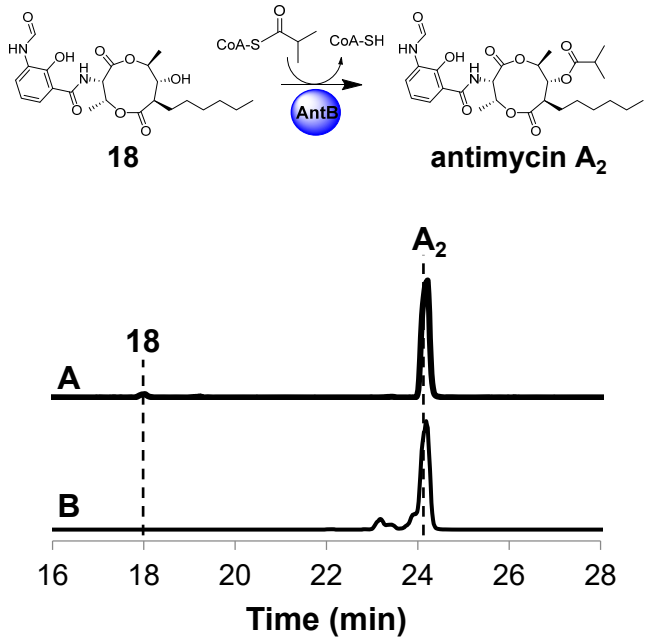


Figure 4-5. *In vitro* generation of antimycin A_2 from **18** and isobutyryl-CoA. Extracted ion chromatograms (**18**: *m/z* 465.2231 $[M+H]^+$, antimycin A_2: *m/z* 535.2650 $[M+H]^+$) showing the product from the AntB-catalyzed reaction (A) and the antimycin A_2 standard (B). The calculated mass with a 10 ppm mass error tolerance was used.

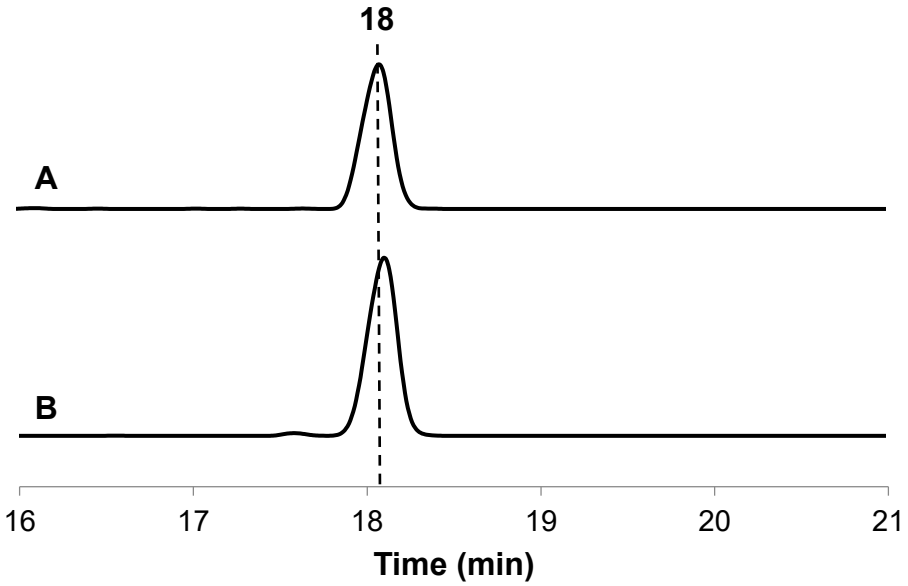


Figure 4-6. Extracted ion chromatograms showing that AntN is not required for the production of **18**. Production of **18** (*m/z* 465.2231 $[M+H]^+$) was observed from both (A) *E. coli* expressing AntCDEFGHIJKLMO and (B) *E. coli* strain JL2 expressing AntCDEFGHIJKLMNO. The calculated mass with a 10 ppm mass error tolerance was used.

Subsequently, strains lacking either *antHIJKL* (JL3) or *antO* (JL4) were constructed. LC-HRMS analysis of the culture extracts showed that the production of **18** was

completely abolished in both of these strains (Figure 4-4), confirming the essential roles of AntHIJKL and AntO in the synthesis of the 3-formamidosalicylate moiety. Careful analysis of the product profile of JL3 and JL4 provided additional insight into the reaction timing of these two enzymes. Analysis of the culture extracts of JL4 showed the formation of **17** as well as trace amounts of **19**, an antimycin-type compound proposed to contain a 3-aminosalicylate moiety resulting from the action of AntHIJKL (Figure 4-3 and Appendix C). AntHIJKL has high sequence similarity to PaaABCDE, a multicomponent oxygenase that catalyzes the epoxidation of the aromatic ring of phenylacetyl-CoA.[155-157] AntHIJKL presumably catalyzes the formation of the 3-aminosalicyloyl-*S*-AntG from anthraniloyl-*S*-AntG by an epoxidation and a 1,2-shift of the thioester group, which has been suggested by feeding experiments with isotope- and fluorine-labeled precursors and gene disruption experiments in *Streptomyces*.[51, 148] We here have reconstituted the oxidation activity of AntHIJKL in *E. coli* for the first time. Analysis of the culture extracts of JL3 then showed the formation of **17** as the major product without the presence of any deoxy-isoantimycin analogues, suggesting that oxygenation by AntHIJKL likely precedes the *N*-formylation by AntO (Figure 4-3). It is further proposed that *N*-formylation of the 3-aminosalicyloyl moiety occurs on the assembly line, as no *N*-formylated antimycin-type compounds were detected upon the feeding of **17** and **19** to *E. coli* cultures overexpressing *antO* or upon the incubation of **17** and **19** with the cell lysate of these cultures (Figure 4-7).

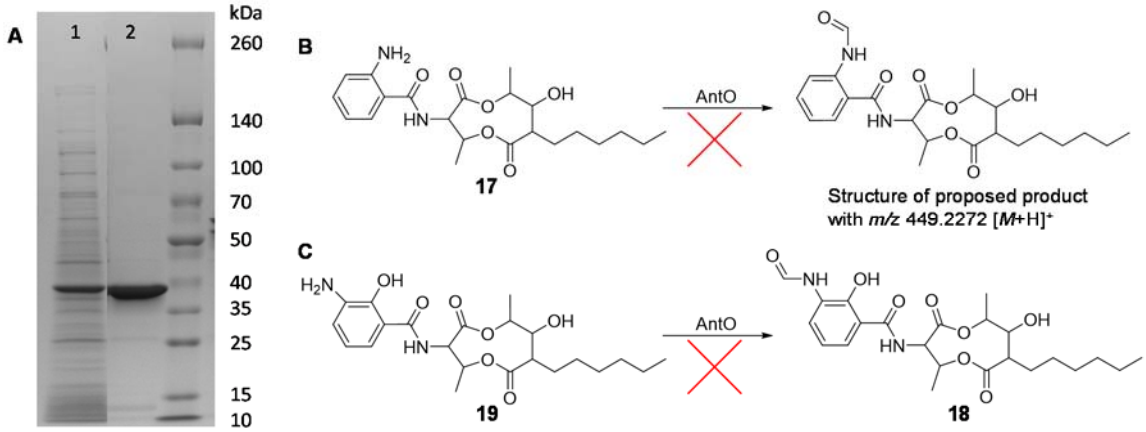


Figure 4-7. Formylation studies of **17** and **18** with *E. coli* overexpressing AntO. **(**A) SDS-PAGE analysis of lysate from *E. coli* expressing AntO (lane 1) and AntO with an *N*-terminal hexahistidine tag purified from *E. coli* (lane 2). (B) The proposed formylated product was not observed in the extracts of *E. coli* cultures expressing AntO fed with **17** or in the cell lysate incubated with **17**. (C) Compound **18** was not observed in the extracts of *E. coli* cultures expressing AntO fed with **19** or in the cell lysate incubated with **19**.

4.2.3 Fed-batch fermentation for antimycin production

To demonstrate the potential scalability of this heterologous antimycin production system, we then performed a fed-batch fermentation of the antimycin-producing JL2 strain in F1 medium supplemented with 1 mM anthranilic acid, octanoic acid, and threonine. Upon the coexpression of the 13 *ant* genes, an antimycin titer of approximately 2 mg/L was achieved when the OD_{600} plateaued at ~13 (Figure 4-8). It is anticipated that

this titer can be further increased by the manipulation of protein expression levels and cultivation conditions. As JL2 produced comparable amounts of **17** and **18**, we reasoned that the activity of AntHIJKL in *E. coli* might be the limiting aspect of the biosynthetic pathway.

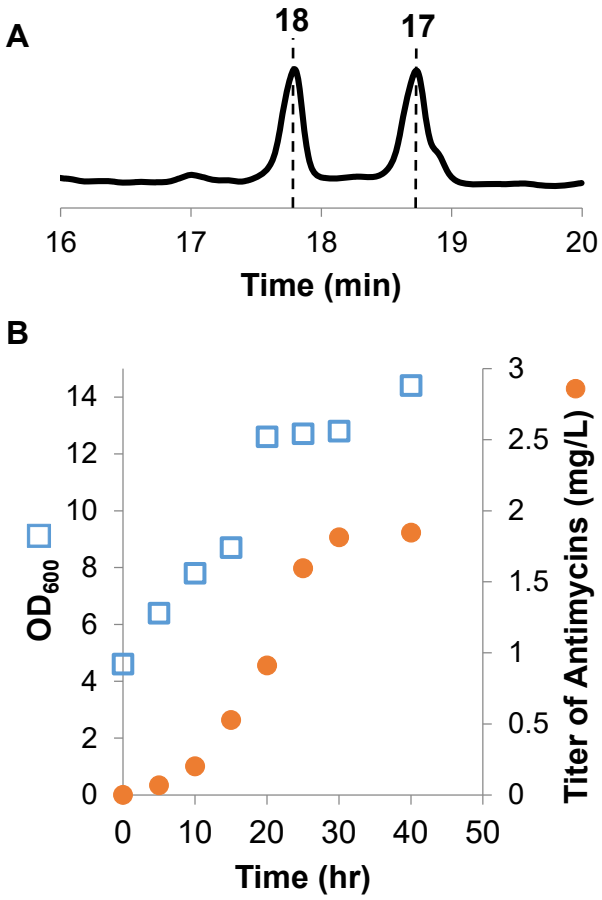


Figure 4-8. Fed-batch fermentation production of antimycins by *E. coli* expressing AntCDEFGHIJKLMNO. (A) HPLC analysis (330 nm) showing the production of **17** and **18**. (B) The maximum titer achieved was ~2 mg/L at OD600 ≈ 13, 30 h after IPTG induction.

4.2.4 Comparison of AntHIJKL homologs

We thus tested two additional AntHIJKL homologs from the antimycin biosynthetic gene cluster in *S. ambofaciens*[144] and the neoantimycin biosynthetic gene cluster in *S. orinoci*[129, 158] in our *E. coli* expression system. The subunits of these homologs share upwards of 50-80% sequence identity with AntHIJKL from *S. albus* J1074, and the coexpression of both homologs with AntCDEFGMNO resulted in the production of **18**. However, comparison of the ratio of the peak areas of **18** to **17** among the extracts of strains with different AntHIJKL homologs showed AntHIJKL from *S. albus* to be the most active followed by the homologs from *S. ambofaciens* and *S. orinoci* (Figure 4-9).

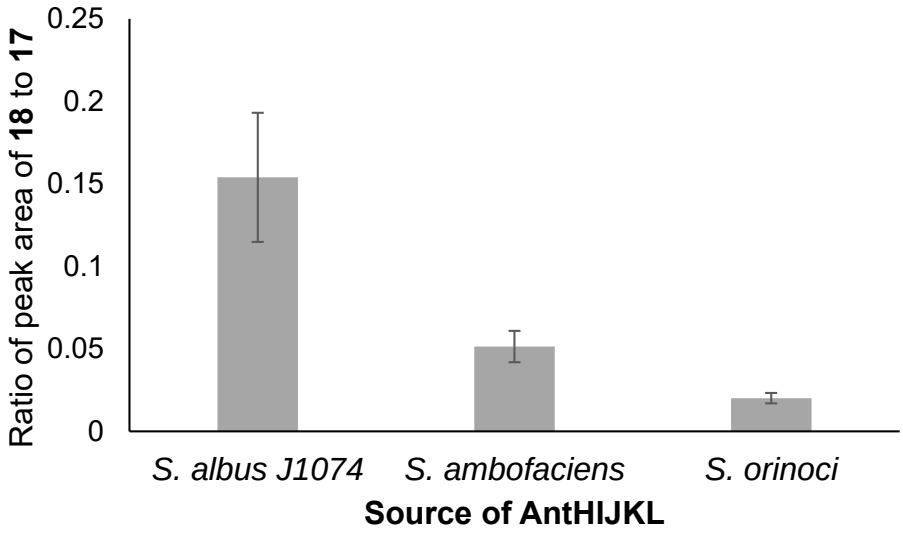


Figure 4-9. Comparison of the production of 18 in *E. coli* using AntHIJKL from different sources. The greatest relative production of **18** resulted from expressing *S. albus* AntHIJKL followed by the *S. ambofaciens* and *S. orinoci* AntHIJKL homologs. The values of the ratios of the peak areas may be underestimated here because **17** is likely to be more easily detected than **18** by mass spectrometry. Error bars represent standard deviations from three replicates.

4.2.5 Characterization of *S. albus* J1074 Δ*antO*

To further probe the function of AntO in antimycin biosynthesis, we also carried out a gene disruption experiment in the antimycin-producing organism, *S. albus* J1074. An *S. albus* J1074 Δ*antO* mutant strain was constructed using the recently reported CRISPR/Cas system for *Streptomyces*.[42] HPLC and LC-HRMS analysis of the culture extracts of this strain showed antimycin production to be abolished (Figure 4-10). While the production of deformylated antimycins was too low to be detected by HPLC, further analysis by LC-HRMS revealed the production of small amounts of putative deformylated antimycins (Figure 4-11). This observation is consistent with the aforementioned proposal that *N*-formylation of the 3-aminosalicylate moiety occurs on the assembly line, likely in the early stage as shown in Figure 4-2, and the disruption of *N*-formylation significantly impairs the NRPS-PKS assembly line. Because AntO also shows homology to the deformylase involved in the conversion of tryptophan to anthranilate, this possible role of AntO was further assessed by feeding anthranilic acid to the cultures of a Δ*antO* mutant.[159] However, the titer of the putative deformylated antimycins did not increase upon the exogenous addition of anthranilic acid, eliminating the possibility that the low product titer is due to a limited supply of anthranilate precursors being generated through the additional function of AntO in antimycin biosynthesis.[159]

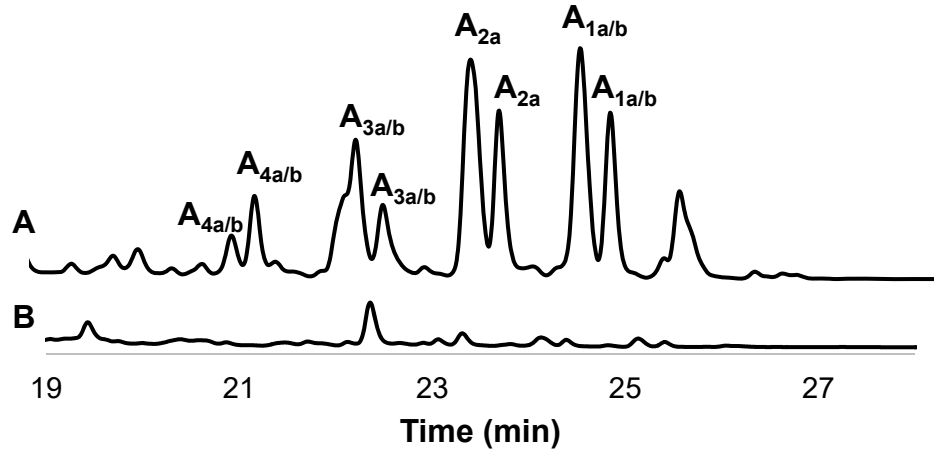


Figure 4-10. HPLC-UV analysis of antimycins produced by wild-type *S. albus* J1074 and *S. albus* J1074 Δ*antO*. Traces (330 nm) show the production of antimycins A_1-A_4 by *S. albus* J1074 (A) but not by *S. albus* J1074 Δ*antO* (B). a/b represents different isomers of antimycins A_1-A_4, the structures of which are shown in Figure 4-1.

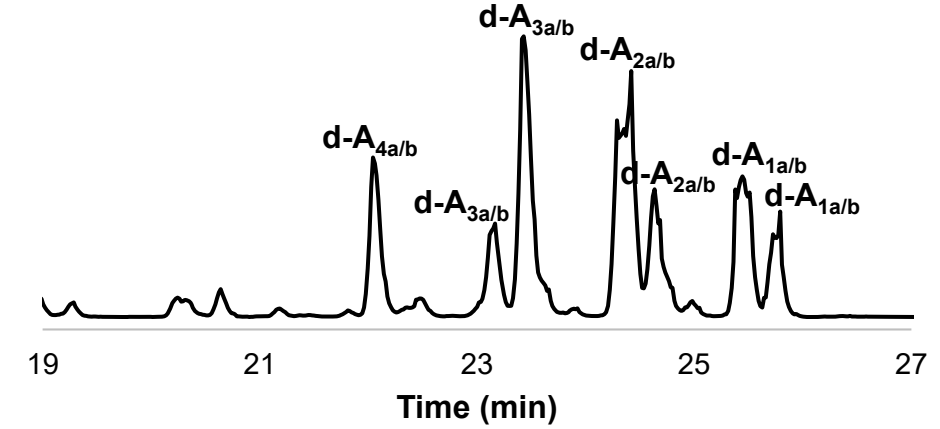


Figure 4-11. Extracted ion chromatograms showing the production of deformylated antimycins by *S. albus* J1074 Δ*antO*. The calculated mass with a 10 ppm mass error tolerance was used.

4.3 Discussion

In summary, we have established a versatile system for the biosynthesis of antimycins in *E. coli*.[159] The heterologous expression and mutagenesis studies reported here have led to a working hypothesis for the biosynthetic pathway of the 3-formamidosalicylate pharmacophore conserved in all antimycin-type depsipeptides. In particular, the substrate, anthranilic acid, is first activated by AntF and loaded onto AntG. The resulting anthraniloyl-*S*-AntG is then subjected to oxidation and rearrangement reactions catalyzed by AntHIJKL and a formylation reaction catalyzed by AntO to generate the 3-formamidosalicylate starter unit for the hybrid NRPS-PKS assembly line (Figure 4-2). Unexpectedly, the *E. coli* production titer of **18**, which contains the natural starter unit, was not significantly increased compared to that of **17**, which is generated from a disrupted modular assembly line with an unnatural starter unit (Figure 4-8A).

Since the overexpression of *antO* in *E. coli* yielded more than 15 mg/L of soluble protein (Figure 4-7A), the low titer of **18** in this heterologous host is possibly due to the low activity of the multicomponent oxygenase, AntHIJKL, and the reconstitution of oxygenase activity *in vitro* using purified proteins from *E. coli* has not yet been successful. Screening additional genes homologous to *antHIJKL* from the dozens of other gene clusters identified for antimycin-type depsipeptides may yield new enzyme candidates with improved efficiency in *E. coli* and lead to much higher titers of antimycins.[122, 129, 144] Nonetheless, our work not only shows additional promise in using *E. coli* as a heterologous host for the production of natural products synthesized through NRPS/PKS enzymes, but it also paves the way for the rapid generation of new antimycin analogues through combinatorial biosynthesis. As an example, this antimycin expression system was recently used in combination with genes encoding terminal alkyne biosynthetic machinery to result in the production of a novel terminal alkyne-bearing antimycin analogue.[53]

4.4 Materials and Methods

Bacterial strains, plasmids, and growth conditions

Bacterial strains and plasmids used in this study are described in Tables 4-1 and 4-2, respectively. LB was used for the propagation of *E. coli* strains at 37°C unless otherwise specified. Mannitol-soy flour (MS) medium (20 g/L soy flour and 20 g/L mannitol) was used to maintain *S. albus* J1074. Growth media was supplemented with antibiotics as required at the following concentrations unless otherwise indicated: carbenicillin (100 μg/mL), kanamycin (50 μg/mL), spectinomycin (50 μg/mL), and apramycin (50 μg/mL).

Table 4-1. Bacterial strains used in this study.

Strain	Function
E. coli XL1-Blue	Cloning using standard molecular biology procedures
E. coli BL21 Gold(DE3)	Overexpression of the individual proteins, AntB and AntO
E. coli BAP1	Heterologous expression of antimycins
S. albus J1074	Construction of Δ*antO* mutant

Table 4-2. Plasmids used in this study.

Plasmid	Derived from	Function
pETDuet-*antC-antE*	pETDuet-1	Co-expression of AntC and AntE in *E. coli*
pCDFDuet-*antD-antGF*	pCDFDuet-1	Co-expression of AntD and AntFG in *E. coli*
pCOLA-*antM*	pCOLADuet-1	Expression of AntM in *E. coli*
pCOLA-*antMO*	pCOLADuet-1	Co-expression of AntM and AntO in *E. coli*
pCOLA-*antHIJKL-antM*	pCOLADuet-1	Co-expression of AntHIJKL and AntM in *E. coli*
pCOLA-*antHIJKL-antMO*	pCOLADuet-1	Co-expression of AntHIJKL and AntMO in *E. coli*

pCOLA-*antHIJKL-antMNO*	pCOLADuet-1	Co-expression of AntHIJKL and AntMNO in *E. coli*
pCOLA-*antHIJKL-antNO*	pCOLADuet-1	Co-expression of AntHIJKL and AntNO in *E. coli*
pET30-*antB*	pET30-Xa/LIC	Expression of AntB in *E. coli*
pET30-*antO*	pET30-Xa/LIC	Expression of AntO in *E. coli*
pCOLA-Samb_*antHIJKL-antMNO*	pCOLADuet-1	Co-expression of AntHIJKL (homolog from *S. ambofaciens*) and AntMNO in *E. coli*
pCOLA-*natHIJKL-antMNO*	pCOLADuet-1	Co-expression of NatHIJKL and AntMNO in *E. coli*
pCRISPomyces2-dantOgRNA-2kb	pCRISPomyces-2	Construction of *S. albus* J1074 Δ*antO* mutant using CRISPR/Cas

Construction of plasmids used for heterologous expression of antimycins

Primers used in this study are listed in Table 4-3. *S. albus* J1074 genomic DNA was used to PCR amplify *antC*, *D*, *GF*, *HIJKL*, and *M*/*MNO*/*MO*, and *S. ambofaciens* genomic DNA was used to PCR amplify *antE*. The *antC* gene was digested from pET30-*antC*[51] using NdeI/XhoI (Thermo Scientific) and ligated into MCS1 of pETDuet-1 (Novagen) using Quick T4 DNA ligase (New England Biolabs). The resulting plasmid was then digested with PstI/HindIII for the ligation of *antE* (PCR product generated using antEam-Duet-F/R primers) into MCS2 to yield the final pETDuet-*antC-antE* construct. The *antD* gene (PCR product generated using antD-Duet-F/R primers) was cloned into MCS1 of pCDFDuet-1 using the EcoRI/HindIII restriction sites, and *antGF* (PCR product generated using AntG_NdeI/antGF-Duet-R primers) was subsequently cloned into MCS2 of the resulting plasmid using the NdeI/XhoI restriction sites to yield the final pCDFDuet-*antD-antGF* construct. The *antM*, *MNO*, or *MO* genes (PCR products generated using antM-Duet-F/R, antM-Duet-F/antNO-Duet-R, or antM-Duet-F/antMO-F2 + antMO-F2/antNO-Duet-R primers, respectively) were cloned into MCS2 of pCOLADuet-1 using the NdeI/XhoI restriction sites to form pCOLADuet-*antM*/*MNO*/*MO*. These plasmids were then further digested with PstI/HindIII for the introduction of *antHIJKL* (PCR product generated using antHIJKL-Duet-F/R primers) into MCS1 to form pCOLADuet-*antHIJKL-antM*/*MNO*/*MO*. Plasmids were isolated using a QIAprep Spin Miniprep Kit (Qiagen) and confirmed by DNA sequencing (UC Berkeley DNA Sequencing Facility).

For the evaluation of AntHIJKL homologs, *antHIJKL* was also amplified from *S. ambofaciens* and *S. orinoci* genomic DNA and cloned into pCOLADuet-antMNO as described above.

Table 4-3. Primers used in this study.

Primer	**Sequence (5' -> 3')***	**Description**
antEam-Duet-F	aaa**CTGCAG**atgagagcgacccaagacctc	Expression of AntE
antEam-Duet-R	tat**AAGCTT**tcacatccggtccccgactc	
antD-Duet-F	aaa**GAATTC**Gatggatgacccgcaggccct	Expression of AntD
antD-Duet-R	tat**AAGCTT**tcagccggaggccgggcgca	
AntG_NdeI	aaa**CATATG**agcaccatcagggagtt	Expression of

antGF-Duet-R	aaa**CTCGAG**tcaggcgcgcagcgccttct	AntFG
antM-Duet-F	aaa**catATG**accaccacccagcggca	Expression of AntM
antM-Duet-R	aaa**CTCGAG**tcacagaccgaggccgacac	
antMO-R2	ggagatgaccctagagcgcatcggggctcctagggaca	Expression of AntMO
antMO-F2	tgtccctaggagccccgatgcgctctagggtcatctcc	
antNO-Duet-F	aaa**catATG**aacgacggcccgtccgc	Expression of AntNO
antNO-Duet-R	aaa**CTCGAG**tcatgcgcgggccatccgcc	
antHIJKL-Duet-F	aaa**CTGCAG**atgaccggctcaccggcgcg	Expression of AntHIJKL
antHIJKL-Duet-R	tat**AAGCTT**tcaggcgtcgaagtcgagcg	
AntB_PET30F	GGTATTGAGGGTCGCatgcgacgcgcgctgtgtcc	Expression of AntB
AntB_PET30R	AGAGGAGAGTTAGAGCCtcagccggcgctggccgg	
AntO_PET30F	GGTATTGAGGGTCGCatgcgctctagggtcatctcc	Expression of AntO
AntO_PET30R	AGAGGAGAGTTAGAGCCtcatgcgcgggccatccgcc	
antH-Samb-Duet-F	aaa**CTGCAG**atgagcgcgccgaccctcgc	Expression of AntHIJKL (homolog from *S.ambofaciens*)
antL-Samb-Duet-R	tat**AAGCTT**caggtgtcgaagtcgaggg	
natHIJKL-Duet-F	aaa**CTGCAG**atggatgacgagggagtggt	Expression of NatHIJKL
natHIJKL-Duet-R	tat**AAGCTT**acacatcgaagtcgaggc	
antO-gRNA-F	ACGCgggagcgtggaggtgtgcca	Construction of *S. albus* J1074 Δ*antO* mutant using CRISPR/Cas
antO-gRNA-R	AAACtggcacacctccacgctccc	
antO-up-F	tcggttgccgccgggcgtttttttatctaga gacggtccgctccacgctcg	
antO-up-R	agacatctggcgggcggtca caccgcgtgcctcccacagc	
antO-down-F	gctgtgggaggcacgcggtg tgaccgcccgccagatgtct	
antO-down-R	gcggccttttttacggttcctggcctctaga cgcttggcgacctcctcggg	
dantO-F	cttcaccgcgctgtacgtgc	Confirmation of *S. albus* J1074 Δ*antO* mutant
dantO-R	gccaggcacaggtgctgttc	

*restriction endonuclease sites are bolded

Biosynthesis of antimycins in E. coli

The plasmids, pETDuet-*antC*-*antE*, pCDFDuet-*antD*-*antGF*, and pCOLADuet-(*antHIJKL)*-*antM*(*N*)(*O*), were electroporated into *E. coli* BAP1, and transformants were selected on LB agar plates supplemented with the appropriate antibiotics. Single colonies were inoculated into 3 mL of LB with antibiotics and grown overnight at 37°C as a seed culture, of which 0.25 mL was used to inoculate 25 mL of fresh LB medium with antibiotics. The cultures were then grown at 37°C to $OD_{600} \approx 0.4 – 0.6$ before induction with 0.12 mM IPTG and the addition of anthranilic acid and octanoic acid at a final concentration of 1 mM. After induction, the temperature was dropped to 20°C, and compound production was allowed to proceed for approximately 2 days.

The F1 fed-batch fermentation was conducted with a DASGIP Parallel Bioreactor System, and methods for the fermentation and medium composition were adopted from methods described previously.[111, 160] A starter culture was grown overnight at 37°C in 3 mL of LB medium supplemented with the appropriate antibiotics, and 1 mL of this culture was used to inoculate 50 mL of LB medium with antibiotics. The culture was then grown overnight at 30°C before being centrifuged and resuspended in 20 mL of F1

medium, of which 8 mL was used to inoculate a 1 L vessel containing 800 mL of F1 medium supplemented with 150 μg/mL carbenicillin, 100 μg/mL spectinomycin, and 75 μg/mL kanamycin. The fermentation was started at 37°C, and the pH was maintained at 7.0 throughout the experiment with concentrated HCl and half-concentrated NH_4OH. Aeration was controlled at 0.2 L/min, and agitation was maintained at 600 rpm. When the OD_{600} reached ~4 – 5, the temperature of the fermentation was reduced to 20°C, followed by the addition of 0.5 mM IPTG, 1 mM anthranilic acid, 1 mM octanoic acid, and 1 mM threonine. At the same time, 0.1 mL/min of feed medium began to be delivered to the fermenter. At each time point, a 1 mL aliquot was removed from the culture for compound extraction and analysis.

HPLC and LC-MS analysis of antimycin production

E. coli culture samples were pelleted by centrifugation (4000 × *g* for 10 min), and the supernatant was extracted with two volumes of ethyl acetate. The solvent was removed by rotary evaporation (or centrifugal evaporation for 1 mL samples), and the residue was redissolved in methanol (400 μL for 25 mL cultures and 100 μL for 1 mL samples) for LC-HRMS, HRMS/MS, and HPLC analysis (20 μL injection). LC-HRMS analysis was performed on an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS instrument with an Agilent Eclipse Plus C18 column (4.6 × 100 mm). A linear gradient of 25-95% CH_3CN (v/v) over 20 min followed by an additional 10 min of 95% CH_3CN (v/v) in H_2O with 0.1% formic acid (v/v) at a flow rate of 0.5 mL/min was used. A collision energy of 20 V was used for all HRMS/MS experiments. Antimycins were quantified by generating a standard curve with antimycin standards and using the Agilent MassHunter Qualitative Analysis software to determine the extracted ion chromatogram peak areas. HPLC analysis was performed on an Agilent Technologies 1260 LC system with DAD using the same column and program as described above but with trifluoroacetic acid in place of formic acid.

Enzymatic synthesis of antimycin A_2

AntB was expressed and purified as described previously.[151] Compound **18** was purified by RP-HPLC (Agilent 1260 HPLC with DAD) using an Agilent Eclipse Plus C18 column (4.6 × 100 mm) with a linear gradient of 25-95% CH_3CN (v/v) in H_2O over 20 min at a flow rate of 0.5 mL/min. Fractions containing antimycins were collected manually and concentrated under vacuum. Assays were then performed in 100 μL of 50 mM HEPES (pH 8.0) containing **18**, 2 mM isobutyryl CoA, and 20 μM AntB. After a 2 h incubation period at room temperature, the reactions were extracted with two volumes of ethyl acetate, and the organic layer was dried under vacuum. The residue was then redissolved in methanol (50 μL) for HPLC, LC-HRMS, and HRMS/MS analysis as described above. The antimycin A_2 standard was obtained from Sigma and used for comparison with the product of the enzymatic assay.

Formylation studies with E. coli overexpressing AntO

The *antO* gene was PCR amplified from *S. albus* J1074 genomic DNA using primers AntO_PET30F/R and cloned into pET-30 XA/LIC (Novagen) following standard protocols. The resulting pET30-*antO* plasmid was transformed into BL21 Gold (DE3), and single colonies were inoculated into 3 mL of LB with kanamycin and grown

overnight at 37°C as a seed culture, of which 0.25 mL was used to inoculate 25 mL of fresh LB with kanamycin. The cultures were then grown at 37°C to $OD_{600} \approx 0.4 – 0.6$ and induced with 0.12 mM IPTG. AntO was expressed at 20°C for 2 h after which compounds **17** and **19**, which were partially purified by running the culture extracts of 250 mL of JL5 through Sep-Pak C18 Plus cartridges (Waters) and eluting with 95% CH_3CN, were added to the cultures. Expression of AntO at 20°C was resumed for approximately 2 days before extraction and product analysis as described above. For assays with *E. coli* lysate, AntO was expressed in BL21 Gold (DE3) for 16-20 h after which the cells from 350 mL of culture were pelleted by centrifugation (6000 × *g* for 15 min), resuspended in 20 mL of lysis buffer (20 mM HEPES, pH 8, 0.5 M NaCl), and lysed by homogenization on ice. Cellular debris was removed by centrifugation (15000 × *g* for 1 h), and the soluble fraction was incubated with **17** and **19** purified from an equivalent amount of culture. After overnight incubation at room temperature, the lysate was subjected to extraction and product analysis as described above.

*Construction and analysis of S. albus J1074 Δ*antO*

The *S. albus* J1047 Δ*antO* mutant strain was constructed using the CRISPR/Cas system engineered for *Streptomyces*.[42] The sequence for the synthetic guide RNA was first cloned into pCRISPomyces-2 by Golden Gate assembly, and a 4 kb editing template consisting of two 2 kb arms homologous to the sequence up and downstream of *antO* was introduced into the resulting plasmid by Gibson assembly. The final pCRISPomyces2-dantOgRNA-2kb plasmid was transformed into *E. coli* WM6026 for conjugation with *S. albus* J1074. Successful transformants were selected by apramycin resistance, and deletion of Δ*antO* was confirmed by PCR with the dantO-F/R primers. Clearance of the plasmid was performed as described by previously[42] with the exception that the strains were cultured in TSB medium.

To check for the production of deformylated antimycins, starter cultures of the confirmed *S. albus* J1047 Δ*antO* mutant strains were inoculated with mycelia and cultured in TSB for 2 days. 0.5 mL of the starter culture was then inoculated into 25 mL of MS, and the cultures were grown at 30ºC, 250 rpm for 5 days. Antimycins were extracted from the cell-free supernatant (25 mL) using ethyl acetate (two volumes). The solvent was removed by rotary evaporation and the residue was redissolved in methanol (0.5 mL) and analyzed by LC-HRMS (10 µL injection) and HPLC (20 µL injection) as described above.

Chapter 5. Expanding on the Biosynthesis of Neoantimycins

Parts of this chapter have been adapted from the following with permission:
Liu, J., Zhu, X., Kim, S., Zhang, W. "Antimycin-type depsipeptides: discovery, biosynthesis, chemical synthesis, and bioactivties." *Nat. Prod. Rep.* DOI: 10.1039/c6np00004e (2016).

5.1 Introduction

While the most well-studied group of antimycin-type depsipeptides are the antimycins with a 9-membered dilactone ring, antimycin-type depsipeptides possessing 12-, 15-, and 18-membered macrocyclic rings have also been reported (Figure 4-1).[122, 129-132, 141] Like the 9-membered ring antimycins, these ring-expanded antimycins also possess a 3-formamidosalicylate moiety, but their biological activities appear to extend beyond the function of this pharmacophore. Specifically, several of the 12- and 15-membered antimycin-type depsipeptides have been shown to downregulate GRP-78,[130, 141, 143, 161] and some of the 15-membered ring neoantimycins have also shown immunosuppressive activity.[162] Moreover, some of the 18-membered ring antimycin-type depsipeptides have exhibited nanomolar inhibition of cancer cell growth though their mechanism of action is currently unknown.[132, 163]

The biosynthetic gene clusters responsible for the formation of the ring-expanded antimycin-type depsipeptides were recently reported by the Magarvey group (Figure 5-1).[129, 158, 164] As expected, all of these clusters feature a hybrid NRPS-PKS assembly line and encode homologs to the antimycin biosynthetic enzymes involved in starter unit formation: AntF, AntG, AntHIJKL, AntN, and AntO. However, the number of modules in the NRPS-PKS assembly line of each of the four classes of antimycin-type depsipeptides differs depending on the size of the macrolactone ring. Specifically, the *ant*, *sml*, *nat*, and *ksn* gene clusters which produce antimycins, JBIR-06/JBIR-52, neoantimycins, and respirantins have a total of 4, 5, 6, and 7 modules, respectively (Figure 5-1). This set of naturally combinatorial gene clusters thus represents a unique system that can be used to study the underlying programming rules for combinatorial biosynthesis, and different domain and module swapping experiments among these naturally combinatorial enzymatic assembly lines can provide new insight into the effective engineering of NRPS-PKS assembly lines. In particular, this system can be used to determine the factors that control ring expansion of these NRPS-PKS hybrid molecules especially with regards to the necessary intermodular protein-protein interactions and the role of the off-loading thioesterase. Moreover, these related but different gene clusters can also be used to elucidate the rules that govern atypical PKS monomer selection, since all antimycin-type depsipeptide gene clusters encode a single module PKS, but the PKS from the antimycin biosynthetic gene cluster is unlike the others in that it uses alkyl-chain monomers rather than the more commonly used malonate or methylmalonate substrates.

Here, we describe our initial efforts toward understanding and reconstituting the biosynthesis of neoantimycins both *in vivo* and *in vitro*, which is imperative for developing an advanced platform that can be used to study the programming rules of this naturally combinatorial system. We begin with a revision of the previously reported neoantimyin gene cluster and discuss our efforts to produce neoantimycins in a more genetically tractable heterologous host. We further show the use of our heterologous

expression system in elucidating the functions of some of the freestanding enzymes encoded by the gene cluster. Finally, we describe work that has been done toward the biochemical characterization of the neoantimycin NRPS-PKS assembly line. These results along with our comprehensive knowledge about the biosynthesis of antimycins with a 9-membered dilactone ring can potentially enable us to further our understanding about the approaches needed to achieve successful combinatorial biosynthesis.

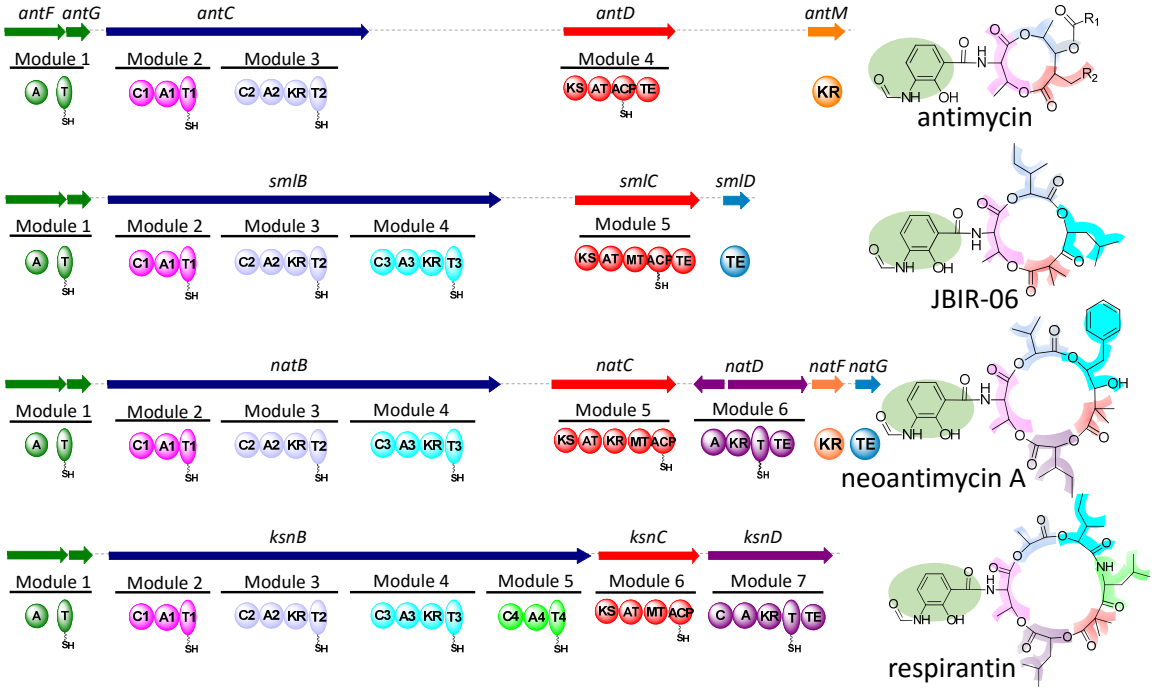


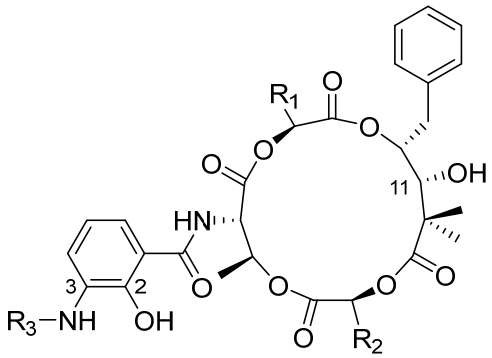
Figure 5-1. Biosynthetic gene clusters for the four classes of antimycin-type depsipeptides. Although not shown here, all clusters also have genes homologous to *antHIJKLNO(P)* for formation of the 3-formamidosalicyloyl moiety as well as a gene encoding an RNA polymerase sigma factor for regulation. With the exception of some antimycin gene clusters, all clusters also have a gene encoding a 4'-phosphopantetheinyl transferase for the post-translational modification of the T and ACP domains.

5.2 Results

5.2.1 Analysis of the neoantimycin biosynthetic gene cluster

The biosynthetic gene cluster that was initially reported for neoantimycin showed the presence of a three-module NRPS with the first module of NatB being an L-threonine-loading unit that is highly conserved among the *ant*, *sml*, and *ksn* clusters (Figure 5-1).[129, 158] Based on the natural structural variation of the neoantimycins reported to be produced by *S. orinoci* (Figure 5-2), the second module of NatB is then expected to select and activate 2-oxo-3-methylbutanoic acid or pyruvate, while the third is expected to select and activate phenylpyruvic acid (Figure 5-1). The downstream PKS, NatC, was proposed to have a domain organization of KS-AT-KR-MT-ACP and load either malonyl- or methylmalonyl-CoA, and the last module of the assembly line was proposed to be a single NRPS module capable of using both 2-oxo-3-methylpentanoic and 2-oxo-3-methylbutanoic acid as substrates (Figures 5-1 and 5-2). The unusual A-KR-T-TE

domain organization of this last NRPS module and the fact that it was encoded across two genes prompted further investigation of this gene cluster through additional sequencing and annotation. Specifically, a cosmid spanning from *natA*, which encodes an RNA polymerase sigma factor for regulation, to the beginning of *natF* was obtained from Dr. Ryan F. Seipke (University of Leeds), and the region between *natC* and *natF* on this cosmid was sequenced by primer walking (Figure 5-3). These results revealed the last module in the neoantimycin gene cluster to be encoded by a single gene, *natD*, and further analysis of this NRPS showed that it has a domain organization of C-A-KR-T-TE, which is similar to the last module of the respirantin gene cluster (Figures 5-1 and 5-3). Closer analysis of the PKS, NatC, also indicated the lack of a KR domain, making it more closely resemble the PKS in the respirantin gene cluster as well. Furthermore, additional annotation of the neoantimycin gene cluster using RAST[165, 166] revealed a small gene, *natE*, located between *natD* and *natF* (Figure 5-3). This gene is predicted to encode an MbtH-like protein, and such proteins have previously been shown to be involved in activating the A domains of NRPSs.[167-169]



	R_1	R_2	R_3
20 neoantimycin A	$CH(CH_3)_2$	$CH(CH_3)CH_2CH_3$	CHO
21 neoantimycin F	$CH(CH_3)_2$	$CH(CH_3)_2$	CHO
22 SW-163A	CH_3	$CH(CH_3)CH_2CH_3$	CHO
23 neoantimycin D	$CH(CH_3)_2$	$CH(CH_3)CH_2CH_3$	H
24 neoantimycin E	$CH(CH_3)_2$	$CH(CH_3)_2$	H
25 JBIR-05	CH_3	$CH(CH_3)CH_2CH_3$	H

Figure 5-2. Neoantimycins produced by *S. orinoci*.

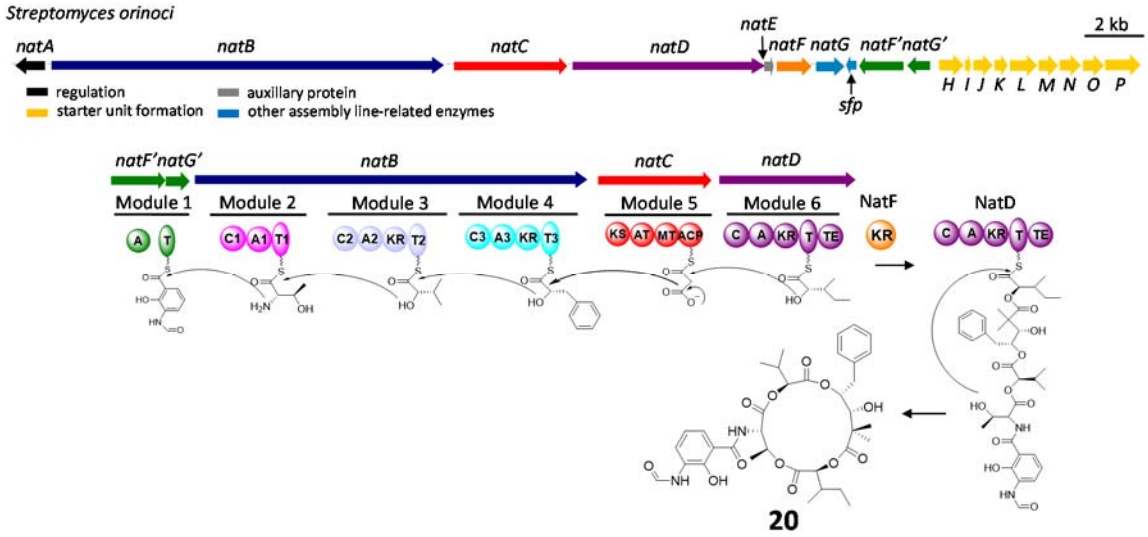


Figure 5-3. Revised neoantimycin biosynthetic gene cluster and proposed biosynthetic pathway to **20**.

5.2.2 Neoantimycin production in *S. albus* J1074

As the native producer of neoantimycins, *S. orinoci*, has been proven to be genetically intractable, the expression of the neoantiymcin biosynthetic gene cluster in a heterologous host can facilitate the further study and manipulation of this NRPS-PKS assembly line. Toward this end, the antimycin producer, *S. albus* J1074, is an ideal host for several reasons: 1) compared to most *Streptomyces*, it is relatively easy to genetically manipulate and culture, and genome editing by CRISPR/Cas has been successfully demonstrated in this strain;[42] 2) given its production of antimycins, it is likely to have all of the necessary precursors for neoantimycin biosynthesis; and 3) it possesses the genes responsible for the formation of the 3-formamidosalicylate starter unit, which are missing from the available cosmid. The cosmid encoding *natA-E* was then subcloned into an *E. coli-Streptomyces* shuttle vector that additionally had *natFG* cloned under a constitutive PermE* promoter, and the resulting plasmid was introduced into *S. albus* J1074 by conjugative transfer. Analysis of the culture extracts of this strain by LC-HRMS and showed the heterologous production of all six of the neoantimycins that have previously been isolated from *S. orinoci* (**20**-**25**),[158] and the titers of these compounds were comparable to those from the native producer (Figures 5-2 and 5-4). The production of neoantimycins by *S. albus* J1074 heterologously expressing *natABCDEFG* also indicates that NatB can recognize AntG in lieu of NatG' (Figure 5-3).

In addition, we also expressed this set of neoantimycin biosynthetic genes in several *S. albus* J1074 mutants that were generated using CRISPR/Cas.[42] Heterologous expression of neoantimycins in the *S. albus* J1074 Δ*antC* mutant strain enabled neoantimycin production with a cleaner metabolite background without antimycins, while the expression of neoantimycins in the *S. albus* J1074 Δ*antO* mutant strain resulted in an approximately ten-fold increase in the production of deformylated neoantimycins, **23**-**25** (Figure 5-5). The selective and increased production of these deformylated neoantimycins can facilitate their purification for more extensive bioassays, since little is

currently known about the mechanism of action of neoantimycins and whether the 3-formamidosalicylate moiety is responsible for their reported anticancer activities.

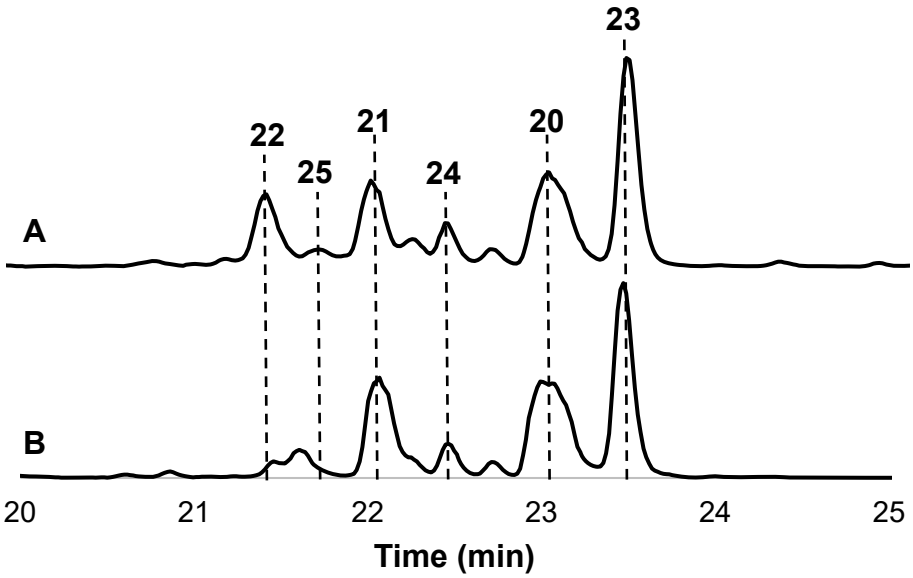


Figure 5-4. Extracted ion chromatograms showing the production of neoantimycins by *S. albus* J1074 (A) and *S. orinoci* (B). The relative amounts of each of the neoantimycins may not be reflected by their peak areas here, since compounds **23**-**25** may be more easily detected by mass spectrometry than compounds **20**-**22**. The calculated mass with a 10 ppm mass error tolerance was used.

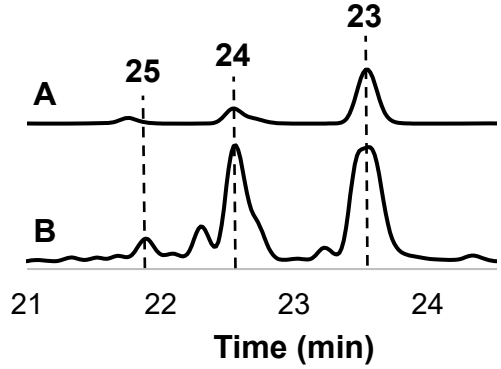
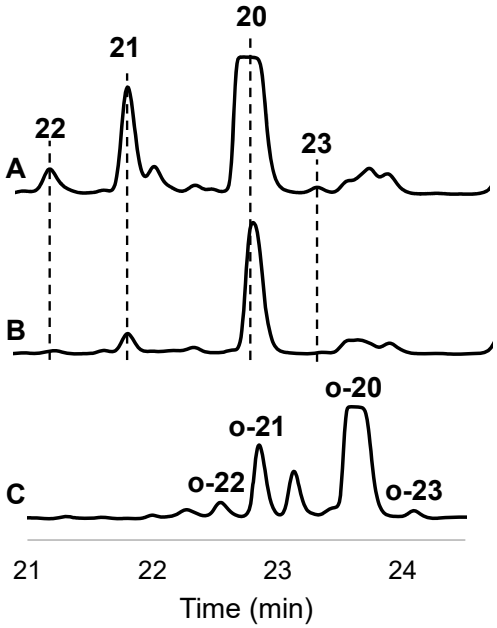


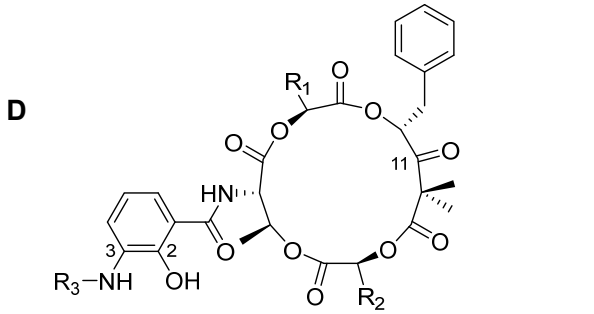
Figure 5-5. Extracted ion chromatograms showing the production of deformylated neoantimycins by *S. albus* J1074 Δ*antC* (A) and *S. albus* J1074 Δ*antO* (B). The calculated mass with a 10 ppm mass error tolerance was used.

5.2.3 Investigation into the functions of NatF and NatG

Two freestanding domains: a ketoreductase encoded by *natF* and a thioesterase encoded by *natG* are also found in the neoantimycin biosynthetic gene cluster, and although they were initially included for heterologous expression, their precise roles in neoantimycin biosynthesis were unclear. To elucidate the functions of these domains, we constructed two additional heterologous expression plasmids with *natABCDEF* and *natABCDEG*. These plasmids were likewise introduced into *S. albus* J1074, and the culture extracts of the resulting strains were analyzed by LC-UV-MS and compared to

the extracts of *S. albus* J1074 heterologously expressing *natABCDEFG*. Neoantimycins were still produced by the strain lacking *natG* albeit at reduced titers (Figure 5-6), leading us to conclude that NatG is not required for neoantimycin biosynthesis and may have an editing role. Similar type II thioesterases have been found in other PKS and NRPS gene clusters and have been shown to hydrolyze incorrect substrates and aberrant intermediates that may block the megaenzyme, leading to increased compound production.[170-172] Conversely, the *S. albus* J1074 strain lacking NatF did not appear to produce neoantimycins and instead showed a suite of compounds with UV absorption spectra identical to those of neoantimycins but with shifted retention times (Figure 5-6). Further analysis of these compounds by HRMS then indicated them to be oxidized variants of **20**-**25**, suggesting that NatF is likely responsible for reducing the C11 keto group to a hydroxyl (Figures 5-2 and 5-6). The production of these oxidized neoantimycins is also consistent with previous reports characterizing similar compounds such as prunustatin A[143, 161] and neoantimycin H (Figure 5-6).[173]





		R_1	R_2	R_3
o-20	neoantimycin H	$CH(CH_3)_2$	$CH(CH_3)CH_2CH_3$	CHO
o-21		$CH(CH_3)_2$	$CH(CH_3)_2$	CHO
o-22	prunustatin A	CH_3	$CH(CH_3)CH_2CH_3$	CHO
o-23		$CH(CH_3)_2$	$CH(CH_3)CH_2CH_3$	H

Figure 5-6. HPLC-UV analysis of neoantimycins produced by *S. albus* J1074 expressing different combinations of NatF and NatG. Traces (320 nm) show neoantimycin (**20**-**23**) production by *S. albus* J1074 expressing *natABCDEFG* (A) and reduced neoantimycin production by *S. albus* J1074 expressing *natABCDEF* (B). The production of oxidized neoantimycins by *S. albus* J1074 expressing *natABCDEG* (C) and their putative structures are also shown (D). Compounds **24** and **25** cannot be clearly seen here because of their low abundance but can be detected by LC-HRMS.

5.2.4 Investigation into the substrate specificity of the neoantimycin PKS

Although bioinformatics analysis has shown high sequence similarity among the AT domains of all four PKSs,[129] AntD-AT incorporates alkyl extender units into the macrocycle while the AT domains of the other three PKSs likely select malonate or methylmalonate, which then undergoes methylation by the MT domain. To distinguish between the use of malonate and methylmalonate by NatC via the number of methylations catalyzed by its MT domain, we fed [methyl-d_3] L-methionine to

neoantimycin-producing *S. albus* J1074 Δ*antC* cultures, since methionine is the precursor of the predicted methyl group donor, *S*-adenosylmethionine (SAM). Malonyl-CoA loading followed by dimethylation would result in an increase in the *M*+6 peak whereas methylmalonyl-CoA loading followed by a single methylation would only result in an increase in the *M*+3 peak. The results of this feeding experiment showed a significant increase in the *M*+6 peak for both **20** and **23**, suggesting that NatC uses malonate as a substrate and its MT domain catalyzes dimethylation (Figure 5-7). Although a large increase in the *M*+3 peak for **20** and **23** was also observed, this can potentially be attributed to dimethylation using one d_3-labeled SAM molecule and one unlabeled SAM molecule.

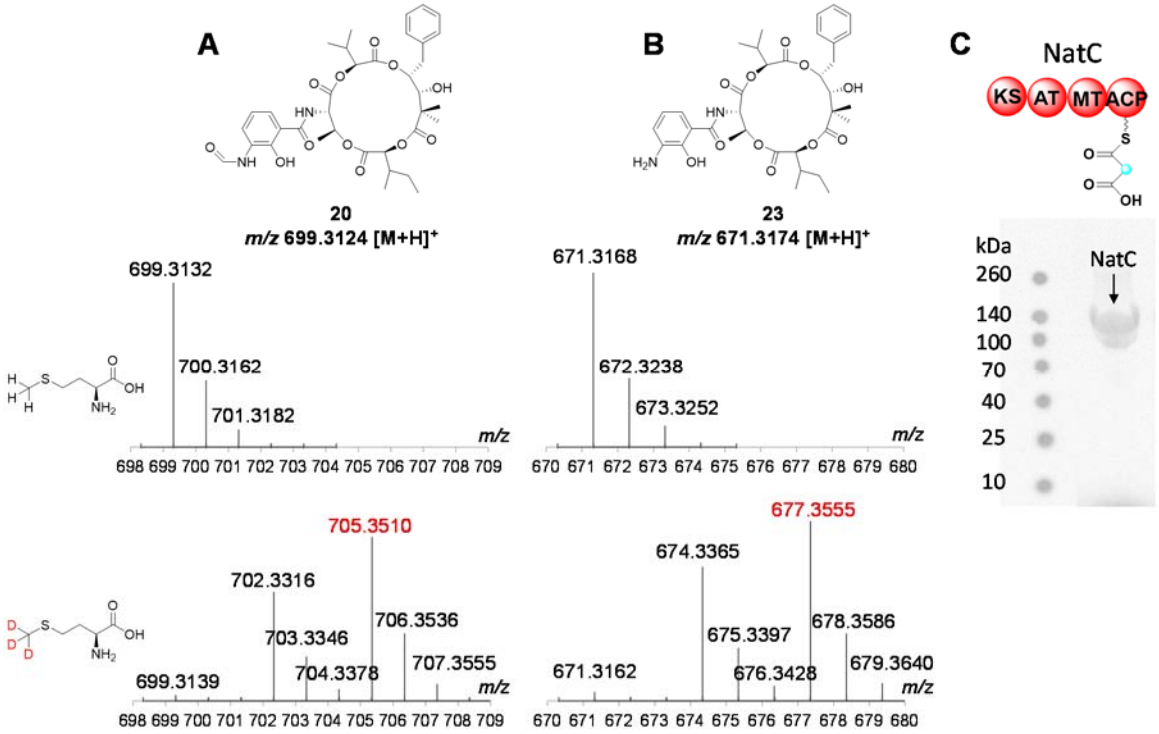


Figure 5-7. Evidence supporting the use of malonate by NatC. (A) Isotopic peak patterns of **20** and **23** (B) from neoantimycin-producing *S. albus* J1074 Δ*antC* cultures fed with unlabeled L-methionine (row 1) and [methyl-d_3] l-methionine. (C) ^{14}C gel autoradiography assay showing the labeling of NatC by [2-^{14}C]malonyl-CoA.

To further confirm this finding, we also purified NatC from *E. coli* BAP1 (Figure 5-8) and performed a ^{14}C gel autoradiography assay to determine if [2-^{14}C]malonyl-CoA could be loaded onto the PKS. The results of this assay showed successful transfer of the radiolabel to NatC, indicating that the NatC-AT domain is functional and capable of recognizing malonyl-CoA (Figure 5-7C).

5.2.5 Attempted *in vitro* reconstitution of neoantimycin

We next attempted to reconstitute the biosynthesis of the neoantimycin scaffold *in vitro*. Toward this end, we overexpressed and purified all of the neoantimycin biosynthetic enzymes (NatBCDFEFG) from *E. coli* (Figure 5-8). For the purification of the three-module NRPS, NatB, the megasynthetase was expressed as two separate

proteins with the first half encoding the first two modules of NatB and the second half encoding the third module. Additionally, NatD was purified from an *E. coli* strain coexpressing NatE, since NatE may be involved in A domain activation. As all of the proteins appeared to be soluble to some extent, we incubated NatB, C, D, E, and F along with AntF, AntG, and their corresponding substrates and cofactors (Table 5-1). NatG was intentionally excluded since it was shown to be unnecessary for neoantimycin production *in vivo*, and its putative function as an editing thioesterase might preclude product formation using anthranilate as the starter unit. Unfortunately, no potential neoantimycin products could be detected from this assay, which may be due to low titers resulting from an impaired assembly line using a modified starter unit. Alternatively, the low solubility and decreased activity of NatB resulting from its dissection may be the issue, and efforts to better purify the first two modules of NatB and to assess the success of substrate transfer from the first two modules to the third module are underway.

Protein	Size (kDa)
NatB (Mod 1&2)	296
NatB (Mod 3)	196
NatC	156
NatD	253
NatE	12
NatF	41
NatG	33

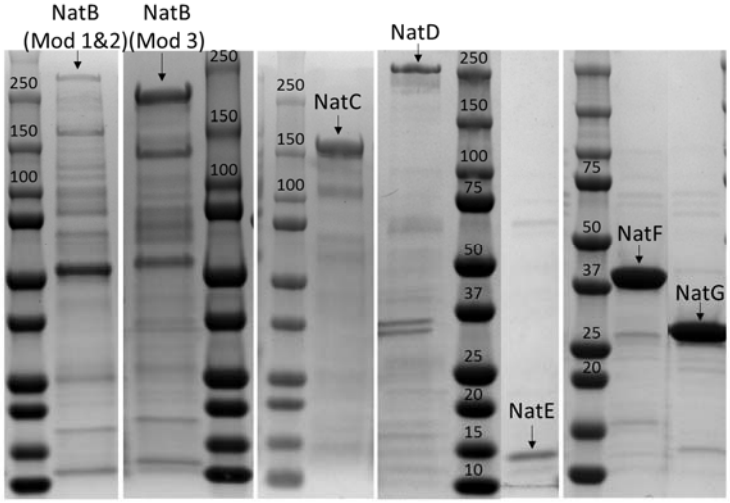


Figure 5-8. Neoantimycin biosynthetic enzymes purified from *E. coli*. The first two modules of NatB appear to have some degradation.

Table 5-1. Assay components for neoantimycin reconstitution.

Enzymes	Substrates	Cofactors
AntF	anthranilate	ATP, $MgCl_2$
AntG		
NatB-Mod 1&2	L-threonine, 2-oxo-3-methylbutanoic acid	ATP, $MgCl_2$, NADPH
NatB-Mod 3	phenylpyruvic acid	ATP, $MgCl_2$, NADPH
NatC	malonyl-CoA/methylmalonyl-CoA	SAM
NatD	2-oxo-3-methylpentanoic acid	ATP, $MgCl_2$, NADPH
NatE		
NatF		NADPH

5.2.6 Forays into modified antimycin-type depsipeptide assembly lines

As we are ultimately interested in using the set of naturally combinatorial antimycin-type depsipeptide assembly lines to better understand the underlying programming rules of NRPS-PKS assembly lines, we have also performed some preliminary studies generating mixed assembly lines composed of both antimycin and neoantimycin biosynthetic machinery. In particular, using CRISPR/Cas, we developed a mutant *S. albus* J1074 Δ*antC*::*natB* strain to determine if any antimycin-type depsipeptides with a 12-membered macrocyclic ring could be produced by this modified assembly line. Although we were unable to detect the presence of such compounds in the culture extracts of this strain, this result is not unexpected, and some of the possible reasons for it include: 1) poor interactions between NatB and AntD due to the lack of compatible docking domains; 2) poor recognition of the non-native substrate by AntD; and 3) inability of the AntD-TE to catalyze the cyclization of antimycin-type depsipeptides with larger macrolactone rings. Further work that can be done to evaluate and address these possible issues is discussed in more detail in the next section

Furthermore, we also attempted to test the *in vitro* activity of the NatD-TE in the context of the antimycin assembly line. Enzymatic synthesis of an antimycin analogue was performed using AntC, D, E, F, G, and M as described previously,[51] and the same compound was produced when using a dissected AntD module, albeit at a titer of nearly two-fold less (Figure 5-9). However, the use of NatD-TE in place of AntD-TE did not result in any detectable antimycin production (Figure 5-9C), suggesting that the TEs of these assembly lines may not be very promiscuous. The reconstitution of the other assembly lines followed by the testing of all four of the TEs can help to further validate this hypothesis, and this work is ongoing.

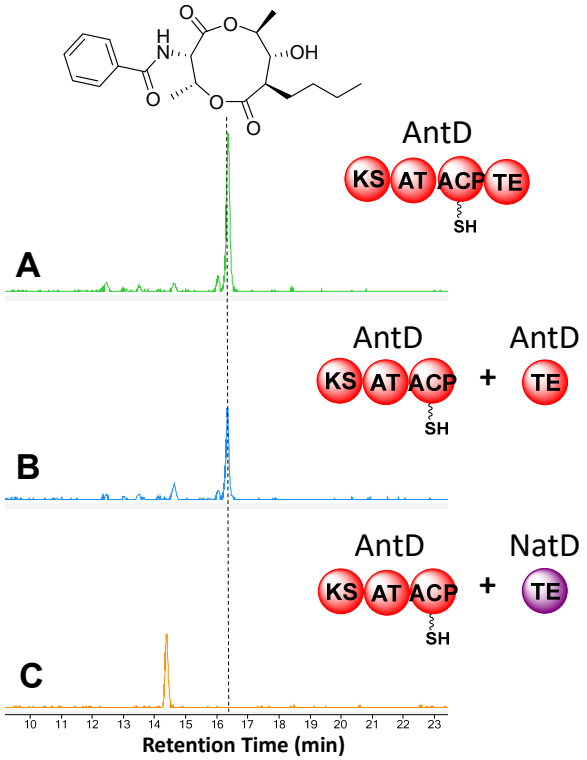


Figure 5-9. *In vitro* reconstitution of an antimycin analogue using different TE domains. Extracted ion chromatograms showing the production of an antimycin analogue with a

benzoate starter unit when using AntD as a single (A) and dissected (B) module but not when using NatD-TE in place of AntD-TE (C). The calculated mass with a 10 ppm mass error tolerance was used.

5.3 Discussion and Future Work

Although neoantimycins are much less well-studied than the antimycins, these compounds are gaining increasing interest because of their promising anticancer properties. Moreover, the unique relationship between the neoantimycin and antimycin biosynthetic gene clusters provides an intriguing system for the study of how these NRPS-PKS assembly lines have evolved to be naturally combinatorial. To advance these studies, we have made progress toward the reconstitution of neoantimycin biosynthesis and have successfully produced neoantimycins in a more convenient heterologous host (Figure 5-4). Our expression of neoantimycins in *S. albus* J1074 not only enabled us to dissect the functions of the two freestanding domains whose roles were previously unknown, but it also allowed us to probe the substrate specificity of the neoantimycin PKS. More recently, we have also developed a more streamlined neoantimycin-producing *S. albus* J1074 strain that will enable us to use CRISPR/Cas to rapidly modify the neoantimycin gene cluster, be it through gene deletion, insertion, or replacement.

In parallel, we are also working towards the reconstitution of neoantimycins *in vitro*, which will provide another platform that can be used for the systematic study of the combinatorial nature of antimycin-type depsipeptide enzymatic assembly lines. While most of the neoantimycin biosynthetic enzymes appear to be somewhat soluble when purified from *E. coli* (Figure 5-8), our initial attempts to produce neoantimycins *in vitro* were unsuccessful. Consequently, we plan to biochemically characterize the individual enzymes and particularly, the intact and dissected NRPS modules. The *in vitro* activities and substrate specificities of these enzymes can be assessed using ATP-[^{32}P]PP$_i$ exchange assays as well as ^{14}C gel autoradiography assays, which can enable the visualization of substrate transfer to downstream modules as in the case of determining whether the dissection of NatB has disrupted the intermodular interactions between its second and third modules. Furthermore, the recent advances in mass spectrometry techniques can also facilitate the examination of carrier protein-bound biosynthetic intermediates,[174-176] and allow us to pinpoint which part of the NRPS-PKS assembly line may be impaired.

Upon developing a platform that can be used to study the programming rules of this set of naturally combinatorial NRPS-PKS assembly lines, we plan to use this system to answer two key questions. First, we seek to elucidate the factors that are integral for engendering NRP-PK ring expansion. Successful macrolactone ring expansion requires both efficient module addition and macrolactone cyclization with the modified ring size, and we hypothesize that the former is governed by intermodular protein-protein interactions, while the latter is controlled by the off-loading TE. To evaluate this hypothesis, we aim to engineer modified assembly lines through the sequential addition of neoantimycin NRPS modules to the antimycin assembly line. As our preliminary attempts to do this via the generation of *S. albus* J1074 Δ*antC*::*natB* did not result in production of antimycin-type depsipeptides with a 12-membered macrolactone ring, the protein-protein interactions between NatB and AntD will be more thoroughly investigated. Specifically, the docking domains (DDs) responsible for the natural interactions between AntC and AntD can first be analyzed using structural modeling (to

determine the residues critical for interaction), size exclusion chromatography (to determine if the DDs can result in copurification of their respective attached domains), and isothermal titration calorimetry (to determine the dissociation constants of these complexes).[177] These strategies will then be used to study the association between the C-terminal DD of NatB and the N-terminal DD of AntD, and if necessary, the DD of NatB may be replaced with that of AntC. In addition, substrate transfer among the modules of these modified assembly lines can be assessed using ^{14}C gel autoradiography and mass spectrometry, which will aid us in determining if intermediate transfer to AntD is limited by its substrate specificity. Because the TE may also be involved in controlling macrolactone ring size, the swapping of this domain may further be necessary for the products from these modified assembly lines to be properly off-loaded and cyclized. The TEs encoded in the four antimycin-type depsipeptide gene clusters have sequence similarities ranging from 50-60%, and testing all of them on the engineered assembly lines will go toward determining whether the TE from each cluster is specific for one ring size. As our preliminary *in vitro* assays showed some loss of activity upon separating the TE from the rest of the PKS module (Figure 5-9), these TE swapping experiments may be performed in *cis* though the solubility of the resulting hybrid modules will need to be examined.

A second area of investigation that can be pursued using the set of naturally combinatorial NRPS-PKS assembly lines is the determination of the rules that govern atypical PKS building block selection. Although the NRPS-PKS hybrid assembly lines for all antimycin-type depsipeptides have a PKS module, only the 9-membered antimycin assembly line incorporates alkyl-chain monomers while the others use the more commonly used malonate or methylmalonate substrates. As a result, we are interested in elucidating the factors responsible for this difference, and notably, the AT domains from these gene clusters have highly conserved active site residues and sequence similarities ranging from 60-70%. Toward this goal, we can use the reconstituted antimycin and neoantimycin assembly lines to systematically study the effects of AT domain modification, including the mutation of residues implicated in substrate selection, the swapping of the entire AT domain, and the swapping of the KS-AT didomain.[178, 179] In performing this work, we can also shed light on the critical intramodular protein-protein interactions between the AT and its adjacent domains and evaluate whether the substrate specificity of the KS domain may lead to the impaired function of modified assembly lines. Ultimately, the knowledge garnered from the combinatorial studies of these related NRPS-PKS assembly lines may provide us with insight that can guide future efforts to perform combinatorial biosynthesis in other NRPS and PKS systems.

5.4 Materials and Methods

Construction of S. albus J1074 mutants

The *S. albus* J1074 Δ*antC* and *S. albus* J1074 Δ*antC*::*natB* mutant strains were constructed using the previously reported CRISPR/Cas system engineered for *Streptomyces*.[42] Plasmids encoding Cas9, two synthetic guide RNAs targeting *antC*, and a 4 kb editing template consisting of two 2 kb arms homologous to the sequence up and downstream of *antC* with and without *natB* between them, are listed in Table 5-2 and were cloned as previously described.[42]. The resulting pCRISPomyces2-dantCgRNA-2kb and pCRISPomyces2-dantCgRNA-2kb-natB plasmids were transformed into *E. coli*

WM6026 for conjugation with *S. albus* J1074. Successful transformants were selected by apramycin resistance, and deletion of Δ*antC* was confirmed by PCR of the extracted genomic DNA with the dantC-F2/R2 primers (Table 5-3). Similarly, the replacement of Δ*antC* by *natB* was confirmed by PCR with the dantC-F2/dantC-B-R2 primers. Clearance of the plasmid was performed as described by previously[42] with the exception that the strains were cultured in TSB medium. Further culturing and product analysis of these strains is described below.

Table 5-2. Plasmids used in this study.

Plasmid	**Derived from**	**Function**
pCRISPomyces2-dantCgRNA-2kb	pCRISPomyces-2	Construction of *S. albus* J1074 Δ*antC* mutant
pCRISPomyces2-dantCgRNA-2kb-*natB*	pCRISPomyces-2	Construction of *S. albus* J1074 Δ*antC*::*natB* mutant
pET30-*natB*-mod1&2	pET30	Purification of NatB (Mod 1&2) from *E. coli*
pET30-natB-mod3	pET30	Purification of NatB (Mod 3) from *E. coli*
pET30-*natC*	pET30	Purification of NatC from *E. coli*
pET24b-*natD*	pET24	Purification of NatD from *E. coli*
pCDF-*natE*	pCDFDuet-1	Purification of NatD from *E. coli*
pET30-*natE*	pET30	Purification of NatE from *E. coli*
pET30-*natF*	pET30	Purification of NatF from *E. coli*
pET30-*natG*	pET30	Purification of NatG from *E. coli*
pET30-*antD*-KS_AT_ACP	pET30	Purification of AntD-KS_AT_ACP from *E. coli*
pET30-*antD*-TE	pET30	Purification of AntD-TE from *E. coli*
pET30-*natD*-TE	pET30	Purification of NatD-TE from *E. coli*
cos813-pIB139-*natFG*	SuperCos1/pIB139	Heterologous expression of neoantimycins in *S. albus* J1074
cos813-pIB139-*natF*	SuperCos1/pIB139	Heterologous expression of neoantimycins in *S. albus* J1074
cos813-pIB139-*natG*	SuperCos1/pIB139	Heterologous expression of neoantimycins in *S. albus* J1074

Table 5-3. Primers used in this study.

Primer	**Sequence (5' -> 3')**	**Description**
supercos1-seq-F2	GCCACCTGACGTCTAAGAAA	initial sequencing primers for primer walking
supercos1-seq-R	GAATGAACAATGGAAGTCAA	
antC-up-R	cgtcggtccttcctctgctg ccgacgatcgtcctcgttgc	Cloning of pCRISPomyces 2-dantCgRNA-2kb
antC-down-F	gcaacgaggacgatcgtcgg cagcagaggaaggaccgacg	
antC-up-F	tcggttgccgccgggcgtttttttatctaga caggtcgttcatcgcctgct	Cloning of

antC-down-R	gcggccttttacggttcctggcctctaga agcgggaggacggtgtcgac	pCRISPomyces 2-dantCgRNA-2kb-(*natB*)
antC-natB-up-R	agtgcgccccggatttccat gacaccaaccctcggttgcg	Cloning of pCRISPomyces 2-dantCgRNA-2kb-*natB*
antC-natB-F	cgcaaccgagggttggtgtc atggaaatccggggcgcact	
antC-natB-R	cgtcggtccttcctctgctg TCAGCCATGGTGAGAGGTGT	
antC-natB-down-F	ACACCTCTCACCATGGCTGA cagcagaggaaggaccgacg	
natB-R2	cctcgacctcctccaccacg	Cloning of pCRISPomyces 2-dantCgRNA-2kb-*natB* and expression of NatB-Mod 1&2
natB-F3	cagccatgagatagaggccg	
natB-R4	gtacggggcagggtggagac	
natB-F5	caggccgtactggtctccac	
dantC-F2	gttcgtggcgcgagtccacc	Confirmation of the *S. albus* J1074 Δ*antC* and Δ*antC*::*natB* mutants
dantC-R2	ctgaccggccgtacggactc	
dantC-B-R2	cagaactcccggtcgcgctc	
pIB-natF-F	GTTGGTAGGATCCA CATATG AGAAGGGAG CGGACATAC atgaaactcctgatcatcgg	Cloning of *natF/G/FG* into pIB139
natF-int-F	cctgaccagacaggccgaactg	
natF-int-R	cttccaggacgctcagttcg	
pIB-natG-R	gattacgaattcgatatcgc tcagggcagccgggcgccgg	
pIB-natF-R	gattacgaattcgatatcgc tcactgcggctgctttccct	
pIB-natG-F	GTTGGTAGGATCCA CATATG AGAAGGGAG CGGACATAC gtgagcaccaccgacctgat	
pSET152L-apr-cos-F	CTCCCCAGCAGGCAGAAGTATGCAAAGCA TGCAT agatccttttggttcatgtg	Cloning of pIB139-*natF*/*G*/*FG* into cos813
pIB-cos-R	GGCGGGACTATGGTTGCTGACTAATTGAGA TGCAT gattacgaattcgatatcgc	
natE-Duet-F	tat**AGATCT**cgtgctggaccggcacccg	Confirmation of uptake of the heterologous expression vectors
natB-F1b	atcaccatcatcaccacagccaggatcc gaattc t atgtctgttcacgaggccgc	Expression of NatB-Mod 1&2
natB-R3	cgtccagcagcacacaggtg	
natB-F4	acatggtgcccagcacctgt	
natB-R5b	tttctgttcgacttaagcattat gcggccgc aagctta ccggacggcgggcaccagcg	
natC-pET30-F	GGT ATT GAG GGT CGC atggctgagcccaccgccca	Expression of

natC-pET30-R	AGA GGA GAG TTA GAG CC tcagccgcgccgggcgctgc	NatC
natDE-pET24-F	gtttaactttaagaaggagatataCATATG CCAACCCCCGTAGGCCG	Expression of NatD
natDE-pET24-R	atctcagtggtggtggtggtggtgCTCGAG cgcgggggttccttccaggg	
Duet-Mbth-F	AAA CATAtgacatccaccagtcccttc	Coexpression of NatE with NatD
Duet-Mbth-R	TAT AGATCTcatgccacggcctccggg	
mbth-pET30-F	GGT ATT GAG GGT CGC gtgacatccaccagtcccttc	Expression of NatE
mbth-pET30-R	AGA GGA GAG TTA GAG CC tcatgccacggcctccggg	
natF-pET30-F	GGT ATT GAG GGT CGC atgaaactcctgatcatcgg	Expression of NatF
natF-pET30-R	AGA GGA GAG TTA GAG CC tcactgcggctgctttccct	
natG-pET30-F	GGT ATT GAG GGT CGC gtgagcaccaccgacctgat	Expression of NatG
natG-pET30-R	AGA GGA GAG TTA GAG CC tcagggcagccgggcgcc	
dantC-gRNA construct	gagacatctt**gaagac**aaacgcCCCTCCTGTGCCCCG GAAAGgttttagagctagaaatagcaagttaaaataaggctagtcc gttatcaacttgaaaaagtggcaccgagtcggtgctttttttagcataacc ccttggggcctctaaacgggtcttgaggggttttttggctgctccttcggt cggacgtgcgtctacgggcaccttaccgcagccgtcggctgtgcgac acggacggatcgggcgaactggccgatgctgggagaagcgcgctg ctgtacggcgcgcaccgggtgcggagcccctcggcgagcggtgtga aacttctgtgaatggcctgttcggttgctttttttatacggctgccagataa ggcttgcagcatctgggcggctaccgctatgatcggggcgttcctgca attcttagtgcgagtatctgaaaggggatacgcTCGGTCCTTC CTCTGCTGCGgtttaa**gtcttc**tttcacgtggc	

*restriction endonuclease sites are bolded and spacer sequences are underlined

Heterologous expression of neoantimycins in wild-type and mutant S. albus J1074 strains

S. orinoci genomic DNA was used to PCR amplify *natF*/*G*/*FG*, and the genes were cloned into pIB139[180] using Gibson assembly. The resulting plasmid was then amplified by PCR and further cloned into the NsiI/CIP-digested cosmid encoding *natA-E* (cos813) using Gibson assembly. Subsequently, the constructs were transformed into *E. coli* WM6026 for conjugation with wild-type or mutant *S. albus* J1074 strains (Δ*antC*, Δ*antO*). Successful transformants were selected by apramycin and kanamycin resistance and confirmed by PCR using the natE-Duet-F/Duet-Mbth-R primers.

For the production of neoantimycins by *S. albus* J1074, 3 mL starter cultures in TSB medium with apramycin (100 μg/mL) and kanamycin (100 μg/mL) were inoculated with mycelia and grown at 30ºC, 250 rpm for 48 h. 0.5 mL of the starter culture was then inoculated into 25 mL of mannitol-soy flour (MS) medium (20 g/L soy flour and 20 g/L mannitol) with the appropriate antibiotics, and the cultures were grown at 30ºC, 250 rpm for 5 days.

LC-MS analysis of neoantimycin production

Neoantimycins were extracted from the cell-free supernatant (25 mL) using ethyl acetate (two volumes). The solvent was removed by rotary evaporation and the residue was redissolved in methanol (0.5 mL) and analyzed by LC-HRMS (10 μL injection) and LC-UV-MS (20 μL injection). LC-HRMS analysis was performed on an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS instrument, and LC-UV-MS analysis was performed on an Agilent Technologies 6120 Quadrupole LC-MS (with DAD) instrument with an Agilent Eclipse Plus C18 column (4.6 × 100 mm). A linear gradient of of 25-95% CH_3CN (v/v) over 20 min followed by an additional 10 min of 95% CH_3CN (v/v) in H_2O with 0.1% formic acid (v/v) at a flow rate of 0.5 mL/min was used.

Isotope-labeled precursor feeding experiments

S. albus J1074 Δ*antC*/cos813-pIB139-*natFG* was cultured for the production of neoantimycins as described above, and 24 h after inoculation of the seed culture into 25 mL of MS with apramycin (100 μg/mL) and kanamycin (100 μg/mL), either unlabeled L-methionine or [methyl-d_3] L-methionine was added to cultures to a final concentration of 1 g/L. Compound extraction and LC-HRMS analysis was performed as described above.

Cloning, overexpression, and purification of proteins

Neoantimycin biosynthetic genes (*natBCDEFG*) were PCR amplified from *S. orinoci* genomic DNA, and *antD* was amplified from *S. albus* J1074 genomic DNA. The genes were cloned into either pET30 (for the introduction of an *N*-terminal His_6-tag) or pET24b (for the introduction of a *C*-terminal His_6-tag) using Gibson assembly, and *natE* was additionally cloned into pCDFDuet-1 (Novagen) using restriction digestion with NdeI/BglII (Thermo Scientific) followed by ligation with Quick T4 DNA ligase (New England Biolabs). The plasmids: pET30-*natB*-mod 1&2, pET30-*natB*-mod 3, pET30-*natC*, pET30-*natE*, pET30-*natF*, pET30-*natG*, pET30-*antD*-KS_AT_ACP, pET30-*antD*-TE, and pET30-*natD*-TE were then individually transformed into *E. coli* BAP1 for protein expression. For the expression of NatD, pET24b-*natD* and pCDF-*natE* were cotransformed into *E. coli* BAP1. Expression and purification for all proteins with a His_6-tag followed the same general procedure and is detailed as follows: cells were grown at 37°C in 1 L of LB with the appropriate antibiotic(s) to an OD_{600} of 0.5. The cells were then cooled on ice for 10 min and induced with 0.12 mM IPTG for 12-16 h at 16°C. Subsequently, cells were harvested by centrifugation (6371 × *g*, 15 min, 4°C), resuspended in 30 mL lysis buffer (25 mM HEPES, pH 8, 0.5 M NaCl, 5 mM imidazole), and lysed by homogenization on ice. Cellular debris was removed by centrifugation (27216 × *g*, 1 h, 4°C). Ni-NTA agarose resin was added to the supernatant (3 mL/L of culture), and the solution was nutated at 4°C for 1 h. The protein-resin mixture was loaded onto a gravity flow column, and proteins were eluted with increasing concentrations of imidazole in Buffer A (20 mM HEPES, pH 8.0, 100 mM NaCl). Purified proteins were concentrated and buffer exchanged into Buffer A + 10% glycerol using Vivaspin centrifugal concentrators. AntD, AntF, and AntG were also expressed and purified as described previously.[51]

^{14}C-labeling of NatC

Assays were performed in 10 μL of 50 mM HEPES (pH 8.0) containing 1 mM TCEP, 4 mM ATP, 4 mM $MgCl_2$, 1 mM CoA, 0.13 mM [2-^{14}C]malonic acid (0.1 mCi/mL; American Radiolabeled Chemicals), 25 μM MatB (malonyl-CoA synthetase), and 90 μM NatC. Reactions were incubated for 2 h at room temperature and quenched with an equal volume of 1X SDS sample buffer before SDS-PAGE analysis with a 4-15% TGX gel (Criterion). The gel was subsequently dried for 2.5 h at 50°C and then exposed on a storage phosphor screen (20 × 25 cm; Molecular Dynamics) for 2-3 days. Phosphor images were captured using a Typhoon 9400 phosphorimager (Storage Phosphor mode, best resolution, 50 μm resolution; Amersham Biosciences).

In vitro reconstitution of neoantimycins and antimycins

For the attempted *in vitro* reconstitution of neoantimycins, assays were performed in 100 μL of 50 mM HEPES (pH 8.0) containing 1 mM TCEP, 4 mM ATP, 4 mM $MgCl_2$, 4 mM NADPH, 4 mM anthranilate, 4 mM L-threonine, 4 mM 2-oxo-3-methylbutanoic acid, 4 mM phenylpyruvate, 4 mM 2-oxo-3-methylpentanoic acid, 2mM SAM, 2 mM malonyl-CoA or methylmalonyl-CoA, and 5-10 μM AntF, AntG, NatB (Mod 1&2), NatB (Mod 3), NatC, NatD, NatE, and NatF. After overnight incubation at room temperature, the reactions were extracted with two volumes of ethyl acetate, and the organic layer was dried under vacuum. The residue was then redissolved in methanol (50 μL) for LC-HRMS analysis as described above.

For the *in vitro* reconstitution of antimycins, assays were performed in 100 μL of 50 mM HEPES (pH 8.0) containing 1 mM TCEP, 4 mM ATP, 4 mM $MgCl_2$, 4 mM NADPH, 33 mM $NaHCO_3$, 4 mM anthranilate, 4 mM L-threonine, 4 mM 2-oxo-3-methylbutanoic acid, 2 mM CoA, 4 mM (2E)-hexenoic acid, 4 mM benzoic acid, 4 mM L-threonine, 4 mM pyruvate, and 10-20 μM Orf35 (acyl-CoA ligase[181]), AntC, AntD (or AntD-KS_AT_ACP + AntD-TE/NatD-TE), AntE, AntF, AntG, and AntM. After overnight incubation at room temperature, the reactions were extracted with two volumes of ethyl acetate, and the organic layer was dried under vacuum. The residue was then redissolved in methanol (50 μL) for LC-HRMS analysis as described above.

Chapter 6. Conclusion

The work described here shows how the biosynthetic study of medicinally active natural products can reveal new enzymatic logic and expand our ability to generate novel natural product analogues. Importantly, the characterization of the unusual enzymes identified from such studies also has the potential to facilitate the design of *de novo* biosynthetic routes to other target natural and unnatural compounds that may be otherwise difficult to obtain. Our studies to elucidate the biosynthesis of the unique pyrroloindole scaffold of physostigmine, the terminal epoxyketone pharmacophore of peptidyl epoxyketone proteasome inhibitors, and the 3-formamidosalicylate pharmacophore of antimycin-type depsipeptides can likewise enable the introduction of these biologically relevant moieties into new molecules though the success of this largely depends on the substrate specificity of the biosynthetic enzymes.

Along these lines, achieving efficient combinatorial biosynthesis, or the manipulation of biosynthetic pathways to produce modified compounds, remains a longstanding aim for those studying natural product biosynthesis. This is particularly true for NRP and PK natural products such as the peptidyl epoxyketone proteasome inhibitors and the antimycin-type depsipeptides discussed here. Due to the assembly line-like mechanism of NRPSs and PKSs as well as the diversity of their building blocks, these pathways are believed to be ideal systems for performing large-scale combinatorial biosynthesis. However, despite the vast potential, the success of engineered NRPS and PKS assembly lines has been limited, and modified assembly lines often show impaired activity. Consequently, further advancement of NRP and PK combinatorial biosynthesis will rely on investigations into the fundamental principles and critical protein-protein interactions that dictate successful assembly line alteration. Rather than studying the extent to which a single NRPS or PKS assembly line can be manipulated, our approach of using a set of naturally combinatorial NRPS-PKS assembly lines may lead to new insights. This ongoing work along with the recent advances in the structural understanding of these biosynthetic megaenzymes can ultimately improve our strategies for engineering NRPS and PKS systems to generate “unnatural” natural products.

Looking to the future, an improved understanding of the NRPS and PKS enzymatic machinery is also critical for the successful engineering of these assembly lines, and biochemical, in particular quantitative and mechanistic, analyses will be important for revealing the underlying programming rules of these enzymatic assembly lines. Specifically, *in vitro* biochemical analysis using purified enzymes can aid in understanding the precise function and substrate specificity of catalytic domains, reaction mechanisms, and the internal kinetics of the catalytic program. Despite improvements in mass spectrometry techniques, directly detecting and quantifying the majority of ACP-bound biosynthetic intermediates remains challenging, and the development of alternative strategies such as those that couple PKS-catalyzed reactions with fluorescent click chemistry may be an area for future investigation. In addition to biochemical assays, structural analysis will also continue to play a vital role in studying megaenzymes and pave the way for a more detailed understanding of their mechanism and dynamics. Accordingly, the continued development of strategies such as small-angle X-ray scattering and cryo-electron microscopy will be important for addressing the technical challenges in obtaining structural information for large, dynamic protein complexes.

With an improved biochemical and biophysical understanding of PKS/NRPS enzymes, the generation of both specific and libraries of "unnatural" natural products may also be more attainable in the future. While PKS-AT domain engineering is becoming more feasible with the identification of some of the key residues involved in dictating the substrate specificity of AT domains and the increasing accessibility of unusual extender units through the use of crotonyl-CoA reductase enzymes, achieving reasonable titers of new compounds is a key challenge that will need to be overcome. Consequently, understanding how the KS may represent another level of substrate selection and how protein-protein interactions between the KS and upstream ACP may affect assembly line turnover are important areas for future investigations. Furthermore, much remains to be understood about the interactions between NRPSs and PKSs in hybrid assembly line systems, as no protein interaction that docks a NRPS to a PKS has been characterized to date though the ability to effectively engineer these hybrid assembly lines would significantly expand the possible structural diversity. Studies in these areas will likely rely on a combination of different methods, including bioinformatics analyses of evolutionarily-related assembly lines, structural analyses of megaenzymes, and biochemical analyses of enzymatic domains. The development of *in vivo*-based approaches in which new functional chimeric assembly lines are generated using recombination or directed evolution may also facilitate these studies.

Through the work reported here, we also demonstrate how heterologous expression can be a powerful tool for studying and engineering biosynthesis, especially when the natural producers are difficult to cultivate or genetically intractable. The expression of biosynthetic gene clusters in more versatile and robust heterologous hosts can not only enable the rapid identification of the genes that are necessary for biosynthesis but it can also enable the elucidation of the roles of biosynthetic enzymes with unclear functions, and this is evident from our studies on eponemycin, antimycin, and neoantimycin biosynthesis. In addition, conducting labeled precursor feeding experiments to gain insight into biosynthetic mechanism may also be easier to do in heterologous hosts whose metabolism is better understood. Finally, as in the case of antimycins, heterologous expression can facilitate the production of new natural product analogues derived from combinatorial biosynthesis. Such efforts can further benefit from the advent of new genome editing tools and cheaper gene synthesis technologies, which enable the expeditious engineering of heterologous hosts and biosynthetic pathways.

References

1. Newman, D.J. & Cragg, G.M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **75**, 311-335 (2012).
2. Rosen, J., Gottfries, J., Muresan, S., Backlund, A. & Oprea, T.I. Novel chemical space exploration via natural products. *J. Med. Chem.* **52**, 1953-1962 (2009).
3. Demain, A.L. & Fang, A. The natural functions of secondary metabolites. *Adv. Biochem. Eng. Biotechnol.* **69**, 1-39 (2000).
4. Floss, H.G. Combinatorial biosynthesis - potential and problems. *J. Biotechnol.* **124**, 242-257 (2006).
5. Dewick, P.M. Medicinal Natural Products: A Biosynthetic Approach, Edn. 3rd. (John Wiley & Sons Ltd, West Sussex, United Kingdom; 2009).
6. Peterson, E.A. & Overman, L.E. Contiguous stereogenic quaternary carbons: A daunting challenge in natural products synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11943-11948 (2004).
7. Pickens, L.B., Tang, Y. & Chooi, Y.H. Metabolic engineering for the production of natural products. *Annu. Rev. Chem. Biomol. Eng.* **2**, 211-236 (2011).
8. Kennedy, J. & Turner, G. δ-(L-α-aminoadipyl)- L-cysteinyl-D-valine synthetase is a rate limiting enzyme for penicillin production in *Aspergillus nidulans*. *Mol. Gen. Genet.* **253**, 189-197 (1996).
9. Malla, S., Niraula, N.P., Liou, K. & Sohng, J.Y. Enhancement of doxorubicin production by expression of structural sugar biosynthesis and glycosyltransferase genes in *Streptomyces peucetius*. *J. Biosci. Bioeng.* **108**, 92-98 (2009).
10. Madduri, K., Kennedy, J., Rivola, G., Inventi-Solari, A., Filippini, S., Zanuso, G., Colombo, A.L., Gewain, K.M., Occi, J.L., MacNeil, D.J. & Hutchinson, C.R. Production of the antitumor drug epirubicin (4'-epidoxorubicin) and its precursor by a genetically engineered strain of *Streptomyces peucetius*. *Nat. Biotechnol.* **16**, 69-74 (1998).
11. Kim, H.J., Ruszczycky, M.W., Choi, S.H., Liu, Y.N. & Liu, H.W. Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A. *Nature* **473**, 109-112 (2011).
12. Liu, P.H., Murakami, K., Seki, T., He, X.M., Yeung, S.M., Kuzuyama, T., Seto, H. & Liu, H.W. Protein purification and function assignment of the epoxidase catalyzing the formation of fosfomycin. *J. Am. Chem. Soc.* **123**, 4619-4620 (2001).
13. Chang, W.C., Dey, M., Liu, P.H., Mansoorabadi, S.O., Moon, S.J., Zhao, Z.B.K., Drennan, C.L. & Liu, H.W. Mechanistic studies of an unprecedented enzyme-catalysed 1,2-phosphono-migration reaction. *Nature* **496**, 114-118 (2013).
14. Watanabe, K., Hotta, K., Praseuth, A.P., Koketsu, K., Migita, A., Boddy, C.N., Wang, C.C.C., Oguri, H. & Oikawa, H. Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*. *Nat. Chem. Biol.* **2**, 423-428 (2006).
15. Ro, D.K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., Chang, M.C.Y., Withers, S.T., Shiba, Y., Sarpong, R. & Keasling, J.D. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-943 (2006).
16. Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., Polichuk, D.R., Teoh, K.H., Reed, D.W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D.,

Dorin, G., Ellens, K.W., Fickes, S., Galazzo, J., Gaucher, S.P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L.F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P.S., Keasling, J.D., Reiling, K.K., Renninger, N.S. & Newman, J.D. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528-532 (2013).
17. Paddon, C.J. & Keasling, J.D. Semi-synthetic artemisinin: A model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.* **12**, 355-367 (2014).
18. Eustaquio, A.S., Gust, B., Li, S.M., Pelzer, S., Wohlleben, W., Chater, K.F. & Heide, L. Production of 8'-halogenated and 8'-unsubstituted novobiocin derivatives in genetically engineered *Streptomyces coelicolor* strains. *Chem. Biol.* **11**, 1561-1572 (2004).
19. Eustaquio, A.S., Gust, B., Galm, U., Li, S.M., Chater, K.F. & Heide, L. Heterologous expression of novobiocin and clorobiocin biosynthetic gene clusters. *Appl. Environ. Microbiol.* **71**, 2452-2459 (2005).
20. Wolpert, M., Heide, L., Kammerer, B. & Gust, B. Assembly and heterologous expression of the coumermycin A_1 gene cluster and production of new derivatives by genetic engineering. *Chembiochem* **9**, 603-612 (2008).
21. Zhang, H.R., Wang, Y., Wu, J.Q., Skalina, K. & Pfeifer, B.A. Complete biosynthesis of erythromycin A and designed analogs using *E. coli* as a heterologous host. *Chem. Biol.* **17**, 1232-1240 (2010).
22. Zhang, G., Li, Y., Fang, L. & Pfeifer, B.A. Tailoring pathway modularity in the biosynthesis of erythromycin analogs heterologously engineered in *E. coli*. *Sci. Adv.* **1**, e1500077 (2015).
23. Minas, W., Brunker, P., Kallio, P.T. & Bailey, J.E. Improved erythromycin production in a genetically engineered industrial strain of *Saccharopolyspora erythraea*. *Biotechnol. Prog.* **14**, 561-566 (1998).
24. Rutledge, P.J. & Challis, G.L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **13**, 509-523 (2015).
25. Yamanaka, K., Reynolds, K.A., Kersten, R.D., Ryan, K.S., Gonzalez, D.J., Nizet, V., Dorrestein, P.C. & Moore, B.S. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1957-1962 (2014).
26. Shao, Z.Y., Rao, G.D., Li, C., Abil, Z., Luo, Y.Z. & Zhao, H.M. Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. *ACS Synth. Biol.* **2**, 662-669 (2013).
27. Ongley, S.E., Bian, X.Y., Neilan, B.A. & Muller, R. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep.* **30**, 1121-1138 (2013).
28. Luo, Y.Z., Li, B.Z., Liu, D., Zhang, L., Chen, Y., Jia, B., Zeng, B.X., Zhao, H.M. & Yuan, Y.J. Engineered biosynthesis of natural products in heterologous hosts. *Chem. Soc. Rev.* **44**, 5265-5290 (2015).
29. Jaitzig, J., Li, J., Sussmuth, R.D. & Neubauer, P. Reconstituted biosynthesis of the nonribosomal macrolactone antibiotic valinomycin in *Escherichia coli*. *ACS Synth. Biol.* **3**, 432-438 (2014).

30. Penn, J., Li, X., Whiting, A., Latif, M., Gibson, T., Silva, C.J., Brian, P., Davies, J., Miao, V., Wrigley, S.K. & Baltz, R.H. Heterologous production of daptomycin in *Streptomyces lividans*. *J. Ind. Microbiol. Biotechnol.* **33**, 121-128 (2006).
31. Liu, H., Jiang, H., Haltli, B., Kulowski, K., Muszynska, E., Feng, X.D., Summers, M., Young, M., Graziani, E., Koehn, F., Carter, G.T. & He, M. Rapid cloning and heterologous expression of the meridamycin biosynthetic gene cluster using a sersatile *Escherichia coli-Streptomyces* artificial chromosome vector, pSBAC. *J. Nat. Prod.* **72**, 389-395 (2009).
32. Shao, Z.Y., Luo, Y.Z. & Zhao, H.M. Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. *Mol. Biosyst.* **7**, 1056-1059 (2011).
33. Li, C.X., Hazzard, C., Florova, G. & Reynolds, K.A. High titer production of tetracenomycins by heterologous expression of the pathway in a *Streptomyces cinnamonensis* industrial monensin producer strain. *Metab. Eng.* **11**, 319-327 (2009).
34. Boghigian, B.A. & Pfeifer, B.A. Current status, strategies, and potential for the metabolic engineering of heterologous polyketides in *Escherichia coli*. *Biotechnol. Lett.* **30**, 1323-1330 (2008).
35. Yuzawa, S., Kim, W., Katz, L. & Keasling, J.D. Heterologous production of polyketides by modular type I polyketide synthases in *Escherichia coli*. *Curr. Opin. Biotech.* **23**, 727-735 (2012).
36. Kim, E., Moore, B.S. & Yoon, Y.J. Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat. Chem. Biol.* **11**, 649-659 (2015).
37. Zhang, M.M., Wang, Y., Ang, E.L. & Zhao, H. Engineering microbial hosts for production of bacterial natural products. *Nat. Prod. Rep.* (2016).
38. Gomez-Escribano, J.P. & Bibb, M.J. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **4**, 207-215 (2011).
39. Komatsu, M., Uchiyama, T., Omura, S., Cane, D.E. & Ikeda, H. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2646-2651 (2010).
40. Zhou, M., Jing, X.Y., Xie, P.F., Chen, W.H., Wang, T., Xia, H.Y. & Qin, Z.J. Sequential deletion of all the polyketide synthase and nonribosomal peptide synthetase biosynthetic gene clusters and a 900-kb subtelomeric sequence of the linear chromosome of *Streptomyces coelicolor*. *FEMS Microbiol. Lett.* **333**, 169-179 (2012).
41. Luo, Y.Z., Zhang, L., Barton, K.W. & Zhao, H.M. Systematic identification of a panel of strong constitutive promoters from *Streptomyces albus*. *ACS Synth. Biol.* **4**, 1001-1010 (2015).
42. Cobb, R.E., Wang, Y.J. & Zhao, H.M. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* **4**, 723-728 (2015).
43. Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790-1792 (2001).
44. Betancor, L., Fernandez, M.J., Weissman, K.J. & Leadlay, P.F. Improved catalytic activity of a purified multienzyme from a modular polyketide synthase after

coexpression with *Streptomyces* chaperonins in *Escherichia coli*. *Chembiochem* **9**, 2962-2966 (2008).
45. Hur, G.H., Vickery, C.R. & Burkart, M.D. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat. Prod. Rep.* **29**, 1074-1098 (2012).
46. Khosla, C., Herschlag, D., Cane, D.E. & Walsh, C.T. Assembly line polyketide synthases: Mechanistic insights and unsolved problems. *Biochemistry* **53**, 2875-2883 (2014).
47. Winn, M., Fyans, J.K., Zhuo, Y. & Micklefield, J. Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* **33**, 317-347 (2016).
48. Fischbach, M.A. & Walsh, C.T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468-3496 (2006).
49. Walsh, C.T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat. Prod. Rep.* **33**, 127-135 (2016).
50. Mo, S., Kim, D.H., Lee, J.H., Park, J.W., Basnet, D.B., Ban, Y.H., Yoo, Y.J., Chen, S.W., Park, S.R., Choi, E.A., Kim, E., Jin, Y.Y., Lee, S.K., Park, J.Y., Liu, Y.A., Lee, M.O., Lee, K.S., Kim, S.J., Kim, D., Park, B.C., Lee, S.G., Kwon, H.J., Suh, J.W., Moore, B.S., Lim, S.K. & Yoon, Y.J. Biosynthesis of the allylmalonyl-CoA extender unit for the FK506 polyketide synthase proceeds through a dedicated polyketide synthase and facilitates the mutasynthesis of analogues. *J. Am. Chem. Soc.* **133**, 976-985 (2011).
51. Sandy, M., Rui, Z., Gallagher, J. & Zhang, W. Enzymatic synthesis of dilactone scaffold of antimycins. *ACS Chem. Biol.* **7**, 1956-1961 (2012).
52. Yan, Y., Chen, J., Zhang, L., Zheng, Q., Han, Y., Zhang, H., Zhang, D., Awakawa, T., Abe, I. & Liu, W. Multiplexing of combinatorial chemistry in antimycin biosynthesis: Expansion of molecular diversity and utility. *Angew. Chem. Int. Ed.* **52**, 12308-12312 (2013).
53. Zhu, X., Liu, J. & Zhang, W. *De novo* biosynthesis of terminal alkyne-labeled natural products. *Nat. Chem. Biol.* **11**, 115-120 (2015).
54. Sattely, E.S., Fischbach, M.A. & Walsh, C.T. Total biosynthesis: *In vitro* reconstitution of polyketide and nonribosomal peptide pathways. *Nat. Prod. Rep.* **25**, 757-793 (2008).
55. Walsh, C.T., Brien, R.V.O. & Khosla, C. Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew. Chem. Int. Ed.* **52**, 7098-7124 (2013).
56. McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., Betlach, M. & Ashley, G. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1846-1851 (1999).
57. Nguyen, K.T., Ritz, D., Gu, J.Q., Alexander, D., Chu, M., Miao, V., Brian, P. & Baltz, R.H. Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 17462-17467 (2006).
58. Baltz, R.H. Combinatorial biosynthesis of cyclic lipopeptide antibiotics: A model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways. *ACS Synth. Biol.* **3**, 748-758 (2014).

59. Menzella, H.G., Reid, R., Carney, J.R., Chandran, S.S., Reisinger, S.J., Patel, K.G., Hopwood, D.A. & Santi, D.V. Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171-1176 (2005).
60. Udwary, D.W., Merski, M. & Townsend, C.A. A method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase. *J. Mol. Biol.* **323**, 585-598 (2002).
61. Wu, N., Cane, D.E. & Khosla, C. Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. *Biochemistry* **41**, 5056-5066 (2002).
62. Broadhurst, R.W., Nietlispach, D., Wheatcroft, M.P., Leadlay, P.F. & Weissman, K.J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**, 723-731 (2003).
63. Whicher, J.R., Smaga, S.S., Hansen, D.A., Brown, W.C., Gerwick, W.H., Sherman, D.H. & Smith, J.L. Cyanobacterial polyketide synthase docking domains: A tool for engineering natural product biosynthesis. *Chem. Biol.* **20**, 1340-1351 (2013).
64. Hahn, M. & Stachelhaus, T. Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 275-280 (2006).
65. Siewers, V., San-Bento, R. & Nielsen, J. Implementation of communication-mediating domains for non-ribosomal peptide production in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **106**, 841-844 (2010).
66. Weissman, K.J. Uncovering the structures of modular polyketide synthases. *Nat. Prod. Rep.* **32**, 436-453 (2015).
67. Weissman, K.J. The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* **11**, 660-670 (2015).
68. Dutta, S., Whicher, J.R., Hansen, D.A., Hale, W.A., Chemler, J.A., Congdon, G.R., Narayan, A.R.H., Hakansson, K., Sherman, D.H., Smith, J.L. & Skiniotis, G. Structure of a modular polyketide synthase. *Nature* **510**, 512-517 (2014).
69. Whicher, J.R., Dutta, S., Hansen, D.A., Hale, W.A., Chemler, J.A., Dosey, A.M., Narayan, A.R.H., Hakansson, K., Sherman, D.H., Smith, J.L. & Skiniotis, G. Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560-564 (2014).
70. Drake, E.J., Miller, B.R., Shi, C., Tarrasch, J.T., Sundlov, J.A., Allen, C.L., Skiniotis, G., Aldrich, C.C. & Gulick, A.M. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235-238 (2016).
71. Witkop, B. From the "ordeal bean" (*Physostigma venenosum*) to the ordeal of Alzheimer's disease - some of the legacy of Percy Lavon Julian (1899-1975). *Heterocycles* **49**, 9-27 (1998).
72. Granacher, R.P. & Baldessarini, R.J. Usefulness of physostigmine in neurology and psychiatry. *Clin. Neuropharmacol.* **1**, 63-79 (1976).
73. Araki, T., Ozawa, T., Yokoe, H., Kanematsu, M., Yoshida, M. & Shishido, K. Diastereoselective intramolecular carbamoylketene/alkene [2+2] cycloaddition: Enantioselective access to pyrrolidinoindoline alkaloids. *Org. Lett.* **15**, 200-203 (2013).

74. El-Azab, A.S., Taniguchi, T. & Ogasawara, K. An expedient route to the calabar bean alkaloids (-)-physovenine and (-)-physostigmine. *Org. Lett.* **2**, 2757-2759 (2000).
75. Kulkarni, M.G., Dhondge, A.P., Borhade, A.S., Gaikwad, D.D., Chavhan, S.W., Shaikh, Y.B., Ningdale, V.B., Desai, M.P., Birhade, D.R. & Shinde, M.P. A novel and efficient total synthesis of (+/-)-physostigmine. *Tetrahedron Lett.* **50**, 2411-2413 (2009).
76. Ruiz-Sanchis, P., Savina, S.A., Albericio, F. & Alvarez, M. Structure, bioactivity and synthesis of natural products with hexahydropyrrolo[2,3-*b*]indole. *Chem. Eur. J.* **17**, 1388-1408 (2011).
77. Steven, A. & Overman, L.E. Total synthesis of complex cyclotryptamine alkaloids: Stereocontrolled construction of quaternary carbon stereocenters. *Angew. Chem. Int. Ed.* **46**, 5488-5508 (2007).
78. Amador, T.A., Verotta, L., Nunes, D.S. & Elisabetsky, E. Antinociceptive profile of hodgkinsine. *Planta Med.* **66**, 770-772 (2000).
79. Amador, T.A., Verotta, L., Nunes, D.S. & Elisabetsky, E. Involvement of NMDA receptors in the analgesic properties of psychotridine. *Phytomedicine* **8**, 202-206 (2001).
80. Schuller, J.M., Zocher, G., Liebhold, M., Xie, X.L., Stahl, M., Li, S.M. & Stehle, T. Structure and catalytic mechanism of a cyclic dipeptide prenyltransferase with broad substrate promiscuity. *J. Mol. Biol.* **422**, 87-99 (2012).
81. Ma, J., Wang, Z., Huang, H., Luo, M., Zuo, D., Wang, B., Sun, A., Cheng, Y., Zhang, C. & Ju, J. Biosynthesis of himastatin: Assembly line and characterization of three cytochrome P450 enzymes involved in the post-tailoring oxidative steps. *Angew. Chem. Int. Ed.* **50**, 7797-7802 (2011).
82. Zhang, J., Marcin, C., Shifflet, M.A., Salmon, P., Brix, T., Greasham, R., Buckland, B. & Chartrain, M. Development of a defined medium fermentation process for physostigmine production by *Streptomyces griseofuscus*. *Appl. Microbiol. Biotechnol.* **44**, 568-575 (1996).
83. Yoshida, T. & Nagasawa, T. ε-Poly-L-lysine: Microbial production, biodegradation and application potential. *Appl. Microbiol. Biotechnol.* **62**, 21-26 (2003).
84. Berger, M., Gray, J.A. & Roth, B.L. The expanded biology of serotonin. *Annu. Rev. Med.* **60**, 355-366 (2009).
85. Tosini, G., Ye, K.Q. & Iuvone, P.M. *N*-Acetylserotonin: Neuroprotection, neurogenesis, and the sleepy brain. *Neuroscientist* **18**, 645-653 (2012).
86. Yano, J.M., Yu, K., Donaldson, G.P., Shastri, G.G., Ann, P., Ma, L., Nagler, C.R., Ismagilov, R.F., Mazmanian, S.K. & Hsiao, E.Y. Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell* **161**, 264-276 (2015).
87. Rudolf, J.D., Wang, H. & Poulter, C.D. Multisite Prenylation of 4-Substituted Tryptophans by Dimethylallyltryptophan Synthase. *J. Am. Chem. Soc.* **135**, 1895-1902 (2013).
88. Liu, J., Ng, T., Rui, Z., Ad, O. & Zhang, W. Unusual acetylation-dependent reaction cascade in the biosynthesis of the pyrroloindole drug physostigmine. *Angew. Chem. Int. Ed.* **53**, 136-139 (2014).

89. Keller, S., Wage, T., Hohaus, K., Holzer, M., Eichhorn, E. & van Pee, K.H. Purification and partial characterization of tryptophan 7-halogenase (PrnA) from *Pseudomonas fluorescens*. *Angew. Chem. Int. Ed.* **39**, 2300-2302 (2000).
90. Heemstra, J.R. & Walsh, C.T. Tandem action of the O_2- and $FADH_2$-dependent halogenases KtzQ and KtzR produce 6,7-dichlorotryptophan for kutzneride assembly. *J. Am. Chem. Soc.* **130**, 14024-14025 (2008).
91. Yin, W.B., Yu, X., Xie, X.L. & Li, S.M. Preparation of pyrrolo[2,3-b]indoles carrying a β-configured reverse C3-dimethylallyl moiety by using a recombinant prenyltransferase CdpC3PT. *Org. Biomol. Chem.* **8**, 2430-2438 (2010).
92. Saruwatari, T., Yagishita, F., Mino, T., Noguchi, H., Hotta, K. & Watanabe, K. Cytochrome P450 as dimerization catalyst in diketopiperazine alkaloid biosynthesis. *Chembiochem* **15**, 656-659 (2014).
93. Gerken, T. & Walsh, C.T. Cloning and sequencing of the chaetocin biosynthetic gene cluster. *Chembiochem* **14**, 2256-2258 (2013).
94. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. & Hopwood, D.A. Practical Streptomyces Genetics. (John Innes Centre Ltd, Norwich, England; 2000).
95. Gust, B., Challis, G.L., Fowler, K., Kieser, T. & Chater, K.F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1541-1546 (2003).
96. Sugawara, K., Hatori, M., Nishiyama, Y., Tomita, K., Kamei, H., Konishi, M. & Oki, T. Eponemycin, a new antibiotic active against B16 melanoma. I. Production, isolation, structure and biological activity. *J. Antibiot.* **43**, 8-18 (1990).
97. Hanada, M., Sugawara, K., Kaneta, K., Toda, S., Nishiyama, Y., Tomita, K., Yamamoto, H., Konishi, M. & Oki, T. Epoxomicin, a new antitumor agent of microbial origin. *J. Antibiot.* **45**, 1746-1752 (1992).
98. Pereira, A.R., Kale, A.J., Fenley, A.T., Byrum, T., Debonsi, H.M., Gilson, M.K., Valeriote, F.A., Moore, B.S. & Gerwick, W.H. The carmaphycins: New proteasome inhibitors exhibiting an α,β-epoxyketone warhead from a marine cyanobacterium. *Chembiochem* **13**, 810-817 (2012).
99. Koguchi, Y., Kohno, J., Suzuki, S., Nishio, M., Takahashi, K., Ohnuki, T. & Komatsubara, S. TMC-86A, B and TMC-96, new proteasome inhibitors from *Streptomyces* sp. TC 1084 and *Saccharothrix* sp. TC 1094. I. Taxonomy, fermentation, isolation, and biological activities. *J. Antibiot.* **52**, 1069-1076 (1999).
100. Owen, J.G., Charlop-Powers, Z., Smith, A.G., Ternei, M.A., Calle, P.Y., Reddy, B.V., Montiel, D. & Brady, S.F. Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 4221-4226 (2015).
101. Keller, L., Plaza, A., Dubiella, C., Groll, M., Kaiser, M. & Muller, R. Macyranones: Structure, biosynthesis, and binding mode of an unprecedented epoxyketone that targets the 20S proteasome. *J. Am. Chem. Soc.* **137**, 8121-8130 (2015).
102. Kim, K.B. & Crews, C.M. From epoxomicin to carfilzomib: Chemistry, biology, and medical outcomes. *Nat. Prod. Rep.* **30**, 600-604 (2013).

103. Groll, M., Kim, K.B., Kairies, N., Huber, R. & Crews, C.M. Crystal structure of epoxomicin: 20S proteasome reveals a molecular basis for selectivity of α',β'-epoxyketone proteasome inhibitors. *J. Am. Chem. Soc.* **122**, 1237-1238 (2000).
104. Huber, E.M., Basler, M., Schwab, R., Heinemeyer, W., Kirk, C.J., Groettrup, M. & Groll, M. Immuno- and constitutive proteasome crystal structures reveal differences in substrate and inhibitor specificity. *Cell* **148**, 727-738 (2012).
105. Grawert, M.A. & Groll, M. Exploiting nature's rich source of proteasome inhibitors as starting points in drug development. *Chem. Commun.* **48**, 1364-1378 (2012).
106. Meng, L., Mohan, R., Kwok, B.H., Elofsson, M., Sin, N. & Crews, C.M. Epoxomicin, a potent and selective proteasome inhibitor, exhibits in vivo antiinflammatory activity. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10403-10408 (1999).
107. Li, H., Ponder, E.L., Verdoes, M., Asbjornsdottir, K.H., Deu, E., Edgington, L.E., Lee, J.T., Kirk, C.J., Demo, S.D., Williamson, K.C. & Bogyo, M. Validation of the proteasome as a therapeutic target in *Plasmodium* using an epoxyketone inhibitor with parasite-specific toxicity. *Chem. Biol.* **19**, 1535-1545 (2012).
108. Kuhn, D.J., Chen, Q., Voorhees, P.M., Strader, J.S., Shenk, K.D., Sun, C.M., Demo, S.D., Bennett, M.K., van Leeuwen, F.W., Chanan-Khan, A.A. & Orlowski, R.Z. Potent activity of carfilzomib, a novel, irreversible inhibitor of the ubiquitin-proteasome pathway, against preclinical models of multiple myeloma. *Blood* **110**, 3281-3290 (2007).
109. McCormack, P.L. Carfilzomib: In relapsed, or relapsed and refractory, multiple myeloma. *Drugs* **72**, 2023-2032 (2012).
110. Schorn, M., Zettler, J., Noel, J.P., Dorrestein, P.C., Moore, B.S. & Kaysser, L. Genetic basis for the biosynthesis of the pharmaceutically important class of epoxyketone proteasome inhibitors. *ACS Chem. Biol.* **9**, 301-309 (2014).
111. Zhang, W., Li, Y. & Tang, Y. Engineered biosynthesis of bacterial aromatic polyketides in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20683-20688 (2008).
112. Matthes, D., Richter, L., Muller, J., Denisiuk, A., Feifel, S.C., Xu, Y., Espinosa-Artiles, P., Sussmuth, R.D. & Molnar, I. *In vitro* chemoenzymatic and *in vivo* biocatalytic syntheses of new beauvericin analogues. *Chem. Commun.* **48**, 5674-5676 (2012).
113. Li, J. & Neubauer, P. *Escherichia coli* as a cell factory for heterologous production of nonribosomal peptides and polyketides. *N. Biotechnol.* **31**, 579-585 (2014).
114. Janssen, H.J. & Steinbuchel, A. Fatty acid synthesis in *Escherichia coli* and its applications towards the production of fatty acid based biofuels. *Biotechnol. Biofuels* **7**, 7 (2014).
115. Tautenhahn, R., Patti, G.J., Rinehart, D. & Siuzdak, G. XCMS Online: A web-based platform to process untargeted metabolomic data. *Anal. Chem.* **84**, 5035-5039 (2012).
116. Bryce, G.F. & Brot, N. Studies on the enzymatic synthesis of the cyclic trimer of 2,3-dihydroxy-*N*-benzoyl-L-serine in Escherichia coli. *Biochemistry* **11**, 1708-1715 (1972).

117. Liu, J., Zhu, X. & Zhang, W. Identifying the minimal enzymes required for biosynthesis of epoxyketone proteasome inhibitors. *Chembiochem* **16**, 2585-2589 (2015).
118. Zettler, J., Zubeil, F., Kulik, A., Grond, S. & Kaysser, L. Epoxomicin and eponemycin biosynthesis involves *gem*-dimethylation and an acyl-CoA dehydrogenase-like enzyme. *Chembiochem* **17**, 792-798 (2016).
119. Zabala, D., Cartwright, J.W., Roberts, D.M., Law, B.J.C., Song, L.J., Samborskyy, M., Leadlay, P.F., Micklefield, J. & Challis, G.L. A flavin-dependent decarboxylase dehydrogenase monooxygenase assembles the warhead of α,β-epoxyketone proteasome inhibitors. *J. Am. Chem. Soc.* **138**, 4342-4345 (2016).
120. Vock, P., Engst, S., Eder, M. & Ghisla, S. Substrate activation by acyl-CoA dehydrogenases: Transition-state stabilization and pKs of involved functional groups. *Biochemistry* **37**, 1848-1860 (1998).
121. Ghisla, S. & Thorpe, C. Acyl-CoA dehydrogenases. A mechanistic overview. *Eur. J. Biochem.* **271**, 494-508 (2004).
122. Liu, J., Zhu, X., Kim, S.J. & Zhang, W. Antimycin-type depsipeptides: Discovery, biosynthesis, chemical synthesis, and bioactivities. *Nat. Prod. Rep.* (2016).
123. Dunshee, B.R., Leben, C., Keitt, G.W. & Strong, F.M. The isolation and properties of antimycin A. *J. Am. Chem. Soc.* **71**, 2436-2437 (1949).
124. Hosotani, N., Kumagai, K., Nakagawa, H., Shimatani, T. & Saji, I. Antimycins A_{10} - A_{16}, seven new antimycin antibiotics produced by *Streptomyces* spp. SPA-10191 and SPA-8893. *J. Antibiot.* **58**, 460-467 (2005).
125. Strangman, W.K., Kwon, H.C., Broide, D., Jensen, P.R. & Fenical, W. Potent inhibitors of pro-inflammatory cytokine production produced by a marine-derived bacterium. *J. Med. Chem.* **52**, 2317-2327 (2009).
126. Yan, L.L., Han, N.N., Zhang, Y.Q., Yu, L.Y., Chen, J., Wei, Y.Z., Li, Q.P., Tao, L., Zheng, G.H., Yang, S.E., Jiang, C.X., Zhang, X.D., Huang, Q., Habdin, X., Hu, Q.B., Li, Z., Liu, S.W., Zhang, Z.Z., He, Q.Y., Si, S.Y. & Sun, C.H. Antimycin A_{18} produced by an endophytic *Streptomyces albidoflavus* isolated from a mangrove plant. *J. Antibiot.* **63**, 259-261 (2010).
127. Han, Z., Xu, Y., McConnell, O., Liu, L., Li, Y., Qi, S., Huang, X. & Qian, P. Two antimycin A analogues from marine-derived actinomycete *Streptomyces lusitanus*. *Mar. Drugs* **10**, 668-676 (2012).
128. Yan, Y., Zhang, L., Ito, T., Qu, X., Asakawa, Y., Awakawa, T., Abe, I. & Liu, W. Biosynthetic pathway for high structural diversity of a common dilactone core in antimycin production. *Org. Lett.* **14**, 4142-4145 (2012).
129. Vanner, S.A., Li, X., Zvanych, R., Torchia, J., Sang, J., Andrews, D.W. & Magarvey, N.A. Chemical and biosynthetic evolution of the antimycin-type depsipeptides. *Mol. Biosyst.* **9**, 2712-2719 (2013).
130. Kozone, I., Ueda, J.Y., Takagi, M. & Shin-ya, K. JBIR-52, a new antimycin-like compound, from *Streptomyces* sp. ML55. *J. Antibiot.* **62**, 593-595 (2009).
131. Cassinelli, G., Grein, A., Orezzi, P., Pennella, P. & Sanfilippo, A. New antibiotics produced by *Streptoverticillium orinoci*, n. sp. *Archiv. Mikrobiol.* **55**, 358-368 (1967).
132. Pettit, G.R., Tan, R., Pettit, R.K., Smith, T.H., Feng, S., Doubek, D.L., Richert, L., Hamblin, J., Weber, C. & Chapuis, J.C. Antineoplastic agents. 560. Isolation and

structure of kitastatin 1 from an Alaskan *Kitasatospora* sp. *J Nat Prod* **70**, 1069-1072 (2007).
133. Rieske, J.S. in Antibiotics. (eds. D. Gottlieb & P.D. Shaw) 542-584 (Springer Berlin Heidelberg, 1967).
134. Shiomi, K., Hatae, K., Hatano, H., Matsumoto, A., Takahashi, Y., Jiang, C.L., Tomoda, H., Kobayashi, S., Tanaka, H. & Omura, S. A new antibiotic, antimycin A_9, produced by *Streptomyces* sp. K01-0031. *J. Antibiot.* **58**, 74-78 (2005).
135. Wikstrom, M.K. & Berden, J.A. Oxidoreduction of cytochrome b in the presence of antimycin. *Biochim. Biophys. Acta* **283**, 403-420 (1972).
136. Slater, E.C. The mechanism of action of the respiratory inhibitor, antimycin. *Biochim. Biophys. Acta* **301**, 129-154 (1973).
137. Huang, L.S., Cobessi, D., Tung, E.Y. & Berry, E.A. Binding of the respiratory chain inhibitor antimycin to the mitochondrial bc_1 complex: A new crystal structure reveals an altered intramolecular hydrogen-bonding pattern. *J. Mol. Biol.* **351**, 573-597 (2005).
138. Tzung, S.P., Kim, K.M., Basanez, G., Giedt, C.D., Simon, J., Zimmerberg, J., Zhang, K.Y. & Hockenbery, D.M. Antimycin A mimics a cell-death-inducing Bcl-2 homology domain 3. *Nat. Cell Biol.* **3**, 183-191 (2001).
139. King, M.A. Antimycin A-induced killing of HL-60 cells: Apoptosis initiated from within mitochondria does not necessarily proceed via caspase 9. *Cytometry, Part A* **63**, 69-76 (2005).
140. Park, W.H., Han, Y.W., Kim, S.H. & Kim, S.Z. An ROS generator, antimycin A, inhibits the growth of HeLa cells via apoptosis. *J. Cell. Biochem.* **102**, 98-109 (2007).
141. Ueda, J., Nagai, A., Izumikawa, M., Chijiwa, S., Takagi, M. & Shin-ya, K. A novel antimycin-like compound, JBIR-06, from *Streptomyces* sp. ML55. *J. Antibiot.* **61**, 241-244 (2008).
142. Izumikawa, M., Ueda, J., Chijiwa, S., Takagi, M. & Shin-ya, K. Novel GRP78 molecular chaperone expression down-regulators JBIR-04 and -05 isolated from *Streptomyces violaceoniger*. *J. Antibiot.* **60**, 640-644 (2007).
143. Umeda, Y., Chijiwa, S., Furihata, K., Furihata, K., Sakuda, S., Nagasawa, H., Watanabe, H. & Shin-ya, K. Prunustatin A, a novel GRP78 molecular chaperone down-regulator isolated from *Streptomyces violaceoniger*. *J. Antibiot.* **58**, 206-209 (2005).
144. Seipke, R.F., Barke, J., Brearley, C., Hill, L., Yu, D.W., Goss, R.J. & Hutchings, M.I. A single *Streptomyces* symbiont makes multiple antifungals to support the fungus farming ant *Acromyrmex octospinosus*. *PLoS One* **6**, e22028 (2011).
145. Seipke, R.F. & Hutchings, M.I. The regulation and biosynthesis of antimycins. *Beilstein J. Org. Chem.* **9**, 2556-2563 (2013).
146. Neft, N. & Farley, T.M. Studies on the biosynthesis of antimycin A. I. Incorporation of ^{14}C-labeled metabolites into the 3-formamidosalicyl moiety. *J. Antibiot.* **25**, 298-303 (1972).
147. Kurnasov, O., Jablonski, L., Polanuyer, B., Dorrestein, P., Begley, T. & Osterman, A. Aerobic tryptophan degradation pathway in bacteria: Novel kynurenine formamidase. *FEMS Microbiol. Lett.* **227**, 219-227 (2003).

148. Schoenian, I., Paetz, C., Dickschat, J.S., Aigle, B., Leblond, P. & Spiteller, D. An unprecedented 1,2-shift in the biosynthesis of the 3-aminosalicylate moiety of antimycins. *Chembiochem* **13**, 769-773 (2012).
149. Chang, C., Huang, R., Yan, Y., Ma, H., Dai, Z., Zhang, B., Deng, Z., Liu, W. & Qu, X. Uncovering the formation and selection of benzylmalonyl-CoA from the biosynthesis of splenocin and enterocin reveals a versatile way to introduce amino acids into polyketide carbon scaffolds. *J. Am. Chem. Soc.* **137**, 4183-4190 (2015).
150. Zhang, L., Mori, T., Zheng, Q., Awakawa, T., Yan, Y., Liu, W. & Abe, I. Rational control of polyketide extender units by structure-based engineering of a crotonyl-CoA carboxylase/reductase in antimycin biosynthesis. *Angew. Chem. Int. Ed.* **54**, 13462-13465 (2015).
151. Sandy, M., Zhu, X., Rui, Z. & Zhang, W. Characterization of AntB, a promiscuous acyltransferase involved in antimycin biosynthesis. *Org. Lett.* **15**, 3396-3399 (2013).
152. Hayashi, K. & Nozaki, H. Kitamycins, new antimycin antibiotics produced by *Streptomyces* sp. *J. Antibiot.* **52**, 325-328 (1999).
153. Pickens, L.B., Kim, W., Wang, P., Zhou, H., Watanabe, K., Gomi, S. & Tang, Y. Biochemical analysis of the biosynthetic pathway of an anticancer tetracycline SF2575. *J. Am. Chem. Soc.* **131**, 17677-17689 (2009).
154. Rui, Z., Petrickova, K., Skanta, F., Pospisil, S., Yang, Y.L., Chen, C.Y., Tsai, S.F., Floss, H.G., Petricek, M. & Yu, T.W. Biochemical and genetic insights into asukamycin biosynthesis. *J. Biol. Chem.* **285**, 24915-24924 (2010).
155. Teufel, R., Mascaraque, V., Ismail, W., Voss, M., Perera, J., Eisenreich, W., Haehnel, W. & Fuchs, G. Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 14390-14395 (2010).
156. Teufel, R., Friedrich, T. & Fuchs, G. An oxygenase that forms and deoxygenates toxic epoxide. *Nature* **483**, 359-U161 (2012).
157. Grishin, A.M., Ajamian, E., Tao, L.M., Zhang, L.H., Menard, R. & Cygler, M. Structural and functional studies of the *Escherichia coli* phenylacetyl-CoA monooxygenase complex. *J. Biol. Chem.* **286**, 10735-10743 (2011).
158. Li, X., Zvanych, R., Vanner, S.A., Wang, W. & Magarvey, N.A. Chemical variation from the neoantimycin depsipeptide assembly line. *Bioorg. Med. Chem. Lett.* **23**, 5123-5127 (2013).
159. Liu, J., Zhu, X., Seipke, R.F. & Zhang, W. Biosynthesis of antimycins with a reconstituted 3-formamidosalicylate pharmacophore in *Escherichia coli*. *ACS Synth. Biol.* **4**, 559-565 (2015).
160. Pfeifer, B., Hu, Z.H., Licari, P. & Khosla, C. Process and metabolic strategies for improved production of *Escherichia coli*-derived 6-deoxyetythronolide B. *Appl. Environ. Microbiol.* **68**, 3287-3292 (2002).
161. Umeda, Y., Furihata, K., Sakuda, S., Nagasawa, H., Ishigami, K., Watanabe, H., Izumikawa, W., Takagi, M., Doi, T., Nakao, Y. & Shin-ya, K. Absolute structure of prunustatin A, a novel GRP78 molecular chaperone down-regulator. *Org. Lett.* **9**, 4239-4242 (2007).
162. Takahashi, K., Tsuda, E. & Kurosawa, K. SW-163A and B, novel immunosuppressants produced by *Streptomyces* sp. *J. Antibiot.* **54**, 867-873 (2001).

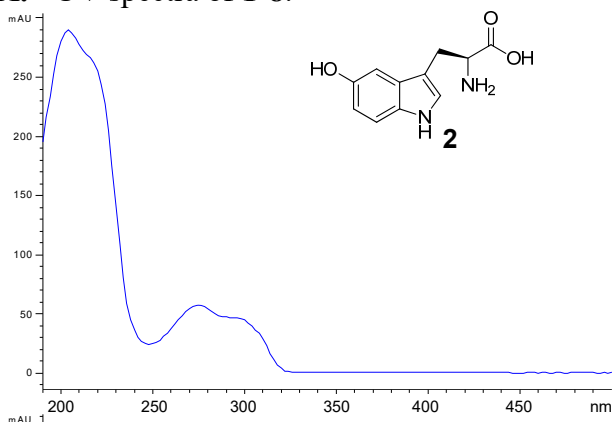
163. Pettit, G.R., Smith, T.H., Feng, S., Knight, J.C., Tan, R., Pettit, R.K. & Hinrichs, P.A. Antineoplastic agents. 561. Total synthesis of respirantin. *J. Nat. Prod.* **70**, 1073-1083 (2007).
164. Li, X., Zvanych, R., Torchia, J. & Magarvey, N.A. Structures and biosynthesis of 12-membered macrocyclic depsipeptides from *Streptomyces* sp. ML55. *Bioorg. Med. Chem. Lett.* **23**, 4150-4153 (2013).
165. Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. & Zagnitko, O. The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* **9** (2008).
166. Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F.F. & Stevens, R. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42**, D206-D214 (2014).
167. Felnagle, E.A., Barkei, J.J., Park, H., Podevels, A.M., McMahon, M.D., Drott, D.W. & Thomas, M.G. MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry* **49**, 8815-8817 (2010).
168. Zhang, W., Heemstra, J.R., Walsh, C.T. & Imker, H.J. Activation of the pacidamycin PacL adenylation domain by MbtH-like Proteins. *Biochemistry* **49**, 9946-9947 (2010).
169. Herbst, D.A., Boll, B., Zocher, G., Stehle, T. & Heide, L. Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes. *J. Biol. Chem.* **288**, 1991-2003 (2013).
170. Kotowska, M. & Pawlik, K. Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Appl. Microbiol. Biotechnol.* **98**, 7735-7746 (2014).
171. Kim, B.S., Cropp, T.A., Beck, B.J., Sherman, D.H. & Reynolds, K.A. Biochemical evidence for an editing role of thioesterase II in the biosynthesis of the polyketide pikromycin. *J. Biol. Chem.* **277**, 48028-48034 (2002).
172. Heathcote, M.L., Staunton, J. & Leadlay, P.F. Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem. Biol.* **8**, 207-220 (2001).
173. Salim, A.A., Cho, K.J., Tan, L., Quezada, M., Lacey, E., Hancock, J.F. & Capon, R.J. Rare *Streptomyces N*-formyl amino-salicylamides inhibit oncogenic K-Ras. *Org. Lett.* **16**, 5036-5039 (2014).
174. Dorrestein, P.C., Bumpus, S.B., Calderone, C.T., Garneau-Tsodikova, S., Aron, Z.D., Straight, P.D., Kolter, R., Walsh, C.T. & Kelleher, N.L. Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry. *Biochemistry* **45**, 12756-12766 (2006).
175. Dorrestein, P.C. & Kelleher, N.L. Dissecting non-ribosomal and polyketide biosynthetic machineries using electrospray ionization Fourier-Transform mass spectrometry. *Nat. Prod. Rep.* **23**, 893-918 (2006).

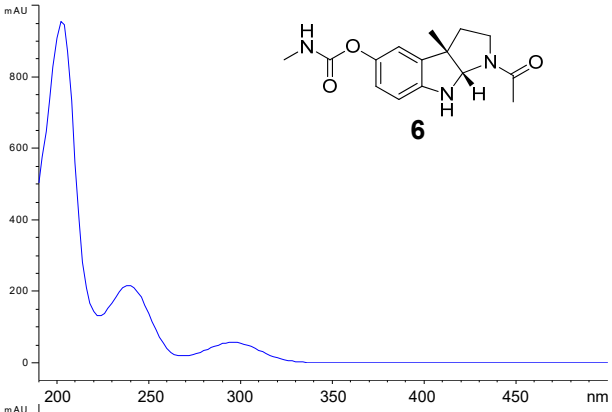
176. Bumpus, S.B. & Kelleher, N.L. Accessing natural product biosynthetic processes by mass spectrometry. *Curr. Opin. Chem. Biol.* **12**, 475-482 (2008).
177. Zeng, J., Wagner, D.T., Zhang, Z., Moretto, L., Addison, J.D. & Keatinge-Clay, A.T. Portability and structure of the four-helix bundle docking domains of trans-acyltransferase modular polyketide synthases. *ACS Chem. Biol.* (2016).
178. Yadav, G., Gokhale, R.S. & Mohanty, B. Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.* **328**, 335-363 (2003).
179. Dunn, B.J. & Khosla, C. Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. *J. R. Soc. Interface* **10** (2013).
180. Wilkinson, C.J., Hughes-Thomas, Z.A., Martin, C.J., Bohm, I., Mironenko, T., Deacon, M., Wheatcroft, M., Wirtz, G., Staunton, J. & Leadlay, P.F. Increasing the efficiency of heterologous promoters in actinomycetes. *J. Mol. Microbiol. Biotechnol.* **4**, 417-426 (2002).
181. Zhang, W., Bolla, M.L., Kahne, D. & Walsh, C.T. A three enzyme pathway for 2-amino-3-hydroxycyclopent-2-enone formation and incorporation in natural product biosynthesis. *J. Am. Chem. Soc.* **132**, 6402-6411 (2010).

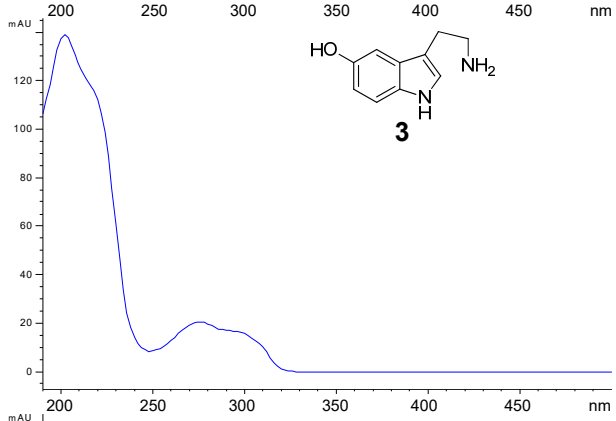
Appendices

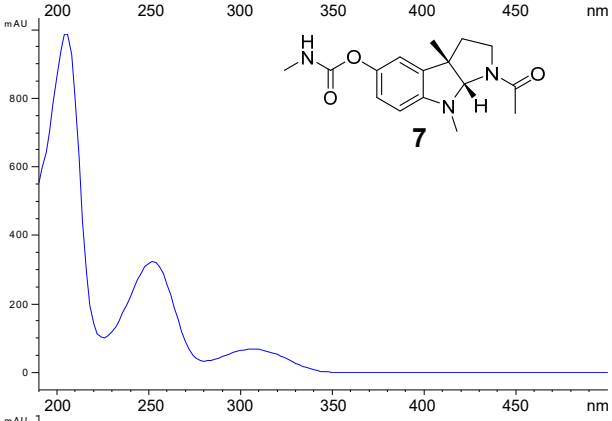
Appendix A. UV spectra of **1**-**8**.
Appendix B. HRMS/MS analysis of peptidyl epoxyketone compounds **15** and **16**.
Appendix C. HRMS/MS analysis of antimycin-type compounds **17**-**19**.
Appendix D. UV spectra of **17**, **18**, and antimycin A_2.

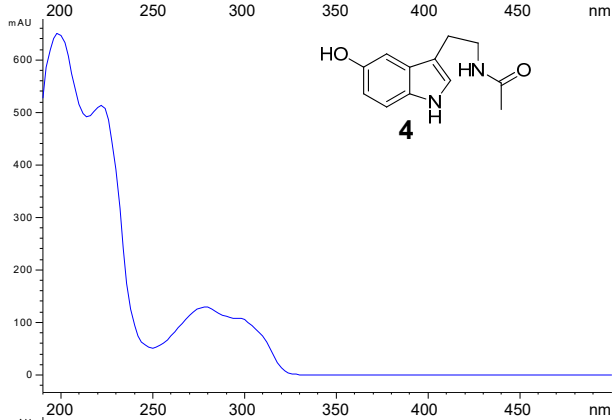
A. UV spectra of **1**-**8**.

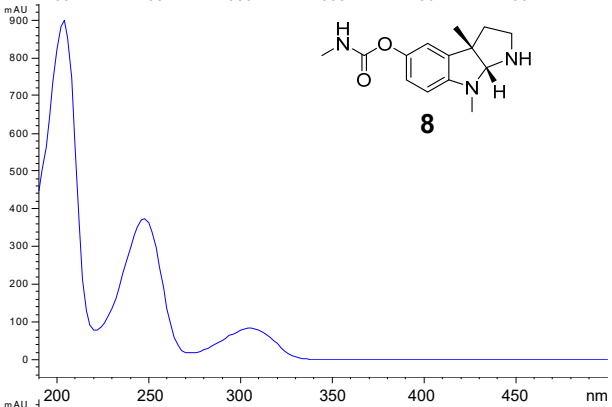


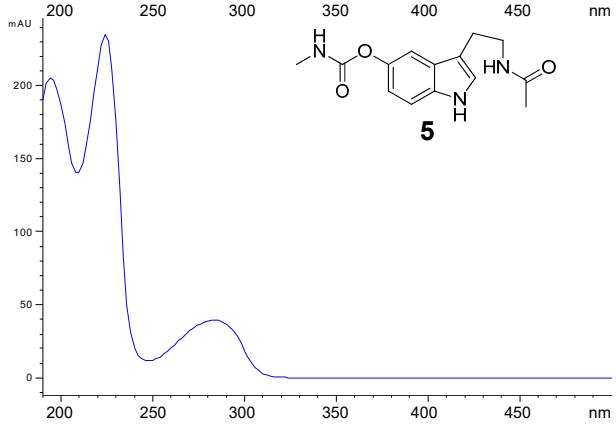


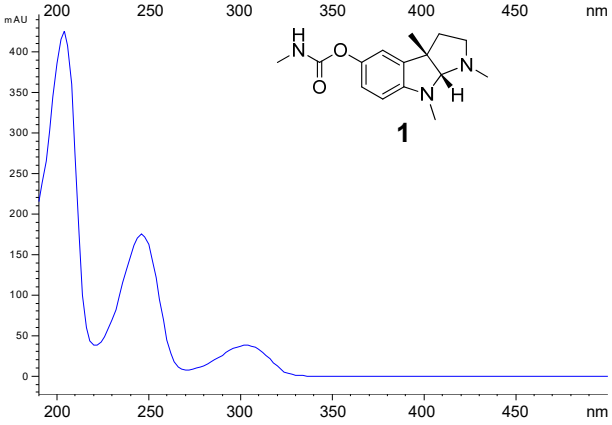




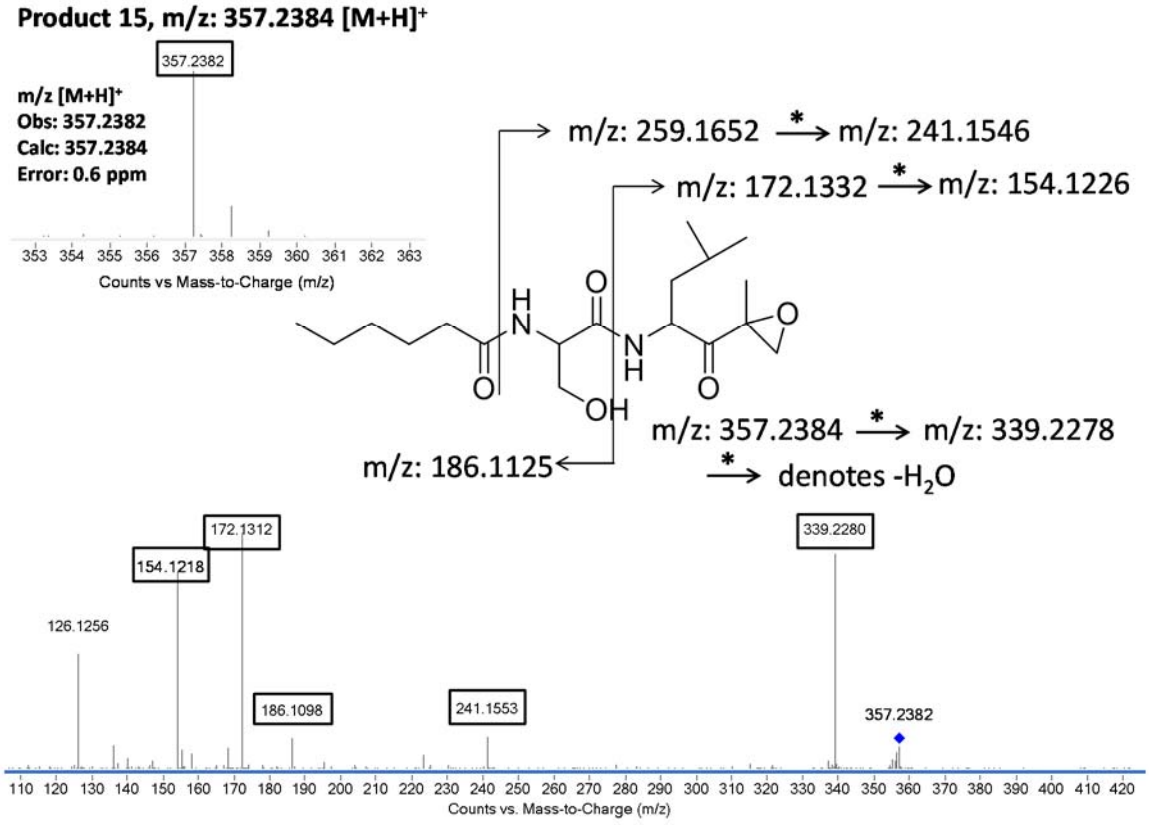


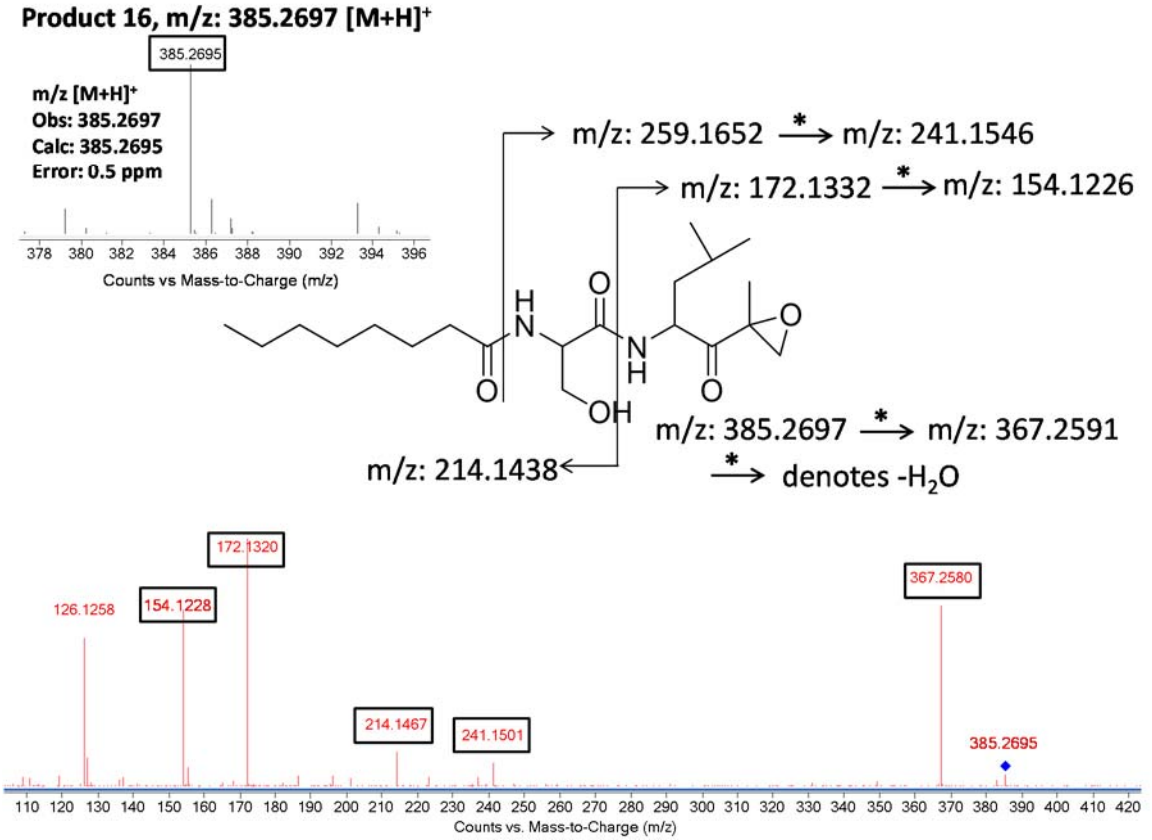




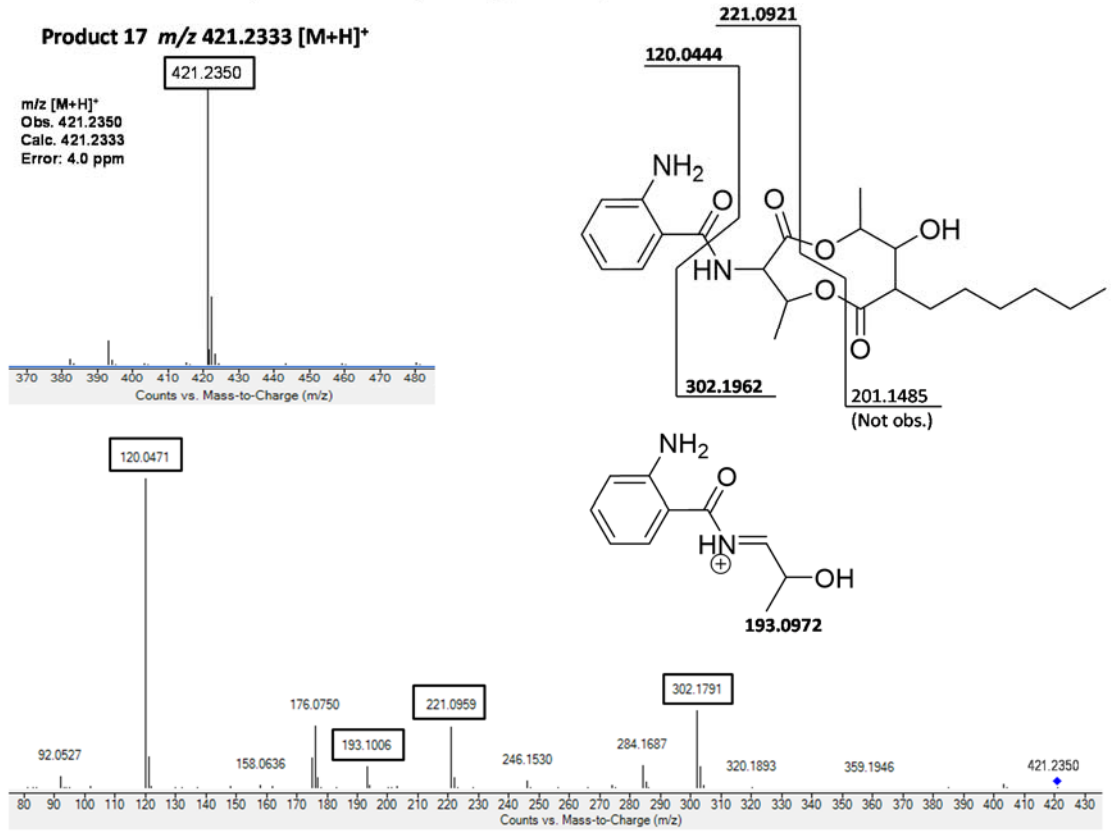


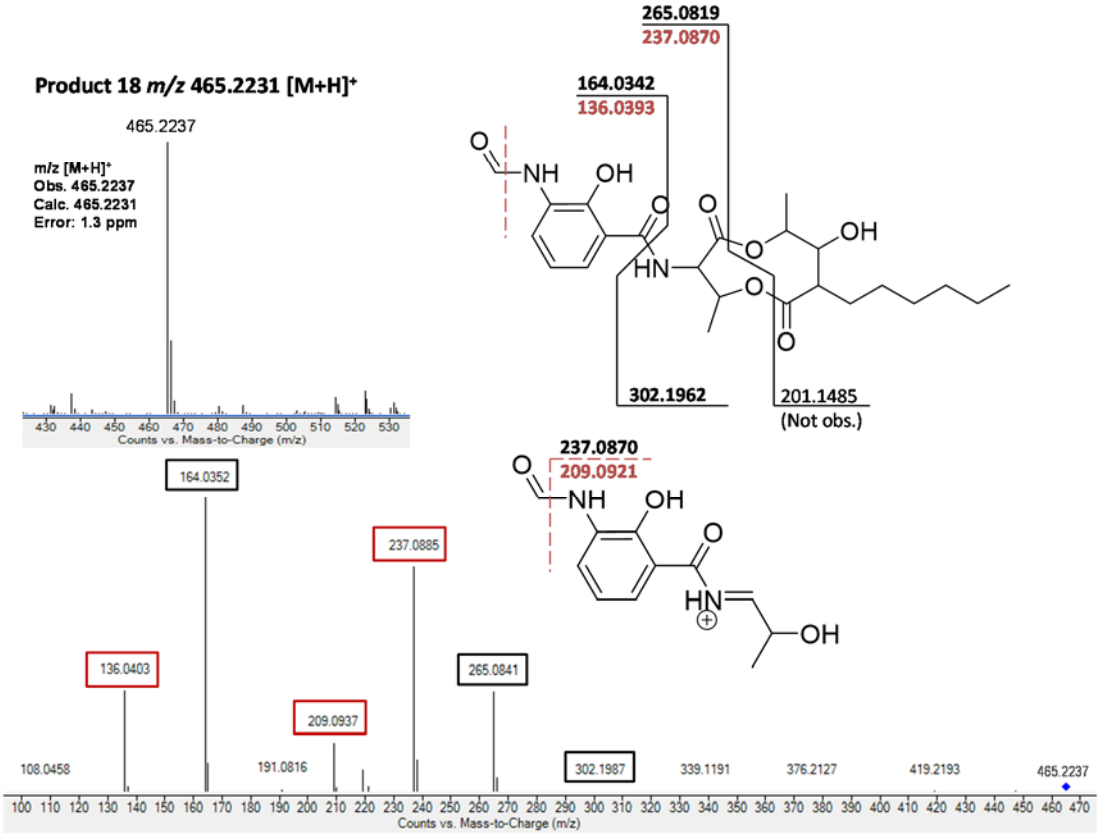
B. HRMS/MS analysis of peptidyl epoxyketone compounds **15** and **16**.

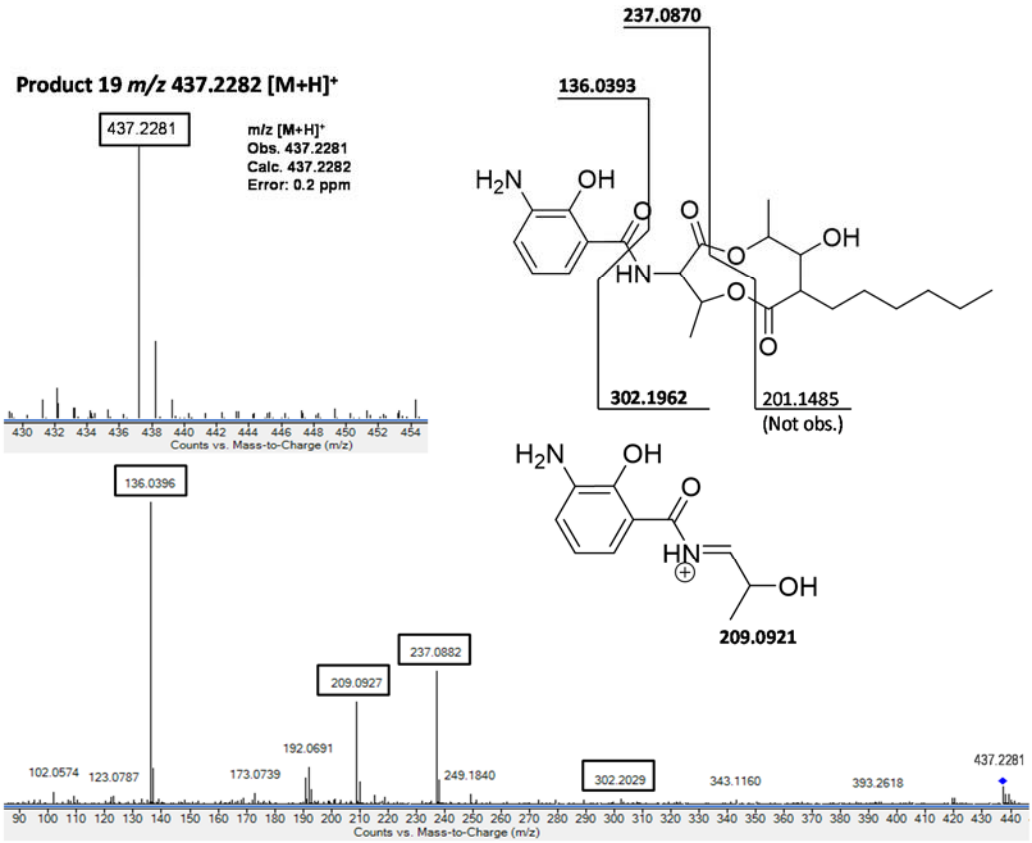




C. HRMS/MS analysis of antimycin-type compounds **17**-**19**.






Product 19 m/z 437.2282 [M+H]+
437.2281
m/z [M+H]+
Obs. 437.2281
Calc. 437.2282
Error: 0.2 ppm
Counts vs. Mass-to-Charge (m/z)
237.0870
136.0393
302.1962
201.1485
(Not obs.)
209.0921
136.0396
209.0927
237.0882
192.0691
102.0574
123.0787
173.0739
249.1840
302.2029
343.1160
393.2618
437.2281
Counts vs. Mass-to-Charge (m/z)

D. UV spectra of **17**, **18**, and antimycin A_2.