UCLA

UCLA Electronic Theses and Dissertations

Title

Biosynthesis of (—)-Sambutoxin, a Fusarium 2-Pyridone Alkaloid Mycotoxin

Permalink

https://escholarship.org/uc/item/0pn7497n

Author

Go, Louise

Publication Date

2021

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
Los Angeles

Biosynthesis of (–)-Sambutoxin, a *Fusarium* 2-Pyridone Alkaloid Mycotoxin

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Chemistry

by

Louise Go

2021

© Copyright by

Louise Go

2021

ABSTRACT OF THE DISSERTATION

Biosynthesis of (–)-Sambutoxin,
a *Fusarium* 2-Pyridone Alkaloid Mycotoxin

by

Louise Go
Doctor of Philosophy in Chemistry
University of California, Los Angeles, 2021
Professor Yi Tang, Chair

This dissertation describes the discovery by genome mining and characterization of the biosynthetic pathway for sambutoxin, a 4-hydroxy-2-pyridone alkaloid mycotoxin from the *Fusarium* fungi. The 4-hydroxy-2-pyridone alkaloid family of fungal natural products has been the subject of numerous biosynthetic studies due to their structurally diverse scaffolds that arise from a common class of biosynthetic precursors, which are transformed by various redox tailoring enzymes and cyclases that have been uncovered by genome mining. Sambutoxin is the first 2-pyridone alkaloid with a *p*-hydroxyphenyl moiety that is derived from L-phenylalanine via a late-stage oxidation of the side chain phenyl group, and we also show that sambutoxin is a biosynthetic precursor to funiculosin, a 4-hydroxy-2-pyridone alkaloid natural product with a unique cyclopentanetetraol motif.

The dissertation of Louise Go is approved.

Neil Kamal Garg

Hosea Martin Nelson

Keriann Marie Backus

Yi Tang, Committee Chair

University of California, Los Angeles

2021

For my parents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This work was supported by grants from the NIH (1R01AI141481) to Yi Tang and the NSF Graduate Research Fellowship to Louise Go.

There are countless people to thank for all their care and support throughout my PhD journey, the first of whom is my advisor, Professor Yi Tang. For me, a great mentor is someone who teaches me to embrace the challenges that I am faced with while providing the wisdom and strength needed to overcome them. I am ever grateful to Professor Tang for having welcomed me into his group as a chemistry graduate student with minimal molecular biology experience and offered me projects that were stimulating and well beyond my comfort zone, all the while supporting me through the challenges and encouraging me in my small wins. Thank you for your concern for not only my academic success but also my physical & mental health and overall well-being, I appreciate everything from the casual drop-in chats in the office, meals that the Faculty Center, to the Friday drop-off at the Union Station for my train ride home. You have absolutely the best one & only guideline for presentations and writing: "make the audience feel like they learned something, that it was worth their time." I most definitely learned a lot from my time in your lab, and it was much more than just worth my time.

In addition to Professor Tang, I would also like to thank all the past & present members of the Tang Lab that I have had the privilege of learing from and working with. Members of the Tang lab including Elizabeth Abramyan, Dr. Undramaa Bat-Erdene, Dr. John Billingsley, Dr. Mengbin Chen, Dr. Wei Cheng, Moriel Dror, Dr. Yang Hai, Dr. Leibniz Hang, Dr. Sunny Hung, Dr. Nicholas Liu, Josh Misa, Colin Johnson, Dr. Masa Ohashi, Ike Okorafor, Dr. Bruno Perlatti, Dr. Zuodong Sun, Jason Williams, Danielle Yee, and Dr. Zhuan Zhang have all offered me guidance in multitude of ways for me to develop as a researcher in the Tang lab. Spending a good portion of my graduate career in the COVID-19 pandemic has made me realize how much I draw from the collaborations with and the company of all Tang

lab members.

A very special shout-out goes to Lee Joon Kim, my eternal research partner and friend. What a fortune and privilege to be able to share in both my undergraduate and graduate research journeys with you—I cannot imagine the last 8 years without your friendship, MicroED expertise, and all our late-night dinners and hours-long conversations. They say that the past predicts the future, and I'm already looking forward to our next collaborative science project after UCLA.

Outside of my research circles, I thank my thesis committee members for their support and inspiration during our committee meetings and beyond—Professor Keri Backus for her wisdom as a guest instructor in the CBI seminar on chemical proteomics, Professor Neil Garg for demonstrating steadfast commitment to both graduate and undergraduate chemistry education and outreach, and Professor Hosea Nelson for his scientific enthusiasm and gift of non-fungible can of Coke that I will treasure. I am grateful for the friendships of Joe Mazzetti, Gina Lee, Woojin Lee, and Susan Joo from the Chemistry Department and my fellow badminton crew Huong Nguyen, Mengyi Tao, Henry Do, and Tam Nguyen for the fun weekend adventures. Lastly, I would like to thank the following individuals who have had an immense positive impact on me as a scientist and a person:

- My Mom and Dad, for their relentless love and sacrifice in support of my interests and passions;
- Prof. David Vosburg, my undergraduate research advisor and mentor in faith, for his teaching and encouragement throughout college and continued mentorship to this day;
- the Southern California chapters of the Korean-American Scientists and Engineers Association and Korean-American Women in Science and Engineering, for their unwavering professional support and friendship for the past 15 years; and last but not least,

- Yukun Lin, my partner, for his friendship and love, keeping me company and being a constant encouragement throughout my grad school journey while never failing to make me laugh.

VITA

2017–2021 NSF Graduate Research Fellow
UCLA Department of Chemistry & Biochemistry, Los Angeles, CA.

2015 B.S. in Joint Chemistry & Biology, Concentration in Spanish
Harvey Mudd College, Claremont, CA.

2014 Undergraduate Research Fellow
Memorial Sloan Kettering Cancer Center, New York, NY.

PUBLICATIONS

Go, E. and Tang, Y. "Natural Product Biosynthesis." *Advanced Chemical Biology* **2021**.

Go, E.; Kim, L.; Ohashi, M.; Tang, Y. "Biosynthesis of the Fusarium mycotoxin (–)-sambutoxin." *Organic Letters* **2021**, *23*, 7819–7823.

Go, E. and Tang, Y. "Fungal Highly Reducing Polyketide Synthases and Associated Natural Products." *Comprehensive Natural Products III: Chemistry and Biology* **2020**, Volume 1, 333–364.

Go, E.; Wetzler, S. P.; Kim, L.; Chang, A. Y.; Vosburg, D. A. M. "Concise, diastereoconvergent synthesis of endiandric-type tetracycles by iterative cross coupling." *Tetrahedron* **2016**, *72*, 3790–3794.

Go, E.; Srisuknimit, V.; Cheng, S. L.; Vosburg, D. A. "Self-assembly, guest capture, and NMR spectroscopy of a metal-organic cage in water." *J. Chem. Ed.* **2016**, *93*, 368–371.

CHAPTER 1

Introduction to Natural Product Biosynthesis

Natural products are broadly defined as small (< 2500 Da) organic molecules produced by living organisms. They include caffeine in a cup of iced coffee, the intermediates and products of citric acid cycle, and the toxins from marine algae that cause shellfish poisoning upon human consumption. For this dissertation, natural products will be defined as secondary metabolites, or small molecules produced as a result of conditional metabolic pathways in living organisms. Unlike primary metabolites (e.g. amino acids, nucleotides, carbohydrates, vitamins), which are present in organisms from all kingdoms of life and are required for host growth and cellular maintenance, secondary metabolites are produced by only a small subset of organisms under specific conditions. Secondary metabolites are not essential for life, although they may confer ecological advantages for the host, such as protection from predators and pathogens.[1]

Natural products are sometimes referred to as the chemical arsenal of living organisms, targeting other invasive, competing organisms by inhibiting key enzymes and pathways. Humans have taken advantage of the biological activities of natural products to use them as therapeutics in various capacities, as demonstrated by the penicillin family of antibiotics (from *Penicillium* fungi), erythromycin (from *Saccharopolyspora* bacteria), and the antimalarial drug artemisin (from *Artemisia annua* plant) (**Figure 1.1**).[2–4] It is estimated that natural products and their derivatives comprise more than half of all drugs approved by the United States Food and Drug Administration within the last forty years.[5]

Due to their biological activity and structural complexity, natural products have at-

tracted much interest from the biomedical and chemical research communities alike. The multitude of chiral centers and dense arrangement of functional groups in complex natural products have inspired generations of chemists seeking to develop efficient strategies for synthesizing them.[6] Efforts into the total synthesis of natural products and semisynthesis of derivatives have resulted in significant advancements in new bond-forming reactions,[7] methods for C–H functionalization,[8,9] protecting-group-free synthesis,[10,11] and atom-economical synthesis with simple and abundant starting materials.[12] While these advances are impressive accomplishments of human innovation, the curious minds ask the following: how does Nature the (bio)chemist synthesize these molecules?

1.1 Natural Product Biosynthetic Logic

Through evolution, Nature has become a versatile and efficient chemist in synthesizing secondary metabolite natural products. Each molecule is synthesized by a set of enzymes in a dedicated biosynthetic pathway that often consists of two stages: assembly and tailoring.

During the assembly stage, simple building blocks from primary metabolism, such as acetyl-CoA, amino acids, and isoprenyl pyrophosphate, are polymerized and often cyclized by the assembly enzymes to form the skeleton of the natural product. Natural products are classified according to their monomers and assembly enzymes—peptides are made of amino acid building blocks, polyketides are synthesized from small carboxylic acids, and terpenes consist of 5-carbon isoprene units (**Figure 1.1**). Occasionally, when biosynthesis requires a building block that is not found in primary metabolism, such as a noncanonical amino acid, the pathway will include dedicated enzymes that synthesize the building block for "on-demand" usage by the assembly enzymes. Following the assembly stage, tailoring enzymes decorate the backbone/skeleton of natural products into their richly-decorated complex structures. Tailoring enzymes include transferases, oxidoreductases, and monooxygenases, and altogether they morph the nascent compounds produced by the assembly enzymes into final products that at times can be unrecognizable. Both the assembly and tailoring en-

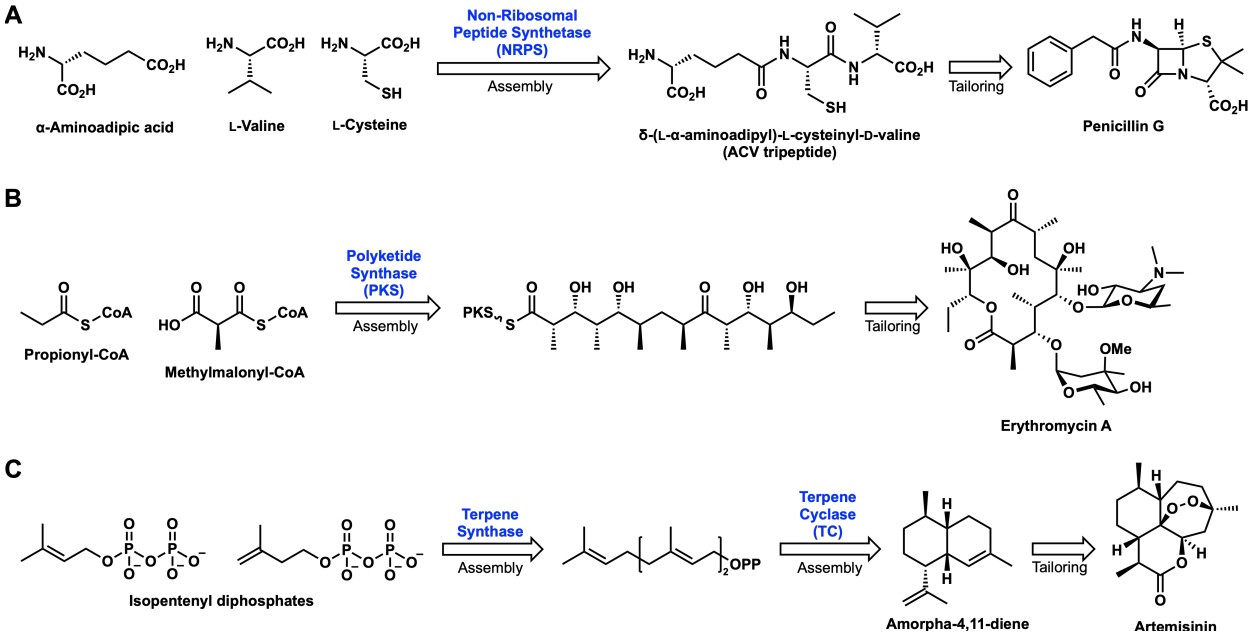


Figure 1.1. Overview of biosynthesis of the three major classes of secondary metabolite natural products. (A) non-ribosomal Peptides, (B) Polyketides, and (C) Terpenoids.

zymes use thermodynamically activated but kinetically stable cellular energy currencies such as adenosine triphosphate (ATP), nicotinamide adenine dinucleotide phosphate (NADPH), and coenzyme A (CoA) to drive the entropically expensive steps.[13] Other cofactors and coenzymes such as *S*-adenosyl methionine (SAM), flavin, and heme are also used. Biosynthetic pathways are full of elegant chemical logic, much like a beautifully planned total synthesis. Nature's biosynthetic strategies and human-designed chemical synthetic strategies for natural products often converge,[14,15] highlighting that the underlying chemical principles that guide both are one and the same.

At the genetic level, the genes encoding the biosynthetic pathway enzymes are often co-localized on the genomes to form a biosynthetic gene cluster (BGC). The clustering of biosynthetic genes is particularly prominent in bacterial and fungi, a phenomenon that is hypothesized to allow a more direct transcriptional co-regulation and horizontal gene transfer between species. At this time, given a sequenced microbial genome, it is possible to identify potential BGCs from co-localization of assembly and tailoring enzymes. Increased

understanding of BGCs and biosynthetic enzymology has resulted in the discovery of many candidate natural product BGCs from genome sequences alone, a discovery catalyzed by the influx of newly-sequenced microbial genomes. The term "genome mining" refers to the annotation and functional characterization of BGCs that may encode for previously unknown natural product biosynthetic pathways,[16,17] and this approach is used to solve the biosynthesis of sambutoxin as will be described in Chapter 3.

1.2 Peptide Natural Products

The first main family of natural products to be discussed are the small-molecule peptides, a broad family that contains many front-line drugs such as nisin, vancomycin, and the penicillin family of antibiotics (**Figure 1.2**). Just like the enzymes and macromolecular peptides in living systems, the scaffolds of natural product peptides are assembled via polymerization of amino acids. This process can take place on the ribosome, or it can also occur via a non-ribosomal pathway. While amino acids are the primary constituents of these compounds, other building blocks can also be incorporated, especially in the non-ribosomal peptides. For example, daptomycin is a peptide-derived natural product containing a fatty acid moiety that is specifically incorporated by its biosynthetic machinery. **Figure 1.2B** lists the key points of comparison between the ribosomal and non-ribosomal routes of peptide biosynthesis, which will be discussed separately in the next two sections.

1.2.1 Ribosomally-synthesized and post-translationally modified peptides

Ribosomally-synthesized and post-translationally modified peptides (RiPPs) are natural products derived from peptide precursors that are synthesized by the ribosome.[18] Some RiPPs can be further classified into subfamilies such as lantipeptides, thiopeptides, lasso peptides, and sactipeptides, which arise from post-translational modifications (PTMs) of their precursor peptides. RiPP precursors are synthesized according to the central dogma of molecular biology, in which the gene encoding the precursor is transcribed into mRNA

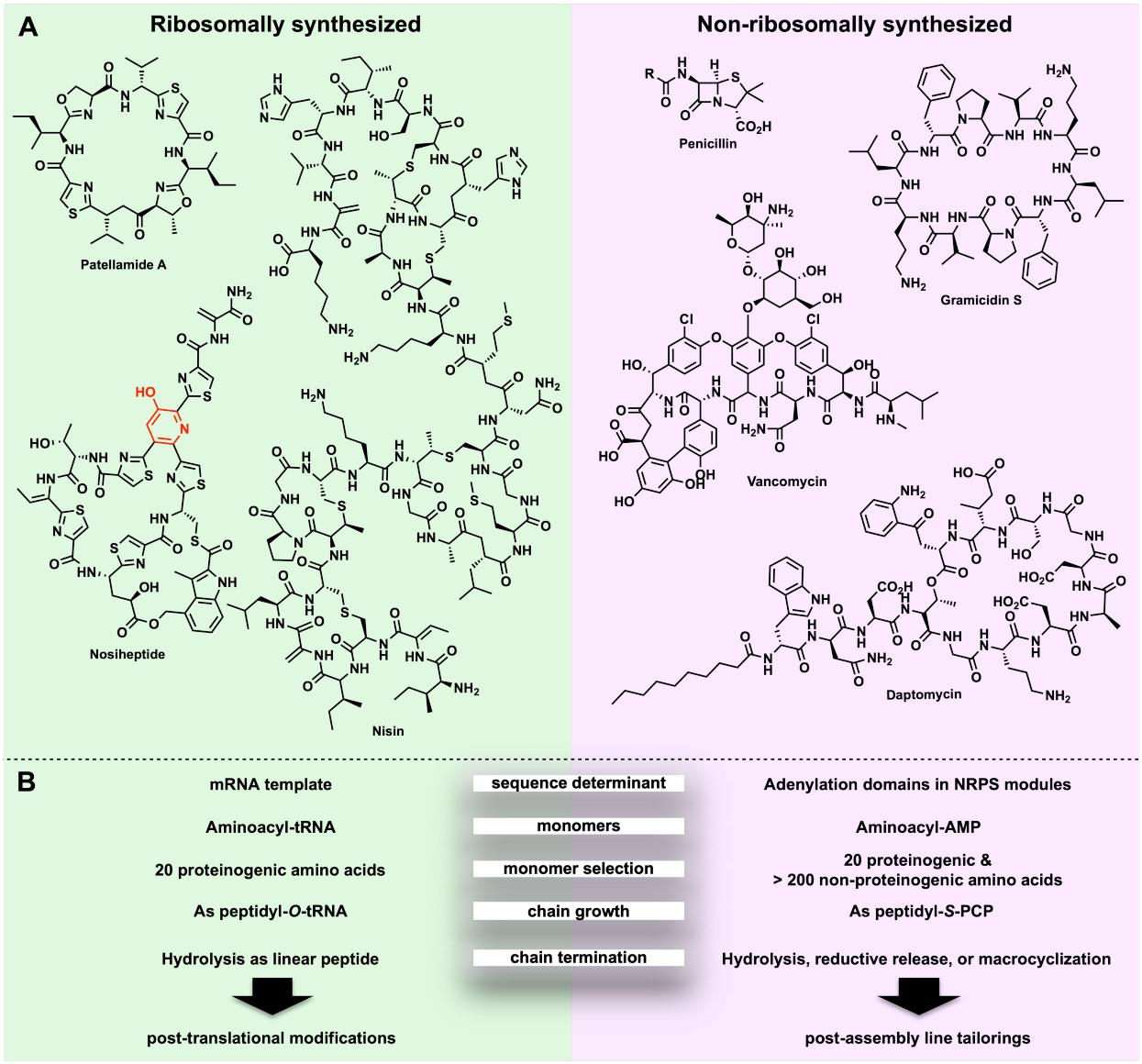


Figure 1.2. Comparison of the ribosomally-synthesized and non-ribosomally-synthesized peptides. The 3-hydroxpyridine motif formed from an aza-Diels-Alder reaction between dehydrated Ser and/or Thr is highlighted in red.

for use as a template by the ribosome to produce the corresponding polypeptide. The 3-nucleotide codons of the mRNA are recognized by the complementary 3-nucleotide anticodons of aminoacyl-tRNA complexes; each tRNA molecule carries a unique amino acid as an oxoester, providing a kinetically stable yet thermodynamically activated *C*-terminus that is primed for nucleophilic attack by the amine group of an incoming amino acid. The ribosome is not directly involved in loading the amino acids onto their cognate tRNAs, a

task that is reserved for aminoacyl-tRNA synthetases—it simply uses them as pre-activated building blocks to synthesize the peptide. While several hundreds of amino acids occur naturally, RiPPs are assembled exclusively from the 20 proteinogenic amino acids (or 22, including selenocysteine and pyrrolysine).[19]

The amino acid sequence of a RiPP (5-30 aa) is specified in the host genome, in a small gene encoding a RiPP precursor peptide that is translated at the start of biosynthesis. The RiPP precursor peptide contains an *N*-terminal leader peptide portion and a core peptide that corresponds to the peptide moiety in the final natural product. The leader peptide is important for recognition by and recruitment of biosynthetic enzymes that carry out the PTMs to morph the core peptide into different classes of RiPPs. The leader peptides are removed by proteases following the PTMs to yield the final natural product RiPPs. The PTMs introduced during RiPP biosynthesis serve primarily to rigidify the peptide chains into their bioactive conformations and to protect from proteolysis (**Figure 1.3**).

Short for "lanthionine-containing antibiotics," lantibiotics comprise the most extensively studied class of ribosomally-synthesized peptides.[20] Lanthionine is a nonproteinogenic amino acid that resembles a dimer of alanine residues that are cross-linked at their β-carbons by a thioether linkage; in the context of lantibiotics, the thioether linkage is formed by addition of the cysteine thiol to C_β in the side chain of dehydrated serine/threonine (**Figure 1.3A**). The dehydration of serine/threonine occurs via phosphorylation of their side chain hydroxyl groups by a kinase, followed by elimination of the H_α and phosphate to generate dehydroalanine (Dha, from serine) or dehydrobutyrine (Dhb, from threonine). The second set of enzymes recruited to the leader peptide subsequently catalyze the intramolecular thioether formation via addition of the cysteine thiolate group, followed by stereospecific protonation to regenerate the stereocenter at C_α of Dha or Dhb . Once all the thioether linkages are formed, the leader peptide is removed, leaving behind the final RiPP. The most famous lantipeptide is nisin, a 34-aa RiPP produced by *Lactococcus lactis* (**Figure 1.2A**). An antibiotic that is used as a food additive to suppress bacterial growth, nisin contains five of

the aforementioned thioester linkages that introduce five macrocyclic rings into the peptide.

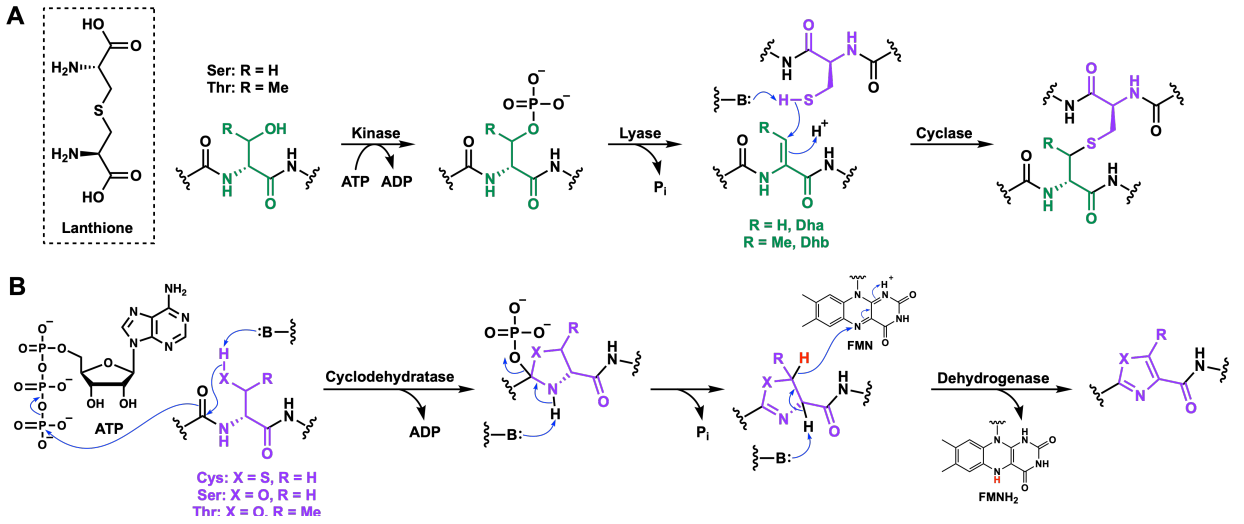


Figure 1.3. Mechanisms for two common RiPP tailoring steps. (A) Lanthiones are formed via nucleophilic attack of a Cys side chain thiol to dehydrated Ser or Thr. (B) Thiazoles and oxazoles are formed via cyclodehydration and dehydrogation of Cys, Ser, and Thr.

Another class of RiPPs, the thiazolyl peptides (thiopeptides) are rich in cysteine, threonine, and serine residues. The key feature that distinguishes thiopeptides from lantibiotics is the PTM of cysteine residues, which are transformed into thiazoles instead of forming lanthionines, as well as serine and threonine residues that form oxazoles (**Figure 1.3B**). Examples of well-known thiopeptides include thiostrepton, thiocillin, and nosiheptide (**Figure 1.2**), which are produced by the *Streptomyces* and *Bacillus* bacteria and are used as antibiotics against other gram-positive bacteria. As noted in the structure of nosiheptide, threonine/serine residues can undergo the same dehydration as in lantibiotics to give Dha and Dhb. While the dehydrated alkenes persist in some residues, in others, they act as Michael acceptors or form a diene-dienophile pair that participates in an aza-Diels-Alder reaction to form dehydropiperidine and hydroxypyridine motifs (**Figure 1.2**).

1.2.2 Non-ribosomal peptides

Similar to their ribosomal counterparts, non-ribosomal peptides are also synthesized from amino acid building blocks, ranging from 3 to more than twenty amino acids in length (**Figure 1.1**). Instead of the ribosome, however, non-ribosomal peptides are synthesized by a class of enzymes known as the non-ribosomal peptide synthetases (NRPSs). These enzymes are distinct from the ribosome in that they function independent of mRNAs or any other "template" biomolecules. In addition, NRPSs use free amino acids as building blocks rather than aminoacyl-tRNAs, and they do not rely on additional enzymes to activate the amino acids like the aminoacyl-tRNA synthetases in the ribosomal pathway. The lack of a biosynthetic template and pre-activated amino acid building blocks means that additional mechanisms must be in place for NRPSs to select the right amino acids, activate the carboxy terminus for peptide bond formation, and polymerize them in the correct order.

As will be discussed in the polyketide synthase section as well, NRPSs are multidomain, multimodule megasynthases that use assembly-line logic to synthesize its peptide. Each module is a functional unit that contains the structural and catalytic domains necessary to add one amino acid during the assembly stage of the peptide. In NRPS systems, each module is responsible for selecting, activating and incorporating an amino acid. Shown in **Figure 1.4A** is a conceptual diagram of an NRPS with its initiation, elongation, and termination modules. The initiation module is the very first module in the NRPS, and it is responsible for loading the peptide's first (*N*-terminal) amino acid onto the NRPS. The initiation module is followed by several elongation modules, one module per amino acid. The last module in the NRPS, the termination module, includes all domains for incorporating the final (*C*-terminal) amino acid and releasing the final peptide from the NRPS as either a linear or a cyclic peptide.

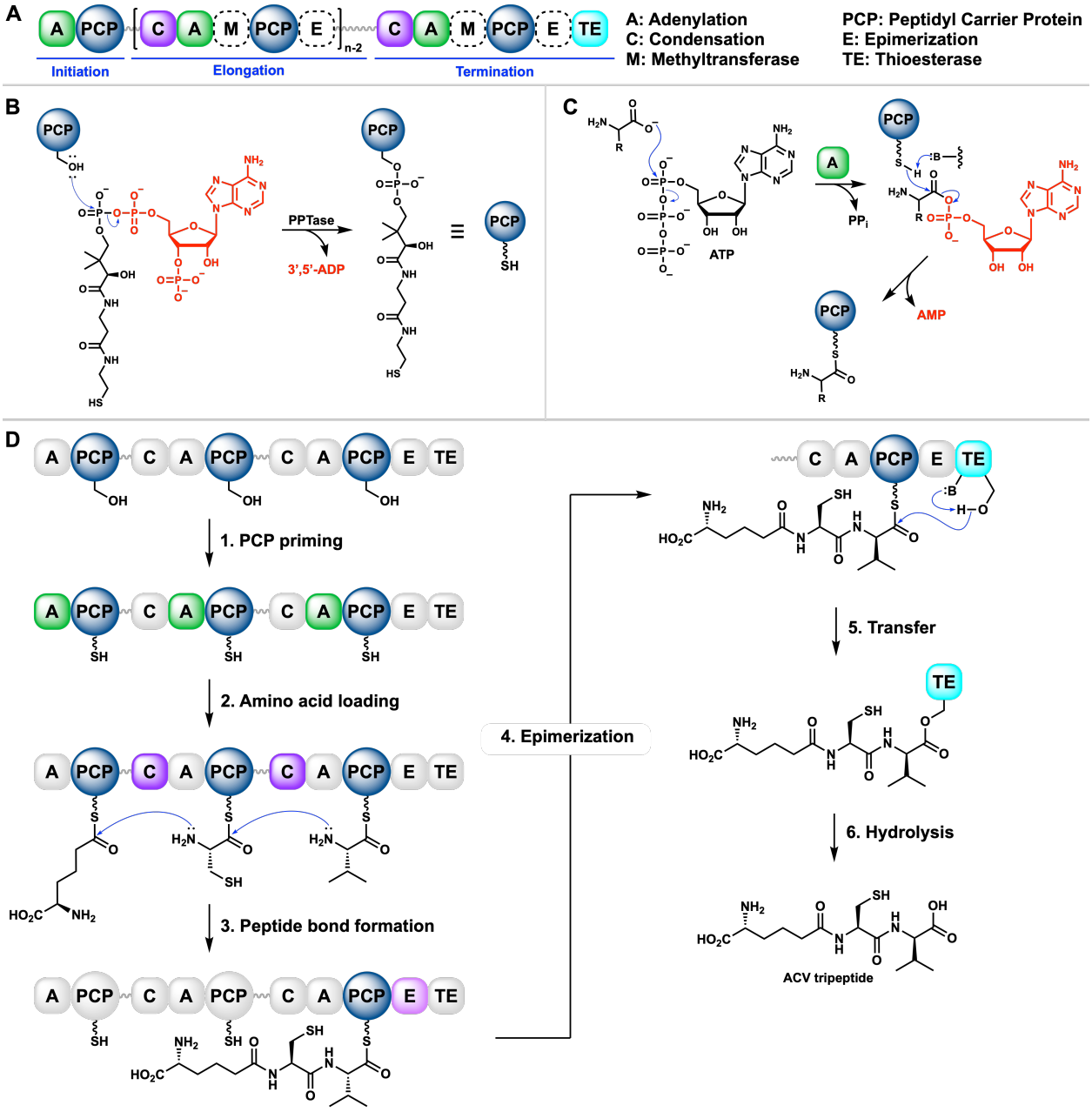


Figure 1.4. Organization and activities of NRPS domains. (A) A general scheme of NRPS modules. (B) Mechanism for PPTase priming of the PCP with Ppant. (C) Mechanism for amino acid activation and loading onto the PCP by the A domain. (D) NRPS assembly line for the biosynthesis of ACV tripeptide.

The minimally required domains for each elongation module are as follows:

- Adenylation (A): selects and activates the amino acid by *C*-terminal adenylation using one equivalence of ATP.
- Peptide carrier protein (PCP or thiolation, T): loads the activated aminoacyl adenylate from the A domain onto its prosthetic 4'-phosphopantetheine (Ppant) arm as an aminoacyl thioester.
- Condensation (C): forms the peptide bond between the amino group of the aminoacyl thioester on the PCP_N domain and the carboxy group of the growing peptide on the PCP_{N-1} domain from the previous module.

In order for the NRPS to be functional, the conserved serine residues in all PCPs must be covalently modified by Ppant, a process commonly referred to as "priming" catalyzed by a phosphopantetheinyl transferase (PPTase) housekeeping enzyme (**Figure 1.4B**). The PPTase is shared with other NPRSs and polyketide synthases in the host as well as with fatty acid synthase in primary metabolism. PCPs are small (80–100 amino acids in length), mostly α-helical proteins that have no other catalytic functions. By modifying the hydroxyl group of the active-site serine into a thiol with Ppant, the amino acids and peptides can be carried on PCPs as thermodynamically activated thioesters. The Ppant arm is approximately 18 Å in length and is flexible and dynamic in solution. These features allow the acyl groups loaded on the Ppant to be transferred from one module to the next, effectively forming an enzymatic assembly line. Biosynthetic intermediates are loaded as thioesters on carrier proteins in polyketide synthases that will be discussed in the next section.

Once the PCP domains are phosphopantetheinylated, the amino acid building blocks are activated by their cognate A domains for loading. Adenylation domains serve as the gatekeepers in NRPS biosynthesis by ensuring that correct amino acids are activated and loaded onto the PCPs, just like the aminoacyl-tRNA synthetases are tasked with loading the correct amino acids onto their cognate tRNAs for ribosomal peptide synthesis. The

amino acid specificity of A domains is determined largely by the identities of ∼10 amino acids lining their active sites, which participate in complementary non-covalent interactions with the side chains of specific amino acids.[21] Unlike aminoacyl-tRNA synthetases, NRPS A domains can activate nonproteinogenic amino acids, tapping into a much wider (>200) pool of amino acid building blocks in addition to the proteinogenic amino acids. Adenylation domains activate amino acids with one molecule of ATP per amino acid to generate the corresponding aminoacyl adenylates (**Figure 1.4C**), the same process by which amino acids are activated by aminoacyl-tRNA synthetases. The activated aminoacyl adenylates are then loaded onto the Ppant-modified PCPs as aminoacyl thioesters.

With the PCPs primed and loaded with amino acid building blocks, the C domains are now ready to assemble the amino acids into a peptide. Peptide biosynthesis on NRPS proceeds linearly starting from the initiation module, with peptide bond formation occurring between the growing peptide in the immediate upstream module and the aminoacyl thioester on the PCP of the current module (**Figure 1.4D**). The C domain catalyzes the nucleophilic attack of the free amine on the upstream thioester, facilitated by a general based histidine that deprotonates the amine. X-ray crystal structures of C domains complexed with PCP domains show that the active site of C domain has two tunnels leading to the reaction site, each accommodating the Ppant arm of the upstream and downstream PCP domains.[22] The C domain function therefore parallels that of the larger ribosomal subunit that catalyzes peptide bond formation.

In some cases, an NRPS module may harbor additional domains besides the minimal A, PCP, and C domains. *N*-methyltransferase (MT) domains methylate the amine group of amino acids, and epimerase (E) domains invert the stereocenter at C_α of the amino acid from its module. *N*-methylation occurs on the aminoacyl-*S*-PCP prior to the condensation step, and depending on the system, epimerization takes place either before or after condensation. In the case of ACV tripeptide, L-valine is converted to D-valine at the tripeptide stage after the condensation (**Figure 1.4D**), and in gramicidin S biosynthesis, the first L-

phenylalanine residue is epimerized to D-phenylalanine immediately after PCP loading prior to condensation.[23] Lastly, similar to RiPPs, certain cysteine, serine, and threonine residues in non-ribosomal peptides may also be transformed into thiazoles and oxazoles by oxidation (Ox) and cyclization (Cy) domains.

As the growing peptide moves down the assembly line as a PCP-bound peptidyl thioester, the C domain in each module elongates the peptide by one amino acid until the termination module, where the final peptide is released from the NRPS. The release is typically catalyzed by a thioesterase (TE) domain at the end of the assembly line. The TE domain active site contains a serine residue that captures the peptide as an oxyester (**Figure 1.4D**). Depending on the function of the specific TE domain, the peptide can be released as a linear peptide through hydrolysis by water (as in the ACV tripeptide), cyclized via attack by the *N*-terminal amine or a nucleophilic side chain from an internal amino acid (e.g. $-NH_2$ of Lys, –OH of Thr/Ser) as in bacitracin and daptomycin (**Figure 1.4A**).

The linear organization of domains and modules in an NRPS enables the assembly of peptide natural products in a ribosome-independent fashion. As will be discussed in the polyketide synthase section, co-linearity is a key feature of the NRPS assembly line. The co-linearity rule specifies that each module is used once and that the assembly line moves in the *N*- to *C*-terminal direction. As a result, the structure of final product can be mapped to the vectoral arrangement of modules and their catalytic domains. The number of total modules in an NRPS corresponds to the length of its product peptide, whose structure can be inferred from the A domain specificity and presence of modification domains.[24] Albeit with exceptions, this feature allows for the general prediction of the product of an NRPS from its sequence. Hence, non-ribosomal peptides can be thought of as being "encoded" by the modules and domains of their synthetases.

Just as the DNA codons in an open-reading frame can be modified to change the amino acid sequence of the translated peptide, catalytic domains of NRPSs can be deleted, rear-

ranged, or added to a module to generate previously unknown un-natural products with varied structures. One popular approach in NRPS engineering is to modify the amino acid specificity of A domains to generate non-ribosomal peptides with new sequences.[25,26] This approach is analogous to introducing a site-specific mutation in the DNA codon so that it hybridizes with a different aminoacylated tRNA, or engineering the aminoacyl-tRNA synthetase to append a new, unnatural amino acid onto its cognate tRNA for incorporation into the peptide. For example, in the NRPS that synthesizes the undecapeptide calcium-dependent antibiotic CdaPS3, the active site of a Glu-specific A domain in module 10 was mutated to preferentially accept glutamine and (2*S*,3*R*)-3-methyl glutamine, a non-natural amino acid.[25] In another study, the Phe-specific A domain in the first module of gramicidin S synthetase was modified to accept *O*-propargyl-tyrosine or para-azido phenylalanine, unnatural amino acids with a terminal alkyne or an azide group for click chemistry in living systems.[26] The change in amino acid specificity from phenylalanine to *O*-propargyl-tyrosine or para-azido phenylalanine was achieved by the single mutation of a tryptophan residue in the A domain active site to a serine, which resulted in an enlarged binding pocket for accommodating the added functional groups on phenylalanine.

1.3 Polyketide Natural Products

In addition to peptides, polyketides are another major class of natural products that are also synthesized in an assembly-line fashion. They are made by polyketide synthases (PKSs), several types of which are found in bacteria and fungi that synthesize structurally diverse molecules. Examples of bacterial polyketides include erythromycin (antibiotic), ivermectin (antiparasitic), amphotericin (antifungal), daunomycin (anticancer), and FK-506 (immunosuppressant) (**Figure 1.5A**). Two of the three groups of bacterial PKSs are discussed in this section: type-I PKSs, which tend to synthesize more reduced, macrocyclic polyketides; and type-II PKSs, which synthesize polycyclic aromatic polyketides.

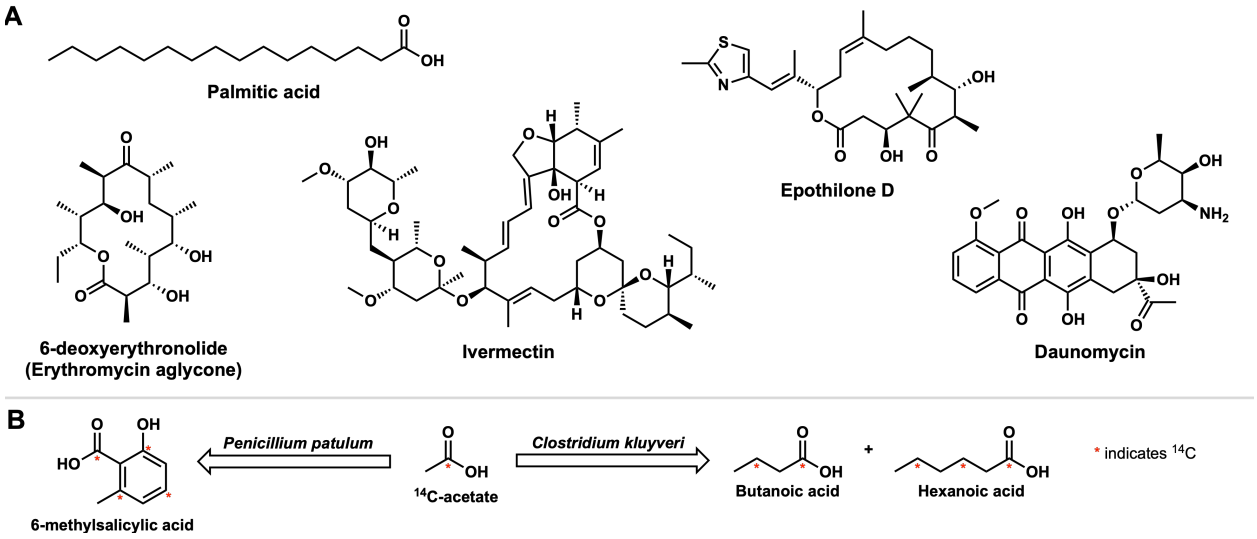


Figure 1.5. Biogenesis of fatty acids and polyketide natural products. (A) Palmitic acid and select examples of bacterial polyketide natural products. (B) Early findings showing that acetate is the building block for biogenesis of 6-methylsalicylic acid and short-chain fatty acids.

One of the earliest fundamental research on polyketide biosynthesis was performed by Arthur Birch in the 1950s, who proposed that polyketides may be formed via head-to-tail condensation of 2-carbon units derived from acetate.[27] Birch tested this hypothesis by feeding $^{14}C_{\alpha}$-acetate to a strain of *Penicillium* fungus that produced 6-methylsalicylic acid (MSA) (**Figure 1.5B**), one of the simplest aromatic polyketides that can be made from a tetra-β-ketyl precursor. A controlled degradation of ^{14}C-labeled 6-MSA revealed that the molecule was uniformly enriched with ^{14}C at the sites anticipated from Birch's hypothesis. Today, we understand this 2-carbon building block of polyketides to be acetyl-CoA.

The core enzymatic machinery for natural product biosynthesis often have precedents in primary metabolism, and this is no exception for polyketides. It is currently understood that the assembly-line machinery of microbial PKSs evolved from fatty acid synthases (FASs).[28] In mammals, some of the acetyl-CoA generated from the metabolism of carbohydrates in diet are assimilated into fatty acids in triglycerides for energy storage, and palmitic acid (**Figure 1.5A**) is synthesized by the mammalian FAS (*m*FAS) during lipogenesis. Palmitic

acid biosynthesis effectively involves polymerizing eight units of acetyl-CoA to give a fully saturated 16-C fatty acid. Early isotopic labeling studies on fatty acid biosynthesis with $^{14}C_{\alpha}$-acetate showed that ^{14}C labeling occurred at alternating carbons starting with the carbonyl carbon of butanoic and hexanoic acids,[29] an outcome consistent with Birch's early hypothesis that acetyl-CoA molecules are adjoined in a head-to-tail manner. For palmitic acid, seven iterations of C–C bond formation occur between the C_{α} of an incoming acetyl-CoA molecule and the carbonyl carbon of the growing fatty acid chain.

Two molecules of acyl thioesters can be joined by a head-to-tail C–C bond via a thio-Claisen condensation reaction, a chemistry that is also used by Nature to polymerize acetyl-CoA in fatty acid biosynthesis. However, FASs do not use acetyl-CoA directly as the building block for fatty acids. Instead, Nature enables this enolate chemistry with malonyl-CoA, an α-carboxylated derivative of acetyl-CoA that is synthesized *in vivo* by acetyl-CoA carboxylase (**Figure 1.6A**). Decarboxylation of malonyl-CoA generates the thioenolate of acetyl-CoA, which adds to the acyl thioester on the FAS/PKS to extend the chain (**Figure 1.6C**). As shown in **Figure 1.6B**, the acyl transferase (AT) domain delivers a molecule of malonyl-CoA to the acyl carrier protein (ACP) domain via transesterification, and the ketosynthase (KS) domain catalyzes the C–C bond formation via thio-Claisen condensation. The KS activity results in a β-keto adduct, which is subsequently reduced to a β-alcohol by the ketoreductase (KR) domain, from β-alcohol to a β-enoyl thioester by the dehydratase (DH) domain, and finally from β-enoyl thioester to a saturated acyl thioester by the enoyl reductase (ER) domain. One complete cycle of reductive β-keto processing results in the overall transformation of a β-ketone to a methylene unit. The mechanism for all steps involved in reducing a β-ketone to methylene is shown in **Figure 1.6D**.

1.3.1 Bacterial Type-I Polyketides

Erythromycin is the first polyketide to be mapped to its biosynthetic genes in the producing bacterium host *Saccharopolyspora erythraea*, and a model for the assembly-line biosynthesis

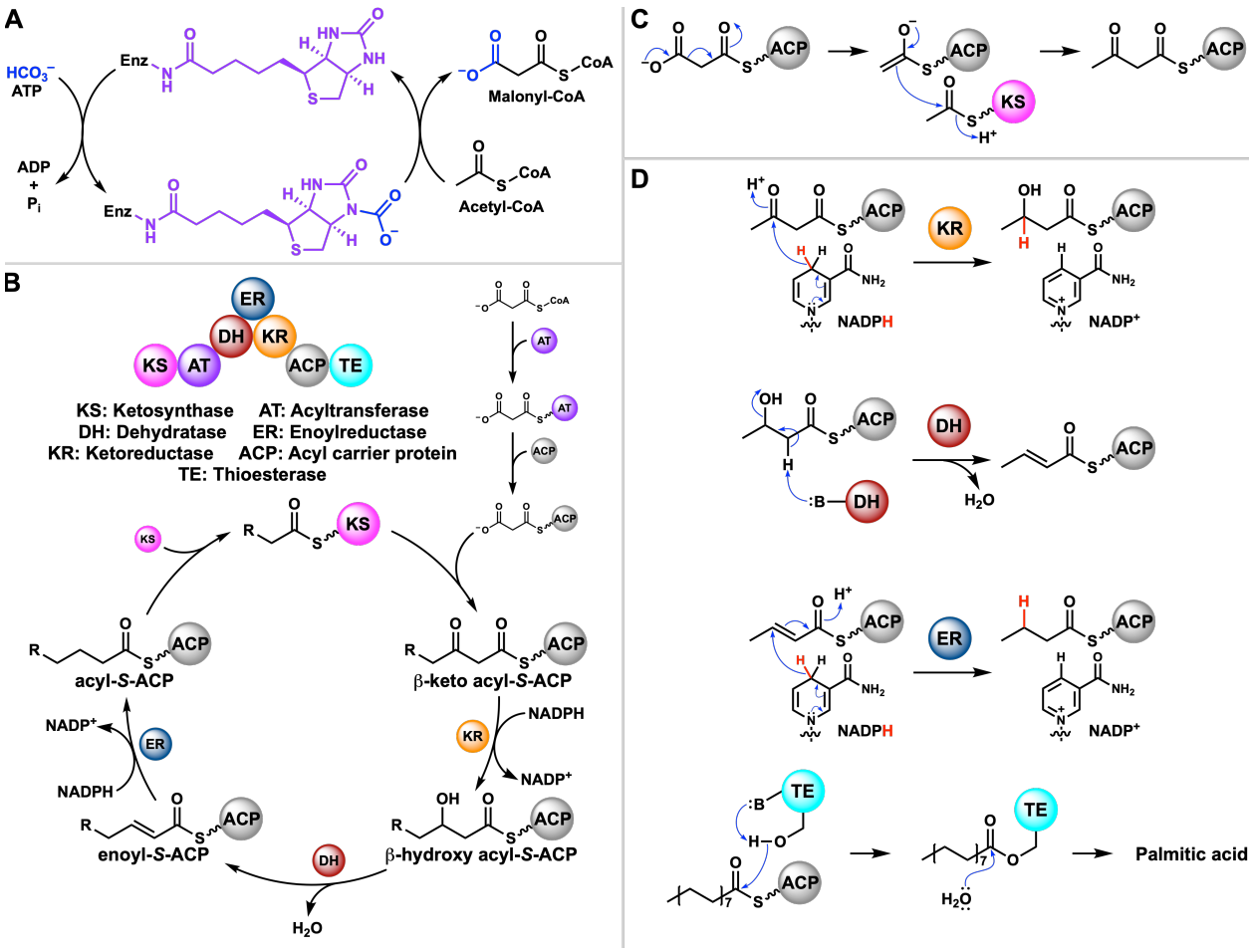


Figure 1.6. Overview of the FAS & bacterial type-I PKS machinery. (A) Biogenesis of malonyl-CoA from acetyl-CoA facilitated by biotin. (B) Type-I PKS domains and their functions. (C) Mechanism of the KS-catalyzed thio-Claisen condensation. (D) Mechanisms of β-reducing domains (KR, DH, ER) and product release by TE.

of its precursor 6-deoxyerythronolide B (6-dEB) was proposed and validated by Donadio *et al.*[30] **Figure 1.7A** shows the DEBS PKS assembly line that synthesize 6-dEB, organized into eight modules distributed among five large polypeptides. With the exception of the loading and termination modules, each module minimally contains the KS, AT and ACP domains. Some elongation modules have no β-modification domains (module 3), while others have only the KR (modules 1, 2, 5 and 6) or all of KR, DH and ER (module 4). The modular assembly line ends with a TE domain that performs the macrocyclization to yield the macrolactone. Once 6-dEB is released from DEBS, post-PKS tailoring enzymes glycosylate the β-alcohols generated by KR domains in modules 1 and 2 with deoxysugars L-mycarose

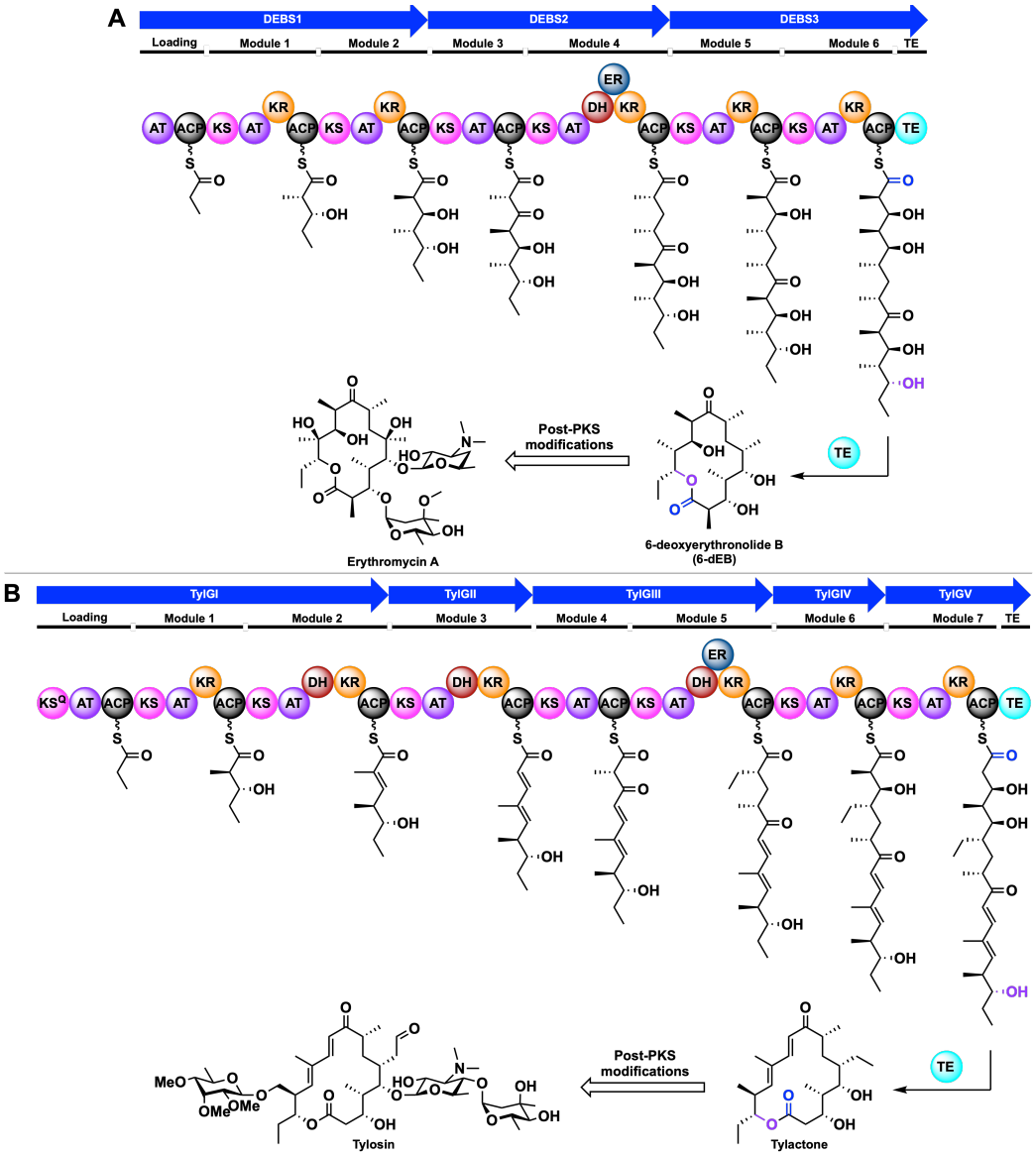


Figure 1.7. Assembly-line type-I bacterial PKS. (A) 6-deoxyerythronolide B (6-dEB) synthase (DEBS) and (B) tylactone synthase assemble the polyketide intermediates that are transformed into the antibiotic erythromycin A and tylosin, respectively.

and D-desosamine and hydroxylate two different positions on the 6-dEB backbone to complete the biosynthesis of erythromycin A. DEBS uses methylmalonyl-CoA as the exclusive building block for 6-dEB, but other type-I PKSs such as the tylactone PKS (**Figure 1.7B**) incorporate a more varied group of α-carboxyl-CoA building blocks.

Just like in NRPSs, the ACP domains in a PKS are post-translationally modified with Ppant for carrying the polyketide intermediates as thioesters. The ACP is analogous to the PCP in NRPS, and both have been referred to as thiolation (T) domain in the literature. Seven out of eight modules in DEBS have ACPs, the first of which is the loading module that initiates 6-dEB biosynthesis with a propionyl starter unit. Some loading modules in PKSs, including DEBS, consist of just the AT and ACP domains, where the AT domain directly transfers an uncarboxylated acyl group from an acyl-CoA molecule to the ACP. Other loading modules, such as in the TylGI, contain a KS^Q domain—a KS-like domain with decarboxylase activity that accepts an α-carboxyacyl substrate such as methylmalonyl-CoA, which is decarboxylated to afford the propionyl unit after ACP loading.[31,32] In DEBS, once the loading module ACP is acylated, the subsequent elongation modules extend the propionate by performing sequential thio-Claisen condensation and β-keto reduction reactions. Building block selection and transacylation on the ACP in each module is controlled by the AT domain, which plays an analogous role to adenylation domains in NRPS. While the *m*FAS AT domain exclusively selects malonyl-CoA to synthesize an unbranched fatty acid, bacterial type-I PKS AT domains can also accept α-substituted carboxyacyl-CoAs in addition to malonyl-CoA. Diversity in building block selection by AT domains can be observed in the tylactone PKS (**Figure 1.7B**): modules 1, 2, 4 and 6 select methylmalonyl-CoA, which correspond to α-methyl groups in the polyketide product; modules 3 and 7 select malonyl-CoA, and module 5 selects ethylmalonyl-CoA, which is reflected in the α-ethyl group in the tylactone backbone. The building block specificity of AT can be mapped to the amino acids in the active site and can now be predicted with reasonable accuracy.[33]

Once the ACP is loaded with the extender unit, the KS domain catalyzes the thio-Claisen condensation between the α-carboxyacyl-*S*-ACP and the polyketide intermediate on the ACP of the upstream module. The polyketide intermediate from upstream ACP is transferred to the KS active site, where it undergoes thio-Claisen condensation with α-carboxyacyl-*S*-ACP. The resulting elongated β-ketoacyl product remains loaded on the ACP of the current module and leaves the KS active site. Depending on the β-keto reduction domains that are

present in the current module, the newly formed β-ketone may be kept intact, partially reduced to a β-alcohol or α,β-unsaturated enone, or completely reduced to a methylene. In the tylactone PKS (TylG), the KR domain in module 1 stereospecifically reduces the β-ketone to β-alcohol (**Figure 1.7B**). It is important to note that if the α-position is substituted and prone to spontaneous epimerization, as in the α-methyl-β-ketone acyl-*S*-ACP at the module 1 KS, the KR reduces the substrate diastereoselectively and determines the stereochemistry at both the α and β positions (note the difference in stereochemistry of products after the first module between the two PKS systems shown in **Figure 1.7**). X-ray structures of various KR domains have led to sequence-activity relationships and identification of fingerprint motifs that determine KR diastereoselectivity.[34] If a module incorporates an α-substituted carboxyacyl-CoA building block and has all of the reductive domains, as in module 4 of DEBS and module 5 of TylG, the stereocenter at C_α is erased during the dehydration step catalyzed by the DH domain and re-established during enoyl reduction by the ER domain. Once the polyketide chain reaches the end of the assembly line, the TE domain catalyzes the release of the final product via hydrolysis and macrocyclization, in a similar fashion to the NPRS TE domains.

Although NRPSs and bacterial type-I PKSs synthesize two very different classes of natural products, the assembly line biosynthetic logic used by both systems is marvelously modular and elegantly simple. Governed by the co-linearity rule, type-I polyketide biosynthesis proceeds in a stepwise manner while chemical functionalities are introduced at both α and β positions in the process. Structural diversity in type-I polyketides is generated by i) the number of modules, which determines the overall length of the polyketide; ii) the AT domains that select carboxyacyl building blocks with varied α substituents; iii) the extent of reductive β-processing and stereochemistry as determined by the β-modification domains in each module; and iv) the cyclization pattern determined by the TE domain. These modular assembly lines are highly (re)-programmable for engineered biosynthesis of un-natural polyketides. Considering the synthetic difficulties associated with accessing and manipulating these polyketide structures, considerable efforts have been made to re-engineer PKS

assembly lines. Using DEBS as a classic example, selective removal of β-modification domains from a module led to predictable changes in the structure of the modified polyketide products: i) deletion of a KR domain in one of the modules resulted a corresponding unreduced ketone in the position consistent with the module with the deleted KR; ii) replacement of an AT domain specific for methylmalonyl-CoA with an AT domain for ethylmalonyl-CoA resulted in the incorporation of ethyl group in the aglycone; iii) removal of entire modules from the C-terminal end led to formation of macrolactones of smaller sizes; etc.[35] In a large scale effort, McDaniel *et al.* generated over 60 analogs of 6-dEB by swapping AT domains and β-modification domain between DEBS and other type-I PKSs.[36] The ease with which new un-natural derivatives of natural products can be obtained simply by modifying the domains in their biosynthetic enzymes in a predictable manner once again underscores Nature's ingenuity in its modular design of natural product biosynthetic processes.

1.3.2 Bacterial Type-II Polyketides

Bacterial type-II polyketides comprise another large class of natural products with therapeutic importance. They include the front-line anticancer drug doxorubicin (and the related daunorubicin) and the antibiotic tetracycline (**Figure 1.8A**). Type-II polyketides are characterized by fused polycyclic aromatic structures, unlike their type-I macrolide counterparts (**Figure 1.8B**), but the underlying assembly chemistry is the same: polymerization of acetate units by the PKS machinery, namely the KS domain. However, there are several notable distinctions between the biochemistry of type-I and type-II PKS systems. While type-I PKSs are made up of catalytic domains organized in a modular assembly line, the catalytic domains of type-II PKSs (KS, AT, ACP, etc.) exist as discrete enzymes, and they cooperate via non-covalent protein-protein interactions. There being only one copy of each catalytic domain means that they are used iteratively throughout the biosynthesis process. Furthermore, type-II PKS systems lack β-keto reductive domains that act on the intermediates during the elongation stage. As a result, the first immediate in a type-II PKS pathway is the poly-β-ketyl-S-ACP, a reactive intermediate that readily undergoes intramolecular aldol cyclizations to form the polycyclic aromatic scaffold. The regioselectivity of aldol con-

densations is controlled by additional accessory enzymes, such as aromatases and cyclases (**Figure 1.8C**).

Type-II PKS systems require the same minimal proteins as type-I PKS systems to support the biosynthesis of poly-β-ketones: a phosphopantetheinylated ACP, AT, and KS domains. Type II PKSs utilize malonyl-CoA as the sole building block, resulting in final products that are not substituted at the corresponding α-positions. Therefore, type-II PKS systems specifically use a malonyl-CoA:ACP acyltransferase (MAT) to load malonyl-CoA onto the ACP. BGCs encoding type-II PKSs tend not to have its own copy of the MAT gene, since the cellular housekeeping MAT enzyme used by bacterial FAS enzymes can also used for type-II polyketide biosynthesis; MAT recognizes ACP from both FAS and PKS for malonyl group transfer.[37,38] The KS in type-II PKSs is a heterodimer of KS_α and KS_β subunits, with KS_α subunit having the bona fide KS activity and an active-site cysteine that is also present in other KS enzymes or domains. KS_α catalyzes a thio-Claisen condensation between the poly-β-ketone intermediate acylated on the active site cysteine and the incoming malonyl-*S*-ACP, and the resulting elongated poly-β-ketyl-*S*-ACP is returned to the cysteine for another round of elongation. In the active site of KS_α, a product channel exists that originates from the cysteine to the dimer interface. As the polyketide grows by two carbons after each chain extension reaction, the product is "extruded" toward the dimer interface. Typical sizes of poly-β-ketones formed include octa- (16-C), nona- (18-C), deca- (20-C) and dodeca-ketides (24-C). The KS_β subunit, which is catalytically inactive and has a mutation at the active site cysteine, is also referred to as the chain-length factor (CLF). As its name suggests, the KS_β subunit has been shown through combinatorial biosynthesis, mutagenesis and X-ray structural analysis to play an important role in determining the length of the poly-β-ketone product, or the number of rounds of elongation by the KS_α-KS_β heterodimer.[39,40] The KS_β subunit helps define the size of the aforementioned product channel and thereby "controls" the chain length.

The assembly and tailoring stages of type-II polyketide biosynthesis are illustrated for the

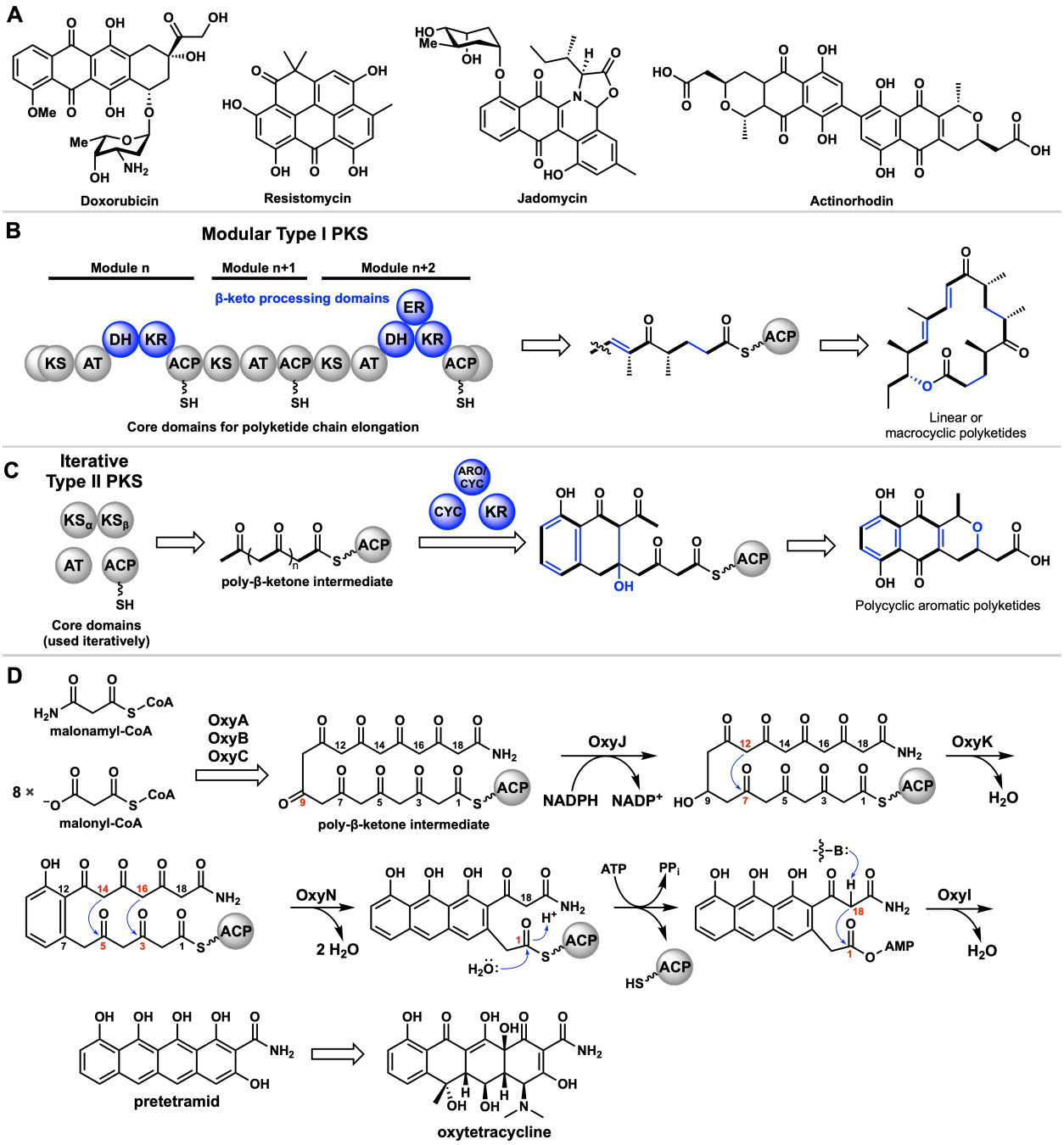


Figure 1.8. Biosynthesis of bacterial type-II polyketides. (A) Examples of aromatic type-II polyketides. Comparison of the domain organization of (B) modular type-I and (C) iterative type-II PKS systems. (D) Overview of the biosynthesis of oxytetracycline, a broad-spectrum inhibitor of bacterial protein synthesis.

oxytetracycline biosynthetic pathway in **Figure 1.8D**. Here, the starter unit that initiates the polymerization of acetate is malonamyl-CoA, which is unique to this pathway. Acetyl-

CoA is the most commonly-used starter unit to prime the KS_α, but other acyl groups may also be used. Following the KS_α priming, the KS_α-KS_β (OxyA-OxyB) and ACP (OxyC) catalyze eight rounds of thio-Claisen condensation to synthesize a 19-carbon poly-β-ketone intermediate. If released into solution without enzymatic control, this highly reactive intermediate will undergo different modes of intramolecular aldol cyclization between C_αs and β-ketones to yield shunt products. To control the reactivity of these poly-β-ketones, nature has encoded several tailoring enzymes for each type-II PKS.[41] In the oxytetracycline pathway, a KR (OxyJ) specifically reduces the C_9 ketone to an alcohol. The resulting alcohol is then captured by cyclases (CYCs)–for oxytetracycline, OxyK is the first ring cyclase that joins C_7 and C_{12}, and the second ring cyclase OxyN encloses the ring at C_5 and C_{14}. The cyclases bind their poly-β-ketyl-*S*-ACP substrates in a specific conformation to ensure the regioselectivity of C–C bond formation. Following the release of the tricyclic intermediate via hydrolysis, the terminal carboxylate group is adenylated to facilitate the formation of the fourth ring by OxyI, resulting in the fully aromatic tetracyclic pretetramid. The pretetramid is transformed to oxytetracycline via a series of tailoring steps involving hydroxylation, reductive amination, and methylation.

Just like oxytetracycline, the aromatic polyketides shown in **Figure 1.8A** can be accessed through similar chemical logic. Structural diversity is introduced via the choice of starter unit, chain length, site of ketone reduction, regioselectivity of aldol condensations, and post-cyclization tailoring. As seen in doxorubicin and jadomycin, many aromatic polyketides are glycosylated with deoxysugars by the action of glycosyltransferases. Sugar groups enhance the solubility of the core polyketide scaffolds and facilitate recognition of specific DNA sites for intercalation in the double helix.

1.4 Terpenoid Natural Products

The largest family of natural products (>50,000), terpenoids are rich in structural diversity and biological activities and are produced by organisms ranging from archaea and bacteria

to fungi, plants, and humans. Well-known terpenoids in human physiology include coenzyme Q10, the respiratory electron carrier in the citric acid cycle; retinal, a light-responsive co-factor that enables vision; and cholesterol, which is necessary for the biosynthesis of hormones and vitamin D but can also lead to heart disease in excess (**Figure 1.9A**). Terpenes are the key constituents of plant essential oils, and they also include key players in modern medicine such as taxol and artemisinin.

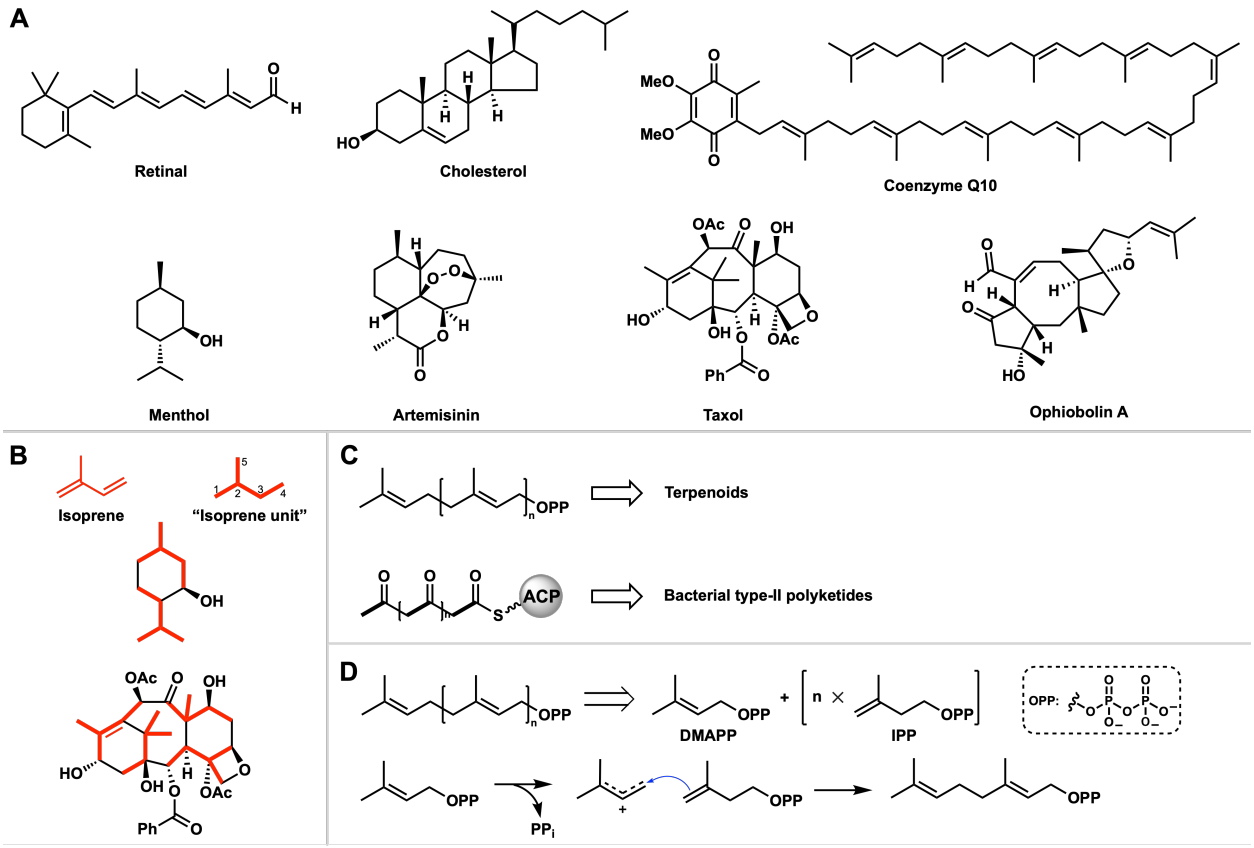


Figure 1.9. Overview of terpenoid natural products. (A) Examples of terpenoid natural products. (B) Isoprene units (red) in the core scaffolds of menthol and taxol. (C) Biosynthesis of terpenoids and bacterial type-II polyketides from linear achiral polymeric precursors. (D) Mechanism for the addition of IPP to DMAPP in the biosynthesis of isoprenyl pyrophosphates.

Prior to the inception of the concept of natural product biosynthetic machinery, early studies on terpenoid biogenesis were carried out by Leopold Ruzicka, who formulated the isoprene rule.[42] According to the isoprene rule, terpenoids are derived from two or more

isoprene units that are linked in a head-to-tail fashion. For example, the core scaffolds of menthol and taxol can be fragmented into 5-carbon isoprene units that are linked between C_1 and C_4 (**Figure 1.9B**). The isoprene rule cannot be universally applied to all terpenoids, since many undergo skeletal rearrangements and modifications in their core scaffolds during cyclization and tailoring. Oxidative modifications of terpenes can rearrange the carbon skeleton of hydrocarbon precursors and introduce oxygenated functional groups. For example, the cholesterol side-chain cleavage enzyme, or P450scc, removes a 6-carbon unit from cholesterol in an oxidative C–C bond cleavage reaction in the first step of the human steroid hormone biosynthesis,[43,44] and artemisinin is synthesized from its immediate precursor via a nonenzymatic oxidation and oxygen-insertion process.[45] However, the isoprene rule is useful for identifying potential terpenoid natural products upon inspection, and it is consistent with the terpenoid biosynthetic process as we currently understand.

The enzymatic machinery and chemistry underlying the assembly of terpenoid core scaffolds differ greatly from those of peptides and polyketides. The one parallel in biosynthetic logic between terpenoids and type-II polyketides is that both are synthesized via enzyme-mediated cyclizations of uniform achiral polymeric precursors (**Figure 1.9D**). Unlike type-II bacterial polyketide scaffolds, which consist largely of sp^2 carbons, terpenoids are rich in chiral centers that are introduced during the cyclization and tailoring steps. Some terpenoid molecules are predominantly linear (i.e. retinal, coenzyme Q10) while others are more elaborately cyclized (i.e. cholesterol, taxol).

Terpenoids are biosynthesized from isoprenyl pyrophosphate precursors that are assembled from 5-carbon dimethylallyl pyrophosphate (DMAPP) and isopentenyl-5-pyrophosphate (IPP) building blocks (**Figure 1.9D**). DMAPP acts as the electrophile in the DMAPP–IPP coupling reaction upon loss of the pyrophosphate group and formation of a resonance-stabilized carbocation. As the nucleophile, IPP quenches the carbocation at C_4 to form a new head-to-tail C-C bond. A single round of DMAPP–IPP coupling gives the 10-carbon geranyl pyrophosphate (GPP), the precursor to limonene and menthol (**Figure 1.10A**).[46]

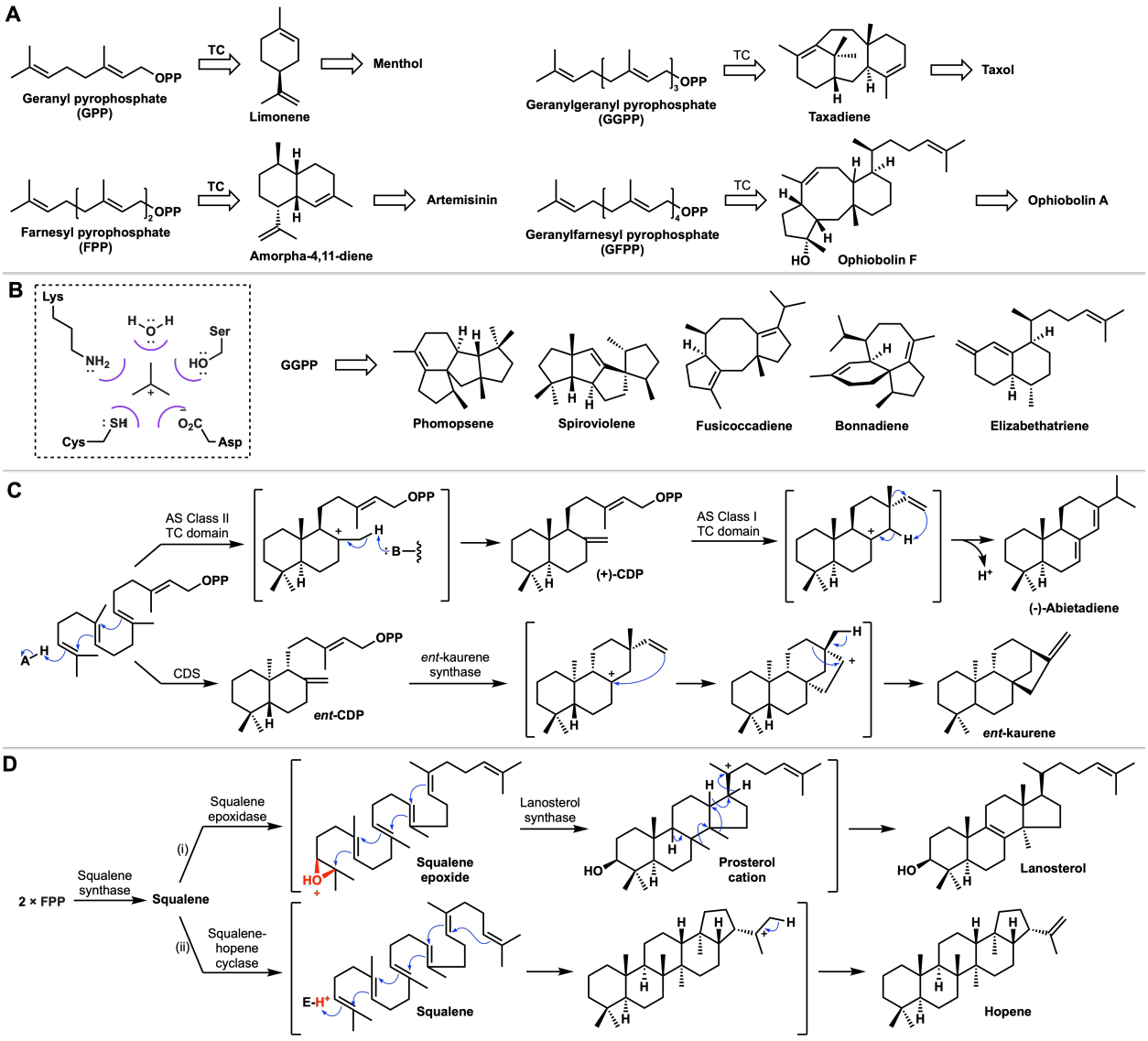


Figure 1.10. Terpene cyclases transform linear polyisoprenyl pyrophosphates into structurally diverse polycyclic scaffolds. (A) Biosynthetic precursors and intermediates for well-known terpenoid natural products. (B) TC active site shields the carbocation intermediate from nucleophiles and controls its conversion into specific terpenoid scaffolds. (C) Conversion of GGPP to (−)-abietadiene by the bifunctional abietadiene synthase (AS) and to ent-kaurene by copalyl diphosphate (CDP) synthase (CDS) follwed by kaurene synthase. AS and CDS make (+)-CDP and (−)-CDP from GGPP, respectively. (D) Electrophilic center to promote the cyclization cascade can be generated by (i) epoxidation or (ii) protonation.

GPP can also serve as electrophile in the next round of coupling with IPP, and this assembly process can continue step-wise, with each coupling step catalyzed by a different

enzyme. GPP coupling with IPP by the farnesyl pyrophosphate (FPP) synthase gives the 15-carbon FPP, which can be cyclized into sesquiterpenes such as the artemisinin-precursor amorpha-4,11-diene. Further elongation by GGPP synthase affords the 20-carbon geranylgeranyl pyrophosphate (GGPP), which can be cyclized into diterpenes, and GFPP synthase gives the 25-carbon geranylfarnesyl pyrophosphate (GFPP) for sesterterpenes, etc. As many as 11 units of isoprene can be polymerized to produce bactoprenols that act as carrier lipids in cell wall assembly in bacteria.[47]

In addition to the head-to-tail coupling of isoprene units, head-to-head coupling between two terpenoids can also take place. Squalene synthase catalyzes the coupling of two FPP molecules to make squalene, the 30-carbon precursor to cholesterol and related hormone molecules.[48] Phytoene synthase catalyzes the head-to-head coupling of two GGPP molecules in the same manner, going through a cyclopropane intermediate to make the 40-carbon phytoene.[49] Phytoene can be isomerized and desaturated to give lycopene, which is the biosynthetic precursor to carotenes and retinal (vitamin A).

Once the linear isoprenyl precursors (GPP, FPP, GGPP, etc.) are synthesized, they can undergo the cation-driven cyclization and rearrangement by terpene cyclases (TCs).[50,51] TCs are intramolecular C–C bond-forming catalysts that contribute greatly to the structural diversity of terpenoids. TCs can generate structurally diverse core scaffolds from the same achiral linear precursor, capitalizing on the reactivity of the allylic carbocation that forms through dissociation of the terminal pyrophosphate. The carbocation can undergo several fates when generated under aqueous environment,such as hydride elimination to generate an olefin, quenching by water to generate an alcohol, and/or bond migration to form rings. In the absence of enzymatic catalysis, it is difficult to control the fate of the carbocation, and many products can form non-selectively. To overcome these limitations, active sites of TCs have a hydrophobic lining that shelters the carbocation from water and other endogenous biochemical nucleophiles to minimize premature quenching of the carbocation (**Figure 1.10B**). The cyclization pathway is partially set by the three-dimensional shape of the active

site cavity, which determines substrate conformation and relative orientation of its olefins that participate in the C–C bond-forming cascade. In this way, a specific TC can suppress the formation of unintended products and lead to exclusive formation of the desired cyclized product. Nature has a large assortment of TCs from diverse organisms, and as a result, a large spectrum of cyclized terpene scaffolds can be generated (**Figure 1.10B**), with each scaffold guided by a dedicated TC in its cyclization pathway from a linear precursor. For example, the α-terpinyl cation can be generated from the simple GPP, which is enzymatically converted to one of at least 14 different monoterpene products, including limonene and α-pinene. The product profile of a TC can be altered through rational mutagenesis of the active site to synthesize differentially cyclized terpenoids, as illustrated in alteration of the product profile of β-humulene synthase, in which 2 to 5 mutations led to the production of diverse sesquiterpenes from FPP.[52]

The terpene cyclization process can be initiated via two distinct mechanisms for isoprenyl precursors with a terminal pyrophosphate group (**Figure 1.10C**).[50,51] In class I TCs, the isoprenyl carbocation is generated upon Mg^{2+}-dependent departure of the pyrophosphate group, which leaves behind a resonance-stabilized allylic carbocation for further cyclization within the active site. One such example is the taxadiene synthase involved in the biosynthesis of the anticancer plant diterpene taxol, which facilitates departure of pyrophosphate from GGPP and subsequent macrocyclization of the cation that forms taxadiene. On the other hand, class II TCs generate carbocations by protonating an existing olefin in the substrate prior to pyrophosphate elimination. The best example of a class II TC is ent-copalyl-diphosphate (CDP) synthase (CDS), which cyclizes GGPP into bicyclic ent-CDP via protonation of the terminal olefin (**Figure 1.10C**). Some diterpene cyclases are bifunctional and contain both class I and class II active sites in a single enzyme,[53,54] like the abietadiene synthase, which first generates CDP using the class II TC domain, followed by cyclization into abietadiene via the allylic cation generated by the class I TC domain. Interestingly, abietadiene synthase generates (+)-CDP in contrast to the ent-CDP generated by CDP synthase,[55] which highlights the power of TCs as enantioselective catalysts in natural product biosynthesis.

The triterpenoid squalene is cyclized into polycyclic five- or six-ring compounds in nearly all kingdoms of life. In bacteria, squalene can be cyclized into hopene. Squalene is the precursor to β-amyrin in plants, whereas in animals, squalene is cyclized into lanosterol, the precursor to sterol molecules such as cholesterol, testosterone, and estradiol. Since squalene is formed from a head-to-head dimerization of two FPP molecules, it lacks a labile anionic leaving group. Cyclizing squalene requires a similar approach to class II TCs with some key differences. The cyclization to hopene by squalene-hopene synthase is initiated by protonation of a terminal olefin by a strong Brønsted acid, facilitated by an aspartic acid residue at the end of the TC hydrophobic channel. The channel also orients the squalene in an all pre-chair conformation for the cyclization cascade. On the other hand, cyclization of squalene into lanosterol occurs via two distinct enzymatic steps. In the first step, the same terminal olefin is stereospecifically epoxidized by squalene epoxidase to give squalene 3*S*-2,3-epoxide. Both the protonation of the terminal olefin by squalene-hopene synthase and epoxidation by squalene epoxidase serve to generate an electrophilic terminus to promote the carbocation-mediated cyclization. The squalene epoxide is protonated within the cyclase active site, and the internal olefins are aligned in a pre chair-boat-chair conformation to allow the electrons to move as shown in **Figure 1.10D** to form the tetracyclic sterol structure. The prosterol cation intermediate generated upon the initial cyclization cascade then undergoes additional round of hydride and methyl shifts and hydride elimination to arrive at lanosterol.

Together with prenyltransferase enzymes that synthesize isoprenyl precursors of different lengths, TCs contribute greatly to the structural diversity of terpenoid natural products. A single isoprenyl precursor can be morphed into multiple scaffolds with different structures based on how the olefins are arranged relative to one another within a TC active site. In the same way that cyclases are critical for the structural diversity of type-II polyketides, terpenoids also rely heavily on their cyclases for structural diversity. A notable similarity between bacterial type-II polyketides and terpenoids is that both classes of natural products are assembled from a single achiral building block, and the cyclase enzymes in their biosyn-

thetic pathways help shape the molecular architecture of the final natural products. Finally, also contributing to the vast structural diversity of terpenoid natural products are the tailoring enzymes that decorate the barren aliphatic scaffolds with heteroatomic functional groups seen in the final natural products.[56]

1.5 Hybrid & Un-Natural Products

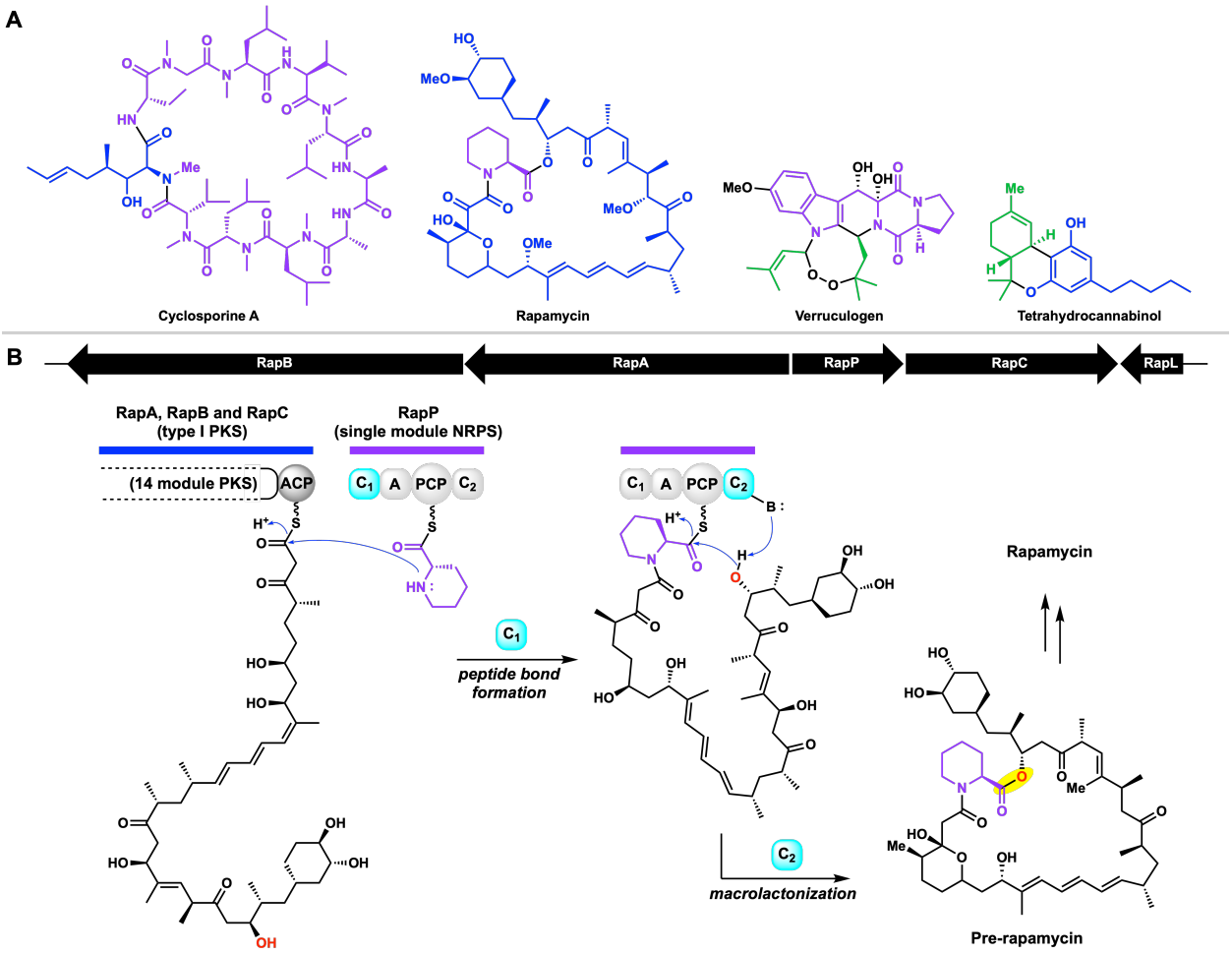


Figure 1.11. Biosynthesis of hybrid natural products. (A) Examples of hybrid natural products. The purple, blue, and green fragments are derived from the non-ribosomal peptide, polyketide, and terpenoid pathways, respectively. (B) Mechanism of acyl transfer of the polyketide chain from PKS to the NPRS for incorporation of pipecolic acid during rapamycin biosynthesis.

Hybrid natural products broadly refer to molecules whose core scaffolds are built by

enzymes from multiple natural product families (**Figure 1.11A**). Hybrid natural products are synthesized when the product of one family of biosynthetic enzymes (e.g. polyketide) is used as a starting material or intermediate for another biosynthetic enzyme from a different family. A well-known example of a hybrid natural product in medicine is cyclosporine A. An immunosuppressive non-ribosomal undecapeptide isolated from a soil fungus, cyclosporine A contains a nonproteinogenic (4*R*)-4-[(*E*)-2-butenyl]-4-methyl-L-threonine, which is synthesized by a designated PKS in the cyclosporine A biosynthetic pathway.[57] Similarly, rapamycin is the product of a PKS-NRPS hybrid assembly line that incorporates pipecolic acid, a noncanonical amino acid, into a polyketide (**Figure 1.11B**).[58] Pipecolic acid is synthesized from lysine by RapL, and it is accepted by the single-module NRPS RapP. Once the polyketide portion is fully synthesized, the C1 domain in RapP transfers the polyketide chain from the PKS RapABC onto the amino group of pipecolic acid-*S*-RapP, resulting in an amide linkage between the polyketide and peptide components. Lastly, the C2 domain catalyzes the intramolecular addition of an alcohol from the polyketide chain to the pipecolate thioester, releasing pre-rapamycin as a PKS-NRPS hybrid macrolactone. Here, the C1 domains of NRPS modules catalyze the addition between the amino group of an amino acid and a thioester, which is peptide-*S*-PCP in the case of non-hybrid NRPS systems, and a polyketide-*S*-ACP for PKS-NRPS hybrid systems. Nature takes advantage of the fact that PKS and NRPS biosynthesis both proceed via thioester intermediates on phosphopantetheinylated carrier proteins to synthesize PKS-NRPS hybrid products such as rapamycin and epothilone (**Figure 1.11A**).

With increased understanding of how different families of natural products are synthesized in nature, there has been a growing interest in re-harnessing the biosynthetic machineries to synthesize new "un-natural" natural products. Now that we understand the functions of key domains in a modular PKS, it would be great to be able to design custom PKSs to synthesize polyketide products of our choosing by adjoining the necessary domains as they occur in native systems. Despite our current understanding of the organization of modular biosynthetic systems and functions of individual catalytic domains, however, we have a

long way to go before we can synthesize "designer polyketides" by simply assembling the modules with the necessary PKS domains. Modular biosynthetic systems are not quite like enzymatic beads on a string, and there are a lot to consider in terms of domain substrate specificity and domain-domain interactions that may influence the outcome of PKS products.

Re-harnessing biosynthetic enzymes as biocatalysts is becoming increasingly popular in organic synthesis due to their selectivity and greenness,[59–61] and with better tools for genome annotation and protein engineering, we anticipate that more enzymes from natural product biosynthetic pathways will find their way to synthetic chemistry hoods.

1.6 References

(1) Maplestone, R. A.; Stone, M. J.; Williams, D. H. *Gene* **1992**, *115*, 151–157.

(2) Gaynes, R. *Emerg. Infect. Dis.* **2017**, *23*, 849–853.

(3) Washington, J. A.; Wilson, W. R. *Mayo Clinic Proceedings* **1985**, *60*, 189–203.

(4) Krishna, S.; Bustamante, L.; Haynes, R. K.; Staines, H. M. *Trends in Pharmacological Sciences* **2008**, *29*, 520–527.

(5) Newman, D. J.; Cragg, G. M. *Journal of Natural Products* **2020**, *83*, 770–803.

(6) Davies, H. M. L.; Itami, K.; Stoltz, B. M. *Chemical Society Reviews* **2018**, *47*, 7828–7829.

(7) Murakami, M.; Ito, Y. In *Activation of Unreactive Bonds and Organic Synthesis*, Murai, S. et al., Eds.; Topics in Organometallic Chemistry; Springer: Berlin, Heidelberg, 1999, pp 97–129.

(8) Cernak, T.; Dykstra, K. D.; Tyagarajan, S.; Vachal, P.; Krska, S. W. *Chemical Society Reviews* **2016**, *45*, 546–576.

(9) Abrams, D. J.; Provencher, P. A.; Sorensen, E. J. *Chemical Society Reviews* **2018**, *47*, 8925–8967.

(10) Young, I. S.; Baran, P. S. *Nature Chemistry* **2009**, *1*, 193–205.

(11) Hui, C.; Chen, F.; Pu, F.; Xu, J. *Nature Reviews Chemistry* **2019**, *3*, 85–107.

(12) Harmange Magnani, C. S.; Thach, D. Q.; Haelsig, K. T.; Maimone, T. J. *Accounts of Chemical Research* **2020**, *53*, 949–961.

(13) Fischbach, M. A.; Walsh, C. T. *Chemical Reviews* **2006**, *106*, 3468–3496.

(14) Yoder, R. A.; Johnston, J. N. *Chemical Reviews* **2005**, *105*, 4730–4756.

(15) Bulger, P. G.; Bagal, S. K.; Marquez, R. *Natural Product Reports* **2008**, *25*, 254–297.

(16) Lautru, S.; Deeth, R. J.; Bailey, L. M.; Challis, G. L. *Nature Chemical Biology* **2005**, *1*, 265–269.

(17) Challis, G. L. *Microbiology* **2008**, *154*, 1555–1569.

(18) Arnison, P. G. et al. *Natural Product Reports* **2012**, *30*, 108–160.

(19) Rother, M.; Krzycki, J. A. Selenocysteine, Pyrrolysine, and the Unique Energy Metabolism of Methanogenic Archaea, en, Review Article, 2010.

(20) Willey, J. M.; van der Donk, W. A. *Annual Review of Microbiology* **2007**, *61*, 477–501.

(21) Challis, G. L.; Ravel, J.; Townsend, C. A. *Chemistry & Biology* **2000**, *7*, 211–224.

(22) Koglin, A.; Mofid, M. R.; Löhr, F.; Schäfer, B.; Rogov, V. V.; Blum, M.-M.; Mittag, T.; Marahiel, M. A.; Bernhard, F.; Dötsch, V. *Science* **2006**, *312*, 273–276.

(23) Luo, L.; Kohli, R. M.; Onishi, M.; Linne, U.; Marahiel, M. A.; Walsh, C. T. *Biochemistry* **2002**, *41*, 9184–9196.

(24) Cane, D. E.; Walsh, C. T.; Khosla, C. *Science* **1998**, *282*, 63–68.

(25) Thirlway, J.; Lewis, R.; Nunns, L.; Al Nakeeb, M.; Styles, M.; Struck, A.-W.; Smith, C. P.; Micklefield, J. *Angewandte Chemie International Edition* **2012**, *51*, 7181–7184.

(26) Kries, H.; Wachtel, R.; Pabst, A.; Wanner, B.; Niquille, D.; Hilvert, D. *Angewandte Chemie International Edition* **2014**, *53*, 10105–10108.

(27) Birch, A. J.; Massy-Westropp, R. A.; Moye, C. J. *Australian Journal of Chemistry* **1955**, *8*, 539–544.

(28) Walsh, C. T.; Wencewicz, T. *Antibiotics: Challenges, Mechanisms, Opportunities* **2016**, 320–342.

(29) Barker, H. A.; Kamen, M. D.; Bornstein, B. T. *Proceedings of the National Academy of Sciences* **1945**, *31*, 373–381.

(30) Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. *Science* **1991**, *252*, 675–679.

(31) Cundliffe, E.; Bate, N.; Butler, A.; Fish, S.; Gandecha, A.; Merson-Davies, L. *Antonie van Leeuwenhoek* **2001**, *79*, 229–234.

(32) Bisang, C.; Long, P. F.; Corte´s, J.; Westcott, J.; Crosby, J.; Matharu, A.-L.; Cox, R. J.; Simpson, T. J.; Staunton, J.; Leadlay, P. F. *Nature* **1999**, *401*, 502–505.

(33) Khayatt, B. I.; Overmars, L.; Siezen, R. J.; Francke, C. *PLOS ONE* **2013**, *8*, e62136.

(34) Keatinge-Clay, A. T. *Chemistry & Biology* **2007**, *14*, 898–908.

(35) Khosla, C.; Gokhale, R. S.; Jacobsen, J. R.; Cane, D. E. *Annual Review of Biochemistry* **1999**, *68*, 219–253.

(36) McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Ashley, G. *Proceedings of the National Academy of Sciences* **1999**, *96*, 1846–1851.

(37) Summers, R. G.; Ali, A.; Shen, B.; Wessel, W. A.; Hutchinson, C. R. *Biochemistry* **1995**, *34*, 9389–9402.

(38) Carreras, C. W.; Khosla, C. *Biochemistry* **1998**, *37*, 2084–2088.

(39) Tang, Y.; Tsai, S.-C.; Khosla, C. *Journal of the American Chemical Society* **2003**, *125*, 12708–12709.

(40) Keatinge-Clay, A. T.; Maltby, D. A.; Medzihradszky, K. F.; Khosla, C.; Stroud, R. M. *Nature Structural & Molecular Biology* **2004**, *11*, 888–893.

(41) Zhang, W.; Tang, Y. In *Methods in Enzymology*; Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates, Vol. 459; Academic Press: 2009, pp 367–393.

(42) Ruzicka, L. *Experientia* **1953**, *9*, 357–367.

(43) Chatuphonprasert, W.; Jarukamjorn, K.; Ellinger, I. *Frontiers in Pharmacology* **2018**, *9*, DOI: 10.3389/fphar.2018.01027.

(44) Spinello, A.; Ritacco, I.; Magistrato, A. *Catalysts* **2019**, *9*, 81.

(45) Wallaart, T. E.; Pras, N.; Quax, W. J. *Journal of Natural Products* **1999**, *62*, 1160–1162.

(46) Turner, G. W.; Croteau, R. *Plant Physiology* **2004**, *136*, 4215–4227.

(47) El Ghachi, M.; Howe, N.; Huang, C.-Y.; Olieric, V.; Warshamanage, R.; Touzé, T.; Weichert, D.; Stansfeld, P. J.; Wang, M.; Kerff, F.; Caffrey, M. *Nature Communications* **2018**, *9*, 1078.

(48) Pandit, J.; Danley, D. E.; Schulte, G. K.; Mazzalupo, S.; Pauly, T. A.; Hayward, C. M.; Hamanaka, E. S.; Thompson, J. F.; Harwood, H. J. *Journal of Biological Chemistry* **2000**, *275*, 30610–30617.

(49) Dogbo, O.; Laferriére, A.; D'Harlingue, A.; Camara, B. *Proceedings of the National Academy of Sciences* **1988**, *85*, 7054–7058.

(50) Christianson, D. W. *Chemical Reviews* **2006**, *106*, 3412–3442.

(51) Christianson, D. W. *Chemical Reviews* **2017**, *117*, 11570–11648.

(52) Cane, D. E. *Nature Chemical Biology* **2006**, *2*, 179–180.

(53) Peters, R. J.; Ravn, M. M.; Coates, R. M.; Croteau, R. B. *Journal of the American Chemical Society* **2001**, *123*, 8974–8978.

(54) Zhou, K.; Gao, Y.; Hoy, J. A.; Mann, F. M.; Honzatko, R. B.; Peters, R. J. *Journal of Biological Chemistry* **2012**, *287*, 6840–6850.

(55) Hayashi, K.-i.; Kawaide, H.; Notomi, M.; Sakigi, Y.; Matsuo, A.; Nozaki, H. *FEBS Letters* **2006**, *580*, 6175–6181.

(56) Denicourt-Nowicki, A.; Rauchdi, M.; Ait Ali, M.; Roucoux, A. *Catalysts* **2019**, *9*, 893.

(57) Offenzeller, M.; Santer, G.; Totschnig, K.; Su, Z.; Moser, H.; Traber, R.; Schneider-Scherzer, E. *Biochemistry* **1996**, *35*, 8401–8412.

(58) Schwecke, T.; Aparicio, J. F.; Molnár, I.; König, A.; Khaw, L. E.; Haydock, S. F.; Oliynyk, M.; Caffrey, P.; Cortés, J.; Lester, J. B. *Proceedings of the National Academy of Sciences* **1995**, *92*, 7839–7843.

(59) Li, J.; Amatuni, A.; Renata, H. *Current Opinion in Chemical Biology* **2020**, *55*, 111–118.

(60) Sandoval, B. A.; Hyster, T. K. *Current Opinion in Chemical Biology* **2020**, *55*, 45–51.

(61) Zetzsche, L. E.; Narayan, A. R. H. *Nature Reviews Chemistry* **2020**, 1–13.

CHAPTER 2

Fungal Highly-Reducing Polyketide Synthases

Fungal iterative polyketide synthases (PKSs) are biosynthetic enzymes that synthesize a diverse group of secondary metabolites, many of which have complex structures and desirable biological activities.[1–3] They are broadly classified into three categories based on their β-keto processing domains and structures of associated products (**Figure 2.1**). Non-reducing PKSs (NRPKSs) lack embedded reductive domains and synthesize aromatic polyketides such as orsellinic acid and alternariol,[4] while partially-reducing PKSs (PRPKSs) catalyze a limited number of reductions to synthesize the prototypical 6-methylsalicylic acid and dihydrocoumarin compounds such as mellein.[5] The topic of this chapter, highly-reducing PKSs (HRPKSs), have the complete or nearly complete ensemble of β-keto processing domains and synthesize compounds such as hypocreolide, botcinic acid, and nafuredin, among many others.

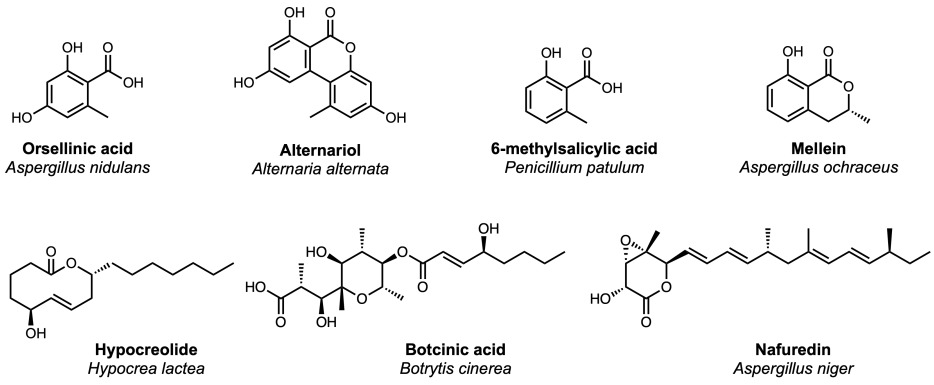


Figure 2.1. Select examples of polyketides synthesized by fungal iterative PKSs.

Similar to bacterial type-II PKSs, fungal iterative PKSs contain a single copy of catalytic domains that are used throughout the course of the biosynthesis. Due to the lack

of co-linearity in fungal HRPKSs, their cannot be predicted from protein sequences alone, unlike the bacterial type-I PKS systems.[6] The organization of domains in fungal HRPKSs does not provide any information that we currently understand as to how their activities are coordinated (# of iteration, combinations of β-keto processing steps, etc.). The iterative programming underlies the vast structural complexity and diversity of metabolites produced by fungal HRPKSs, making them one of the most enigmatic class of enzymes in microbial secondary metabolism.

How is the sequence of domain activities programmed in an iterative HRPKS and reliably executed to afford a precisely-crafted product? The coordination of HRPKS domain activities is critical for synthesizing products that are properly groomed for post-PKS modifications, which are often catalyzed by dedicated enzymes in biosynthetic pathways and essential for the biological activities of the final natural products. Significant advances have been made in the last decade to better understand their programming rules, motivated by the ultimate goal of fully cracking the iterative HRPKS code to predict their products from protein sequences. This chapter provides a brief overview of HRPKS programming and highlights recent advances in our understanding of the biosynthesis of HRPKS-associated fungal natural products.

2.1 General Features of Fungal HRPKSs

2.1.1 HRPKS Domains and their Architectures

Fungal HRPKSs are multidomain enzymes that consist of linearly-fused catalytic domains (**Figure 2.2A**), resembling mammalian fatty acid synthases (mFASs, (**Figure 2.2B**)) and individual modules from bacterial PKSs (**Figure 2.2C**). All fungal HRPKSs are large, with the smallest HRPKS at 200 kDa and the lovastatin nonaketide synthase (LovB) being one of the largest at approximately 335 kDa.

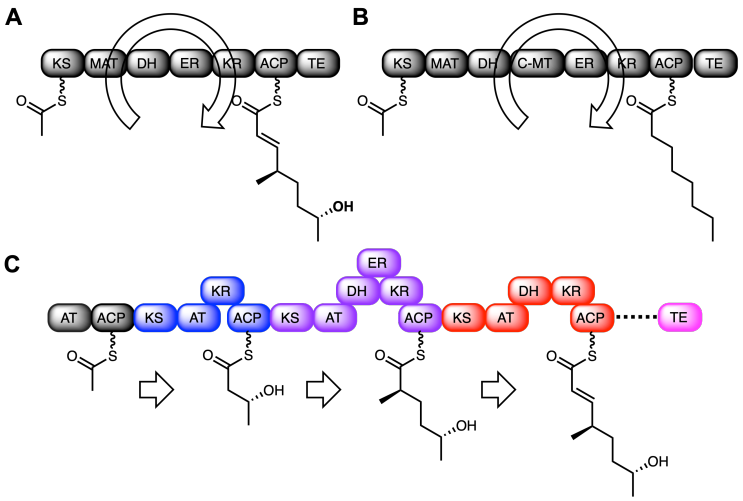


Figure 2.2. Domain organization of (A) *m*FAS, (B) HRPKS, and (C) bacterial modular PKS.

All three classes of fungal iterative PKSs contain the minimal PKS domains. The minimal PKS domains are capable of synthesizing the triketide α-pyrone as the simplest isolable product. In addition to the minimal domains, HRPKSs also have KR and DH domains, and the ER domain is also present in some HRPKSs. As will be noted later, many HRPKSs have an inactive ER domain (ER^{o}) and rely on *trans*-acting auxiliary ER enzymes for the programmed enoyl reduction events. Many HRPKSs also have a *C*-MT domain that methylates the C_{α} of the growing polyketide chain prior to β-ketoreduction.

Although iterative HRPKSs are very similar to *m*FASs in terms of domain organization and activities (**Figures 2.2A,B**), the two synthases differ drastically in the programming of their β-keto processing domains. The KR, DH, and ER domains of *m*FAS are programmed to be active in every iteration to produce saturated fatty acids such as the C-16 palmitic acid; *m*FAS can thus be considered a fully-reducing PKS. On the other hand, the β-keto processing domains of HRPKSs operate to varying extents in each iteration. A ketide unit may be fully reduced to a methylene by the KR, DH, and ER domains; or the β-keto processing may stop at either the KR or DH to give a β-alcohol or an enoyl thioester. The selective use of reductive domains by HRPKSs generates functionalities such as alcohols, alkenes, and

polyenes, which are often elaborated into more complex structures by tailoring enzymes in fungal polyketide biosynthetic pathways. Furthermore, most fungal HRPKSs are capable of synthesizing branched polyketides with the *C*-MT domain, in contrast to *m*FASs that synthesize only linear fatty acids. As will be noted later, α-methylation often serves as a biosynthetic checkpoint for HRPKSs and is important for post-PKS modifications. Lack of proper MT activity can compromise the fidelity of HRPKS programming and lead to shunt PKS products.[7,8]

In addition to the programming of β-keto processing domains, iterative HRPKSs and *m*FASs also differ notably in their mechanisms of product release. *m*FASs contain a dedicated thioesterase (TE) domain that hydrolyzes fatty acids from the ACP once they reach their target lengths.[9] However, most HRPKSs do not have a fused TE domain for product turnover. Some HRPKS products are offloaded by auxiliary enzymes,[10,11] and others undergo reductive release as aldehydes.[12] Many HRPKS products are also transferred as starter units for partnering biosynthetic enzymes such as NRPKSs and non-ribosomal peptide synthases (NRPSs), which introduce further structural complexity and diversity to HRPKS-associated natural products (**Figure 2.3**).

2.1.2 Molecules synthesized by HRPKSs

Even though fungal HRPKSs are highly homologous and their domains perform a limited set of chemical transformations, HRPKS-associated natural products are structurally diverse and range from linear fatty acids to polyenes and decalins, and families of compounds such as resorcylic acid lactones (RALs), 2-pyridone alkaloids, and cytochalasins (**Figure 2.3**). In rare cases, a single HRPKS synthesizes two distinct polyketides that are processed differentially by downstream enzymes.[13,14] Some HRPKSs also synthesize precursors to non-proteinogenic amino acids that are utilized by NRPSs and PKS-NRPS hybrid enzymes.[15,16]

Among the fungal HRPKS-associated natural products, the most famous is perhaps

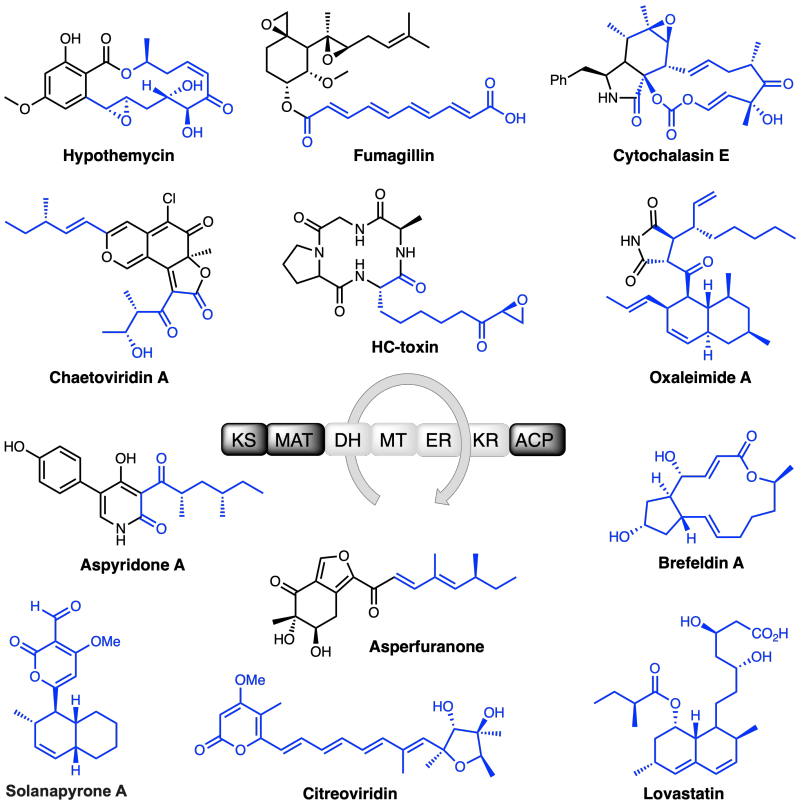


Figure 2.3. Select examples of polyketides known to be associated with fungal HRPKSs. The highly-reduced polyketide portions of the molecules are highlighted in blue.

the cholesterol-lowering drug lovastatin, whose biosynthesis has been studied in great detail.[7,10,17–22] Lovastatin is the product of a collaboration between two HRPKSs–LovB, which synthesizes the decalin-containing dihydromonacolin L; and LovF, which synthesizes the diketide α-methylbutyrate that is transferred to monacolin J acid by the acyltransferase LovD. At first glance, it is difficult to grasp how a HRPKS with a single set of domains, together with an auxiliary ER (LovC), might be programmed to coordinate its β-keto processing activities. The programming of LovB involves eight iterations of polyketide chain elongation and β-keto processing that culminate in dihydromonacolin L acid (**Figure 2.4**).

While the extent of β-keto processing in LovB iterations leading up to the hexaketide may seem inconsequential, a careful examination of programming logic reveals that LovB strategically installs a diene and an olefin for an intramolecular Diels-Alder cycloaddition

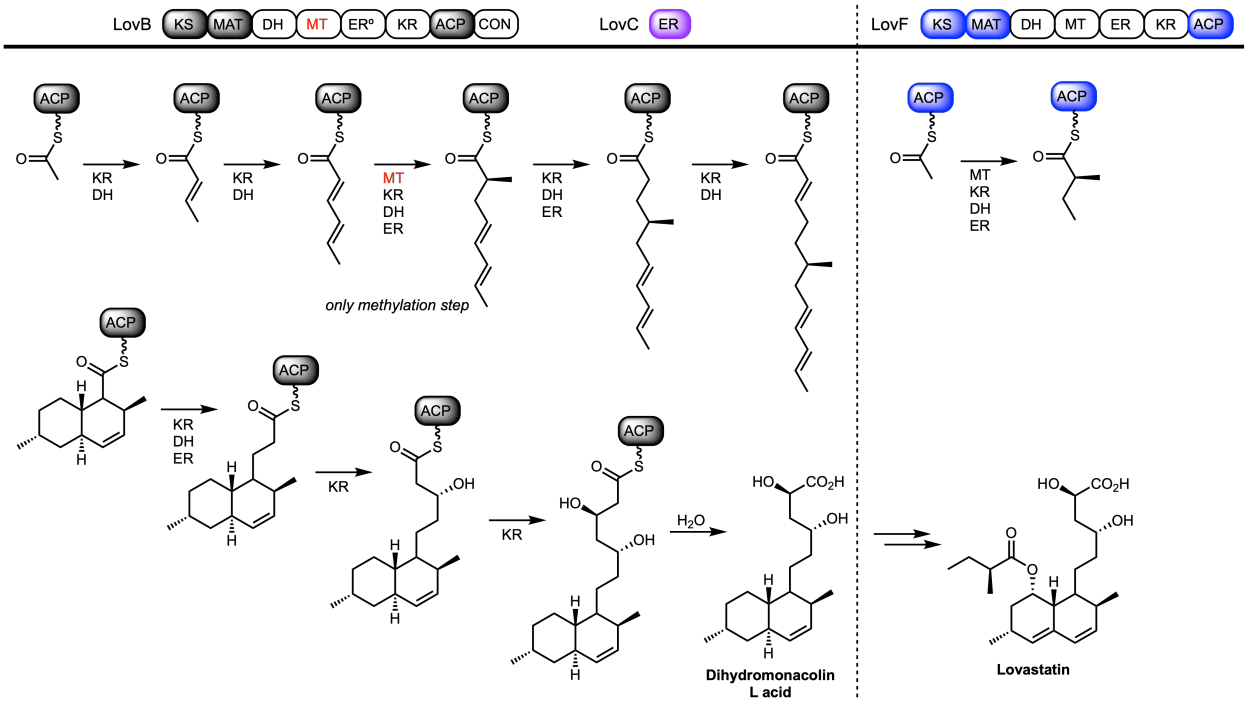


Figure 2.4. The programming of LovB, LovC, and LovF in the biosynthesis of lovastatin.

to afford the decalin ring. Following the Diels-Alder cycloaddition, three more iterations of chain elongation and β-keto processing give dihydromonacolin L acid, the lack of dehydration in the last two iterations setting up the molecular warhead that mimics the 3-hydroxy-3-methylglutarate portion of HMG-CoA. Overall, the single set of domains in LovB and LovC catalyze 26 transformations to synthesize the final nonaketide. This is in stark contrast to modular bacterial PKSs, which would in theory require eight modules, each with all the necessary β-keto processing domains, to achieve the same biosynthesis as LovB.

Notwithstanding this impressive complexity, most fungal HRPKSs have the same domain architecture as LovB but synthesize completely different products. Case in point is the contrast between LovB and LovF. Even though the two HRPKSs in the lovastatin pathway have very similar domain organizations, the size and structures of their products differ drastically. Unlike the nonaketide synthase LovB, LovF is essentially a non-iterative HRPKS whose domains all function once to synthesize the C_4 α-methylbutyrate. What makes LovB

and LovF function so differently despite the similarity in their domain organizations? This is the fundamental question that needs to be answered before we can attempt to predict the structures of HRPKS products from protein sequences.

2.2 References

(1) Schueffler, A.; Anke, T. *Natural Product Reports* **2014**, *31*, 1425–1448.

(2) Schor, R.; Cox, R. *Natural Product Reports* **2018**, *35*, 230–256.

(3) Zhang, X.; Li, S.-J.; Li, J.-J.; Liang, Z.-Z.; Zhao, C.-Q. *Marine Drugs* **2018**, *16*, 194.

(4) Zhou, H.; Li, Y.; Tang, Y. *Natural Product Reports* **2010**, *27*, 839–868.

(5) Sun, H.; Ho, C. L.; Ding, F.; Soehano, I.; Liu, X.-W.; Liang, Z.-X. *Journal of the American Chemical Society* **2012**, *134*, 11924–11927.

(6) Weissman, K. J. In *Methods in Enzymology*; Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates, Vol. 459; Academic Press: 2009, pp 3–16.

(7) Ma, S. M.; Li, J. W.-H.; Choi, J. W.; Zhou, H.; Lee, K. K. M.; Moorthie, V. A.; Xie, X.; Kealey, J. T.; Silva, N. A. D.; Vederas, J. C.; Tang, Y. *Science* **2009**, *326*, 589–592.

(8) Tsunematsu, Y.; Fukutomi, M.; Saruwatari, T.; Noguchi, H.; Hotta, K.; Tang, Y.; Watanabe, K. *Angewandte Chemie International Edition* **2014**, *53*, 8475–8479.

(9) Chakravarty, B.; Gu, Z.; Chirala, S. S.; Wakil, S. J.; Quiocho, F. A. *Proceedings of the National Academy of Sciences* **2004**, *101*, 15567–15572.

(10) Xu, W.; Chooi, Y.-H.; Choi, J. W.; Li, S.; Vederas, J. C.; Da Silva, N. A.; Tang, Y. *Angewandte Chemie International Edition* **2013**, *52*, 6472–6475.

(11) Zabala, A. O.; Chooi, Y.-H.; Choi, M. S.; Lin, H.-C.; Tang, Y. *ACS Chemical Biology* **2014**, *9*, 1576–1586.

(12) Ugai, T.; Minami, A.; Fujii, R.; Tanaka, M.; Oguri, H.; Gomi, K.; Oikawa, H. *Chemical Communications* **2015**, *51*, 1878–1881.

(13) Zhu, X.; Yu, F.; Li, X.-C.; Du, L. *Journal of the American Chemical Society* **2007**, *129*, 36–37.

(14) Winter, J. M.; Sato, M.; Sugimoto, S.; Chiou, G.; Garg, N. K.; Tang, Y.; Watanabe, K. *Journal of the American Chemical Society* **2012**, *134*, 17900–17903.

(15) Offenzeller, M.; Santer, G.; Totschnig, K.; Su, Z.; Moser, H.; Traber, R.; Schneider-Scherzer, E. *Biochemistry* **1996**, *35*, 8401–8412.

(16) Sato, M.; Dander, J. E.; Sato, C.; Hung, Y.-S.; Gao, S.-S.; Tang, M.-C.; Hang, L.; Winter, J. M.; Garg, N. K.; Watanabe, K.; Tang, Y. *Journal of the American Chemical Society* **2017**, *139*, 5317–5320.

(17) Kennedy, J.; Auclair, K.; Kendrew, S. G.; Park, C.; Vederas, J. C.; Hutchinson, C. R. *Science* **1999**, *284*, 1368–1372.

(18) Xie, X.; Watanabe, K.; Wojcicki, W. A.; Wang, C. C. C.; Tang, Y. *Chemistry & Biology* **2006**, *13*, 1161–1169.

(19) Ma, S. M.; Tang, Y. *The FEBS Journal* **2007**, *274*, 2854–2864.

(20) Ames, B. D.; Nguyen, C.; Bruegger, J.; Smith, P.; Xu, W.; Ma, S.; Wong, E.; Wong, S.; Xie, X.; Li, J. W.-H.; Vederas, J. C.; Tang, Y.; Tsai, S.-C. *Proceedings of the National Academy of Sciences* **2012**, *109*, 11144–11149.

(21) Cacho, R. A.; Thuss, J.; Xu, W.; Sanichar, R.; Gao, Z.; Nguyen, A.; Vederas, J. C.; Tang, Y. *Journal of the American Chemical Society* **2015**, *137*, 15688–15691.

(22) Campbell, C. D.; Vederas, J. C. *Biopolymers* **2010**, *93*, 755–763.

CHAPTER 3

Biosynthesis of (–)-sambutoxin en route to funiculosin

3.1 Introduction

4-Hydroxy-2-pyridone alkaloids are a family of fungal natural products exhibiting an impressive range of biological activities and structural diversity (**Figure 3.1A**). They have the general structure shown in **Figure 3.1B**, in which the R_1 group originates from the side chain of either phenylalanine or tyrosine, and R_2 is synthesized by the iterative HRPKS portion of PKS-NRPSs that synthesizes their tetramic acid precursors. They are further classified as the 3-acyl-, 3-alkyl- and 3-ether-modified subfamilies based on the linkage at C_3—while the biosynthesis of 3-acyl- (e.g., ilicicolin H) and 3-alkyl subfamilies (e.g., leporins, citridones, and pyridoxatin) have been well-studied with recent emphasis on ring-forming reactions,[1–3] much less is known about the biosynthesis of the 3-ether subfamily featuring a tetrahydropyran motif such as (–)-sambutoxin (**1**) and funiculosin (**2**).

First isolated from the potato parasite *Fusarium sambucinum* and subsequently co-isolated with structurally-related *N*-desmethylsambutoxin (**3**) and 6-deoxysporidinone (**4**) from *Fusarium oxysporum*, **1** exhibited toxicity in chicken embryos and human tumor cells and caused hemorrhages in rats. **2** was isolated for the first time from *Penicillium funiculosum* and most recently from *Talaromyces cecidicola*, and it was active against pathogenic fungi such as *Candida albicans* and *Trychophyton mentagrophytes*.[4,5] Particularly notable in **2** from a biosynthetic perspective is its cyclopentanetetraol motif, which we hypothesized originates from oxidative ring-contraction of the aromatic side chain group of either phenylalanine or tyrosine. Given their high structural similarity at C_3, we also hypothesized that

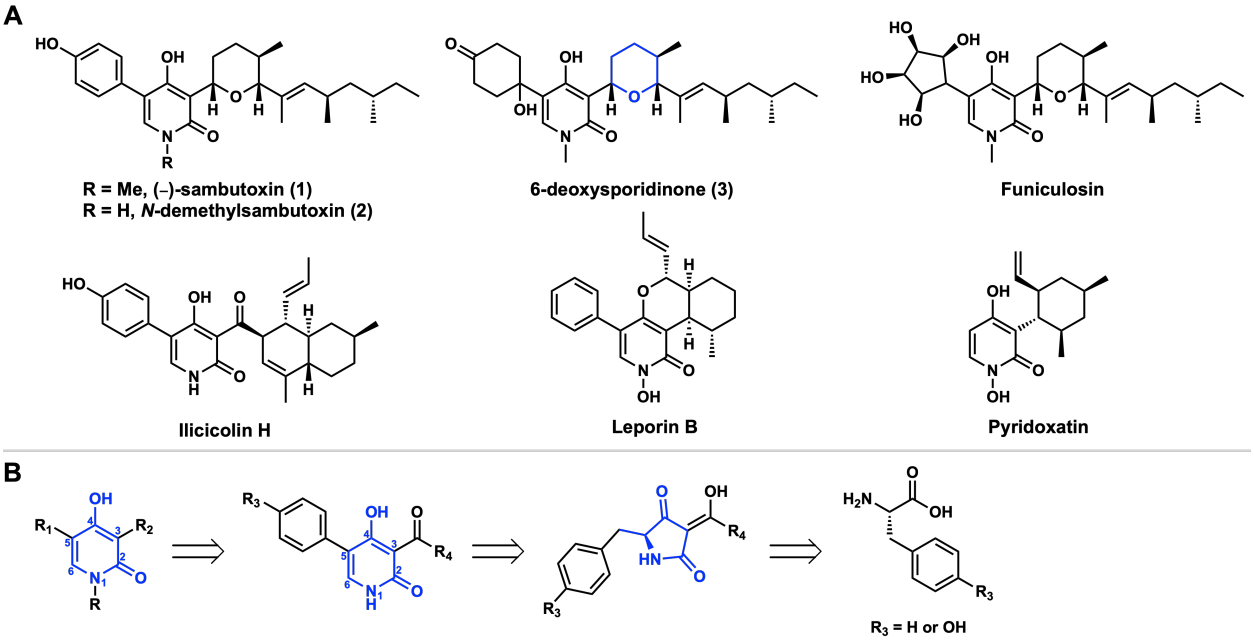


Figure 3.1. Fungal 4-hydroxy-2-pyridone alkaloids. (A) Representative members of the fungal 4-hydroxy-2-pyridone alkaloid family, including the tetrahydropyran-containing **1**–**4**. (B) Biosynthetic origins of the 4-hydroxy-2-pyridone core (blue) and its substituent groups. The R_4 group corresponds to the highly-reduced polyketide chain portion of the molecule synthesized by PKS module of the PKS-NRPS.

1 is a biosynthetic precursor to **2**. This chapter describes the biosynthetic pathway for **1** as determined by heterologous reconstitution studies, as well as the current progress on elucidating the transformation of **1** to **2**.

3.2 Results & Discussion

3.2.1 Retrobiosynthesis and Genome Mining for Sambutoxin and Funiculosin BGCs

Based on the biosynthetic relationship of **1** to other 4-hydroxy-2-pyridone natural products, a retrobiosynthesis for **1** was proposed as shown in **Figure 3.2A**. We initially proposed that the conversion of **3** to **1** is the final step in the sambutoxin biosynthetic pathway. The tetrahydropyran moiety in **3** was expected to form via an oxidative cyclization of the linear

3-acyl-4-hydroxy-2-pyridone precursor **5**, which is structurally similar to aspyridone A[6] and pre-tenellin B.[7] In turn, **5** may be derived from the P450-catalyzed ring expansion (RE) of the **6**, the first intermediate in the pathway synthesized from the collaborative efforts of a polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS) and a *trans*-enoyl reductase (ER). The *p*-hydroxyphenyl group in **1** was expected to be derived from L-tyrosine.

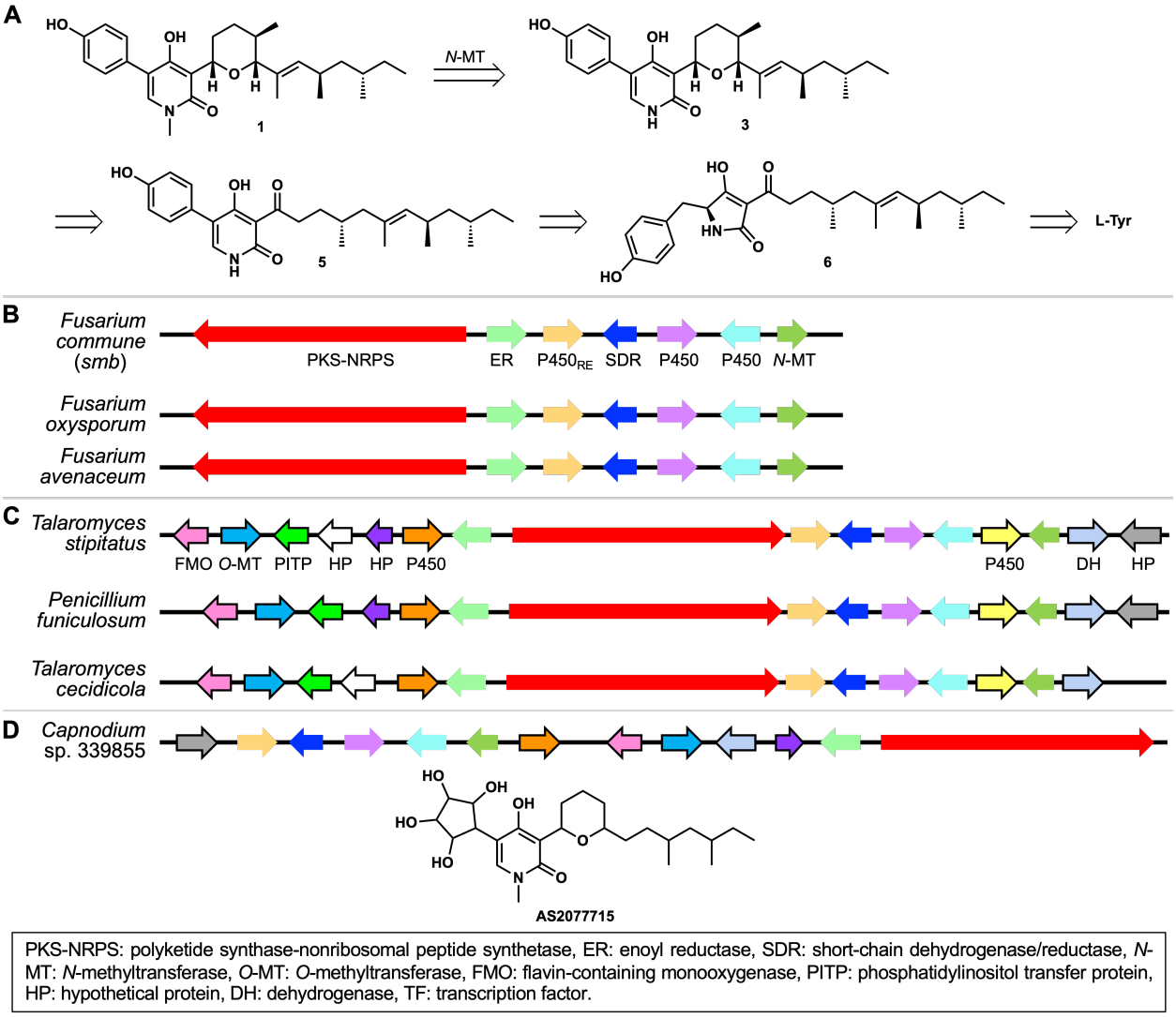


Figure 3.2. Proposed biosynthetic relationship between 1 and 2 as supported by genome mining. (A) Retrobiosynthetic proposal for **1** and **2**. Putative BGCs of (B) **1**, (C) **2**, and (D) AS2077715 identified by genome mining. Bolded block arrows in (C) and (D) indicate genes that are additionally present in the putative funiculosin and AS2077715 BGCs besides the *smb* homologs.

Considering that **1** and its structurally-related co-metabolites have been isolated from several *Fusarium* fungi,[8–10] we used a genome mining approach to identify potential BGCs of **1** in the *Fusarium* genomes in the National Center for Biotechnology Information (NCBI) fungal genome database. Querying the NCBI genome database with protein sequences of the PKS-NRPS (LepA), *trans*-ER (LepG), and $P450_{RE}$ (LepH) from the leporin B biosynthetic pathway,[1] with the added requirement of an *N*-methyltransferase, led to several well-conserved BGC hits (**Figure 3.2B**). Also present in the hits were genes encoding a short-chain dehydrogenase/reductase (SDR) and two additional P450 mono-oxygenases. One of the species housing the cluster was *Fusarium oxysporum*, from which **1**,**3**, and **4** have been previously isolated.[10]

In addition to **1**, we also sought to identify potential BGCs of **2** by mining the genomes of *Penicillium funiculosum* and *Talaromyces cecidicola*, two strains from which **2** had been previously isolated. If **1** is a true biosynthetic intermediate to **2**, we would expect to find homologous BGCs in *P. funiculosum* and *T. cecidicola* that contain the homolog of **1**'s putative BGC (named *smb*) and additional tailoring genes necessary for converting the phenol group of **1** to the cyclopentanetetraol of **2**. Querying the genome of *P. funiculosum* and *T. cecidicola* with the *smb* cluster revealed a gene cluster (named *fun*) that contained its homologue as a sub-cluster. Additional homologous gene clusters to *fun* were also found in *Talaromyces stipitatus* and *Talaromyces funiculosus*. Lastly, the genome of AS2077715-producing *Capnodium* sp.339855 also contained a homologous gene cluster to the *fun* cluster. Together, these results suggested that the *fun* cluster may be responsible for the biosynthesis of **2** by the way of **1**.

3.2.2 Heterologous Expression of the *smb* cluster in *Aspergillus nidulans*

To examine the products of the *smb* cluster from *F. commune*, we chose to reconstitute it in an engineered strain of *A. nidulans* ΔSTΔEM with a reduced endogenous metabolite background.[11] First, co-expression of the PKS-NRPS (SmbA) and trans-ER (SmbB) gave a new

compound **7** at 2 mg/L yield. Scaled-up culturing, isolation and structural characterization revealed **7** to be the tetramic acid shown in **Figure 3.5A** (**Table S4** and **Figures S2-S6**).

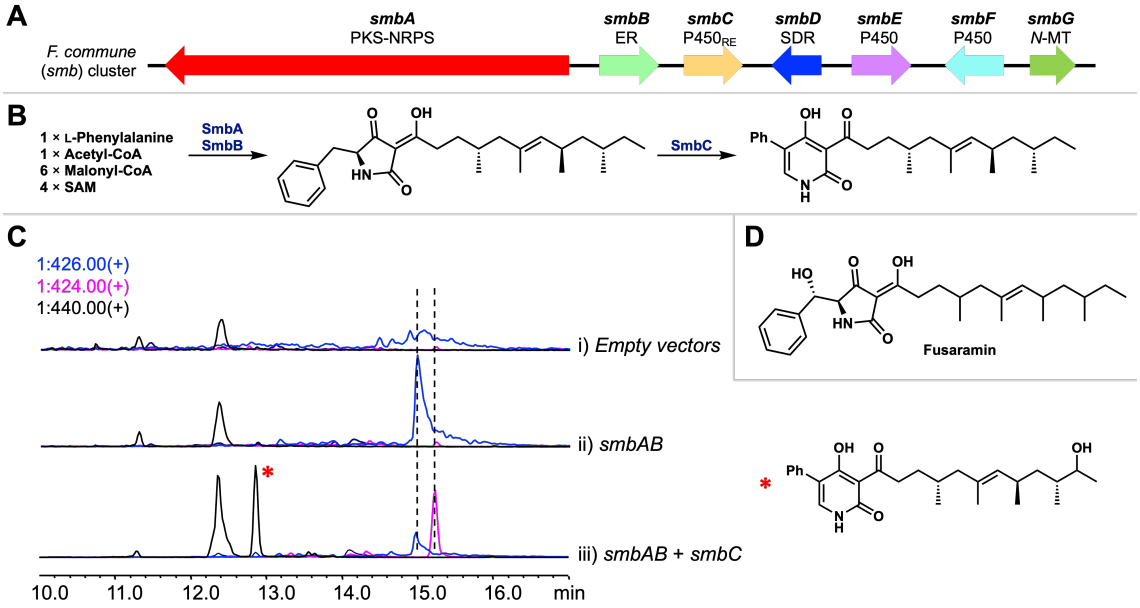


Figure 3.3. Phenylalanine incorporated by the PKS-NRPS SmbA. (A) The *smb* cluster, a putative BGC for sambutoxin. (B) Biosynthetic pathway to **8**. (C) LC-MS extracted ion chromatograms showing production of **7** by SmbA and SmbB and **8** upon co-expression of SmbC with SmbA and SmbB.

The polyketide portion of **7** is consistent that proposed for **6**, but it contained a phenyl group that is presumably derived from phenylalanine instead of the expected *p*-hydroxyphenyl group derived from tyrosine. No MS signal was observed for m/z of 442$[M+H]^+$ corresponding to **6**. This was initially a surprising result, and it suggested that SmbA exclusively accepts phenylalanine to make the early-stage intermediate en route to **1**. While this work was underway, Õmura and Shiomi reported the isolation of fusaramin (**Figure 3.3D**) from a *Fusarium* sp.[12] Fusaramin was a co-metabolite of **1**, **3**, and **4**, presumably derived from hydroxylation of C_β of phenylalanine in **7**. Together with our isolation of **7**, the existence of fusaramin suggested that contrary to our initial proposal, the biosynthesis of **1** starts with phenylalanine, and the *p*-hydroxy group of **1** is introduced later in the pathway.

Next, co-expression of SmbA and SmbB with SmbC, a putative $P450_{RE}$, led to the produc-

tion of **8** (**Figure 3.5** and **Figure S1**, trace iv). **8** was structurally verified as the 2-pyridone ketone resulting from the ring expansion of **7** (**Table S5** and **Figures S7–S11**), consistent with the functions of $P450_{RE}$ enzymes in other 4-hydroxy-2-pyridone alkaloid pathways.[1,13,14] In addition to **8**, a co-metabolite **8'** with +16 mu was isolated and characterized (**Figure 3.5** and **Table S6**, **Figures S12–S16**). **8'** was found to contain a secondary alcohol in polyketide chain of **8**, which may be attributed to the activity of endogenous alkyl hydroxylases in *A. nidulans* that have also observed in previous reconstitution studies.[15] It is worth noting here that **8'** persists throughout the remaining reconstitution work (**Figure S1**) and is not modified by the downstream *smb* enzymes.

3.2.2.1 Biosynthesis of the tetrahydropyran motif in sambutoxin

With **8** in hand, we next investigated the sequence of transformations leading to the formation of the tetrahydropyran. In the biosynthesis of other cyclic 2-pyridone such as leporins and citridones, the C_7 ketone is reduced by a SDR to C_7 alcohol and dehydrated to yield a reactive *ortho*-quinone methide (*o*-QM) that can serve as dienes or (di)enophiles in pericyclic reactions.[1,3,14,16] In the *smb* pathway, SmbD displayed moderate sequence homology to LepF[1] (32%) and PfpC[3] (36%). When co-expressed with SmbA–C, two new metabolites, **9** and **10**, with m/z 426$[M+H]^+$ and 408$[M+H]^+$, respectively, were produced (**Figure S1**, trace v).

Neither compound could be isolated for NMR characterization due to low amounts and relative instability. However, based on the mass and biosynthetic logic, we propose that **9** is the C_7 alcohol following the ketoreduction of **8**, and **10** with more extended conjugation (**Figure 3.4C**) is the [1,5]-hydride shifted shunt product (**Figure 3.5**). These putative assignments were supported by chemical reduction of the ketone **8** with $NaBH_4$, a method that has been used in previous 2-pyridone studies to verify SDR functions.[1,3] Here, chemical reduction of **8** led to the same two compounds **9** and **10**, eluting at the same retention time with the same MS and UV profiles (**Figure S1**, traces v and xii). Compound **10** may be

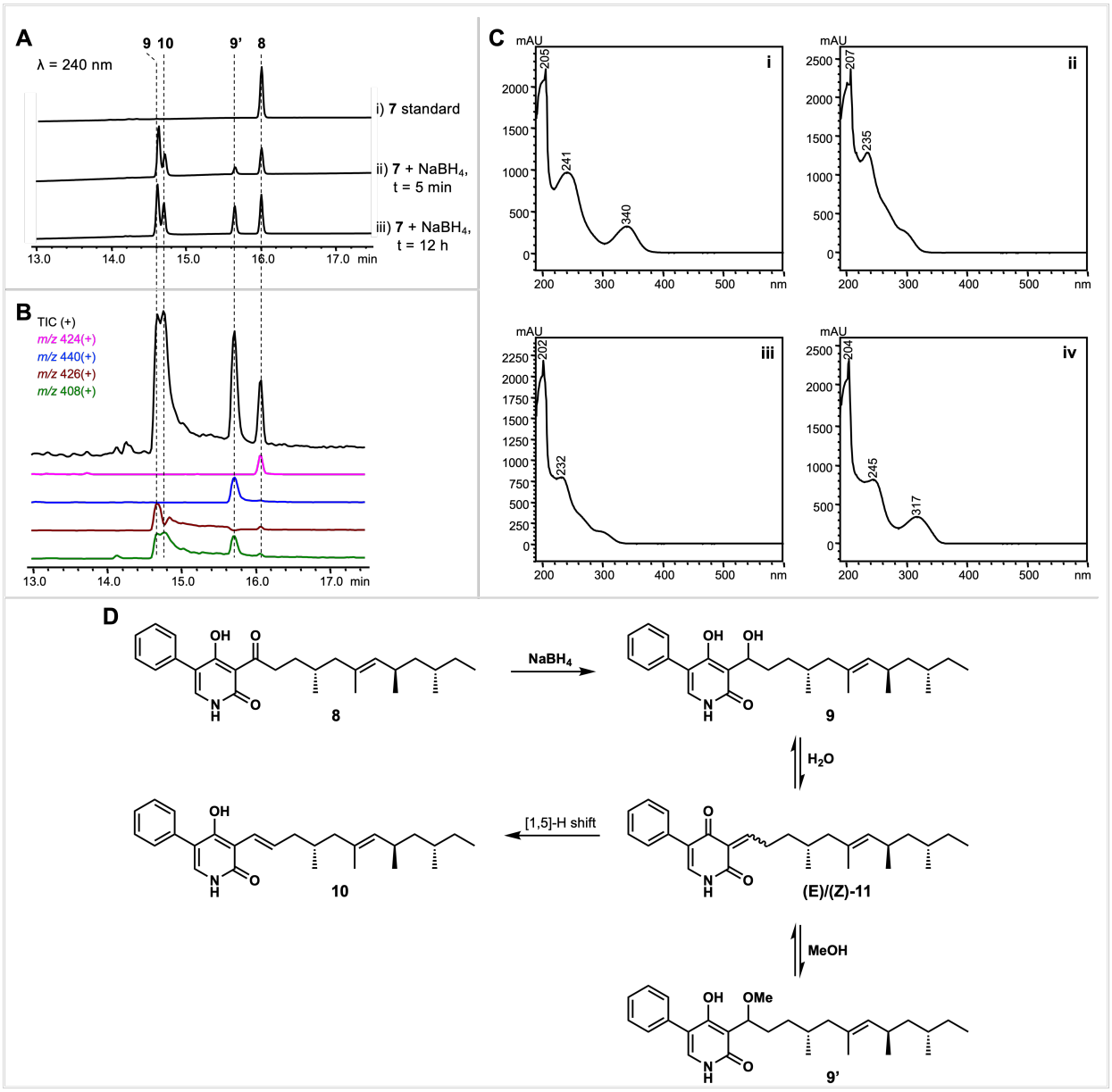


Figure 3.4. $NaBH_4$ reduction of 8 verifies SmbD function as a ketoreductase. (A) Reaction of 10 mg **8** (1 eq.) with 0.8 mg $NaBH_4$ (0.9 eq.) in MeOH (1 mL) resulted in the formation of 3 new products (**9**, **9'**, and **10**) with conversion of **9** to **9'** and **10** over time. (B) Extracted ion chromatograms corresponding to **8** and products of its reduction by $NaBH_4$. (C) The UV-vis absorption spectra of i) **8**, ii) **9**, iii) **9'**, and iv) **10**. (D) Proposed structures for **9**, **9'**, and **10** and their formation.

formed via dehydration of **9** to either the (E)- or (Z)- *o*-QM **11**, which may subsequently undergo a [1,5]-hydride shift as observed in the synthetic study of 2-pyridones.[17] One product of the chemical reduction of **8** that was not detected from *A. nidulans* reconstitution is the

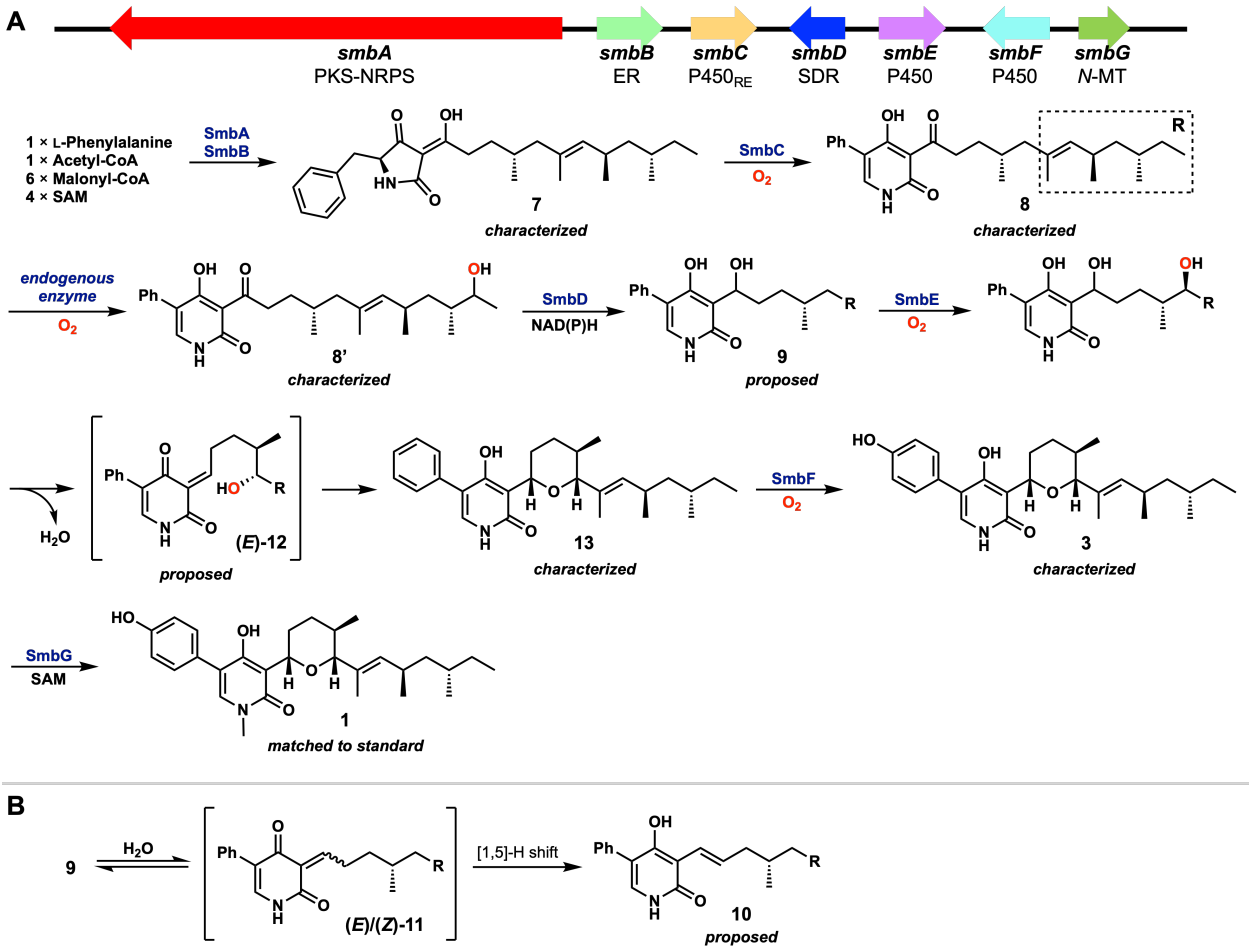


Figure 3.5. Biosynthetic pathway for sambutoxin (1). (A) The annotated *smb* cluster and proposed biosynthetic pathway for **1** based on step-wise reconstitution. (B) Proposed pathway for the generation of shunt [1,5]-hydride shift product upon dehydration of **9** to give **11**.

likely methoxy adduct **8'** (**Figure 3.4**, traces ii and iii) that arises upon Michael addition of solvent methanol to **11**. The co-emergence of **9** and **10** upon co-expression of SmbA-D as well as upon $NaBH_4$ reduction of **8** supports that SmbD functions as a ketoreductase, consistent with previous studies of their homologs,[1,3] and it also suggests that an additional enzyme that can function on **9** is need to minimize shunt product formation.

Given that electrophilic *o*-QMs can serve as Michael acceptors in nucleophilic additions,[18,19] we propose that the (E)-*o*-QM **11** formed by a stereospecific dehydration of **9** serves as a Michael acceptor for the tetrahydropyran formation. This would require the

installation of an (*R*)-hydroxyl group in at C_{11} of the polyketide chain as in (E)-**12** to serve as an internal nucleophile (**Figure 3.5**). The Michael addition could take place in a stereospecific manner without enzymatic control given the bulky groups on C_2 and C_6 of the tetrahydropyran would preferentially arrange equatorially, as demonstrated in the total syntheses of (+)-1 and septoriamycin.[20,21] Hydroxylation of the allylic C_{11} in **9** can be catalyzed by one of the remaining P450 enzymes, SmbE or SmbF. No new metabolite was produced (**9** and **10** remained) when SmbF was co-expressed with SmbA–D (vida infra) (**Figure S1**, trace vi). In contrast, co-expression of SmbE with SmbA–D led to significant changes in the metabolite profile (**Figure S1**, trace vii). A new metabolite **13** with m/z 424$[M+H]^+$ emerged (3.7 mg/L), accompanied by a significant decrease in the level of **9**. Structural characterization of **13** by NMR (**Table S7** and **Figures S17–S21**) revealed the compound to indeed contain the tetrahydropyran moiety found in **1**.

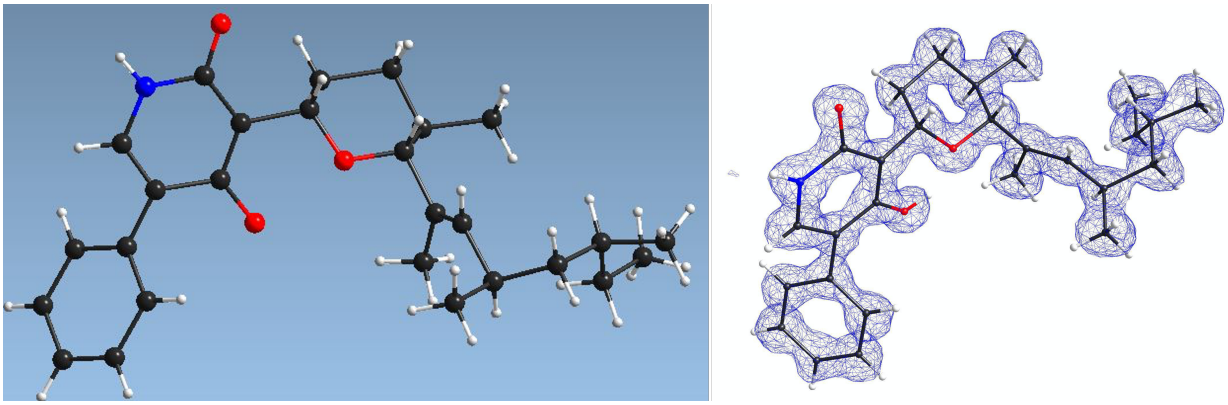


Figure 3.6. MicroED structure of 13.

To further verify that the relative stereochemistries of the tetrahydropyran and methyl groups in the polyketide chain are consistent with the reported structure of **1**, we obtained the three-dimensional structure of **13** using microcrystal electron diffraction (MicroED)[22] (**Figure 3.6**). Microcrystals of **13** were obtained by slow air evaporation of a pure HPLC fraction of **13** in acetonitrile/water. The relative stereochemistry of **13** was identical to that reported for (−)-**1**, confirming that **13** is most likely an on-pathway intermediate, and that SmbE functions as the C_{11} hydroxylase to enable the tetrahydropyran formation via an

intramolecular Michael addition.

3.2.2.2 Late-stage oxidation

It is interesting to note that up to **13** in the sambutoxin biosynthetic pathway, the aromatic ring at $C_{4'}$ has remained an un-oxidized phenyl group, in contrast to the *p*-hydroxyphenyl group in **1**. Given that the remaining enzymes in the cluster include another P450 (SmbF) and an *N*-MT (SmbG), we anticipated that SmbF would oxidize the $C_{4'}$ of the phenyl group to give *N*-desmethylsambutoxin (**3**). Indeed, when SmbF was co-expressed with SmbA–E in *A. nidulans*, the strain produced a new compound (2 mg/L) that was confirmed to be **3** (**Figure S1**, trace x; **Table S8** and **Figures S22-S26**). As a result, SmbF was assigned as the $C_{4'}$-hydroxylase in this pathway. The timing of SmbF activity must follow tetrahydropyran formation by SmbE, as no phenyl hydroxylase activity was observed when SmbF was expressed with SmbA-D in the absence of SmbE (**Figure S1**, trace vi).

Two possible mechanism can be proposed for the SmbF-catalyzed aryl hydroxylation: i) abstraction of hydrogen from the weak amide N_1-H bond generates a free radical that can be delocalized to $C_{4'}$ on the phenyl ring (**Figure 3.7C**), which could further react with the iron-bound hydroxyl radical in SmbF to complete the regiospecific $C_{4'}$-hydroxylation; or ii) epoxidation at $C_{3'}$-$C_{4'}$ of the phenyl group is followed by 1,2-hydride (NIH) shift to form the 4'-OH group in a mechanism that is similar to that reported for cinnamate hydroxylase (**Figure 3.7**D). This late-stage C–H oxidation by SmbF to oxidize the para position of phenylalanine-derived phenyl group is rather unique in the biosynthesis of polyketide-nonribosomal peptide hybrid molecules. For tenellin, aspyridone,[6] and illicicolin H,[2] the *p*-hydroxyphenyl groups are introduced at the start of the biosynthetic pathways via incorporation of tyrosine by the respective PKS-NRPSs. It is thus intriguing that the sambutoxin biosynthetic pathway does not directly incorporate tyrosine, which could lead to a shorter and more efficient pathway. One possible explanation for the choice of phenylalanine by SmbA is that some $P450_{RE}$ enzymes are capable of dephenylating tyrosine-containing tetramic acids in the pro-

cess of ring expansion (**Figure 3.7**B).[3,23] The *p*-hydroxyphenyl group in the tetramic acid is critical for dephenylation, which has been shown to occur in the biosynthesis of pyridoxatin,[3] aspyridone,[24] harzianopyridone.[23] Dephenylation of 5 would derail the biosynthesis of **1**, which could rationalize the strategy employed by the smb pathway to first incorporate a phenylalanine at the early stage of sambutoxin biosynthesis and dedicate a separate enzyme to oxidize the phenyl group at a later stage in the pathway when dephenylation is no longer possible.

3.2.3 *In vitro N*-methylation of *N*-desmethylsambutoxin affords sambutoxin

From **3**, the remaining *N*-MT (SmbG) is proposed to perform the *N*-methylation to afford **1**. When SmbG was co-expressed with SmbA–F in *A. nidulans*, we observed a small peak with m/z 454$[M+H]^+$ that eluted at the same retention time as a commercial standard of **1** (**Figure S1**, traces xi and xiii). Attempts to purify sufficient quantity of **1** for NMR structural characterization from 4 L of cultures was unsuccessful due to low titer, possibly due to unknown degradation/detoxification pathways in the host. However, we were able to demonstrate that recombinantly expressed SmbG is capable of *N*-methylating **3** *in vitro* ($K_M = 54.7 \pm 11.4\ \mu M$, $K_{cat} = 1.70\ \ 0.13\ min^{-1}$, and $k_{cat}/K_M = 3.1\ \ 10^4\ min^{-1}M^{-1}$) to give **1** in the presence of S-adenosylmethionine (SAM) (**Figure 3.8**). SmbG specifically methylated **3** and was inactive on **13**, as no new peak corresponding to the *N*-methylated derivative of **13** was observed upon reaction with SmbG and SAM. These results confirmed that the *N*-methylation of **3** by SmbG is the final step of the biosynthetic pathway for **1**.

3.2.4 From sambutoxin *en route* to funiculosin

It was proposed earlier that **1** may be a biosynthetic precursor to **2** given their highly similar polyketide chain groups. One way to determine the intermediacy of a molecule in a biosynthetic pathway is to genetically knock out a biosynthetic gene for the molecule in

question, which would abolish production of the pathway's final product. If feeding a known precursor to Molecule A that lies downstream of the knocked-out gene restores Molecule A production, then this is a strong evidence that the exogenously introduced precursor molecule is an intermediate in the Molecule A biosynthetic pathway.

When grown in sucrose/nutrient-rich media, the wild-type *Talaromyces coalescens* (TTI-0630) produces **1**, **2**, and other known precursors to sambutoxin (**3**, **7**, and**14**) (**Figure** 3.9D). Knocking out *funM*, the *N*-MT gene in *T. cecidicola* that is homolgous to SmbG (78% sequence identity), led to loss of production of **1** and **2** (**Figure 3.9E**. On the other hand, **3** accumulated at higher levels relative to the earlier intermediates in the sambutoxin pathway. This experiment provided further evidence for the role of the *N*-MT in converting **3** to **1** and that **1** could be an on-pathway biosynthetic intermediate in the funiculosin pathway. The intermediacy of **1** to **2** was verified by feeding **1** to the *funK* (PKS-NRPS)-deficient knockout of *T. cecidicola* (*T. cecidicola*Δ*funG*), which restored the production of **2**.(**Figure 3.9C.**

3.3 Conclusion

In summary, we have uncovered the linear biosynthetic pathway for **1**, providing insight into Nature's strategy for synthesizing the tetrahydropyran motif in 4-hydroxy-2-pyridone alkaloids. This is the first example from 4-hydroxy-2-pyridone alkaloid biosynthesis in which the *p*-hydroxyphenyl group at $C_{4'}$ is derived not from tyrosine but rather via a late-stage oxidation of the phenylalanine side chain group. Full reconstitution of the biosynthesis of **1** enables genome-based mapping of fungi capable of making this toxin and sets the stage for the investigation of related compounds such as 6-deoxysporidinone (**15**). Lastly, we have also demonstrated that **1** is an on-pathway intermediate to **2** by feeding and knock-out studies in the producing host *Talaromyces cecidicola*.

3.4 References

(1) Ohashi, M.; Liu, F.; Hai, Y.; Chen, M.; Tang, M.-c.; Yang, Z.; Sato, M.; Watanabe, K.; Houk, K. N.; Tang, Y. *Nature* **2017**, *549*, 502–506.

(2) Zhang, R. K.; Chen, K.; Huang, X.; Wohlschlager, L.; Renata, H.; Arnold, F. H. *Nature* **2019**, *565*, 67–72.

(3) Ohashi, M. et al. *Nature* **2020**, *586*, 64–69.

(4) Ando, K.; Suzuki, S.; Saeki, T.; Tamura, G.; Arima, K. *The Journal of Antibiotics* **1969**, *22*, 189–194.

(5) Ando, K.; Matsuura, I.; Nawata, Y.; Endo, H.; Sasaki, H.; Okytomi, T.; Saehi, T.; Tamura, G. *The Journal of Antibiotics* **1978**, *31*, 533–538.

(6) Bergmann, S.; Schümann, J.; Scherlach, K.; Lange, C.; Brakhage, A. A.; Hertweck, C. *Nature Chemical Biology* **2007**, *3*, 213–217.

(7) Halo, L. M.; Marshall, J. W.; Yakasai, A. A.; Song, Z.; Butts, C. P.; Crump, M. P.; Heneghan, M.; Bailey, A. M.; Simpson, T. J.; Lazarus, C. M.; Cox, R. J. *ChemBioChem* **2008**, *9*, 585–594.

(8) Kim, J.; Lee, Y.; Yu, S. *Applied and Environmental Microbiology* **1995**, *61*, 3750–3751.

(9) Breinholt, J.; Ludvigsen, S.; Rassing, B. R.; Rosendahl, C. N.; Nielsen, S. E.; Olsen, C. E. *Journal of Natural Products* **1997**, *60*, 33–35.

(10) Jayasinghe, L.; Abbas, H. K.; Jacob, M. R.; Herath, W. H. M. W.; Nanayakkara, N. P. D. *Journal of Natural Products* **2006**, *69*, 439–442.

(11) Liu, N.; Hung, Y.-S.; Gao, S.-S.; Hang, L.; Zou, Y.; Chooi, Y.-H.; Tang, Y. *Organic Letters* **2017**, *19*, 3560–3563.

(12) Sakai, K. et al. *The Journal of Antibiotics* **2019**, *72*, 645–652.

(13) Zhang, Z.; Jamieson, C. S.; Zhao, Y.-L.; Li, D.; Ohashi, M.; Houk, K. N.; Tang, Y. *Journal of the American Chemical Society* **2019**, *141*, 5659–5663.

(14) Zhang, Z.; Qiao, T.; Watanabe, K.; Tang, Y. *Angewandte Chemie International Edition* **2020**, *59*, 19889–19893.

(15) Liu, N.; Abramyan, E. D.; Cheng, W.; Perlatti, B.; Harvey, C. J.; Bills, G. F.; Tang, Y. *Journal of the American Chemical Society* **2021**, *143*, 6043–6047.

(16) Chen, Q.; Gao, J.; Jamieson, C.; Liu, J.; Ohashi, M.; Bai, J.; Yan, D.; Liu, B.; Che, Y.; Wang, Y.; Houk, K. N.; Hu, Y. *Journal of the American Chemical Society* **2019**, *141*, 14052–14056.

(17) Fotiadou, A. D.; Zografos, A. L. *Organic Letters* **2011**, *13*, 4592–4595.

(18) Purdy, T. N.; Kim, M. C.; Cullum, R.; Fenical, W.; Moore, B. S. *Journal of the American Chemical Society* **2021**, *143*, 3682–3686.

(19) Doyon, T. J.; Perkins, J. C.; Baker Dockrey, S. A.; Romero, E. O.; Skinner, K. C.; Zimmerman, P. M.; Narayan, A. R. H. *Journal of the American Chemical Society* **2019**, *141*, 20269–20277.

(20) Williams, D. R.; Turske, R. A. *Organic Letters* **2000**, *2*, 3217–3220.

(21) Nakamura, T.; Harachi, M.; Kano, T.; Mukaeda, Y.; Hosokawa, S. *Organic Letters* **2013**, *15*, 3170–3173.

(22) Jones, C. G.; Martynowycz, M. W.; Hattne, J.; Fulton, T. J.; Stoltz, B. M.; Rodriguez, J. A.; Nelson, H. M.; Gonen, T. *ACS Central Science* **2018**, *4*, 1587–1592.

(23) Bat-Erdene, U.; Kanayama, D.; Tan, D.; Turner, W. C.; Houk, K. N.; Ohashi, M.; Tang, Y. *Journal of the American Chemical Society* **2020**, *142*, 8550–8554.

(24) Wasil, Z.; Pahirulzaman, K. A. K.; Butts, C.; Simpson, T. J.; Lazarus, C. M.; Cox, R. J. *Chemical Science* **2013**, *4*, 3845–3856.

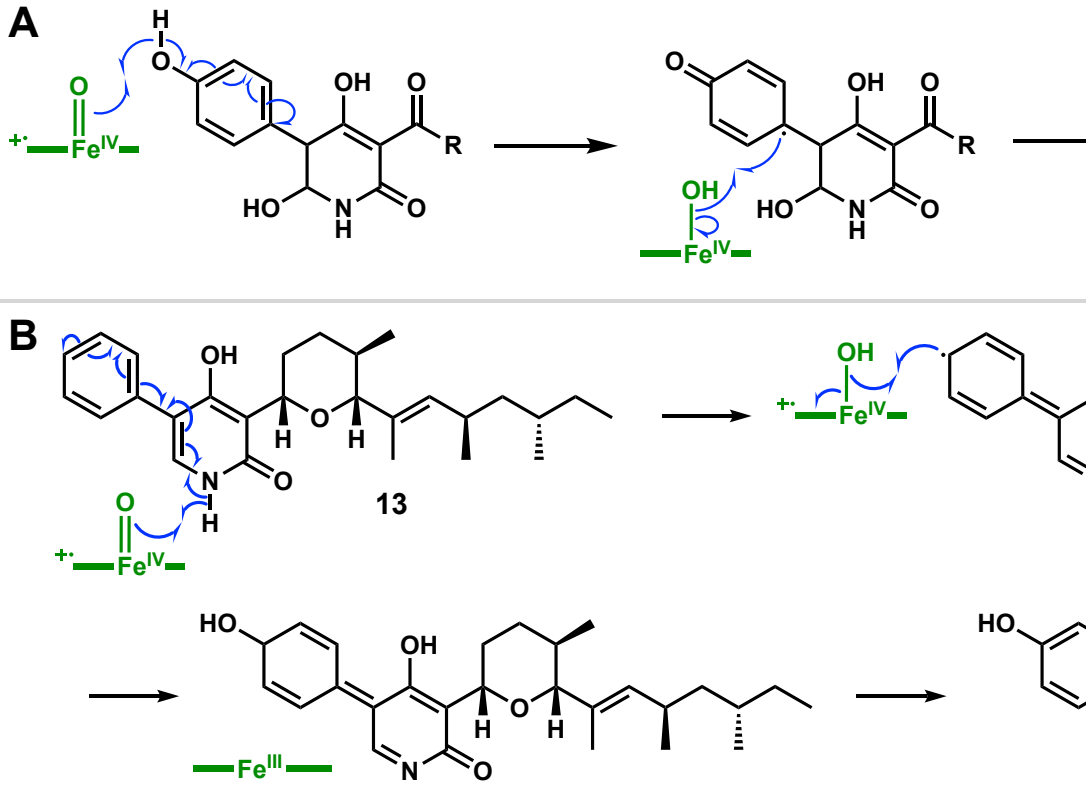


Figure 3.7. Proposed mechanisms for ring expansion, dephenylation, and aryl hydroxylation. (A) Overview of 2-pyridone biosynthesis by PKS-NRPS, *trans*-ER, and ring-expansion P450 ($P450_{RE}$). (B) Dephenylation has been proposed to be driven by sequential oxygen rebound processes that hydroxylate the C_6 and $C_{1'}$ positions, well-poised to release a pbenzoquinone that results in the elimination of the 6-OH group to release the dephenylated 2-pyridone alkaloid.[3,23] (C) Hydroxylation by SmbF may be facilitated by ab-

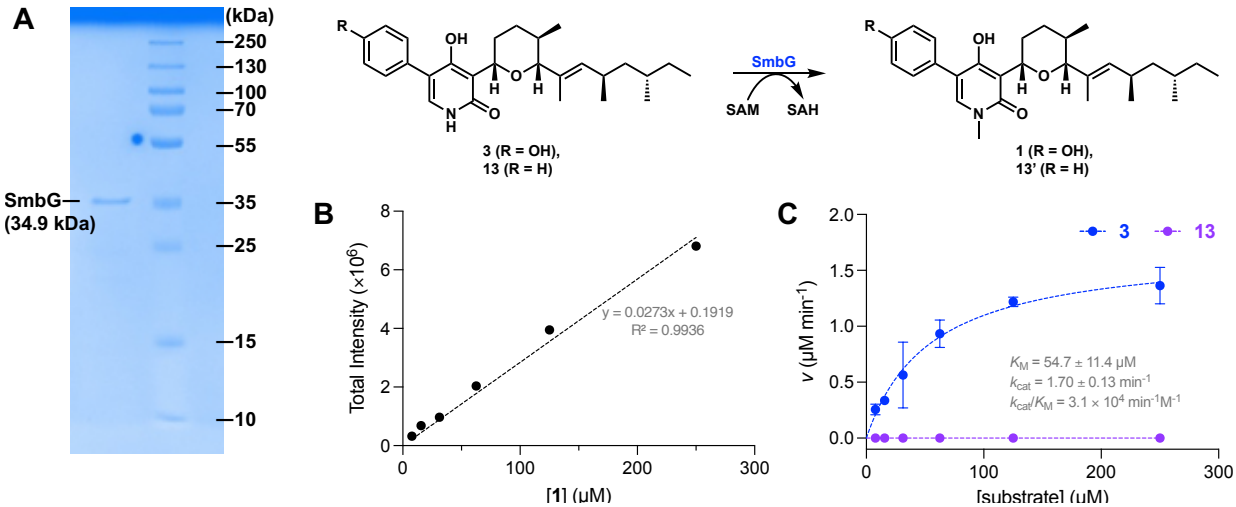


Figure 3.8. Biochemical characterization of SmbG. (A) SDS-PAGE analysis of SmbG expressed and purified from *E. coli*. (B) Standard curve of **3**. (C) Steady-state kinetic parameters for SmbG. Reactions were performed with 1 μM SmbG, 1 mM SAM, and varying concentrations of **3** and **13**. No *N*-methylation was observed for **13** to give 14.

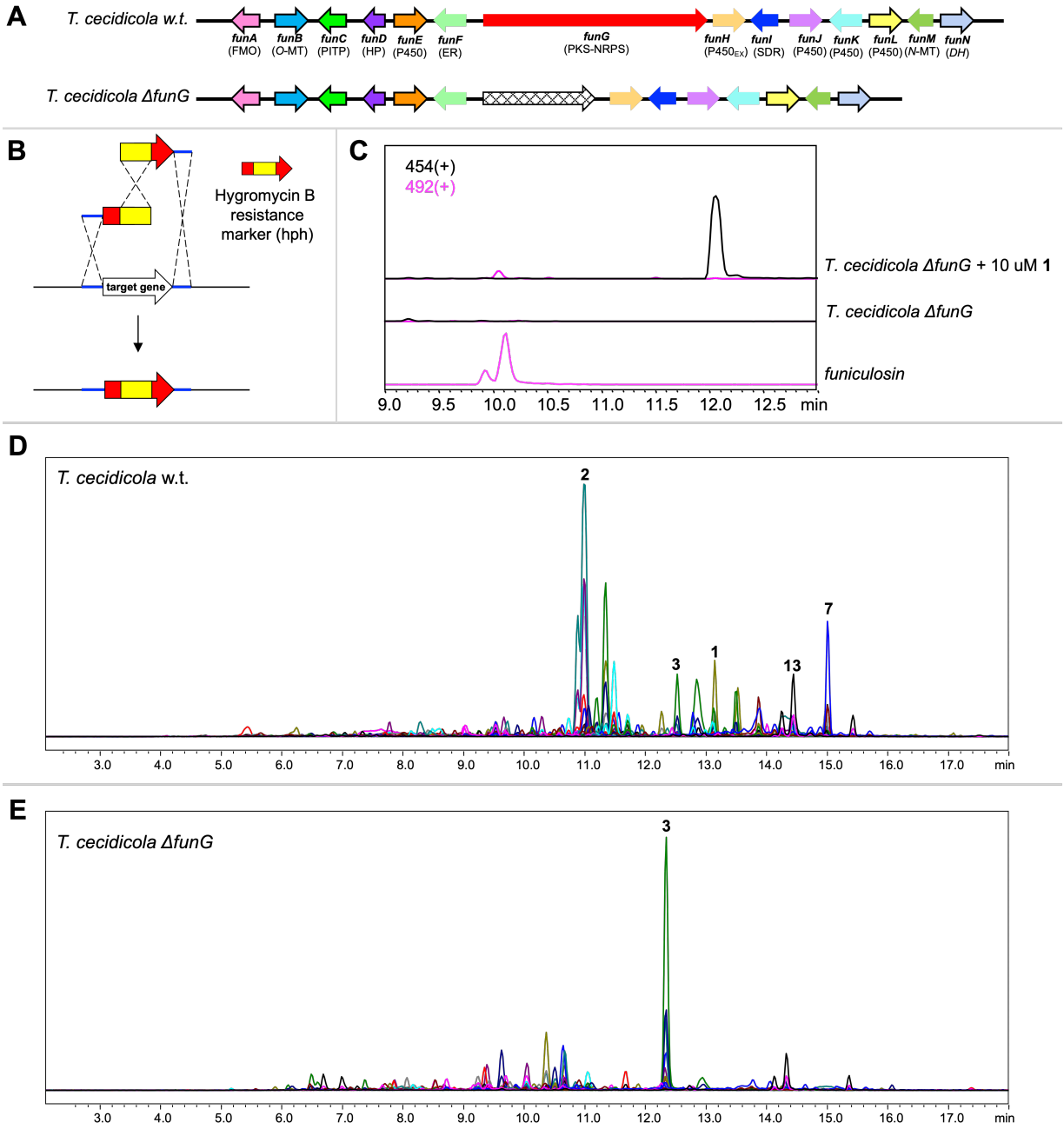


Figure 3.9. Genetic knockouts in *T. cecidicola*. (A) Scheme showing the wild-type *T. cecidicola* funiculosin BGC (top) and knockout of the PKS-NRPS (*funG*). (B) Split-marker approach to target gene knockout via *in vivo* homologous recombination. (C) **2** production restored upon feeding **1** to *T. cecidicola*Δ*funG*. (D and E) Metabolic profiles of wild-type and Δ*funN T. cecidicola*.

CHAPTER 4

Appendix

4.1 Growth media, buffers and solutions used in this study

Nitrate salts	120 g $NaNO_3$, 10.4 g KCl, 10.4 g $MgSO_4 \cdot 7H_2O$, 30.4 g KH_2PO_4 in 1 L de-ionized H_2O
Trace elements	2.20 g $ZnSO_4 \cdot 7H_2O$, 1.10 g H_3BO_3, 0.50 g $MnCl_2 \cdot 4H_2O$, 0.16 g $FeSO_4 \cdot 7H_2O$, 0.16 g $CoCl_2 \cdot 5H_2O$, 0.16 g $CuSO_4 \cdot 5H_2O$, 0.11g $(NH_4)_6Mo_7O_{24} \cdot 7H_2O$ in 100 mL de-ionized H_2O; pH 6.5
Liquid CD	0.1% glucose, 5 % v/v nitrate salts, 0.1 % v/v trace elements
CD agar	Liquid CD with 2% agar
YPD media	1% yeast extract, 2% peptone, 2% glucose
CD-ST soft agar	2% starch, 2% casamino acids, 5% v/v nitrate salts, 0.1% v/v trace elements, 1% agar
Osmotic medium	1.2 M $MgSO_4$, 10 mM sodium phosphate; pH 5.8
Trapping buffer	0.6 M sorbitol, 0.1 M Tris-HCl; pH 7.0
STC buffer	1.2 M sorbitol, 10 mM $CaCl_2$, 10 mM Tris-HCl; pH 7.5
TE buffer	1 mM EDTA, 10 mM Tris-HCl; pH 7.8

4.2 Strains and general culture conditions

Aspergillus nidulans A1145ΔEMΔST was grown on CD agar supplemented with 10 mM uridine, 5 mM uracil, 0.5 μg/mL pyridoxine·HCl and 2.5 μg/mL riboflavin for sporulation or on CD-ST soft agar for heterologous expression of biosynthetic genes, compound production,

and RNA isolation. *Talaromyces cecidicola* (TTI-0630) was grown on CD agar with no added supplements. All *Escherichia coli* strains were cultured in liquid LB media at 37 °C—strain DH5α was used for plasmid propagation and BL21 (DE3) for protein expression. Yeast was cultured in YPD media.

4.3 General DNA manipulation techniques

DNA restriction enzymes were used as recommended by the manufacturer (New England Biolabs, NEB). PCR was performed with Phusion or Q5 DNA polymerase (NEB). *E. coli* DH5α cells were used for cloning, following standard recombinant DNA techniques. The primers used in this study are listed in Table S1.

4.3.1 Fungal nucleic acid isolation

Fungal gDNA was isolated by beating the cell body with 50 μL 0.5 mm Zirconia beads in 1 mL 1:1 vol. mixture of LETS buffer (0.1 M LiCl, 0.01 M Na_2EDTA, 0.01 M Tris-HCl at pH 7.4, 0.2 % SDS) and phenol:$CHCl_3$ solution (10 min) to rupture the cells. The organic and aqueous layers were separated by centrifugation, and the aqueous layer was added to 1 mL i-PrOH to precipitate the gDNA. The gDNA was pelleted, washed with 70% EtOH/H_2O, resuspended in H_2O, purified on the DNA Clean Concentrator™-5 (Zymo Inc. USA) and stored in TE buffer at 4 °C.

Fungal RNA was isolated using the RiboPure™-Yeast Kit (Invitrogen™ AM1926), digested with DNase I (NEB M0303L) (37 °C, 3 hours), purified on the DNA Clean & Concentrator™-5 and stored in TE buffer at –20 °C.

4.3.2 Fungal cDNA library synthesis

SuperScript III First-Strand Synthesis System (Invitrogen™ 18080051) was used to synthesize the cDNA library from fungal RNA via reverse transcription with oligo-dT primers

according to the manufacturer's protocol.

4.3.3 Construction of plasmids for heterologous expression in *A. nidulans*

Plasmids for heterologous expression in *A. nidulans* were constructed by *in vivo* homologous recombination in the yeast strain Saccharomyces cerevisiae BJ5464-NpgA (*MATα ura3-52 his3-Δ200 leu2-Δ1 trp1 pep4::HIS3 prb1 Δ1.6R can1 GAL*). For each plasmid, gene inserts were amplified by PCR and co-transformed into yeast with the appropriate double-digested vectors. Assembled plasmids were isolated from yeast using the Zymoprep™ Yeast Plasmid Miniprep I Kit (Zymo Inc. USA) and transformed into *E. coli* DH5α cells by electroporation. The sequences of plasmids were verified by sequencing (Laragen Inc.). The plasmids used in this study are listed in **Table** S2.

4.3.4 Construction of plasmids for preparation of DNA fragments for *T. cecidicola* knockouts

Plasmids to serve as templates for PCR to generate DNA fragments for *T. cecidicola* knockout experiments were prepared via HiFi assembly of the following components: i) PCR-amplified XW55 backbone, ii) PCR-amplified pTrpC-hph-tTrpC marker, and iii) the upstream & downstream homology regions.

4.4 Preparation and transformation of *A. nidulans* protoplasts for heterologous expression

Spores of *A. nidulans* A1145 ΔSTΔEM[1] were inoculated into 50 mL liquid CD supplemented with 10 mM uridine, 5 mM uracil, 0.5 μg/mL pyridoxine·HCl and 2.5 μg/mL riboflavin in a 250-mL flask and germinated (30 °C, 250 rpm, 16 h). Mycelia were harvested by centrifugation (24 °C, 4,300 rpm, 10 min), washed with 10 mL osmotic medium, re-suspended in 10 mL osmotic medium containing 30 mg lysing enzymes from *Trichoderma* (MilliporeSigma) and 20 mg Yatalase (Takara) in a 125-mL flask and digested overnight (30 °C, 80 rpm, 15

h). The digest mixture was transferred to a 15-mL conical tube and gently overlaid with 5 mL trapping buffer, and protoplasts were collected from the interface of the buffer layers after centrifugation (4 °C, 4,300 rpm, 15 min). The protoplasts were transferred to a new 15-mL conical tube, washed with 10 mL STC buffer, resuspended in 2 mL STC buffer and stored as 40 μL aliquots at –80 °C. The protoplasts remained competent for transformation for more than 7 months.

For each transformation, 1 μL of each plasmid was added to a freshly thawed 40-μL aliquot of *A. nidulans* protoplasts and incubated on ice for 45 minutes. 600 μL PEG solution was added to the protoplast/plasmid mixture and mixed by gentle shaking to form a homogenous mixture, which was spread on the regeneration medium (CD agar with 1.2 M sorbitol) and incubated at 37 °C for 2-3 days until colonies of transformants appeared. The transformants were transferred to CD agar and stored at 28 °C. The *A. nidulans* transformants prepared in this study are listed in **Table S3**.

4.5 Preparation and transformation of *T. cecidicola* protoplasts for targeted gene knockout

Spores of *T. cecidicola* were inoculated into 250 mL liquid CD and germinated (30 °C, 250 rpm, 5 days). Mycelia were harvested by centrifugation (24 °C, 4,300 rpm, 10 min), washed with 10 mL osmotic medium, re-suspended in 10 mL osmotic medium containing 30 mg lysing enzymes from *Trichoderma* (MilliporeSigma) and 20 mg Yatalase (Takara) in a 125-mL flask and digested overnight (30 °C, 80 rpm, 15 h). The digest mixture was transferred to a 15-mL conical tube and gently overlaid with 5 mL trapping buffer, and protoplasts were collected from the interface of the buffer layers after centrifugation (4 °C, 4,300 rpm, 15 min). The protoplasts were transferred to a new 15-mL conical tube, washed with 10 mL STC buffer, resuspended in 2 mL STC buffer and aliquoted into 40 μL for immediate use.

For the $\Delta funG$ knockout, split marker fragments were PCR-amplied with primer pairs

{KO_FunG_upF, hph_R} & {KO_FunG_downR, hph_F} with pEB116 as template. For the Δ*funL* knockout, split marker fragments were PCR-amplified with primer paris {KO_FunL_upF, hph_R} & {KO_FunL_downR, hph_F} with pEB117 as template. For each transformation, the split markers were combined and purified in 5 μL STC buffer, added to a tube of protoplasts, incubated on ice for 45 minutes, and spread on the regeneration medium (CD agar with 1.2 M sorbitol, 100 μg/mL Hygromycin B) and incubated at 37 °C for 2-3 days until colonies of transformants appeared. Single colonies of transformants were transferred to CD agar with 50 μL Hygromycin B, and colonies with successful knockout was identified by diagnostic PCR on purified genomic DNA with primers targeting the inserted split marker fragments.

4.6 General protocol for LC-MS analysis

Samples were analyzed on the Shimadzu 2020 EV LC-MS with a reversed-phase C18 column (Phenomenex Kinetex, 1.7 μm, 100 Å, 2.1 × 100 mm) with positive- and negative-mode electrospray ionization. A linear gradient of 5-95% MeCN/H_2O in 15 minutes followed by 95% MeCN/H_2O for 3 minutes with a flow rate of 0.3 mL/min. was applied for each sample. The MeCN and H_2O were supplemented with 0.1 % v/v formic acid.

4.7 Small-scale (25 mL) cultures of *A. nidulans* transformants for LC-MS analysis

A. nidulans transformants were grown on 25 mL CD-ST soft agar (28 °C, 3–4 days) and extracted overnight with 25 mL EtOAc. For each LC-MS analysis, 750 μL of the EtOAc extract was washed with 750 μL H_2O, and the organic layer was dried and re-dissolved in 50 μL MeOH, of which 10 μL was used for LC-MS analysis.

4.8 Large-scale (4 L) cultures of *A. nidulans* transformants for compound purification

For isolation of **7**, **8**, **8'**,**13** and **3**, each *A. nidulans* transformant (**Table S3**) was grown on 4 L soft-agar CD-ST (28 °C, 3–4 days) and extracted overnight with 2 L acetone. The acetone extracts were concentrated by rotary evaporation, and the residual aqueous layer was extracted with EtOAc (3 × 750 mL). The EtOAc layers from all extractions were combined and concentrated by rotary evaporation, re-dissolved in MeOH, and adsorbed onto 5 g Celite for dry loading for separation on the CombiFlash system. A normal phase EtOAc/hexanes gradient (25% EtOAc for 5 min, followed by a linear gradient of 25%-100% EtOAc for 15 min, followed by 100% EtOAc for 5 min) was applied for the initial separation of target compounds. Fractions containing the target compounds were identified by LC-MS and combined for further purification by HPLC.

HPLC purification was performed on the Thermo Scientific™ UltiMate™ 3000 Basic Automated System with a semi-preparative reversed-phase C18 column (Phenomenex Kinetex, 5 μm, 100 A, 10 × 250 mm). A linear gradient of 75-95% MeCN/H_2O in 11.5 minutes, followed by 95% MeCN/H_2O for 10 minutes, followed by 75% MeCN/H_2O for 3.5 minutes with a flow rate of 3 mL/min. was applied for each sample. The MeCN and H_2O were supplemented with 0.1 % v/v formic acid. HPLC fractions containing pure samples of the desired compounds were identified by LC-MS, combined, and dried under reduced pressure to yield the final compounds.

4.9 Purification and biochemical characterization of SmbG

The intron-free cDNA of SmbG was amplified from the cDNA library of AnEB106 and ligated into PCR-amplified modified pET28a(+) vector (Addgene plasmid 29656) backbone by HiFi DNA Assembly (NEB). The assembly reaction mixture was transformed into chemically competent *E. coli* DH5α cells to obtain pEB115.

E. coli BL21 (DE3) cells harboring pEB115 was grown overnight in LB medium (5 mL,

supplemented with 50 μg/mL kanamycin) and inoculated into larger-scale LB medium (1 L, supplemented with 50 μg/mL kanamycin). Once the optical density (OD600) of the culture reached 0.6, it was chilled on ice for 30 min, and protein expression was induced with 0.1 mM IPTG (250 rpm, 16 °C, 16 h).

Cells were harvested by centrifugation (4300 rpm, 15 min), re-suspended in 30 mL A10 buffer (10 mM imidazole, 50 mM Tris-HCl, 100 mM NaCl; pH 8.0) and lysed by sonication on ice. The insoluble debris was removed from the lysate by centrifugation (17,000 g, 15 min, 4 °C), and the soluble fractions were incubated with the Ni-NTA resin (Qiagen) (4 °C, 12 h) for protein loading. The protein was eluted by sequentially washing the resin with buffers containing increasing concentrations of imidazole (50 mM, 100 mM, 200 mM; 50 mL per wash). The purified proteins were concentrated and exchanged into storage buffer (50 mM phosphate buffer, 100 mM NaCl, 10% glycerol, pH 8.0) with Centriprep filters (Amicon) and flash-frozen in liquid N_2 as 10-μL aliquots for storage at –80 °C until use.

To verify the proposed *N*-methyltransferase activity of SmbG, conversion of **3** to **1** in the presence of SmbG and *S*-adenosylmethionine (SAM) was monitored by LC-MS. 10 μL reaction mixtures with 1 mM SAM and varying concentrations of **3** were prepared, and SmbG was added at a final concentration of 1 μM to initiate the reactions. The reactions were also run with varying concentrations of **13** to explore the substrate specificity of SmbG.

4.10 Spectrometric analysis

NMR spectra were obtained on the Bruker AV500 spectrometer with 5 mm dual cryoprobe at the UCLA Molecular Instrumentation Center (^{H} 500 MHz, ^{13}C 125 MHz). Chemical shifts are reported in parts per million (ppm) using the resonance of deuterated solvent as reference ($CDCl_3$: $\delta_H = 7.26$, $\delta_C = 77.16$; DMSO-d_6: $\delta_H = 2.50$, $\delta_C = 39.52$). Optical rotation was measured on the Rudolph Research Autopol III automatic polarimeter.

4.11 Electron diffraction data collection, refinement and statistics for compound 13

Compound thp was recrystallized by slow evaporation of the HPLC fraction in MeCN/H_2O. Microcrystals were deposited onto a pure carbon EM grid and transferred to the FEI Tecnai F200C transmission electron microscope (operating voltage of 200 keV, wavelength of 0.025 Å) on a Gatan 626 cryo-holder at ambient temperature (273 K). The sample was cooled to cryogenic temperature (100 K) after insertion. Diffraction data acquisition, conversion and reduction were performed as described previously.[2] Data from two crystals were scaled and merged together to produce the final data set for **13**, before converting the intensities to SHELX format. Structure of **13** was solved ab initio using direct methods in SHELXD[3] and refined with SHELXLin ShelXle,[4] refining all non-hydrogen atoms anisotropically and placing hydrogen atoms using the riding model. Residual density belonging to solvent molecules was observed, but solvent molecules could not be placed due to disorder.

Table S1. Primers used in this study.

Primer	Sequence 5'→3'
pYTU_SmbA_F1	CTTCATCCCCAGCATCATTACACCTCAGCAATGCCATCCCCAGGCGAG
pYTU_SmbA_R1	TTGAGATCAGCTTGCTCAGG
pYTU_SmbA_F2	ATGGCTATGGCTATCCAAGC
pYTU_SmbA_R2	AAAAGCGAGGGTTGTTGACC
pYTU_SmbA_F3	ACTTGTATGAGGGCATCAGC
pYTU_SmbA_R3	ACTGTCGTCTCCATCATCGT
pYTU_SmbA_F4	TTACGCCGAACAATGACACG
pYTU_SmbA_R4	GAGGACATACCCGTAATTTTCTGGCGGCCGCCTACGTGTTAATTAACACATTCCAGTCTG
pYTR_SmbB_F	ATTACCCCGCCACATAGACACATCTAAACAATGGCAGTCCAATCAATCATC
pYTR_SmbB_R	TGCTAAAGGGTATCATCGAAAGGGAGTCATCCAGGATCCCAAAAGCAAGGTTTGAAGTGC
pYTR_SmbC_F	GCATACAGAACACTTCAAACAATCGCAAAAATGTCATCTCGTCTCTCCG
pYTR_SmbC_R	AGACCCAACAACCATGATACCAGGGGGCTAGCTTAATTAAAGCCATTGATGGTGCTTGTG
pYTP_SmbD_F	GGGGTGTGATGACGACATTGTTTAGATGTGTCTATGTGGC
pYTP_SmbD_R	AACAACCATGATACCAGGGGGCTAGCCGATACTTAATTAACTGTAACCGCTATCTGATGC
pYTP_SmbE_F	CTTATACATGATCTAACAATTCTAGTAAACCGCAATCATGACAAACCCCACAAACATTAC
pYTP_SmbE_R	AACAACCATGATACCAGGGGGGATCCTAGTCCTTAATTAAAAGGTTCCGTAAGCTTGACG
pYTP_SmbF_F	ACAGAACACTTCAAACAATCGCAAAAATGGAAAATTTGCAATTGAACG
pYTP_SmbF_R	AACAACCATGATACCAGGGGCTAGCGGTACCTTAATTAATTGATATTGCGCTTACGTCTG
pYTR_SmbG_F1	ACTCCGGTGAATTGATTTGGG
pYTR_SmbG_R1	GGGAGATTGGTTCACTGCTTCAGTAGCCATTGTTTAGATGTGTCTATGTGGC
pYTR_SmbG_F2	ATGGCTACTGAAGCAGTG
pYTR_SmbG_R2	AGACCCAACAACCATGATACCAGGGGGGATCCTTAATTAAAATCCTACCAATATTAGCGTTGC
pYTR_SmbBC_F	CCAACTTTAATACGCAATAAGACACCATAGTTTGCTCCAGGAATACATGTG
pYTR_SmbBC_R	TCATCGAAAGGGAGTCATCCAGGATCCAGCGCTTAATTAATGTACTAGCTGGGAAGATCG
pYTU_SmbABC_F1	ATTCAGAGAAGGTACTAAGGATGTCGATGGACTCCGGTGAATTGATTTGGG
pYTU_SmbABC_R1	TCGAGCTGCATTCATAGTCC
pYTU_SmbABC_F2	ACACTTCAGGGTTGTGGAAAATATCAGGCTTTTGCTCCAGGAATACATGTG
pYTU_SmbABC_R2	GGACATACCCGTAATTTTCTGGCGGCCGCTACTTAATTAAACCTCTGCACCAGTTTATCC
pYTP_SmbDE_F	CAGGGTCGCGATGCATCAGATAGCGGTTACAGGATTCGTCCAGGGCTTCC
pYTP_SmbDE_R	AACAACCATGATACCAGGGGGGATCCTAGTCCTTAATTAAAAGGTTCCGTAAGCTTGACG
pYTP_SmbDEF_F	GGACTGATTCGTCAAGCTTACGGAACCTTTTTGCTCCAGGAATACATGTG
pYTP_SmbDEF_R	ACAACCATGATACCAGGGGGGATCCTAGTCCTTAATTAATTGATATTGCGCTTACGTCTG
pET28a_SmbG_F	TAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGCTACTGAAGCAGTGAAC
pET28a_SmbG_R	GCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCCTTCTTAGGCTTCAATCCTC
XW55_backbone_F	CTCGAGGGTACCGAGCTCG
XW55_backbone_R	GGCCGCTAAATGGATCCGAGC
pTrpC_F	TCGACAGAAGATGATATTGAAGGAGCACT
tTrpC_R	AAGAAGGATTACCTCTAAACAAGTGTACCTGT
hph_F	GAGTTGGTCAAGACCAATGC
hph_R	AATGTCCTGACGGACAATGG
KO_FunL_upF	AAGCTGGAGCTCGGATCCATTTAGCGGCCCGTTCTTCCTTTCACTCCACG
KO_FunL_upR	AAGTGCTCCTTCAATATCATCTTCTGTCGATGTGCAAGTCTCGATTGCGA
KO_FunL_downF	CACTTGTTTAGAGGTAATCCTTCTTGTTTTCATATTCATTTAAATCCAAG
KO_FunL_downR	GCCAGTGAATTCGAGCTCGGTACCCTCGAGGACGACCTTAGCTTTGATCG
KO_FunG_upF	AGCTGGAGCTCGGATCCATTTAGCGGCCCCGCTTCGCATTTGATACTCAG
KO_FunG_upR	AAGTGCTCCTTCAATATCATCTTCTGTCGAGGTGTACCCAGATGTGATAG
KO_FunG_downF	GTACACTTGTTTAGAGGTAATCCTTCTTACGGAACATCTTCATCGAATTC
KO_FunG_downR	CCAGTGAATTCGAGCTCGGTACCCTCGAGCCCAATGAGCATCTTCATTCG

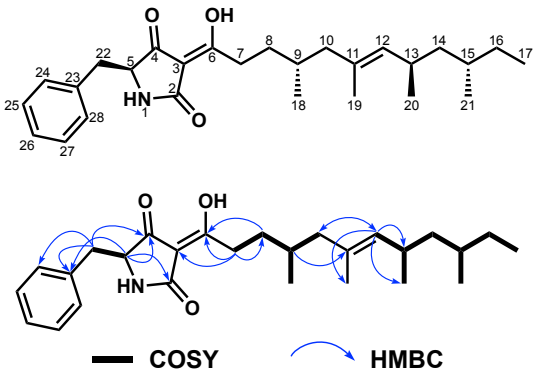
Table S2. Plasmids used in this study.

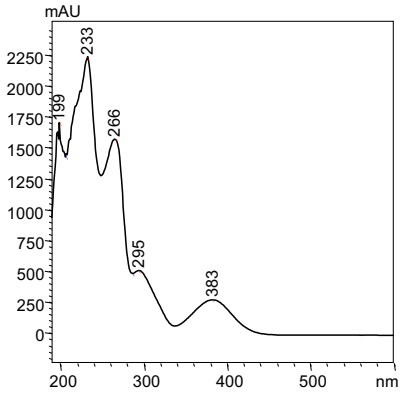
Plasmid	Vector	Description
pEB101	pYTP	Vector only
pEB102	pYTR	Vector only
pEB103	pYTU	Vector only
pEB104	pYTU	glaA-SmbA
pEB105	pYTR	ANgpdA-SmbB
pEB106	pYTP	POgpdA-SmbC
pEB107	pYTP	ANgpdA-SmbD
pEB108	pYTP	PEgpdA-SmbE
pEB109	pYTP	POgpdA-SmbF
pEB110	pYTR	ANgpdA-SmbG
pEB111	pYTR	ANgpdA-SmbB; POgpdA-SmbC
pEB112	pYTU	glaA-SmbA; ANgpdA-SmbB; POgpdA-SmbC
pEB113	pYTP	ANgpdA-SmbD; PEgpdA-SmbE
pEB114	pYTP	ANgpdA-SmbD; PEgpdA-SmbE; POgpdA-SmbF
pEB115	pET28a	T7-SmbG-*N*-His
pEB116	XW55	pTrpC-hph-tTrpC marker flanked by 2,000-bp homology to regions downstream & upstream of funG in TTI-0630
pEB117	XW55	pTrpC-hph-tTrpC marker flanked by 2,000-bp homology to regions downstream & upstream of funL in TTI-0630

Table S3. *A. nidulans* protoplasts prepared in this study.

Transformant	Description (plasmids)	Major compounds produced
AnEB101	pEB101, pEB104, pEB105	**7**
AnEB102	pEB104, pEB105, pEB106	**8**
AnEB103	pEB104, pEB111, pEB107	
AnEB104	pEB104, pEB111, pEB113	**13**
AnEB105	pEB104, pEB111, pEB114	**3**
AnEB106	pEB112, pEB110, pEB114	**1** (detected)
AnEB107	pEB104, pEB111, pEB108	
AnEB108	pEB104, pEB111, pEB109	

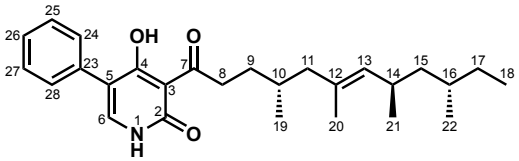
Table S4. Structural characterization of 7. ^{1}H-NMR (500 MHz), ^{13}C-NMR (125 MHz), DMSO-d_6.

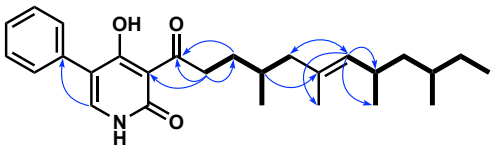


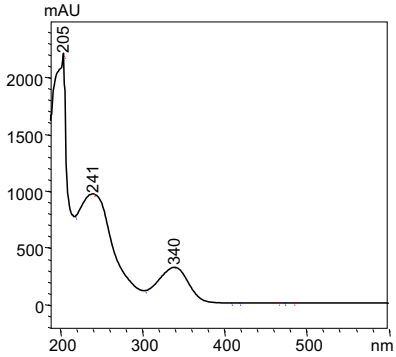


Position	δ_H	δ_C
1	--	--
2	--	174.6
3	--	101.1
4	--	194.6
5	4.16 (t, *J* = 5.0 Hz)	61.9
6	--	188.7
7	2.73 (m), 2.65 (m)	29.9
8	1.47 (m), 1.24(m)	32.2
9	1.55	29.4
10	1.94 (m), 1.68 (m)	46.9
11	--	131.0
12	4.85 (d, *J* = 9.5 Hz)	133.2
13	2.41 (m)	29.4
14	1.17 (m), 1.02 (m)	44.6
15	1.32 (m)	31.5
16	1.32 (m), 1.02 (m)	28.4
17	0.81 (t, *J* = 6.0 Hz)	11.1
18	0.77 (d, *J* = 6.5 Hz))	18.8
19	1.51 (d, *J* = 1.0 Hz)	15.7
20	0.85 (d, *J* = 7.0 Hz))	21.2
21	0.81 (d, *J* = 7.0 Hz)	19.6
22	2.96 (m), 2.90 (m)	36.6
23	--	135.8
24/28	7.14-7.25 (m)	129.7
25/27	7.14-7.25 (m)	128.0
26	7.14-7.25 (m)	126.5

Table S5. Structural characterization of 8. ^{1}H-NMR (500 MHz), ^{13}C-NMR (125 MHz), $CDCl_3$.

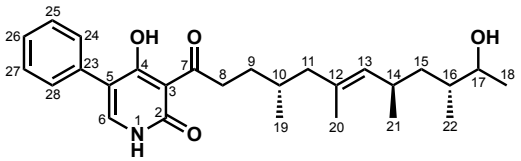


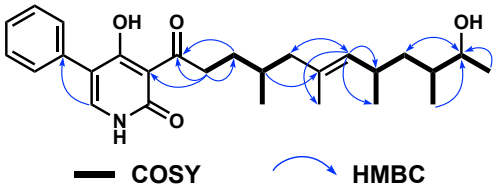


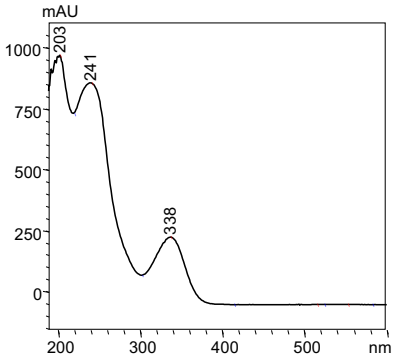


Position	δ_H	δ_C
1	--	--
2	--	163.1
3	--	106.2
4	--	176.3
5	--	115.3
6	7.42 (s)	138.5
7	--	209.2
8	3.25 (m)	40.7
9	1.49 (m), 1.72 (m)	31.3
10	1.71 (m)	30.5
11	1.07 (m), 1.77 (m)	47.8
12	--	131.4
13	4.89 (d, *J* = 9.8 Hz)	133.8
14	2.45 (m)	29.8
15	1.02, 1.18	45.1
16	1.29 (m)	32.0
17	1.03 (m), 1.33 (m)	29.1
18	0.84 (m)	11.3
19	0.74 (d, *J* = 6.6 Hz)	19.6
20	1.56 (s)	15.9
21	0.90 (d, *J* = 6.6 Hz)	21.1
22	0.87 (m)	19.2
23	--	132.8
24/28	7.35-7.46 (m)	129.0
25/27	7.35-7.46 (m)	128.5
26	7.35-7.46 (m)	128.0

Table S6. Structural characterization of 8'. ^{1}H-NMR (500 MHz), ^{13}C-NMR (125 MHz), $CDCl_3$.

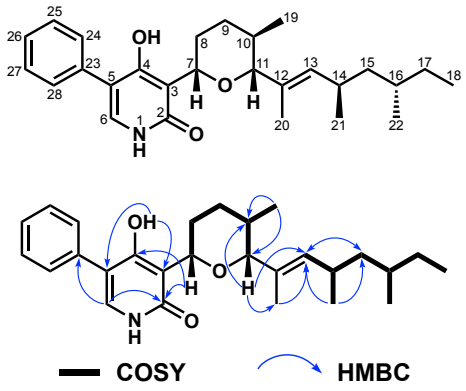


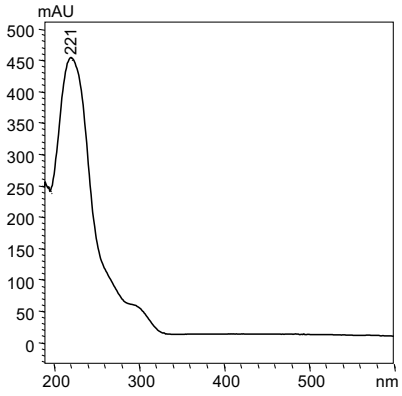




Position	δ_H	δ_C
1	--	--
2	--	163.7
3	--	106.7
4	--	176.2
5	--	115.4
6	7.43 (s)	138.8
7	--	209.2
8	3.25 (m)	40.8
9	1.47 (m), 1.75 (m)	31.3
10	1.73 (m)	30.6
11	1.80 (m), 2.04(m)	47.7
12	--	131.8
13	4.93 (d, J = 9.4 Hz)	138.8
14	2.45 (m)	29.8
15	1.07 (m), 1.36 (m)	40.8
16	1.48 (m)	37.1
17	3.72 (m)	70.7
18	1.13 (m)	20.8
19	0.87 (m)	19.4
20	1.57 (s)	16.1
21	0.90 (d, J = 6.6 Hz)	20.5
22	0.87 (m)	14.1
23	--	132.6
24/28	7.35-7.46 (m)	129.0
25/27	7.35-7.46 (m)	128.5
26	7.35-7.46 (m)	128.0

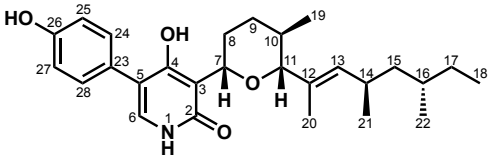
Table S7. Structural characterization of 16. ^{1}H-NMR (500 MHz), ^{13}C-NMR (125 MHz), $CDCl_3$.

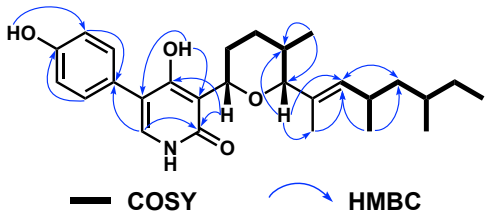


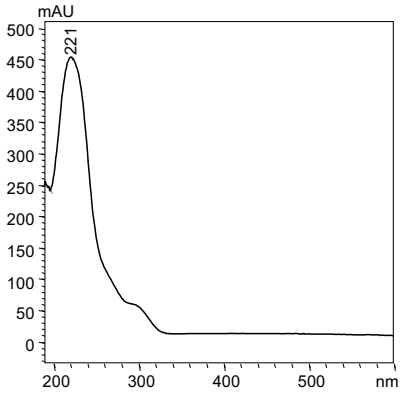


Position	δ_H	δ_C
1	--	--
2	--	163.7
3	--	110.4
4	--	163.8
5	--	116.0
6	7.28 (s)	134.1
7	5.03 (d, *J* = 11.5 Hz)	76.9
8	1.69 (m), 2.09 (m)	30.9
9	1.46 (m), 1.93 (m)	32.1
10	1.70 (m)	32.4
11	3.54 (d, *J* = 10.0 Hz)	92.8
12	--	130.3
13	5.20 (d, *J* = 9.5 Hz)	138.2
14	2.48 (m)	29.7
15	1.05 (m), 1.21 (m)	44.8
16	1.33 (m)	32.3
17	1.05, 1.33 (m)	29.0
18	0.84 (m)	11.7
19	0.76 (d, *J* = 6.6 Hz)	17.8
20	1.63 (s)	11.3
21	0.91 (d, *J* = 6.6 Hz)	20.8
22	0.84 (m)	19.7
23	--	133.1
24/28	7.39 (d, *J* = 7.8 Hz)	129.3
25/27	7.45 (m)	128.4
26	7.32 (m)	127.5

Table S8. Structural characterization of 3. ^{1}H-NMR (500 MHz), ^{13}C-NMR (125 MHz), $CDCl_3$.







Position	δ_H	δ_C
1	--	--
2	--	163.5
3	--	110.4
4	--	164.4
5	--	116.4
6	6.74 (s)	132.8
7	5.03 (d, *J* = 11.5 Hz)	77.0
8	1.66 (m), 2.10 (m)	31.1
9	1.44 (m), 1.91(m)	32.1
10	1.68 (m)	32.7
11	3.54 (d, *J* = 10.0 Hz)	92.9
12	--	130.2
13	5.21 (d, *J* = 9.5 Hz)	138.4
14	2.45 (m)	29.8
15	1.02 (m), 1.18 (m)	44.9
16	1.29 (m)	32.1
17	1.03, 1.33 (m)	29.0
18	0.83 (m)	11.8
19	0.74 (d, *J* = 6.6 Hz)	17.8
20	1.62 (s)	11.4
21	0.90 (d, *J* = 6.6 Hz)	20.9
22	0.84 (m)	19.8
23	--	125.8
24/28	7.20 (d, *J* = 8.2 Hz)	130.7
25/27	6.93 (d, *J* = 8.4 Hz)	115.7
26–OH	10.21 (s)	--

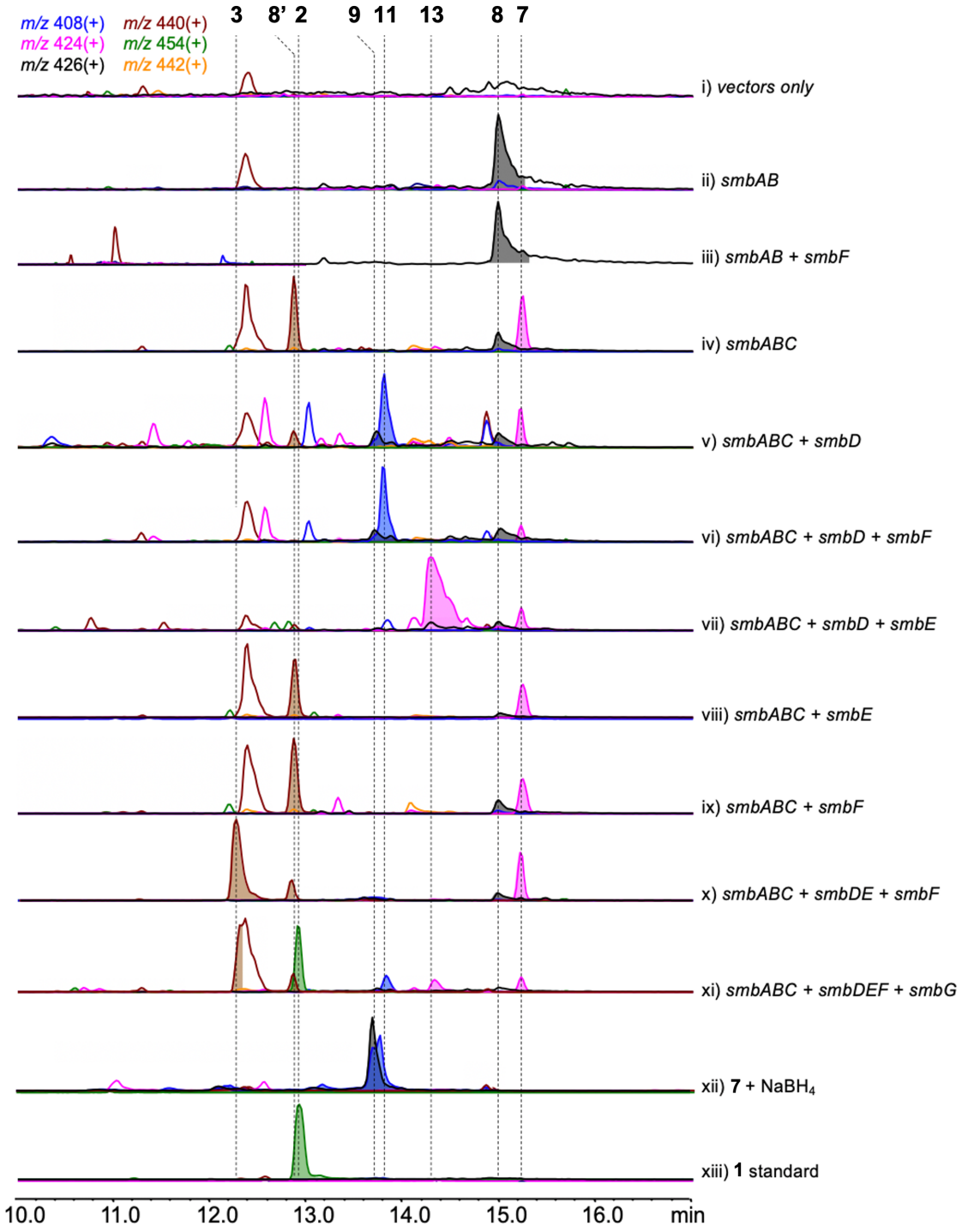


Figure S1. LC-MS analysis of metabolic extracts of *A. nidulans* expressing *smbA–G*.

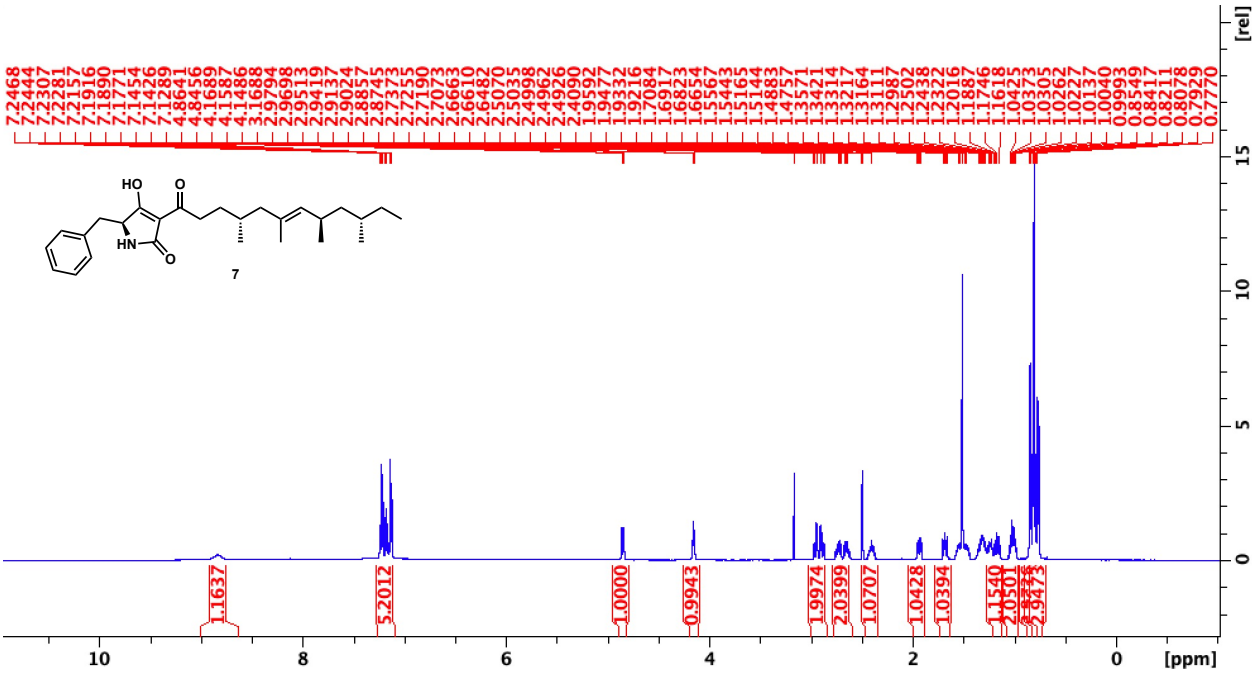


Figure S2. ^{1}H-NMR (500 MHz) spectrum of compound 6 in DMSO-d$_6$.

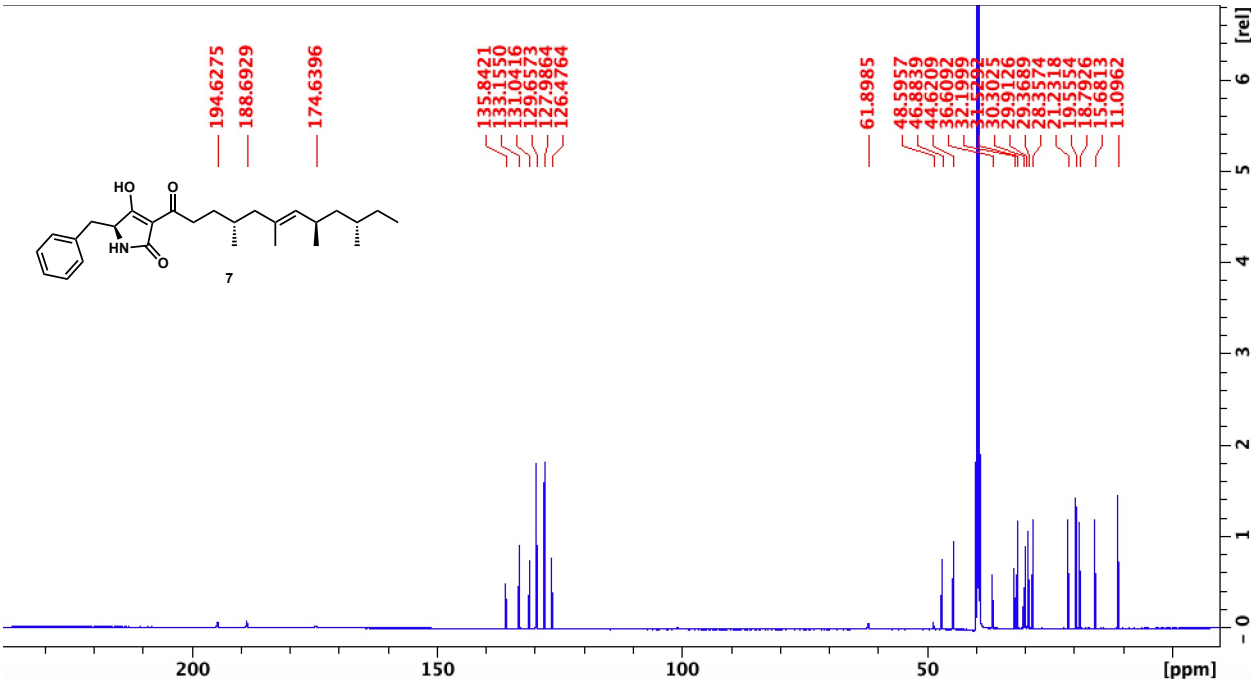


Figure S3. ^{13}C-NMR (500 MHz) spectrum of tetramic acid in DMSO-d_6.

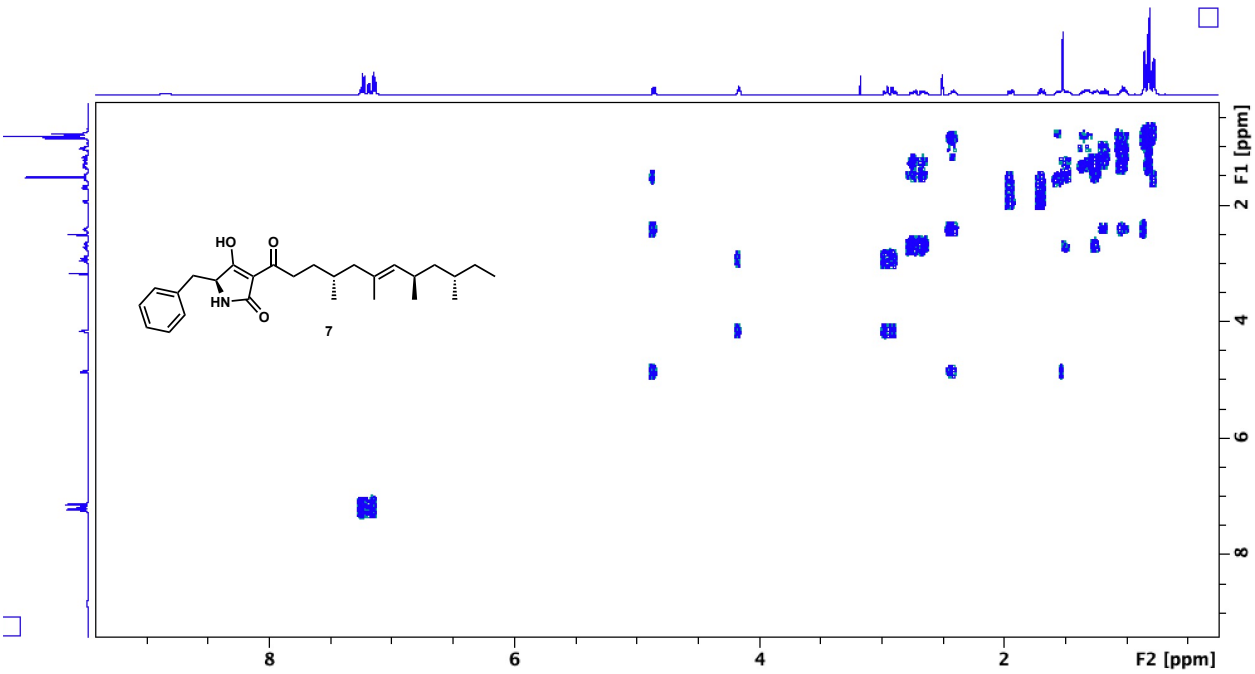


Figure S4. ^{1}H-^{1}H COSY spectrum of tetramic acid in DMSO-d_6.

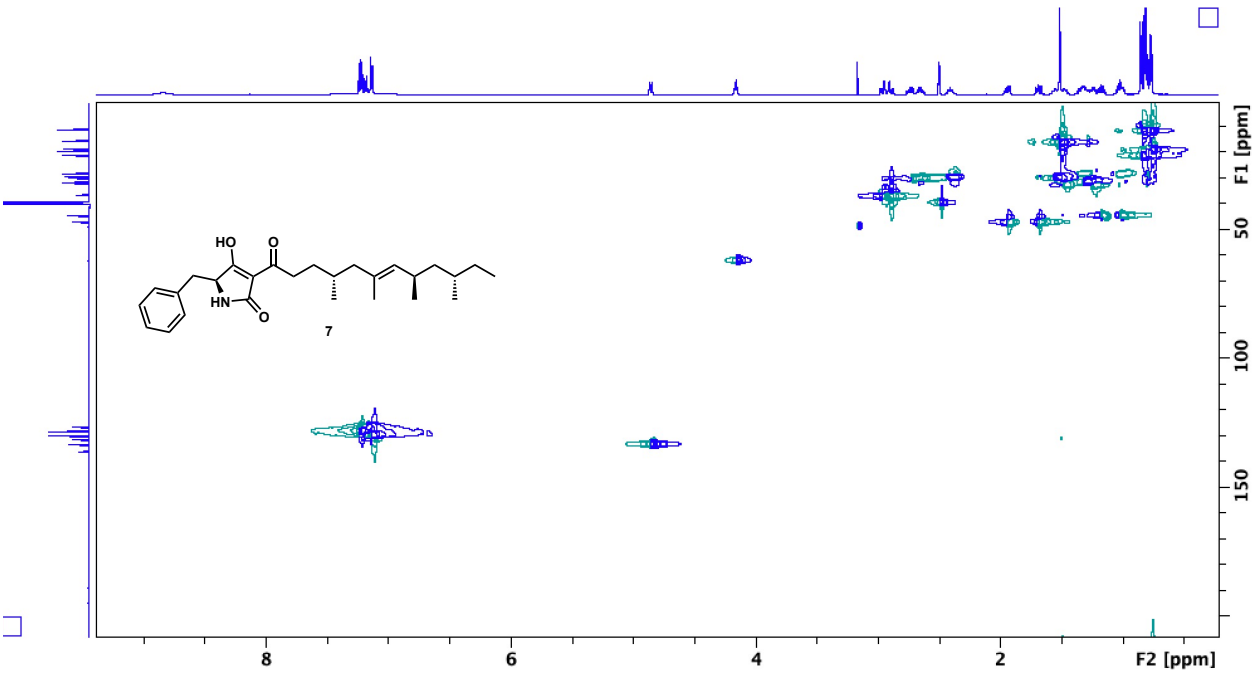


Figure S5. ^{1}H-^{13}C HSQC spectrum of tetramic acid in $CDCl_3$.

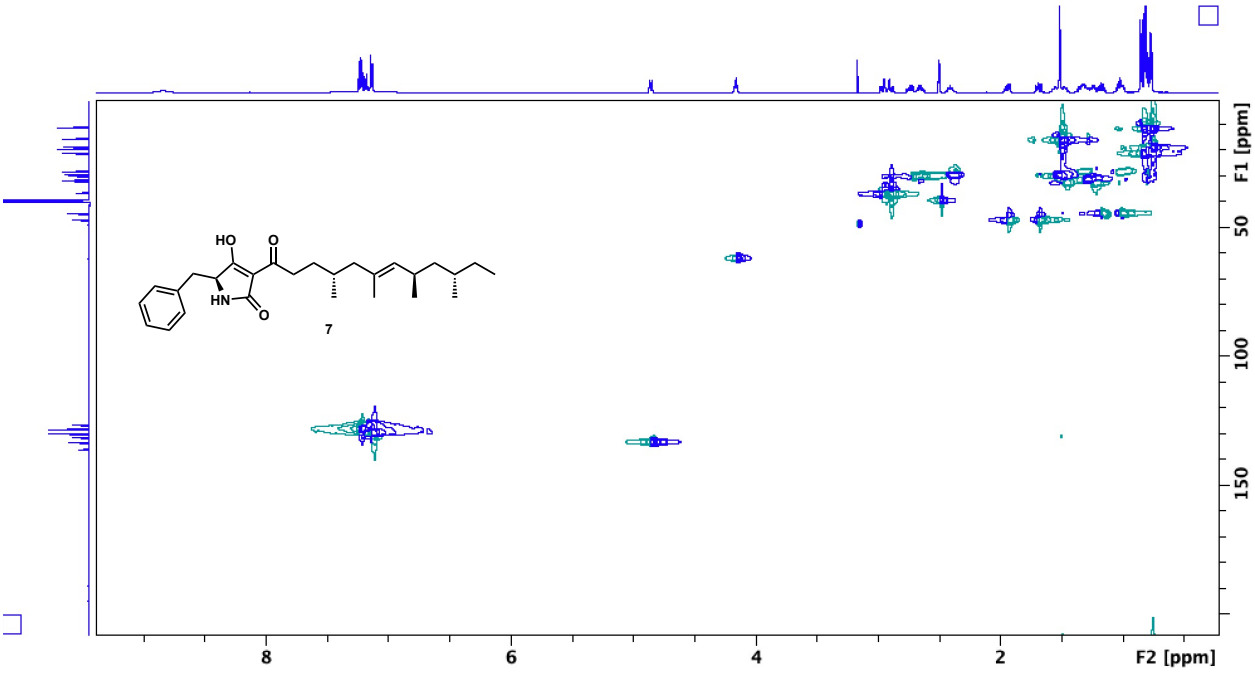


Figure S6. ^{1}H-^{13}C HMBC spectrum of tetramic acid in DMSO-d_6.

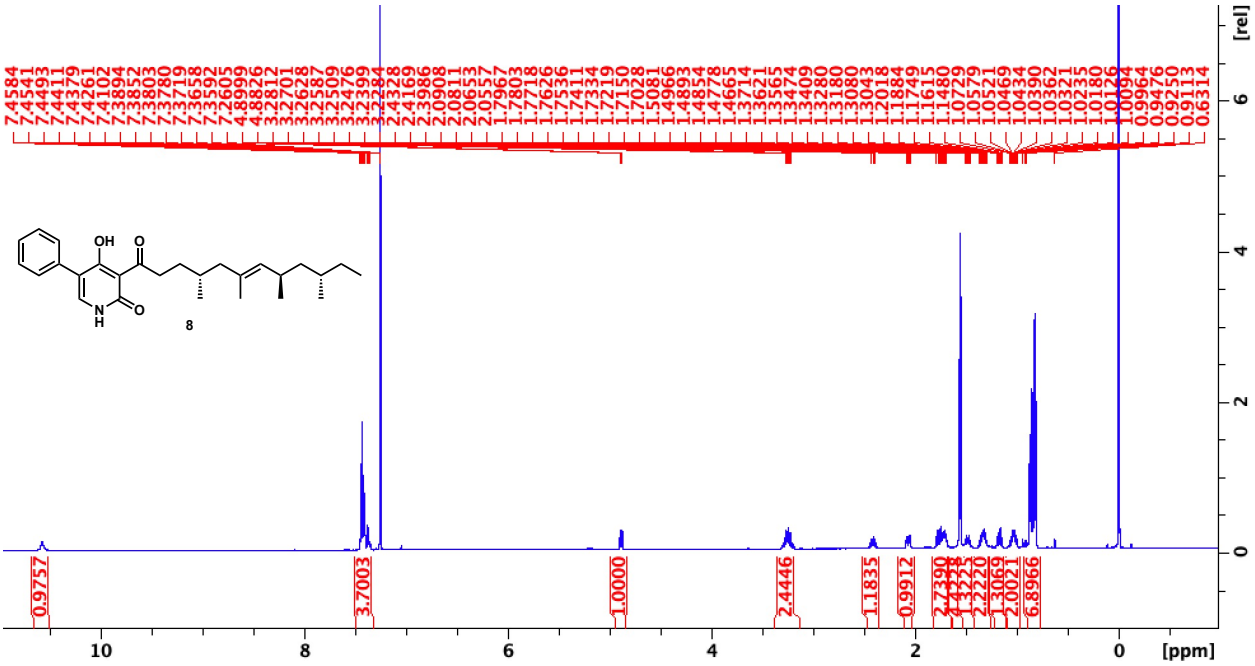


Figure S7. ^{1}H-NMR (500 MHz) spectrum of ketone in $CDCl_3$.

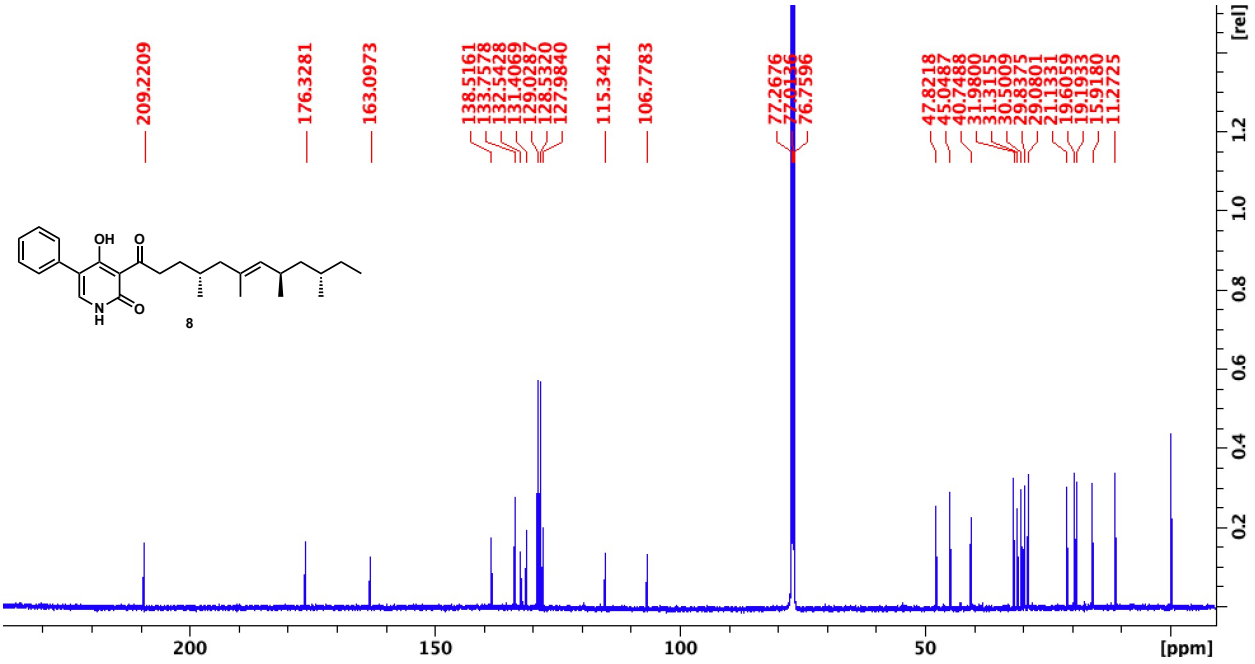


Figure S8. ^{13}C-NMR (500 MHz) spectrum of ketone in $CDCl_3$.

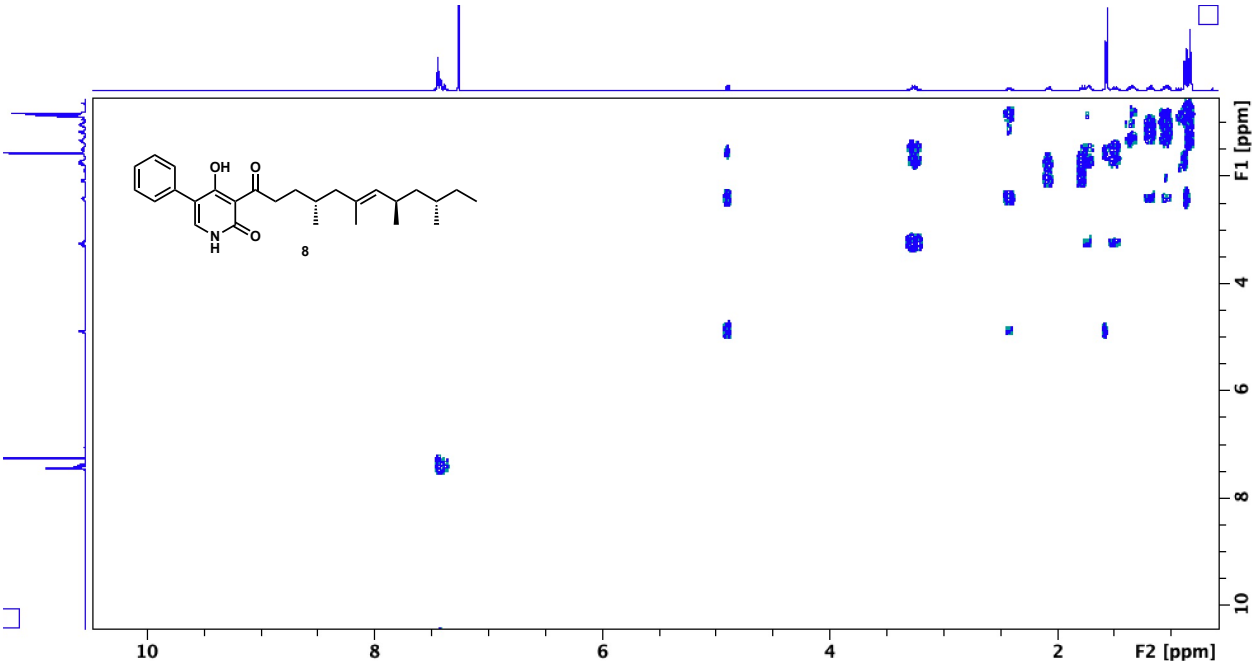


Figure S9. ^{1}H-^{1}H COSY spectrum of ketone in $CDCl_3$.

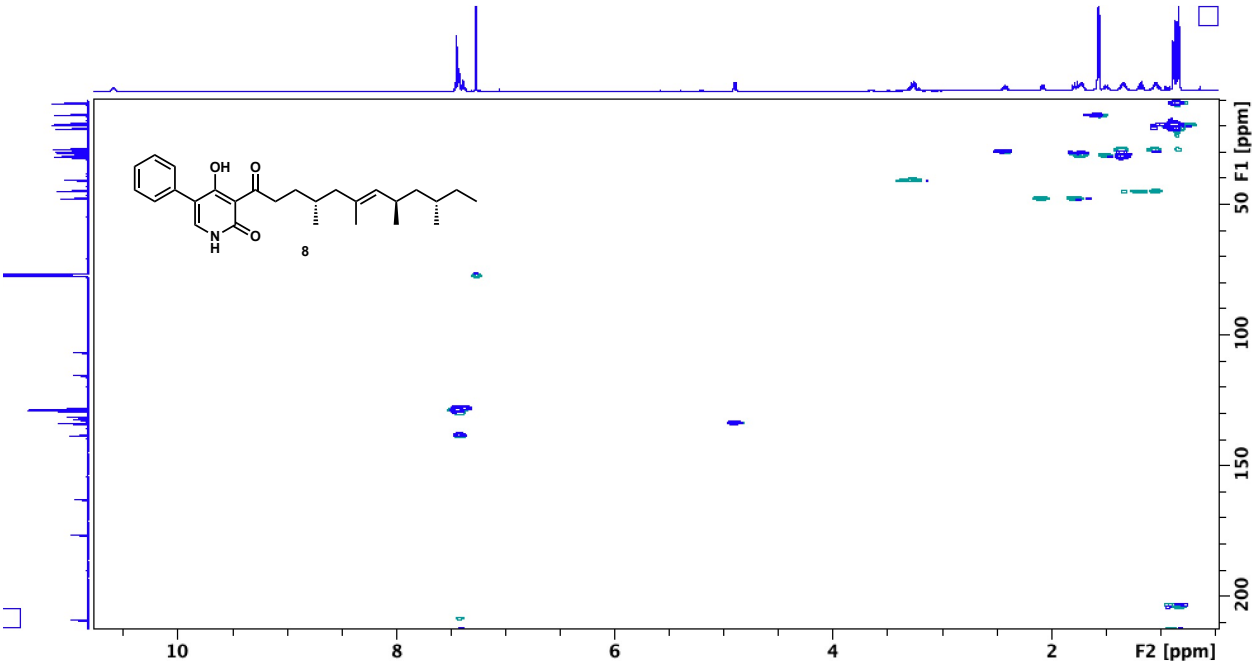


Figure S10. ^{1}H-^{13}C HSQC spectrum of ketone in $CDCl_3$.

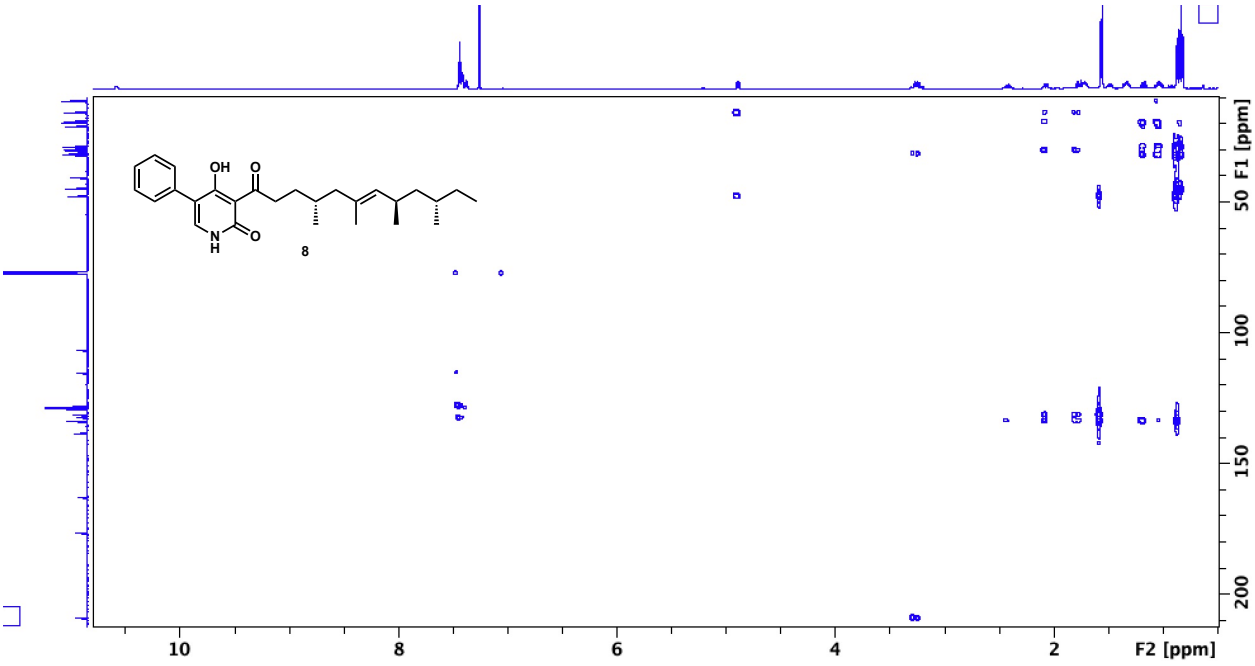


Figure S11. ^{1}H-^{13}C HMBC spectrum of ketone in $CDCl_3$.

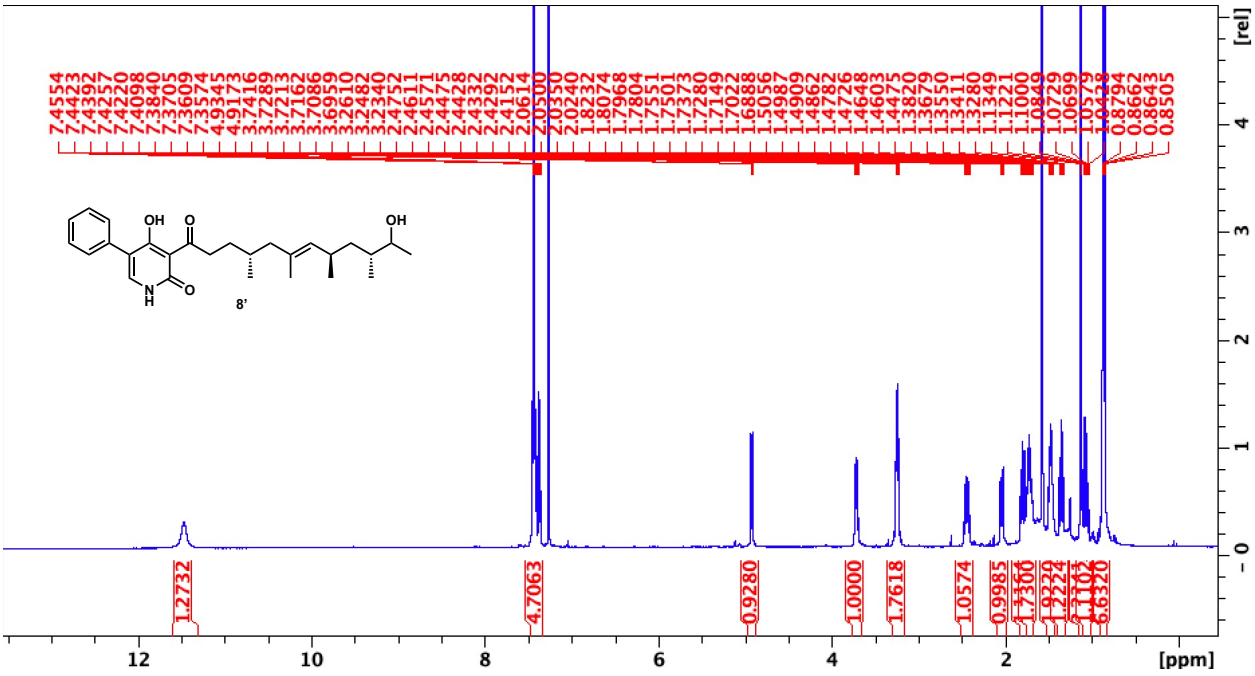


Figure S12. ^{1}H-NMR (500 MHz) spectrum of ketone-OH in $CDCl_3$.

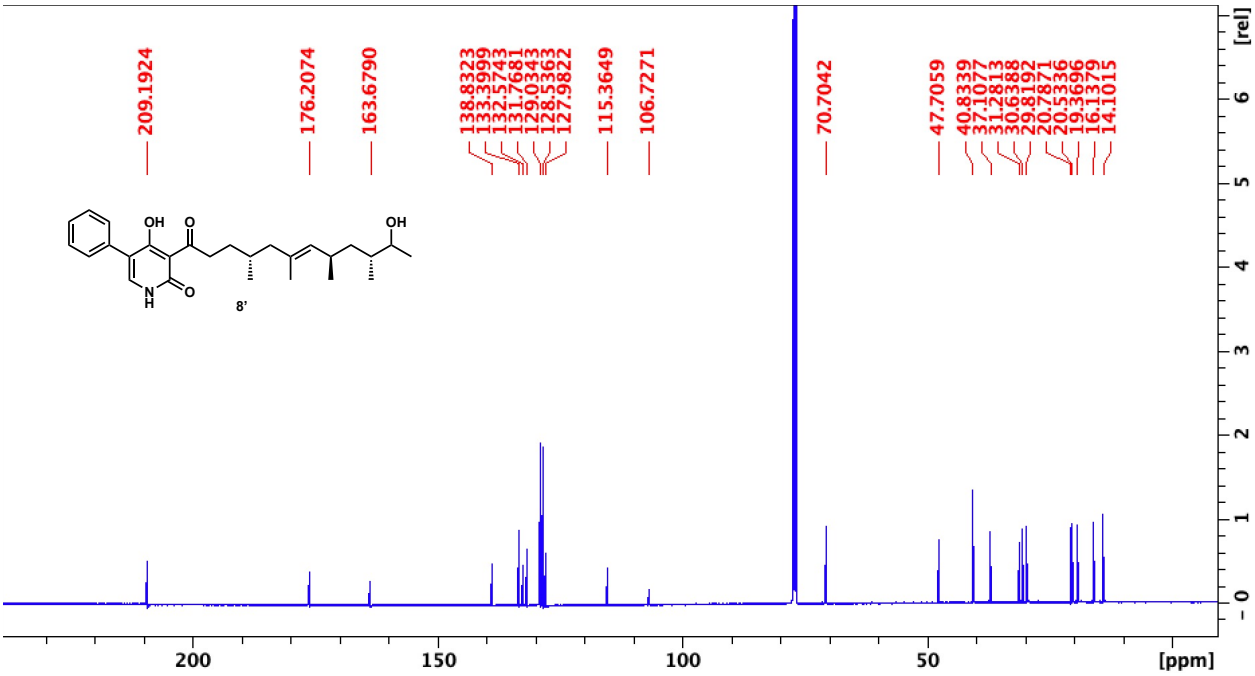


Figure S13. ^{13}C-NMR (500 MHz) spectrum of ketone-OH in $CDCl_3$.

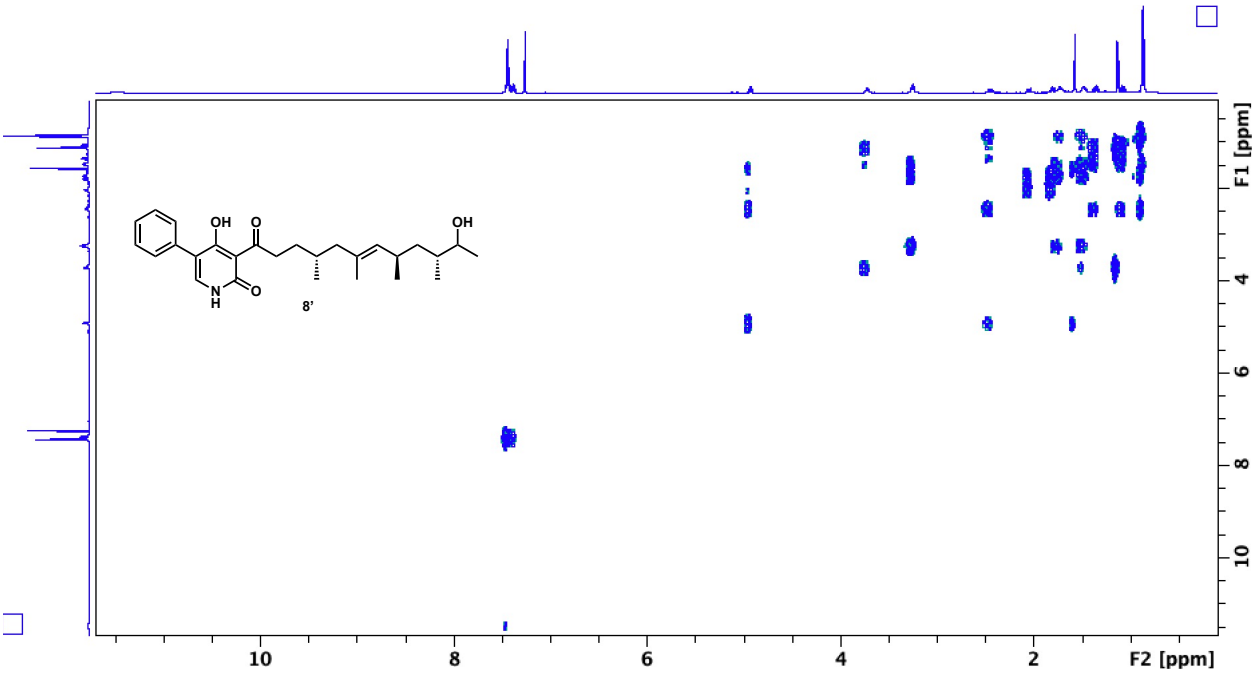


Figure S14. ^{1}H-^{1}H COSY spectrum of ketone-OH in $CDCl_3$.

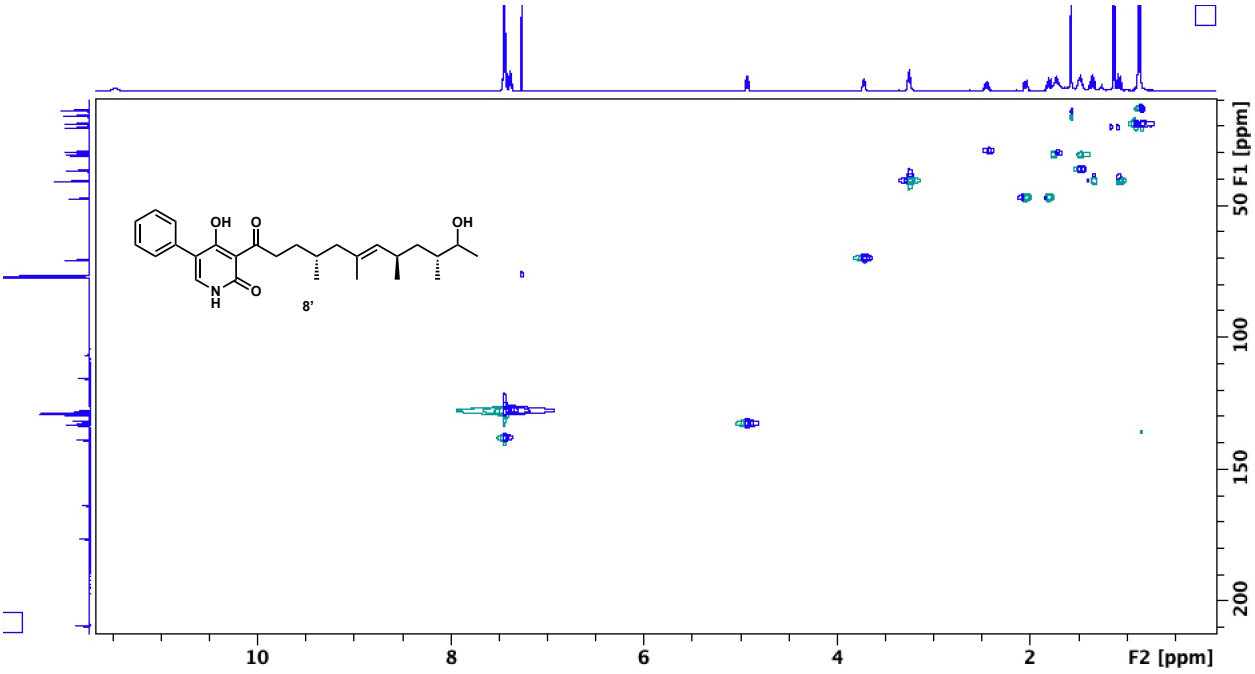


Figure S15. ^{1}H-^{13}C HSQC spectrum of ketone-OH in $CDCl_3$.

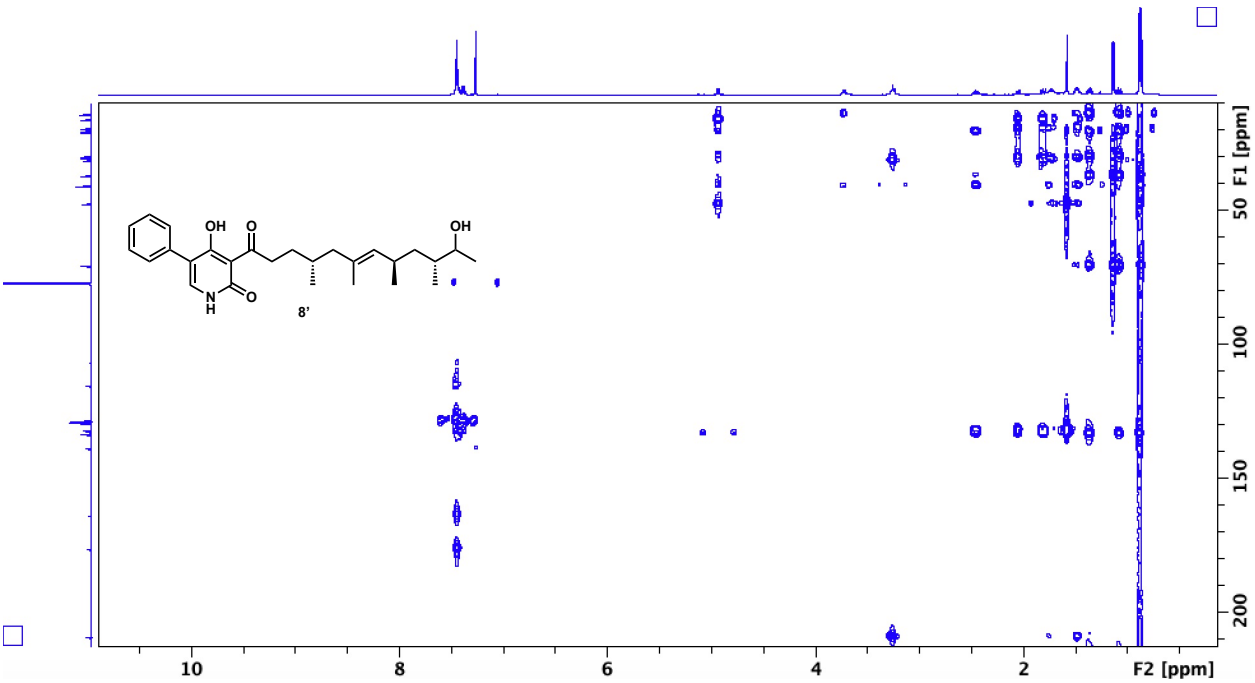


Figure S16. ^{1}H-^{13}C HMBC spectrum of ketone-OH in $CDCl_3$.

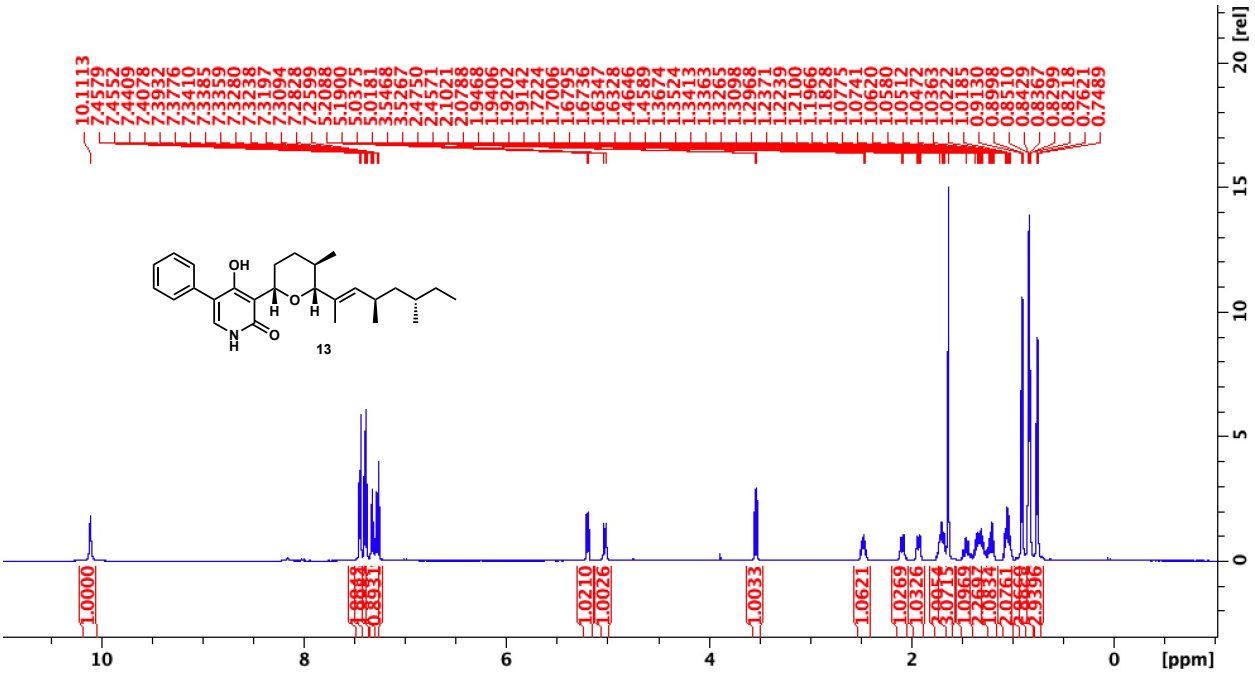


Figure S17. ^{1}H-NMR (500 MHz) spectrum of 424thp in $CDCl_3$.

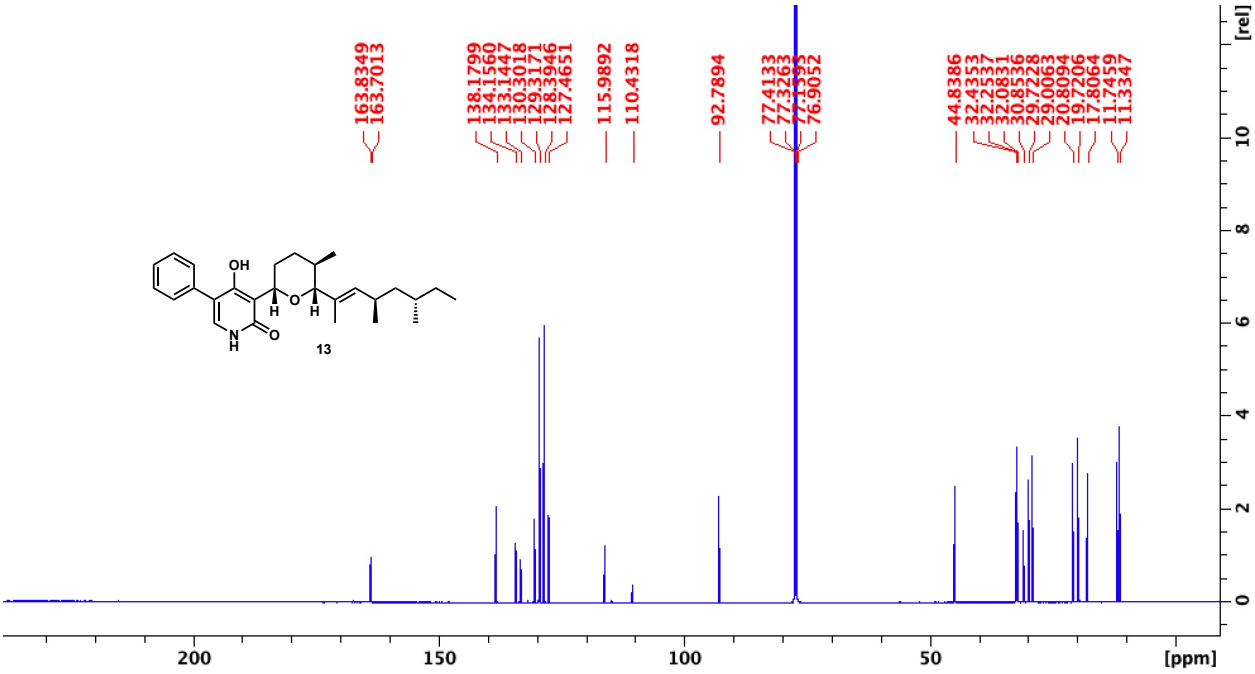


Figure S18. ^{13}C-NMR (500 MHz) spectrum of 424thp in $CDCl_3$.

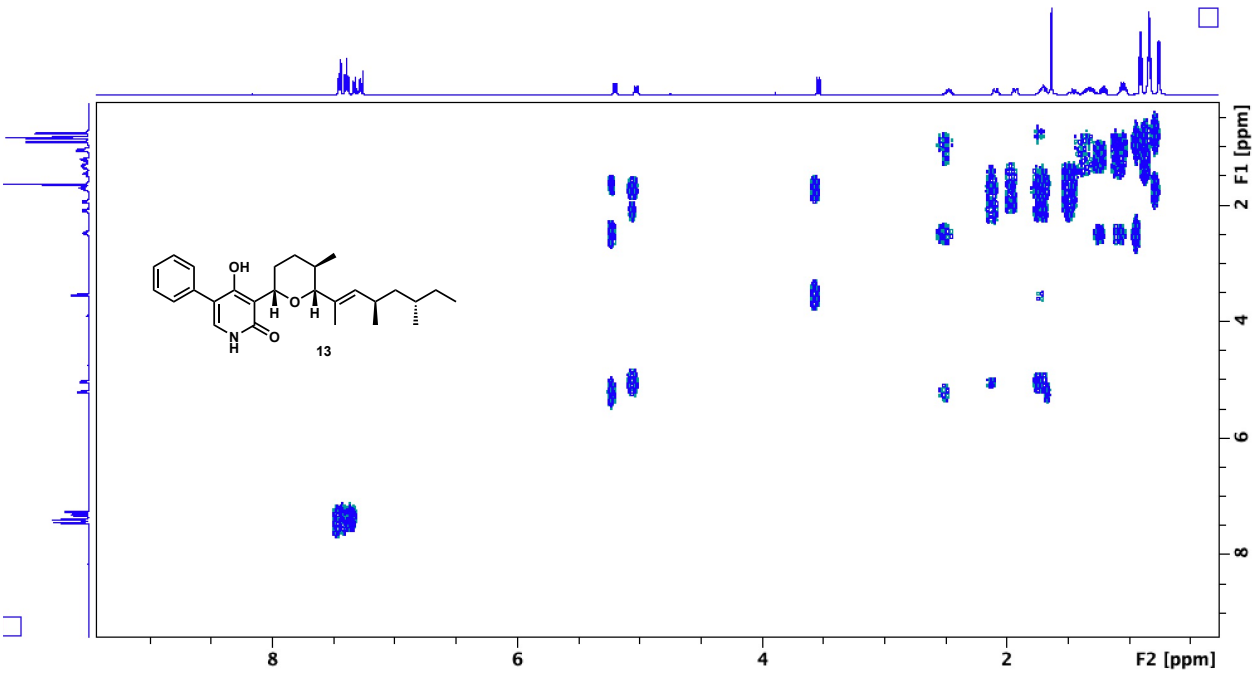


Figure S19. ^{1}H-^{1}H COSY spectrum of 424thp in $CDCl_3$.

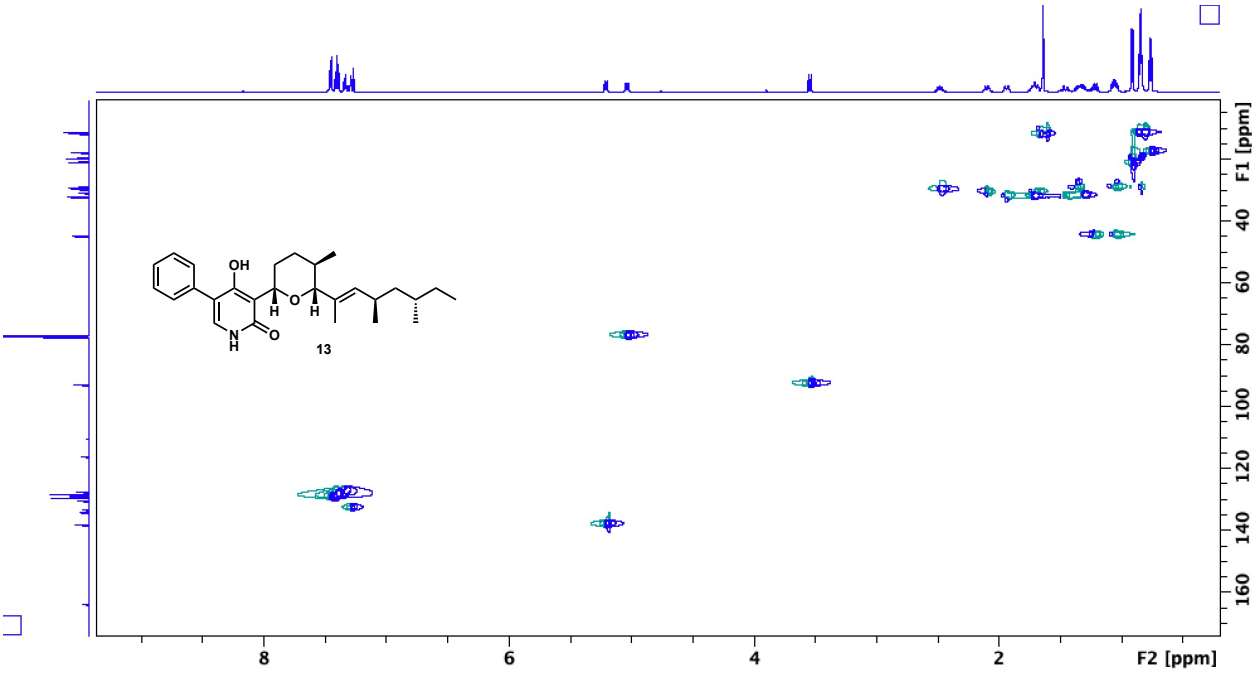


Figure S20. ^{1}H-^{13}C HSQC spectrum of 424thp in $CDCl_3$.

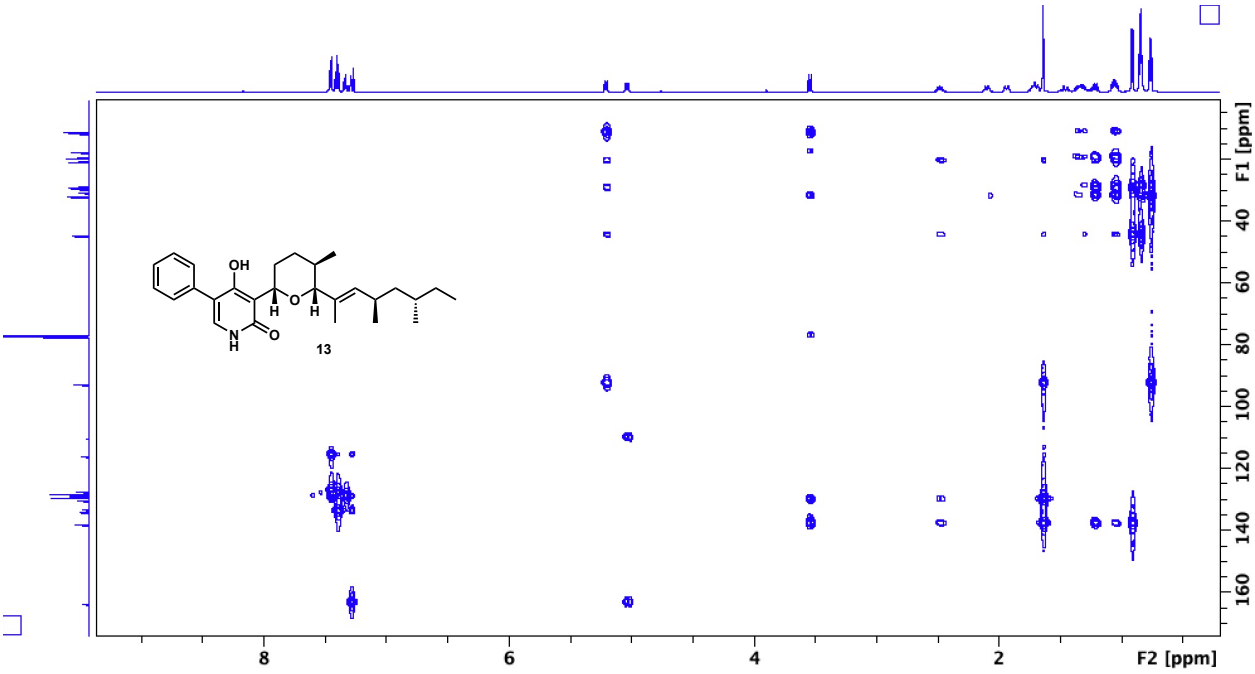


Figure S21. ^{1}H-^{13}C HMBC spectrum of 424thp in $CDCl_3$.

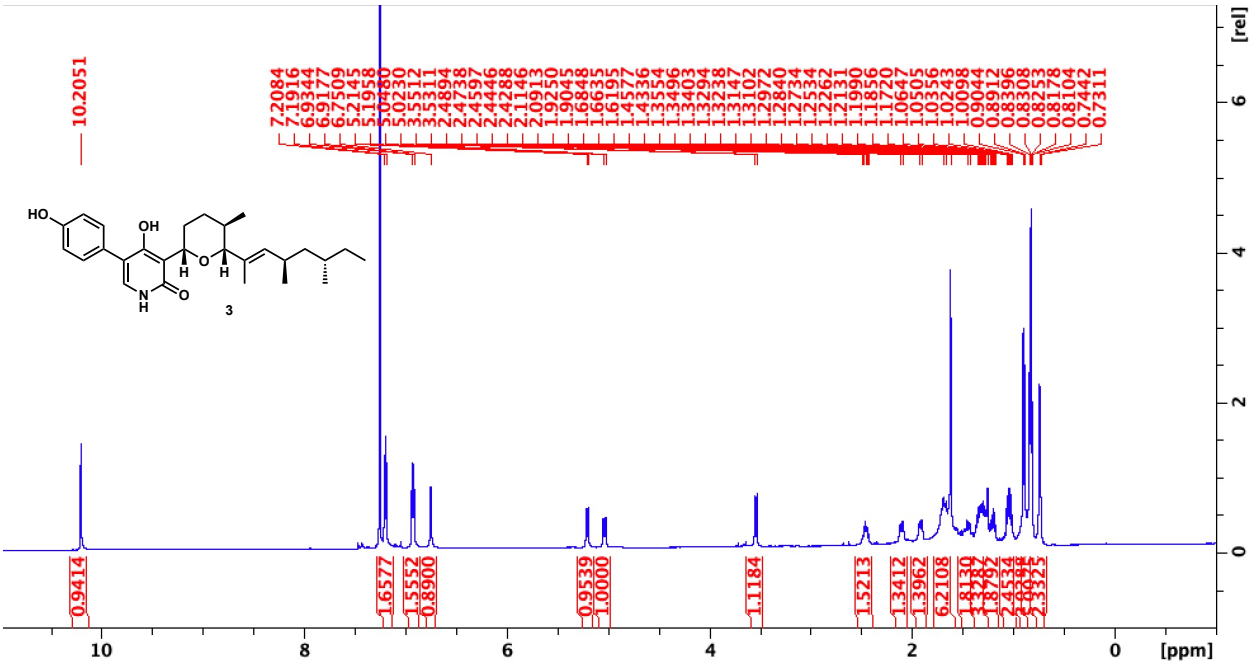


Figure S22. ^{1}H-NMR (500 MHz) spectrum of 440thp in $CDCl_3$.

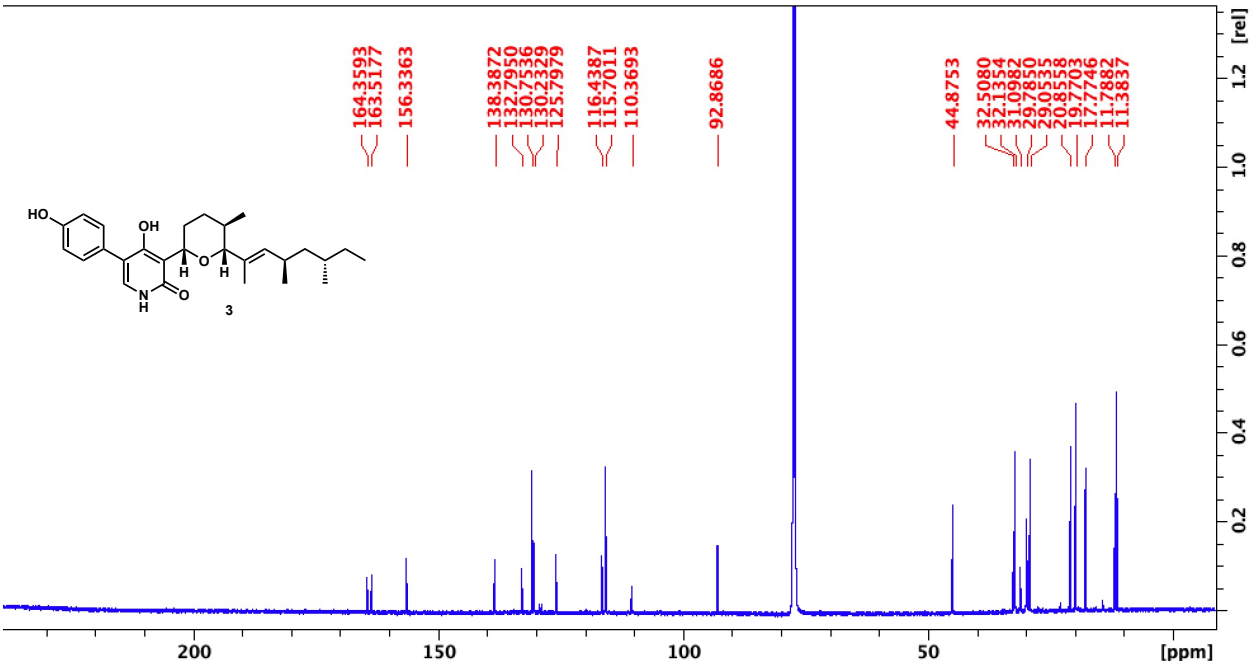


Figure S23. ^{13}C-NMR (500 MHz) spectrum of 440thp in $CDCl_3$.

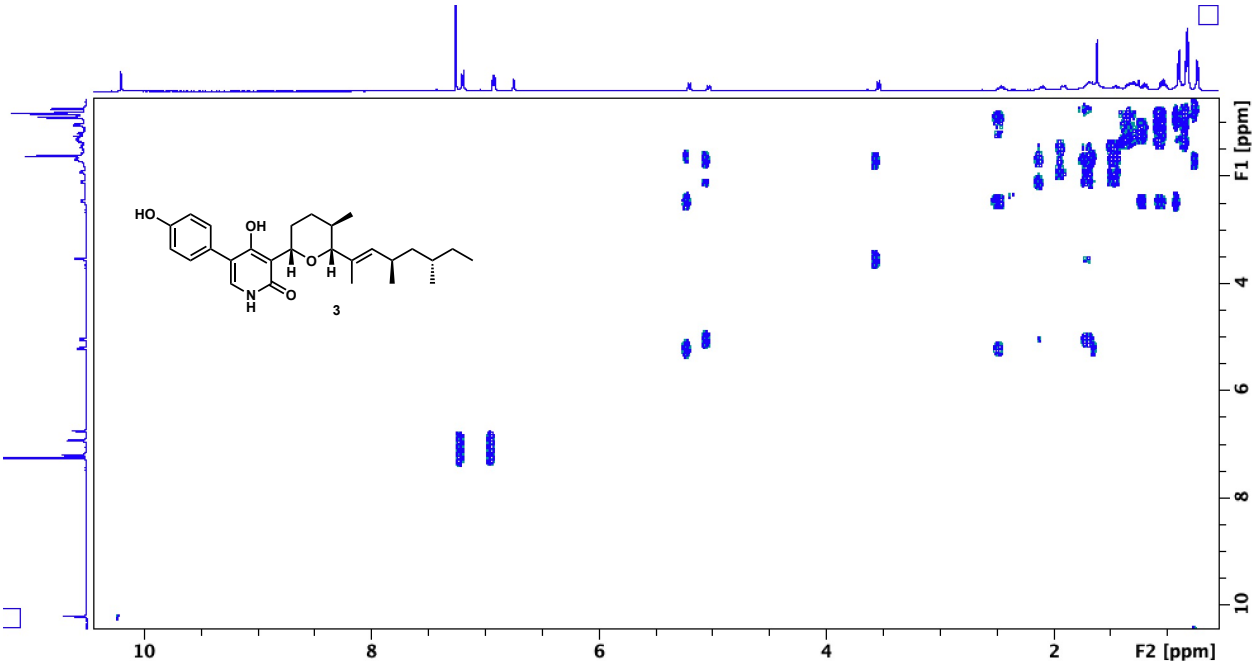


Figure S24. ^{1}H-^{1}H COSY spectrum of 440thp in $CDCl_3$.

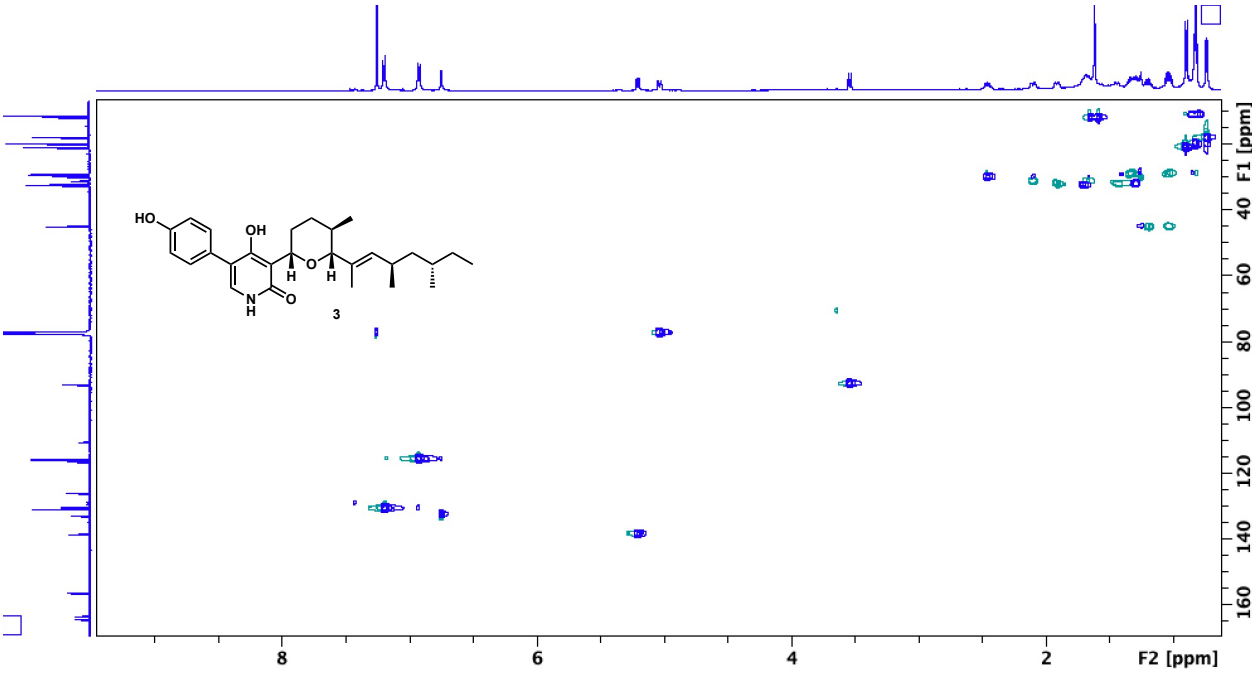


Figure S25. ^{1}H-^{13}C HSQC spectrum of 440thp in $CDCl_3$.

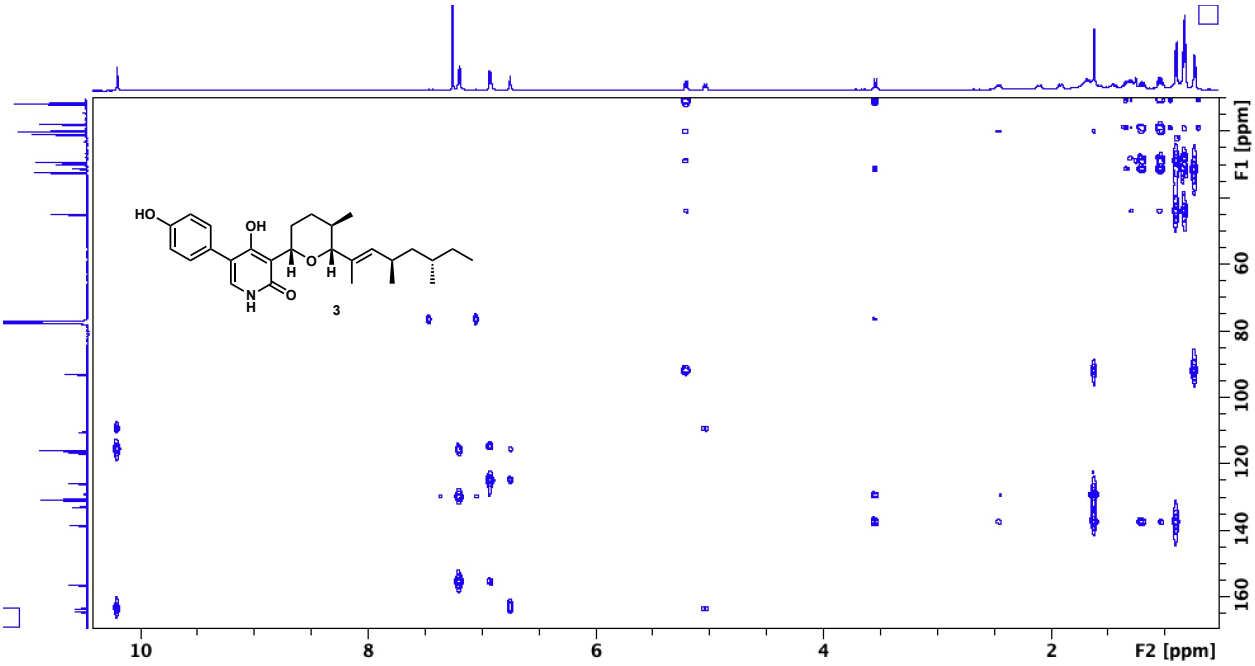


Figure S26. ^{1}H-^{13}C HMBC spectrum of 440thp in $CDCl_3$.

4.12 References

(1) Liu, N.; Hung, Y.-S.; Gao, S.-S.; Hang, L.; Zou, Y.; Chooi, Y.-H.; Tang, Y. *Organic Letters* **2017**, *19*, 3560–3563.

(2) Kim, L. J.; Xue, M.; Li, X.; Xu, Z.; Paulson, E.; Mercado, B.; Nelson, H. M.; Herzon, S. B. *Journal of the American Chemical Society* **2021**, *143*, 6578–6585.

(3) Sheldrick, G. M. *Acta Crystallographica Section A: Foundations of Crystallography* **2008**, *64*, 112–122.

(4) Sheldrick, G. M. *Acta Crystallographica Section C: Structural Chemistry* **2015**, *71*, 3–8.